IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received January 19, 2021, accepted February 1, 2021, date of publication February 22, 2021, date of current version March 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3060863

The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities

M. MAZHAR RATHORE[1], (Member, IEEE), SYED ATTIQUE SHAH[2], DHIRENDRA SHUKLA[3], ELMAHDI BENTAFAT[1], AND SPIRIDON BAKIRAS[1], (Member, IEEE)

[1]Division of Information and Computing Technology, College of Science and Engineering, Hamad Bin Khalifa University, Doha, Qatar
[2]Data Systems Group, Institute of Computer Science, University of Tartu, 51009 Tartu, Estonia
[3]Dr. J. Herbert Smith Centre, University of New Brunswick, Fredericton, NB E3B 5A3, Canada

Corresponding author: Spiridon Bakiras (sbakiras@hbku.edu.qa)

ABSTRACT Digital twinning is one of the top ten technology trends in the last couple of years, due to its high applicability in the industrial sector. The integration of big data analytics and artificial intelligence/machine learning (AI-ML) techniques with digital twinning, further enriches its significance and research potential with new opportunities and unique challenges. To date, a number of scientific models have been designed and implemented related to this evolving topic. However, there is no systematic review of digital twinning, particularly focusing on the role of AI-ML and big data, to guide the academia and industry towards future developments. Therefore, this article emphasizes the role of big data and AI-ML in the creation of digital twins (DTs) or DT-based systems for various industrial applications, by highlighting the current state-of-the-art deployments. We performed a systematic review on top of multidisciplinary electronic bibliographic databases, in addition to existing patents in the field. Also, we identified development-tools that can facilitate various levels of the digital twinning. Further, we designed a big data driven and AI-enriched reference architecture that leads developers to a complete DT-enabled system. Finally, we highlighted the research potential of AI-ML for digital twinning by unveiling challenges and current opportunities.

INDEX TERMS Digital twin, artificial intelligence, machine learning, big data, industry 4.0.

I. INTRODUCTION

Digital twinning is a process that involves the creation of a virtual model (i.e., a twin) of any physical object, in order to streamline, optimize, and maintain the underlying physical process. Theoretically, the digital twin concept was first presented in 2002 by Grieves *et al.* [1] during a special meeting on product life-cycle management at the University of Michigan Lurie Engineering Center. In his subsequent article [2], he further defined digital twinning as a combination of three primary components: 1) a virtual twin; 2) a corresponding physical twin (a physical object that can be a product, a system, a model, or any other component such as, a robot, a car, a power turbine, a human, a hospital, etc.); and 3) a data flow cycle that feeds data from a physical twin to its virtual twin and takes back the information and processes from the virtual twin to the physical twin. The virtual twin is nothing but an algorithm that replicates the behavior (fully or partially) of the corresponding physical counterpart, by generating the same output as does the physical object on given input values. Mostly, it is considered as part of the smart manufacturing process, but it can be used in any domain, such as construction, education, business, transport, power and electronics, human and healthcare, sports, and networking and communications.

Digital twinning was first adopted by Tuegel *et al.* [3] in 2011 to digitally reproduce the structural behavior of an aircraft. Initially, digital twinning was used as a maintenance tool to continuously monitor the craft's structure. Then, it was replicated as a complete twin in order to simulate its entire life-cycle and predict its performance [3]. Later, digital twinning started gaining popularity in several industries that aimed at making their processes smarter, intelligent, and optimally dynamic, based on the operating conditions. The technology raises its global demand, as it facilitates in finding the product flaws, reducing production cost, real-time

The associate editor coordinating the review of this manuscript and approving it for publication was Claudio Zunino.

monitoring of resources, and increasing the life of the product by predicting product failure. On this account, digital twinning became one of the top-ten technology trends [4].

Several surveys have been published, highlighting the current research trends of digital twinning in various fields. For instance, Wanasinghe *et al.* [5] pointed out the state-of-the-art works of digital twinning in the oil and gas industries. Lu *et al.* [6] and Cimino *et al.* [7] reviewed the current reference models, applications, and research issues in manufacturing. Qi and Tao [8] emphasized on the role of data and digital twinning in achieving smart manufacturing. DT-related patents are discussed by Tao *et al.* [9] in different industries. And, the modeling perspective of digital twinning is explored by Rasheed *et al.* [10].

Recently, the use of IoT, big data, and AI-ML technologies have brought new potentials in digital twinning. The adoption of these techniques ensures a perfect digital twin and introduces new research challenges and opportunities. Since 2015, several digital twins have been developed in various industries using AI-ML and big data analytics, and the number of related research articles is growing rapidly. Despite the growing popularity, adaptability, and applicability of AI-enabled digital twinning in the industrial sector, exploited by IoT and big data technologies, no systematic review has been performed that explicitly focuses on the role of these technologies in digital twinning. The above-mentioned surveys do not fully cover the importance of these technologies in the DT domain. Therefore, there is an exigency of a systematic approach towards the thorough review of the current developments in AI-enabled digital twinning using IoT technology and big data. This can drive both academia and industry towards further research, by highlighting the current findings, future potentials, challenges, and applications of AI-enabled digital twinning in the industrial sector.

In this article, we carried out a systematic literature review that incorporates all the research work in the form of articles, patents, and web-reports, covering digital twinning and its integration with state-of-the-art AI-ML and big data analytics techniques. We highlighted the role of big data, AI, machine learning, and IoT technologies in the process of digital twin creation, by listing examples from current deployments in various industrial domains. We introduced the digital twin paradigm, by explaining its basic concepts and highlighting its applications in several industrial areas. After a thorough literature survey, we identified 1) tools that can be used for digital twin creation; 2) the criteria for successful digital twinning; and 3) research opportunities and challenges in digital twinning for diverse industrial sectors. Finally, we designed a reference model for digital twinning that exploits IoT, big data, and AI-ML approaches.

The rest of the paper is organized as follows. Section II briefly presents the survey methodology. Section III formally defines digital twinning, its creation method, and other basic concepts. Section IV summarizes the application of digital twinning in various industries. Section V briefly describes big data and AI, while Section VI discusses the relationship between IoT, big data, AI, and digital twinning. Section VII summarizes the role of AI in digital twinning with state-of-the-art research developments. Section VIII outlines the important data-driven patents in digital twinning. Section IX presents the evaluation criteria for an ideal digital twinning, and Section X lists the tools that may be required in the process of digital twinning. The design details of the reference architecture for AI-enabled DT creation is presented in Section XI, while the current research opportunities and research challenges in digital twinning are described in Section XII. The article is concluded in Section XIII.

II. METHODOLOGY

To the best of our knowledge, the survey at hand is the first of its kind in terms of reviewing AI-ML and big data analytics techniques for digital twinning. The systematic literature review (SLR) carried out in this study is based on the guidelines recommended by [11], [12], with the aim of summarizing the current literature and establishing the basis for qualitative synthesis and information extraction. SLR is an organized, efficient, and widely recognized method that is comparatively better than the traditional literature review process [13].

We identified the following six research questions that directed our entire review process:

1) What is digital twinning, how does it work, and what are the standards and technologies to create a digital twin (DT)?
2) What is the relationship between AI-ML, big data, IoT, and digital twinning?
3) What is the role of AI-ML and big data analytics in digital twinning, its related applications, and current deployments in different industrial sectors?
4) What are the tools required for the creation of AI-enabled DT?
5) What is the criteria for a successful DT or DT-based system?
6) What are the main challenges, market opportunities, and future directions in digital twinning?

To capture the wide range of digital twinning applications, we searched eight multidisciplinary electronic bibliographic databases, including 1) IEEE Xplore (IEEE, IET); 2) ACM digital library; 3) Scopus (ScienceDirect, Elsevier); 4) SpringerLink (Springer); 5) Hindawi; 6) IGI-Global; 7) Taylor & Francis online; and 8) Wiley online library. We also searched the US patents database. Using suitable search strings is crucial to extracting the appropriate literature from the electronic bibliographic databases. Due to the diverse nature of this study, we used a set of appropriate keywords that assures the inclusion of AI-ML and big data analytics in industrial digital twinning. Specifically, as shown in Table 1, we defined various keywords, combined with logical operators, to search the electronic bibliographic databases.

The search was carried out just before August 2020. Prior to 2015, we found very few papers on digital twinning.

TABLE 1. Search strings.

"Digital twin"	**AND** ("Artificial intelligence" **OR** "Big data")
	AND ("Machine learning" **OR** "Deep learning")
	AND ("Industry" **OR** "Production" **OR** "Manufacturing")
	AND ("Automotive industry" **OR** "Transportation")
	AND ("Energy" **OR** "Power sector")
	AND ("Healthcare" **OR** "Fitness")
	AND ("Predictive analysis" **OR** "Maintenance")

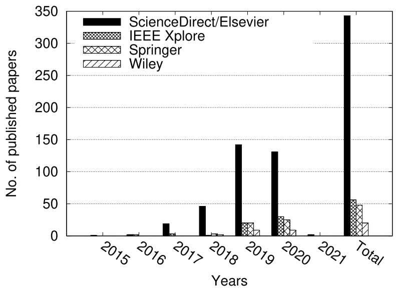

FIGURE 1. Number of journal papers published by different libraries.

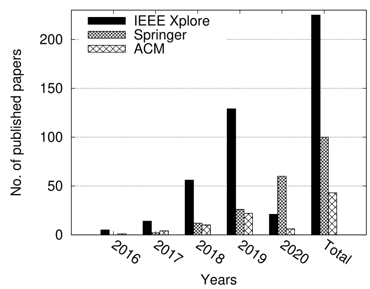

FIGURE 2. Number of conference papers published by different libraries.

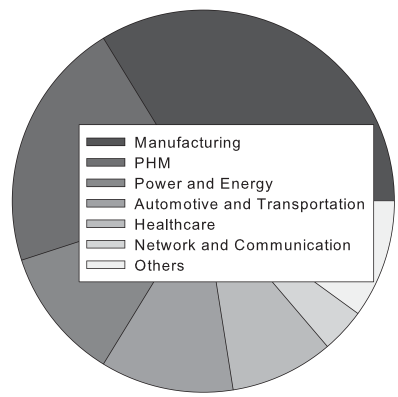

FIGURE 3. AI-ML driven digital twinning research statistics in different fields.

In 2017, the topic gained popularity and became one of the top 10 trends in strategic technology [14]–[16]. In the period 2015–2020, more than 2000 Scopus-indexed journal articles, more than 1000 patents, 250 book chapters, and 20 books have been published, discussing digital twinning technology. However, we identified over 850 articles that match the search criteria defined in Table 1. Fig. 1 and 2 show the total number of journal and conference papers published on the topic of digital twinning by the different libraries. Among other publishers, IGI-Global published seven articles, Hindawi published three articles, and ACM published only two articles in their journals. Additionally, Fig. 3 illustrates the pie chart of published articles related to various applications of DT (it includes both conference and journal papers). Clearly, manufacturing is the dominant application area for digital twinning.

Considering the aforementioned research questions, we defined a set of inclusion and exclusion criteria for an article as follows:

1) The study is written in English.
2) The study is published in a scientific journal, magazine, book, book chapter, conference, or workshop.
3) The journal article is included only if the journal's impact factor is > 1.0.
4) The conference article is included only if the conference is mature enough (it has already published at least fifteen versions of its proceedings).
5) Publications such as dissertations, in-progress research, guest editorials, poster sessions, and blogs are excluded.
6) Duplicate papers that appear in several electronic databases will only be considered once.
7) The study is excluded if not fully focusing on the digital twinning concept or any of its specified applications.

Among the 850 articles that matched the designated keywords, a total of 213 papers were selected after applying the above inclusion and exclusion criteria. IEEE ACCESS and Elsevier Journal of Manufacturing Systems are the top two journals that have published the most articles within the set criteria. The selected publications were first evaluated on the basis of their titles and abstracts. The concept of digital twinning in relation to the research questions was critically examined, and a total of 63 papers were excluded in this phase. Some paper-abstracts were not clear enough to be directly evaluated, hence a full-text screening was performed on 150 papers, resulting in the exclusion of 52 additional papers. Snowball sampling was performed on the remaining set of 98 papers. Then, we used the references and citations of the selected papers to perform backward and forward search, respectively, for identifying new potential papers.

Finally, a total of 117 papers concerning digital twinning, its applications, and related technologies, were selected for data extraction and synthesis of this study. Among the 117 articles, 61 articles discussed AI-ML based digital twins. For each selected article, metadata forms were maintained to categorize the information about the articles and to note the observations assessed. The extracted metadata was then coded for analysis, according to the year of publication, authors' names, affiliated universities or organizations, keywords, name of journal or conference, research model, area of focus, data source, and opportunities/issues highlighted. The categories were derived according to the data needed to answer the research questions and for identifying the paper's main research areas. In addition to journal and conference articles, we included 20 US patents, 15 technical web-reports, and 5 standards, focusing on digital twinning. Some other articles that indirectly relate to digital twinning, such as supporting tools, technologies, and survey methodologies, are also referred in our study.

III. DIGITAL TWIN: INTRODUCTION AND BACKGROUND

A. DIGITAL TWIN: DEFINITIONS AND CONCEPT

Researchers define digital twins in several ways. The pioneers of digital twinning, Grieves and Vickers [17], define a digital twin as "a set of virtual information constructs that fully describes a potential or actual physical manufactured product from the micro atomic level to the macro geometrical level. At its optimum, any information that could be obtained from inspecting a physical manufactured product can be obtained from its Digital Twin." In their opinion, the digital twin can be any of the following three types: 1) digital twin prototype (DTP); 2) digital twin instance (DTI); and 3) digital twin aggregate (DTA). A DTP is a constructed digital model of an object that has not yet been created in the physical world, e.g., 3D modeling of a component. The primary purpose of a DTP is to build an ideal product, covering all the important requirements of the physical world. On the other hand, a DTI is a virtual twin of an already existing object, focusing on only one of its aspects. Finally, a DTA is an aggregate of multiple DTIs that may be an exact digital copy of the physical twin. For example, the digital twins of a spacecraft structure and a spacecraft engine are considered DTIs that may be aggregated into a DTA.

In this article, we assume the concepts of DTI and DTA when referring to a DT. Note that, the majority of academic scholars and industries follow similar definitions for a digital twin. For instance, Glaessgen and Stargel [18] defined it from the perspective of vehicles as "A digital twin is an integrated multiphysics, multiscale, probabilistic simulation of an as-built vehicle or system that uses the best available physical models, sensor updates, fleet history, etc., to mirror the life of its corresponding flying twin." Similarly, Tao *et al.* [19] considered the aspect of product life cycle and interpreted the digital twin as "a real mapping of all components in the product life cycle using physical data, virtual data and interaction data between them." Söderberg *et al.* [20] focused on the application of optimization while defining a digital twin. According to them, digital twinning is an approach to perform a real-time optimization to a physical system using its digital copy. Finally, Bacchiega [21] made it simpler by defining it as "a real-time digital replica of a physical device."

With our understanding, shown in Fig. 4, digital twinning is a process that involves the construction of 1) a cyber twin that digitally projects a living or non-living physical entity or a process (a system); and 2) a physical connection between cyber and physical twins to share data (and information) between them aimed at dynamic optimization, real-time monitoring, fault diagnostics and early prediction, or health monitoring of the physical counterpart. A physical twin can be a process, a human, a place, a device, or any other object with a special purpose, and which is able to be replicated in the digital world as either a partial twin with limited functionalities, or a complete twin that incorporates the full behavior of its physical peer. Digital twinning is mostly employed in industries for physical objects in their units. However, there exist some digital twins that are mirrors of processes in the physical world, such as digital twins of a mobile-edge computing (MEC) system [22], human protein–protein interaction (PPI) [23], supply chain [24], components-assembly at a manufacturing unit, and job scheduling [25].

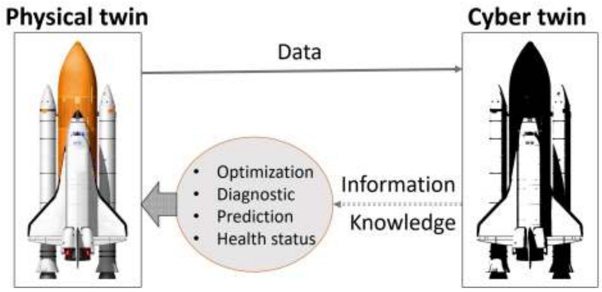

FIGURE 4. Digital twinning concept.

B. A BRIEF HISTORY OF DIGITAL TWINS

The idea of creating a digital copy of a physical entity was introduced in the early 2000s. However, the term "digital twin" originated around ten years ago. Michael Grieves, in one of his articles [2], claimed that the concept of digital twins was first presented during a lecture on product life-cycle management (PLM) in 2003. Whereas, in his other book chapter [1], he stated that the concept was originally proposed, without a name, in 2002 while presenting a paper in a special meeting at the University of Michigan Lurie Engineering Center. Grieves mentioned in this book chapter, "While the name has changed over time, the concept and model has remained the same." He added that it was given the name "mirrored spaces model (MSM)" in 2005 and changed to "information mirroring model" in 2006. NASA started using this concept of virtual and physical models in their technology roadmaps [26] and proposals for sustainable space exploration [27] since 2010. However, the name "digital twin" was first coined in 2011 by John Vickers of NASA. Practically, the first digital twin was developed by Tuegel *et al.* [3] for the next-generation fighter aircraft, in order to predict its structural life.

C. OPERATIONAL MECHANISM

Although the digital twin concept was introduced in 2002, it became a popular trend due to the advancement in sensor technology and IoT, which play a vital role in digital twinning by collecting real-time data from the physical world and sharing it with the digital world. The twinning can be viewed as a bridge between a physical twin and the corresponding virtual twin. The physical-to-virtual connection is established with a technology that allows the transfer of information from the physical environment to its virtual twin, including web services, cellular technology, WiFi, etc. The virtual twin is adjusted gradually with the functioning of the physical twin by continuously collecting the differences between the two environments. These connections allow the monitoring of responses to both conditions and interventions. The conditions mainly occur in the physical environment, whereas the interventions take place within the virtual twin. Thus, a digital twin holds a real-time status of the physical counterpart.

The virtual-to-physical connections represent the information circulating from the virtual to the physical environment. This information may change the state of the physical twin by displaying some data or changing the system's parameters (for optimization, diagnostics, or prognostics). Although virtual-to-physical connections are very helpful in DT modeling, they are not always included in the description. Instead, it is common to consider a one-way connection, i.e., physical-to-virtual. Finally, the data and the information from both physical and virtual worlds are stored and analyzed at a centralized server—or a cloud computing platform—where the final decisions related to optimization, diagnostics, or prognostics, are made.

D. DIGITAL TWIN STANDARDS

Currently, there is no particular standard that solely focuses on the technical aspects of digital twinning. Standardization efforts are under-development by the joint advisory group (JAG) of ISO and IEC on emerging technologies [28]. However, the ISO standard ISO/DIS 23247-1 [29] is the only standard that offers limited information on digital twins. In addition, there are other related standards that may facilitate DT creation. For example, the ISO 10303 STEP standard [30], the ISO 13399 standard [31], and the OPC unified architecture (OPC UA) [32] technically describe ways to share data between systems in a manufacturing environment.

IV. DIGITAL TWINNING IN INDUSTRIES: APPLICATIONS

Digital twinning is becoming apparent in various industries, including manufacturing, medical, transportation, business, education, and many more. In this section, we present the role of digital twinning and the current research followed in these areas.

A. MANUFACTURING

Digital twinning is conceived as a major tool in the manufacturing industry to carry out smart manufacturing, fault diagnosis, robotic assembly, quality monitoring, job shop scheduling, and meticulousness management. In this way, Rosen *et al.* [33] emphasizes the use of digital twinning in manufacturing. Modules in a computerized system communicate with each other during every step of the production, thus depicting a realistic model of its physical counterpart. Similarly, the work by Qi and Tao [8] explains the benefits of big data-driven DT in smart manufacturing. The DT combines all the manufacturing processes, starting from product design to maintenance and repair. The virtual model is capable of identifying the constraints of the virtual design in the physical world, which are iteratively improved by the designers. Data produced by sensors and IoT devices are then analyzed and processed using big data analytics and AI applications to enable the manufacturers to select a satisfactory plan.

On the other hand, DT is also used to monitor a component or a product, considering its usage, health, and performance during the life-cycle of manufacturing. Real-time data provided to the virtual model allows it to self-update and predict any abnormal behaviors. Optimal solutions are developed for problems found in the virtual models, and the actual manufacturing model is adjusted accordingly. Maintenance and repair of the physical system can also be scheduled timely, based on the predictions of the virtual models. One of such digital twin projects is originated by Slovak University of Technology in Bratislava [34] for a physical production line of pneumatic cylinders, where they defined the continuous optimization of production processes and performed proactive maintenance, based on the real-time monitoring data. Similarly, a digital twin of manufacturing execution system (MES) was developed by Negri *et al.* [35] that enables the supervisory control over the physical MES system using sensor technology, by allowing the multi-directional communication between digital and physical sides of manufacturing assets.

Several state-of-the-art works highlight that DTs should be capable of self-healing and predictions. These predictions play a vital role in an important aspect of smart manufacturing, i.e., fault diagnosis, since a minor issue during production can cause irreparable damages. A variety of technologies used in fault diagnosis like Support Vector Machines [36], Bayesian Networks [37], Deep Learning [38]–[40], and many others [41]–[44] are capable of enhanced fault diagnosis. However, Xu *et al.* [45] highlight that, in production systems, conditions are constantly changing. Therefore, the same training model cannot be applied throughout the process, but creating a new model requires a lot of time and resources. As such, they proposed a digital twin-assisted fault diagnosis using deep transfer learning (DFDD) approach. DFDD has been applied to fault diagnosis in smart and complex manufacturing. The framework involves two phases. In the first phase, the virtual model of the system is constructed. Repeated designs of the model are tested and evaluated in the virtual space until all anomalies are discovered. Simulation data during design testing is provided to an embedded fault diagnosis model in the virtual space. The diagnosis model keeps learning from the simulation data using Deep Neural Networks, in order to increase its efficiency for failure prediction during the start of the production phase when there is insufficient training data. The second phase starts when the virtual model achieves acceptable performance. The physical entity is constructed and linked to its corresponding virtual model. Data is transferred from a physical entity to the virtual model through sensors during production. A diagnosis model is formed and updated using the current data from the physical entity and the knowledge learned from the previous phase, which is transferred using deep transfer learning (DTL).

Robotic assembly, in industrial manufacturing, is responsible for handling a notable amount of work [46]. It is involved in packaging, labeling, painting, welding, and many others. With the advancements in the complexity of manufacturing, these robotic assemblies have become more error-prone. The concept of DT is being utilized in this area to monitor and optimize the assembly process. In [47], a multisource model-driven digital twin system (MSDTS) is designed for robotic assembly. The MSDTS model consists of three parts. The physical space consists of sensors, its associated data, and the robotic arm for moving and gripping objects. The virtual space consists of a server, a multisource model, and a virtual reality display and control (VRDC). A communication interface offers the exchange of data between two spaces in real-time. Initially, a 3D model of the entire physical space is constructed using a depth sensor that is mounted on the robot arm. During operation, the VRDC provides a complete view of the physical system by receiving a video stream from an RGB camera. When the robot arm moves, angular data is sent to the virtual twin through the communication interface in real-time, and the graphical model in the virtual system follows the same trajectory. The physical contact of the robot arm with the surrounding object is simulated in the virtual system using the Kelvin-Voigt model (KVM), where parameters of the model are estimated through the data of contact force and relative motion of contact point. A surface-based deformation algorithm is used to simulate the deformation of an object using the data generated by KVM. The results of the models are rendered in the VRDC. A complete view of the system is provided to the operator via a head mount. Interaction with the physical space is done using a control handle.

Another important element in manufacturing is job shop scheduling, which makes efficient use of resources to reduce production time and maximize production efficiency. In real-life situations, due to errors and anomalies, the scheduling process can be rendered inefficient. With the introduction of smart manufacturing and digital twins, new DT-based job shop scheduling methods are introduced to overcome scheduling plan deviation and provide a timely response. One such model is proposed in [48]. A DT-based job shop consists of a physical and a virtual space, which communicate through a CPS. Scheduling data from the physical space is sent to the virtual space, and multiple scheduling strategies are simulated and retrieved from the virtual models. The finalized scheduling plan is fed into the physical space. Since a physical system has many modules, the plan is divided and categorized based on the respective modules. Continuous communication between the physical and virtual space results in achieving precise scheduling parameters, as well as prediction of any disturbances in the schedule. The scheduling plan can hence be updated and fed to the physical system for increased efficiency and timely response.

Digital twin and big data are playing an important role in smart manufacturing starting from product life-cycle to maintenance and repair. Some of the stated research articles highlighted the importance of digital twinning in the areas of smart manufacturing. The concept of utilizing a variety of data and integrating it with IoT, virtual reality, and data analytics, results in high fidelity monitoring, timely prediction and diagnosis of faults in assembly or production, and overall optimization and improvement of the manufacturing process.

B. MEDICAL

Applications of DT in medical include the maintenance of medical devices and their performance optimization. DT, along with AI applications, are also used to optimize the life-cycle of hospitals by transforming a large amount of patient data into useful information. The ultimate aim of the digital twinning in healthcare is to help authorities in managing and coordinating patients. Mater private hospitals in Dublin (for cardiology and radiology) were facing problems regarding increased services, patient demand, deteriorating equipment, deficiency of beds, increased waiting time, and queues. These problems indicated the call for the improvement in the current infrastructure to cater to increasing needs. Mater private hospitals (MPH) partnered with Siemens Healthineers to develop an AI-based virtual model of their radiology department and its operations [49]. As a result, the simulations of the model provided insights towards the optimization of workflows and layouts. The realistic 3D models of the radiology department, provided by DT techniques, allowed for the prediction of operational scenarios and the evaluation of the best possible alternatives to transform care delivery.

In recent years, with the introduction of "precision medicine," the focus of DT technology is shifted towards a human DT. Precision medicine is the branch of healthcare that promotes tailored treatments on an individual level. The human DT would be linked to its physical twin and would display the processes inside the human body. It can result in an easier and accurate prediction of illness with proper context, and bring a paradigm shift in the way patients are treated. Virtual physiological human (VPH) was the earliest human DT that was developed [50]. VPHs would act as a "Virtual Human Laboratory" where each VPH was modified based on the specific patient, and different treatments would be tested on the modified VPH platform.

Apart from human DTs, organs or human body parts digital twins have also been developed. Data from Fitbit devices, smartphones, and IoT devices are sent in real-time to such DTs, in order to provide constant feedback regarding human organ activity. Some organs' DTs have been used by experts to perform clinical analysis, whereas many others are under development. In a study, a 3D digital twin of a heart was developed by Siemens Healthineers [51], after performing a comprehensive research on approximately 250 million images, functional reports, and data. The model exhibited the physical and electrical structure of a human heart. This

DT is currently under research at the Heidelberg university hospital (HUH), Germany, where DTs of 100 patients have been created, who had a history of heart diseases within a period of six years. Simulations of these DTs were compared with the ground truth, which provided promising results.

Another DT of the heart has been developed by researchers at the Multimedia Communications Research Lab in Ottawa, Canada. It is called a Cardio Twin and targets the detection of ischemic heart disease (IHD) [52]. IHD is a condition characterized by reduced blood flow to the heart, which can lead to chest pain or mortality in case of delayed treatment. The researchers developed the DT on the concept of edge computing/analytics, where the time is considered very critical. Data is collected from social networks, sensors, and medical records. The accumulated data is fed to an AI-inference engine, where data fusion, formatting, and analytics are performed using TensorFlow Lite to discover new information. The Cardio Twin can communicate with the physical twin in the real world, using a multimodal interaction component that employs WiFi/4G or Bluetooth communication. Cardio Twin performed a sample classification of 13420 ECG segments with an accuracy of 85.77%, in a short span of 4.84 seconds. However, no method to evaluate Cardio Twin in the real world has been introduced.

Sim&Cure, a company based in Montpellier, France, developed a simulation model for the treatment of aneurysm. Aneurysm is an outward bulging of blood vessels, typically caused by an abnormally weakened vessel wall. A serious case of aneurysm can result in clotting, strokes, or death. The last option for treating aneurysm is surgery. However, endovascular repair (EVAR) is generally used, since it is less invasive and low-risk. In EVAR, a stent-graft/catheter is placed into the affected area to minimize the pressure. In many cases, choosing the stent-graft/catheter is difficult and depends on the size of the blood vessels. The Sim&Cure's DT helps surgeons in selecting an ideal implant to cater to the size of the aneurysm as well as the blood vessels. A 3D model of the affected area and surrounding vessels is created, and multiple simulations are run on the personalized DT, which allows surgeons to have a better picture. Promising results have been presented in preliminary trials [53], [54].

Researchers at the Oklahoma State University developed a human airway DT—named "virtual human"—in their computational biofluidics and biomechanics laboratory (CBBL). They tracked the flow of air particles in aerosol-delivered chemotherapy and found out that, the aerosol-based drug hit the cancerous cells with less than 25% accuracy [55]. This caused more harm than benefits to patients, as the remaining drug would fall on healthy tissue. The version 1.0 of "virtual human" was based on a 47-year-old standing male, containing the entire respiratory system. V1.0 also allowed patient-specific structural modifications, e.g., creating a respiratory system of a standing/seated female or a kid with/without respiratory conditions. Following the success of V1.0, CBBL researchers developed its successor version 2.0. The V2.0 was patient-specific, and was created by performing an MRI/CT scan of the patient. The scanned data was used to construct a 3D model of the lungs. The researchers at CBBL then created a virtual population group (VPG), which was a large group of human DTs. The VPG exhibited trends within different groups/sub-groups. Simulations to analyze the trends of aerosol particle movement were conducted on the VPG, by varying the particle sizes, inhalation rate, and initial position of the medication. These simulations indicated that the drug's effectiveness would increase to 90% if the drug delivery method was personalized to each patient, rather than distributing the drug evenly for every patient [55].

In another study, Liu *et al.* [56] proposed a cloud-based DT healthcare solution (CloudDTH) for elderly people. The cloud-based solution provides a fusion of physical and virtual systems to address real-time interaction between patients and medical institutions, and personalized healthcare for the entire life-cycle of the elderly. CloudDTH has a layered architecture, providing health resources, identification of medical personnel, user interface, virtualization, and security services to users. CloudDTH obtains real-time data from sensors for ECG, BP, pulse rate, and body temperature. These sensors are already implemented in the CloudDTH framework. The sensor data are then transmitted to the cloud server, using TCP. In case of an incident, such as patient falling, heart attack, stroke, etc., the monitoring model, after performing analysis on the received data, sends a high-frequency and multi-attribute monitoring order of the patient to medical personnel. A case study was performed by researchers, where data from two patients with normal and abnormal heart rates was input to the system. The simulation results indicated symptoms of arrhythmia in one patient, and recommended the dosage of medication based on their physical conditions. The CloudDTH framework simulations also provided a feasible scheduling mechanism for elderly patients in hospitals, in order to avoid long queues.

C. TRANSPORTATION

Numerous innovative technologies have been brought forward with the development of IoT, including digital twins, autonomous things, immersive technology, etc. Various types of digital twins are developed in transportation sector, including DTs for automobile components, vehicles, vehicular networks, and road infrastructures. However, the purpose remains the same i.e., monitoring, optimization, and prognostics and health management. For example, Wang *et al.* [57] developed a framework for connected vehicles based on digital twins. The framework used vehicle-to-cloud (V2C) communication to provide advisory speed assistance (ADSA) to the driver. Real-time data from sensors was obtained in the physical system, which was sent to the cloud through the V2C module. All processing of the data from V2C was performed on the cloud server. The computed results were sent back to the physical system and served as a guidance system for components within the physical world. The authors demonstrated the effectiveness of their framework with a case study of cooperative ramp merging involving three passenger vehicles, and the results showed that the digital twin can indeed assist transportation systems.

Cioroaica *et al.* [58] worked on the context of connected vehicles in smart ecosystems. The establishment and achievement of goals in smart ecosystems are possible when smart entities within the ecosystem co-operate with each other. This is achieved when the systems have a level of trust. The authors developed a virtual hardware-in-the-loop (vHiL) testbed model to evaluate the trust-building capability of smart systems within an ecosystem. A smart agent, capable of interacting with the vehicle's electronic control unit (ECU), is installed at the vehicle along with its corresponding DT. In Phase 1, the trustworthiness of the smart agent is evaluated by simulation in its corresponding virtual twin. Phase 2 involves trust-building, where the smart agent is executed on the ECU. Evaluation of simulated and actual results identifies the obstacles. These obstacles are overcome by the collaboration of virtual and physical entities to achieve trustworthiness in a smart ecosystem.

Chen *et al.* [59] studied the use of unmanned aerial vehicles (UAVs) as complementary computation resources in a mobile edge computational (MEC) environment for mobile users (MU). MEC provides computational capabilities to MUs within a radio access network (RAN). Mobile users send computational tasks to UAVs by creating the corresponding VMs. The tasks arriving at the UAVs are stored in queues and, due to limited resources, the MUs have to compete for them. The authors proposed deep reinforcement learning (DRL) techniques for the scheduling of tasks on the UAV, and for minimizing the response delay from the UAV to the MUs. The training of the DRL network in an offline manner is achieved by creating a digital twin of the entire MEC system. Simulations with varying parameters were conducted and the best results were selected. The results of the DRL scheme trained on digital twins ensured significant performance gains when compared to other baseline approaches.

Digital twins have also been utilized in transportation systems for traffic congestion management, congestion prediction, and avoidance. Kumar *et al.* [60] worked on intelligent transport systems, leveraging technologies such as fog/edge analytics, digital twins, machine learning, data lakes, and blockchain. The authors captured situational information from cameras, and performed edge analytics on the acquired data. An entire virtual vehicle model was created via a digital twin to replicate the real-world scenario. Driver intentions were predicted using machine and deep learning algorithms to avoid traffic congestion. This virtual vehicle model allowed autonomous vehicles to make decisions regarding optimal paths, but also helped drivers of non-autonomous vehicles to make better decisions based on the traffic situation and the mined driver intentions.

Digital twins have also been used for the maintenance of different systems. The work implemented by Venkatesan *et al.* [61] monitored and projected the health conditions of electric motor vehicles using an intelligent digital twin (i-DT). The framework tracked the health of the electric motor in an electric vehicle using fuzzy logic and artificial neural networks (ANNs). The average speed of the vehicle and the duration of travel was fed into the ANN i-DT and fuzzy logic i-DT for training purposes. In addition, simulations carried out on a digital twin tested the performance of the entire framework. Parameters such as winding and casing temperature, deterioration in magnetic flux, and lubricant refill time were set for the digital twin. The comparison of theoretical and i-DT computations indicated that an i-DT can effectively be used in electric vehicles to foresee their health.

D. EDUCATION

Another important area where digital twins can play a crucial role, is education. Digital twins of physical entities such as labs, construction, mechanical equipment, can be created and provided to students for online learning. However, there has not been a lot of research effort on the use of DT in the education domain. One such work was performed by Sepasgozar [62] that used digital twins and virtual gaming for online education. The authors created a digital twin of an excavator along with a virtual game for the course of construction management and engineering. The project contained four modules named 1) group wiki project and role play (GWiP); 2) interactive construction tour 360 (ICRT 360); 3) virtual tunnel boring machine (VTBM); and 4) piling augmented reality and digital twin (PAR-DT). GWiP was used for doing group projects online. ICRT 360 consisted of recorded videos to provide details on construction sites and machinery. VTBM was a virtual game-based environment that helped students to learn about the working of a tunnel boring machine. Virtual equipment was introduced in the game, where a student or a group of students could explore their interests. PAR was developed for smartphones and Oculus headsets to provide students an augmented environment to collaborate and understand the importance of piling in construction. The final module involved a digital twin of an excavator, which was also linked to a physical instance. The DT provided hands-on learning about the functions and movements of an excavator. The students' feedback emphasized the importance of an immersive environment in online education.

E. BUSINESS

Business is also one of the areas where DT is playing an important role. According to PropTechNL [63], the real estate sector is fragmented in terms of architects, installation, construction, transport, and management. This fragmentation results in an inefficient system that has a negative impact on people living in a society. Digital twins can provide huge opportunities in real state, and facilitate the creation of smart societies. For example, a wide range of sensors can collect data, and the performance of a building can be measured and improved. Digital twins in real estate may add significant value by re-positioning buildings to the needs and requirements of customers, hence improving the customer experience. The design of buildings, the usage, effectiveness, and strength of raw materials, as well as maintenance and running costs, can be managed through digital twins. Thus, it provides a cost-effective, fast, and smart way of developing a real estate. For instance, an American multinational company, GE Healthcare, has incorporated the use of DT to redesign its systems, in order to run new hospitals more efficiently.

Kampker *et al.* [64] introduced a framework for the development of successful business models in smart services. The scenario of crop (potato) harvesting was taken into consideration during their research. In traditional harvesting mechanisms, the harvesting machines are set up based on historical data and the experiences of individual operators. However, the lack of standard procedures may cause damage to the crop. Therefore, the authors developed a framework, based on a digital twin, to reduce the damage to the crop during harvesting. Specifically, a digital twin is set up near the physical field. The virtual model then passes through the same stages as the real crop. During the simulation, the condition of the neighboring crop is analyzed for potential damage. The results of the analysis lead to adjusting the parameters, and repeated simulations continue until the optimal settings are found. Tests carried out by the authors indicated that more damage to the crop is caused by its impact on multiple conveyor belts during the transition. Hence, adjustment to the height and position of conveyor belts can reduce the risk of damage. This framework can also tweak the settings of autonomous harvesting machines, apart from providing recommendations to operators.

F. OTHER INDUSTRIES

Digital twinning can be a part of smart construction, where a DT may be designed for buildings, roads, or any other infrastructure development. For example, a virtual twin was developed for office buildings [65] that manages the building's life-cycle, by collecting data through sensors. Furthermore, DT technology may advance the disaster management approaches in smart cities [66]. Possibly, the technology also has a potential to protect industrial control systems and data from cyber attacks. On this account, Dietz and Pernul [67] proposed the use of digital twinning technology to identify security threats that target industrial control systems (ICSs), and rectify their effects. Theoretically, they focused on the Stuxnet worm [68] that compromised the speed of centrifuge, and Triton [69] that digitally invaded a petrochemical plant in Saudi Arabia. Deitz *et al.*, indicated in the Stuxnet example that the outliers in the historical network traffic would have detected a threat. Similarly, in the case of simulations, the deviation of network traffic between the virtual and physical systems would have identified the attack.

V. AI-ML AND BIG DATA: AN INTRODUCTION

Big data remains one of the top research trends from last few years. It is different from an ordinary data because of its high volume, high velocity, and heterogeneous variety, as interpreted in Fig. 5. Researchers named these characteristics as "the 3Vs of big data," i.e., volume, velocity, and variety. Later, two more Vs—value and veracity—were added to the list. Thus, we refer to any data as big data, if it is of significant size (volume), it is being produced at very high-speed (velocity), and it is heterogeneous with structured, semi-structured, or unstructured nature (variety). The worth of big data analytics brings the fourth V (i.e., value) into its characteristics, thus making it an asset to the organization. Big data analytics is a process that analyzes big data and converts it to valuable information, using state-of-the-art mathematical, statistical, probabilistic, or artificial intelligence models. However, the 3Vs of big data lead us to a new world of challenges, including capturing, storing, sharing, managing, processing, analyzing, and visualizing such high-volume, high-velocity, and diverse variety of data. To this end, various frameworks [70]–[73] have been designed to handle big data for effective analytics in different applications.

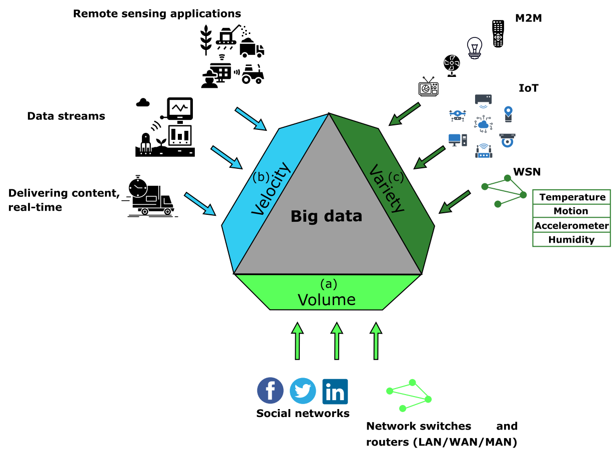

FIGURE 5. Big data definition.

Artificial intelligence (AI) is the digital replication of three human cognitive skills: learning, reasoning, and self-correction. Digital learning is a collection of rules, implemented as a computer algorithm, which converts the real historical data into actionable information. Digital reasoning focuses on choosing the right rules to reach a desired goal. Whereas, digital self-correction is the iterative process of adopting the outcomes of learning and reasoning. Every AI model follows this process to build a smart system, which performs a task that normally requires human intelligence. Most of the AI systems are driven by machine learning, deep learning, data mining, or rule-based algorithms, where others follow logic-based and knowledge-based methods. Nowadays, machine learning and deep learning are widely used AI approaches.

It is often confusing to differentiate between artificial intelligence, machine learning, and deep learning techniques. Machine learning is an AI method, which searches for particular patterns in historical data to facilitate decision-making. The more data we collect, the more accurate is the learning process (reflects the value of big data). Machine learning can be 1) supervised learning, which accepts data sets with labeled outputs in order to train a model for classification or future predictions; 2) unsupervised learning, which works on unlabeled data sets and is used for clustering or grouping; and 3) reinforcement learning, which accepts data records with no labels but, after performing certain actions, it provides feedback to the AI system. Examples of supervised learning techniques are regression, decision trees, support vector machines (SVMs), naive Bayes classifiers, and random forests. Similarly, K-means and hierarchical clustering, as well as mixture models, are examples of unsupervised learning. Finally, Monte Carlo learning and Q-learning fall under the reinforcement learning category. On the other hand, deep learning is a machine learning technique that is motivated by biological neural networks with one or more hidden layers of digital neurons. During the learning process, the historical data are processed iteratively by different layers, making connections, and constantly weighing the neuron inputs for optimal results. In this article, we mainly focus on digital twin systems based on machine learning.

VI. RELATIONSHIP BETWEEN IoT, BIG DATA, AI-ML, AND DIGITAL TWINS

The emerging sensor technologies and IoT deployments in industrial environments have paved the way for several interesting applications, such as real-time monitoring of physical devices [74], indoor asset tracking [75], and outdoor asset tracking [76]. IoT devices facilitate the real-time data collection—that is necessary for the creation of a digital twin of the physical component [77], [78]—and enable the optimization [79] and maintenance [80] of the physical component by linking the physical environment to its virtual image (using sensors and actuators). Note that, the above-mentioned IoT data is big in nature [81] (as explained in Section V), so the big data analytics can play a key role in the development of a successful digital twin. The reason is that industrial processes are very complex, and identifying potential issues in early stages is cumbersome, if we use traditional techniques. On the other hand, such issues can easily be extracted from the collected data, which brings efficiency and intelligence into the industrial processes. However, handling this enormous amount of data in the industrial and DT domains requires advanced techniques, architectures, frameworks, tools, and algorithms. For instance, Zhang et al. [82], [83] proposed a big data processing framework for smart manufacturing and maintenance in a DT environment.

Oftentimes, cloud computing is the best platform for processing and analyzing big data [84]. Additionally, an intelligent DT system can only be developed by applying advanced AI techniques on the collected data. To this end, intelligence is achieved by allowing the DT to detect (e.g., best process strategy, best resource allocation, safety detection, fault detection) [85], predict (e.g., health status and early maintenance) [80], [86], optimize (e.g., planning, process control, scheduler, assembly line) [87], [88], and take decisions dynamically based on physical sensor data and/or virtual twin data. In short, IoT is used to harvest big data from the physical

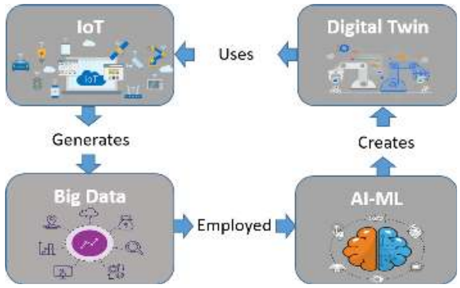

FIGURE 6. Relationship between IoT, big data, AI-ML, and digital twins.

environment. Later, the data is fed to an AI model for the creation of a digital twin. Then, the developed DT can be employed to optimize other processes in the industry. The overall relationship among IoT, big data, AI, and digital twins is presented in Fig. 6.

VII. CURRENT DEPLOYMENTS OF DIGITAL TWINS USING BIG DATA AND MACHINE LEARNING

We have identified the primary sectors where DT-based systems are developed with the help of AI-ML techniques. In the following sections, we discuss the current deployments in these sectors, including smart manufacturing, prognostics and health management (PHM), power and energy, automotive and transport, healthcare, communication and networks, smart cities, and others.

A. SMART MANUFACTURING

Smart manufacturing involves 1) the acquisition of data from manufacturing cells through a variety of sensors; 2) the management of the acquired data; and 3) the data exchange between different devices and servers. In a DT environment, the data is collected from a physical manufacturing cell and/or its corresponding virtual cell. Such data can be further utilized for manufacturing process optimization, efficient assembly line, fault diagnosis, etc., using AI approaches. The AI-ML based digital twinning process for smart manufacturing is depicted in Fig. 7.

Manufacturing is the top industry where most digital twins are being developed. Xia et al. [91] proposed a manufacturing cell digital twin to optimize the dynamic scheduler for smart manufacturing. An intelligent scheduler agent, called digital engine, was developed and trained for optimization using deep reinforcement learning algorithms (DRLs), such as natural deep Q-learning [101], double deep Q-learning [102], and prioritized experience replay (PER) [103]. The underlying features were captured from both the physical and virtual environments of the cell by an open platform communications (OPC) server. The training of the DRL network was done through a gradient descent process, which requires finite learning iterations and is sufficiently intelligent, reliable, and robust. The developed DT-based dynamic scheduler optimizes the manufacturing process by accelerating the training, testing, and validation of smart control systems. The system was tested on a robot cell to optimally select the strategy

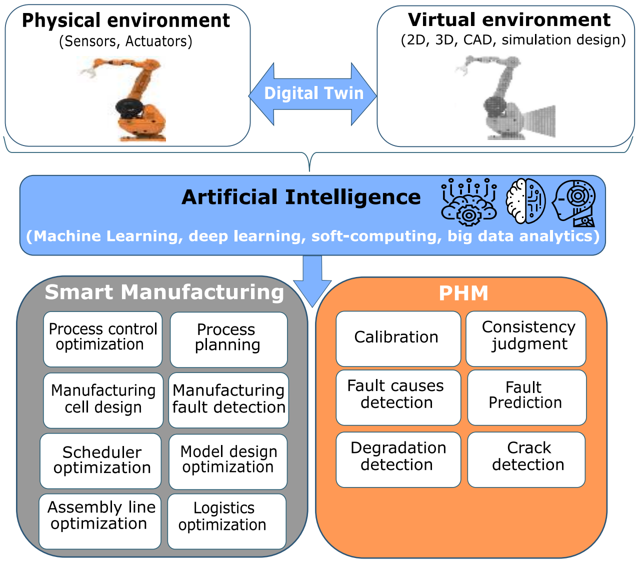

FIGURE 7. DT-based smart manufacturing using big data analytics and AI-ML.

for performing the lower level tasks that are necessary to accomplish the higher level manufacturing goal.

Zhou *et al.* [79] performed a geometric optimization of centrifugal impeller (CI) by collecting features, such as meridional section (MS), straight generatrix vectors (SGV), and set of streamlines (SSL), from both the physical and CAD-based digital model of the CI. However, with the improvement in machinability, the DT-based geometric optimization reduces the aerodynamic performance. Thus, the best model for the CI is selected by training the deep deterministic policy gradient (DDPG) reinforcement learning model [104] to iteratively select the fair geometry of the CI-design with optimum values of machinability and aerodynamic performance. For the DDPG algorithm, they used two actor networks (online and target network) as the strategy function π to control the agent-actions, and two critic networks (online and target network) to evaluate these actions and give rewards. The proposed DT-based optimization sped up significantly the design and manufacturing of the impeller. Similarly, Zhang *et al.* [95] also developed an impeller DT, but for the purpose of manufacturing process planning. They employed a knowledge reuse deep learning network (PKR-Net) [105], which takes data from dynamic knowledge base, views from 3D computer-aided impeller design (CAD), 2D drawings, and process knowledge. The objective is to optimize the theoretical processes and generate the best process plan for product manufacturing, by considering both manufacturing time and monetary cost.

Furthermore, Lee *et al.* [106] designed a deep learning and cyber-physical system based digital twinning (DTDL-CPS) architecture for smart manufacturing, that can be used in shop floor optimization, fault diagnosis, product design optimization, and predictive maintenance. BDHDTPREMfg [84] is a similar CPS-based big data-driven model for DT-enabled re-manufacturing. Several other digital twins have been developed in the manufacturing industry using AI approaches that could not be fully discussed in this article. Rather, Table 2 summarizes these digital twins with respect to the problem they solved (i.e., the application), the ML-approach they used to solve the problem, and the DT use-case they developed.

B. PROGNOSTICS AND HEALTH MANAGEMENT

The persistent use of a product degrades its performance over time, which may lead to malfunctioning. Thus, prognostics and health management (PHM) is very crucial in all industries. PHM process involves the prediction of the remaining useful life of a product and the consistent monitoring of its health. This is the second most important application of DT, following smart manufacturing. Note that, several alternative terms, such as "predictive modeling" [86], "structural life prediction" [3], "remaining useful life", and "predict and act" [107] have also been used in place of PHM. DT-based PHM regularly monitors the physical equipment based on the data generated by the equipment-sensors, performs diagnosis and prognosis operations on the data using big data analytics and AI, and recommends design rules for immediate maintenance. The process of DT-enabled PHM is depicted in Fig. 7.

Tao *et al.* [108] developed a digital twin for a wind turbine in a power plant, in order to monitor its health by performing gearbox prognosis and fault detection. The proposed DT-driven PHM can be applied to any complex equipment in harsh environments, such as aircraft, ships, and wind turbines. The wind turbine DT is built based on various geometry levels, physics, behavior, and rules. The DT can detect the disturbances in the turbine environment, as well as potential faults in itself and its designed model. The data is collected from the DT model (both physical and digital) and is matched against the thresholds for degradation detection. In addition, past DT-data is used to train a single hidden layer neural network for better prediction of gradual faults and detection of its causes, using extreme learning machine (ELM) [109]. The abrupt fault in the turbine is detected by comparing the data from the physical and virtual environments. Similarly, to improve ship efficiency and avoid unnecessary maintenance operations, a data-driven ship digital twin was developed by Coraddu *et al.* [110]. Their goal was to determine the speed loss due to marine fouling. Multilayered-deep extreme learning (DELM) [111] predicts the ship's speed, based on the features collected from on-board sensors, such as designed and ground speed, draft, engine and shaft generator power, wind speed, temperature, fuel consumption, etc. The expected ship speed is compared with the measured speed to compute the speed loss. Finally, robust linear regression is applied to the speed loss information to determine whether the speed loss is due to marine fouling.

Numerous other digital twins have been developed for PHM of industrial components, including photovoltaic energy conversion unit [112], battery system [113], vehicle motor [61], UAV [115], spacecraft [116],

TABLE 2. State-of-the-art AI-ML developments in digital twinning for smart manufacturing.

Paper	Application	AI-ML approach	DT use-case / Impl. environment
[89]	Product quality	Artificial neural network (ANN), stacked auto encoder (SAE)	CNC bending machine
[87]	Dynamic scheduling optimization, machine availability prediction, disturbance detection	Multi-layer neural network, heuristic algorithm (genetic algorithm), data fusion	Job shop-milling machine
[90]	Robot optimization (avoiding obstacles)	Ant colony optimization	Robots
[91]	Process control, assembly line, scheduling optimization	Deep reinforcement learning (deep Q-learning, double deep Q-learning, and prioritized experience replay (PER))	Robots manufacturing cell
[84]	AGV's pallet optimization design and multi-life-cycle process forecast, AGV fault diagnosis	Deep learning	Automated guided vehicle (AGV)
[45]	Development fault diagnosis	Deep neural network (transfer learning, DTL)	Shop floor
[92]	Product design, strategies, and process optimization	Machine learning, deep learning	Shop floor
[93]	Job scheduling optimization, optimal resource allocation	Genetic algorithm, PSO, evolution algorithm	Shop floor of aircraft engine (blisk machining)
[25]	Resource management, product quality control and optimization, scheduling	Genetic algorithm, PSO	Satellite assembly shop floor
[94]	Forecast work-in process time, optimum resource allocation	Time-weighted multiple linear regression models	Shop floor
[79]	Aerodynamic performance optimization	Deep reinforcement learning (deep deterministic policy gradient algorithm)	Centrifugal impeller
[95]	Optimal process planning	Deep learning (reuse network (PKR-Net), deep residual networks)	Impeller
[96]	Achieved dynamic geometric and physical properties	Biomimicry principles (biological mimicry)	Missile air rudder-machining process (geometry-DT, behavior-DT and context-DT)
[97]	Process planning and optimization	Big data analytics (mathematical, statistical)	Marine diesel engine (connecting rod)
[98]	Quality improvement for product assembly, quality defects and causes detection	Deep learning (convolutional neural network)	Remote laser welding (aluminium doors)
[99]	Performance optimization	Feed-forward neural network, multi objective evolutionary optimization (genetic algorithm)	Dew-point cooler
[100]	Collaborative data management, AM defect analysis	Deep learning	Project MANUELA

aircraft [3], [118]–[120], gillnet [122], gearbox, aircraft-turbofan engine, rotating shaft-bearing [121], etc. All these systems are summarized in Table 3.

C. POWER AND ENERGY

In the power and energy sector, most of the DTs are developed in electronic systems, wind-power farms, cooling systems, and fuel-related systems. The digital twin of an inverter model [125] was developed by imitating the voltage controller, the current control loop, and the controlled plant, based on three distinct neural networks (NNs). Each of the three NNs is trained on real data collected from the physical model, where the back propagation (BP) algorithm is deployed to tune, in real-time, the proportional–integral (PI) controller. Also, Andryushkevich *et al.* [126] introduced the digital twin of power-system using ontological modeling. The developed DT selects the optimal configuration of the hybrid power supply system, by utilizing genetic algorithms [127]. Likewise, a digital twin framework for power grids was designed by Zhou *et al.* [128] to perform real-time analysis. Specifically, NN-based learning was applied to predict the grid operational behavior for fast security assessment, based on the voltage stability and oscillation damping.

In addition, a DT for a dew-point cooler was developed [99] to improve its cooling performance, by optimizing operational and design parameters, including cooling capacity, coefficient of performance (COP), dew point efficiency, wet-bulb efficiency, supply air temperature, and surface area. The DT of the cooler is developed with feed-forward neural networks (FFNNs), and digitally mimics the cooler's behavior by utilizing the air characteristics (i.e., temperature, relative humidity) as well as the main operational and design parameters (i.e., air velocity, air fraction, HMX height, channel gap) as inputs to the FFNN. Later, the DT-collected data are supplied to a genetic algorithm (GA) for multi-objective evolutionary optimization, in order to maximize cooling, COP, and wet-bulb efficiency, and minimize the surface area within four diverse climates (i.e., tropical rainforest, arid, Mediterranean hot summer, and hot summer continental climates).

Apart from design and performance optimizations, ML-based PHM is accomplished for power and energy related components with the use of DTs, such as wind-turbine, [108], electric vehicle motor [61], photovoltaic systems [112], battery systems [113], plasma radiation detection in metal absorber–metal resistor bolometer [114], as discussed in Section VII-B.

TABLE 3. State-of-the-art AI-ML research in industrial digital twinning for PHM.

Paper	Application	AI-ML approach	DT use-case
[108]	Gearbox prognosis, performance degradation, and fault detection	Neural networks (extreme learning machine (ELM))	Wind turbine
[112]	Faults diagnosis	Holistic fault diagnosis approach	Photovoltaic energy conversion unit
[113]	Battery's aging level (capacity fade and power fade) prediction	Particle swarm optimization	Lithium and lead-acid battery system
[61]	Faults in motor (degradation of winding insulation, short circuit, etc.)	Artificial neural network (ANN) and fuzzy logic	Electric vehicle motor
[114]	Plasma radiation detection	Fuzzy logic	Bolometer
[80]	Health monitoring, anomalies in control rod systems	SVM (classification),PCA (feature reduction), K-means algorithm (clustering)	Control element drive mechanism
[115]	Dynamically detect structural damage or degradation and adopt strategy	Static-condensation reduced-basis-element (SCRBE) method, Bayesian state estimation	12ft wingspan unmanned aerial vehicle (UAV)
[110]	Ship speed-loss prediction due to marine fouling	Deep learning (multilayered-deep extreme learning, robust regression model)	Ship
[116]	Spacecraft structural life prediction	Dynamic Bayesian network [117]	Spacecraft
[3]	Aircraft structural life prediction	Mathematical modeling, machine learning	Aircraft
[118]	Crack state estimation, fatigue life prediction	Probabilistic models (Monte Carlo and high-fidelity finite element models)	Aircraft
[119]	Expert fault diagnosis	Artificial neural network (ANN), information fusion	Aircraft engine
[120]	Fault prediction and maintenance in shaft bearing (inner race, outer race, rolling parts faults)	Neural network, deep learning (extreme learning machine (ELM))	Aero-engine
[121]	Track asset degradation, detect faults	Generative adversarial network (GAN)	Gearbox, aircraft-turbofan engines, rotating shaft bearing
[122]	Damage detection	Artificial neural networks	Bottom-set gillnet
[123]	Fault prediction and optimization	Machine learning (artificial neural networks)	CNC machine tool (CNCMT)
[124]	Cutting tool fault, life prediction	Regression model, Bayesian model, Kalman filter, particle filter	CNC machine tool (CNCMT)

D. VEHICLES AND TRANSPORTATION

A vehicle digital twin was developed by Alam and El Saddik [85] in a vehicular cyber-physical system (VCPS), by mimicking its speed behavior, fuel consumption, and airbag status. The system utilized fuzzy rule base with a Bayesian network [129], in order to build a reconfiguration model for driving assistance. Similarly, Kumar *et al.* [60] built virtual models of running vehicles in the cloud, which obtained real-time road and vehicular data through fog or edge devices, in order to avoid traffic congestion. The driver behavior and intention are predicted using machine learning on historical data. LSTM-based recurrent neural networks (RNNs) [130] are applied on the data to obtain the best route for a particular vehicle. Besides, digital twins have also been developed for vehicular network system, itself. For instance, the digital twin of a mobile edge computing (MEC) system was developed [59] for resource allocation in unmanned aerial vehicle (UAV) networks, using deep recurrent Q-networks (DRQNs) [131]. Likewise, the digital twin of software-defined vehicular networks (SDVNs) [132] allows for the predictive verification and maintenance diagnosis of running vehicles network, using machine learning. Furthermore, prognostics and health management is conducted by developing digital twin of aircraft [118] and spacecraft [116], ship [110], and electric vehicle motor [61]. All of these PHM approaches employ machine learning techniques.

E. HEALTHCARE

In healthcare, the majority of AI-ML enabled DTs are human digital twins [23], [56], [133]–[136]. Mimicking the full functionalities of a human body is not currently possible, thus, a human digital twin can only focus on limited aspects of human biology. For example, the digital twin by Barricelli *et al.* [133] focuses on fitness-related measurements of athletes. Specifically, their virtual patient classified physical athletes and predicted their behavior using KNN classifiers [137] and support vector networks [138], by training models on physical patient data collected by IoT devices. Björnsson *et al.* [23] concentrated on protein–protein interaction (PPI) networks to diagnose and treat patients of a particular disease. Their model is implemented as an AI system that monitors the effect of drugs on the human body, using machine learning tools, such as Bayesian networks, deep learning, and decision trees.

Furthermore, Chakshu *et al.* [135] mimicked the patient's head behavior for detecting the severity of carotid stenosis. Their model selects components from a patient video and applies principal component analysis (PCA) to identify the severity of carotid stenosis, by comparing it with the virtual model components. The authors also recommended the use of deep learning, machine learning, and other AI techniques for better detection accuracy. Similarly, Mazumder *et al.* [134] digitally replicated the process of generating synthetic PPG signals to create the digital twin of a cardiovascular system. In the virtual model, parameters are optimized using a particle-swarm-optimization (PSO) algorithm. The algorithm minimizes the integral-squared-error (ISE) in the feature set, in order to generate the synthetic PPG signal. On the other hand, Laamarti *et al.* [136] and Liu *et al.*'s [56] models are generic ML-enabled frameworks for providing health services to elderly people.

F. COMMUNICATIONS AND NETWORKS

In the networking and communications domain, the digital twin of an indoor space environment [139] is implemented to model, predict, and control the terahertz (THz) signal propagation characteristics in an indoor space. The DT selects the best THz signal path from the base station to the mobile target, by avoiding obstacles. The DT identifies the obstacle, its position, and dimensions, by applying a you only look once (YOLO) machine learning algorithm [140] on the monochromatic image of the obstacle. Furthermore, deep learning algorithms are used for material texture recognition and classification. On the other hand, a new network architecture, equipped with ML-based virtual twin of a software-defined vehicular network (SDVN) [132], is designed to benefit from intelligent networking and adaptive routing. Dong *et al.* [22] developed a similar digital twin of a real network for mobile edge computing. The virtual model of the MEC is equipped with a deep neural network that is frequently updated based on the variation of the real network. The model then selects the optimal resource allocation and offloading policy at each access point.

G. SMART CITIES

In the smart city sector, a Zurich city digital twin [141] was developed by transforming 3D spatial data and city models, including buildings, bridges, vegetation, etc., to a virtual world. The authors discussed the effects of urban climate, which can be predicted by machine learning techniques based on the current weather and air-quality data. Similarly, a Vienna city digital geoTwin [142] can be linked with city data, such as socioeconomic, energy consumption, and maintenance management data, in order to make it a living digital twin with the aim of AI technologies. Furthermore, a vision for integrating artificial and human intelligence for a disaster city digital twin is introduced by Fan *et al.* [66]. Finally, a geospatial digital twin [143] is the digital replica of a spatial entity, where machine learning and deep learning techniques are used for interpretation, analysis, and organization of 3D point clouds.

H. OTHER INDUSTRIES

DT systems that utilize AI-ML techniques have been deployed in other industries as well. For instance, the supply chain DT by Marmolejo-Saucedo [24] was developed for a pharmaceutical company, using machine learning and pattern recognition algorithms. The objective was to identify the behavior, dynamics, and changing trends in the supply chain. Data management for DT environments is another area of active research. Specifically, a DT-enabled collaborative data management framework was proposed, using edge and cloud computing [100]. The goal was to perform advanced data analytics in additive manufacturing (AM) systems, in order to reduce the development time and cost, and improve the product quality and production efficiency. To this end, the authors introduced cloud-DTs and edge-DTs, developed at different product life-cycle stages, which communicate with each other in order to support intelligent process monitoring, control, and optimization. As a use case, the framework was implemented within the MANUELA project, where layer defect analysis was performed by a deep learning model on product life-cycle data. Moreover, Tong *et al.* [144] introduced an intelligent machine tool (IMT) digital twin model for machining data acquisition and processing, using data fusion and ML approaches.

VIII. DATA-DRIVEN DIGITAL TWINNING PATENTS

The importance of DT technologies can be verified by the number of patents in this field. In particular, more than one thousand patents have been awarded on AI-enabled digital twinning in all around the world. A wind-power farm digital twin was filed as a U.S. Patent in 2016 by General Electric (GE) [145], where the DT is composed of two communication networks: 1) a farm-based communication network, which enables the coupling of control systems from individual wind turbines with the main wind farm control system and with other wind turbines; and 2) a cloud-based communication network that is composed of an infrastructure of digital wind-turbine models, where the plurality of the models are continuously changing during farm operation, by investigating and analyzing data generated by the farm-based communication network using machine learning. Furthermore, they provided a fully functional graphical user interface (GUI) of the digital wind-turbines, where the user can control the input features of the DT model to optimize the performance of the wind farm using machine learning algorithms. In another patent, Shah *et al.* [146] developed the digital twin of a vehicle cooling system, by using status data (such as health scores) to predict cooling system failures and optimize its performance. Similar data-driven digital twinning systems have been designed in the energy and power sector [147].

In predictive analytics for machine maintenance, GE's Hershey *et al.* [148] developed a system to predict the lifetime of a component in the electromechanical industry (such as an aircraft engine), by developing a digital twin of the physical system. The component is monitored by IoT-based sensors and its remaining life is assessed based on the monitoring conditions. In this process, they developed a stress analysis model, a fluid dynamics model, a structural dynamic model, a thermodynamic model, and a fatigue cracking model. Then, they utilized probabilistic models, such as a Kalman filter, to predict the lifetime and detect component faults. Similarly, the Siemens corporation designed a generic digital twin model [149] for a variety of machines, including heating, ventilation, and air conditioning (HVAC). They utilized data-driven approaches for energy-efficient machine maintenance, utilizing sensor data and model-based analytics. Several other patents focus on predictive analytics with AI-enabled digital twins [150]–[152].

A few digital twin patents have also been developed in the healthcare sector. GE researchers designed a patient DT [153] to diagnose diseases, treat, and prescribe medicines. The digital representation of the patient (i.e., the DT) consists of medical record data structures, medical images, and historical patient information. The DT is equipped with healthcare software applications (such as expert systems), patient medical data, and AI models (neural networks, machine learning) that can diagnose, identify health issues, and prescribe treatments (e.g., medication, surgery, etc.). Also, Nagesh [154] build an X-ray tube DT to predict tube-liquid bearing failures. He used X-ray tube housing vibration data, collected by a sensor in a free run mode of an X-ray tube, and applied AI-based prediction. There are also patents in DT-based surgery for the healthcare industry that utilize data-driven approaches [155], [156].

Finally, there are hundreds of additional patents that emphasize AI-enabled data-driven digital twinning, which could not be covered here. These digital twinning systems belong to a variety of industrial sectors, including manufacturing [157], [158], run-time environment [159], transport and automotive industry [160]–[162], building and construction systems [163], etc.

IX. EVALUATING A SUCCESSFUL DIGITAL TWIN

A successful digital twin can only be justified when its virtual twin closely matches the functionality of its physical counterpart. This justification can be presented by comparing the outputs of the physical and virtual models, and computing the loss. On this account, accuracy is the main factor to consider when evaluating digital twins. On the other hand, the purpose of building a digital twin also matters in evaluating its success. This can be justified by the performance improvement of the corresponding physical system that is attributed to its digital twin. For example, for a DT whose purpose is to optimize the assembly line, the improvement can be measured by computing the number of actions (or subtasks) and the time taken to manufacture a full component (or to complete a main task/goal) with the DT and without DT. This is also the case with other applications, including product design optimization, product performance optimization, process optimization, control optimization, scheduler optimization, resource management, component PHM, etc. In addition, the processing time and efficiency of the digital twinning system can also be one of the success criteria.

In addition, when using AI or machine learning approaches, the accuracy of the selected model affects the efficacy of the DT. Specifically, the accuracy of the underlying ML model, as well as the feature selection process and the amount of training data, may greatly affect the outcome of the DT. Therefore, when designing a DT-based system that employs ML techniques, we have to select the model with the higher accuracy and efficiency. The same approach should be taken with the selection of other technologies for DT-development, such as IoT, edge computing, and cloud computing.

To this end, only a few state-of-the-art digital twinning systems have been fully evaluated in the literature. For instance, Zhang *et al.* [87] assessed their job-floor digital twin by comparing the performance of the job-floor with and without digital twinning. They selected job scheduling time, utility rate, and job tardiness as performance parameters. Similarly, Zhang *et al.* [93] highlighted the importance of digital twinning by showing the performance improvement in process time, fault time, and maintenance time of blisk machining due to its digital twin. Likewise, Min *et al.* [164] conveyed a rise in the oil yield ratio due to a petrochemical industry DT. Furthermore, Xu *et al.* [45] used the accuracy of fault diagnosis as a metric to assess the performance of the developed virtual twin. Finally, Akhlaghi *et al.* [99] verified the accuracy of the developed twin by comparing the outputs of the digital and physical twins. They also showed the effectiveness of their digital twinning mechanism, by pointing out the optimization achieved for the dew point cooler. All the aforementioned DTs were developed using various machine learning models and, in each case, the authors selected the model that provided the best accuracy.

X. DIGITAL TWIN DEVELOPMENT TOOLS

There is no standalone technology for DT implementation, rather, there is an integration of multiple technologies, including big data, AI-ML, IoT, CPS, edge computing, cloud computing, communication technologies, etc. Every technological component can be implemented with a variety of tools. Here, we only focus on the tools that facilitate components integration, digital twin simulation, twins bridging, physical twin control, data storage and processing, and machine learning. Table 4 presents the summary of widely used tools that may provide support in different stages of digital twinning.

Integrating physical components for data collection and then digitally mimicking them in a virtual environment are two important stages of digital twinning. There are various tools available to accomplish these tasks in an industrial unit. Siemens MindSphere is one of the widely used tools to integrate components in a manufacturing industry. Siemens also developed an object-oriented-based Tecnomatix API to simulate physical components in a virtual environment, as used by [91]. The Open Simulation Platform (OSP) is another one, which is jointly developed by the Det Norske Veritas Germanischer Lloyd group (DNV GL), the Norwegian University of Science and Technology (NTNU), Rolls-Royce, and SINTEF Ocean. OSP can digitally mimic any component of the maritime industry. Other popular integration and simulation tools are FIWARE, Predix (a cloud-based platform from

TABLE 4. Digital twinning supporting tools.

Type	Tool-name	Owned-by	Additional services & details	Reference
Integration and simulation	MindSphere	Siemens	Facilitates bridging and twin control	https://siemens.mindsphere.io
	Tecnomatix API	Siemens	Object-oriented	[165]
	Open Simulation Platform	DNV-GL	For any component of the maritime industry	https://opensimulationplatform.com/
	FIWARE	Naeva Tec	Open source platform–facilitates bridging and twin control	[166]
	The Predix System™	GE digital	Cloud-based platform–facilitates data processing, bridging, and twin control	https://www.predix.io/
	IndraMotion MTX	Rexroth	CNC machine tools control platform–facilitates bridging and twin control	[167]
	Beacon	Fii-Foxconn	Data storage, processing, bridging, and twin control	https://www.iotone.com/term/beacons/t80
	Thingworx	PTC	DT-modeling, data storage, bridging, and twin control	https://www.ptc.com/en/products/thingworx
Digital twin modeling	ANSYS Twin Builder	ANSYS	Extensive features and libraries–facilitates simulation, bridging, and twin control	https://www.ansys.com/products/systems/ansys-twin-builder
	Mworks	TONGYUAN	Behavioral modeling–facilitates simulation	http://en.tongyuan.cc/
	Siemens NX software	Siemens	Integrated tool set–facilitates design, simulation, and manufacturing solutions	https://www.plm.automation.siemens.com/global/en/products/nx/
	SolidWorks	Dassault Systèmes	Geometric modeling and design	https://www.solidworks.com/
	Autodesk tools (AutoCAD, 3D Max, Maya)	Autodesk	Geometric modeling	https://www.autodesk.com/products/
	FreeCAD	Freecadweb	Open source, geometric modeling	https://www.freecadweb.org/
Bridging and twin control	TwinCAT software system	Beckhoff	Facilitates physical twin optimization	https://www.beckhoff.com/en-us/.../twincat/
	SAP with trenitalia	SAP	Train monitoring and PHM	https://news.sap.com/sap-tv/sap-trenitalia-iot-express-all-aboard/
	Codesys	CODESYS Group	Engineering control systems	https://www.codesys.com/
	FANUC CNC	FANUC GLOBAL	Control	https://www.fanuc.com/
	Flexium CNC	Flexium	Monitoring, bridging, and twin control	https://num.com/products/tools
	Hauzhing CNC system	Hauzhing CNC	Twin control	https://www.hcnc-group.com/
	Guangzhou CNC system	Guangzhou CNC	Twin control	http://www.gz-cnc.com/
	Aspera	IBM	Fast data transmission	https://www.ibm.com/products/aspera
	RaySync	Raysync	Fast data transmission	https://www.raysync.io/
Big data processing tools	Hadoop	Apache	Big data ecosystem for distributed processing	[168]
	MapReduce	Apache	Programming framework	https://www.tutorialspoint.com/hadoop/hadoop_mapreduce.htm
	Apache tools (Spark™, Storm, S4, Hive, Mahout, Flink, Pig, Impala)	Apache	Big data processing tools, with built-in modules for streaming, SQL, machine learning and graph processing. Mostly open-source.	https://www.apache.org/index.htmlprojects-list
	HPCC	LexisNexis Risk Solution	Data management and processing–open-source solution	https://hpccsystems.com/
	Qubole platform	Qubole	Open data lake platform (streaming & offline)	https://www.qubole.com/
	Statwing	Qualtrics	Statistical data analysis	https://www.statwing.com/
	Pentaho	Hitachi Vantara	Data analytics platform	https://marketplace.hitachivantara.com/pentaho/
	VoltDB	VoltDB	Data storage–facilitates data processing and in-memory database support	https://www.voltdb.com/
	Akka	Lightbend	Open-source toolkit and runtime	https://akka.io/
	Predix platform	GE Digital	Monitoring, bridging, processing	https://www.predix.io/
Big data storage	Hadoop-HDFS	Apache	Distributed storage	https://hadoop.apache.org/docs/../hdfs_design.html
	HBase	Apache	Real-time read/write access to disk	https://hbase.apache.org/
	Oracle	Oracle Corporation	Cloud support + relational DB	https://www.oracle.com/index.html
	Kafka	Apache	Open-source distributed event streaming platform	https://kafka.apache.org/
AI-ML tools and APIs	TensorFlow	Google	Free and open-source ML-library	[169]
	CNTK	Microsoft	Deep learning toolkit	[170]
	Caffe	Berkeley AI Research (BAIR)	Deep learning framework	[171]
	Keras	François (MIT license)	Easier and user friendly interfaces (basic models)	[172]
	Weka	University of Waikato	Easier and user friendly interfaces (basic models)	[173]
	Matlab	Mathworks	Commercial tools–vast libraries for ML, Microsoft-Azure ML models	https://ch.mathworks.com/products/matlab.html
	Gym	OpenAI	Reinforcement learning (standardized interfaces)	[174]
	rllab	UC Berkeley/ OpenAI	Reinforcement learning (standardized interfaces)	[175]

GE digital), CNC machine tools control platform IndraMotion MTX, Beacon, Thingworx, and others.

Next, bridging physical and virtual twins is another primary aspect of digital twinning. This bridge is used by a virtual twin to harvest the real-time data from the corresponding physical peer using sensors. On the other side, the physical peer is controlled (optimized) based on the output of the virtual twin. Popular tools in the market to facilitate the bridging between physical and virtual twins are TwinCat, SAP, Codesys, CNC tools, Aspera, and RaySync. Similarly, there are few applications that are used in initial modeling and twin design, such as ANSYS Twin Builder, MWorks, Siemens NX software, SolidWorks, Autodesk tools, and FreeCAD.

In the machine learning domain, there are hundreds of models available for tasks such as optimization, prediction, classification, and clustering. However, there is no single platform that offers APIs for all existing ML models. The most widely used and well-known libraries for implementing, training, and testing supervised ML-models are Tensorflow, CNTK and Caffe. Keras and Weka provide easier and user-friendly interfaces for developing basic machine learning models. There are also commercial tools available, such as Mathworks Matlab, which is equipped with vast libraries of neural networks and Microsoft-Azure implemented ML models. Reinforcement learning is one of the most popular techniques that is widely used for dynamic optimization and process planning in digital twinning. To this end, OpenAI's Gym and rllab are tools with standardized interfaces for reinforcement learning.

Industrial components produce large amounts of data, termed as big data, which are hard to process with standard data management tools in a digital twin environment. Hadoop is one of the most popular ecosystems for big data processing that offers parallel processing capabilities with multiple compute nodes. Apache has also developed several tools for big data processing and effective analysis, including Cassandra, Spark, Storm, S4, Hive, Mahout, Flink, and HBase. Most of the Apache tools are open-source and support machine learning APIs. Similar tools include HPCC by LexisNexis Risk Solution, Qubole, Statwing, Pentaho, and VoltDB.

XI. DATA-DRIVEN REFERENCE ARCHITECTURE FOR DIGITAL TWINNING

To effectively exploit the value-added capabilities offered by the integration of big data analytics and AI-ML within the scope of digital twinning, we present a novel reference model derived from the conducted systematic literature review. Fig. 8 shows the designed reference layered-architecture for the efficient handling of big data analytics in DT-based industrial environments. The process starts with the collection of data from the physical environment (using sensors and actuators) or from the virtual environment (using computer-aided software and/or simulations). The data is fed to the data analysis and decision-making layer, where AI models, statistical and probabilistic approaches, or mathematical models are employed to create the DT-based system or the digital twin itself. During the entire process, various big data processing tools may be utilized, such as Hadoop, Storm, S4, Spark, etc., that allow for parallel processing on multiple compute nodes. Fig. 9 depicts the overall data flow for creating an ML-enabled digital twin, and then using it for optimization, PHM, or other purposes. First, the virtual model is created by deploying one of the AI models on the data generated by the physical twin. Once the digital twin is produced, the data from both the physical and virtual twins are given to other AI models to achieve the given industrial goals, such as design optimization, dynamic process planning, healthcare, or PHM. Moreover, the results can be further used to update and improve both the physical and virtual twins.

XII. MARKET OPPORTUNITIES AND RESEARCH CHALLENGES

A. MARKET OPPORTUNITIES AND RESEARCH AREAS

Based on the detailed literature survey, we have summarized the following major application areas where DT research can play a vital role.

1) OPTIMIZATION

Optimization is required in almost every industrial process, including product design, product performance, process planning, assembly line, task-scheduling, and resource-allocation. Digital twinning is an emerging technology that provides a direct pathway to optimization with little effort. However, careful consideration of the optimization algorithm (i.e., ML model) and the underlying feature set (for the optimization algorithm) is desired for better results.

2) PROCESS MONITORING, DIAGNOSTICS, AND PREDICTION

Digital twins can be developed for industrial process monitoring, defect diagnosis (i.e., product quality assurance), dynamic process or product design updating for time and cost savings, industrial process surveillance (e.g., robot DT for obstacle avoidance), product time-to-complete prediction, and damage detection.

3) PREDICTIVE ANALYTICS FOR MANUFACTURED PRODUCTS

The quality of every physical entity degrades over time, thus affecting its performance. Early detection of failures may promote on-time maintenance, fatigue avoidance, as well as time and cost savings. Such failures can be attributed to faults and cracks in the product, performance degradation due to aging, and other minor or major complications. Moreover, health monitoring is crucial for certain components that may potentially cause human casualties, e.g., brake systems in cars, vehicles, aircraft, and ship engines, fueling systems, gearboxes, etc. Digital twinning is the most powerful technology for predictive analytics and health monitoring of physical components. This is also an area where AI-ML techniques can have a significant impact.

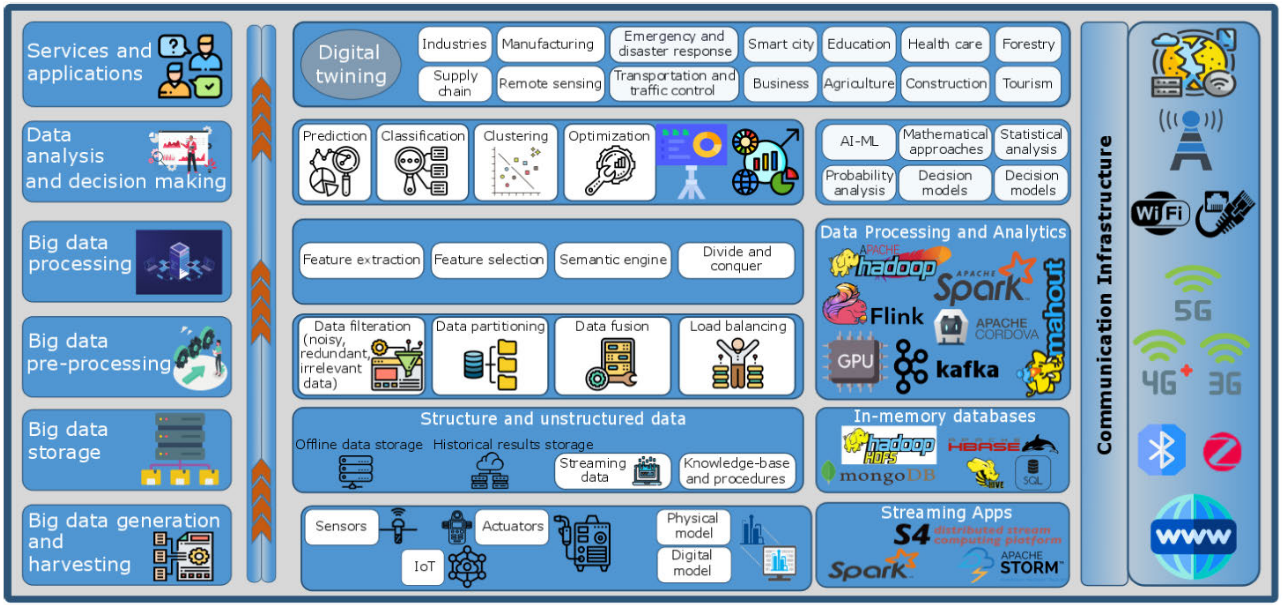

FIGURE 8. Data-driven reference architecture for digital twinning.

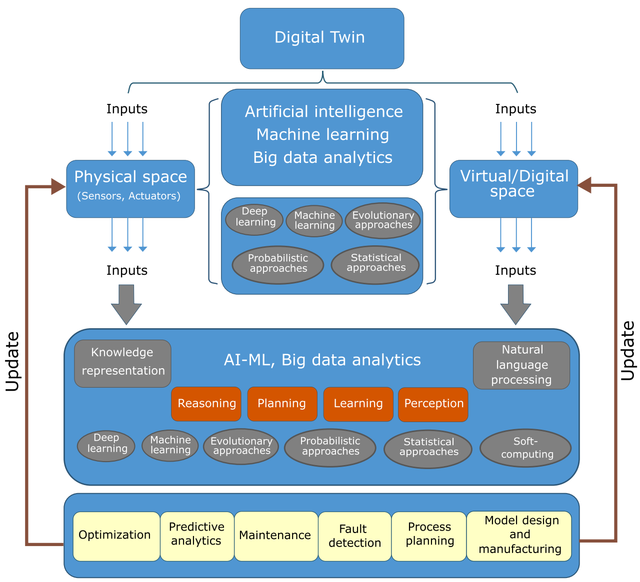

FIGURE 9. Overall data-flow framework for digital twinning using big data analytics and AI-ML.

4) HEALTHCARE

Digital twinning has a wider scope in the healthcare sector where human-DTs assist in day-to-day human fitness and health monitoring, early disease diagnosis, and the overall well-being of individuals, especially for the elderly and infants. In addition, it can be used for the treatment or surgery of patients, by developing a patient-DT. Developing digital twins for human organs or biological systems will bring a revolution in the healthcare sector, such as DTs for lungs, liver, pregnant female womb or uterus, cardiac system, digestion system, neural system, reproductive system, etc. Other than biological digital twins, the healthcare sector can benefit by developing DTs for hospitals, medical and surgical instruments, remote surgery, surgical processes, etc.

5) SMART CITIES

In the context of smart cities, DT technologies can be implemented for traffic systems, smart homes and devices, parking, buildings, livestock, lighting systems, and renewable energy. Furthermore, 3D virtual city models may facilitate urban planning and monitoring in various smart city areas, including road monitoring and construction, city garbage management, bridge and housing constructions, etc.

6) OTHER APPLICATIONS

Research opportunities are not limited to the above-mentioned sectors, but the potential is there in every field, including education, construction, mining, communications and networks, food and agriculture, sports, and so on.

B. RESEARCH CHALLENGES AND ISSUES

The rapidly increasing DT popularity and scope, as well as the involvement of IoT, big data, and AI technologies, broaden the research challenges of digital twinning. These challenges are categorized in the following five areas.

1) DATA COLLECTION

IoT facilitates data harvesting from a physical twin (using sensors), data integration, and data sharing with the corresponding virtual twins. This process can amount to a considerable cost. Sometimes, the digital twin may be more costly than the asset itself, in which case it does not make sense to create the DT. On the other hand, the collected data is large (big data), heterogeneous in nature, unstructured, and noisy. Thus, further processing on the data is required to ensure its effective use. Specifically, we need to apply data cleaning techniques, and also organize, restructure, and make the data homogeneous. Furthermore, controlling the flow of such large amount of data is also a significant challenge. Finally, to improve the accuracy of the DT model, the underlying machine learning algorithms require a certain amount of data for training purposes.

2) BIG DATA CHALLENGES

The explosive growth of IoT technologies in the industrial sector has led to the generation of large amounts of monitoring (sensor) data. To this end, big data analytics requires advanced architectures, frameworks, technologies, tools, and algorithms to capture, store, share, process, and analyze the underlying data. There is also a potential for edge and cloud computing platforms to handle DT-related data. Specifically, edge computing enables the distributed processing at the network's edge, while the aggregate processing is accomplished in the cloud. However, the aggregation of data in the cloud may cause an increase in response time.

3) DATA ANALYSIS

AI-algorithms for data analytics played a major role in DT for decision-making, as discussed in the literature. However, the selection of a particular model among hundreds of ML-models with customized configuration is challenging. Every AI-approach has diverse accuracy and efficiency levels with different applications and datasets (feature set). On the other hand, accuracy may affect the efficiency on the other side. Hence, depending on the motive and application of a DT, the selection of the best ML-algorithm and features is challenging. Besides, fewer practical implementations of AI-techniques for digital twinning in the literature raises more challenges.

4) DT STANDARDIZATION CHALLENGES

Even though many digital twins have been developed in various industries, the creation of a complex and reliable digital twin demands standardization. Currently, there is no single standard that solely focuses on digital twinning. The ISO/DIS 23247-1 standard [29] has only limited information on digital twinning and, therefore, DT deployment challenges grow due to the lack of standardization. Standardization efforts are underway by the joint advisory group (JAG) of ISO and IEC on emerging technologies [28].

5) SECURITY AND PRIVACY ISSUES

Some DT systems, such as human-DTs, product PHM, or defense-related DTs, are considered critical and may require stringent security and privacy guarantees. First, due to the involvement of IoT devices in digital twinning, a lot of emphasis has to be placed on the security of the underlying communication protocols. Additionally, the large collection of asset-related data needs to be stored securely, in order to prevent data breaches from insider and outsider threats.

XIII. CONCLUSION

We performed a systematic literature review of the state-of-the-art DT systems that employ machine learning and AI technologies. In particular, we focused on papers published in top multidisciplinary electronic bibliographic and patent libraries, and summarized the current DT deployments in a variety of industries. With the immersion of AI-ML and big data, digital twinning is evolving at a rapid rate and, with it, a lot of unique challenges and new opportunities are emerging. This article highlighted the research challenges and potentials in many diverse areas, for both academia and industry. Furthermore, we identified the DT criteria and tools that aid its successful development. Finally, we designed a reference model for an AI-ML and big data-enabled digital twinning system to further guide industrial developers in establishing DTs that can make their systems smarter, intelligent, and dynamically adaptable to changing conditions.

REFERENCES

[1] M. W. Grieves, "Virtually intelligent product systems: Digital and physical twins," *Complex Syst. Eng., Theory Pract.*, pp. 175–200, 2019.

[2] M. Grieves, "Digital twin: Manufacturing excellence through virtual factory replication," White Paper, 2014, pp. 1–7, vol. 1.

[3] E. J. Tuegel, A. R. Ingraffea, T. G. Eason, and S. M. Spottswood, "Reengineering aircraft structural life prediction using a digital twin," *Int. J. Aerosp. Eng.*, vol. 2011, pp. 1–14, Aug. 2011.

[4] D. Cearley, B. Burke, D. Smith, N. Jones, A. Chandrasekaran, and C. Lu, "Top 10 strategic technology trends for 2020," Gartner, Stamford, CT, USA, Tech. Rep., 2019.

[5] T. R. Wanasinghe, L. Wroblewski, B. K. Petersen, R. G. Gosine, L. A. James, O. De Silva, G. K. I. Mann, and P. J. Warrian, "Digital twin for the oil and gas industry: Overview, research trends, opportunities, and challenges," *IEEE Access*, vol. 8, pp. 104175–104197, 2020.

[6] Y. Lu, C. Liu, K. I.-K. Wang, H. Huang, and X. Xu, "Digital twin-driven smart manufacturing: Connotation, reference model, applications and research issues," *Robot. Comput.-Integr. Manuf.*, vol. 61, Feb. 2020, Art. no. 101837.

[7] C. Cimino, E. Negri, and L. Fumagalli, "Review of digital twin applications in manufacturing," *Comput. Ind.*, vol. 113, Dec. 2019, Art. no. 103130.

[8] Q. Qi and F. Tao, "Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison," *IEEE Access*, vol. 6, pp. 3585–3593, 2018.

[9] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital twin in industry: State-of-the-art," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2405–2415, Apr. 2019.

[10] A. Rasheed, O. San, and T. Kvamsdal, "Digital twin: Values, challenges and enablers from a modeling perspective," *IEEE Access*, vol. 8, pp. 21980–22012, 2020.

[11] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Keele Univ., Durham Univ., Keele, U.K., Tech. Rep. EBSE 2007-001, 2007.

[12] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering—A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, Jan. 2009.

[13] C. Okoli and K. Schabram, "A guide to conducting a systematic literature review of information systems research," SSRN, Tech. Rep., 2010. [Online]. Available: http://dx.doi.org/10.2139/ssrn.1954824

[14] D. Cearley, B. Burke, S. Searle, and M. Walker, "Top 10 strategic technology trends for 2017: A gartner trend insight report," *Gartner*, vol. 23, Jun. 2017, Art. no. 6595640781. [Online]. Available: https://www.gartner.com/doc/3645332

[15] D. Cearley, B. Burke, S. Searle, and M. J. Walker, "Top 10 strategic technology trends for 2018," *Gartner*, 2017.

[16] D. Cearley and B. Burke, "Top 10 strategic technology trends for 2019," *Gartner*, 2018.

[17] M. Grieves and J. Vickers, "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems," in *Transdisciplinary Perspectives on Complex Systems*. Cham, Switzerland: Springer, 2017, pp. 85–113.

[18] E. Glaessgen and D. Stargel, "The digital twin paradigm for future NASA and US air force vehicles," in *Proc. 53rd AIAA/ASME/ASCE/AHS/ASC Struct., Struct. Dyn. Mater. Conf., 20th AIAA/ASME/AHS Adapt. Struct. Conf., 14th AIAA*, 2012, p. 1818.

[19] F. Tao, F. Sui, A. Liu, Q. Qi, M. Zhang, B. Song, Z. Guo, S. C.-Y. Lu, and A. Nee, "Digital twin-driven product design framework," *Int. J. Prod. Res.*, vol. 57, no. 12, pp. 3935–3953, 2019.

[20] R. Söderberg, K. Wärmefjord, J. S. Carlson, and L. Lindkvist, "Toward a digital twin for real-time geometry assurance in individualized production," *CIRP Ann.*, vol. 66, no. 1, pp. 137–140, 2017.

[21] G. Bacchiega, "Creating an embedded digital twin: Monitor, understand and predict device health failure," *Inn4mech-Mechatronics Ind.*, vol. 4, 2018.

[22] R. Dong, C. She, W. Hardjawana, Y. Li, and B. Vucetic, "Deep learning for hybrid 5G services in mobile edge computing systems: Learn from a digital twin," *IEEE Trans. Wireless Commun.*, vol. 18, no. 10, pp. 4692–4707, Oct. 2019.

[23] B. Björnsson, C. Borrebaeck, N. Elander, T. Gasslander, D. R. Gawel, M. Gustafsson, R. Jörnsten, E. J. Lee, X. Li, S. Lilja, D. Martínez-Enguita, A. Matussek, P. Sandström, S. Schäfer, M. Stenmarker, X. F. Sun, O. Sysoev, H. Zhang, and M. Benson, "Digital twins to personalize medicine," *Genome Med.*, vol. 12, no. 1, pp. 1–4, Dec. 2020.

[24] J. A. Marmolejo-Saucedo, "Design and development of digital twins: A case study in supply chains," *Mobile Netw. Appl.*, vol. 25, no. 6, pp. 2141–2160, Dec. 2020.

[25] C. Zhuang, J. Liu, and H. Xiong, "Digital twin-based smart production management and control framework for the complex product assembly shop-floor," *Int. J. Adv. Manuf. Technol.*, vol. 96, nos. 1–4, pp. 1149–1163, Apr. 2018.

[26] R. Piascik, J. Vickers, D. Lowry, S. Scotti, J. Stewart, and A. Calomino, "Technology area 12: Materials, structures, mechanical systems, and manufacturing road map," *NASA Office Chief Technol.*, 2010.

[27] P. Caruso, D. Dumbacher, and M. Grieves, "Product lifecycle management and the quest for sustainable space exploration," in *Proc. AIAA SPACE Conf. Expo.*, Aug. 2010, p. 8628.

[28] JETI. *Which Technologies is Jeti Considering?* Accessed: May 8, 2020. [Online]. Available: https://jtc1info.org/technology/advisory-groups/jeti/

[29] *Automation Systems and Integration Digital Twin Framework for Manufacturing—Part 1: Overview and General Principles*, Standard ISO/DIS 23247-1, 2020. [Online]. Available: https://www.iso.org/standard/75066.html

[30] *Industrial Automation Systems and Integration-Product Data Representation and Exchange—Part 1: Overview and Fundamental Principles*, Standard ISO 10303-1, 1994. [Online]. Available: https://www.iso.org/standard/20579.html

[31] *2014 Cutting Tool Data Representation and Exchange—Part 3: Reference Dictionary for Tool Items, Int. Org. Standard*, Standard ISO 13399-3, 2014. [Online]. Available: https://www.iso.org/standard/54168.html

[32] O. Foundation. *Unified Architecture*. Accessed: 2008. [Online]. Available: https://opcfoundation.org/about/opc-technologies/opc-ua/

[33] R. Rosen, G. von Wichert, G. Lo, and K. D. Bettenhausen, "About the importance of autonomy and digital twins for the future of manufacturing," *IFAC-PapersOnLine*, vol. 48, no. 3, pp. 567–572, 2015.

[34] J. Vachálek, L. Bartalský, O. Rovný, D. Šišmišová, M. Morhác, and M. Lokšík, "The digital twin of an industrial production line within the industry 4.0 concept," in *Proc. 21st Int. Conf. Process Control (PC)*, Jun. 2017, pp. 258–262.

[35] E. Negri, S. Berardi, L. Fumagalli, and M. Macchi, "MES-integrated digital twin frameworks," *J. Manuf. Syst.*, vol. 56, pp. 58–71, Jul. 2020.

[36] Z. Yin and J. Hou, "Recent advances on SVM based fault diagnosis and process monitoring in complicated industrial processes," *Neurocomputing*, vol. 174, pp. 643–650, Jan. 2016.

[37] L. Bennacer, Y. Amirat, A. Chibani, A. Mellouk, and L. Ciavaglia, "Self-diagnosis technique for virtual private networks combining Bayesian networks and case-based reasoning," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 1, pp. 354–366, Jan. 2015.

[38] P. Tamilselvan and P. Wang, "Failure diagnosis using deep belief learning based health state classification," *Rel. Eng. Syst. Saf.*, vol. 115, pp. 124–135, Jul. 2013.

[39] Y. Qi, C. Shen, D. Wang, J. Shi, X. Jiang, and Z. Zhu, "Stacked sparse autoencoder-based deep network for fault diagnosis of rotating machinery," *IEEE Access*, vol. 5, pp. 15066–15079, 2017.

[40] W. Lu, Y. Li, Y. Cheng, D. Meng, B. Liang, and P. Zhou, "Early fault detection approach with deep architectures," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 7, pp. 1679–1689, Jul. 2018.

[41] Y. Qi Chen, O. Fink, and G. Sansavini, "Combined fault location and classification for power transmission lines fault diagnosis with integrated feature extraction," *IEEE Trans. Ind. Electron.*, vol. 65, no. 1, pp. 561–569, Jan. 2018.

[42] H. Darong, K. Lanyan, M. Bo, Z. Ling, and S. Guoxi, "A new incipient fault diagnosis method combining improved RLS and LMD algorithm for rolling bearings with strong background noise," *IEEE Access*, vol. 6, pp. 26001–26010, 2018.

[43] Y. Wang, Z. Wei, and J. Yang, "Feature trend extraction and adaptive density peaks search for intelligent fault diagnosis of machines," *IEEE Trans. Ind. Informat.*, vol. 15, no. 1, pp. 105–115, Jan. 2019.

[44] S. Yin, X. Zhu, and O. Kaynak, "Improved PLS focused on key-performance-indicator-related fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 62, no. 3, pp. 1651–1658, Mar. 2015.

[45] Y. Xu, Y. Sun, X. Liu, and Y. Zheng, "A digital-twin-assisted fault diagnosis using deep transfer learning," *IEEE Access*, vol. 7, pp. 19990–19999, 2019.

[46] Y. Wang, R. Xiong, H. Yu, J. Zhang, and Y. Liu, "Perception of demonstration for automatic programing of robotic assembly: Framework, algorithm, and validation," *IEEE/ASME Trans. Mechatronics*, vol. 23, no. 3, pp. 1059–1070, Jun. 2018.

[47] X. Li, B. He, Y. Zhou, and G. Li, "Multisource model-driven digital twin system of robotic assembly," *IEEE Syst. J.*, early access, Jan. 3, 2020, doi: 10.1109/JSYST.2019.2958874.

[48] Y. Fang, C. Peng, P. Lou, Z. Zhou, J. Hu, and J. Yan, "Digital-twin-based job shop scheduling toward smart manufacturing," *IEEE Trans. Ind. Informat.*, vol. 15, no. 12, pp. 6425–6435, Dec. 2019.

[49] S. Scharff. (2019). *From Digital Twin to Improved Patient Experience*. Accessed: May 8, 2020. [Online]. Available: https://www.siemens-healthineers.com/news/mso-digital-twin-mater.html

[50] T. Marchal. (Sep. 2016). *VPH: The Ultimate Stage Before Your Own Medical Digital Twin*. Accessed: May 8, 2020. [Online]. Available: https://www.linkedin.com/pulse/vph-ultimate-stage-before-your-own-medical-digital-twin-marchal/?trk=mp-reader-car

[51] C. Copley. (Aug. 2018). *Medical Technology Firms Develop 'Digital Twins' for Personalized Health Care*. Accessed: May 8, 2020. [Online]. Available: https://www.theglobeandmail.com/business/article-medical-technology-firms-develop-digital-twins-for-personalized/

[52] R. Martinez-Velazquez, R. Gamez, and A. El Saddik, "Cardio twin: A digital twin of the human heart running on the edge," in *Proc. IEEE Int. Symp. Med. Meas. Appl. (MeMeA)*, Jun. 2019, pp. 1–6.

[53] J. M. Ospel, G. Gascou, V. Costalat, L. Piergallini, K. A. Blackham, and D. W. Zumofen, "Comparison of Pipeline embolization device sizing based on conventional 2D measurements and virtual simulation using the Sim&Size software: An agreement study," *Amer. J. Neuroradiol.*, vol. 40, no. 3, pp. 524–530, Feb. 2019.

[54] M. Holtmannspotter, M. Martinez-Galdamez, M. Isokangas, R. Ferrara, and M. Sanchez, "Simulation in clinical practice: First experience with Sim&Cure before implantation of flow diverter (pipeline) or web-device for the treatment of intracranial aneurysm," in *Proc. ABC/WIN*, 2017.

[55] Y. Feng, J. Zhao, X. Chen, and J. Lin, "An in silico subject-variability study of upper airway morphological influence on the airflow regime in a tracheobronchial tree," *Bioengineering*, vol. 4, no. 4, p. 90, Nov. 2017.

[56] Y. Liu, L. Zhang, Y. Yang, L. Zhou, L. Ren, F. Wang, R. Liu, Z. Pang, and M. J. Deen, "A novel cloud-based framework for the elderly healthcare services using digital twin," *IEEE Access*, vol. 7, pp. 49088–49101, 2019.

[57] Z. Wang, X. Liao, X. Zhao, K. Han, P. Tiwari, M. J. Barth, and G. Wu, "A digital twin paradigm: Vehicle-to-cloud based advanced driver assistance systems," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–6.

[58] E. Cioroaica, T. Kuhn, and B. Buhnova, "(Do Not) trust in ecosystems," in *Proc. IEEE/ACM 41st Int. Conf. Softw. Eng., New Ideas Emerg. Results (ICSE-NIER)*, May 2019, pp. 9–12.

[59] X. Chen, T. Chen, Z. Zhao, H. Zhang, M. Bennis, and J. I. Yusheng, "Resource awareness in unmanned aerial vehicle-assisted mobile-edge computing systems," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, May 2020, pp. 1–6.

[60] S. A. P. Kumar, R. Madhumathi, P. R. Chelliah, L. Tao, and S. Wang, "A novel digital twin-centric approach for driver intention prediction and traffic congestion avoidance," *J. Reliable Intell. Environ.*, vol. 4, no. 4, pp. 199–209, Dec. 2018.

[61] S. Venkatesan, K. Manickavasagam, N. Tengenkai, and N. Vijayalakshmi, "Health monitoring and prognosis of electric vehicle motor using intelligent-digital twin," *IET Electr. Power Appl.*, vol. 13, no. 9, pp. 1328–1335, Sep. 2019.

[62] S. M. E. Sepasgozar, "Digital twin and Web-based virtual gaming technologies for online education: A case of construction management and engineering," *Appl. Sci.*, vol. 10, no. 13, p. 4678, Jul. 2020.

[63] M. Lammers. (Jun. 2018). *Opinion | Digital Twin Offers Huge Opportunities for Real Estate Life Cycle*. Accessed: May 8, 2020. [Online]. Available: https://www.proptech.nl/blog/digital-twin/

[64] A. Kampker, V. Stich, P. Jussen, B. Moser, and J. Kuntz, "Business models for industrial smart services—The example of a digital twin for a product-service-system for potato harvesting," *Procedia CIRP*, vol. 83, pp. 534–540, Jan. 2019.

[65] S. H. Khajavi, N. H. Motlagh, A. Jaribion, L. C. Werner, and J. Holmström, "Digital twin: Vision, benefits, boundaries, and creation for buildings," *IEEE Access*, vol. 7, pp. 147406–147419, 2019.

[66] C. Fan, C. Zhang, A. Yahja, and A. Mostafavi, "Disaster city digital twin: A vision for integrating artificial and human intelligence for disaster management," *Int. J. Inf. Manage.*, vol. 56, Feb. 2021, Art. no. 102049.

[67] M. Dietz and G. Pernul, "Unleashing the digital Twin's potential for ICS security," *IEEE Secur. Privacy*, vol. 18, no. 4, pp. 20–27, Jul. 2020.

[68] R. Langner, "To kill a centrifuge: A technical analysis of what stuxnet's creators tried to achieve," The Langner Group, Tech. Rep., 2013.

[69] S. Miller, N. Brubaker, D. K. Zafra, and D. Caban, "Triton actor TTP profile, custom attack tools, detections, and ATT&CK mapping," *Fireeye Threat Res. Blog*, Apr. 2019.

[70] M. M. U. Rathore, M. J. J. Gul, A. Paul, A. A. Khan, R. W. Ahmad, J. Rodrigues, and S. Bakiras, "Multilevel graph-based decision making in big scholarly data: An approach to identify expert reviewer, finding quality impact factor, ranking journals and researchers," *IEEE Trans. Emerg. Topics Comput.*, early access, Sep. 10, 2018, doi: 10.1109/TETC.2018.2869458.

[71] M. M. Rathore, H. Son, A. Ahmad, and A. Paul, "Real-time video processing for traffic control in smart city using Hadoop ecosystem with GPUs," *Soft Comput.*, vol. 22, no. 5, pp. 1533–1544, Mar. 2018.

[72] M. M. Rathore, A. Ahmad, A. Paul, and S. Rho, "Exploiting encrypted and tunneled multimedia calls in high-speed big data environment," *Multimedia Tools Appl.*, vol. 77, no. 4, pp. 4959–4984, Feb. 2018.

[73] S. A. Shah, D. Z. Seker, M. M. Rathore, S. Hameed, S. Ben Yahia, and D. Draheim, "Towards disaster resilient smart cities: Can Internet of Things and big data analytics be the game changers?" *IEEE Access*, vol. 7, pp. 91885–91903, 2019.

[74] X. Yuan, C. J. Anumba, and M. K. Parfitt, "Cyber-physical systems for temporary structure monitoring," *Autom. Construct.*, vol. 66, pp. 1–14, Jun. 2016.

[75] F. Thiesse, M. Dierkes, and E. Fleisch, "LotTrack: RFID-based process control in the semiconductor industry," *IEEE Pervas. Comput.*, vol. 5, no. 1, pp. 47–53, Jan. 2006.

[76] H. Choi, Y. Baek, and B. Lee, "Design and implementation of practical asset tracking system in container terminals," *Int. J. Precis. Eng. Manuf.*, vol. 13, no. 11, pp. 1955–1964, Nov. 2012.

[77] Y. Zheng, S. Yang, and H. Cheng, "An application framework of digital twin and its case study," *J. Ambient Intell. Humanized Comput.*, vol. 10, no. 3, pp. 1141–1153, Mar. 2019.

[78] K. Ding, H. Shi, J. Hui, Y. Liu, B. Zhu, F. Zhang, and W. Cao, "Smart steel bridge construction enabled by BIM and Internet of Things in industry 4.0: A framework," in *Proc. IEEE 15th Int. Conf. Netw., Sens. Control (ICNSC)*, Mar. 2018, pp. 1–5.

[79] Y. Zhou, T. Xing, Y. Song, Y. Li, X. Zhu, G. Li, and S. Ding, "Digital-twin-driven geometric optimization of centrifugal impeller with free-form blades for five-axis flank milling," *J. Manuf. Syst.*, Jul. 2020.

[80] A. Oluwasegun and J.-C. Jung, "The application of machine learning for the prognostics and health management of control element drive system," *Nucl. Eng. Technol.*, vol. 52, no. 10, pp. 2262–2273, Oct. 2020.

[81] A. Gandomi and M. Haider, "Beyond the hype: Big data concepts, methods, and analytics," *Int. J. Inf. Manage.*, vol. 35, no. 2, pp. 137–144, Apr. 2015.

[82] Y. Zhang, S. Ma, H. Yang, J. Lv, and Y. Liu, "A big data driven analytical framework for energy-intensive manufacturing industries," *J. Cleaner Prod.*, vol. 197, pp. 57–72, Oct. 2018.

[83] Y. Zhang, S. Ren, Y. Liu, and S. Si, "A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products," *J. Cleaner Prod.*, vol. 142, pp. 626–641, Jan. 2017.

[84] Y. Wang, S. Wang, B. Yang, L. Zhu, and F. Liu, "Big data driven hierarchical digital twin predictive remanufacturing paradigm: Architecture, control mechanism, application scenario and benefits," *J. Cleaner Prod.*, vol. 248, Mar. 2020, Art. no. 119299.

[85] K. M. Alam and A. El Saddik, "C2PS: A digital twin architecture reference model for the cloud-based cyber-physical systems," *IEEE Access*, vol. 5, pp. 2050–2062, 2017.

[86] E. A. Patterson, R. J. Taylor, and M. Bankhead, "A framework for an integrated nuclear digital environment," *Prog. Nucl. Energy*, vol. 87, pp. 97–103, Mar. 2016.

[87] M. Zhang, F. Tao, and A. Y. C. Nee, "Digital twin enhanced dynamic job-shop scheduling," *J. Manuf. Syst.*, May 2020.

[88] M. Schluse, M. Priggemeyer, L. Atorf, and J. Rossmann, "Experimentable digital twins—Streamlining simulation-based systems engineering for industry 4.0," *IEEE Trans. Ind. Informat.*, vol. 14, no. 4, pp. 1722–1731, Feb. 2018.

[89] S. Zhang, C. Kang, Z. Liu, J. Wu, and C. Ma, "A product quality monitor model with the digital twin model and the stacked auto encoder," *IEEE Access*, vol. 8, pp. 113826–113836, 2020.

[90] R. Bansal, M. A. Khanesar, and D. Branson, "Ant colony optimization algorithm for industrial robot programming in a digital twin," in *Proc. 25th Int. Conf. Autom. Comput. (ICAC)*, Sep. 2019, pp. 1–5.

[91] K. Xia, C. Sacco, M. Kirkpatrick, C. Saidy, L. Nguyen, A. Kircaliali, and R. Harik, "A digital twin to train deep reinforcement learning agent for smart manufacturing plants: Environment, interfaces and intelligence," *J. Manuf. Syst.*, Jul. 2020.

[92] F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui, "Digital twin-driven product design, manufacturing and service with big data," *Int. J. Adv. Manuf. Technol.*, vol. 94, nos. 9–12, pp. 3563–3576, Feb. 2018.

[93] H. Zhang, G. Zhang, and Q. Yan, "Digital twin-driven cyber-physical production system towards smart shop-floor," *J. Ambient Intell. Humanized Comput.*, vol. 10, no. 11, pp. 4439–4453, Nov. 2019.

[94] W. Wang, Y. Zhang, and R. Y. Zhong, "A proactive material handling method for CPS enabled shop-floor," *Robot. Comput.-Integr. Manuf.*, vol. 61, Feb. 2020, Art. no. 101849.

[95] C. Zhang, G. Zhou, J. Hu, and J. Li, "Deep learning-enabled intelligent process planning for digital twin manufacturing cell," *Knowl.-Based Syst.*, vol. 191, Mar. 2020, Art. no. 105247.

[96] S. Liu, J. Bao, Y. Lu, J. Li, S. Lu, and X. Sun, "Digital twin modeling method based on biomimicry for machining aerospace components," *J. Manuf. Syst.*, May 2020.

[97] J. Liu, H. Zhou, G. Tian, X. Liu, and X. Jing, "Digital twin-based process reuse and evaluation approach for smart process planning," *Int. J. Adv. Manuf. Technol.*, vol. 100, nos. 5–8, pp. 1619–1634, Feb. 2019.

[98] P. Franciosa, M. Sokolov, S. Sinha, T. Sun, and D. Ceglarek, "Deep learning enhanced digital twin for remote laser welding of aluminium structures," *CIRP Ann. Manuf. Technol.*, vol. 69, no. 1, 2020.

[99] Y. Golizadeh Akhlaghi, A. Badiei, X. Zhao, K. Aslansefat, X. Xiao, S. Shittu, and X. Ma, "A constraint multi-objective evolutionary optimization of a state-of-the-art dew point cooler using digital twins," *Energy Convers. Manage.*, vol. 211, May 2020, Art. no. 112772.

[100] C. Liu, L. Le Roux, C. Körner, O. Tabaste, F. Lacan, and S. Bigot, "Digital twin-enabled collaborative data management for metal additive manufacturing systems," *J. Manuf. Syst.*, May 2020.

[101] V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, and M. Riedmiller, "Playing atari with deep reinforcement learning," 2013, *arXiv:1312.5602*. [Online]. Available: http://arxiv.org/abs/1312.5602

[102] H. Van Hasselt, A. Guez, and D. Silver, "Deep reinforcement learning with double q-learning," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 1–7.

[103] T. Schaul, J. Quan, I. Antonoglou, and D. Silver, "Prioritized experience replay," 2015, *arXiv:1511.05952*. [Online]. Available: http://arxiv.org/abs/1511.05952

[104] J. Leng, C. Jin, A. Vogl, and H. Liu, "Deep reinforcement learning for a color-batching resequencing problem," *J. Manuf. Syst.*, vol. 56, pp. 175–187, Jul. 2020.

[105] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 630–645.

[106] J. Lee, M. Azamfar, J. Singh, and S. Siahpour, "Integration of digital twin and deep learning in cyber-physical systems: Towards smart manufacturing," *IET Collaborative Intell. Manuf.*, vol. 2, no. 1, pp. 34–36, Mar. 2020.

[107] M. Tomko and S. Winter, "Beyond digital twins—A commentary," *Environ. Planning B, Urban Analytics City Sci.*, vol. 46, no. 2, pp. 395–399, Feb. 2019.
[108] F. Tao, M. Zhang, Y. Liu, and A. Y. C. Nee, "Digital twin driven prognostics and health management for complex equipment," *CIRP Ann.*, vol. 67, no. 1, pp. 169–172, 2018.
[109] G.-B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.
[110] A. Coraddu, L. Oneto, F. Baldi, F. Cipollini, M. Atlar, and S. Savio, "Data-driven ship digital twin for estimating the speed loss caused by the marine fouling," *Ocean Eng.*, vol. 186, Aug. 2019, Art. no. 106063.
[111] J. Tang, C. Deng, and G.-B. Huang, "Extreme learning machine for multilayer perceptron," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 4, pp. 809–821, Apr. 2016.
[112] P. Jain, J. Poon, J. P. Singh, C. Spanos, S. R. Sanders, and S. K. Panda, "A digital twin approach for fault diagnosis in distributed photovoltaic systems," *IEEE Trans. Power Electron.*, vol. 35, no. 1, pp. 940–956, Jan. 2020.
[113] W. Li, M. Rentemeister, J. Badeda, D. Jöst, D. Schulte, and D. U. Sauer, "Digital twin for battery systems: Cloud battery management system with online state-of-charge and state-of-health estimation," *J. Energy Storage*, vol. 30, Aug. 2020, Art. no. 101557.
[114] A. Piros, L. Trautmann, and E. Baka, "Error handling method for digital twin-based plasma radiation detection," *Fusion Eng. Design*, vol. 156, Jul. 2020, Art. no. 111592.
[115] M. G. Kapteyn, D. J. Knezevic, D. B. P. Huynh, M. Tran, and K. E. Willcox, "Data-driven physics-based digital twins via a library of component-based reduced-order models," *Int. J. Numer. Methods Eng.*, Jun. 2020.
[116] Y. Ye, Q. Yang, F. Yang, Y. Huo, and S. Meng, "Digital twin for the structural health management of reusable spacecraft: A case study," *Eng. Fract. Mech.*, vol. 234, Jul. 2020, Art. no. 107076.
[117] K. P. Murphy, "Dynamic Bayesian networks: Representation, inference and learning, dissertation," Ph.D. dissertation, Dept. Comput. Sci., Univ. California, Berkeley Berkeley, CA, USA, 2002.
[118] P. E. Leser, J. E. Warner, W. P. Leser, G. F. Bomarito, J. A. Newman, and J. D. Hochhalter, "A digital twin feasibility study (Part II): Non-deterministic predictions of fatigue life using *in-situ* diagnostics and prognostics," *Eng. Fract. Mech.*, vol. 229, Apr. 2020, Art. no. 106903.
[119] H. Zhang, Q. Yan, and Z. Wen, "Information modeling for cyber-physical production system based on digital twin and automationml," *Int. J. Adv. Manuf. Technol.*, pp. 1–19, Mar. 2020.
[120] Z. Liu, W. Chen, C. Zhang, C. Yang, and H. Chu, "Data super-network fault prediction model and maintenance strategy for mechanical product based on digital twin," *IEEE Access*, vol. 7, pp. 177284–177296, 2019.
[121] W. Booyse, D. N. Wilke, and S. Heyns, "Deep digital twins for detection, diagnostics and prognostics," *Mech. Syst. Signal Process.*, vol. 140, Jun. 2020, Art. no. 106612.
[122] H. Kim, C. Jin, M. Kim, and K. Kim, "Damage detection of bottom-set gillnet using artificial neural network," *Ocean Eng.*, vol. 208, Jul. 2020, Art. no. 107423.
[123] W. Luo, T. Hu, C. Zhang, and Y. Wei, "Digital twin for CNC machine tool: Modeling and using strategy," *J. Ambient Intell. Humanized Comput.*, vol. 10, no. 3, pp. 1129–1140, Mar. 2019.
[124] W. Luo, T. Hu, Y. Ye, C. Zhang, and Y. Wei, "A hybrid predictive maintenance approach for CNC machine tool driven by digital twin," *Robot. Comput.-Integr. Manuf.*, vol. 65, Oct. 2020, Art. no. 101974.
[125] X. Song, T. Jiang, S. Schlegel, and D. Westermann, "Parameter tuning for dynamic digital twins in inverter-dominated distribution grid," *IET Renew. Power Gener.*, vol. 14, no. 5, pp. 811–821, Apr. 2020.
[126] S. K. Andryushkevich, S. P. Kovalyov, and E. Nefedov, "Composition and application of power system digital twins based on ontological modeling," in *Proc. IEEE 17th Int. Conf. Ind. Informat. (INDIN)*, vol. 1, Jul. 2019, pp. 1536–1542.
[127] D. Gong, J. Sun, and Z. Miao, "A set-based genetic algorithm for interval many-objective optimization problems," *IEEE Trans. Evol. Comput.*, vol. 22, no. 1, pp. 47–60, Feb. 2018.
[128] M. Zhou, J. Yan, and D. Feng, "Digital twin framework and its application to power grid online analysis," *CSEE J. Power Energy Syst.*, vol. 5, no. 3, pp. 391–398, 2019.
[129] C. C. Lee, "Fuzzy logic in control systems: Fuzzy logic controller. II," *IEEE Trans. Syst., Man, Cybern.*, vol. 20, no. 2, pp. 419–435, Apr. 1990.
[130] J. Morton, T. A. Wheeler, and M. J. Kochenderfer, "Analysis of recurrent neural networks for probabilistic modeling of driver behavior," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 5, pp. 1289–1298, May 2017.
[131] X. Chen, C. Wu, T. Chen, H. Zhang, Z. Liu, Y. Zhang, and M. Bennis, "Age of information aware radio resource management in vehicular networks: A proactive deep reinforcement learning perspective," *IEEE Trans. Wireless Commun.*, vol. 19, no. 4, pp. 2268–2281, Apr. 2020.
[132] L. Zhao, G. Han, Z. Li, and L. Shu, "Intelligent digital twin-based software-defined vehicular networks," *IEEE Netw.*, vol. 34, no. 5, pp. 178–184, Sep. 2020.
[133] B. R. Barricelli, E. Casiraghi, J. Gliozzo, A. Petrini, and S. Valtolina, "Human digital twin for fitness management," *IEEE Access*, vol. 8, pp. 26637–26664, 2020.
[134] O. Mazumder, D. Roy, S. Bhattacharya, A. Sinha, and A. Pal, "Synthetic PPG generation from haemodynamic model with baroreflex autoregulation: A digital twin of cardiovascular system," in *Proc. 41st Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2019, pp. 5024–5029.
[135] N. K. Chakshu, J. Carson, I. Sazonov, and P. Nithiarasu, "A semi-active human digital twin model for detecting severity of carotid stenoses from head vibration—A coupled computational mechanics and computer vision method," *Int. J. Numer. methods Biomed. Eng.*, vol. 35, no. 5, p. e3180, 2019.
[136] F. Laamarti, H. Faiz Badawi, Y. Ding, F. Arafsha, B. Hafidh, and A. El Saddik, "An ISO/IEEE 11073 standardized digital twin framework for health and well-being in smart cities," *IEEE Access*, vol. 8, pp. 105950–105961, 2020.
[137] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *Amer. Statistician*, vol. 46, no. 3, pp. 175–185, Aug. 1992.
[138] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[139] M. Pengnoo, M. Taynnan Barros, L. Wuttisittikulkij, B. Butler, A. Davy, and S. Balasubramaniam, "Digital twin for metasurface reflector management in 6G terahertz communications," *IEEE Access*, vol. 8, pp. 114580–114596, 2020.
[140] R. Zhang, Y. Yang, W. Wang, L. Zeng, J. Chen, and S. McGrath, "An algorithm for obstacle detection based on YOLO and light filed camera," in *Proc. 12th Int. Conf. Sens. Technol. (ICST)*, Dec. 2018, pp. 223–226.
[141] G. Schrotter and C. Hürzeler, "The digital twin of the city of Zurich for urban planning," *PFG, J. Photogramm., Remote Sens. Geoinf. Sci.*, pp. 1–14, Feb. 2020.
[142] H. Lehner and L. Dorffner, "Digital geoTwin Vienna: Towards a digital twin city as Geodata Hub," *PFG, J. Photogramm., Remote Sensing Geoinformat. Sci. volume*, vol. 88, pp. 63–75, 2020.
[143] J. Döllner, "Geospatial artificial intelligence: Potentials of machine learning for 3D point clouds and geospatial digital twins," *PFG, J. Photogramm., Remote Sens. Geoinformation Sci.*, pp. 1–10, 2020.
[144] X. Tong, Q. Liu, S. Pi, and Y. Xiao, "Real-time machining data application and service based on IMT digital twin," *J. Intell. Manuf.*, vol. 8, pp. 1–20, Oct. 2019.
[145] A. M. Lund, K. Mochel, J. Lin, R. Onetto, J. Srinivasan, P. Gregg, J. E. Bergman, K. D. Hartling, A. Ahmed, and S. Chotai, "Digital system and method for managing a wind farm having plurality of wind turbines coupled to power grid," U.S. Patent 10 132 295, Nov. 20, 2018.
[146] T. Shah, S. Govindappa, P. Nistler, and B. Narayanan, "Digital twin system for a cooling system," U.S. Patent 9 881 430, Jan. 30, 2018.
[147] H. Wang, "Digital twin based management system and method and digital twin based fuel cell management system and method," U.S. Patent 10 522 854, Dec. 31, 2019.
[148] J. E. Hershey, F. W. Wheeler, M. C. Nielsen, C. D. Johnson, M. J. Dell'Anno, and J. Joykutti, "Digital twin of twinned physical system," U.S. Patent App. 15 087 217, Oct. 5, 2017.
[149] Z. Song and A. M. Canedo, "Digital twins for energy efficient asset maintenance," U.S. Patent App. 15 052 992, Aug. 25, 2016.
[150] C. J. Yates, M. Stankiewicz, J. Alexander, and C. Softley, "Industrial safety monitoring configuration using a digital twin," U.S. Patent App. 16 189 116, May 14, 2020.
[151] T. Masuda, B. Kim, and S. Shiraishi, "Proactive vehicle maintenance scheduling based on digital twin simulations," U.S. Patent App. 15 908 768, Aug. 29, 2019.
[152] H. Goldfarb, A. Pandey, and W. Yan, "Feature selection and feature synthesis methods for predictive modeling in a twinned physical system," U.S. Patent App. 15 350 665, May 17, 2018.

[153] J. Zimmerman, C. Dodd, and M. Peterson, "Methods and systems for generating a patient digital twin," U.S. Patent App. 15 635 805, Jan. 3, 2019.
[154] S. Nagesh, "X-ray tube bearing failure prediction using digital twin analytics," U.S. Patent 10 524 760, Jan. 7, 2020.
[155] M. Peterson, "Surgery digital twin," U.S. Patent App. 15 711 786, Mar. 21, 2019.
[156] L. G. E. Cox, C. P. Hendriks, M. Bulut, V. Lavezzo, and O. van der Sluis, "Digital twin operation," U.S. Patent App. 16 704 495, Jun. 11, 2020.
[157] K. Fischer and M. Heintel, "Examining a consistency between reference data of a production object and data of a digital twin of the production object," U.S. Patent App. 15 750 538, Aug. 9, 2018.
[158] M. G. Burd and P. F. McLaughlin, "Integrated digital twin for an industrial facility," U.S. Patent App. 15 416 569, Jul. 26, 2018.
[159] K. Deutsch, S. Pal, R. Milev, and K. Yang, "Contextual digital twin runtime environment," U.S. Patent 10 564 993, Feb. 18, 2020.
[160] S. Shiraishi, Z. Jiang, and B. Kim, "Digital twin for vehicle risk evaluation," U.S. Patent App. 16 007 693, Dec. 19, 2019.
[161] S. Shiraishi and Y. Zhao, "Sensor-based digital twin system for vehicular analysis," U.S. Patent App. 15 925 070, Sep. 19, 2019.
[162] A. Yousif, A. Ayyagari, D. T. Kirkland, E. C. Owyang, J. Apanovitch, and T. W. Anstey, "Aircraft communications system with an operational digital twin," U.S. Patent App. 16 100 985, Feb. 13, 2020.
[163] Y. Park, S. R. Sinha, V. Venkiteswaran, V. S. Chennupati, and E. S. Paulson, "Building system with digital twin based data ingestion and processing," U.S. Patent 10 854 194, Dec. 1, 2020.
[164] Q. Min, Y. Lu, Z. Liu, C. Su, and B. Wang, "Machine learning based digital twin framework for production optimization in petrochemical industry," *Int. J. Inf. Manage.*, vol. 49, pp. 502–519, Dec. 2019.
[165] S. Bangsow, *Tecnomatix Plant Simulation*. Springer, 2015.
[166] A. Glikson, "Fi-Ware: Core platform for future Internet applications," in *Proc. 4th Annu. Int. Conf. Syst. Storage*, 2011.
[167] A. Bosch Rexroth. *Indramotion Mtx*. Accessed: 2010. https://www.boschrexroth.com/en/us/products/product-groups/electric-drives-and-controls/topics/cnc/indramotion-mtx-standard-performance-and-advanced/index
[168] T. White, *Hadoop: The Definitive Guide*. Newton, MA, USA: O'Reilly Media, 2012.
[169] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, and M. Kudlur, "TensorFlow: A system for large-scale machine learning," in *Proc. 12th USENIX Symp. Operating Syst. Design Implement. (OSDI)*, 2016, pp. 265–283.
[170] F. Seide and A. Agarwal, "CNTK: Microsoft's open-source deep-learning toolkit," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2016, p. 2135.
[171] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, Nov. 2014, pp. 675–678.
[172] A. Gulli and S. Pal, *Deep Learning With Keras*. Birmingham, U.K.: Packt, 2017.
[173] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Explor. Newslett.*, vol. 11, no. 1, pp. 10–18, 2009.
[174] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba, "OpenAI gym," 2016, *arXiv:1606.01540*. [Online]. Available: http://arxiv.org/abs/1606.01540
[175] Y. Duan, X. Chen, R. Houthooft, J. Schulman, and P. Abbeel, "Benchmarking deep reinforcement learning for continuous control," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1329–1338.

M. MAZHAR RATHORE (Member, IEEE) received the master's degree in computer and communication security from the National University of Sciences and Technology, Pakistan, in 2012, and the Ph.D. degree in computer science and engineering from Kyungpook National University, South Korea, in 2018. He is currently working as a Postdoctoral Researcher with the College of Science and Engineering, Hamad Bin Khalifa University, Qatar. His research interests include big data analytics, the Internet of Things, smart systems, network traffic analysis and monitoring, remote sensing, smart city, urban planning, intrusion detection, and information security and privacy. He is a professional member of ACM. He received the Best Project/Paper Award in the 2016 Qualcomm Innovation Award at Kyungpook National University, for his paper "IoT-Based Smart City Development Using Big Data Analytical Approach." He was also a nominee for the Best Project Award in the 2015 IEEE Communications Society Student Competition, for his Project "IoT-Based Smart City." He is serving frequently as a Reviewer for various IEEE, ACM, Springer, and Elsevier journals.

SYED ATTIQUE SHAH received the Ph.D. degree from the Institute of Informatics, Istanbul Technical University, Istanbul, Turkey. During his Ph.D. degree, he was a Visiting Scholar with National Chiao Tung University, Taiwan, The University of Tokyo, Japan, and the Tallinn University of Technology, Estonia, where he completed the major content of his thesis. He was an Assistant Professor and the Chairperson of the Department of Computer Science, BUITEMS, Quetta, Pakistan. He is currently a Lecturer with the Data Systems Group, Institute of Computer Science, University of Tartu, Estonia. His research interests include big data analytics, cloud computing, information management, and the Internet of Things.

DHIRENDRA SHUKLA is currently a Professor and the Dr. J Herbert Smith ACOA Chair in technology management and entrepreneurship of the University of New Brunswick (UNB), Canada. He utilizes his expertise from the telecom sector and extensive academic background in the areas of entrepreneurial finance, masters of business administration, and engineering, to promote a bright future for New Brunswick. Recognition of his tireless efforts and vision are demonstrated through the UNB's 2014 Award from Startup Canada as the "Most Entrepreneurial Post-Secondary Institution of the Year," his nomination as a Finalist for the Industry Champion by KIRA, and his nomination as a Finalist for the Progress Media's Innovation in Practice Award. He was nominated for the RBC Top 25 Canadian Immigrant Award and selected by a panel of judges as a Top 75 finalist. Most recently, he received the Entrepreneur Promotion Award by Startup Canada in 2017, as well as the Outstanding Educator Award by the Association of Professional Engineers and Geoscientists of New Brunswick in 2018.

ELMAHDI BENTAFAT received the bachelor's and M.Sc. degrees in computer science from the Ecole Nationale Supérieure d'Informatique, Algeria, in 2012 and 2016, respectively. He is currently pursuing the Ph.D. degree with the College of Science and Engineering, Hamad Bin Khalifa University, Qatar. His research interests include applied cryptography, privacy, information security, and network security.

SPIRIDON BAKIRAS (Member, IEEE) received the B.S. degree in electrical and computer engineering from the National Technical University of Athens, in 1993, the M.S. degree in telematics from the University of Surrey, in 1994, and the Ph.D. degree in electrical engineering from the University of Southern California, in 2000. He is currently an Associate Professor with the College of Science and Engineering, Hamad Bin Khalifa University, Qatar. Before that, he held teaching and research positions at Michigan Technological University, The City University of New York, The University of Hong Kong, and The Hong Kong University of Science and Technology. His current research interests include security and privacy, applied cryptography, mobile computing, and spatiotemporal databases. He is a member of ACM. He was a recipient of the U.S. National Science Foundation (NSF) CAREER Award.