Journal of Industrial Information Integration 35 (2023) 100483

Contents lists available at ScienceDirect

Journal of Industrial Information Integration

journal homepage: www.elsevier.com/locate/jii

Review article

Towards big data driven construction industry

Fangyu Li [a,b,c,d,*], Yuanjun Laili [e], Xuqiang Chen [a,b,c,d], Yihuai Lou [f], Chen Wang [g], Hongyan Yang [a,b,c,d], Xuejin Gao [a,b,c,d], Honggui Han [a,b,c,d]

[a] *Faculty of Information Technology, Beijing University of Technology, Beijing, 100124, China*
[b] *Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing University of Technology, Beijing, 100124, China*
[c] *Engineering Research Center of Digital Community, Ministry of Education, Beijing University of Technology, Beijing, 100124, China*
[d] *Beijing Artificial Intelligence Institute, Beijing University of Technology, Beijing, 100124, China*
[e] *School of Automation Science and Electrical Engineering, Beihang University, Beijing, 100191, China*
[f] *Center for Hypergravity Experimental and Interdisciplinary Research, College of Civil Engineering and Architecture, Zhejiang University, Hangzhou, 310058, Zhejiang, China*
[g] *National Engineering Research Center for Big Data Software, Tsinghua University, Beijing, 100084, China*

ARTICLE INFO

Keywords:
Construction
Big data
Data analytics
Data engineering

ABSTRACT

The construction industry is currently going through an intelligent revolution. The profound transformation of the Industry 4.0 era is made possible by contemporary technologies such as Internet of Things (IoT), cloud computing, and robotics. Essentially, the vast amount of diverse big data from many sources should be properly utilized to enhance the entire life-cycle construction process. Construction efficiency can be enhanced while material waste and construction expenses are reduced, planning and decision-making processes can be improved while errors are lowered, and applications of big data in construction analytics will make construction sites safer. This article not only offers a comprehensive review of the advantages of associated big data approaches, but it also assesses the current state of the art in the construction industry. Several unresolved difficulties are also discussed. In the end, we express our thoughts on the potential future of big data in the construction industry.

1. Introduction

Predicted by *Statista*, the Big Data market in 2021 and 2022 is anticipated to increase in value by US $30 billion, making it one among the most highly prized commodities worldwide. According to a study by *NewVantage*, 97.2% of businesses spend money on big data and artificial intelligence (AI), and a survey by *Sage* found that 57% of construction companies seek access to dependable financial and project data. Furthermore, according to a *BARC* analysis, using big data enhances a company's chances of coming up with superior strategic decisions by 69%. Data endow great productivity to different disciplinary businesses and industries, which generates opportunities to make significant progresses [1]. The construction industry is seeing a significant increase in data generation and utilization, which indicates a digital era of the construction industry has come [2]. Big data can be used to inform project planning and design, monitor construction progress, and optimize operations and maintenance. By collecting and analyzing data on past projects, construction companies can use predictive analytics to identify potential risks and opportunities for improvement; Sensors can be used to monitor equipment performance, worker productivity, and environmental conditions in real-time; By using data analytics to optimize the construction supply chain, companies can reduce waste, improve logistics, and increase efficiency; By collecting and analyzing data on building performance, construction companies can optimize energy usage, reduce operating costs, and improve sustainability. Overall, big data is becoming increasingly important in the construction industry as companies seek to improve project outcomes, reduce costs, and increase efficiency.

However, the digitization revolution of construction companies is often still in the early stages. According to an FMI study [3], 95.5% of all data captured in 2018 goes unused when it comes to engineering and construction. Digitization and information & communication technology (ICT) provide a huge amount of promise [4–7]. The associated Big Data will bring significant benefits to the construction industry, because Big Data would make the construction processes in the construction sites more efficient and safe, and additionally, it will provide beneficial services and resources for every stage of the construction project life cycle in Fig. 1, including project planning, management, prefabrication manufacture, and the building process.

* Corresponding author at: Faculty of Information Technology, Beijing University of Technology, Beijing, 100124, China.
E-mail address: fangyu.li@bjut.edu.cn (F. Li).

https://doi.org/10.1016/j.jii.2023.100483
Received 8 December 2022; Received in revised form 23 April 2023; Accepted 14 June 2023
Available online 28 June 2023
2452-414X/

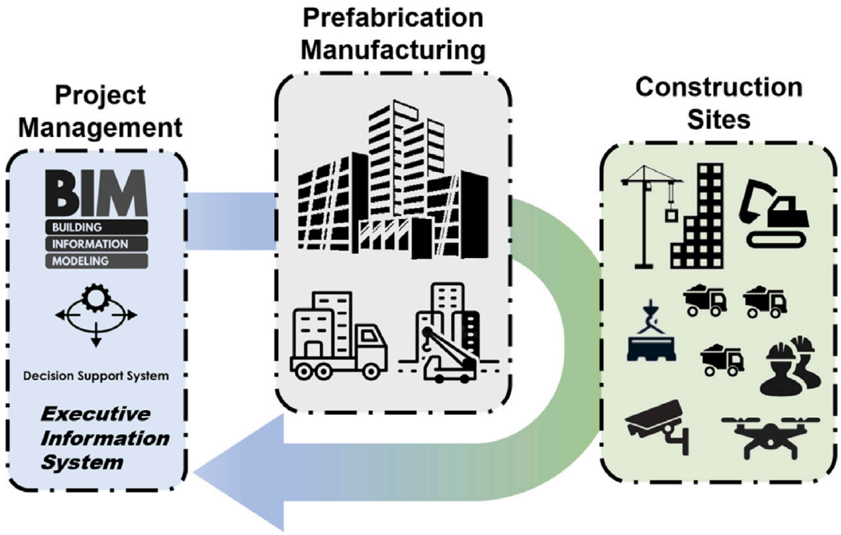

Fig. 1. Big data are produced throughout the duration of an entire construction project and can be advantageous for the industry in the other way round.

Besides common-sense construction processes including the prefabrication manufacturing and construction sites, construction planning and management software, such as BIM (building information modeling), EIS (executive information system), DSS (decision support system), also produce big data. The sources can be summarized as simulation dataset, experimental dataset, organization dataset, company dataset, government dataset, project field survey, project monitoring dataset, and unstructured sources such as social networks, statistic yearbooks, and crawler data from websites. The main benefits brought from big data include but not limit to building efficiency improvement, environmental impact reduction, collaboration promotion, building and infrastructure sustainability improvement and so on.

Many industrial applications, such as large-scale distributed device, long-lasting production, company operation, production supply chain, and external cooperation sources, can provide industrial big data [8–11]. For example, energy Internet [12], distributed modeling [13,14], sensor network monitoring [15], high dimensional process monitoring [16], and so on, generate big data. For modern industries in the Industry 4.0 era, data are being generated by all kinds of isolated and networked machines and devices, cloud-based solutions, business planning and management, etc. Big data is a distinguishing feature of the current generation of intelligent construction because IoT (Internet of Things) uses widespread sensors and microprocessors throughout the whole construction site, generating a massive volume of data that is well above that of traditional size. And the data volume is expected to be continuously increasing in the next decades. Due to the streaming nature of data sources, construction data is also dynamic. However, the term "big data" does not simply mean "a big amount" of data containing abundant information, which would be called "very large data" or "massive data". Essentially, "big data" stands for the data which are generally unstructured, heterogeneous, and therefore extremely complex to deal with. Fortunately, information sharing, processing, and application are becoming more organized as a result of digitization and standardization of many sorts of data formats and types. IoT solutions including mobile, pervasive computing, and smart devices have also seen huge increases in popularity along with the big data industry's explosive growth.

Real-time or nearly real-time analysis is frequently needed for industrial big data. Therefore, it is challenging even for the commercial database software and data analytics software tools to efficiently and effectively capture, store, manage, and analyze the data sets [17]. The construction industry – particularly intelligent construction – generates enormous volumes of data every day, which keep growing with information on everything from building models and designs to communications and cost prediction as well as management. Because of the often unstructured data which could be difficult to access without the appropriate tools, harnessing big data in construction is important. Furthermore, AI paradigms are also driving the digital revolution of the traditional construction towards intelligent construction [18]. Therefore, it is essential to research and create big data approaches for the construction industry, including engineering and analytics methods and strategies [19,20]. For example, big data algorithm can produce a promotion in construction project quality and reduce the incidence of quality problems [21]; multi-party construction project can be improved by a big data service platform [22]; the capacity of the construction companies can be assessed using a predictive and prescriptive big data platform [23].

Data analytics in the construction industry contributes to many different aspects, including building design, construction cost management, energy consumption prediction and pattern identification, material performance prediction, safety management, cloud framework establishment, decision making and control system, and other areas [2]. In the other way, the vast data gathered on the progress of construction projects during their entire life cycles, machinery conditions, worker activities, material positions, vehicle trajectories, energy consumption, weather conditions, and so on can improve the AI models [19], providing wiser and broader insights into the construction industry. Utilizing the data from construction projects to optimize design, build, and run the next wave of industrial innovation. It becomes possible to develop better scheduling and planning strategies as well as efficient construction rules and techniques could be derived and developed to solve the potential issues and improve the efficiency. Besides, thanks to the more streamlined data and information flows, it is also feasible to benefit from the decision-making and strategy development for the management of construction projects. To sum up, data analytics and AI contribute to parts of the construction process.

Construction is a fairly fragmented procedure with an ad hoc organizational structure and non-linear workflow, in contrast to the assembly line approach used in manufacturing. Tasks do not typically link in a straight line. Contrarily, shared resources build connections for the activity in between or within tasks and other activity. Because subcontractors usually receive a variety of projects and frequently have variable degrees of information literacy, getting precise information from them is challenging for general contractors and project owners. Conflicting information flows cause participants to perceive the project differently, and their coordination consumes a significant amount of labor and resources, which is detrimental to project management and team organization [24]. As a result, unstructured data, which are data

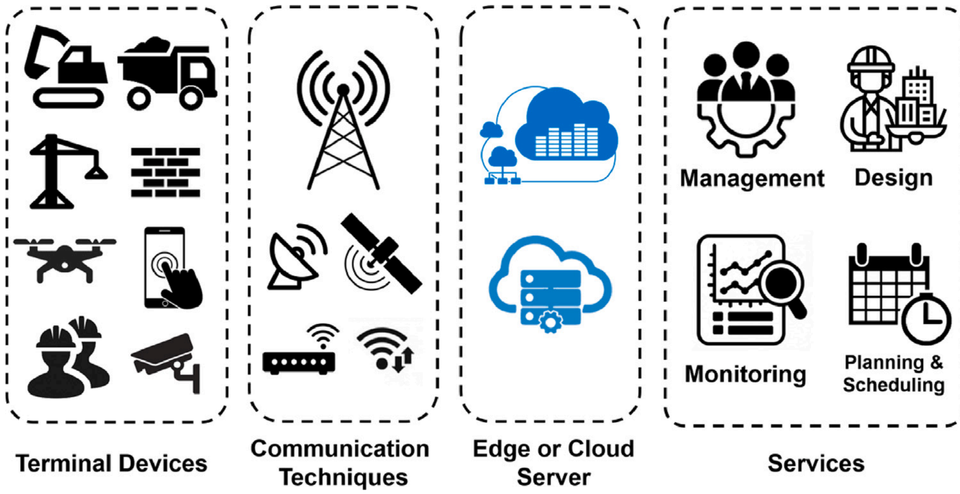

Fig. 2. Big data generation scenarios in the construction process, covering terminal devices, communication techniques, edge/cloud servers, and associated services.

that include information but lack a clear structure, must be handled by data analytics throughout the construction process, including images, emails, plans, websites and reports [25]. Fig. 2 shows the typical big data generation scenarios, including the pervasive sensors installed on the machinery used in the construction sites or terminal devices used in the whole construction process, communication techniques such as satellite, WiFi, RoLa, etc., edge or cloud servers, and all kinds of construction services, including management, design, monitoring, planning, scheduling and so on.

As far as we are aware, this is the first comprehensive, cutting-edge study of how big data paradigm has affected every stage of the construction industry's life cycle in the new era of digitization. Although there have been works describing big data analytics [10,17,19], existing surveys have mainly focused on certain isolated aspects of the construction process. In terms of different construction perspectives, researchers have provided overviews on BIM storage [26], IoT intelligent management [27], construction site management [5], prefabricated building [28], CPS (cyber physical system) based construction organization [24], data mining in construction [2], bibliometric analysis [29], and so on. This article aims to help readers have a better understanding of the role of big data engineering and analytics in modern construction by observing the sources, analytics tools, and benefits.

This survey's remaining sections are organized as follows. In Section 2, we first introduce the data generation possibilities for the different stages of the construction project life cycle. We discuss the associated big data engineering and analytics approaches and procedures in Section 3. The cutting-edge advantages of big data in the construction sector are illustrated in Section 4. Section 5 discusses the current problems and difficulties with big data in terms of the construction applications. Section 6 concludes the survey.

2. Big data generated from construction process

Big data is a broad term for a complex and substantial collection of data, which needs advanced engineering strategies and analytics systems to process, store and manage. Big data also includes a data transformation flow and a data security architecture in addition to the 5Vs (volume, velocity, variety, value, and veracity) [30]. In the construction industry, not just in construction sites big data are produced, but also other related procedures within the entire construction process, consisting of the modeling, designing, planning, scheduling, and management. Therefore, various sources in the construction industry provide data from a variety of structured and unstructured formats, including cameras, sensors, wearables, mobile devices, designs, log and management files, and so forth. We can only significantly improve the construction process and reduce loss and waste if information is shared throughout the design, manufacture, transportation, assembly, construction, and maintenance phases.

2.1. Construction modeling and planning

BIM (Building Information Modeling) was developed and derived from computer-aided design (CAD) [31]. However, BIM is a more comprehensive concept right now, which assists a construction project throughout its entire life cycle. Generally speaking, BIM offers a virtual model and the necessary data about buildings. Planning, design, building, and operation are the four steps that BIM typically supports. Phase planning, site study, cost budgeting, and status quo modeling are all involved in the planning step; Design process includes assessment of design modeling, cost accounting, structural analysis, performance analysis, resource analysis, energy analysis, pipeline and data integration. In the construction stage, model building, site simulation, construction simulation, dynamic collision detection, construction optimization, design coordination, monitoring and adjustment are implemented by BIM; In the operation stage, BIM includes the study of building systems, maintenance schedule, equipment management, record model, space management, disaster warning, and document production. As a result, BIM is defined as "sharing of knowledge resources for information about a facility, providing a trustworthy foundation for decisions throughout its life-cycle; BIM exists from the earliest conception to the demolition of a construction project" by US National Institute of Building Science [32].

BIM has become more than a modeling and simulation tool. It offers a digital replica representation of the building under construction, which seems similar to "digital twin" [33]. Through the interactions between physical and cyber worlds, engineering project monitoring, modeling, and decision-making are now possible with the help of BIM. At the component level, BIM may provide a highly accurate representation of a project by including geometric, topological, and metadata features [34]. Specifically, BIM models make it possible for building blueprints to be quickly accessed via digital devices and offer a way to track the current status of construction projects in real time [35]. As geographical data acquisition and retrieval advance, Integration of geographical and semantic data is related to BIM on the many stages of the construction process, including land planning, cadastral survey, and other applications related to geographic information systems (GIS). In the construction site, BIM acts as a real-time standard safety planning and automated hazard checker to automatically identify and prevent construction worker fall hazards on the construction site [36]. The amount of available BIM data is readily growing. BIM models are designed to contain a significant amount of data details from multiple sources and disciplines such as architecture, engineering, and construction. BIM models of large buildings, complex infrastructures, and multi-disciplinary projects which contain high-level of details occupy large storage spaces. In addition, BIM models are often used for collaboration and coordination among project stakeholders, such as architects, engineers, contractors, and building owners, which requires sharing and updating large amounts of data between multiple parties. BIM files of a building model can easily exceed 50 GB [37] or 100 GB [38] in size. Big data management and analytics are therefore crucial.

A full collection of BIM data is produced that includes physical model and attribute details for the whole life cycle, from design, survey, and construction through to operation and maintenance. How to store, integrate, and use BIM models has become a key issue because of the increased volume of data and the longer period of the visualization preparation of the BIM model. The ground, underground, inside, outdoor, building, street, real-time, history, and prediction data are all included in the BIM big data. Input, acquisition, modification, and integration of information are done by project participants throughout different phases of the construction project life cycle, leading to information exchange as well as integration. The arrangement, storage, and maintenance of BIM data should all be done with high efficiency [26]. It is necessary and urgent to handle large BIM files more efficiently, allowing for greater collaboration and faster decision-making in the design and construction process.

2.2. Construction site sensing

Instinctively, construction site stands for the construction. Multidimensional information about a construction site can be collected by IoT sensors. With the use of Industry 4.0 technology, it is possible to create a "connected" construction site that integrates off-site, onsite, and post-construction activities to increase productivity, uphold sustainability, and enhance worker safety [39]. Onsite IoT network brings all kinds of possibilities in terms of construction site monitoring [5] and the possibility to build an IoT empowered intelligent construction monitoring system for prefabricated buildings [28]. Distributed and smart sensors have been deployed in the construction sites. For example, the RFID (radio-frequency identification) tags can be implanted into prefabricated construction materials, then the construction quality control can be achieved and the data can be integrated with BIM for more convenient search and positioning [40]. Project managers employ a variety of advanced instruments, including video equipment, facial recognition technology, RFID technology, wireless sensor technology, and terminal location devices, to achieve dynamic site supervision. As an illustration, dynamically grasp the surroundings of the construction site, the state of personnel, equipment, and vehicles having access to it, and the attendance status of the labor force. Project employees can quickly spot issues, correct deviations, and make sure the project is running smoothly by examining critical data information in the BIM model [41].

In addition, risk perception is critical for workers' safety. Data collected from wearable sensors is used to improve the worker safety. For instance, wearable sensors have made it possible to evaluate the health and happiness of employees in a personalized, objective manner [42]. And, when a construction worker is actively working, physiological data can be acquired from them [43]. In essence, there are an enormous number of significant use cases that make the collecting, processing, and analysis of big data on construction sites necessary.

3. Big data approaches and methods

The construction industry is being transformed by big data, which has been known for being slow to accept new technology. This change will boost productivity, collaboration, worker safety, and material waste reduction while lowering risks. Big data paradigm efficiently process multi-sourced and diverse data and extract valuable information to analyze and improve the existing workflow. Big data makes it possible for engineering and construction companies to gather and analyze data on costs, site-based transactions, photos, conversations, changes to plans, and more. There are literally thousands of bits of data created for each project in the construction business. Many processes are either not tracked at all, or only when reporting is done using paper-based document sheets. Without digital technology, it is practically impossible to locate crucial data items that would allow for an immediate response to anticipated issues or the application of successful outcomes to future projects, such as the construction industry's use of data mining [2]. Nowadays, information, knowledge and models of the real world are usually represented in a virtual system. Other elements of the virtual world can be software, data-driven models, data mining, AI models, and various simulation and processing techniques [44]. In this context, AI approaches are more and more preferred by practitioners for data processing and analytics to support decision-making and provide feedback to the physical construction system and projects [45,46]. The virtual twin can imitate and manage the real system, improve a procedure, and foresee problems that are not yet emerged in the physical system [47,48].

Big data engineering and analytics are the two primary functions of big data, as shown in Fig. 3. Big data engineering includes critical steps to acquire, regularize and manage the semi-structured and unstructured data with or without interference. Acquisition, processing, storage, database and pipeline construct the main components of big data engineering. Big data analytics are used to categorize, characterize, consolidate, predict, infer, and classify data in order to produce useful information. Prediction, classification, clustering, inference, and optimization can be summarized as the core operations, whereas statistics, data mining, expert knowledge, machine learning, and deep learning are frequently utilized techniques [49,50].

3.1. Big data engineering

Big data are advantageous to the construction industry, highlighting the importance of data engineering. The streaming and heterogeneity of big data are characteristics of the construction process. The continuous large data streaming flows are frequently used and processed over time in the construction industry. Because big data is produced in a streaming manner, it needs to be properly structured and stored to facilitate data analytics. Data engineering, including acquisition, processing, storage, database and pipeline, serves as the foundation for effective data analysis by arranging the raw data into a form that is both analysis- and store-friendly. The tools for data engineering have recently improved, making them more broadly accessible. The engineering of big data should meet the following criteria from the perspective of data analysis: (1) accessibility needs to be implemented for different types of data; (2) massive data needs to be stored efficiently; (3) the engineering system needs to be capable of scalability; (4) the data needs to be controlled globally. From data acquisition to the whole pipeline, big data engineering is a systematic task, as shown in Fig. 4.

The data that is used in the construction process is collected from multi-modal sources and transmitted using a variety of methods and protocols, according to the data acquisition perspective. Besides the different types of sensor data, global positioning system (GPS) [40], RFID [51], location-based service (LBS) [52], etc. are often used. After being acquired from many sources, the heterogeneous data is brought into a unified process. Both cable and wireless technologies, including WiFi, ultra-wideband (UWB), etc., would be used to transfer the local data. The Open Platform Communications United Architecture (OPC-UA), the data exchange standards for industrial communication (M2M or PC-to-Machine connection), is used to manage the acquired data by programmable logic controllers (PLC), remote terminal units (RTU), and other devices [53]. All forms of data will eventually be managed by the supervisory control and data acquisition (SCADA) server. The

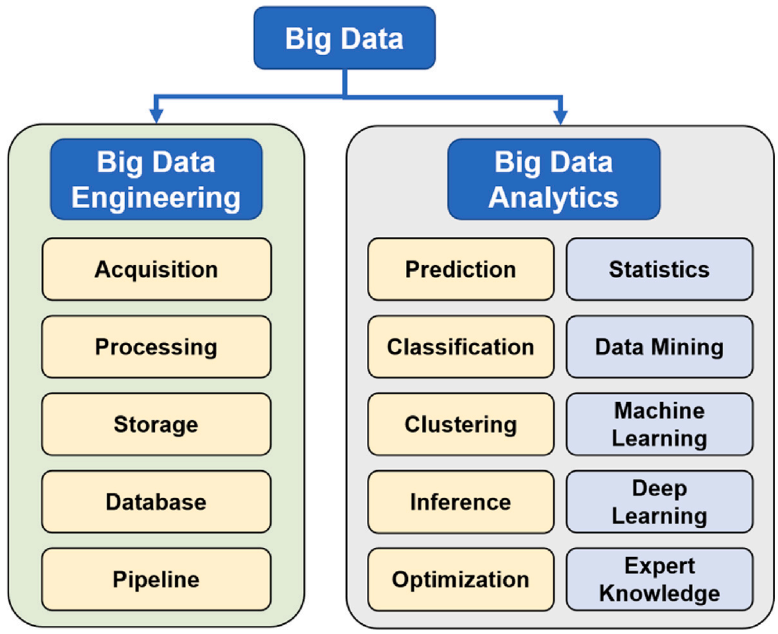

Fig. 3. Big data paradigm components in the upcoming digital construction industry.

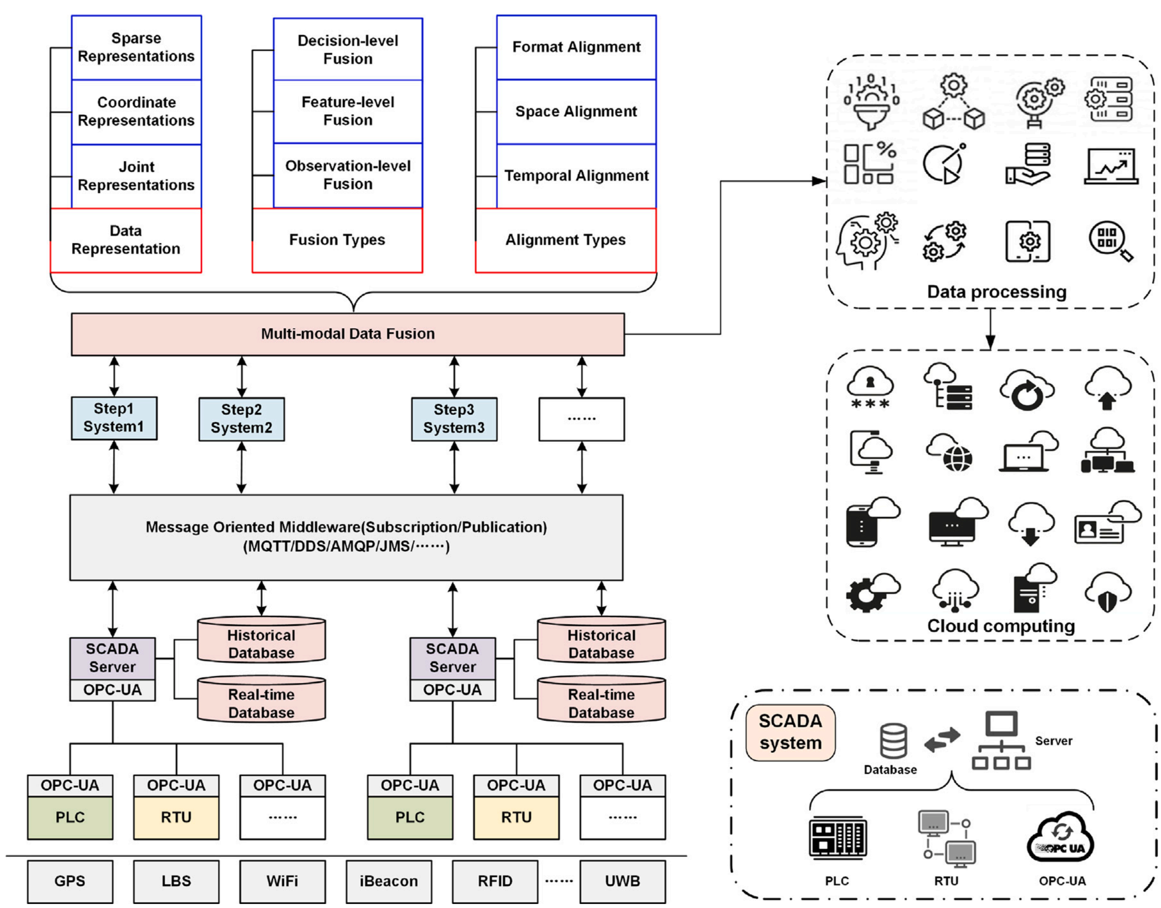

Fig. 4. Big data engineering diagram covering from data acquisition to cloud computing.

SCADA system is essential to data engineering's data collecting process (as shown in the bottom right corner of Fig. 4). This system, which manages many kinds of big data in the construction industry, includes sensor readings like velocity, temperature, frequency data, and so on. It is a centralized system that keeps a check on and regulates the entire data collecting process. Data are continuously recorded and then transformed into live and historical data by connecting and monitoring the dispersed devices throughout the entire site and throughout the entire construction process. Separate databases are used to hold these two types of data. While live real-time data are utilized to fine-tune the pre-trained system model to better synchronize the instantaneous system responses or detect abnormal events, historical data are used to construct empirical models for system analysis. The resilience and sensitivity of the system model in terms of maintenance and responses are improved by the merging of historical and real-time data.

Message-oriented middleware (MOM), a crucial element of Industrial Internet of Things (Industrial IoT), is specifically created to flexibly integrate distributed applications and services that are running on heterogeneous computing and communication devices [54,55], which also improves the interoperability among peer-to-peer communication parties. (Message Queuing Telemetry Transport (MQTT), Data Distribution Service (DDS), Advanced Message Queuing Protocol (AMQP), Java Message Service (JMS), and other open message protocols are frequently used by MOM for M2M communications.) The MOM connects with and regulates the systems or sub-systems that set up the entire construction process, providing services like identifying, authentication, permission, and security.

Data representation, distinct level fusion, and data alignments are used to realize the data fusion with diverse modalities. Data presentations extract and create the standard presentations for the integrated data analysis using the multi-modal data features. In general, data presentation includes coordinate representation based on spatial and distribution information [56,57], sparse representations based on dictionary learning [58,59], joint representation based on graphs and cross-media data [60,61], and so on. Data fusion focuses on a dynamic multi-level and multi-faceted data analysis of big data originating from heterogeneous sources into a unified architecture, in order to transform various types of information into a format that is simpler to handle [62]. Combining data from diverse data sources, data fusion is a technique for detecting, correlating, estimating, and combining data and information from several sources [63]. Careful data fusion can reduce the inconsistent and redundant data that can come from several database descriptions of the same attribute. Observation-level fusion merges data straight from the source; For feature-level fusion, representative features need to be first extracted from the raw sensor data; Only when a first round determination by each terminal device independently has been made is decision-level fusion used. Data alignment, which includes fundamental format alignment [64] as well as advanced space alignment [65] and temporal alignment [66], is another method for bridging the varied semantic gaps between various data sources. It is possible to efficiently extract and evaluate the information and features after the multi-modal data are uniformly aligned.

Data processing is required upon the formatted data, due to the noise, missing data, and inconsistent nature of real-world data. Processing data to increase quality is crucial because low-quality data yield low-quality information. The primary methods for preparing data include cleaning, transformation, and discretization [67]. Data cleaning purifies the original data by removing noise and redundant information, closing gaps, identifying and treating outliers, and smoothing noise. A smaller-volume, identically analytic representation of the original data collection is produced by data reduction. Examples of data reduction methods include data compression and dimensionality reduction [68]. Discretization and data transformation incorporate or modify the original data to enhance the mining process' efficiency and the ability to understand the patterns that are extracted. Examples of data transformation techniques include normalization, aggregating, smoothing, and attribute building [69].

Because of cloud computing, users leverage Internet to instantly access a shared pool of flexible resources. The appropriate cloud-based construction management platform can handle and process tasks related to building design, costs, safety compliance, and accounting more swiftly. Development of an information management platform that manages all facets of a project's operation for big data is thus required for project management in the construction process. The cloud platform is a great technical and data resource for the clever administration of engineering projects due to its potential to be enhanced and combined. For instance, using big data cloud calculations to increase computing capacity might be a great approach to do so without having to worry about setting up or maintaining a system. Edge computing has also had an impact in addition to the cloud, particularly when processing requires only a portion of the data to be transmitted to the cloud [70]. In today's intelligent integration of engineering projects, such as construction projects, the cloud and networks are commonly used. The development of cutting-edge information technology and the ubiquitous availability of high-speed networks theoretically make intelligent building project management via cloud platforms feasible. Therefore, as we progress to Industry 4.0's cloud-based distributed and real-time big data processing, it is no longer appropriate to employ traditional data analytics methodologies for intelligent construction. The bulk of AI algorithms and open source tools struggle with scalability, usability, extensibility, and generalization capabilities when facing big data problems [71].

Therefore, the pipeline is an essential skill for building a complete and fluent workflow. For distributed computing, a popular and highly fault tolerant programming model is map-reduce [72]. Big data algorithms and libraries can be created using platforms like Hadoop [73] or Spark [74] in order to extract useful information from massive data collections. It becomes essential to develop methods for predicting big data parameters, specify strategies for their structure, aggregation, testing, and storage, as well as understand how formats relate to their streaming. The strategic advantage is the capacity to use the results of big data analysis to identify hidden trends, comprehend them, and take appropriate action. While contrasting the various stream processing platforms, such as Apache Storm, Flink, Kafka Streams, or Samza [75], we may also choose platforms that allow processes to run continuously, allowing new data to be processed as soon as it arrives. Public clouds and private servers can both use the technology. Amazon Lambda (also known as AWS Lambda), one of the most well-known cloud services, enables communication with other AWS services and the execution of code written in a variety of different languages, including Node.js, Python, C#, or Java. It is built using Amazon CloudWatch [76], which enables system monitoring and alterations in response to changes. With Azure Data Factory, Microsoft offers a service that is comparable to this. Access to their own big data and cloud storage services is available through both the Google Cloud Platform [77] and Oracle Cloud [78], and now other significant corporations are starting to follow the trend.

3.2. Big data analytics

Construction industry under the big data paradigm can profit a lot from the data analysis tools. In the construction industry, as shown in Fig. 3, prediction, classification, clustering and inference are often utilized data analytics techniques. Descriptive, diagnostic, predictive, and prescriptive analytics are the most used types of data analysis, but outlier identification is rarely utilized [2]. It is feasible to have an overall and occasionally real-time prediction and monitoring during the life cycle of the building sector by using several data analytics methodologies. Fig. 5 shows that various analytics methods could be used in the construction business to uncover latent information in big data.

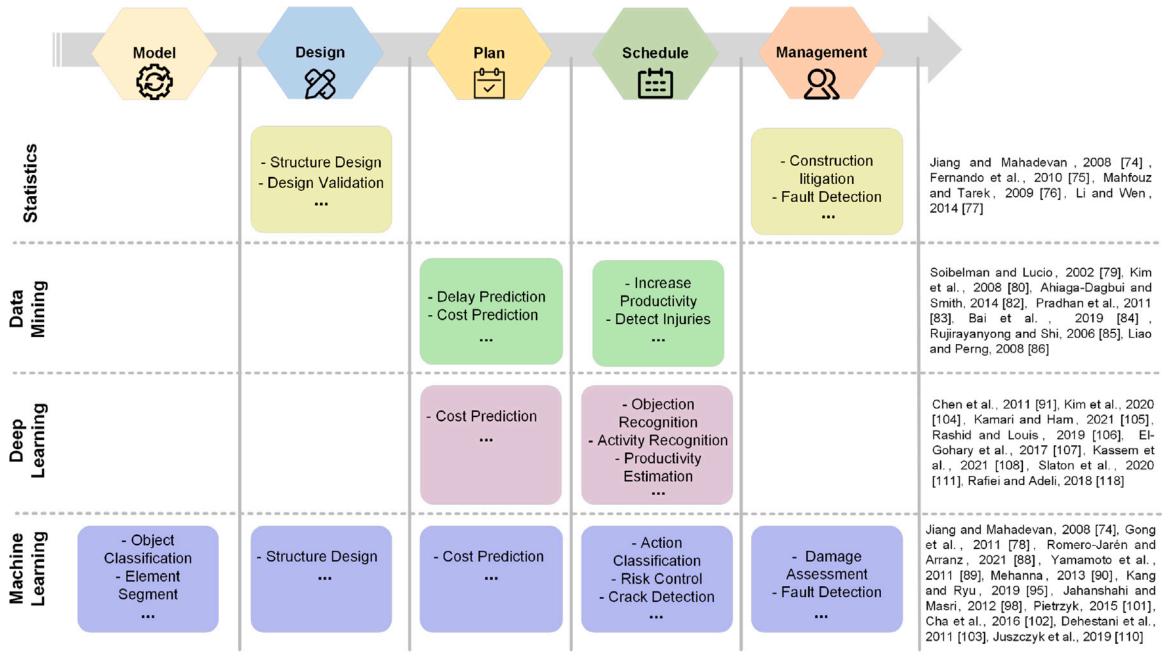

Fig. 5. Big data analytics techniques in the model, design, plan, schedule and management stages of a construction project.

The whole construction process can take advantages of big data analytics. For instance, whether a project will succeed or fail depends greatly on how well the plan is executed and how well the project information is understood. Most construction projects eventually cost more than anticipated attributed to schedule issues, poor planning, and a lack of knowledge in the early stages of the project. Each employee's output is inevitably decreased as a result, which also decreases project productivity as a whole. Because the main structural systems, major building techniques, and the majority of construction materials are chosen during this stage, cost planning during the design phase is essential for the successful completion of a construction project [79]. The mastery of project information regarding the operational status of the construction process is also crucial for risk management, schedule modification, staff restructuring, etc. Furthermore, big data analytics methods must be used to create a fully complete picture of the operating scenario. In terms of methodology, the construction industry has made extensive use of statistics, data mining, machine learning, and deep learning. Therefore, we shall discuss such methods in the subsequent paragraphs.

Statistics is the most conventional way for solving problems since it employs inherent information in a rigorous and efficient manner. Trends and patterns at the data level can be analyzed and used to guide decisions regarding the building process, depending on the statistical nature of the data. A manner of interpretable model is also offered by statistics. In the vast majority of instances, the management of the construction industry makes use of statistical methodologies, while the use of statistical methods for construction design exists [80,81]. For instance, the statistical approach was utilized to help legal decision-making in construction lawsuits [82]. Li and Jin [83] proposed a fault detection system based on pattern matching and principal component analysis (PCA) methods. Additionally, statistical techniques were used to schedule the construction delay as well as to track the movements of heavy equipment and workers [84].

Data mining combines vast amounts of data from various sources and finding hidden patterns or laws. For the sake of analytics, data mining deconstructs a huge batch of data's inherent linkages. For planning and scheduling in the construction industry, mining the information in big data frequently yields more accuracy. Data mining helps the construction industry prepare by allowing companies to uncover knowledge that has been hidden in databases, improving estimates of building costs [85] and delays [86]. Buchheit et al. [87] offered an illustration of the knowledge discovery procedure using data mining methods for an "intelligent" construction. Data mining was used by Dagbui et al. [88] to extract the information included in the construction data and predict the cost of building. Kim et al. [86] chose the most important causes of the construction delay and used data mining to analyze it. Data mining tools increasingly place a greater emphasis on construction productivity during scheduling. By using data mining, Pradhan et al. [89] made it possible to track the productivity of the construction industry and identify areas for improvement. Bai et al. [90] applied intelligent data mining methods to assess the effective productivity of construction equipment. Additionally, data mining was employed to increase construction productivity. The integration of construction data for increasing productivity was accomplished by Rujirayanyong and Shi [91]. Meanwhile, there are other aspects such as injury detection [92] that data mining could make a difference and explore.

Machine learning is a type of technology that can learn from experience. Machine learning is receiving more attention than ever before because of its capacity to learn as the evaluation of industry data grows easier and easier. Machine learning is extensively employed throughout the whole construction life cycle because of its great generalization capabilities. BIM modeling is essential for the success of the entire construction process. The digital construction process is made easier by machine learning thanks to BIM capability. For instance, each object in the Industry Foundation Classes (IFC) model might be categorized into the appropriate categories using data-driven classifications [93].

A method to automatically segment and categorize BIM objects from point cloud data was put forth by Romero-Jarén [94]. Machine learning contributes to a more thorough BIM model ling. Machine learning algorithms could identify phenomena that we can only evaluate from a statistical perspective rather than an intuitionistic perspective throughout the design and construction. For instance, one of the elements that could benefit from data-driven analytics is structural design. The researcher can design the construction of the building while taking into account its benefits and drawbacks by analyzing several data sets from various fields [95,96]. To find a more compact structure for tensegrity structures, Yamamoto et al. [95] recommended using a genetic algorithm. Be aware that machine learning is a useful tool for planning building projects as well. A model-based planning system in BIM is presented by Chen et al. [97]. using genetic algorithms to determine the best crew assignment configuration. In a manner similar to this, Shin et al. [98] likewise got the best layout for the hoist. To forecast the cost of the building, Hwang et al. [99] created a dynamic regression model. Machine learning is used in construction industry planning to help find the best solution given the resource constraints. Natural frequencies, damping ratios, and mode shapes are important model properties for monitoring the development of complex structures and the structural health of those systems. These data were used to characterize and identify dynamic structural features using very efficient Bayesian algorithms [100]. Additionally, machine learning can be used in daily tasks such as construction scheduling to improve efficiency. Using a random forest model to reduce risks, Kang and Ryu [101] forecast workplace accidents. Kale and Baradam [102] used logistic regression to study the modeling of construction injuries. For identifying bridge deterioration, Liu and Jiao [103] devised a genetic algorithm-support vector machine technique. And the crack of bridge can be detected more precisely using machine learning algorithms [104]. On-site construction projects can be supported by data-driven solutions based on automated earthwork planners and cost prediction tools, which ensures the effective use of vehicles and route planning during on-site building and results in considerable time and cost savings [105]. Similarly to the above, Ardeshir et al. [106] developed a fuzzy inference system for the risk assessment of health, safety, and the environment (HSE). In the management phase of the construction industry, machine learning algorithms also play a unique role on the tasks like damage assessment [80,107] and fault detection [108,109]. Machine learning is capable of handling such problems that need to excavate information from data as above.

Deep learning is a cutting-edge framework that incorporates many aspects of the human brain. Data's inherent patterns and hierarchical representations are obtained using deep learning. Deep learning is now frequently used as a powerful monitoring and prediction tool. Such a task, requiring significantly more non-linear and non-anthropogenic than ever before, can be handled by deep learning. Deep learning's distinctive capabilities, including object recognition and detection [110, 111], cost calculation, activity recognition [112], productivity estimation [113,114], safety performance, etc., come into play during the scheduling step.

A unique framework that recognizes worker behaviors and determines if they fall within the parameters of their certification for safety control was introduced by Fang et al. [115]. A deep neural network approach by Juszczyk et al. [116] enables for cost forecasts of a particular kind of construction object. Slaton et al. [117] used a model made up of convolutional neural networks (CNN) and long short-term memory networks (LSTM) layers to identify the activities of heavy machinery. This kind of AI technology might be used to manage swarms of robots and self-driving vehicles to increase productivity on construction sites. All of the aforementioned tasks can only be completed when sensory data streams are networked to data warehouses, analytics platforms, and platforms for developing insights and subsequently using by humans, robots, and automation devices. Many studies also used big data analytics technologies to pinpoint the contributing elements or causes of accidents. For instance, Zhang et al. [118] analyzed the reports of the construction accident using natural language processing (NLP). Tixier et al. [119] creatively developed a conceptual framework employing NLP and graph mining to uncover factors related to injuries. In the interim, data analytics techniques can also be used to implement or enhance risk monitoring [120], event prediction [121], and hazard identification [122].

Furthermore, by integrating data from earlier initiatives, integrated analytics will provide more precise cost estimation. For instance, Lowe et al. [123] found that cost estimation using a neural network was superior to cost estimation using more conventional techniques. Rafiei et al. [124] estimated construction costs using deep models while taking into account the effects of external economics (DBM-SoftMax). Bayram et al. [125] developed models utilizing artificial neural networks and radial basis functions for the initial estimation of construction expenses.

For better decision-making, the construction industry also needs to estimate risk in addition to costs. Asadi et al. [105], for instance, investigated the best approach to creating a prediction model that might pinpoint the reasons for building project delays. To forecast project success, Cheng et al. [126] introduced the evolutionary support vector machine inference model (ESIM). These researchers frequently use machine learning and statistical techniques to anticipate the course of construction using integrated data. These data are obtained either by using nonintrusive approaches or by installing sensors on the under-construction buildings in an invasive manner. AI has been used in the construction sector to evaluate historical data for the estimation and forecasting of the compressive strength of concrete in order to accomplish the aforementioned objectives. It has been useful to comprehend how the five fundamental components of concrete – water, cement, metakaolin, fine aggregate, and coarse aggregate – affect the construction's quality [127].

Information modeling can make use of predictability. Unexperienced builders could be guided and helped by an AI-powered system in the absence of highly skilled experts. Alternatively, because of the data-driven paradigm, big data analytics technologies can bring expert knowledge to the construction practice.

4. Big data benefits

For industries, a survey pointed out that those companies using big data had an 8% increase in profit and a 10% reduction in cost [128]. Specifically, McKinsey & Company claimed that using big data analytics in construction can reduce project costs by 5%–10% and shorten project schedules by 10–20% [129]. And using BIM-like digital tools, the time to generate reports has been reduced by 75%, and document transmittal is sped up by 90%. In addition, big data analytics can also improve construction safety [130].

With the decline in the cost of data acquisition, real-time massive data collecting and processing is becoming more and more prevalent. A dynamic relationship exists between information and decision-making. Data technology is commonly used for optimization. Data collection, update, pattern recognition, and correlation will all be automatic. Big data systems are used by AI approaches to provide crucial information and insight even before a project has even started. This allows for the early management of potential concerns including coordination issues on construction sites, conflicts between different disciplines and trades, and even the impact of the weather. The ability to pivot in reaction to data insights may have a major influence on cost reduction and time overruns. For instance, big data and BIM alter the construction industry during the design phase. Data can be gathered and used to help in the design process. When a suitable data analytics platform is in place, large amounts of data may be quickly evaluated and used to uncover probabilities and trends that can help anticipate potential issues that may influence building projects throughout the construction process. Data analytics technology reduces construction delays and material-related costs by providing comprehensible data and early detection of potential structural issues. Because of this, there are less human errors and project managers are able to decide more swiftly and wisely. Big data advantages in construction projects promote collaboration, boost output, speed up construction, lower risk, decrease waste, improve worker safety, and so on.

4.1. Efficiency improvement

Nowadays, simplified activities are frequently used in the construction process. As soon as the operating plan is established, initiatives become simpler to evaluate. Team members are equipped with the knowledge and resources necessary to make decisions, reduce risk, and increase efficiency. One of the major problems that construction industry workers typically encounter is a lack of communication. Fortunately, big data solutions make it simple and accessible for crew members to exchange information. IoT's visual remote monitoring feature boosts management effectiveness while cutting costs. Moreover, by doing this, the likelihood of misunderstanding-related mistakes is reduced, working relationships are improved, and everyone is kept informed in the event of an unexpected change or interruption.

Big data can also help increase the productivity of construction sites and machinery. In order to increase efficiency, sensors are utilized on modern construction sites to collect data about the site and the equipment. The performance and use of instrumented machines can be studied in great detail thanks to the information produced by these devices when they are connected to on-site operational equipment. Sensor data can reveal when construction equipment is idle and in use, allowing contractors to increase fuel efficiency and productivity and determine whether it is more cost-effective to buy, lease, or rent such equipment.

4.2. Risk mitigation

On-site and off-site real-time data analytics can enhance risk identification and evaluation. Drones, BIM, and prefabricated building techniques are examples of recent, crucial technologies that can assist reduce or even completely remove typical risks including inadequate time management, safety issues, and weather. By predicting problems and using smart wearables to send out warnings and alarms in advance, it can provide workers more capabilities [131]. Additionally, it can enable systems that use automated learning to control hazards in real time and on site. Big data analysis from construction projects can be used to identify possible pitfalls and issues. For instance, big data solutions can alert the project team to potential delays, potential weariness, and overall project time and cost overruns by assessing the productivity of key resources like people and equipment. Additionally, it is possible to overlay project-centric data with corporate data to assist spot trends by gathering both structured and unstructured data. Negative trends should not be disregarded because the entire company could be put at unacceptably high risk.

4.3. Waste reduction

35% of construction expenses are made up of waste materials and corrective work [132]. Companies are using analytics tools to reduce material waste as a result of a recent emphasis on lean construction concepts. These technologies give the entire project team access to real-time information, allowing for more effective supply and usage of materials, plant, and equipment. One of the worst pollutants in the world is the construction industry [133]. The building and engineering industries were really responsible for 39% of process-related carbon dioxide emissions in 2018.

More environmentally friendly building materials and construction methods are required as the sector expands. Big data integration in construction can assist in resolving this issue. In order to better accurately predict the materials and energy needed for a future project, BIM technology may use construction data from prior projects. This lessens unnecessary trash generated during construction and enables planners to look into energy-efficient options whenever available. Construction waste and CO_2 emissions are decreased by accurate building predictions. For example, AI and RGB data streams might guarantee that earthworks are finished quickly and effectively. This helps decrease the amount of greenhouse gases discharged into the atmosphere [134].

4.4. Working condition improvement

Compared to experts in other industries, construction workers are more likely to sustain an occupational injury [135]. Wearable sensors are having a significant impact on enhancing working conditions for site staff in addition to offering vital insights on the productivity and efficiency of machinery and equipment. The wearable's biometric sensors allow for the monitoring of worker health in addition to environmental variables that may affect workplace safety. A pleased workforce produces more. Fortunately, the sector is quickly adopting technology such as smart construction devices and safety management software. Through the collection of health and activity data, the detection of safety risks, and the notification of construction teams of safety protocol violations, these devices take advantage of the potential of big data. By providing this knowledge, project managers may be ready for any safety concerns on upcoming projects in addition to protecting the people who use safety technologies.

5. Open issues and challenges

Big data in the construction sector are gathered in a variety of ways in the Industry 4.0 era. Although significant efforts have been made to digitize the construction industry, there are still unresolved problems and associated difficulties, such as those related to data security [136] and data structure [137,138]. We will highlight some current problems, difficulties, and potential future directions in this article.

5.1. Light-weight security assurance

One of the top issues for construction industry management in the sector is information security. Due to the enormous amounts of data they generate, process, and transfer that is privacy- and security-sensitive, the actual building programs themselves make for appealing targets for attacks. However, IoT devices usually lack security and are susceptible to assaults like distributed denial of service (DoS) operations, as a result of the low-cost requirement [139]. Data assurance and security become challenging issues. The following criteria must be met by an approach that is effective for ensuring privacy and security: data and user confidentiality, user access control, service availability, data freshness, integrity protection, forward and backward secrecy ensuring [140]. Since cryptographic techniques are frequently highly heavy with unacceptable computation overhead, to safeguard the security and privacy of IIoT, it is necessary to create lightweight algorithms [141]. The federated learning architecture can be used to design privacy-preserving yet lightweight implementations of complex calculations [142]. Distributed frameworks might be created in addition to the framework to assist the front-end hardware in doing the difficult tasks needed by security services. As another innovative technology, Blockchain essentially works as a distributed, point-to-point ledger of records that is updated in real-time and unconstrained by a single central authority to improve security and privacy [143]. Due to blockchain's decentralization and high efficiency, decentralized devices may track transactions and store data without having to worry about the data being altered.

5.2. Edge intelligence

Thanks to the recent communication technology development, the network would need to concurrently handle a large number of heterogeneous services used in the systems in 5G or even 6G scenarios [144,145]. However, the wide variety of industrial IoT, in this survey especially construction IoT, applications can conflict with the need for extremely dependable transmissions and massively steady connections. Integrating edge computing and creating energy-efficient mechanisms is of utmost relevance in order to satisfy all these demanding needs while keeping a self-sustainable characteristic. High accuracy and low latency are main demands for the smart construction. A new trend is enabling computation at the network edge. For example, the IoT sensors in the construction sites can not only implement the data acquisition, but also have the abilities to accomplish computation tasks for data engineering and analytics. Distributed sensors continue to produce valuable data, and AI makes inferences from it. In essence, the edges with strong computing capabilities bring the AI closer to the end consumers and application scenarios. Local edge servers can improve privacy and security while reducing bandwidth and latency at the edge of IoT networks by utilizing edge AI. The distributed BIM software can use the local edge AI engines to improve the modeling performances without losing the sensitive information. When coordinating heterogeneous resources across several domains, the edge intelligence paradigm that is produced by organically integrating global and local intelligence can always apply AI techniques. Federated learning and blockchain techniques provide the security communication protocols to leverage both local and global sources. The extracted local information or knowledge and even edge intelligence overcomes the traditional limits of model training. In other words, although global intelligence negatively impacts the accuracy and speed of local intelligence, local intelligence really enhances global intelligence. Edge intelligence consequently becomes a crucial concern, necessitating further edge structure design that is tailored for the whole construction process.

5.3. Heterogeneous devices coexisting

In the course of building, a variety of CPS and IoT systems coexist and function, frequently utilizing heterogeneous hardware and software. Critical issues arise as a result of this heterogeneity. Non-linear interactions between the many subsystems. The interactions, particularly the data and information flows, between the diverse subsystems are crucial. Often, the subsystems must work together to maintain a working flow. Workers and managers need to have a comprehensive understanding of the entire building project, which necessitates understanding the subsystems' workflow. Negative interference brought on by diverse subsystems. The restricted resources available to subsystems like the machinery at a building site make it possible for such subsystems to severely interact with one another. Therefore, additional work is needed. Utilizing cross technology communication (CTC), which permits communication between many communication systems, is one potential remedy [146]. System's overall scalability and flexibility are unavailable. Network nodes should be able to join and depart the system at any moment due to its dynamic nature, for instance, a truck leaving a construction site or a new worker joining the job at any time. Thus, ensuring the flexibility and scalability is crucial. Potential solutions include developing adaptive network growing/pruning strategies and incremental service orchestration, which reuses the existing network resources.

6. Conclusion

The modern construction industry and the next wave of digitization require big data processing and analytics in real-world industrial applications during a construction project's lifespan. Big data techniques aid the construction industry by enhancing efficiency and reducing waste by utilizing cutting-edge information technology and data management systems. Modern technologies, like as AI, sophisticated statistical and optimization models, and big data analytics, offer further opportunities for process improvement. The advancement of big data analytics enhances the capacity to track, record, and analyze data to forecast and recommend the best course of action for the management of building projects. Big data applications in construction yet face a number of unresolved problems and difficulties at various levels. We provide a broad overview of big data in construction, focusing on the entire life cycle. We started by outlining the history of large data production in the construction process. Then, topics of data engineering and data analytics are discussed. Additionally, the advantages of big data are also illustrated. Finally, in order to shed light on the contemporary intelligent construction, key issues are emphasized along with future research and application areas. We hope our paper can help researchers and practitioners to better understand the state of the art in the modern construction industry, identify big data related areas for further investigation, and ultimately drive innovations and progresses in the construction industry.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Honggui Han reports financial support was provided by National Key Research and Development Project.

Data availability

No data was used for the research described in the article.

Acknowledgments

The research is supported by National Key Research and Development Project under Grant 2022YFB3305800-5, National Science Foundation of China under Grants 92267107, 62125301, 61890930-5, 61903010, and 62021003, Beijing Outstanding Young Scientist Program under Grant BJJWZYJH01201910005020, and Beijing Natural Science Foundation under Grant KZ202110005009, and Beijing Youth Scholar under Grant No. 037.

References

[1] C.P. Chen, C.-Y. Zhang, Data-intensive applications, challenges, techniques and technologies: A survey on Big Data, Inform. Sci. 275 (2014) 314–347.
[2] H. Yan, N. Yang, Y. Peng, Y. Ren, Data mining in the construction industry: Present status, opportunities, and future trends, Autom. Constr. 119 (2020) 103331.
[3] J. Snyder, A. Menard, N. Spare, Big data: Big questions for the engineering and construction industry, 2018.
[4] Y. Lu, Industry 4.0: A survey on technologies, applications and open research issues, J. Ind. Inf. Integr. 6 (2017) 1–10.
[5] A. Bucchiarone, M. De Sanctis, P. Hevesi, M. Hirsch, F.J.R. Abancens, P.F. Vivanco, O. Amiraslanov, P. Lukowicz, Smart construction: remote and adaptable management of construction sites through IoT, IEEE Internet Things Mag. 2 (3) (2019) 38–45.
[6] P.K.R. Maddikunta, Q.-V. Pham, B. Prabadevi, N. Deepa, K. Dev, T.R. Gadekallu, R. Ruby, M. Liyanage, Industry 5.0: A survey on enabling technologies and potential applications, J. Ind. Inf. Integr. 26 (2022) 100257.
[7] E. Hämäläinen, T. Inkinen, Industrial applications of big data in disruptive innovations supporting environmental reporting, J. Ind. Inf. Integr. 16 (2019) 100105.
[8] Y. Cheng, K. Chen, H. Sun, Y. Zhang, F. Tao, Data and knowledge mining with big data towards smart production, J. Ind. Inf. Integr. 9 (2018) 1–13.
[9] X. Zhou, Y. Hu, W. Liang, J. Ma, Q. Jin, Variational LSTM enhanced anomaly detection for industrial big data, IEEE Trans. Ind. Inform. 17 (5) (2020) 3469–3477.
[10] Z. Lv, H. Song, P. Basanta-Val, A. Steed, M. Jo, Next-generation big data analytics: State of the art, challenges, and future research topics, IEEE Trans. Ind. Inform. 13 (4) (2017) 1891–1899.
[11] A. Al-Abassi, H. Karimipour, H. HaddadPajouh, A. Dehghantanha, R.M. Parizi, Industrial big data analytics: challenges and opportunities, in: Handbook of Big Data Privacy, Springer, 2020, pp. 37–61.
[12] K. Wang, H. Li, Y. Feng, G. Tian, Big data analytics for system stability evaluation strategy in the energy Internet, IEEE Trans. Ind. Inform. 13 (4) (2017) 1969–1978.
[13] D. Wang, F. Li, K. Liu, Modeling and monitoring of a multivariate spatio-temporal network system, IISE Trans. (2021) 1–17.
[14] F. Li, R. Xie, Z. Wang, L. Guo, J. Ye, P. Ma, W. Song, Online distributed IoT security monitoring with multidimensional streaming big data, IEEE Internet Things J. 7 (5) (2020) 4387–4394.
[15] S. Rani, S.H. Ahmed, R. Talwar, J. Malhotra, Can sensors collect big data? An energy-efficient big data gathering algorithm for a WSN, IEEE Trans. Ind. Inform. 13 (4) (2017) 1961–1968.

[16] K. Huang, Y. Wu, C. Wang, Y. Xie, C. Yang, W. Gui, A projective and discriminative dictionary learning for high-dimensional process monitoring with industrial applications, IEEE Trans. Ind. Inform. 17 (1) (2021) 558–568.
[17] S. Yin, O. Kaynak, Big data for modern industry: challenges and trends [point of view], Proc. IEEE 103 (2) (2015) 143–146.
[18] W. Yu, T. Dillon, F. Mostafa, W. Rahayu, Y. Liu, A global manufacturing big data ecosystem for fault detection in predictive maintenance, IEEE Trans. Ind. Inform. 16 (1) (2020) 183–192.
[19] M. Bilal, L.O. Oyedele, J. Qadir, K. Munir, S.O. Ajayi, O.O. Akinade, H.A. Owolabi, H.A. Alaka, M. Pasha, Big Data in the construction industry: A review of present status, opportunities, and future trends, Adv. Eng. Inform. 30 (3) (2016) 500–521.
[20] S.A. Ismail, S. Bandi, Z.N. Maaz, An appraisal into the potential application of big data in the construction industry, Int. J. Built Environ. Sustain. 5 (2) (2018).
[21] D. Wang, J. Fan, H. Fu, B. Zhang, Research on optimization of big data construction engineering quality management based on RNN-LSTM, Complexity 2018 (2018).
[22] X. Zhang, X. Ming, D. Yin, Reference architecture of common service platform for Industrial Big Data (I-BD) based on multi-party co-construction, Int. J. Adv. Manuf. Technol. 105 (5) (2019) 1949–1965.
[23] J. Ngo, B.-G. Hwang, C. Zhang, Factor-based big data and predictive analytics capability assessment tool for the construction industry, Autom. Constr. 110 (2020) 103042.
[24] Z. You, L. Feng, Integration of industry 4.0 related technologies in construction industry: A framework of cyber-physical system, IEEE Access 8 (2020) 122908–122922.
[25] H. Baars, H.-G. Kemper, Management support with structured and unstructured data—an integrated business intelligence framework, Inf. Syst. Manage. 25 (2) (2008) 132–148.
[26] Z. Lv, X. Li, H. Lv, W. Xiu, BIM big data storage in WebVRGIS, IEEE Trans. Ind. Inform. 16 (4) (2020) 2566–2573.
[27] Y. Zhao, Q. Wang, X. Wang, Refined and intelligent management mode of construction project based on BIM and IOT technology, in: The Sixth International Conference on Information Management and Technology, 2021, pp. 1–5.
[28] X. Wang, S. Wang, X. Song, Y. Han, IoT-based intelligent construction system for prefabricated buildings: Study of operating mechanism and implementation in China, Appl. Sci. 10 (18) (2020) 6311.
[29] Y. Lu, J. Zhang, Bibliometric analysis and critical review of the research on big data in the construction industry, Eng. Constr. Archit. Manag. (2021).
[30] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, M. Gidlund, Industrial internet of things: Challenges, opportunities, and directions, IEEE Trans. Ind. Inform. 14 (11) (2018) 4724–4734.
[31] S. Azhar, Building information modeling (BIM): Trends, benefits, risks, and challenges for the AEC industry, Leadersh. Manag. Eng. 11 (3) (2011) 241–252.
[32] H.-M. Chen, K.-C. Chang, T.-H. Lin, A cloud-based system framework for performing online viewing, storage, and analysis on big data of massive BIMs, Autom. Constr. 71 (2016) 34–48.
[33] F. Tao, Q. Qi, Make more digital twins, Nature 573 (7775) (2019) 490–491.
[34] T.D. Oesterreich, F. Teuteberg, Understanding the implications of digitisation and automation in the context of Industry 4.0: A triangulation approach and elements of a research agenda for the construction industry, Comput. Ind. 83 (2016) 121–139.
[35] J.K. Whyte, T. Hartmann, How digitizing building information transforms the built environment, Build. Res. Inf. 45 (6) (2017) 591–595.
[36] S. Zhang, K. Sulankivi, M. Kiviniemi, I. Romo, C.M. Eastman, J. Teizer, BIM-based fall hazard identification and prevention in construction safety planning, Saf. Sci. 72 (2015) 31–45.
[37] M. Bilal, L.O. Oyedele, J. Qadir, K. Munir, O.O. Akinade, S.O. Ajayi, H.A. Alaka, H.A. Owolabi, Analysis of critical features and evaluation of BIM software: towards a plug-in for construction waste minimization using big data, Int. J. Sustain. Build. Technol. Urban Dev. 6 (4) (2015) 211–228.
[38] N. Garyaev, V. Garyaeva, Big data technology in construction, in: E3S Web of Conferences, Vol. 97, EDP Sciences, 2019, p. 01032.
[39] C.J. Turner, J. Oyekan, L. Stergioulas, D. Griffin, Utilizing industry 4.0 on the construction site: Challenges and opportunities, IEEE Trans. Ind. Inform. 17 (2) (2020) 746–756.
[40] N. Pradhananga, J. Teizer, Automatic spatio-temporal analysis of construction site equipment operations using GPS data, Autom. Constr. 29 (2013) 107–122.
[41] S. Gong, X. Gao, Z. Li, L. Chen, Developing a dynamic supervision mechanism to improve construction safety investment supervision efficiency in China: Theoretical simulation of evolutionary game process, Int. J. Environ. Res. Public Health 18 (7) (2021) 3594.
[42] W. Lee, K.-Y. Lin, E. Seto, G.C. Migliaccio, Wearable sensors for monitoring on-duty and off-duty worker physiological status and activities in construction, Autom. Constr. 83 (2017) 341–353.
[43] B. Choi, H. Jebelli, S. Lee, Feasibility analysis of electrodermal activity (EDA) acquired from wearable sensors to assess construction workers' perceived risk, Saf. Sci. 115 (2019) 110–120.
[44] E. Winsberg, Simulated experiments: Methodology for a virtual world, Philos. Sci. 70 (1) (2003) 105–125.
[45] S. Dick, Artificial intelligence, 2019.
[46] A.I. AI, The fourth industrial revolution, in: Presented at the American Psychological Association in Minneapolis, Vol. 4, 2022, p. 6.
[47] F. Tao, H. Zhang, A. Liu, A.Y. Nee, Digital twin in industry: State-of-the-art, IEEE Trans. Ind. Inform. 15 (4) (2018) 2405–2415.
[48] S. Haag, R. Anderl, Digital twin–Proof of concept, Manuf. Lett. 15 (2018) 64–66.
[49] J. Evans, C. Lindner, Business Analytics: The Next Frontier for Decision Sciences, Vol. 21, College of Business, University of Cincinnati, Decision Science Institute, 2012, p. 2018, http://www.cbpp.uaa.alaska.edu/afef/business_analytics.htm. Zugegriffen Am. (12).
[50] G. Wang, A. Gunasekaran, E.W. Ngai, T. Papadopoulos, Big data analytics in logistics and supply chain management: Certain investigations for research and applications, Int. J. Prod. Econ. 176 (2016) 98–110.
[51] L.-C. Wang, Enhancing construction quality inspection and management using RFID technology, Autom. Constr. 17 (4) (2008) 467–479.
[52] G. Malacarne, G. Toller, C. Marcher, M. Riedl, D. Matt, Investigating benefits and criticisms of bim for construction scheduling in SMEs: An Italian case study, Int. J. Sustain. Dev. Plan. 13 (1) (2018) 139–150.
[53] S.-H. Leitner, W. Mahnke, OPC UA–Service-Oriented Architecture for Industrial Applications, Vol. 48, ABB Corporate Research Center, 2006, p. 22, (61–66).
[54] E. Curry, Message-oriented middleware, in: Middleware for Communications, Wiley Online Library, 2004, pp. 1–28.
[55] J. Yongguo, L. Qiang, Q. Changshuai, S. Jian, L. Qianqian, Message-oriented middleware: A review, in: 2019 5th International Conference on Big Data Computing and Communications, BIGCOM, IEEE, 2019, pp. 88–97.
[56] R.A. Andersen, L.H. Snyder, C.-S. Li, B. Stricanne, Coordinate transformations in the representation of spatial information, Curr. Opin. Neurobiol. 3 (2) (1993) 171–176.
[57] F. Zhang, X. Zhu, H. Dai, M. Ye, C. Zhu, Distribution-aware coordinate representation for human pose estimation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 7093–7102.
[58] R. Gribonval, M. Nielsen, Sparse representations in unions of bases, IEEE Trans. Inform. Theory 49 (12) (2003) 3320–3325.
[59] X.X. Zhu, R. Bamler, A sparse image fusion algorithm with application to pan-sharpening, IEEE Trans. Geosci. Remote Sens. 51 (5) (2012) 2827–2836.
[60] H. Liu, T. Li, R. Hu, Y. Fu, J. Gu, H. Xiong, Joint representation learning for multi-modal transportation recommendation, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33, 2019, pp. 1036–1043.
[61] X. Zhai, Y. Peng, J. Xiao, Learning cross-media joint representation with sparse and semisupervised regularization, IEEE Trans. Circuits Syst. Video Technol. 24 (6) (2013) 965–978.
[62] M. Schmitt, X.X. Zhu, Data fusion and remote sensing: An ever-growing relationship, IEEE Geosci. Remote Sens. Mag. 4 (4) (2016) 6–23.
[63] U.S. Department of Defense, Data fusion lexicon: data fusion subpanel of the joint directors of laboratories technical panel for C3, 1991.
[64] G.J. Barton, et al., ALSCRIPT: a tool to format multiple sequence alignments, Protein Eng. Des. Sel. 6 (1) (1993) 37–40.
[65] H. Pang, S. Wei, G. Zhang, S. Zhang, S. Qiu, Y. Zhao, Heterogeneous feature alignment and fusion in cross-modal augmented space for composed image retrieval, IEEE Trans. Multimed. (2022).
[66] T.D. Wang, C. Plaisant, A.J. Quinn, R. Stanchak, S. Murphy, B. Shneiderman, Aligning temporal data by sentinel events: discovering patterns in electronic health records, in: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2008, pp. 457–466.
[67] S. García, S. Ramírez-Gallego, J. Luengo, J.M. Benítez, F. Herrera, Big data preprocessing: methods and prospects, Big Data Anal. 1 (1) (2016) 1–22.
[68] K. Sayood, Introduction to Data Compression, Morgan Kaufmann, 2017.
[69] A.A. Bakar, Z.A. Othman, N.L.M. Shuib, Building a new taxonomy for data discretization techniques, in: 2009 2nd Conference on Data Mining and Optimization, IEEE, 2009, pp. 132–140.
[70] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, D. Sabella, On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration, IEEE Commun. Surv. Tutor. 19 (3) (2017) 1657–1681.
[71] S. Sharma, Expanded cloud plumes hiding Big Data ecosystem, Future Gener. Comput. Syst. 59 (2016) 63–92.
[72] J. Dean, S. Ghemawat, MapReduce: Simplified data processing on large clusters, 2004.
[73] T. White, Hadoop: The Definitive Guide, " O'Reilly Media, Inc.", 2012.
[74] H. Karau, A. Konwinski, P. Wendell, M. Zaharia, Learning Spark: Lightning-Fast Big Data Analysis, " O'Reilly Media, Inc.", 2015.
[75] C. Prakash, Spark streaming vs flink vs storm vs kafka streams vs samza: Choose your stream processing framework, 2018, Why Not Learn Something Blogspot.
[76] J. Varia, S. Mathew, et al., Overview of Amazon Web Services, Vol. 105, Amazon Web Services, 2014.
[77] E. Bisong, An overview of google cloud platform services, in: Building Machine Learning and Deep Learning Models on Google Cloud Platform, Springer, 2019, pp. 7–10.
[78] M.T. Jakóbczyk, Practical Oracle Cloud Infrastructure, Springer, 2020.

[79] J.R. Meredith, S.M. Shafer, S.J. Mantel Jr., Project Management: A Strategic Managerial Approach, John Wiley & Sons, 2017.
[80] X. Jiang, S. Mahadevan, Bayesian probabilistic inference for nonparametric damage detection of structures, J. Eng. Mech. 134 (10) (2008) 820–831.
[81] R. Fernando, R. Drogemuller, F. Salim, J. Burry, Patterns, heuristics for architectural design support: Making use of evolutionary modelling in design, in: A.I. Li, N. Gu, B. Dave, H.J. Park (Eds.), New Frontiers: Proceedings of the 15th International Conference on Computer-Aided Architectural Design Research in Asia, The Association for Computer-Aided Architectural Design Research in Asia, Hong Kong, 2010, pp. 283–292.
[82] T.S. Mahfouz, Construction Legal Support for Differing Site Conditions (DSC) through Statistical Modeling and Machine Learning (ML), Iowa State University, 2009.
[83] S. Li, J. Wen, Application of pattern matching method for detecting faults in air handling unit system, Autom. Constr. 43 (2014) 49–58.
[84] J. Gong, C.H. Caldas, C. Gordon, Learning and classifying actions of construction workers and equipment using Bag-of-Video-Feature-Words and Bayesian network models, Adv. Eng. Inform. 25 (4) (2011) 771–782.
[85] L. Soibelman, H. Kim, Data preparation process for construction knowledge generation through knowledge discovery in databases, J. Comput. Civ. Eng. 16 (1) (2002) 39–48.
[86] H. Kim, L. Soibelman, F. Grobler, Factor selection for delay analysis using Knowledge Discovery in Databases, Autom. Constr. 17 (5) (2008) 550–560.
[87] R.B. Buchheit, J. Garrett, S.R. Lee, R. Brahme, A knowledge discovery case study for the intelligent workplace, 2012, pp. 914–921.
[88] D.D. Ahiaga-Dagbui, S.D. Smith, Dealing with construction cost overruns using data mining, Constr. Manag. Econ. 32 (7–8) (2014) 682–694.
[89] A. Pradhan, B. Akinci, C.T. Haas, Formalisms for query capture and data source identification to support data fusion for construction productivity monitoring, Autom. Constr. 20 (4) (2011) 389–398.
[90] S. Bai, M. Li, R. Kong, S. Han, H. Li, L. Qin, Data mining approach to construction productivity prediction for cutter suction dredgers, Autom. Constr. 105 (2019) 102833.
[91] T. Rujirayanyong, J.J. Shi, A project-oriented data warehouse for construction, Autom. Constr. 15 (6) (2006) 800–807.
[92] C.-W. Liao, Y.-H. Perng, Data mining for occupational injuries in the Taiwan construction industry, Saf. Sci. 46 (7) (2008) 1091–1102.
[93] J. Wu, J. Zhang, New automated BIM object classification method to support BIM interoperability, J. Comput. Civ. Eng. 33 (5) (2019).
[94] R. Romero-Jarén, J.J. Arranz, Automatic segmentation and classification of BIM elements from point clouds, Autom. Constr. 124 (2021) 103576.
[95] M. Yamamoto, B.S. Gan, K. Fujita, J. Kurokawa, A genetic algorithm based form-finding for tensegrity structure, Procedia Eng. 14 (2011) 2949–2956.
[96] R. Mehanna, Resilient structures through machine learning and evolution, in: ACADIA 13: Adaptive Architecture [Proceedings of the 33rd Annual Conference of the Association for Computer Aided Design in Architecture (ACADIA) ISBN 978-1-926724-22-5] Cambridge 24-26 October, 2013), CUMINCAD, 2013, pp. 319–326.
[97] Y.-J. Chen, C.-W. Feng, Y.-R. Wang, H.-M. Wu, Using BIM model and genetic algorithms to optimize the crew assignment for construction project planning, Int. J. Technol. (3) (2011) 179–187.
[98] Y. Shin, H. Cho, K.-I. Kang, Simulation model incorporating genetic algorithms for optimal temporary hoist planning in high-rise building construction, Autom. Constr. 20 (5) (2011) 550–558.
[99] S. Hwang, Dynamic regression models for prediction of construction costs, J. Constr. Eng. Manag. 135 (5) (2009) 360–367.
[100] F. Zhang, H. Xiong, W. Shi, X. Ou, Structural health monitoring of Shanghai Tower during different stages using a Bayesian approach, Struct. Control Health Monit. 23 (11) (2016) 1366–1384.
[101] K. Kang, H. Ryu, Predicting types of occupational accidents at construction sites in Korea using random forest model, Saf. Sci. 120 (2019) 226–236.
[102] Ö.A. Kale, S. Baradan, Identifying factors that contribute to severity of construction injuries using logistic regression model, Tek. Dergi 31 (2) (2020) 9919–9940.
[103] H.-B. Liu, Y.-B. Jiao, Application of genetic algorithm-support vector machine (GA-SVM) for damage identification of bridge, Int. J. Comput. Intell. Appl. 10 (4) (2011) 383–397, Cited by: 42.
[104] M.R. Jahanshahi, S.F. Masri, Adaptive vision-based crack detection using 3D scene reconstruction for condition assessment of structures, Autom. Constr. 22 (2012) 567–576.
[105] A. Asadi, M. Alsubaey, C. Makatsoris, A machine learning approach for predicting delays in construction logistics, Int. J. Adv. Logist. 4 (2) (2015) 115–130.
[106] A. Ardeshir, P. Farnood Ahmadi, H. Bayat, A prioritization model for hse risk assessment using combined failure mode, effect analysis, and fuzzy inference system: A case study in iranian construction industry, Int. J. Eng. 31 (9) (2018) 1487–1497.
[107] K. Pietrzyk, A systemic approach to moisture problems in buildings for mould safety modelling, Build. Environ. 86 (2015) 50–60.
[108] Y.-J. Cha, K. You, W. Choi, Vision-based detection of loosened bolts using the Hough transform and support vector machines, Autom. Constr. 71 (2016) 181–188.
[109] D. Dehestani, F. Eftekhari, Y. Guo, S.S. Ling, S. Su, H.T. Nguyen, Online support vector machine application for model based fault detection and isolation of HVAC system, 2011.
[110] D. Kim, S. Lee, V.R. Kamat, Proximity prediction of mobile objects to prevent contact-driven accidents in co-robotic construction, J. Comput. Civ. Eng. 34 (4) (2020).
[111] M. Kamari, Y. Ham, Vision-based volumetric measurements via deep learning-based point cloud segmentation for material management in jobsites, Autom. Constr. 121 (2021) 103430.
[112] K.M. Rashid, J. Louis, Times-series data augmentation and deep learning for construction equipment activity recognition, Adv. Eng. Inform. 42 (2019) 100944.
[113] K.M. El-Gohary, R.F. Aziz, H.A. Abdel-Khalek, Engineering approach using ANN to improve and predict construction labor productivity under different influences, J. Constr. Eng. Manag. 143 (8) (2017) 04017045.
[114] Measuring and benchmarking the productivity of excavators in infrastructure projects: A deep neural network approach, Autom. Constr. 124 (2021) 103532.
[115] Q. Fang, H. Li, X. Luo, L. Ding, T.M. Rose, W. An, Y. Yu, A deep learning-based method for detecting non-certified work on construction sites, Adv. Eng. Inform. 35 (2018) 56–68.
[116] M. Juszczyk, K. Zima, W. Lelek, Forecasting of sports fields construction costs aided by ensembles of neural networks, J. Civ. Eng. Manag. 25 (7) (2019) 715–729.
[117] T. Slaton, C. Hernandez, R. Akhavian, Construction activity recognition with convolutional recurrent networks, Autom. Constr. 113 (2020) 103138.
[118] F. Zhang, H. Fleyeh, X. Wang, M. Lu, Construction site accident analysis using text mining and natural language processing techniques, Autom. Constr. 99 (2019) 238–248.
[119] A.J.-P. Tixier, M.R. Hallowell, B. Rajagopalan, D. Bowman, Construction safety clash detection: identifying safety incompatibilities among fundamental attributes using data mining, Autom. Constr. 74 (2017) 39–54.
[120] J. Wu, N. Cai, W. Chen, H. Wang, G. Wang, Automatic detection of hardhats worn by construction personnel: A deep learning approach and benchmark dataset, Autom. Constr. 106 (2019) 102894.
[121] B.U. Ayhan, O.B. Tokdemir, Predicting the outcome of construction incidents, Saf. Sci. 113 (2019) 91–104.
[122] Y. Zhao, Q. Chen, W. Cao, J. Yang, J. Xiong, G. Gui, Deep learning for risk detection and trajectory tracking at construction sites, IEEE Access 7 (2019) 30905–30912.
[123] D.J. Lowe, M.W. Emsley, A. Harding, Predicting construction cost using multiple regression techniques, J. Constr. Eng. Manag. 132 (7) (2006-07) 750–758.
[124] M.H. Rafiei, H. Adeli, Novel machine-learning model for estimating construction costs considering economic variables and indexes, J. Constr. Eng. Manag. 144 (12) (2018) 04018106.
[125] S. Bayram, M.E. Ocal, E. Laptali Oral, C.D. Atis, Comparison of multi layer perceptron (MLP) and radial basis function (RBF) for construction cost estimation: the case of Turkey, J. Civ. Eng. Manag. 22 (4) (2016) 480–490.
[126] M.-Y. Cheng, N.-D. Hoang, A.F. Roy, Y.-W. Wu, A novel time-depended evolutionary fuzzy SVM inference model for estimating construction project at completion, Eng. Appl. Artif. Intell. 25 (4) (2012) 744–752.
[127] J.-B. Yu, Y. Yu, L.-N. Wang, Z. Yuan, X. Ji, The knowledge modeling system of ready-mixed concrete enterprise and artificial intelligence with ANN-GA for manufacturing production, J. Intell. Manuf. 27 (4) (2016) 905–914.
[128] Business Application Research Center (BARC), Big Data Use Cases 2015 – Getting Real On Data Monetization, 2016.
[129] McKinsey & Company, Imagining Construction's Digital Future, 2016.
[130] Dodge Data & Analytics, Improving Performance with Project Data - How Improved Collection and Analysis is Leading the Digital Transformation of the Construction Industry, 2019.
[131] J. Barata, P.R.d. Cunha, Safety is the new black: the increasing role of wearables in occupational health and safety in construction, in: International Conference on Business Information Systems, Springer, 2019, pp. 526–537.
[132] M.S. Aslam, B. Huang, L. Cui, Review of construction and demolition waste management in China and USA, J. Environ. Manag. 264 (2020) 110445.
[133] X. Deng, G. Liu, J. Hao, A study of construction and demolition waste management in Hong Kong, in: 2008 4th International Conference on Wireless Communications, Networking and Mobile Computing, IEEE, 2008, pp. 1–4.
[134] O. Moselhi, A. Alshibani, Optimization of earthmoving operations in heavy civil engineering projects, J. Constr. Eng. Manag. 135 (10) (2009) 948–954.
[135] N. Van Tam, N.L. Huong, N.B. Ngoc, Factors affecting labour productivity of construction worker on construction site: A case of Hanoi, J. Sci. Technol. Civ. Eng. (STCE)-HUCE 12 (5) (2018) 127–138.
[136] F. Li, R. Xie, B. Yang, L. Guo, P. Ma, J. Shi, J. Ye, W. Song, Detection and identification of cyber and physical attacks on distribution power grids with pvs: An online high-dimensional data-driven approach, IEEE J. Emerg. Sel. Top. Power Electron. 10 (1) (2022) 1282–1291.
[137] D. Lazer, R. Kennedy, G. King, A. Vespignani, The parable of Google Flu: traps in big data analysis, Science 343 (6176) (2014) 1203–1205.

[138] J. Fan, F. Han, H. Liu, Challenges of big data analysis, Natl. Sci. Rev. 1 (2) (2014) 293–314.
[139] W. Wang, H. Xu, M. Alazab, T.R. Gadekallu, Z. Han, C. Su, Blockchain-based reliable and efficient certificateless signature for IIoT devices, IEEE Trans. Ind. Inform. (2021).
[140] C.-C. Sun, A. Hahn, C.-C. Liu, Cyber security of a power grid: State-of-the-art, Int. J. Electr. Power Energy Syst. 99 (2018) 45–56.
[141] F. Li, Y. Shi, A. Shinde, J. Ye, W.-Z. Song, Enhanced cyber-physical security in internet of things through energy auditing, IEEE Internet Things J. 6 (3) (2019) 5224–5231.
[142] L. Zhao, J. Li, Q. Li, F. Li, A federated learning framework for detecting false data injection attacks in solar farms, IEEE Trans. Power Electron. 37 (3) (2021) 2496–2501.
[143] Y. Lu, X. Huang, Y. Dai, S. Maharjan, Y. Zhang, Blockchain and federated learning for privacy-preserved data sharing in industrial IoT, IEEE Trans. Ind. Inform. 16 (6) (2019) 4177–4186.
[144] S. Li, L. Da Xu, S. Zhao, 5G Internet of Things: A survey, J. Ind. Inf. Integr. 10 (2018) 1–9.
[145] Y. Lu, X. Zheng, 6G: A survey on technologies, scenarios, challenges, and the related issues, J. Ind. Inf. Integr. 19 (2020) 100158.
[146] Y. Chen, M. Li, P. Chen, S. Xia, Survey of cross-technology communication for IoT heterogeneous devices, IET Commun. 13 (12) (2019) 1709–1720.