

# Current State and Future Opportunities of Data Mining for Construction 4.0

K. Wu<sup>a</sup> and B. García de Soto<sup>a</sup>

<sup>a</sup>S.M.A.R.T. Construction Research Group, Division of Engineering, New York University Abu Dhabi (NYUAD), Experimental Research Building, Saadiyat Island, P.O. Box 129188, Abu Dhabi, United Arab Emirates  
E-mail: [keyi.wu@nyu.edu](mailto:keyi.wu@nyu.edu), [garcia.de.soto@nyu.edu](mailto:garcia.de.soto@nyu.edu)

## Abstract –

In the context of Construction 4.0, the data-intensive nature of the AEC/FM industry puts data-related issues at its core. The powerful ability of data mining to process and utilize data makes it the preferred option for solving data-related problems. Currently, there is no summary of data mining applications in terms of Construction 4.0. To figure out the current state and future opportunities of data mining for Construction 4.0, this study conducts a bibliometric analysis with three steps: (1) determining research context and scope, (2) retrieving literature from Web of Science, and (3) modeling and visualizing the word similarity network using VOSviewer. Three main research areas, namely data mining for intelligence, digitalization, and automation, are identified. The main research topics and objects are summarized for each research area. Furthermore, two promising research fields, namely construction robots and construction cybersecurity, are discussed as future opportunities. This study reveals the current body of knowledge concerning the applications of data mining and points out its future development directions in the context of Construction 4.0.

## Keywords –

Knowledge discovery; Digitalization; Automation; Virtualization; Decentralization; Artificial intelligence; BIM; Robot

## 1 Introduction

With the emergence and evolution of digitalization, automation, virtualization, and decentralization technologies, the world is experiencing a new industrial revolution, commonly labeled as Industry 4.0 [1]. The counterpart of Industry 4.0 in the Architecture, Engineering and Construction, and Facilities Management (AEC/FM) industry is known as

Construction 4.0 [2]. The AEC/FM industry has many opportunities to benefit from Construction 4.0, and traditional industry practices are expected to be highly smart to avoid excessive human intervention for achieving concerned targets [3].

The AEC/FM industry is a typical data-intensive domain. Especially, it undergoes rapid growth in terms of data generation and collection in the information age [4]. Effective data utilization can contribute to the AEC/FM industry's added value. In the context of Construction 4.0, data-related issues are central [5]. Due to the adoption of digitalization, automation, virtualization, and decentralization technologies, large-sized, multi-attributed, and unstructured data from diverse information sources (e.g., text, graph, image, audio, video) significantly increases the difficulty of data utilization [6][7]. Data mining is a process of discovering knowledge such as patterns from large data sets, which incorporates multiple fields, including statistics, pattern recognition, and machine learning [8]. Data mining can powerfully deal with a wide variety of data, and the knowledge it discovers can be used for information management, query optimization, decision support, etc. Therefore, data mining is a promising instrument for Construction 4.0.

In recent years, the interest in data mining application status in the AEC/FM industry has increased, and several studies have explored it [4][6][7][9]. The potential of using data mining over traditional basic statistical methods and pure analytical methods to provide quick and useful insights for the AEC/FM industry has been indicated. However, there is a lack of a summary of data mining applications in terms of Construction 4.0. To fill this gap, this study conducts a literature review to identify the current state and future opportunities of data mining for Construction 4.0. The remainder of the paper is structured below. First, each step of the bibliometric analysis is explained. Then, the current state and future opportunities of data mining for Construction 4.0 are discussed. Finally, conclusions and outlook are provided.

Table 1. Search rules used in Web of Science

| Root        | AND | Where (OR)        | AND | When (OR)    | AND | What (OR)                     |
|-------------|-----|-------------------|-----|--------------|-----|-------------------------------|
| Data mining |     | Infrastructure    |     | Planning     |     | Construction 4.0              |
|             |     | Building          |     | Design       |     | Digitalization                |
|             |     | Road              |     | Fabrication  |     | Automation                    |
|             |     | Railway           |     | Construction |     | Virtualization                |
|             |     | Bridge            |     | Operation    |     | Decentralization              |
|             |     | Tunnel            |     | Maintenance  |     | Building information modeling |
|             |     | Water supply      |     | Renovation   |     | Geographic information system |
|             |     | Sewage            |     | Demolition   |     | Digital twin                  |
|             |     | Electrical grid   |     |              |     | Artificial intelligence       |
|             |     | Telecommunication |     |              |     | Big data                      |

## 2 Research Methodology

The bibliometric analysis process consists of three steps, and they are clarified as follows.

The first step is to determine the research context and scope used to identify search keywords and narrow the research focus. From the perspective of technology applications, Construction 4.0 is seen as a means of finding a coherent complementarity between the main emerging technological approaches in the AEC/FM industry [5]. Because digitalization, automation, virtualization, and decentralization technologies are closely related to Construction 4.0, the literature on those technologies falls within the research scope. When studying a technology-related research topic, it is necessary to focus on a recent publication period to highlight the latest developments [10]. Therefore, the publication period of literature is limited to the past ten years (January 2012 to December 2021) to summarize recent technological advances and identify the current state.

The second step is to retrieve literature from Web of Science according to the search rules presented in Table 1. The dimensions data mining could be achieved, namely “Where”, “When”, and “What”, are investigated in parallel using the logical operator “AND”. In the dimension “Where”, ten keywords, such as “Building” and “Railway”, are connected using the logical operator “OR”. In the dimension “When”, eight keywords, such as “Design” and “Construction”, are connected using the logical operator “OR”. In the dimension “What”, 19 keywords, such as “Building information modeling” and “Robot”, are connected using the logical operator “OR”. The initial search returned 3,653 pieces of literature, and

the number was reduced to 274 by filtering to specific categories on Web of Science. The selected categories were Engineering Civil, Construction Building Technology, and Architecture. Fields such as Engineering Mechanical and Engineering Aerospace were excluded.

The third step is to model and visualize the word similarity network using VOSviewer for identifying research areas. The words used for constructing the similarity network are extracted from the titles and abstracts of the retrieved literature [11].

The similarity between two words is calculated using a similarity measure known as the association strength [12], as expressed in Equation (1).  $s_{ij}$  refers to the similarity between words  $i$  and  $j$ ;  $c_{ij}$  refers to the number of co-occurrences of words  $i$  and  $j$ ; and  $w_i$  and  $w_j$  refer to the total number of occurrences of words  $i$  and  $j$ , respectively.

$$s_{ij} = \frac{c_{ij}}{w_i w_j} \quad (1)$$

The similarity of words is visualized by their distance from each other in VOSviewer. The higher the similarity between two words, the shorter their distance. VOSviewer is to minimize a weighted sum of the squared Euclidean distances of all words for mapping and to impose that the average distance between two words equals 1 for avoiding all words having the same location [12], as expressed in Equations (2) and (3), respectively.  $n$  refers to the number of words; and  $\mathbf{x}_i = (x_{i1}, x_{i2})$  refers to the location of word  $i$  in a two-dimensional map.

$$V(\mathbf{x}_1, \dots, \mathbf{x}_n) = \sum_{i < j} s_{ij} \|\mathbf{x}_i - \mathbf{x}_j\|^2 \quad (2)$$

$$\frac{2}{n(n-1)} \sum_{i < j} \| \mathbf{x}_i - \mathbf{x}_j \| = 1 \quad (3)$$

The weight of words is visualized by the size of their corresponding nodes. The larger a word and the corresponding node, the higher the weight of the word. The nodes of words are grouped in clusters of different colors. Words are assigned to clusters by maximizing  $V(c_1, \dots, c_n)$  [13], as expressed in Equation (4).  $n$  refers to the number of words;  $c_i$  and  $c_j$  refer to the clusters to which words  $i$  and  $j$  are assigned, respectively;  $\delta(c_i, c_j)$  refers to a function equaling 1 if  $c_i = c_j$  and 0 otherwise; and  $\gamma$  refers to a resolution parameter determining the detail level of clusters.

$$V(c_1, \dots, c_n) = \sum_{i < j} \delta(c_i, c_j) (s_{ij} - \gamma) \quad (4)$$

### 3 Findings

The word similarity network based on the retrieved literature is presented in Figure 1. The blue (upper portion), red (right side), and green (left side) nodes indicate the words associated with clusters 1, 2 and 3, respectively. It is worth noting that the words clustered in one specific cluster can share similar concepts with another cluster through strong connections, meaning that they are not exclusively isolated from each other. These clusters could be seen as three main research areas in terms of Construction 4.0: (1) data mining for intelligence, (2) digitalization, and (3) automation, as listed in Table 2. The three research areas and future data mining opportunities for Construction 4.0 are interpreted and analyzed below.

### 3.1 Data Mining for Intelligence

Intelligence is defined as the ability to acquire and apply knowledge and skills [14]. In the context of artificial intelligence (AI), it allows to make decisions through data processing and get feedback in a manner similar to human thinking when facing complex and random environments. Data mining can nontrivially extract previously unknown and potentially useful information from data, and the application of data mining is one of the mainstream in intelligence. In the research area of data mining for intelligence, there are four main research topics: descriptive intelligence [17][18], diagnostic intelligence [19][20], predictive intelligence [21][22], and prescriptive intelligence [23][24], as listed in Table 2.

Descriptive intelligence refers to describing data, and it answers the question “What happened in the past?”. Descriptive intelligence can be used to understand situations, performances, levels, and so on [17][18]. Occupant presence status is essential information for the simulation of building energy use. For gathering the actual information on occupant presence, a recognition method was designed using C4.5 Decision Tree (C4.5) and Curve Description (CD) from environmental data and the usage information of light and air conditioning [17]. Bad conditions of roads are one of the causes of fatal traffic accidents, and it is required to monitor the road state and detect road damages to enhance the safety of road traffic. Aiming to characterize the road condition, the classifiers applying Support Vector Machine (SVM) were developed to distinguish different types of road surfaces and obstacles [18].

Diagnostic intelligence refers to explaining data, and it answers the question “Why something happened in the past?”. Diagnostic intelligence can be used to determine the causes of trends, correlations between variables, why

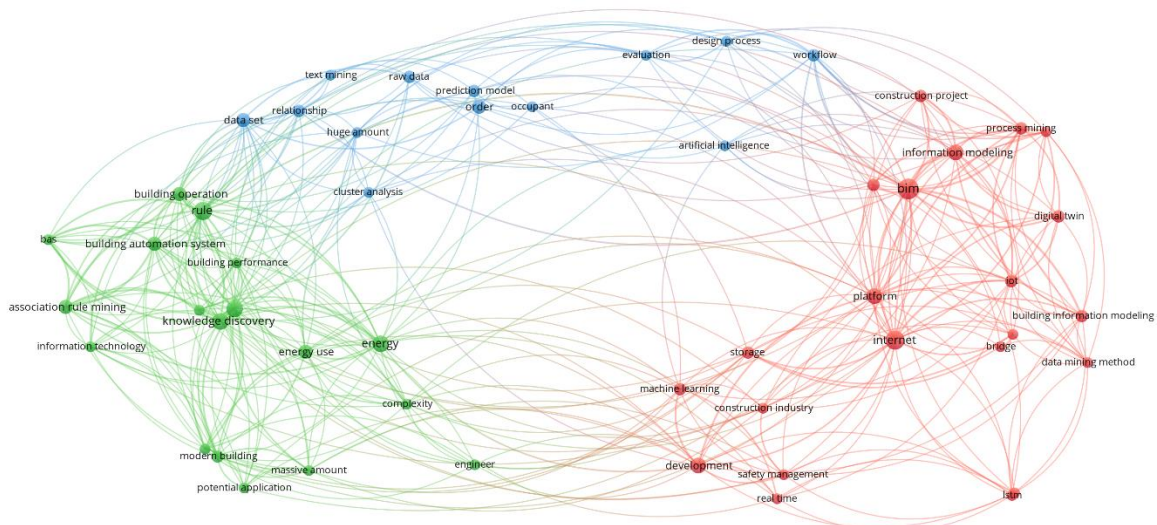


Figure 1. Word similarity network based on the retrieved literature

Table 2. Research areas, topics, and objects identified and summarized from the retrieved literature

| Research area                                | Research topic             | Research object   |
|--|----------------------------|---|
| Data mining for intelligence (blue cluster)  | Descriptive intelligence   | Occupant presence status [17]; Road conditions [18]   |
|  | Diagnostic intelligence    | Building operation behaviors [19]; Electricity load patterns [20]   |
|  | Predictive intelligence    | Rebar amount [21]; Rock mass types [22]   |
| Data mining for digitalization (red cluster) | Prescriptive intelligence  | Sustainable building design decision-marking [23]; Building maintenance management [24]                               |
|  | Digitalization demand      | BIM innovation directions [25]; BIM user needs [26]; BIM manager role viability [27]; BIM labor costs [28]            |
|  | Digitalization modeling    | BIM modeling progress [29]; BIM and GIS schemas mapping [30]  |
| Data mining for automation (green cluster)   | Digitalization management  | BIM-based collaborative design [31]; BIM construction data analytics [32]; BIM-based facility management [33]         |
|  | Automation detection       | Road cracks [34]; Railway tunnel elements [35]  |
|  | Automation equipment       | Panelized home prefabrication facility production [36]; Tunnel boring machine construction safety and efficiency [37] |
|  | Building automation system | Building energy management [38]; BAS alarm management [39]  |

anomalies occur, and so on [19][20]. Discovering underlying building operation data structures and relationships is beneficial to understanding building operation behaviors. In order to realize such abilities, a generic framework with Quantitative Association Rule Mining (QARM) was established, which helps detect and diagnose building operation strategies, non-typical and abnormal building operations, and sensor faults [19]. A considerable amount of real-time electricity consumption data provides a promising way to figure out energy usage patterns and improve building energy management. With this in mind, a general framework integrating Density-Based Spatial Clustering Application with Noise (DBSCAN), K-means, and Classification and Regression Tree (CART) was proposed to extract typical electricity load patterns and discover insightful information hidden in the patterns [20].

Predictive intelligence refers to identifying the likelihood of outcomes based on data, and it answers the question “What is likely to happen in the future?”. Predictive intelligence can be used to forecast unknown future features, activities, trends, and so on [21][22]. The amount estimation of rebar is essential for the cost determination of reinforced concrete structures during the design stage. Taking into account the existing limitations of rebar in 3D modeling, Decision Tree (DT) and Case-Based Reasoning (CBR) were adopted to estimate the amount of rebar in reinforced concrete structures, and the amount can be statistically classified by parameters through generating decision tree nodes [21]. Achieving safe and efficient tunneling needs geological conditions (i.e., rock mass types) ahead of the

tunnel face. To determine rock mass types, Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH), K-means++, and Support Vector Classifier (SVC) were employed based on the operation data of a tunneling boring machine (i.e., cutterhead speed, cutterhead torque, thrust, and advance rate) [22].

Prescriptive intelligence refers to recommending actions to affect likely outcomes based on data, and it answers the question “What is the best course of action?”. Prescriptive intelligence can be used to suggest decision options and show the implication of decision options [23][24]. In sustainable building design, an interplay between multidisciplinary input and the fulfillment of diverse criteria is required for aligning into one high-performing whole. However, design decision-making still relies heavily on rules of thumb. Therefore, the design decision-making based on knowledge discovery in disparate building data was proposed with Multivariate Motif Discovery (MMD) and Temporal Association Rule Mining (TARM) [23]. Building maintenance data such as maintenance requests is a valuable means to assess building performance and gain insights for preventive maintenance actions. To allow facility managers to shrink and limit the area where faults usually occur and determine what building elements and systems are the most problematic, a Text Mining (TM) approach was applied by analyzing textual data contained in unstructured maintenance management systems [24].

### 3.2 Data Mining for Digitalization

Digitalization is defined as the adaptation of a system,

process, etc., to be operated using computers and the internet [15]. Digitalization greatly facilitates data generation and collection, thus driving an unprecedented growth rate of data. Traditional data analysis cannot handle a large amount of data, which promotes the application of data mining in digitalization. In the research area of data mining for digitalization, there are three main research topics: digitalization demand [25][26][27][28], digitalization modeling [29][30], and digitalization management [31][32][33], as listed in Table 2.

To enable digitalization growth, extra demands, such as innovative directions, user needs, personnel roles, and labor costs, are appeared [25][26][27][28]. The awareness of application hotspots and the forecast of development trends can drive BIM innovations. Based on patent analysis of BIM applications, a framework integrating Social Network Analysis (SNA) and Latent Dirichlet Allocation (LDA) was established to identify the technological development and innovation of BIM [25]. The high threshold of domain knowledge and the information asymmetry between developers and users make it difficult for technology developers of BIM applications to fully understand user needs. Combined with domain knowledge, text mining techniques, including Sentiment Analysis (SA) and Topic Modeling (TM), were utilized to capture user needs from BIM app attributes and user comments [26]. In BIM-enabled projects, the BIM manager has emerged as a necessary adjunct role. To test the likelihood of a long-term market demand for the BIM manager as a distinct role, Singular Value Decomposition (SVD) and Frequent Pattern Growth (FP-growth) were conducted according to BIM-related job advertisements [27]. When adopting BIM in a project, additional labor costs need to be involved. To reduce the prediction risk of BIM labor costs, a hybrid approach integrating Random Forest (RF) and Simple Linear Regression (SLR) was used to improve the prediction accuracy of a project's BIM labor costs [28].

Digitalization changes traditional modeling methods, so new concerns are raised from a modeling perspective, including modeling progress and interoperability [29][30]. In building design practices, the close monitoring of modeling processes and the correct measurement of modelers' performance are required. Having an objective measurement system to quantify modeling progress contributes to performance monitoring. Hence, a sequence mining algorithm based on Generalized Suffix Tree (GST) was implemented to identify implicit 3D modeling patterns from unstructured temporal BIM log data [29]. The functionality between BIM and GIS can be enhanced through their interoperability, and data mapping is critical for seamless information sharing between BIM and GIS models. Given the complexity of BIM and GIS schemas, mapping candidates were generated using text mining techniques

such as Cosine Similarity (CS), Market Basket Analysis (MBA), and Jaccard Coefficient (JC) [30].

Digitalization creates a new work mode and generates new types of information; thus, some management-related thinking emerges [31][32][33]. BIM technically supports multiple designers to model together and exchange opinions. Considering the network-enabled event log mining is beneficial for a deep understanding of the BIM-based collaborative design work, a novel algorithm combining node2vec and Gaussian Mixture Model (GMM) was proposed to discover and analyze potential clusters of designers within a network from BIM log data, which provides support for BIM-based design monitoring and reliable decisions to increase collaboration opportunities [31]. BIM is an effective tool that improves communication and information flow between construction parties. To efficiently retrieve useful information from raw project data within the BIM environment, association, clustering, and trend analyses were performed to identify hidden patterns and detect relationships between different attributes (e.g., the correlation between construction elements or correspondence subjects) [32]. In current building operation and maintenance activities, complex and non-intuitive data records and inaccurate manual inputs raise difficulties in making full use of the information stored in BIM models. For improving facility management, K-means, local density-based outlier detection, and Apriori were conducted to extract meaningful patterns and detect improper records in a data warehouse transformed from the BIM database [33].

### 3.3 Data Mining for Automation

Automation is defined as the use of largely automatic operations, equipment, or systems [16]. To achieve predetermined goals in accordance with human requirements, automation needs to be driven by data. It is common to collect, process, and utilize various types of complex data in automation, and the application of data mining is widespread. In the research area of data mining for automation, there are three main research topics: automation detection [34][35], automation equipment [36][37], and building automation system [38][39], as listed in Table 2.

Traditional manual detection is time-consuming, error-prone, and in some cases, dangerous. To overcome these shortcomings, data mining is used to automate the detection process [34][35]. Road cracks potentially reduce road performance and threaten traffic safety. For road crack detection automation, there exist several challenges such as intense inhomogeneity along cracks, topology complexity of cracks, and inference of noises with a similar texture to cracks. Taking these issues into consideration, an automatic road crack detection framework was built using structured Random Forest

(RF) and Support Vector Machine (SVM) upon road crack image datasets [34]. Performing regular inspections of railways is essential to avoid extreme events. In order to enhance the automation of railway inspections to reduce the human component, Principal Component Analysis (PCA) and Support Vector Machine (SVM) were applied for the detection and decomposition of railway tunnels according to mobile laser scanning datasets [35].

Automation equipment can facilitate workflow and make work-related tasks perform independently. Data mining can play a positive role in the functional improvement of automation equipment [36][37]. Due to the complex and unique nature of the home building process, existing manufacturing concepts do not apply to the panelized home prefabrication facility. Accordingly, production planning and control were developed for the characteristics of a panelized home prefabrication facility, and Random Sample Consensus (RANSAC) was used to extract models from the data collected using RFID [36]. The tunnel boring machine (TBM) has become a preferred equipment in the construction of long and large tunnels. However, its inability to mine massive information leads to a prevalence of unsafe and uneconomical TBM construction. Aiming to realize real-time safety warnings, deviation corrections, and excavation controls, the rules, such as the interaction between rock mechanics properties and machine characteristics, were mined [37].

Building automation system (BAS) provides a network-based platform for automatically monitoring and controlling various complex building systems, and data mining is applied to enhance the performance of BAS [38][39]. Building operations are typically dynamic, and therefore BAS data is multivariate time-series data in essence. Because the temporal knowledge discovery in BAS data receives little attention, a time-series data mining methodology, including Symbolic Aggregate approXimation (SAX), Motif Discovery (MD), and Temporal Association Rule Analysis (TARA), was presented for building energy management [38]. When building systems behave differently from design values, BAS will raise alarms, usually generating an excessive number of alarms every day. The lack of actionable alarm information makes it difficult for building operators to take action, so a data mining framework was constructed to preprocess raw alarm data, categorize the alarms based on affected objects, and prioritize the alarms with quantitative impacts [39].

### 3.4 Future Opportunities

There are two promising research fields, i.e., construction robots and construction cybersecurity, to be exploited using data mining for Construction 4.0.

The construction robot is an important part of

Construction 4.0-related technologies. With the continuous development of robotics in recent years, more and more projects have used, or are considering using, construction robots. Construction robots can effectively improve work productivity and reduce safety risks by replacing or assisting workers in performing construction tasks. In a complex and dynamic on-site environment, localization and navigation are the main challenges for construction robots. Localization refers to the robot's ability to identify its location, and navigation refers to its ability to monitor and control its movement from one place to another [40]. To avoid obstacles, especially moving ones, these abilities require construction robots to be equipped with extra sensors to collect data from the on-site environment. Depending on intended applications, the collected sensory data may be visual, thermal, and so forth. A high volume of such unstructured data with noise needs to be processed and utilized in real-time, which provides opportunities for data mining. Although some researchers have begun to apply data mining techniques to deal with these problems, there is still great space for research exploration, considering the random nature of the on-site environment [41].

Construction 4.0 is making the AEC/FM industry more vulnerable to cyber attacks, significantly increasing concerns about construction cybersecurity [42]. Construction cybersecurity is the practice of protecting the critical systems (e.g., digital twin system) and sensitive information (e.g., BIM model) used in projects from cyber threats. These potential threats include denial-of-service, functional modification, reading forgery, and data theft. Intrusion detection is one of the primary means for construction cybersecurity, which has two basic approaches: misuse detection and anomaly detection. Misuse detection refers to matching monitored events with attack signatures in the database, and anomaly detection refers to identifying events that mismatch expected patterns. Undoubtedly, data mining can provide an effective solution for these detections. Although there has been existing research in the cybersecurity domain, the AEC/FM industry still has unique challenges. The application of data mining in construction cybersecurity is worth investigating in this cutting-edge field generally overlooked [43].

## 4 Conclusions and Outlook

With the increased generation and accumulation of data in the era of Construction 4.0, the computing paradigm is shifting to data-oriented. Large enough data contains valuable information. However, finding this valuable information is not trivial. Data mining provides the ability to discover knowledge from large data.

This study takes a systematic literature review to provide an overall view of data mining for Construction 4.0. A bibliometric analysis was conducted based on the

literature retrieved from Web of Science, and the word similarity network was generated using VOSviewer. The outcome shows three main research areas identified in terms of Construction 4.0: data mining for intelligence, digitalization, and automation. Data mining for intelligence is characterized by four research topics: descriptive intelligence, diagnostic intelligence, predictive intelligence, and prescriptive intelligence. Similarly, data mining for digitalization has three main research topics: digitalization demand, digitalization modeling, and digitalization management. Finally, data mining for automation includes three main research topics: automation detection, automation equipment, and building automation system. Moreover, the fundamental issues of construction robots and construction cybersecurity that can be investigated by data mining techniques are discussed. This study summarizes the current state and future opportunities of data mining for Construction 4.0, reveals the current body of knowledge concerning data mining applications, and proposes development directions in the context of Construction 4.0.

It should be highlighted that this study focuses on the application of data mining for Construction 4.0 from a global perspective. In future research, non-obvious local details (e.g., comparison between data mining techniques applied to a specific field) will be further investigated.

## References

- [1] Ghobakhloo, M., Fathi, M., Iranmanesh, M., Maroufkhani, P. and Morales, M.E. Industry 4.0 ten years on: A bibliometric and systematic review of concepts, sustainability value drivers, and success determinants. *Journal of Cleaner Production*, 302:127052, 2021.
- [2] Garc ía de Soto, B., Agust íJuan, I., Joss, S. and Hunhevicz, J. Implications of Construction 4.0 to the workforce and organizational structures. *International Journal of Construction Management*, 1-13, 2019.
- [3] Oesterreich, T.D. and Teuteberg, F. 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. *Computers in Industry*, 83:121-139, 2016.
- [4] Yan, H., Yang, N., Peng, Y. and Ren, Y. Data mining in the construction industry: Present status, opportunities, and future trends. *Automation in Construction*, 119:103331, 2020.
- [5] Boton, C. and Forgues, D. Construction 4.0: The next revolution in the construction industry. *CanBIM Innovation Spotlight Publication 2020*, 2020.
- [6] Ahmed, V., Aziz, Z., Tezel, A. and Riaz, Z. Challenges and drivers for data mining in the AEC sector. *Engineering, Construction and Architectural Management*, 25(11):1436-1453, 2018.
- [7] Bilal, M., Oyedele, L.O., Qadir, J., Munir, K., Ajayi, S.O., Akinade, O.O., Owolabi, H.A., Alaka, H.A. and Pasha, M. Big Data in the construction industry: A review of present status, opportunities, and future trends. *Advanced Engineering Informatics*, 30(3):500-521, 2016.
- [8] Dean, J. *Big data, data mining, and machine learning: Value creation for business leaders and practitioners*. John Wiley & Sons, 2014.
- [9] Bilge, E.Ç. and Yaman, H. Research trends analysis using text mining in construction management: 2000-2020. *Engineering, Construction and Architectural Management*, 2021.
- [10] Sonkor, M.S. and Garc ía de Soto, B. Operational technology on construction sites: A review from the cybersecurity perspective. *Journal of Construction Engineering and Management*, 147(12):04021172, 2021.
- [11] Chen, Q., Garc ía de Soto, B. and Adey, B.T. Construction automation: Research areas, industry concerns and suggestions for advancement. *Automation in Construction*, 94:22-38, 2018.
- [12] Van Eck, N.J. and Waltman, L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2):523-538, 2010.
- [13] Van Eck, N.J. and Waltman, L. Visualizing bibliometric networks. *Measuring Scholarly Impact*, 285-320, Springer, Cham, 2014.
- [14] Oxford University Press. Lexico. On-line: <https://www.lexico.com/en/definition/intelligence>, Accessed: 20/02/2022.
- [15] Oxford University Press. Lexico. On-line: <https://www.lexico.com/en/definition/digitalization>, Accessed: 20/02/2022.
- [16] Oxford University Press. Lexico. On-line: <https://www.lexico.com/en/definition/automation>, Accessed: 20/02/2022.
- [17] Zhou, H., Yu, J., Zhao, Y., Chang, C., Li, J. and Lin, B. Recognizing occupant presence status in residential buildings from environment sensing data by data mining approach. *Energy and Buildings*, 252:111432, 2021.
- [18] Masino, J., Thumm, J., Levasseur, G., Frey, M., Gauterin, F., Mikut, R. and Reischl, M. Characterization of road condition with data mining based on measured kinematic vehicle parameters. *Journal of Advanced Transportation*, 2018.
- [19] Fan, C., Xiao, F. and Yan, C. A framework for knowledge discovery in massive building automation data and its application in building diagnostics. *Automation in Construction*, 50:81-90, 2015.
- [20] Liu, X., Ding, Y., Tang, H. and Xiao, F. A data

- mining-based framework for the identification of daily electricity usage patterns and anomaly detection in building electricity consumption data. *Energy and Buildings*, 231:110601, 2021.
- [21] Cho, J. and Chun, J. Cost estimating methods for RC structures by quantity takeoff and quantity prediction in the design development stage. *Journal of Asian Architecture and Building Engineering*, 14(1):65-72, 2015.
- [22] Zhang, Q., Liu, Z. and Tan, J. Prediction of geological conditions for a tunnel boring machine using big operational data. *Automation in Construction*, 100:73-83, 2019.
- [23] Petrova, E., Pauwels, P., Svidt, K. and Jensen, R.L. Towards data-driven sustainable design: Decision support based on knowledge discovery in disparate building data. *Architectural Engineering and Design Management*, 15(5):334-356, 2019.
- [24] Marocco, M. and Garofolo, I. Operational text-mining methods for enhancing building maintenance management. *Building Research & Information*, 49(8):893-911, 2021.
- [25] Pan, X., Zhong, B., Wang, X. and Xiang, R. Text mining-based patent analysis of BIM application in construction. *Journal of Civil Engineering and Management*, 27(5):303-315, 2021.
- [26] Zhou, S., Ng, S.T., Lee, S.H., Xu, F.J. and Yang, Y. A domain knowledge incorporated text mining approach for capturing user needs on BIM applications. *Engineering, Construction and Architectural Management*, 27(2):458-482, 2020.
- [27] Hosseini, M.R., Martek, I., Papadonikolaki, E., Sheikhhoshkar, M., Banihashemi, S. and Arashpour, M. Viability of the BIM manager enduring as a distinct role: Association rule mining of job advertisements. *Journal of Construction Engineering and Management*, 144(9):04018085, 2018.
- [28] Huang, C.H. and Hsieh, S.H. Predicting BIM labor cost with random forest and simple linear regression. *Automation in Construction*, 118:103280, 2020.
- [29] Yarmohammadi, S., Pourabolghasem, R. and Castro-Lacouture, D. Mining implicit 3D modeling patterns from unstructured temporal BIM log text data. *Automation in Construction*, 81:17-24, 2017.
- [30] Cheng, J.C., Deng, Y. and Anumba, C. Mapping BIM schema and 3D GIS schema semi-automatically utilizing linguistic and text mining techniques. *Journal of Information Technology in Construction*, 20(14):193-212, 2015.
- [31] Pan, Y., Zhang, L. and Skibniewski, M.J. (2020). Clustering of designers based on building information modeling event logs. *Computer-Aided Civil and Infrastructure Engineering*, 35(7):701-718, 2020.
- [32] Marzouk, M. and Enaba, M. Analyzing project data in BIM with descriptive analytics to improve project performance. *Built Environment Project and Asset Management*, 9(4):476-488, 2019.
- [33] Peng, Y., Lin, J.R., Zhang, J.P. and Hu, Z.Z. A hybrid data mining approach on BIM-based building operation and maintenance. *Building and Environment*, 126:483-495, 2017.
- [34] Shi, Y., Cui, L., Qi, Z., Meng, F. and Chen, Z. Automatic road crack detection using random structured forests. *IEEE Transactions on Intelligent Transportation Systems*, 17(12):3434-3445, 2016.
- [35] Sanchez-Rodriguez, A., Riveiro, B., Soilan, M. and Gonzalez-deSantos, L.M. Automated detection and decomposition of railway tunnels from mobile laser scanning datasets. *Automation in Construction*, 96:171-179, 2018.
- [36] Altaf, M.S., Bouferguene, A., Liu, H., Al-Hussein, M. and Yu, H. Integrated production planning and control system for a panelized home prefabrication facility using simulation and RFID. *Automation in construction*, 85:369-383, 2018.
- [37] Li, J., Jing, L., Zheng, X., Li, P. and Yang, C. Application and outlook of information and intelligence technology for safe and efficient TBM construction. *Tunnelling and Underground Space Technology*, 93:103097, 2019.
- [38] Fan, C., Xiao, F., Madsen, H. and Wang, D. Temporal knowledge discovery in big BAS data for building energy management. *Energy and Buildings*, 109:75-89, 2015.
- [39] Li, H., Aziz, A., Cochran, E. and Lasternas, B. Building automation system alarm management for operation and maintenance decision making. *ASHRAE 2018 Winter Conference*, 2018.
- [40] Brunl, T. (2008). *Embedded robotics: Mobile robot design and applications with embedded systems*. Springer, Berlin, Heidelberg, 2008.
- [41] Liang, C.J., Wang, X., Kamat, V.R. and Menassa, C.C. Human-Robot collaboration in construction: Classification and research trends. *Journal of Construction Engineering and Management*, 147(10):03121006, 2021.
- [42] Mantha, B.R.K. and Garca de Soto, B. Cybersecurity in construction: Where do we stand and how do we get better prepared. *Frontiers in Built Environment*, 7:612668, 2021.
- [43] Yao, D. and Garca de Soto, B. A preliminary SWOT evaluation for the applications of ML to cyber risk analysis in the construction Industry. *In IOP Conference Series: Materials Science and Engineering*, 1218(1):012017, 2022.