Al-Azhar University Civil Engineering Research Magazine (CERM)
Vol. (45) No. (2) October. 2023

The Impacts of Data Science Applications on Construction Management

Elbadr O. Elgendi
Construction and Building Engineering Department, College of Engineering and Technology,
Arab Academy for Science, Technology and Maritime Transport (AASTMT), Egypt,
Email: elbadrosman@aast.edu

الملخص العربى :

يسلط الضوء على الاستخدام المتزايد لعلوم البيانات في إدارة التشييد منذ عام 2014 على كيفية تطوير صناعة التشييد والبناء بواسطة التكنولوجيا. تعكس المساهمات الأكاديمية المتزايدة في هذا المجال هذا التقدم. لذا، هدف هذا البحث هو تعزيز وتطوير المعرفة بالتأثيرات الحالية والمستقبلية لعلوم البيانات في إدارة التشييد من خلال: (1) وصف كيف يتم استخدام علوم البيانات حاليا في إدارة التشييد، (2) دراسة التطبيقات المهمة للتعلم الآلي، والتعلم العميق، وتعدين البيانات التي تؤدي إلى البناء " الذكي"، (3) تقييم فوائد وتحديات استخدام علوم البيانات في صناعة التشييد والبناء، و(4) تحديد المجالات التي تحتاج إلى مزيد من البحث والتطوير في المستقبل. يقدم البحث تحليل كمي ونوعي، عن طريق تقديم إطار تحليلي قوي بناء على منهجية من مرحلتين للتحليل. حددت نتائج هذا البحث الاتجاهات والممارسات والعقبات المعاصرة والمستقبلية المرتبطة باستخدام علوم البيانات في إدارة التشييد. وعلى الرغم من تزايد التركيز على استخدام تقنيات علوم البيانات في صناعة التشييد والبناء، فإن نتائج هذا البحث أثبتت الحاجة إلى المزيد من العلوم المنهجية التي تركز على علوم البيانات. وعلاوة على ذلك، اكدت النتائج على ضرورة فهم أوجه الضعف المرتبطة بتطبيقات البيانات الرقمية في مجال العلوم العلمية والتغلب على المعوقات والسيطرة عليها على نحو كامل، مع تعزيز إمكاناتها في الوقت نفسه. وأخيرا، فإن نتيجة هذا البحث تشكل نقطة مرجعية أساسية لفهم تطور ومستقبل وآثار علم البيانات الرقمية في مجال إدارة التشييد الذي يتوقع أن ييسر المزيد من البحوث والتقدم التكنولوجي في هذا المجال.

ABSTRACT

The increasing use of Data Science (DS) in Construction Management (CM) since 2014 highlights how technology can transform the construction industry. The growing scholarly contributions in this interdisciplinary field reflect this progress. Therefore, the aim of this research is to enhance and develop the knowledge of the present and future impacts of DS in CM by: (1) describing how DS is currently being used in CM, (2) examining important applications of machine learning, deep learning, and data mining that are leading to "smart" construction, (3) evaluating the benefits and challenges of using DS in the industry, and (4) identifying areas that need more research and development in the future. Quantitative and qualitative perspectives provided in this research, which proposed a strong analytical framework based on two-phase approach for the analysis processes. The findings of this research outlined contemporary and future trends, practices, and obstacles associated with DS utilization in CM. Althought there is a growing emphasis on using DS techniques in the construction industry, The research results proved the need for more data-focused methods. Furthermore, results confirmed the necessary to fully understand, reduce, and control the vulnerabilities associated with DS applications, while also enhancing their potential. Finally, the result of this research serves as a fundamental reference point for comprehending the evolution, future, and impacts of DS in the domain of CM which is expected to facilitate further research and technological advancements in this field.

Keywords: Data Science, Machine Learning, Data Mining, Deep Learning, Current Trends, Future Trends, Construction Management.

1. INTRODUCTION

The construction industry is undergoing a significant transformation by integrating DS across its value chain, encompassing the stages of project planning, execution, monitoring, control, operation, and maintenance. From 2014, DS emerged as a crucial driver in construction engineering and management, owing to the exponential growth of data and escalating project complexities. DS offers considerable benefits such as time efficiencies, resource optimization, and reduced reliance on human resources [1]. Data science, a subset of computer science, empowers computers to emulate human cognitive functions like information processing, reasoning, problem-solving, feature extraction, planning, and decision-making, thereby addressing complex and intentionally defined challenges in an intelligent, flexible manner. The realms of DS, including machine learning (ML), specifically deep learning (DL), and Data Mining (DM), have seen substantial technological advancements over the past 40 years, catalyzing notable changes across various industries [2]. However, a report by Purdy and Daugherty (2016) suggests that the CM domain lags in the adoption of DS techniques compared to other sectors [3,4]

DS has emerged as a potent instrument for the construction sector, enhancing process automation and decision-making support. Amidst the evolving landscape of "Industry 4.0", CM is witnessing a relentless shift towards digitization and intelligence, with DS at its foundation facilitating substantial enhancements in automation, productivity, health and safety, and sustainability [5-13]. The total number of articles on DS has been increased. However, many studies have been conducted, most of them focusing on a specific subfield, such as building information modeling (BIM) [14], automated construction manufacturing (ACM) [15], or computer vision (CV) [16], and others. There's an evident research gap in the broad application of DS in CM, with limited exploration of appropriate algorithm selection for diverse scenarios or foundational research for optimizing DS usage in CM. Consequently, there's an overreliance on human evaluation and appraisal that could potentially yield misleading conclusions, especially considering the growing data volumes in complex projects [17].

The aim of this research is to provide a thorough and organized investigation into the present and future impacts of practical use of DS in the field of CM. A strong framework is being proposed for both present and future studies to comprehend the relationship between DS and ML, DL, and DM through a scientometric analysis of pertinent scholarly articles. Furthermore, the research provides a comprehensive analysis of the current use of DS in CM, identifies the main benefits and challenges associated with its implementation, and showcases the most recent and advanced research articles in this area. The adoption of a results-driven methodology provides a strong foundation for understanding and further research in this developing field.

2. DATA SCIENCE (DS)

The complexity of construction projects has led to a surge in data volume, rendering human-driven analysis increasingly challenging [10]. DS equips stakeholders, researchers, and project managers with the necessary tools to navigate this data explosion, facilitating efficient information extraction, predictive analytics, and data-informed decision-making processes. In this section, we delve into DS and its interplay with ML, DL, and DM, as illustrated in Fig 1.

Data science, a field born out of statistical science, leverages scientific methodologies, processes, algorithms, and systems to distill knowledge and insights from diverse, labeled, and unlabeled data, for application across a broad array of domains [10]. The evolution of DS has given rise to three prominent areas examined in this research: 1) ML, 2) DL, and 3) DM. ML is a crucial subset of DS, training machines to unearth patterns within vast data sets and generate data-driven predictions about future events.

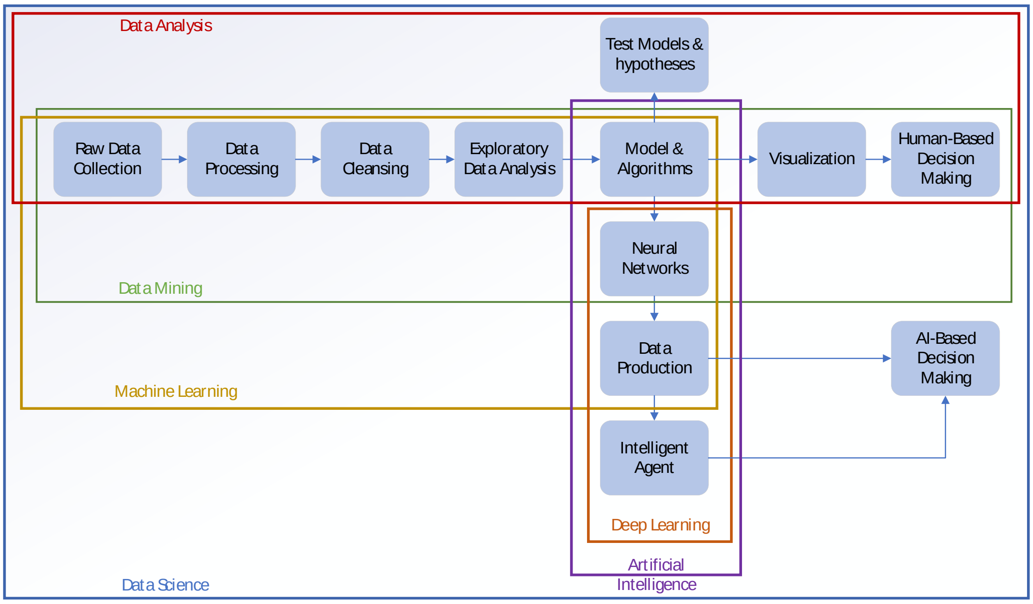

Fig 1: Data Science and Its Sub ML, DL, Data Analysis, and DM.

As ML advances, new trends such as DL and Q-learning had been established at a higher level. DL is a subset of ML that consists of neural networks (NN) with several hidden layers. In comparison to applications based on ML, DL models require a substantial amount of training and testing data [18]. Moreover, DM is a relatively new area that aims to connect process management and DS. DM makes extensive use of data logs in order to monitor, diagnose, analyze, and improve actual processes [19].

2.1. Machine Learning (ML)

ML harnesses statistical and mathematical models to identify patterns and make predictions, ideal for discovering optimal solutions that would be hard to obtain via human trial-and-error. Four types of ML exist: supervised learning, unsupervised learning, semi-supervised learning, and Q-learning [7]. Supervised learning extracts features from labeled data to predict the outcomes of new inputs. Unsupervised learning gleans features and information from unlabeled data, excelling at data reduction and clustering tasks. Semi-supervised learning handles datasets with many inputs and scarce output data. Q-learning, or reinforcement learning, utilizes an agent, states, and action sets for each state, aiming to maximize the total rewards obtained from its learning process [20].

2.2.Deep Learning (DL)

DL is a part of ML, which is a subset of NN. Backpropagation was initially postulated as the foundation of a comprehensive NN theory, sparking the 1st ML movement [21]. Previously, NNs lacked appropriate algorithm support, preventing them from training multi-layer NNs. By following backpropagation, as a method to update the weights in the neural network by taking into account the actual output and the desired output as shown in Fig. 2, two frameworks for classical NNs were developed: LeNet [22] and Long-Short-Term Memory Networks (LSTM) [23].

As with the majority of NN designs, layers, neurons, activation functions, and weights are used in DL techniques. The neurons that act as feature detectors are layered. First layers detect essential features and deliver them to the next layers and so on, which identify more complex features. While the majority of DL techniques are adaptable to a broad range of regression, classification, or clustering problems [2,12], they are occasionally connected via ensemble modeling to increase performance.

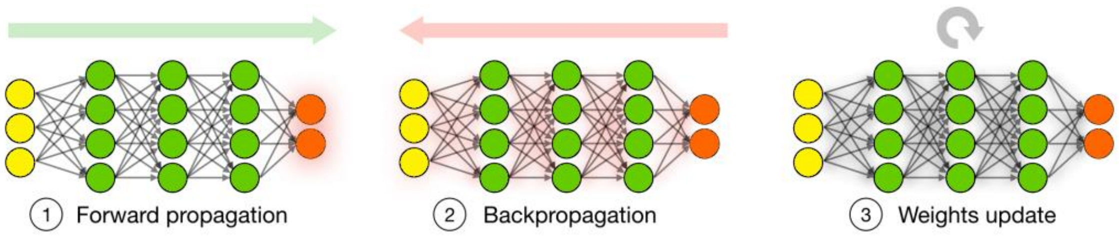

Fig. 2: The Steps of Updating Weights of NN.

The type of DL model depends on the input data. The simplest form of NN that developed in the form of feed-forward NNs was artificial neural networks (ANN) [24] that is using simple data in regression and classification models. Recurrent Neural Networks (RNN) specialize in dealing with sequential data like time series ML. They outperform other types of DL when it comes to analyzing time-dependent data [25]. ANN was used in CM like [26–28]. RNNs are mostly used in video and speech processing because they can retain knowledge about previously processed audio chunks or video frames in order to forecast subsequent input [24,29,30]. RNNs were used in construction for cost, safety management [31–34].

3. RESEARCH METHODOLOGY

The research's objective was to conduct scientometric analyses and visualizations to offer a comprehensive understanding of the structure, research areas, and trends of DS in the CM field, using a results-based approach and empirical findings. Scientometric analysis, leveraging mathematical formulas and visualizations, identifies structural patterns and prominent research boundaries, allowing for an efficient yet exhaustive mapping of scientific knowledge [6,35–37].

Various researchers have utilized this method across diverse research subjects, including building information modelling [38], and sustainability [36,37]. Scopus, a widely encompassing bibliographic database, was chosen as the primary source of data for this research, over alternatives like the Web of Science (WOS), Dimensions, Google Scholar, or Research Gate [36,39]. Scopus's indexing technique is also faster, enabling the retrieval of more recent materials [35,37]. While multiple literature databases could enhance the analysis, issues with identifying and eliminating duplicated articles across the databases due to the dataset's size and the time constraints necessitated the sole use of Scopus, which aligns with many previous science mapping studies [12,40].

To augment search efficacy, this research incorporated recent review articles [2,6,10,12,13] to find relevant Scopus database search terms. As a consequence, a list of DS-related keywords were generated. These keywords, in addition to "*ALL ('artificial intelligence' OR 'Data Mining' OR 'AI' OR 'computational intelligence') AND (LIMIT-TO [SUBJAREA, 'ENGI']) AND (LIMIT-TO [EXACTKEYWORD, 'Project Management' OR 'Construction Management' OR Construction Industry'] AND (EXCLUDE [EXACTKEYWORD, 'Optimization' OR 'Genetic Algorithms' OR 'Multiobjective Optimization' OR 'Genetic Algorithm' OR 'Particle Swarm Optimization (PSO)']) AND (LIMIT-TO [DOCTYPE, 'ar'])."*

A total of 1,752 scholarly papers and articles were initially retrieved for this research. To ensure the inclusion of high-quality research, the document type was limited to "article" since journal articles are considered reliable sources of certified knowledge in science mapping [41]. Other document types could potentially introduce noise and hinder analysis and interpretation. Papers referring to the application of DS in the CM domain were selected based on their abstracts. Optimization methods like genetic algorithms and fuzzy approaches were omitted, ensuring the focus remains on DS. The stringent screening yielded a dataset of 140 DS in CM articles, exported in plain-text format as complete records Fig. 3. VOSviewer, a Java-based scientific visualization tool, was used for the primary analysis. It aids in creating, visualizing, and exploring bibliometric networks. The software was used to conduct co-occurrence analyses to generate keyword and information maps. The analysis spanned four categories: a) Annual publish: This category represented the number of published articles per year, allowing for an understanding of the publication trends over time. b) Co-occurrence keyword using index keywords and authors' keywords: This category focused on the relationships between index keywords and authors' keywords, providing further insights into the specific topics and areas of interest within the field.

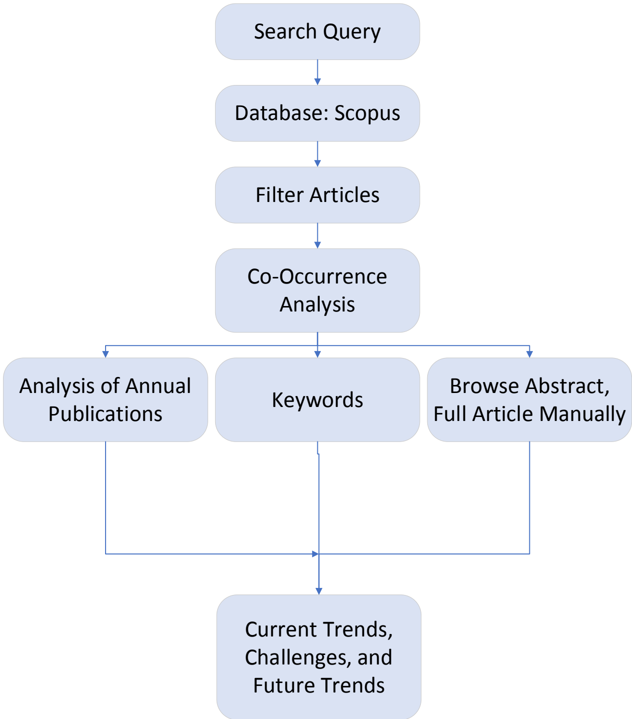

Fig. 3: Methodology Map

.

4. DATA ANALYSIS

4.1. Current Analysis of Data Science Articles

Annual Articles on Data Science in Construction Management

The annual number of relevant articles is increasing between the start of publication and 2024 as shown in Fig. 4, demonstrating that the application of developing DS in the CM domain is becoming a popular issue at the moment. As illustrated in Fig. 4, the number of relevant articles has increased quickly during the last 23 years. Around 50% of articles were published after 2018, indicating that the popularity of DS in the CM domain increased significantly around 2018. It should be emphasized that the final 100 represents only the papers published from 2015 to the present, surpassing the annual articles from 1997 to 2015. That is, DS techniques in the field of the CM domain are gaining increasing traction with the hope of introducing digital innovation to the construction management area. This expansion appears to correspond to the growing interest in DL, which began in 2012 with the publication of a study introducing AlexNet as a deep NN optimized for image classification [42].

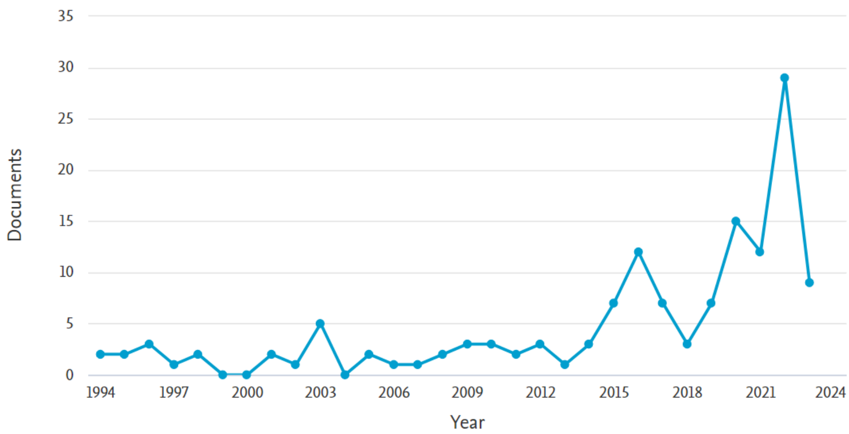

Fig. 4: Annual Articles of Data Science in Construction Management Domain

Co-Occurrence Network of Keywords

Crucial areas of study were identified through the application of co-occurrence keyword analysis, as reported in reference [9]. The authors employed a methodology that facilitated the creation of a network comprising commonly appearing keywords in the analyzed articles. This network served as a representation of the primary research themes [10]. According to a study cited in reference 43, the approach utilized by the researchers aided in the clear visualization of the knowledge domain. Additionally, this approach provided valuable insights into the discussed topics and their cognitive interconnections. The construction of the keyword co-occurrence network was carried out through the utilization of the VOSviewer application. The application employs the smart local moving (SLM) algorithm for cluster analysis, utilizing bibliographic data sourced from Scopus. The selection of index keywords, as opposed to author keywords, was made in order to guarantee a consistent and comprehensible depiction [43]. Utilizing the fractional counting method, a total of 2026 keywords were identified across 284 articles. To eliminate extraneous terms, a minimum occurrence threshold of 15 was established. A total of 23 keywords that met the specified criterion were identified and subsequently displayed. The degree of association between two terms was ascertained based on the frequency of their co-occurrence in scholarly articles, thereby indicating their relevance to their respective domains of inquiry. In the network visualization, a more robust connection is depicted by a thicker line. This approach was documented in reference [44].

From our analysis, it was evident that "project management" was the most prevalent research subject, indicating its frequent usage in this field. Furthermore, the prominence of the research subjects was indicated by the frequent appearance of the phrases “decision-making,” “information management,” “artificial intelligence,” and “construction management” in the 284 articles examined, as demonstrated in Figure 5. The knowledge of commonly used keywords can assist researchers in selecting the most appropriate ones for their articles, which can improve indexing and retrieval [41]. The interconnections among the aforementioned terminologies were subsequently depicted via three discrete co-occurring networks.

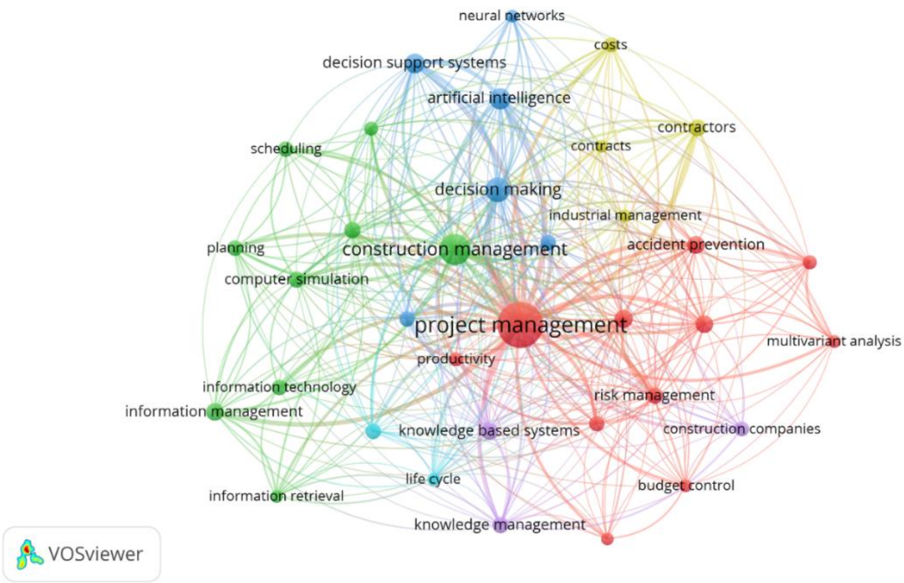

Fig. 5: Data Science and Artificial Intelligence Co-occurrence Network

4.2.Current Application of Data Science in Construction Management

Some of the main existing applications of the DS techniques that were mentioned in the preceding part are addressed in this section. The DS issues addressed by these techniques are discussed.

Schedule Management

According to existing research, the ultimate result of a project is significantly impacted by the caliber of planning during its initial phase [45] A ground-breaking schedule-learning platform was developed by the authors, which utilizes DL and ML techniques. The platform in question utilizes a vast database of historical project schedules to improve the efficacy of DS applications within the realm of CM. This innovative and scalable approach contributes to elevating the accuracy and reliability of project planning. In their research, the authors of reference [45] utilized initial planning statuses as variables to forecast the achievement of project cost and schedule objectives. The researchers employed ML techniques to construct a collection of ANNs and support vector machine (SVM) classification models. The authors of the study analyzed a specific construction project by utilizing statistical data and ML algorithms to pinpoint the primary factors contributing to delays in the project's completion. They proposed a predictive DM model to determine the residual value of heavy construction equipment, providing significant benefits for construction equipment management decision-making. The proposed technique demonstrates advantages over traditional residual value analyses, offering simplicity, superior interpretability, and suitable accuracy [46].

In a different approach, [47] devised a DL method that employs image segmentation to autonomously examine the wall construction progress of an entire floor. The findings were integrated into BIM in a seamless manner. Further, [48] introduced an immersive virtual reality (iVR) system that combines 3D scanning, extended reality, and visual programming to simulate interactive onsite inspections for indoor activities and provide numerical statistics. In a previous study [49], a system was proposed that integrates videography with

Matrix Laboratory (MATLAB) and BIM through the use of an accessible Internet protocol (IP) camera. The integration of an IP camera and a software program has enabled the automated extraction of as-built quantities for columns, beams, and block masonry. These elements are integral to building operations and represent a substantial portion thereof. The tangible evidence of the transformative impact of DS on CM is provided by these real-world applications.

Cost Management

Cost estimation in construction is frequently influenced by elements such as construction time, building types, labor, and equipment. These variables, together with changes in economic matrices and key performance indexes (KPIs), are frequently neglected by traditional building cost estimators. These characteristics, on the other hand, are critical for forecasting an already difficult-to-predict construction cost. [50] provided a model for estimating building costs based on DBMs that takes economic matrices and KPIs into consideration. To validate the model's accuracy, test data from multi-story and mid-rise buildings was used. Also, there are some frameworks that cover cost analysis and life cycle costs but with limited data like [34,51–61]

5. PROSPECTIVE CHALLENGES OF DATA SCIENCE APPLICATIONS

Despite DS' huge success as mentioned before in a variety of fields, certain problems continue in its use. Several of these challenges are expected to arise when solutions in the CM area are developed. Data availability, data privacy, and data ethics, cybersecurity, implementation cost, as well as a shortage of in-house DS skills. The list of issues described in this section is not exhaustive, and further challenges may arise as a result of the implementation of DS in CM.

5.1.Data Availability

DS works best when huge amounts of data are available; otherwise, models trained on small amounts of data will struggle. Accessing data for a specific purpose is frequently challenging, much more so now that the data ethics and General Data Protection Regulation (GDPR) standards have been implemented. To replace limited training data, methods such as picture rotation and flipping may be necessary. However, data augmentation may result in the loss of critical data or outliers necessary for training [62,63]. Additionally, the majority of ML models operate in a black-box way, which means they do not disclose how they arrived at their results. To foster trust in such systems, building practitioners must first understand how the system makes judgments [64].

5.2. Lack of One Model to Fit All

A single general DS model cannot address the majority of the previously described and recommended applications. Each issue must be addressed individually using a model trained expressly for that purpose. Transfer learning techniques or reusing existing models proven to perform well in classification or prediction can be utilized to jumpstart model training, although sufficient weight optimization and hyper parameter adjustment are still required. This is not a one-size-fits-all answer in its totality, and this is one of the primary research issues for DS.

5.3. Cybersecurity

Machine learning (ML) has been identified as a promising tool for enhancing security measures and intrusion detection. However, it is noteworthy that ML also poses a significant risk for cyber threats, such as privacy infringements and hacking activities. The matter at hand bears significant economic and financial ramifications. Minor errors during construction processes can have significant repercussions, impacting project quality, cost, and timeline. These effects can extend to other aspects of project planning, such as supply chain logistics and procurement. Instances have been discovered through research where ML has the potential to be exploited for the development of malware threats [65]. Notable security vulnerabilities that incorporate artificial intelligence have been identified, such as automated password-generating systems [66] and voice cloning systems driven by ML [67]. The significance of incorporating strong cybersecurity measures in the integration of machine learning and other DS techniques into CM practices is emphasized. The findings of the study emphasize the necessity for further research and inventive measures to reconcile the benefits of ML with the crucial requirement for secure applications in the construction sector.

5.4. Lack of Specialists and Implications Cost

It may be difficult to find engineering employees with proper expertise in information technology to execute DS methodologies. A suitable solution would be for these engineers to be trained on the job. This, however, may not be feasible because these professionals are best qualified to comprehend building challenges and optimal solutions, but ML and DL are beyond their domain of expertise. A similar issue arises when outsourcing construction tasks to ML or DL experts with no prior experience in construction. These specialists have no concept of how the building industry operates. Additionally, there is a shortage of DS specialists with the technical ability and experience essential to find creative solutions to construction challenges.

On the other hand, the installation of cutting-edge technology is always costly. To avoid months of training, training as a model requires strong computer resources. However, these resources are expensive, even though they may reduce the entire training procedure to a few hours. Businesses considering adopting DS techniques should also be aware of the financial costs. It is worth mentioning that hiring DS professionals can be pretty expensive. The exact cost of using DS is difficult to quantify since it is highly dependent on the amount of expertise required and the training resources used.

6. FUTURE TRENDS

The effectiveness of DS is apparent in numerous industries, however, its integration into the realm of CM presents distinct obstacles. The present section presents empirical evidence that elucidates the expected challenges in the CM domain with regards to the development and execution of DS. The hindrances that impede the progress of DS implementation encompass concerns related to the accessibility of data, ethical and privacy considerations, cybersecurity, the expenses involved in implementation, and the limited availability of in-house DS proficiency. It is noteworthy that the aforementioned list is not comprehensive and may undergo modifications as the discipline progresses in its utilization of DS. Table 1 presents an analysis of frequently used keywords in titles and abstracts of relevant literature. Terms like "Time", "Cost", "Decision", and "Safety", among others, are included. The significance of these keywords in CM, especially in the context of DS application, is highlighted by their frequency. In addition, relevance scores indicate the significance of each term in the field, identifying important areas and potential directions for future research. The integration of DS into CM will likely keep these keywords and themes at the forefront of the discourse, driving research and innovation in the field. The aim is to improve the effectiveness of DS methodologies in CM by examining common terms and their implications.

Table 1. The Top domain for the DS in construction industry according to the analysis

Term	**Occurrences**	**Relevance Score**
Time	49	0.4824
Cost	41	1.338
Decision	38	0.5291
Activity	35	0.5206
Work	35	0.3477
Safety	28	1.0444
Schedule	27	1.8516
Uncertainty	25	0.9815
Productivity	22	1.1813
Stakeholder	20	0.5889
BIM	19	1.5811

6.1. Public Dataset.

The utilization of data is of utmost importance in DS applications pertaining to the field of construction. The analysis conducted indicates a pressing requirement for a database that is accessible to the public and tailored to the field of CM. By drawing parallels with the significant contributions made by ImageNet to DL in image processing, it is suggested that a comparable database for construction could potentially revolutionize the field of construction learning. The availability of these public datasets would enable researchers to concentrate on the development of DS systems. Furthermore, the presence of data collections of superior quality may alleviate the issue of model overfitting, thereby enhancing dependable performance across a range of scenarios.

6.2. Cash Flow

The research's findings emphasize the significance of cash flow within the construction industry, as it has a notable impact on both project profitability and stability. According to previous research [12], contractor performance may be adversely affected by inadequate cash flow, which can result in a shortage of funds for routine business activities. The findings of the research indicate that it is imperative for contractors to possess cash flow projections that encompass the entire duration of the project, commencing from the initial stages of the tendering process. The aforementioned information enables individuals to anticipate forthcoming challenges and assess diverse impact factors that may affect the success of a project. Construction cash flow forecasting often involves the use of various predictive models, such as NNs [45,68]. Nevertheless, the models frequently concentrate exclusively on the weights of variable costs. According to the research conducted, in order to enhance the credibility of forecasts, it is recommended to incorporate uncertainty into these projections. This can be achieved by utilizing interpretable models that are trained on time-dependent real-world data.

6.3. Quality Management

The empirical data presented indicates that the use of DS can effectively improve the design process, resulting in the creation of more desirable environments for users. Through the utilization of DS, entities have the ability to scrutinize and anticipate usage trends, which can subsequently inform the development of facilities that are closely aligned with their objectives. The results indicate that the utilization of DS can facilitate the detection of design errors or omissions prior to the construction phase, thereby enabling the allocation of additional time towards other productive activities. Furthermore, the employment of DS enables the evaluation of the model's performance across a range of environmental conditions and scenarios.

6.4. Public-Private-Partnership (P3)

Public-private partnerships have become increasingly common, especially in the face of complex funding mechanisms [71]. Our analysis reveals a need for more research to identify risk factors in P3 projects and incorporate those factors into ML or DL models for predictive evaluation of future projects. For example, [72] developed nine ML models to predict the litigation outcomes of primary causes of P3 disputes, while [73] introduced a neuro-fuzzy model based on a framework combining fuzzy logic and artificial neural network techniques to model the risk allocation decision-making process. These empirical results indicate the potential benefits and challenges of using DS techniques in P3 projects.

7. CONCLUSIONS

The potential of DS to bring significant changes and improvements in various sectors, including CM, has been demonstrated through impacts of its applications and implementation. DS's capability to handle, analyze, and extract meaningful patterns from large quantities of data is credited for this transformation. This proved by the quantitative and qualitative examination employed in the present research based on the developed two-phase approach for the proposed strong analytical framework a multi-dimensional approach, to offer robust assessment, analysis, and predication processes. The findings of the research indicated a noticeable increase in topics related to DS since 2014. This timeline coincides with the emergence of cloud databases that have facilitated the training of ML algorithms. However, several potential challenges and considerations that could affect the application of DS in CM have been identified by this research, despite its capabilities. Data availability as revealed is the most focal challenge as Abundant and high-quality data are crucial for the performance of DS, although obtaining such data can be challenging due to concerns about data privacy and compliance with GDPR. Furthermore, it was confirmed that a universal model for CM scenarios is lacking in the field. As a distinct DS model is required for each situation, tailored to the specific task. However, some benefit can be gained from reusing models and transfer learning. Moreover, cybersecurity presents another significant challenge. Robust security measures are required in CM due to the risks of cyber threats, such as hacking and privacy violations, that come with the introduction of DS and ML. The research also revealed a deficiency of DS experts in the CM domain, which may impede the effective execution of DS approaches. In addition, the results showed that companies are concerned about the cost of implementing DS methodologies, obtaining the required computing resources, and recruiting qualified DS professionals.

However, the importance of public datasets, cash flow forecasting, quality management, and public-private partnerships is predicted by results to increase in the application of DS in CM in the future. The creation of a public dataset specifically designed for the CM field is expected to encourage innovative research, similar to the impact of ImageNet on image processing. Cash flow forecasting has been proved as the most important direction as it has a direct impact on project profitability and stability. The recognition of DS's potential to enhance design processes, anticipate design errors, and enable superior project planning were also presented as the second important directions. Furthermore, the results revealed that DS have a significant role in identifying and predicting risk factors related to public-private partnerships, which are becoming increasingly common. These potential future directions and prospects outlined by the results of this research will be very helpful and informative for professionals and scholars working in this area. In conclusion, although incorporating DS in the CM industry has many benefits, it is important to recognize the obstacles and possible remedies. As we move forward in the technology era, comprehending these essential elements will act as a roadmap for better and more efficient DS techniques in condition monitoring.

Declaration of Interest

The author states that he has no conflicting financial interests or personal relationships that might seem to have influenced the work presented in this research.

REFERENCES:

[1] Z. You, C. Wu, A framework for data-driven informatization of the construction company, Adv. Eng. Inform. 39 (2019) 269–277.

[2] T.D. Akinosho, L.O. Oyedele, M. Bilal, A.O. Ajayi, M.D. Delgado, O.O. Akinade, A.A. Ahmed, Deep learning in the construction industry: A review of present status and future innovations, J. Build. Eng. 32 (2020) 101827. https://doi.org/10.1016/j.jobe.2020.101827.

[3] N. Thompson, W. Squires, N. Fearnhead, R. Claase, Digitalisation in Construction-Industrial Strategy Review, supporting the Government’s Industrial Strategy, Univ. Coll. Lond. Lond. (2017).

[4] N. Van Tam, N.Q. Toan, V. Van Phong, S. Durdyev, Impact of BIM-related factors affecting construction project performance, Int. J. Build. Pathol. Adapt. (2021).

[5] M. Bolpagni, I. Bartoletti, Artificial Intelligence in the Construction Industry: Adoption, Benefits and Risks, in: Proc Conf. CIB W78, 2021: pp. 11–15.

[6] A. Darko, A.P.C. Chan, M.A. Adabre, D.J. Edwards, M.R. Hosseini, E.E. Ameyaw, Artificial intelligence in the AEC industry: Scientometric analysis and visualization of research activities, Autom. Constr. 112 (2020) 103081. https://doi.org/10.1016/j.autcon.2020.103081.

[7] K. KOÇ, A.P. GURGUN, Machine learning applications in construction safety literature, Proc. Int. Struct. Eng. Constr. 8 (2021) 1.
[8] S. Liu, R. Chang, J. Zuo, R.J. Webber, F. Xiong, N. Dong, Application of Artificial Neural Networks in Construction Management: Current Status and Future Directions, Appl. Sci. 11 (2021) 9616.
[9] S. Liu, R. Chang, J. Zuo, R.J. Webber, F. Xiong, N. Dong, Application of Artificial Neural Networks in Construction Management: Current Status and Future Directions, Appl. Sci. 11 (2021) 9616. https://doi.org/10.3390/app11209616.
[10] Y. Lu, J. Zhang, Bibliometric analysis and critical review of the research on big data in the construction industry, Eng. Constr. Archit. Manag. (2021).
[11] B. Manzoor, I. Othman, S. Durdyev, S. Ismail, M.H. Wahab, Influence of Artificial Intelligence in Civil Engineering toward Sustainable Development—A Systematic Literature Review, Appl. Syst. Innov. 4 (2021) 52. https://doi.org/10.3390/asi4030052.
[12] Y. Pan, L. Zhang, Roles of artificial intelligence in construction engineering and management: A critical review and future trends, Autom. Constr. 122 (2021) 103517.
[13] Y. Xu, Y. Zhou, P. Sekula, L. Ding, Machine learning in construction: From shallow to deep learning, Dev. Built Environ. 6 (2021) 100045. https://doi.org/10.1016/j.dibe.2021.100045.
[14] Y. Zou, A. Kiviniemi, S.W. Jones, A review of risk management through BIM and BIM-related technologies, Saf. Sci. 97 (2017) 88–98.
[15] M. Hatami, I. Flood, B. Franz, X. Zhang, State-of-the-art review on the applicability of AI methods to automated construction manufacturing, Comput. Civ. Eng. 2019 Data Sens. Anal. (2019) 368–375.
[16] G. Zhang, Y. Pan, L. Zhang, R.L.K. Tiong, Cross-scale generative adversarial network for crowd density estimation from images, Eng. Appl. Artif. Intell. 94 (2020) 103777.
[17] M.R. Hosseini, I. Martek, E.K. Zavadskas, A.A. Aibinu, M. Arashpour, N. Chileshe, Critical evaluation of off-site construction research: A Scientometric analysis, Autom. Constr. 87 (2018) 235–247.
[18] R. Ilin, T.P. Watson, R. Kozma, Abstraction hierarchy in deep learning neural networks, 2017 Int. Jt. Conf. Neural Netw. IJCNN. (2017). https://doi.org/10.1109/IJCNN.2017.7965929.
[19] V. Ahmed, Z. Aziz, A. Tezel, Z. Riaz, Challenges and drivers for data mining in the AEC sector, Eng. Constr. Archit. Manag. (2018).
[20] A.K. Lakshmanan, R. Mohan, B. Ramalingam, A. Le, P. Veerajagadeshwar, K. Tiwari, M. Ilyas, Complete coverage path planning using reinforcement learning for Tetromino based cleaning and maintenance robot, (2020). https://doi.org/10.1016/j.autcon.2020.103078.
[21] D.E. Rumelhart, G.E. Hinton, Learning representations by back-propagating, (1986).

[22] Y. LeCun, B. Boser, J. Denker, D. Henderson, R. Howard, W. Hubbard, L. Jackel, Backpropagation Applied to Handwritten Zip Code Recognition, Neural Comput. (1989). https://doi.org/10.1162/neco.1989.1.4.541.

[23] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comput. 9 (1997).

[24] I. Flood, N. Kartam, Neural Networks in Civil Engineering. II: Systems and Application, J. Comput. Civ. Eng. 8 (1994) 149–162. https://doi.org/10.1061/(ASCE)0887-3801(1994)8:2(149).

[25] T. Mikolov, M. Karafiát, L. Burget, J. Cernocký, S. Khudanpur, Recurrent neural network based language model, in: INTERSPEECH, 2010.

[26] M. Badawy, A. Hussein, S.M. Elseufy, K. Alnaas, How to predict the rebar labours' production rate by using ANN model?, Int. J. Constr. Manag. 21 (2021) 427–438. https://doi.org/10.1080/15623599.2018.1553573.

[27] S.S. Bangaru, C. Wang, S.A. Busam, F. Aghazadeh, ANN-based automated scaffold builder activity recognition through wearable EMG and IMU sensors, Autom. Constr. 126 (2021). https://doi.org/10.1016/j.autcon.2021.103653.

[28] 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 (2017). https://doi.org/10.1061/(ASCE)CO.1943-7862.0001340.

[29] J. Schmidhuber, Deep Learning in Neural Networks: An Overview, IDSIA, 2014. https://repository.supsi.ch/4902/ (accessed November 27, 2021).

[30] P.J. Werbos, Generalization of backpropagation with application to a recurrent gas market model, Neural Netw. 1 (1988) 339–356. https://doi.org/10.1016/0893-6080(88)90007-X.

[31] Y. Jang, I. Jeong, Y.K. Cho, Business Failure Prediction of Construction Contractors Using a LSTM RNN with Accounting, Construction Market, and Macroeconomic Variables, J. Manag. Eng. 36 (2020). https://doi.org/10.1061/(ASCE)ME.1943-5479.0000733.

[32] G.H. Scavariello, A.R.S. Picanco, C. Torezzan, Combining maximum likelihood estimation and LSTM neural network to forecast reliability distributions: A study based on real data from the sugar-energy sector, IEEE Lat. Am. Trans. 20 (2022) 608–615. https://doi.org/10.1109/TLA.2022.9675466.

[33] 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). https://doi.org/10.1155/2018/9691868.

[34] H. Weytjens, E. Lohmann, M. Kleinsteuber, Cash flow prediction: MLP and LSTM compared to ARIMA and Prophet, Electron. Commer. Res. 21 (2021). https://doi.org/10.1007/s10660-019-09362-7.

[35] D. De Filippo, M.L. Lascurain, A. Pandiella-Dominique, E. Sanz-Casado, Scientometric Analysis of Research in Energy Efficiency and Citizen Science through Projects and Publications, Sustainability. 12 (2020) 5175. https://doi.org/10.3390/su12125175.

[36] T.O. Olawumi, D.W.M. Chan, A scientometric review of global research on sustainability and sustainable development, J. Clean. Prod. 183 (2018) 231–250. https://doi.org/10.1016/j.jclepro.2018.02.162.
[37] R.F. de Toledo, H.L. Miranda Junior, J.R. Farias Filho, H.G. Costa, A scientometric review of global research on sustainability and project management dataset, Data Brief. 25 (2019) 104312. https://doi.org/10.1016/j.dib.2019.104312.
[38] Z. Liu, Y. Lu, L.C. Peh, A Review and Scientometric Analysis of Global Building Information Modeling (BIM) Research in the Architecture, Engineering and Construction (AEC) Industry, Buildings. 9 (2019) 210. https://doi.org/10.3390/buildings9100210.
[39] L. Meho, Y. Rogers, Citation counting, citation ranking, and h-index of human-computer interaction researchers: A comparison of Scopus and Web of Science, J. Am. Soc. Inf. Sci. Technol. 59 (2008) 1711–1726. https://doi.org/10.1002/asi.20874.
[40] L. Hou, H. Chen, G. (Kevin) Zhang, X. Wang, Deep Learning-Based Applications for Safety Management in the AEC Industry: A Review, Appl. Sci. 11 (2021) 821. https://doi.org/10.3390/app11020821.
[41] X. Xiao, M. Skitmore, H. Li, B. Xia, Mapping Knowledge in the Economic Areas of Green Building Using Scientometric Analysis, Energies. 12 (2019) 3011. https://doi.org/10.3390/en12153011.
[42] A. Krizhevsky, I. Sutskever, Imagenet classification with deep convolutional, (2012).
[43] N.J. van Eck, L. Waltman, Visualizing Bibliometric Networks, in: Y. Ding, R. Rousseau, D. Wolfram (Eds.), Meas. Sch. Impact Methods Pract., Springer International Publishing, Cham, 2014: pp. 285–320. https://doi.org/10.1007/978-3-319-10377-8_13.
[44] N.J. Van Eck, L. Waltman, VOSviewer manual, Leiden Univeristeit Leiden. 1 (2013) 1–53.
[45] Y.-R. Wang, C.-Y. Yu, H.-H. Chan, Predicting construction cost and schedule success using artificial neural networks ensemble and support vector machines classification models, (2012). https://doi.org/10.1016/J.IJPROMAN.2011.09.002.
[46] H. Kim, L. Soibelman, F. Grobler, Factor selection for delay analysis using Knowledge Discovery in Databases, Autom. Constr. 17 (2008) 550–560. https://doi.org/10.1016/j.autcon.2007.10.001.
[47] W. Wei, Y. Lu, T. Zhong, P. Li, B. Liu, Integrated vision-based automated progress monitoring of indoor construction using mask region-based convolutional neural networks and BIM, Autom. Constr. 140 (2022). https://doi.org/10.1016/j.autcon.2022.104327.
[48] A.K. Ali, O.J. Lee, D. Lee, C. Park, Remote indoor construction progress monitoring using extended reality, Sustain. Switz. 13 (2021) 1–24. https://doi.org/10.3390/su13042290.

[49] F. Arif, W.A. Khan, Smart Progress Monitoring Framework for Building Construction Elements Using Videography–MATLAB–BIM Integration, Int. J. Civ. Eng. (2021). https://doi.org/10.1007/s40999-021-00601-3.
[50] Q. Jiang, Estimation of construction project building cost by back-propagation neural network, J. Eng. Des. Technol. 18 (2019) 601–609. https://doi.org/10.1108/JEDT-08-2019-0195.
[51] F. Costantino, G. Di Gravio, F. Nonino, Project selection in project portfolio management: An artificial neural network model based on critical success factors, Int. J. Proj. Manag. 33 (2015) 1744–1754. https://doi.org/10.1016/j.ijproman.2015.07.003.
[52] S.R. Dastgheib, M.R. Feylizadeh, M. Bagherpour, A. Mahmoudi, Improving estimate at completion (EAC) cost of construction projects using adaptive neuro-fuzzy inference system (ANFIS), Can. J. Civ. Eng. 49 (2022) 222–232. https://doi.org/10.1139/cjce-2020-0399.
[53] H.H. Elmousalami, Artificial Intelligence and Parametric Construction Cost Estimate Modeling: State-of-the-Art Review, J. Constr. Eng. Manag. 146 (2020) 03119008. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001678.
[54] H.H. ElMousalami, A.H. Elyamany, A.H. Ibrahim, Predicting Conceptual Cost for Field Canal Improvement Projects, J. Constr. Eng. Manag. 144 (2018) 04018102. https://doi.org/10.1061/(ASCE)CO.1943-7862.0001561.
[55] X. Gao, P. Pishdad-Bozorgi, A framework of developing machine learning models for facility life-cycle cost analysis, Build. Res. Inf. 48 (2020) 501–525. https://doi.org/10.1080/09613218.2019.1691488.
[56] K.H. Hyari, A. Al-Daraiseh, M. El-Mashaleh, Conceptual Cost Estimation Model for Engineering Services in Public Construction Projects, J. Manag. Eng. 32 (2016) 04015021. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000381.
[57] A. Leśniak, M. Juszczyk, Prediction of site overhead costs with the use of artificial neural network based model, Arch. Civ. Mech. Eng. 18 (2018) 973–982. https://doi.org/10.1016/j.acme.2018.01.014.
[58] A. Mahmoudi, S.A. Javed, X. Deng, Earned duration management under uncertainty, Soft Comput. 25 (2021) 8921–8940. https://doi.org/10.1007/s00500-021-05782-6.
[59] J. Meng, J. Yan, B. Xue, J. Fu, N. He, Reducing construction material cost by optimizing buy-in decision that accounts the flexibility of non-critical activities, Eng. Constr. Archit. Manag. 25 (2018) 1092–1108. https://doi.org/10.1108/ECAM-12-2016-0263.
[60] M.H. Rafiei, H. Adeli, Novel Machine-Learning Model for Estimating Construction Costs Considering Economic Variables and Indexes, J. Constr. Eng. Manag. (2018). https://doi.org/10.1061/(ASCE)CO.1943-7862.0001570.
[61] C.G. Wilmot, B. Mei, Neural Network Modeling of Highway Construction Costs, J. Constr. Eng. Manag. 131 (2005) 765–771. https://doi.org/10.1061/(ASCE)0733-9364(2005)131:7(765).

[62] J.L. Hong, Automated Data Extraction with Multiple Ontologies, (2016). https://doi.org/10.14257/IJGDC.2016.9.6.33.
[63] C.A. Knoblock, K. Lerman, S. Minton, I. Muslea, Accurately and Reliably Extracting Data from the Web: A Machine Learning Approach, in: IEEE Data Eng Bull, 2000. https://doi.org/10.1007/978-3-7908-1772-0_17.
[64] S.O. Abioye, L.O. Oyedele, L. Akanbi, A. Ajayi, J.M. Davila Delgado, M. Bilal, O.O. Akinade, A. Ahmed, Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges, J. Build. Eng. 44 (2021) 103299. https://doi.org/10.1016/j.jobe.2021.103299.
[65] L. Huang, A.D. Joseph, B. Nelson, B.I.P. Rubinstein, J.D. Tygar, Adversarial machine learning, Proc. ACM Conf. Comput. Commun. Secur. (2011). https://findanexpert.unimelb.edu.au/scholarlywork/740414-adversarial-machine-learning (accessed November 27, 2021).
[66] B. Hitaj, P. Gasti, G. Ateniese, F. Pérez-Cruz, PassGAN: A Deep Learning Approach for Password Guessing, in: ACNS, 2019. https://doi.org/10.1007/978-3-030-21568-2_11.
[67] J. Lorenzo-Trueba, F. Fang, X. Wang, I. Echizen, J. Yamagishi, T. Kinnunen, Can we steal your vocal identity from the Internet?: Initial investigation of cloning Obama's voice using GAN, WaveNet and low-quality found data, Odyssey. (2018). https://doi.org/10.21437/Odyssey.2018-34.
[68] M. Cheng, H.-C. Tsai, C.-L. Liu, Artificial intelligence approaches to achieve strategic control over project cash flows, (2009). https://doi.org/10.1016/J.AUTCON.2008.10.005.
[69] K. Dadteev, B. Shchukin, S. Nemeshaev, Using artificial intelligence technologies to predict cash flow, (2020). https://doi.org/10.1016/j.procs.2020.02.163.
[70] L. Zhu, M. Yan, L. Bai, Prediction of Enterprise Free Cash Flow Based on a Backpropagation Neural Network Model of the Improved Genetic Algorithm, Inf. (2022). https://doi.org/10.3390/info13040172.
[71] G.A. Hodge, C. Greve, Public–private partnerships: an international performance review, Public Adm. Rev. 67 (2007) 545–558.
[72] X. Zheng, Y. Liu, J. Jiang, L.M. Thomas, N. Su, Predicting the litigation outcome of ppp project disputes between public authority and private partner using an ensemble model, J. Bus. Econ. Manag. 22 (2021) 320–345. https://doi.org/10.3846/jbem.2021.13219.
[73] X.-H. Jin, Model for efficient risk allocation in privately financed public infrastructure projects using neuro-fuzzy techniques, J. Constr. Eng. Manag. 137 (2011) 1003–1014. https://doi.org/10.1061/(ASCE)CO.1943-7862.0000365.