IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received July 6, 2021, accepted July 26, 2021, date of publication August 3, 2021, date of current version August 13, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3102215

Towards an AEC-AI Industry Optimization Algorithmic Knowledge Mapping: An Adaptive Methodology for Macroscopic Conceptual Analysis

CARLOS MAUREIRA[1,2], HERNÁN PINTO[3], VÍCTOR YEPES[4], AND JOSÉ GARCÍA[3]
[1]Facultad de Economía y Negocios, Universidad de las Américas, Santiago 7500975, Chile
[2]Facultad de Ingeniería, Ciencias y Tecnología, Universidad Bernardo O'Higgins, Santiago 8370993, Chile
[3]Escuela de Ingeniería en Construcción, Pontificia Universidad Católica de Valparaíso, Valparaíso 2362804, Chile
[4]Institute of Concrete Science and Technology (ICITECH), Universitat Politècnica de València, 46022 València, Spain

Corresponding author: Carlos Maureira (carlos.maureira@gmail.com)

This work was supported by the CONICYT/FONDECYT/INICIACION under Grant 11180056 to Jose Garcia and the Spanish Ministry of Science and Innovation through the FEDER Funding under Project PID2020-117056RB-I00 to Victor Yepes.

ABSTRACT The Architecture, Engineering, and Construction (AEC) Industry is one of the most important productive sectors, hence also produce a high impact on the economic balances, societal stability, and global challenges in climate change. Regarding its adoption of technologies, applications and processes is also recognized by its status-quo, its slow innovation pace, and the conservative approaches. However, a new technological era - Industry 4.0 fueled by AI- is driving productive sectors in a highly pressurized global technological competition and sociopolitical landscape. In this paper, we develop an adaptive approach to mining text content in the literature research corpus related to the AEC and AI (AEC-AI) industries, in particular on its relation to technological processes and applications. We present a first stage approach to an adaptive assessment of AI algorithms, to form an integrative AI platform in the AEC industry, the AEC-AI industry 4.0. At this stage, a macroscopic adaptive method is deployed to characterize "Optimization," a key term in AEC-AI industry, using a mixed methodology incorporating machine learning and classical evaluation process. Our results show that effective use of metadata, constrained search queries, and domain knowledge allows getting a macroscopic assessment of the target concept. This allows the extraction of a high-level mapping and conceptual structure characterization of the literature corpus. The results are comparable, at this level, to classical methodologies for the literature review. In addition, our method is designed for an adaptive assessment to incorporate further stages.

INDEX TERMS Architecture, engineering and construction, AEC, artificial intelligence, literature corpus, machine learning, optimization algorithms, knowledge mapping and structure.

I. INTRODUCTION

The architecture, Engineering, and Construction (AEC) industry is speeding up in adopting the latest technological advances in the Artificial Intelligence (AI) and Internet of Things (IoT) ecosystems [1]–[3]. Applications of AI in the AEC industry (AEC-AI) have been developing from the early 70s in the last century, experimenting a substantial speed-up in the last 10 years [2]. The application of AI in the AEC industry brings considerable benefits across the AEC ecosystem in almost every aspect of the AEC workflow, therefore, an important transformation of the sector is expected. Communication and Information Technologies (CIT) will boost and reshape the industry in almost every aspect: from Edge Computing to 5G and blockchain technologies processes and operations will be substantially transformed. The results will be a migration towards an ever more important role of data generation, analysis, and integration [3], [4]. AI optimization will reach the more complex areas in the AEC industry, and robotic automation promise to recapture interest in an increasingly more sophisticated workforce along with new skills and talents

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

emerging in the technological sectors [5]. Security, adaptive business modeling, and sustainable technologies will increase pressure to redefine traditional construction and building life cycles. Edge computing, smart construction objects, cryptography, and data-block technologies, added to the new cultural workforce, social and sustainable responsibility will be some challenges that will be faced in a process of competitive transformation [6], [7]. The AEC industry ecosystem will converge into incorporating technological systems as much as health, finance, and high-tech sectors.

Several recent reviews and special issues conferences have addressed the state-of-the-art in Artificial Intelligence (AI) in the Architecture Engineering and Construction (AEC) Industry [1]–[3], [8]–[11]. Despite several efforts to advance in a systematic understanding of the mixture and interactions between the different processes been addressed by AI-*algorithms* and how they work in the AEC industry. At the time of also revealing where more effort and challenges need further developments, is still missing [2], [11], [12]. However, the research efforts addressing the strength and weakness in developing the AEC-AI industry requires advancing in multiple fronts in an integrated way across a very complex infrastructure of digital and physical components and processes [4], [13]. Here we develop a review in the state-of-the-art, complementary to previous efforts, addressing more specifically the developments and scenarios of algorithms' operation. Systematic development of these types of methods will detect and expose how algorithms are being developed and used, allowing the formation of a repertory of solutions for the AEC-AI Industry ecosystem. Providing empirical evidence in how the broad variety of algorithms and processes can be addressed and integrated into a multi-platform AEC-AI ecosystem allows an open and more integrated platform as well as a more easy, quick, and robust deployment of solutions.

Towards to characterize and analyze the AEC industry workflow, to reveal the different computational algorithms used, and how the AEC industry can be powered by a fully integrated platform using AI and the challenges to be solved, here an efficient global adaptive methodology is developed. In this article, a first step is implemented in characterizing one of the most frequent terms used in the AEC-AI industry in relation to the developments in AI [2]. We address what are the associated conceptual structures underlying a corpus of the literature associated with AEC-AI industry and `Optimization`. A difference from other algorithms and concepts used in AI, `Optimization` is involved in many aspects of computer sciences. Therefore, there are many processes and interactions where `Optimization` is used. `Optimization` can be used in its general assertion, as in any process where improvement and efficiencies are in aim, for example, in the optimization of the supply chain. However `Optimization` is a key term for mathematical extremization, then mathematical optimization is present in almost every algorithm. However, specialized areas as metaheuristics explore optimization in some specific areas. Therefore, `Optimization` requires a more in-depth analysis in order to gather specific algorithmic contributions to the AEC-AI Industry. In particular, we introduce an adaptive method starting from macroscopic or high-level conceptual mapping to progressively improve more detailed information as well as also allowing to adjust input/output data.

Methodologically, the rest of the work is grouped into 5 sections with the aim of developing a brief state of the traditional art and subsequently strengthening it with the proposed macroscopic methodology. In section II, the state of AEC-AI traditional art is recast along with their applications in the AEC sector. Later in section III, we develop the methods of our macroscopic approach. In section III-A, the stages of the proposed methodology are developed in a general way, and later in section III-B, it is detailed in the context of the analysis of the `optimization` techniques in AEC-AI. The results are detailed later in section IV. First in section IV-A, a brief summary of the results is provided for assessment of the adaptive method at the macroscopic level. In section IV-B, a more specific account of results is provided for general bibliometric assessment. Section IV-C, a general analysis of the references is developed for the last 10 years. Later in section IV-D, a more focused specialized characterization is carried out. Finally, in IV-E, a specialized characterization is developed, but looking at the last 3 years. Later in section V a discussion and comparison (V-A) with the traditional results obtained are developed to finally in section VI generate our conclusions and next steps.

II. AI AND AEC INDUSTRY

We complement previous results by proposing and developing an adaptive methodology to characterize context and AI algorithms' uses. Then, we present a first stage of a macroscopic adaptive approach to mapping the AEC-AI industry contents and their relationship to particular concepts, in this case, `Optimization`. To properly contextualize our results in the next section we will first recast the state of the AEC-AI industry. We review some antecedents already in the literature regarding the use of AI in the AEC industry. We summarize the results from few reviews directly addressing the AEC and AI topic [1]–[4], [9]–[11], [13].

In section II-A, we revise the AEC-AI industry to present an initial landscape. In section II-A1, we overview the main areas that are developed in a scientometric review of the AEC-AI industry. Next, in section II-A2 we discuss the most frequent keyword used in the context of the AEC-AI industry, revealing the major themes in a classical review process. Section II-A3, describes some of the main algorithms, methods, and use context in the AEC-AI industry. After that, section II-A4 recast a summary of the Knowledge Map in the AEC-AI industry formed by classical literature review methods: clusters of topics, the impact of themes, top journals, collaboration networks, and uses are included. In section II-A5, we review prospective development in the most promising and needed directions.

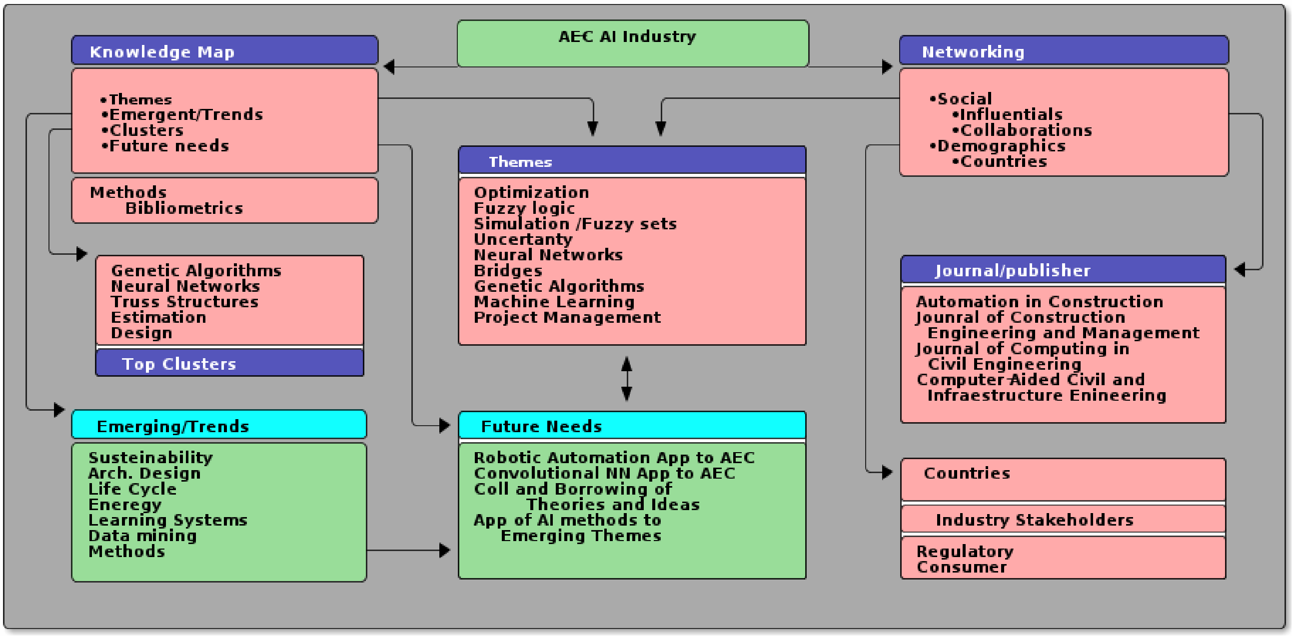


FIGURE 1. **Artificial intelligence in the AEC industry as is revealed by scientometric methods, modified from [2].**

Next, we make a brief introduction to the main platform in use in the AEC industry, building information modeling, which sets basic technology to interact with (section II-B). In this section, we include reviews pointing to key areas: sustainability (section II-B1), workforce (section II-B2) and infrastructure (section II-B3).

In the last section, II-C, we include recent reviews related to the AEC-AI industry, from the point of view of several other technologies called to interact in the development of an integrated AI-AEC platform (section II-C1). Internet of things (IoT, section II-C2), virtual and augmented reality with Vision-AI (section II-C4), blockchain and cryptography (section II-C3), and geographic information systems (GIS, (section II-C5)) will be briefly presented. We further include some areas where special interest is captured: energy, manufacture, and health and safety (section II-C6).

A. RECASTING AEC-AI

Advances taking place in the AEC industry and AI-related technologies have motivated several reviews, most of them focusing either on a specialized topic or in broad bibliometric overview [1]–[4], [10]. In [2], scientometric analysis on networks of selected keywords searches in Scopus resulted in almost 42.000 publications. An analysis of keywords' Networks of co-occurrence was obtained for document co-citation, citation burst, outlets, direct citation, and co-authorship. These analyzes were complemented with maps from network measures to extract conceptual, intellectual, or social evolution of the research field, discovering patterns, trends, seasonality, and outliers.

1) KNOWLEDGE MAP OVERVIEW

A scientometric literature analysis suggests a state-of-the-art in AEC-AI as depicted in the Figure 1 (modified from [2]). Two major areas are delineated in developing an assessment of the AEC-AI Industry: one referent to knowledge map and other characterized by networking of actors and institutions. It reveals areas of knowledge contents including major themes/topics/concepts (Figure 1, center box), trends and emerging fields and areas of developments, cluster of closely related concepts and themes, and finally needs to future developments can be estimated (left and middle branches in figure 1). For the case of networking, Social and demographics are the major areas (Figure 1, right column). In social areas, the most influential actors and collaborative teams are mapped, evaluating the impact of research. In the demographics characterization, the assessments values the institutions, organizations, and country's status in relation to the workforce in the global concert of research. Both knowledge map and networking jointly define the themes and topics more developed or more frequently used in the sourced data.

2) THE MOST USED KEYWORDS APPEARING

When the key-terms are analyzed in an independent methodology, several terms related to algorithms in AI are recovered. The most frequent concepts/terms correspond to Optimization, Genetic Algorithms, Neural Networks, Fuzzy logic, Fuzzy sets, Machine Learning [2], [12]. Notably, Optimization becomes the most frequent followed by genetic algorithms, both terms of which are broadly used in metaheuristics algorithms. In particular, optimization is a sort of general term used in a special context of algorithms and in the general problem of extremization involved in solving many computation and functional problems. In some sense, optimization is equivalent to machine learning. Therefore, optimization in the AEC-AI industry is a key term that needs a more in-depth characterization.

3) METHODS & ALGORITHMS

For use cases, most AI applications in the AEC industry rely on computer vision-related algorithms. Vision, virtual and artificial reality methodologies are more extensively used in the context of building information modeling (BIM). One important area of development is under the umbrella of Smart Construction and Smart cities [12]. BIM associated processes and workflows are central in the AEC industry and relay this importance to the development and integration of AI genealogies. Two main paradigms are used for BIM generation by AI: as-planned-BIM and as-built-BIM. In this context, recurrent convolutional neural networks (R-CNN) and deep neural networks (DNN) are the AI algorithms most used. Recurrent neuronal network (RNN) and hybrid RNN/LSTM models are used to classify human activity recognition (HAR) [14], [15], in some cases with 80-90% accuracy [14]. Deep Learning for image based inspection of concrete defects for example, where deep learning-based classifiers can reach 95.6% accuracy [16].

4) STRUCTURE IN AEC-AI KNOWLEDGE DOMAIN

Here we summarize the main finding after a scientometric analysis of Scopus derived data [2]. *Clusters.* 13 Major clusters were detected, well defined, and isolated. The top (5) most prominent corresponds to: Genetic algorithms (GAs), used in a variety of optimization and search algorithms. Neural Networks (NNs) which although there are some of the first implementations in AI, new paradigms as deep neuronal networks make emerge a new landscape of use and theory. Truss Structures, which in relation to AI, some optimization applications associated to design have been developed. Similarly, the design concept is present in many objectives in the AEC industry, from automated design in BIM to structural and architectural optimization design for energy efficiency [17]–[19]. Whereas, estimation is a general concept that, in the context of AEC-AI, is related to prediction and assessment. From this cluster analysis, the newest to appear in the cluster #10, corresponding to convolutional neural networks (CNNs) in 2017. The smallest set in clustering contains 10-20 studies and includes forecasting systems (for construction material prices) and Structural Control Implementation [2]. *Impact of themes.* As when the impact of different themes is examined, it's found that Expert Systems (ES), correlate with the rising start of IT in the eighties (1985-2003). Whereas, Knowledge-based systems (KBS) received much attention, but it does not reach many applications to AEC problems (1988-2010). The latest burst occurs for: sustainable development, architectural design, life cycle, energy-related issues (e.g., energy efficiency), learning system, data mining, intelligent systems, numerical method, performance assessment, stochastic systems, Monte Carlo method, reliability analysis, strength, and particle swarm optimization. *Top Journals.* Publishers in the most influential spots are Automation in Construction, Journal of Construction Engineering and Management, Journal of Computing in Civil Engineering, Computer Aided Civil and Infrastructure Engineering. Besides the previously well positioned publishers, others with specialized journals with less impact also offer high-quality works with direct connection to the AEC-AI industry. *Collaboration.* At the level of networking and social characterization, the field appears with low levels of collaboration between institutions. The countries better positioned are the USA and China at the top, followed by Australia, United Kingdom, Hong-Kong, there are 67 countries publishing more than 20 publications and Spain is 15th, whereas Chile does not reach the 30th. *Uses.* When use cases are addressed, Optimization shapes optimization problems in structural design. Other algorithms as Particle Swarm Optimization (PSO), Genetic Algorithms (GA), Harmony Search Algorithms (HSA) do not appear with simple uses or are missing in their use. The most conspicuous use cases appear related to: Simulation, Uncertainty, Project Management, and Bridges.

5) APPLICATIONS AND AREAS OF DEVELOPMENTS

There are many areas where the application of AI along with partner technologies (section II-C) can bring considerable advances, and where it is important to also generate a relationship between the algorithm and the decision-maker, making explainable-AI (XAI) more in demand [12], [20]–[22]. We will briefly present some of them as most relevant in prospective developments:

Interestingly, in the AEC industry, there are some concepts that are considered key performance indicators. However, they are not yet addressed by AI methodologies or they are been weakly linked to AI concepts [2]. Some key performance indicators are cost, productivity, safety, risk management, and quality. Costs have been weakly linked to case-based reasoning (CBR) and GA, but no links to BIM have been found; whereas risk management has no link to BIM, which has been shown to be an efficient alternative to risk management technologies [2].

Earthquake Engineering: In this area, active and passive engineering methods are assisted by AI, and combined with smart monitoring applications are welcome in the complex scenarios of catastrophic events as well as methods for quick assessment and damage mitigation systems. *Civil infrastructure:* infrastructure is crossed by many important elements, from connectivity and development of isolated regions to rapid deployment of hospitals and shelters. Similar to previously prospective applications in Earthquake engineering, integrated health, and monitoring system (HMS) with smart capabilities operating across civil infrastructure networks allows resilience and mitigation plans as well as logistic efficiencies in economic systems. *Building structures:* for the case of building structures, energy efficiency, security, comfort, and management, receive the most attention. Again, recent developments incorporating technologies in the IoT spectrum will boost and consolidate trends in the not too far-away horizon.

Although that applications emerge along with the AEC industry, there is more to be made in specific areas, including: Energy, Thermal comfort, Life cycle, costs, Optimum design, Life cycle assessment, Sustainability, and Quality [19], [23].

What to expect for tomorrow, between the most notable projection the arrival of Robotic Automation. Developing advanced, usable, human-friendly, and smart robots may be around already. Convolutional Neural Networks probably will expand as much as quality data become available for different applications. Damage detection, advanced techniques are always put to test in such as delicate fronts, anticipatory techniques, and damage evaluation for mitigation and preventive management. Similarly, cost efficiency, automated monitoring, and maintenance systems for facility operations and management.

The building information modeling system is one anchoring many developments, from all in one technology solution [1], or for specific practical purposes [24]. However, the most advanced prospected uses of AI are still in need of further integration into BIM technology. Methods including laser scanners and digital photogrammetry, BIM modeling, IoT sensors, all produce data that may be shareable through Cloud-based solutions. Although, still the industry seems to lack a coordinated effort to identify new streamlined workflow paradigms whereby the entire process, creating and managing the BIM model, would form a comprehensive tool [3]. In any case, data will be also at the center of the new technologies been incorporated, transforming also the AEC industry towards a complex ecosystem where DATA will be another valuable construction material with the less hard division between classical constructive processes and digital realm applications [3]. Around 42000 records were found related to AEC and AI industry [2], which makes desirable complementary methods to get more granular information on the contents of the articles. Next, we will overview some of the literature corpus in AEC AI Industry, from a point of view that contributes to the AEC AI industry development.

B. BIM: AEC INDUSTRY BACKBONE

Many AEC firms have reported significant competitive advantages following Building Information Modeling (BIM) implementation. However, BIM adoption requires an important change in organizational structure. Enablers of this process are being identified, including strategic initiatives, cultural readiness, learning capacity, knowledge capability and IT leveragability, network relationships, process and performance management, and changes in management practices [25]. New methodologies focused on sociotechnical dimensions as context, process, organization, task and actors may also be useful for BIM-enabled projects and BIM-based construction networks [26] and safety [25]–[27]. Scientometric ontology investigation has revealed how development from "project management" and "knowledge management" to "building information modeling," and "compliance control" leads to the successful adoption of information techniques in the construction industry. Ontological perspective identifies, using a combination of cluster analysis and critical review, "Domain ontology," "Industry foundation classes," "Automated compliance checking," and "Building information modeling" as the key research themes [28]. AEC industry is still engaging in incorporating and exploiting BIM, although the confrontation of the industry 4.0 and the challenges and benefits are on the horizon, decisive implementation across the industry is not spread [29]. Particular aspects of adopting BIM in the AEC have been reviewed, from ontological perspectives and enabling factors for adoption, to sustainable practices [28]–[32].

1) SUSTAINABILITY

Another area where BIM can be potentiated because of the awareness of the increasingly important impacts of the construction process is sustainability. BIM integration with life cycle assessment (LCA) improves the whole-building life cycle [23]. There are advantages and disadvantages to BIM use. Tendering and initial states enable quick and detailed assessment, as well as the possibility to check along with stakeholder correctness of projects [33]. In particular, quantification of materials with different alternatives selection of sustainable materials in the early design phase can contribute to quick and accurate quantification and evaluation. However, reusability and recyclability of the materials and refurbishment and demolition process of a building are not comprehensively considered in LCA [30]. Although that holistic green BIM is still in early phases, the global challenges in sustainability and controlled reduction in CO_2 emission will hopefully incentive the use, as well as incorporation and development of methodologies to overcome the barriers for environmentally responsible developments [31]. One important way to advance in sustainability efforts is by rationalization of process and products; for example, the analysis of construction's demolition waste and reverse logistics to propose plans of action [34].

2) WORKFORCE

Although the level of technological adoption in the AEC industry has been variable, from low levels to increasingly more compromised, new global status makes more likely acceleration in a new competitive phase, and low levels of adoption may start to change [35]. Industry 4.0 implementation implies not only achieve technical objectives in the development and integration of technologies. The new professional profiles of the workforce also need an evaluation. The range of skill need is wide and complex, whereas knowledge of fields may need a type of factory of knowledge platform [5]. An important component in the AEC industry that will also suffer transformative innovation is the workforce. The profile of workers under the new paradigm of Industry 4.0 results in a worker 4.0 conceived to streamline productivity performance and management [36]. Recommendations for future education include: BIM as the digital platform, involving BIM-standalone and BIM-embedded courses; collaborative BIM aimed to reduce the fragmentation among different AEC disciplines (e.g., interdisciplinary project-based learning). Increase information sharing between BIM and other digital technologies to improve integration (e.g., laser scanning). Similarly, continuously performing educational innovation programs to update technical and managerial digitization capabilities as well as digital transformation [5], [35], [37], [37]–[39].

3) INFRASTRUCTURE

Structural Health Monitoring (SHM) is one area where demand will increase related to industry 4.0 and the AEC industry; maintenance, monitoring, resilience, and disaster preparation will push smart monitoring networks as cyber-physical infrastructure in modern cities. Wireless sensor network technologies for SHM will be central to future developments connected to the AEC industry [40]. Design limitation in the development of IFC also needs to play an active role in implementing AEC industry technologies as for example BIM into particular applications or disciplines as the case for specialized infrastructure [41]. In the fatigue condition assessment for steel bridges infrastructure, for example, data-driven assessment approaches may also benefit from AEC industry technological advances [42]. Another interesting development comes from an integration of visualization and communication technologies, it is expected that 5G added to other core technologies like AI and cloud computing can improve the use of virtual/extended reality using BIM-based workflows across the entire spectrum of stakeholders [43]. BIM-based XR technologies' potential application include areas such as design management, safety management, progress monitoring, conflict avoidance, and decision making, whereas the challenges are defined by technology, content, and users [43].

C. PARTNER TECHNOLOGIES

Most of the published work recognizes the synergistic interaction occurring in construction with the potential to transform the productive sector. Different technologies such as blockchain, virtual and enhanced reality, cloud, fog, and distributed computing as well as sensing, all of them with enormous potential to be integrated into AI concert. Similarly, integrated systems using BIM platforms in combination with safety and risk monitoring and assessment are required [4], [44]–[46]. Probably these efforts and analyses make a bolder requirement to introducing Big-data into the AEC-AI industry as a cognitive platform for human development.

1) AI INTEGRATED NEEDS

Is expected that the integration of AI technologies will decrease the risks of delay and cost overrun; problems of lack of trust in traditional construction projects; online management efficacy; as well as in benefit and efficiency in progress management. Similarly, safety and quality in construction projects, especially for remote management, are very attractive. Trusted transactions mediated by blockchain will affect automation of contracts, payment, real-time monitoring, and government regulations and compliance. However, more studies are needed to make each required technique meet the requirements in the real-life construction projects [4]. In BIM integration, because of the large dataset required, a common problem for deep learning applications translates into expensive computing costs and the low ability of generalization of the deep learning models. For as-built BIM model generation, end-to-end algorithms need more efficiency to achieve production. Similar problems also occur in safety and quality management [4]. Importantly, for AI in the AEC industry, the fragmentation of knowledge remains unresolved. A platform for integrating multiple technologies is still needed to achieve the ''end-to-end'' construction management [4].

Safety recognizes four crucial stages of action in analyzing risk management: identifying hazards, assessing risks, controlling risks, and reviewing control measures. In these areas, data-driven prevention systems can be implemented using: wireless sensor networks, RFID, UWB, IMUs (accelerometers and gyroscopes), and vision-based techniques [47]. Then, location tracking, gait analysis, object detection, activity recognition, ergonomic assessment, etc. can be integrated into addressing risk and safety [44]. Several areas of improvement are being identified: an integrated framework combining information obtained using a variety of methods, which also may include real-time monitoring and analysis, including Big Data Engineering and Data Analytics. Another is, for example, near-misses assessment of safety, which may include a vision-based monitoring system with BIM platforms [47]. Similarly, the availability of large datasets can allow for AI algorithmic training in construction safety monitoring for benchmarks. Quantitative analysis and monitoring systems for accident precursors for fall, struck-by, caught in or between, and electrocution [44].

Lack of interoperability between devices, high complexity of setting up a smart home network, and absence of unified interfaces for device management are issues that have to be solved. Similarly, privacy and security are probably on top requirements to advance in smart infrastructure all the way from smart-housing to smart planet [45]. The construction industry generates massive amounts of data throughout the life cycle of a building, however, the adoption of Big Data technology lags in relation to other fields. While data-driven analytics have long been used in the construction industry, the adoption of the recent agiler and powerful, Big Data technology has been slow. However, Big Data applicability is amplified by many emerging trends, such as BIM, IoT, cloud computing, smart buildings, and augmented reality [46].

2) IOT
Narrow sub-domains for Big Data applications, although big data technologies are the future. Similarly, hardware, software, and algorithms are recognized as elements in smart building development, but still, efforts are required to integrate technologies [13], [48]–[51] For the domains of Construction Operation and Monitoring, Health & Safety Management, Construction Logistic & Management, and Facility Management, integration methods are using BIM tools' APIs and relational database with new data schema and query language using semantic web technologies and hybrid approaches [49]. One area that can be very reshaped by industry 4.0 technologies is the construction supply chain, where continuous interchange of information enables sophisticated workflows [50].

3) BLOCKCHAIN
Blockchain or Distributed Ledger Technologies (DLT) is being recognized as a key technology in the orchestration of many deficiencies in the AEC industry trust. Blockchain/DLT defined as a decentralized ledger that registers every transactional operation made in encrypted data format. Consequently, several areas of use, as well as plans and recommendations to adoption, have been made [52], [53]. Recent reviews of uses and developments of Distributed Ledge Technology, identify seven areas of use of blockchain. Five areas of blockchain expect application in the SMART variety including energy, cities and economy, government, homes, and transport. Whereas that another two areas are in the core of the AEC industry, BIM and construction management as well as a business model and organizational structures [52]. There are also proposing conceptual models for a roadmap to advance the adoption of DLT in the construction sector [52]. Commissioning distributed ledger technology (DLT) in the BIM working processes provides increased security and allows automation of processes. Which overall, reinforces network security, including reliable data storage and management of permissions and ensuring change tracing and data ownership. This way enables an enhanced framework for automating code conformance checking process [53].

4) POINTCLOUD, VR & VISION
Four review works were obtained in relation to the application of VR, computational vision and point cloud in the AEC industry [47], [54]–[56]. This covers improving communication, geometry reconstruction, and quality evaluation in as is BIM. Geometric diagnostics, starting with edge detection and blob detection, coupled with color segmentation, operation boundaries, and occupied areas (by materials and equipment) can be measured; and facilitate scene understanding and global reasoning in construction and building domains. These quantitative measurements are useful, leading to improved monitoring, safety, productivity, and quality [54]. When the case of applications of VR in the communication area of the AEC industry is made, improvements were mainly found in building inspection, facility management, safety training, construction education, and design and review [17], [18], [55]. A fifteen-year review of 3D point cloud identifies two key areas of development, 3D model reconstruction, and geometry quality inspections. 3D reconstruction includes geometry models and semantic models. Geometry quality inspection applied to construction works includes dimensional evaluation, surface quality, and displacement inspection. Additionally, a spectrum of other applications can be found as: construction progress tracking, building performance analysis, construction safety management, building renovation, construction automation, heritage applications, and robot navigation [56].

5) WORKSITE & GIS
GIS can contribute to the AEC industry, however is probably one area where more efforts are needed to be made from practical issues to ontology development in the standards [57], [58]. GIS and BIM were originally developed for different purposes and their integration faces many challenges. Geometric and semantic information transferred from building modeling to a geospatial context will positively influence a series of current activities, such as site selection, safety management, and environmental impact assessment. Whereas GIS can be enriched with its true 3D by being integrated with BIM. However, standards also need further development. For example, in an IFC model, additional semantics to show how some aspects are to be modeled in CityGML [57]. BIM provides rich geometric and semantic information through the building life cycle, while GIS covers implementations in geovisualization-based decision making and geospatial modeling. BIM-GIS integration benefits and helps management methods and coordination mechanisms. Directly affecting quality and progress management, reduction of time and cost improves control and HSE performance. Allowing more integrated information management and the coordination of various sectors [58].

6) SPECIAL INTEREST
a: ENERGY
Energy efficiency and comfort probably will become very interesting areas of development, from the design phase in architecture and engineering to parameter tuning in building operations [17], [19]. New and classical algorithms will play along with the most sophisticated area of integration in conceptual design [17], [59], [60]. BIM-based design processes in structural engineering and life cycle energy domains may allow evaluation frameworks that compute the performance criteria of sustainable and energy efficient structural systems, at the same time consolidating intuitive decision workflows. Nevertheless, intelligent and efficient design of building components needs the integration of several components including building and construction automation processes, business systems, and information technologies. In addition,

enhanced knowledge transfer between the various disciplines and organizations in the building industry is required [60]. Recently, [59], energy consumption optimization has been surveyed in the context of IoT and smart buildings. Many classical as well new algorithms have been tested in comfort and energy consumption, however, it's been found that genetic optimization algorithms (GOA) obtained the best performances. However, there is plenty of room to test new and innovative approaches as there is a gap at the moment in considering parameters and preferences by users. Additional deficiencies were also detected in relation to meteorological or external environmental variables at the time of optimizations. Notably, also new nature-inspired optimization algorithm where introduced, as for example a crow search optimization algorithm [59].

b: MANUFACTURE

Benefits of manufacturing practices in the construction industry require framework/platforms developments, however, the benefits in social and environmental responsibility may start to integrate a considerable factor in relation to purely economic considerations [7], [61]. There also important advances proposed for a conceptual framework for synchronization of prefabricated housing production (PHP) and BIM use in smart BIM platforms, smart work packages, and smart PHP processes. This proposed integration tries to articulate from design to deconstruction and recycling amid to facilitate decision-making; modular and manageable work processes to link and handling. In this way, a chain of prefabricated products/resources with tracking, sensing, processing, storage, and communication capability are obtained [7]. At least 32 Lean Construction Practices (LPCs) implemented in the construction industry were identified, with last planner systems and just-in-time (LPS/JIT) procedures the most globally documented. Overall, LPCs can allow to obtaining close to 20 different benefits in economic, social, and environmental domains. LPCs were implemented at planning, design, and construction stages, both in buildings and infrastructure projects [61].

c: HEALTH SAFETY

A very important area of research is concerned with health and safety, in construction sites as well as in the area of occupational health. The availability of new technologies in wearable monitoring systems will allow for rapid and opportune decision-making for decreasing risk in construction sites and in worker operations [62], [63]. Direct measurement sensors are the most common method used for the investigation of occupational health and safety (OHS) in the construction industry. Sensing- and warning-based technology applications provide complementary sensing networks for construction site safety management and monitoring. The areas to incorporate these technologies include safety risk identification and assessment, intrusion warnings and proximity detection, physiological status monitoring, activity recognition and classification accuracy, and structural health monitoring (SHM). Between the areas of developments of sensing- and warning-based technologies, integration of other advanced information technologies is also required [62]. Wearable wireless system based on inertial measurement units (IMUs) provides non-invasive, long-term, and ubiquitous tracking of body postures and motions; derived data analysis can recognize postures and evaluate them in real-time. This type of technology may have a great impact on the construction industry, alleviating the poor record of occupational health, in particular regarding work-related musculoskeletal disorders (WMSDs) [63].

III. METHODS

A. MACROSCOPIC METHODOLOGY: GENERAL APPROACH

Search Engine: as a first approach, we gathered records from the open multidisciplinary CrossRef platform. CrossRef adhesion has been growing continuously and contains between 70-90% of the records provided by other paid platforms [64]. The metadata available in CrossRef also provides support for analysis, assessment, and development for license use and contains a larger collection of open-access articles [64]. Although we choose to CrossRef in attention to its features related to multidisciplinary, open-access, and availability of tools, these methodologies can be adapted and used with other sources of data.

Analysis Platform: We used R (v4.04) along with text analysis tools and rOpensci [65]–[75].

Constrained Queries: We created a basis search query construction to retrieve records from the CrossRef service. CrossRef engine does not seem to handle direct boolean operators at the moment.[1] Because of these constraints in the implementation of the searching engine, we constructed three different but related search queries plus a basic query. We used a `basis query` including full wording and abbreviation for the AEC-AI Industry[2]:

```
(architecture+ engineering+
construction+ AEC+ industry+ building+
information+ modeling+ BIM+ artificial+
intelligence+ AI+ machine+ learning+
ML+)
```

To this `basis query` we can add specific KEY-TERMS and constraints to sources and publications dates as we have made for assess the systematic impact of them.

Here we present analysis and results obtained for three main constrained queries (see figure 2):

- `General Query`: The broadest general query, the only constraint being the publication dates beginning in 2010-01-01.
- `Specialized Query`: Specialized journals beginning in 2010-01-01. Specialized records related to a set of journals (table 1) from the last 10 yrs.
- `Specialized Recent Query`: Specialized journals beginning in 2019-01-01. Specialized records

[1] https://community.crossref.org/t/crossref-api-questions-test-queries-example/999

[2] Actual strings were constructed with no blank spaces.

TABLE 1. Main journal with publication in AEC.

Journal	issn
Automation in Construction	0926-5805
Journal of Construction Engineering and Management	1943-7862
Journal of Computing in Civil Engineering	1943-5487
Engineering Structures	0141-0296
Mathematics	2227-7390

related to a set of journals (table 1) from the last <3 yrs.

An important way we found to constraint and getting more aim-related searches is using publisher-associated metadata. We specified a specialized set of journal-related groups by including a set `issn` codes. Associated with the journals in table 1, we choose this set of journals by its related content in AEC Industry and its relation to mathematical `Optimization`. Obviously, these can be adjusted as it seems necessary. The basic bibliometrics characterization includes the journal retrieved in the search, the *time cited index*, publication dates, and *relevance score*, as is shown in figures.

1) DIAGRAMMATIC PROCEDURE

We summarize our method in the diagram in figure 2; we begin by setting a KEY-TERM which defines the aimed concept for this work `Optimization`, stage (1) in the diagram. Next, the KEY-TERM is composed with the basis query defining the AEC-AI industry. In this work, as is mentioned before, we have defined the basis query by: "information+ modeling+ BIM+ artificial+ intelligence+ AI+ machine+ learning+ ML+," stage (2). The composed search query is then used to query the database, stage (3), in this work CrossRef was used, or any other database via API (ex. Google scholar, WoS, etc.). The retrieved query results are then used to perform full-text analysis, as in stage (5) branching, on one hand. Basic bibliometric analysis is performed over metadata, to define constrained queries as described before (general, specialized, and specialized recent), towards an adaptive refinement in the branch staged (4).

After a KEY-TERM is defined, we perform the query, and the retrieved results are analyzed. Often, depending on the number of retrieved records, a sample of one thousand records is retrieved for analysis. Retrieved results are chosen based on the relevance index provided by the CrossRef algorithm. A figure is formed to summarize the relevance score and times cited indexes and the subjects disclosed by the authors/publishers and publishing journals. After that, we usually performed one of the established constrained searches described above, or any other defined necessary. This process is iterated according to the research requirements. We try to leave as open as possible the search criteria to achieve some level the noise acceptance in the retrieved records.

Using `Optimization` as a KEY-TERM the results available allow us to constrain the search to dates for records in the last 10 or less than 3 years, and to a group of highly leading journals; as this stage, we did not attempt to further characterize the combined bias generated in the method yet, mainly because it is not the limiting step in the targeted aim (see results sections). We observed that for the `Optimization` KEY-TERM, a considerable fraction of records did not contain `Abstract` section (see results section). Therefore, after a basic characterization of the literature corpus derived from the search in the CrossRef platform, we evaluated the availability of full-text articles as open-access and in platforms as arXiv.org, researchgate.org, personal collections, etc. We manually sorted the full text to separate them into articles and reviews type (Thesis, patents, or other documents were discarded and not further analyzed). Then, we performed extractions of `Abstract` section, `Conclusion` section and `All` sections using techniques for data science and ML&AI, see results section.

2) ANALYSIS

Analysis: a basic analysis of textual data content derived from the extracted sections (`Title` section, `Abstract` section, `Conclusion` section, `All` section) is performed depending on the methods. Extracted sections, directly from CrossRef metadata or by extracting the utf-8 text from full-text pdf:

- `Title` Section: Title section of articles directly retrieved from CrossRef title field.
- `Abstract` Section: Abstract section retrieved from full-text articles.
- `Conclusion` Section: Conclusions section retrieved from full-text articles. We consider the "Conclusion" section any section labeled or variations as "Final Summary," "Conclusions and Perspectives," etc.
- `All` Section: Full-Text, all sections from Abstract to end of Conclusions, i.e. from a full text we discarded all ancillary sections, title and author and authors related information and references.

3) VISUALIZATIONS & THEMES MAPPING

Hierarchical clustering, conceptual structure, and thematic mapping were performed and visualization created for figures [67]; see figure 2, and results section. Other analyses using NLP techniques, Machine learning, literature review are coupled on an ad-hoc basis (figure 2).

Hierarchical Clustering: We performed agglomerative clustering, each observation is initially considered as a cluster of its own (leaf), then the most similar clusters are successively merged until there is just one big cluster (root; [76]). The result of hierarchical clustering is a tree-based representation of the objects, which is also known as a dendrogram (figure 6) [76].

Phylogenic Dendrogram: Is a phylogenic tree, a branching diagram or a tree showing the similarity relationships among the various entities based upon measurement of distance of similarities and differences in their characteristics.

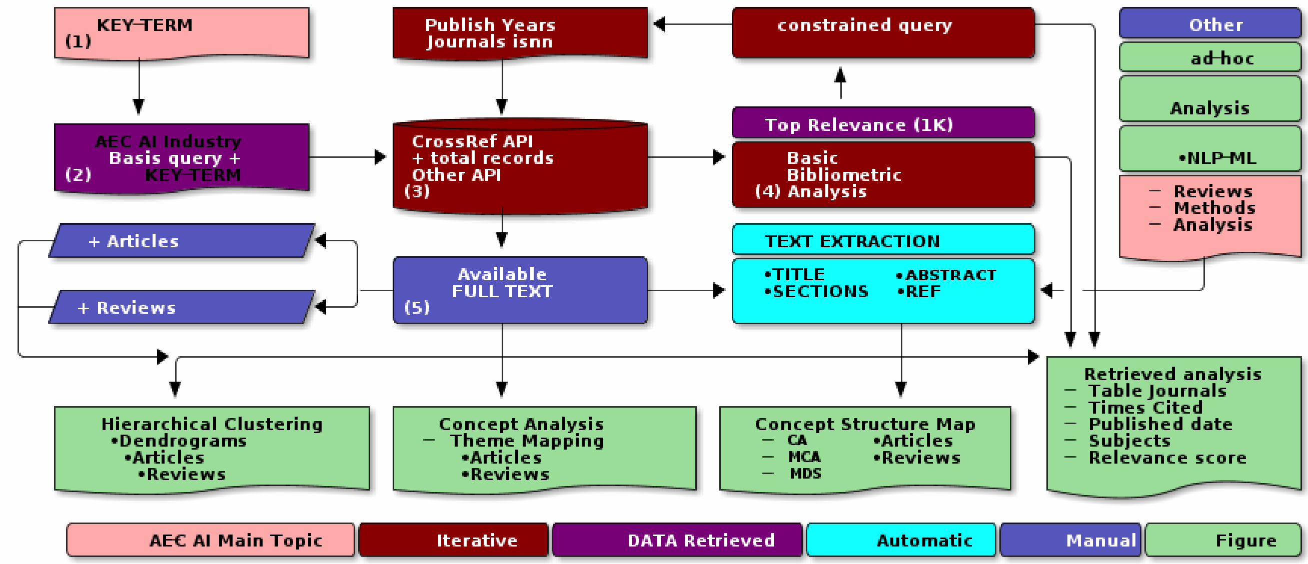


FIGURE 2. **Diagrammatic implementation of the proposed adaptive method for CrossRef database analysis. To aid in grasping the working flow of the method, we have numbered key stages from (1) to (5). First, a target KEY-TERM is formed in (1), in the case presented here is** `Optimization`**. After the key-term definition, it is combined in (2) with the** `basis query` **defining the AEC-AI Industry, which is used to perform the query upon the CrossRef database (3). At this point an iteration process performs the adaptive core by i) using other API, ii) operating the loop towards (4), and iii) implementing a full-text analysis towards (5), see methods section (III) for further details.**

CA & MCA: Correspondence analysis and multi correspondence analysis was performed using bibliometrix after some changes or supplementation.

B. MACROSCOPIC METHODOLOGY: AEC-AI OPTIMIZATION

In [2], it is reported that one of the most used terms in AEC-AI developments was `Optimization`; a general term associated with many stages in AI algorithms and process. Here we present a macroscopic adaptive method which can be extended algorithmically towards meso and microscopic levels of assessment analysis for topical structure map, content modeling and knowledge map of themes of interest in the AEC-AI literature and research. In particular, we begin here the evaluation of the most prominent concepts used in the elaboration of AI&ML algorithms used in the AEC industry. The proposed method complements and extends reviews of literature that uses classical methodologies. The proposed method presents several advantages as decreasing biases, big-data extensibility, and adaptive control of representation detail. Therefore, the first KEY-TERM to study was `Optimization`. Specifically, we study how some of the top-most frequent terms associated with `Optimization` and AEC-AI are being used, and then elaborate an initial landscape of operations which can be integrated to the needs of the AEC Industry.

We describe the major results of our Macroscopic Approach in sections IV-C, IV-D and IV-E. We used a source refereed approach in combination with a still representation for recent developments. This macroscopic approach is chosen in relation to several important constraints. In section IV-C, we characterized the retrieved results using a broad search criteria, only limited to the time lapsed from publication. In this case, the publications were limited to the last 10 years to maximize the number of full text available over the time period. In section IV-D, we performed the basic same pipeline of analysis as in section IV-C, however in this case we constraint the results to a set of related journals. Similarly, in this case, we also limited the retrieved records to the last 10 years. Finally, in section IV-E the same pipeline of analysis was performed on the last constraint used, comprising the same set of related journals but only for the <3 years. This overall macroscopic approach allows: i) Use an open-access multi-disciplinary platform. ii) Take advantage of the underlying relevance index implemented by CrossRef. To our best knowledge, the relevance score uses fuzzy algorithms for content matching. iii) The method allows to characterize any field/area with minimum constraint specific to the field. iv) It also allows to combine this macroscopic approach with complementary methods and analysis. v) Combine integrated metadata and downstream analysis.

Overall, the macroscopic approach allows progressively increases the level of detail in the characterization/search process. We discuss the obtained results in V, where the major advantages and disadvantages are delineated. We compare to other literature review characterization in section V-A, a specific relation is made to bibliometrics analysis and classical systematic reviews.

Finally, in section VI we integrate the results obtained, and the approach used, as well as future developments.

IV. RESULTS

Next, we will briefly review the main results obtained as an illustration of the use of the adaptive method in the context of the macroscopic stage approach.

A. SUMMARY OF RESULTS: MACROSCOPIC APPROACH

The number of records associated with the search query AEC-AI industry with `Optimization`, yield `591090` documents, many more than other related searches using a different platform [2]. We have retrieved records for the top one thousand search results, provided by CrossRef, for each query performed. A basic bibliometric analysis and description was implemented, as CrossRef does not provide methods to retrieve boolean connected records. Disclosed subjects and containing journals show a broad range of retrieved documents. Predominant journals and the average times cited index for retrieved records were used to constrain the search queries. Automation in Construction appears as the journal containing the highest impact (times cited) index. In agreement with other studies, the best-ranked journals appear consistently across searches, either as the most cited and highest impact [2]. We implement complementary queries to limit journals and dates. The mode of records was disclosed in the subject of engineering, construction, and AI. No relationship was clear between relevance score and time cited index from CrossRef metadata. We found that the fraction of articles that contained abstracts in the metadata was around 30%. We collect a more reduced set of full-text articles from the 1K top relevant works retrieved in the search. A coherent sample of subjects and journals was obtained for the available full-text articles. We got a representative sample of full-text articles to analyze with ML and textual-NLP methods.

From the available articles collected, we separate them into Review and Articles types and performed the analysis in each set. We have collected about 100 full-text articles and `43` Reviews type articles that were in the Top 1K retrieved from CrossRef. Therefore, from the 1K more relevant records associated with the search, we got close to 15% of available full-text documents. Hierarchical clustering allows describing the conceptual content and relative differences between the underlying literature corpus. We used the `Title` section and `Abstract` section to perform aggregation clustering from the set of full-text documents. A review set of articles were used as a reference point to compare the textual content analysis as a curated source of data. Reviews derived analysis of `Title` section and `Abstract` section contents, allows comparing to a curated reference source the overall content structure in the related articles. Successful conceptual structure map, by the correspondence analysis (CA), multi-correspondence analysis (MCA), and multidimensional scaling (MDS) methods were formed. Conceptual structure and theme mapping revealed materials and properties linked to the AEC-AI industry and `Optimization`, for example: `material, strength, mortar, concrete, properties`. Other clusters were determined for `construction, industry, study`; covariant with the cluster `optimization, concrete, approach`, interpreted as the aim contents. Whereas that the `bim, build, model, design, system`, and `machine, learning` cluster along with `approach, optimization, concrete` to name a few examples recovered by the macroscopic approach to high-level contents in the literature corpus analyzed.

Hierarchical clustering and dimensional reduction allow describing the conceptual content and relative differences between the underlying literature corpus. `Title` section and `Abstract` section derived analysis from the Articles type set of full-text documents display more heterogeneity of clustered terms, showing a probably higher variety of topics than the carried by Reviews type. Overall, these clustering schemes agree with a setting of articles' objectives/aims/problems and probably two of the main approaches expected: `machine learning` associated and `optimization` related. Consistent with these interpretations, equivalent content analysis derived from `Abstract` section from Reviews type of the articles literature shows a main variation axis covering the `machine learning` to `optimization algorithms`, and `material-structure` component of less variance. The results of the analysis show that `optimization` and `machine learning` related works are been used in the design of material properties and in the process of integrating "bim technologies." In agreement with these, the analysis of the reviews associated with the retrieved search also shown emphasis on these aspects as well as their advantages and developments.

Using constrained search queries for specialized journals with high impact (about 70% of the sample) allowed a more specific conceptual description. The sample contained articles published from 2014 to 2021, showing a clear correlation between time lapsed from publication and the TC index. In some cases, residual variance projections display a continuous scattering, making it difficult to extract concepts in clusters forms. We observed that for the most influential journals considered in the AEC-AI industry field [2], the last few years (<3yrs) trend is slightly different for the case of the `Optimization`.

A third constrained search query was used to isolate, from a sample of the top one thousand most relevant documents, a sample from the top 3 journals accounting for more than 60% of the records in the sample. From a sample of the available-full text documents (`76`) we found most of the available full-text articles in the sample are filed in "civil engineering" and "construction management"; Automation in construction is the leading journal with highly times cited index by article. Correspondence analysis and clustering techniques help to visualize and extract the conceptual structure of the literature corpus, however, it becomes more difficult to visualize as the number of documents increases, in the visualization and in the computation requirements. Specialized and recently constrained type of search queries allows to better define the conceptual structure of the field research by selecting the most predominant sources and the newest developments.

Feature detection by low dimensional projection with machine learning methods for dimensional reduction allows a broad sense of the content of a literature body derived from a combined analysis derived from `Title` and `Abstract` sections. `Title` section derived analysis produces less populated classes, allowing more specific concept allocations. In some cases, process and/or materials can appear as the case of `prefabrication` and `concrete`. `Abstract` section derived analysis produces an equivalent amount of clusters, enforcing the notion that the methods used allow extracting low dimensional features with low conceptual resolution. A low number of clusters (3-4) were systematically produced from `Title` section and `Abstract` section analysis, the last one most probably reflecting the structure of research content with problem definition, main results, and method in a compressed sentence structure.

B. DETAILED RESULTS

We selected the CrossRef platform to perform a selection of queries addressing the AEC-AI industry. CrossRef is a leading multidisciplinary platform offering a range of useful metadata for scientific and bibliometric research as well as policy creation and evaluation, including one of the biggest collections and access-point to open-access articles.

We have found that from `general query` there are several thousands of articles containing and/or related to the `Optimization`. Therefore, we have made an analysis of `Optimization` thematic content in the AEC Industry literature, covering the top (1000) more relevant articles from the last decade (beginning in 2010-01-01). However that the CrossRef platform allows an open multidisciplinary and fully-featured database, we have found a significant amount of records incomplete to the level of abstracts.

Specifically, we have found that for the AEC-AI industry - `Optimization` -, a low fraction of the records contained data in the `Abstract` section field, with over 70% missing. Then, we carry out text analysis using NLP methods in a sample of full-text articles, from the top 1K cited and relevant records.

1) INITIAL BIBLIOMETRIC ANALYSIS

We look at the set of documents retrieved in a general search for AEC-AI Industry and `Optimization` as KEY-TERM (see figure 2). When we tested for the content in the general space of AEC, AI, IoT, and MH-Structure, we observed that most of the documents contained, in the abstracts, AEC-related concepts with little content of AI-related concepts. Whereas those other ones with high content of AI-related concepts. We have extracted and analyzed `Title` section and `Abstract` sections to characterize the sample literature and its relation to AI and ML developments applied to the AEC-Industry. For the `Optimization`, we have found that it is possible to perform some qualitative and quantitative inferences on the contents' structure in the literature's body sample.

2) JOURNALS CONTENTS

These inferences, derived from quantitative concepts decomposition, on the `Title` and `Abstract` sections, show that most of the articles contain some relevance associated mainly with AEC-Industry. Which agrees with `Optimization` as relevant to these areas [2]. However, `Optimization` is broadly used in AEC-Industry intrinsically in their process as well as an AI/Machine Learning associated technologies. In particular, from the journal sources, retrieved in the query, and the subjects disclosed for the records, the two major areas - AEC-Industry and AI/ML - may form sets with variable degrees of overlapping. We noted, from manual screening many, either exclusively or independently on each specific area, i.e. some articles were purely on AEC-Industry with not content of AI/ML or vice versa.

C. GENERAL CHARACTERIZATION OF `OPTIMIZATION` IN THE AEC-AI INDUSTRY

We obtained, for the `general query`, `591090` records associated with AEC-AI Industry and `Optimization`; from which we retrieved the top 1K more relevant records by CrossrRef's score index (figures 3– 4). In [2], 42000 works were analyzed for the whole AEC-AI Industry, here the simple `general query` search yields more than 10 times more records, showing that there are probably some caveats ahead.

1) JOURNALS TABLE 1K SAMPLE

We carry out a basic bibliometric analysis on the sample of the top 1K retrieved records from the search query with the key-term, `Optimization`. In figure 3, the number of documents in the retrieved 1K Top sample and the times cited (TC) metrics are shown for the top 25 journals. We observed a broad range of journals with an equivalent broad range of focus/aims/editorial emphasis. Most of the journals shown are in the area of AI and/or engineering. In addition, we noticed that documents spanning across several journals in the field of AI and ML (figure 3). In more detail we have found: The number of records associated with the search query with `Optimization`, yield `591090` documents, many more than other related searches using a different platform [2]. We retrieved records for the top one thousand and analyzed their basic bibliometric description. As CrossRef does not implement methods to retrieve boolean connected records, therefore an initial assessment of the content retrieved is fundamental.

Disclosed subjects and containing journals show a broad range of retrieved documents, with many probably not in the interest, but fairly related. It seems that publisher/author combined efforts allow the generation of the record to contain some noise as not available values show a significant prevalence (see next sections). However, the predominant journals and the average times cited index for them suggest they are more likely to contain works related to the aimed search queries. An important piece of informational metadata is the relevance of the records in the search and how they recover another metric as it is the impact of the work. CrossRef provides its fuzzy logic relevance score, and the times cited metrics which allows to extract the main journals in the field and assess the impact of the retrieved works, which will be used as criteria to isolate/decrease the noise of related documents not in boolean searches.

Automation in Construction appears as the journal containing the highest impact (times cited) index. In agreement with other studies, the best-ranked journals appear consistently across searches, either as the most cited and highest impact [2]. We will implement other constrained queries to limit journals and dates. We implemented a comparative search approach to defining the best set of retrieved records. Therefore, we use the publishing date to constraint the time span from which is useful to limit the number of records retrieved and to analyze trends and temporal patterns in the development of the literature corpus [67], [77].

2) SUBJECTS & SCORES 1K SAMPLE

We evaluated the disclosed subject in the 1K retrieved search, for `Optimization` constrained to `general query`, figure 4a. Notably, a significant number of records did not contain subject data, as the metadata includes not available (NA) values. `Civil structural engineering, building and construction, and artificial intelligence`, were the most frequent subjects disclosed. We found no clear relationship between time cited and relevance score as provided by CrossRef (figure 4b), however, we noticed a slightly tend to get scores in "low, middle and high" modes. In summary, our results show:

a: BROAD SCOPE OF THE SUBJECT IS PRESENT IN THE 1K TOP RELEVANCE SCORE SAMPLE

In an initial assessment of quality and content retrieved in the CrossRef platform, we obtained a sign that the pool of records contained a wide set of contents. Corresponding to a broad set of related results with variable degrees of relevance and importance to the aimed objectives.

b: MANY RECORDS WITHOUT DISCLOSING IN THE AVAILABLE SUBJECTS

An initial sign of the type of records provided by the disclosed subjects, showing that most of the data are disclosed as not available value (NA), which in some sense can be interpreted as a too far connection to the search aim as well as defined subjects were also disclosed. We cannot attribute the lack of subjects either to the available options allowed by the publisher or to a specific disclosure by the authors, especially because more than a single option seems currently in use.

c: THE MODE OF RECORDS WAS DISCLOSED IN THE SUBJECT OF ENGINEERING, CONSTRUCTION, AND AI

Discarded the observation with NA data. We obtained that most of the records disclosed relevant and connected subjects to the search query. However, this result does not show the number of records specifically addressing the subjects in a boolean way or logic. The results also contain disclosed subjects far from what in the first approach a connection to the query can be expected. Although, the problem complexity and generality of the `Optimization` defining the query can allow a broad range of subjects related to the records.

d: NO RELATIONSHIP WAS CLEAR BETWEEN RELEVANCE SCORE AND TIME CITED INDEX FROM CROSSREF METADATA

Relevance score (reference matching) can be a little tricky for the structured/unstructured records deposited in the CrossRef platform. Although they have high precision and accuracy, noise in the records may affect the search results. The relationship between times cited values and the relevance score values did not display a direct correlation, instead, they display a broad scattering pattern, with as expected high density in the lower range of values.

e: BROAD RANGE OF RELEVANCE AND TIME CITED INDEX WERE OBSERVED IN THE 1K SAMPLE OF DOCUMENTS

The time cited, the number of times a referenced is issued to a particular record, display a skewed distribution towards low values, the typical behavior in many research fields, mostly of articles are cited a few times or nothing at all [78]. With the records obtained for the `general query` search and the combined query for `Optimization`, we observed the overall structure expected for the impact of articles with an exponential decline in the number of citations received by the articles. Interestingly, in the case of the relevance score computed for the query, the skewed distribution is observed towards a higher frequency of low score values, and an additional clustering pattern for low, middle, and high citation times values.

f: SOME DOCUMENTS WERE DISCLOSED IN FAR-AWAY CONNECTION WITH THE EXPECTED SUBJECT

Intriguingly, some subjects disclosed were not in the immediate expected connection with a search query, for example, is not evident as first the relationship between `Language and Linguistics` to the AEC Industry, however, it can be very much related to AI.

Next, from this sample of 1K records, we evaluated the amount of abstract containing in the metadata. We found that the fraction of articles which contained abstract in the metadata was `0.276`.

3) SUBJECTS & IMPACTS FOR FULL-TEXT ARTICLES

As we get an important fraction of works without abstract section (fraction with abstract = `0.276`, we evaluated a

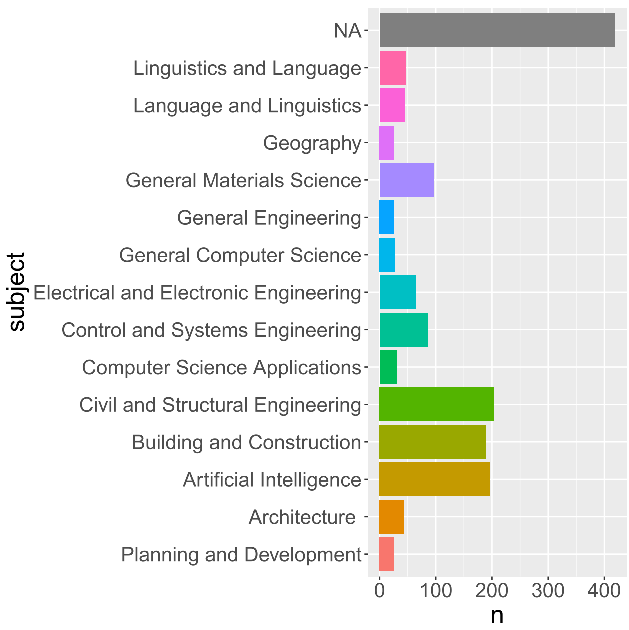
Journal	Num	Mean TC	sd
Engineering Applications of Artificial Intelligence	109	20.93	33.2
Construction and Building Materials	98	10.01	11.4
Artificial Intelligence	45	7.07	14.8
Automation in Construction	39	45.13	39.6
Radiology: Artificial Intelligence	26	2.81	3.9
AI & SOCIETY	20	1.25	2.9
Sustainability	20	4.50	6.0
ACADEMIC JOURNAL Series: Industrial Machine Building, Civil Engineering	18	0.00	0.0
Journal of Asian Architecture and Building Engineering	16	0.44	1.0
Advances in Civil Engineering	15	3.93	4.1
Applied Artificial Intelligence	15	9.27	29.6
Artificial Intelligence Review	15	7.67	13.0
Building Research & Information	15	9.67	13.6
Engineering, Construction and Architectural Management	15	4.13	5.3
Nature Machine Intelligence	14	7.14	10.7
Construction Economics and Building	13	5.69	8.2
Journal of Experimental & Theoretical Artificial Intelligence	12	1.83	3.5
IEEE Access	11	10.73	28.5
Information	11	4.18	2.8
Journal of Chemical Information and Modeling	11	18.00	24.4
Journal of Construction Engineering and Management	11	5.64	8.3
AI	10	3.00	5.8
Advances in Artificial Intelligence	9	3.78	5.9
Artificial Intelligence in Medicine	9	6.89	7.0
IAES International Journal of Artificial Intelligence (IJ−AI)	9	0.00	0.0

FIGURE 3. Main journals present in the 1K sample retrieved query for "AEC-AI Industry" and `Optimization` for the `general query`. The number of articles by journal (Num) and mean citation times (Mean TC) and standard deviations (sd) are shown for top 25 more frequent.

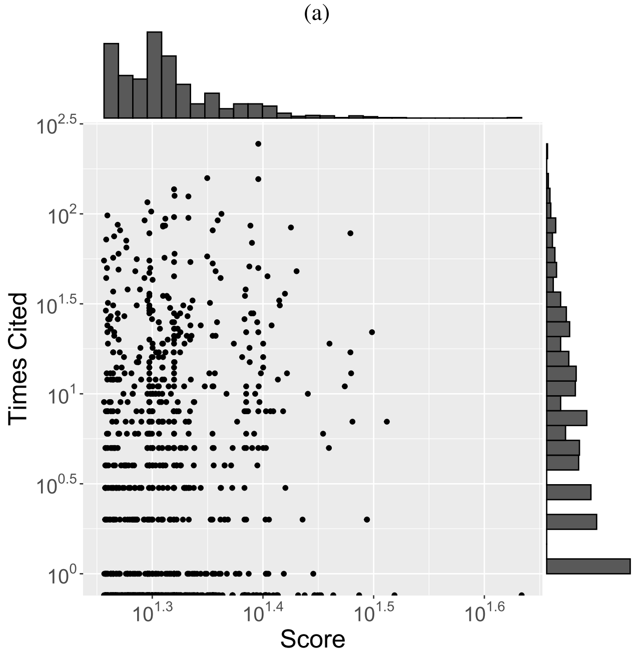
sample of full-text articles available from the top 1K most relevant. From the sample of full-text articles, we found a more adjusted set of disclosed subjects for the articles analyzed (5a). Most of the subjects were in the area of engineering, with a field in `civil and structural`, `building and construction` and `control and systems`. Similarly, artificial intelligence and general material science were the most represented besides engineering. In the same direction, we found that most of the articles were published in a more limited ad hoc set of journals (5b).

Specifically, we have found: `General query` derived search allows us to collect a more reduced set of full-text articles from the 1K top relevant works retrieved in the search. A coherent sample of subjects and journals was obtained for the available full-text articles. Similar to the content for the 1K records, we collected a sample of full-text articles closely related to the properties exhibited by the 1K sample. The sample of full articles seems well representing the retrieved 1K top relevant records. The main journals and their impact footprint are preserved in the sample of full-text articles, although this is not a core requirement it follows that some minor bias can be expected.

We got a representative sample of full-text articles to analyze with ML and textual-NLP methods. Given the features in the full-text sample of articles available will define further the results, it would be noticed that, as the literature corpus is changed, the conclusion as method extends to particular cases of exploration. As such, we will use the sample of



(a)

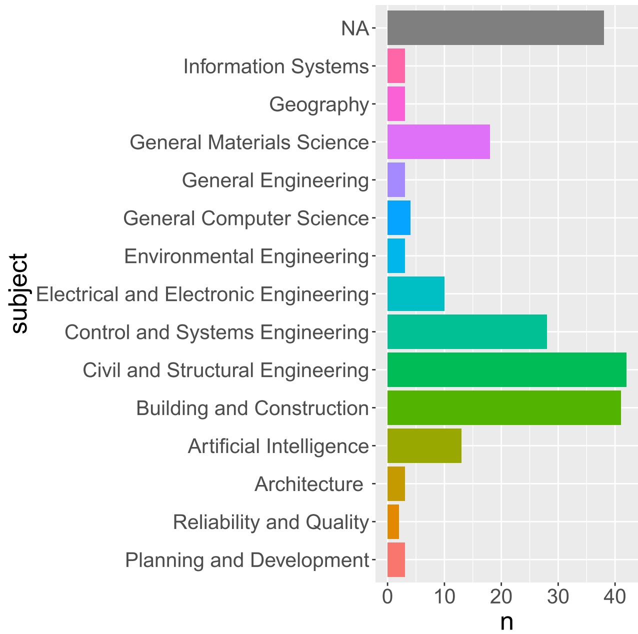


(b)

FIGURE 4. **Disclosed subjects, impact and relevance metrics for "AEC-AI Industry" and** `Optimization` **query. (a) Subjects disclosed for a pool of 1K more relevant articles from CrossRef. (b) Scatter and distribution histogram of relevance score index and times cited index respectively.**

available articles to develop a combined strategy using extensible techniques of machine learning and textual NLP. From the available articles collected, we separate them into Review and Articles types and performed the analysis in each set.

From the 1K reduced sample retrieved for the query `general query` and AEC-AI Industry associated with the `Optimization`, we got `104` available full-text articles. These articles were all format pdf and will be the basis



(a)

Journal	Num	Mean CT	sd
Construction and Building Materials	26	19	13
Automation in Construction	25	51	40
Engineering Applications of Artificial Intelligence	14	42	63
Advanced Engineering Informatics	5	24	25
Computers in Industry	4	27	19
International Journal of Naval Architecture and Ocean Engineering	3	1	0

(b)

FIGURE 5. **Subjects and journal metrics from a sample of full-text articles sample retrieved for "AEC-AI Industry" and** `Optimization` **for the** `general query`**. Number of documents (works) with the reported topic/theme (a) and the number (Num) and mean times cited (mean/sd) of documents in the retrieved full-text sample set, grouped by journal, (b).**

for content extraction and conceptual definition forming the literature corpus analyzed for the associated query.

We have collected about 100 full-text articles and `43` Reviews type articles that were in the Top 1K retrieved from CrossRef. Therefore, from the 1K more relevant records associated with the search, we obtained close to 15% of available full-text documents. It should be noticed that these samples are suited to test and improve the methods and they can be scaled to handle more data once new full-text articles become available.

4) `TITLE` SECTION DERIVED ANALYSIS

Next, we assessed the similarity of the content of the retrieved record by hierarchical clustering analysis of the `Title`'s section containing words for the sample of the full article (figure 6). In the figure the dendrogram of words cluster/groups are shown, in this case over 5 major similarity groups can be detected with a rich repertory of terms related to the AEC industry. We further separated manually the Review type articles in the sample. We used this sample of reviews articles as reference content as it is well appreciated that Reviews aim to represent a curated literature compendium with special focus.

In figure 6-b, an equivalent hierarchical cluster analysis is performed in the set of reviews separated from the retrieved sample. We observed a more trimmed dendrogram tree for the sample of reviews, with a clear group-set of well-associated words for the main topical reviews themes. The `engineering` and `learning` group segregates, probably reflecting the general context in the focus of the reviews' title. The next major group reflects mainly the aim of the articles in the form of `construction, industry, practices` related to `bim` which is the basis of a standard framework for the industry. Whereas that the remaining cluster reveals methodological and aims content as `systematic, literature/implementation, review/trend`.

Distribution histogram of citation times by article is displayed in the figure 6-c (upper panel histogram). The average TC index for the articles grouped by journal and number of articles can be found in the table in figure (6-c, lower panel table). The sampled full-text articles were published between 2014 and 2021, showing a correlation between time lapsed from publication and the TC index.

From these results, we summarize: a) Hierarchical clustering allows describing the conceptual content and relative differences between the underlying literature corpus. b) We used the `Title` section to perform aggregation clustering from the set of full-text documents. c) `Title` section derived analysis shows a good likelihood of underlying content in relation to the search query. d) Given that we obtained a Review set of articles, they were used to comparing the textual content analysis as a curated source of data. e) Reviews derived analysis of `Title` section contents, allows to compare to curated reference source the overall content structure in the related articles.

To get more insights into the content of the set of retrieved references, we performed a textual analysis on the titles. Figure 6-d and 6-e shown contents theme analysis for articles and reviews, respectively.

a: ARTICLES

In thematic analysis, the four quadrants in the plane - Density vs Centrality - are usually labeled as emerging/declining-niche and basic-motor themes, according to the combined degree of development/density and relevance/centrality projection.

The topological themes map for the articles group-set shows more defined themes with predominance of `learning, modeling` and `construction` as the most frequent terms; `optimization, project` are followed by `properties` and `cost` to end in `emergency`. In the examination of the co-occurring terms, we find:

BASIC: The dominant basic theme in the quadrant, with a lower degree of development, is `learning` associated with: `machine`, `approach`, `concrete`, `aggregate`, `prediction`, `strength`, `adaptive`, `asphalt`, and `deep`. More developed appears `modeling`, associated to `bim`, `building`, `design`, `based`, `system`, `data`, `automated`, `systems`, and `control`.

MOTOR: `optimization` displays a middle degree of developing with higher centrality, and associated with co-occurring with `algorithm`, `multi-objective`, `big`, `model`, `production`, `recognition`, `structures`, `classes`, and `discrete`. Slightly more developed but more peripherally is `construction` associated with `industry`, `study`, `case`, `performance`, `analysis`, `materials`, `process`, `knowledge`, and `potential`.

NICHE: In this quadrant, `properties` is associated with `mortar`, `durability`, `engineering`, `geopolymers`, `influence`, `powder`. Whereas in the same quadrant we found, `cost` associated with `applications`, `estimation`, `techniques`, `approaches`, `impact`, and `parameters`. The case of `project` display more centrality and associates with `buildings`, `evolutionary`, `hybrid`, `management`, `algorithms`, `energy`, `artificial`, `composite`, and `developing` also in the niche quadrant.

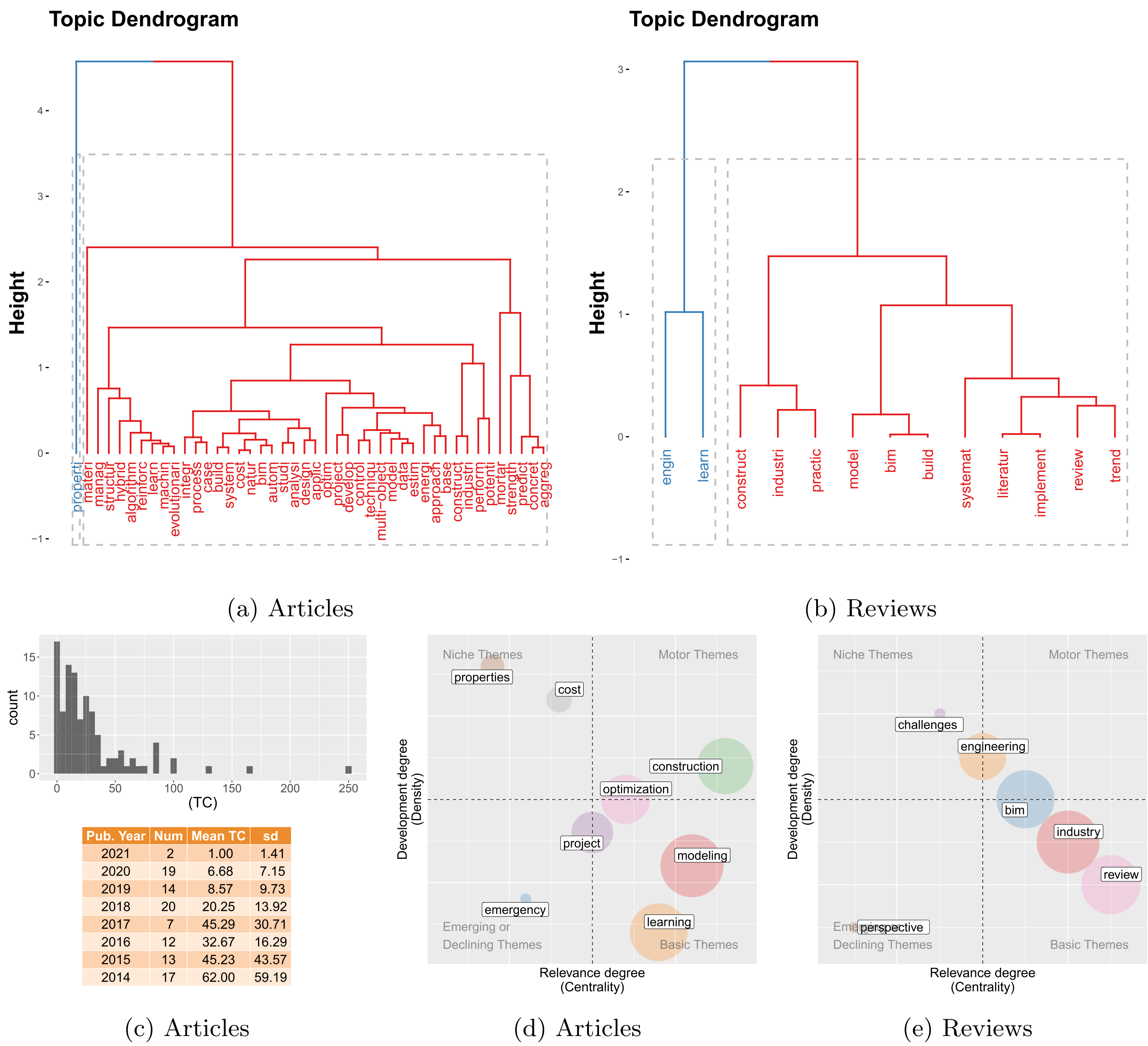
EMERGING: In turn, `emergency`, associated with `indoor`, `localization`, `response`, `support`, is the emerging theme in the set.

Together these results showed that the sample of literature corpus contains an enriched body of basic `construction, building, and optimization` associated concepts, covering main approaches using machine learning and artificial intelligence, including some specialized topics in development. We further note that material science as well as management as likely areas where AI is finding applications. Moreover, the literature sample corresponds to highly cited works published in leading journals in the fields of engineering, construction, material science. Likewise, it is possible to perform clustering of the works for fewer groups.

b: REVIEWS

For the case of reviews type articles in analysis, the topological themes map, shows more defined themes with a predominance of `bim`, `engineering` and `industry` as subject themes in the map, with `review` as the more basic theme concept. `Perspective` and `challenges` appear as emergent and niche themes, respectively. In more detail we have:

BASIC: The basic dominant theme in growth is `industry` associated/co-occurring with: `construction`, `analysis`, `architecture`, `history`, `practices`, `data`, and `sustainable`. In the case of `review`, is associated with `literature`, `systematic`, `implementation`, `aec`, `adoption`, and `green`.



Pub. Year	Num	Mean TC	sd
2021	2	1.00	1.41
2020	19	6.68	7.15
2019	14	8.57	9.73
2018	20	20.25	13.92
2017	7	45.29	30.71
2016	12	32.67	16.29
2015	13	45.23	43.57
2014	17	62.00	59.19

FIGURE 6. `Title` **section derived textual terms analysis for the full-text sample retrieved for "AEC-AI Industry" and** `Optimization` **for the** `general query`**. Dendrogram of words in the title for a sample of full articles derived from the most relevant retrieved documents: (a) Articles (b) Reviews. (c) Distributions histograms of times cited TC (upper panel) and articles by year with associated TC stats (lower panel). Documents terms mapping derived from titles words, for Articles (d) and Reviews (e).**

MOTOR: A *basic motor* theme is `bim` associated to co-occurring with, `building`, `modeling`, `modelling`, `trends`, `infrastructure`, `integration`, and `safety`. For `engineering` we have, `applications`, `learning`, `intelligence`, `machine`, and `materials`. Whereas that `challenges` and `perspective` are present as single occurrences in the niche and emerging quadrants, respectively.

This analysis shows that the literature's reviews sample contains an enriched body of basic construction terms. As it is recognized in the area, `bim` is the central term and platform where a variety of aims and developments in the AEC-AI industry occur as a central topic integrating efforts. Similarly, engineering aspects are developed along with machine learning/intelligence. The reviews, also highly the methodological aspects and some of the findings and areas of applications.

Next we performed a correspondence analysis to detect structural components in the `Title` section of the articles retrieved (figure 7a-d).

In summary, correspondence analysis shows: The conceptual structure map, by the correspondence analysis method, shows one axis describing the material's properties, for example: `material`, `strength`, `mortar`, `concrete`, `properties`. This shows that there are been some interest and developments in using `Optimization` methods, probably as AI/ML application to a structural components in the form of design as well as improving some performance in material properties (figure 7-a). This interpretation is further developed in the analysis of the `Title` section content by multi-correspondence analysis (MCA), where four structural clusters are clearly evidenced by: a cluster associated to `construction, industry, study` which reflect the commonality of the subject in the documents. Covariant with this cluster, we observe one defined by `optimization, concrete, approach`, which can be interpreted as the aimed contents. Whereas in the opposed quadrant, `bim, build, model, design, system`, reflects design applications associated with the BIM framework. Then, `machine-learning` appear as a fourth structural component, showing the methods as well as their application with the other clusters (figure 7).

Similarly, the non-linear multi-dimensional scaling (MDS) method recovers equivalent clusters, however it merges the `machine, learning` along with `approach-concrete-optimization` (figure 7-c). Using CA, in the textual content of the Reviews' `Title` section, yields two clusters, one defined by `engineering, learning`. Whereas another spanning the axis `bim-build-model` extends towards `construction-industry-practice` leaving in between `review, literature, systematic, implement, trend`. Which mostly reflects the aim of providing a review of literature on the themes (figure 7-d).

5) ABSTRACT SECTION DERIVED ANALYSIS

Next, we further examine textual content on the full-text available sample by analyzing the `Abstract` section derived content from the articles and reviews (figure 8-a-e). Figure 8-b, display the basic bibliometric analysis from the reviews group.

Hierarchical clustering analysis, for `Abstract` section derived, becomes more crowded on its visualization. To decrease the degree of words overlapping (whiteout cutting sections) we display the dendrogram in phylogenic form. Figure 8-a, shows the dendrogram derived from `Abstract` section for the Article type documents. Reviews' equivalent analysis derived from `Abstract` section is presented in figure 8-b. Hierarchical clustering allows describing the conceptual content and relative differences between the underlying literature corpus. `Abstract` section derived aggregation clustering from the set of full-text documents display more heterogeneity of clustered terms, showing a probably higher variety of topics than the carried by reviews type. `Abstract` section derived analysis indicates a good likelihood of underlying content in relation to the search query. Given that we obtained a Reviews set of articles, they were used to comparing the textual content analysis as a curated source of data.

Reviews derived analysis of `Abstract` section contents, reveal closer relation between `machine learning` and `optimization algorithms` as they branch close to the root with ML separating a single related cluster of `material` and `epoxy` probably in material composition and properties. Reviews' derived hierarchical clustering displays considerably fewer concepts and clusters as compared to the Articles set. Showing significant differences and variations of concepts underlying respective groups and scope of the contents. The times cited metric for the Reviews documents are shown in figure 8; we observe a significant impact of this type of document in the sample. Next we assessed the corresponding theme mapping for each type of article: Articles (8 and Reviews (8).

a: ARTICLES

The topological themes map shows more defined themes with a predominance of `model` and `performance`, with the former as central domain theme whereas that the latter falling in motor/active domain themes. The emergent `study` concept appears as a single representation, most probably reflecting the economy of words and constraints in the `Abstract` section of the article's structure.

- The dominant theme at the center is `model` associated/co-occurring with: `design`, `building`, `model, construction, system, process`, `approach`, `management`, and `project`.
- In the case of `performance`, is associated to `proposed, algorithm, learning, method, energy, machine, multi, methods`, and `support`.

b: REVIEWS

- The next dominant theme in the motor quadrant is `industry` associated /co-occurring with: `technologies`, `applications`, `infrastructure`, `modeling`, `future`, `algorithms`, `application`, `areas`, and `challenges`.
- In the case of `construction`, is associated to `bim, aec, engineering, implementation, building, safety. learning, machine`, and `materials`.
- A *basic motor* theme is `study` associated with, `management, sustainable, knowledge`, `model`, `models`, `systems`, `practices`, `process`, and `three`.
- For emergent themes, `analysis` we found its associates to: `potential, analysis, epoxy, green`, `identified. performance`, `problems`, `optimization`, and `studies`.
- In the case of `review`, is associated to `literature`, `design`, `system`, `systematic`, `identify`, `web. quality`, `directions`, and `lean`.

Reviews content appears more defined for the analyses derived from `Abstract` section, whereas for the type of the article a higher variety of content is observed at the time of trying hierarchical clustering. However, for themes mapping derived from `Abstract` section of Articles display less variety, indicating a common mapping, probably reflecting in the structural approach where `model` is central, some

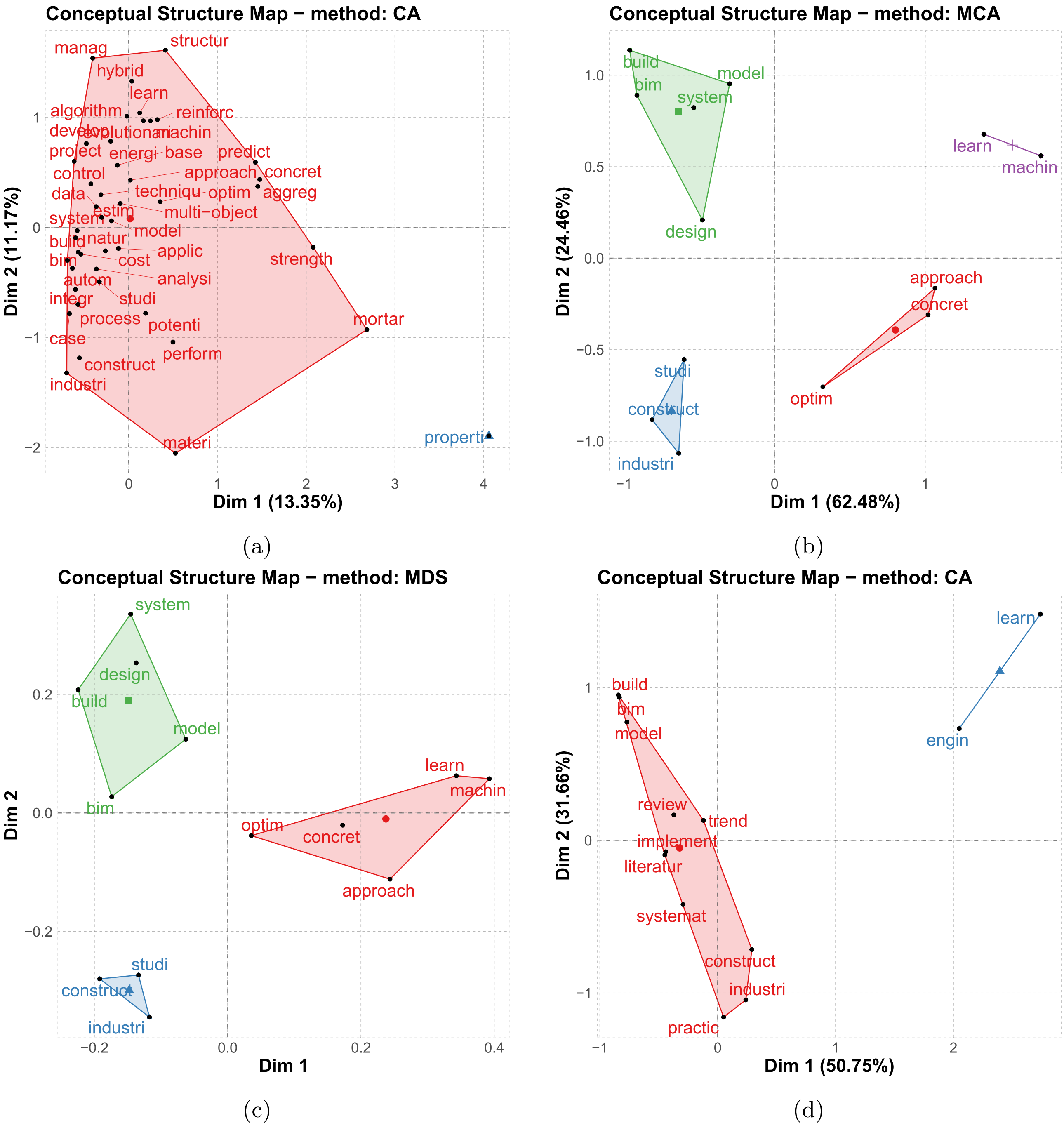
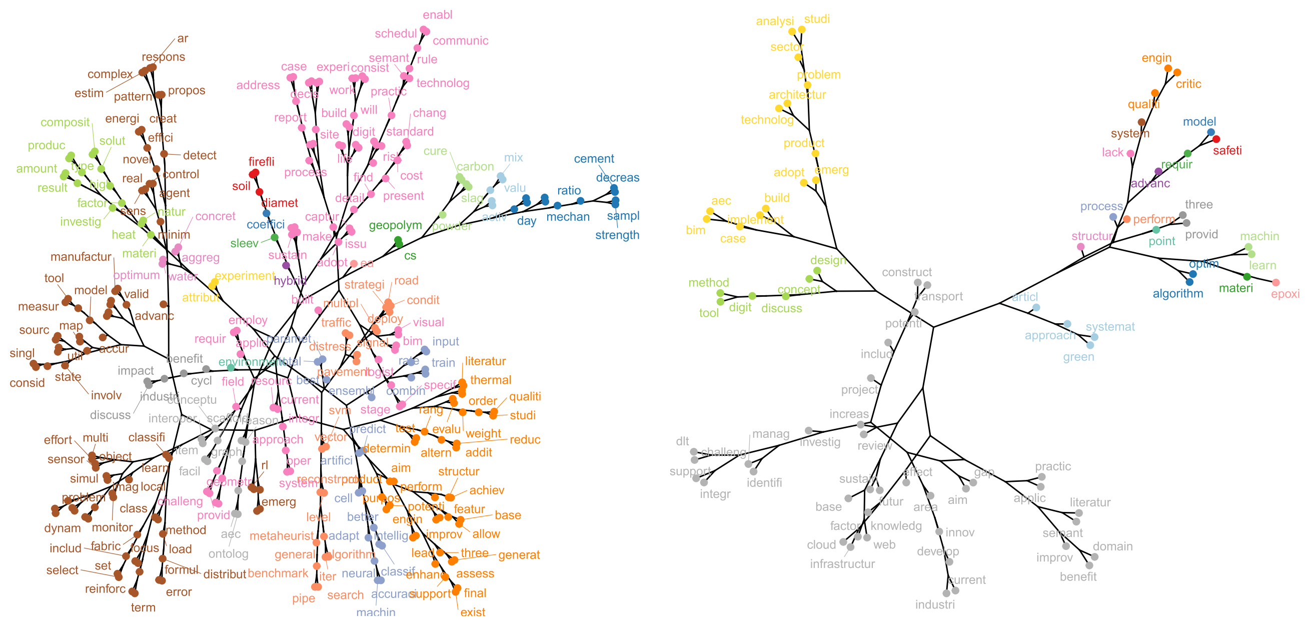


FIGURE 7. Dimensionality reduction analysis for `Title` section derived content from full-text sample documents for the "AEC-AI Industry" and `Optimization` for the `general query`. Clustering analysis for abstract contents from a sample of full-text documents. (a) Low dimensional projection of concepts derived from the `Title` section by the CA method. (b) Clustering projection for documents `Title` section derived content with MCA method. (c) Nonlinear MDS method for `Title` section derived clustering. (d) `Title` section content analysis, by the CA method, for the sub-sample of reviews documents.

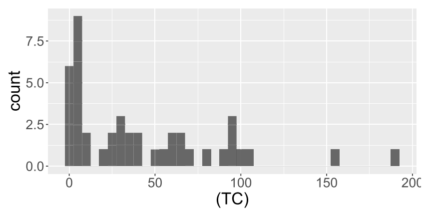
other intrinsic commonality of content is the aiming towards `performance`.

Next, we performed further analysis on the textual content in `Abstract` section for the constrained search `general query`. Figure 9a – e, display correspondence analysis for `Abstract` section derived content conceptual structure.

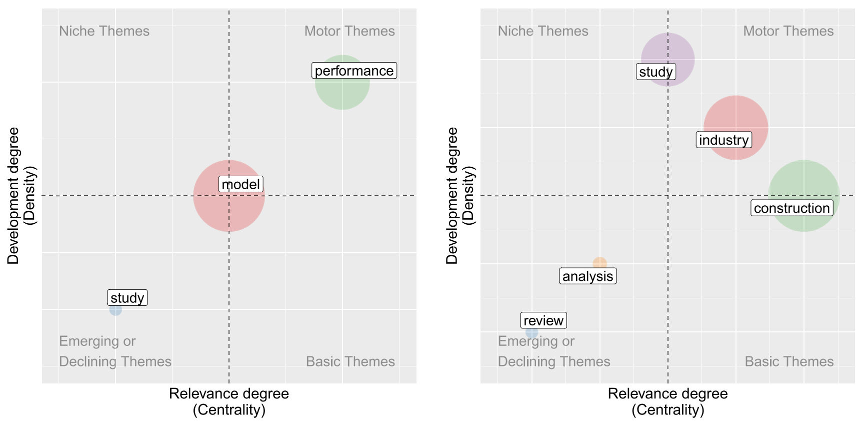
The conceptual structure map using the CA method shows a broad spanning axis from `optimization`, `ontology`, `knowledge` up to `geopolym`, `slag`, `carbon`, `mortar` forming a construction material method axis (9). The orthogonal direction with less variance appears to display methods in ML as `optimization`, `SVM`,



(a) Articles (b) Reviews



Pub. Year	Num	Mean CT	sd
2021	2	3.50	2.12
2020	11	7.45	12.13
2019	6	42.67	29.98
2018	5	63.20	27.12
2017	3	95.00	4.58
2016	1	104.00	NA
2015	2	68.00	5.66



(c) Reviews (d) Articles (e) Reviews

FIGURE 8. `Abstract` **section derived analysis of textual terms for the full-text sample retrieved for "AEC-AI Industry" and** `Optimization` **for the** `general query`**. The constrained search query related full-text available documents were analyzed and** `Abstract` **section extracted. (a)** `Abstract` **section derived hierarchical clustering dendrogram is displayed in circular form to decrease the overlapping of labels. The data corresponds to articles type document with available full-text. (b) Dendrogram of words in the** `Abstract` **section for the available full-text documents (c) Histograms of times cited index (upper panel) for Reviews type document found in the retrieved sample. (d)** `Abstract` **section derived theme mapping. (e) Documents terms mapping derived from titles words.**

`firefly`, `hybrid`. The MCA method reveals two main clusters: a small peripheral cluster, again, emphasizing material construction elements and properties, and a core conceptual cluster with a variety of concepts in the AEC-AI industry.

Figure 9, shows `Abstract` section derived content analysis using MDS method, the results from this analysis yield three main clusters. A main central cluster containing many target objectives was observed; the peripheral superior cluster contains `optimization` closely related concepts, whereas the inferior peripheral cluster reflects more AI-related concepts. Overall, this clustering schema agrees with a setting of articles' objectives/aims/problems and probably two of the main approaches expected: `machine learning` associated and `optimization` related. Consistent with this interpretation, equivalent content analysis derived from `Abstract` section from Reviews type of the articles' literature shows a main variation axis covering the `machine learning` to `optimization algorithms`, and `material-structure` component of less variance (CA method, figure 9).

In turn, two main clusters are revealed by the MCA and MDS methods, one revealing "use" enclosed in `architecture`, `engineering`, `advantage`, `bim`, `build`, `development` with `current`, `tool`, technology. The accompanying cluster reveals, the study, analysis, benefits associated with the AEC-Industry enclosed by concept, implementation, improve, increase, performance, practice, model, potentiate, for the case of the MCA method. The MDS method yields a minor cluster associated with "projection/discussions" with a major cluster of heterogeneous content (figure 9).

The results analysis, show that optimization and machine learning related works are being used in the design of material properties as well as in the process of integrating "bim technologies." In agreement with these, the analysis of the reviews associated with the retrieved search also shown emphasis on these aspects as well as their advantages and developments.

D. SPECIALIZED CHARACTERIZATION OF OPTIMIZATION IN THE AEC-AI INDUSTRY

As it was shown previously, for implementing our adaptive macroscopic method, we designed the use of a combined constrained search query to get a comparative approach to content assessment. Therefore, we characterized, for the constrained search specialized query, the literature pool retrieved from CrossRef as is shown in figure 10–14. We obtained 13739 records associated with AEC-AI Industry and Optimization; from which we retrieved the top 1K more relevant records by CrossrRef relevance index (figures 10– 11b).

1) JOURNALS TABLE 1K SAMPLE

The basic bibliometric analysis on the sample of the top 1K retrieved records from the search query with the key-term, Optimization for the selected publishing journals associations is shown in figure 10.

We observe the following features in the records retrieved by the search:

a: LOW PROPORTION OF RECORDS WITH ABSTRACT FIELD DATA

First, we have found that similar to the previous search, most retrieved articles did not contain the Abstract (AB) field. In the case of the constrained specialized query search, the fraction of articles with abstract was 0.276.

b: SPECIALIZED JOURNALS WITH HIGH IMPACT (ABOUT 70% OF THE SAMPLE)

Notably, the topmost represented journal associated with the selection of journals revealed that in the sample of records associated with the search query account in about 2/3 of the total combined composition of works in the 1K more relevant sample. Interestingly, some journals in the top most influential [2] were displaced in their contribution to the search with Optimization.

In this case, we obtained mostly Construction and Engineering subjects journals, as the most numerous in the retrieved search, with a predominance of the well-established journals in the AEC-Industry. As it was previously noticed, the score index includes a mixture of associations between records and the search query, in this case, we observed journals retrieved with a broad range of times cited impact, some of the journals did not appear with a clear connection with the AEC-AI Industry or they may be inferred distant.

2) SUBJECTS & SCORES 1K SAMPLE

Then, we analyzed the basic bibliometric characterization for the top 1K more relevant records in the retrieved search using the constrained query specialized query. Figure 11a, display the subject disclosed in the retrieved 1K records sample for Optimization, in the constrained specialized query search. Times cited (impact) and relevance metrics index (score) associated with the retrieved records are shown in figure 11b.

It can be observed that most retrieved documents are filed as in "civil and structural engineering" type, followed by "building and construction," in addition to "strategy and management," "industrial relations" and "control and system engineering." These were followed by a journal with more focus on machine intelligence and other engineering journals (figure 11a).

Metadata analysis of the relevance score and times cited to show the same pattern as previous search queries. Figure 11b, displays a scattering pattern visualized in the double-log (log-log) plot for time cited vs relevance score, both present low values skewed distributions, with the relevance score tend to cluster in low, middle, and high scores values. Interestingly, a small group of works appears as a cluster displaying the highest relevance, whereas that another cluster of 3 items displays indexes of highest impact and low relevance.

Constrained specialized query search associated with most influential publishers contains a high fraction of relevant journals and impact associated with Optimization. Although some unexpected residuals are noticed in relation to the use of the fuzzy relevance score implemented by the CrossRef platform.

As for the more general search results, bibliometric parameters are also preserved. The sample covers the standard skewed distribution in the field for impact and relevance provided by the CrossRef platform. Similar to the general query search pattern, relevance score trends consolidate in low, middle, and high values levels.

3) SUBJECTS & IMPACT FOR FULL-TEXT ARTICLES

As with the previous general search queries, and because of the low level of abstract fields present, we gather a sample of full-text articles available. From the sample of full-text articles, we found a more adjusted set of disclosed subjects for the articles analyzed (12a). In the same direction; we found that most of the articles were limited to a more ad hoc set of journals (12b).

From the specialized query constrained search, we obtained 119 full-text available documents. This fraction

FIGURE 9. `Abstract` **section derived clustering for "AEC-AI Industry" and** `Optimization` **for the** `general query`**. Clustering analysis of textual content from** `Abstract` **sections derived from the sample of full-text documents. (a) Low dimensional projection of concepts derived from the** `Abstract` **sections by the CA method. (b) Clustering projection for documents** `Abstract` **derived content with MCA method. (c) Nonlinear MDS method for** `Abstract` **derived clustering. The upper row corresponds to Articles and the lower row (d-f) corresponds to review documents.**

is similar to the proportion of full-text available for the more general search query, which consolidates around 10% of the original sample (1K).

The available sample of full-text articles displays similar metadata distribution and values as the containing distribution. Most of the records are filed under "civil and structural engineering," "building and construction," and "control and system engineering," followed by "Industrial relations" and "strategy and management."

We noticed that, although disclosed subjects follow a similar proportion as in the more general search query, in this case, the range of the disclosed subjects is narrow, i.e. more focused and therefore more probably related to the aimed objective.

The full-text sample contains similar proportions of high-impact journals as well as impact metrics. Dominant journal in the available full-text sample is the leader in the field, Automation in Construction, with almost 50% of documents. The higher impact by journal occurs in the most frequent and the fourth-ranked, with a broad range of impact of content works.

4) `TITLE` AND `ABSTRACT` SECTIONS' HIERARCHICAL ANALYSIS

Next, we assessed the similarity of the retrieved records by hierarchical clustering (and correspondence analysis) of the `Title` section~ containing words for the sample of the full-text article (figure 13). A similar analysis, hierarchical clustering of `Abstract` section derived text is found in figure 13-b.

Hierarchical clustering of `Title` section derived representations recovers a `machine learning` defined group close to the root, where also other two major groups are defined. These results show that the constrained search query can effectively isolate records in the expected group of works in the full-text available sample.

The distribution histogram of citation times by article is displayed in the figure 13-c (upper panel histogram). The average TC index for the articles grouped by journal and number of articles can be found in the table in figure (13-c, lower panel table). This result for the full-text sample articles, shown that they correspond to the overall metric distribution of the impact of research articles.

Journal	Num	Mean TC	sd
Automation in Construction	333	24.54	35.34
Journal of Construction Engineering and Management	293	2.93	5.06
Mathematics	127	1.72	2.88
Journal of Computing in Civil Engineering	61	4.77	7.66
Engineering Structures	54	10.80	14.11
IEEE Transactions on Computational Intelligence and AI in Games	28	4.14	5.02
Computational Intelligence	23	0.22	0.42
Journal of Legal Affairs and Dispute Resolution in Engineering and Construction	22	1.59	3.10
Journal of Management in Engineering	7	6.71	8.20
Journal of Hydrologic Engineering	5	3.60	4.04
Journal of Optical Communications and Networking	5	29.20	36.13
Practice Periodical on Structural Design and Construction	4	0.50	0.58
Journal of Architectural Engineering	3	1.67	2.08
Journal of Energy Engineering	3	2.33	2.08
Journal of Environmental Engineering	3	7.00	6.08
Journal of Structural Engineering	3	0.33	0.58
Processes	3	0.33	0.58
Computer−Aided Civil and Infrastructure Engineering	2	0.50	0.71
IEEE Embedded Systems Letters	2	2.00	0.00
IEEE Photonics Journal	2	0.50	0.71
Journal of Professional Issues in Engineering Education and Practice	2	5.50	3.54
Dermatologic Therapy	1	0.00	NA
Healthcare	1	0.00	NA
International Journal of Financial Studies	1	2.00	NA
Journal of Bridge Engineering	1	0.00	NA

FIGURE 10. Main journals present in the 1K sample retrieved query for "AEC-AI Industry" and `Optimization` for the `specialized query`. The number of articles by journal (Num) and mean citation times (Mean TC) and standard deviations (sd) are shown for top 25 more frequent.

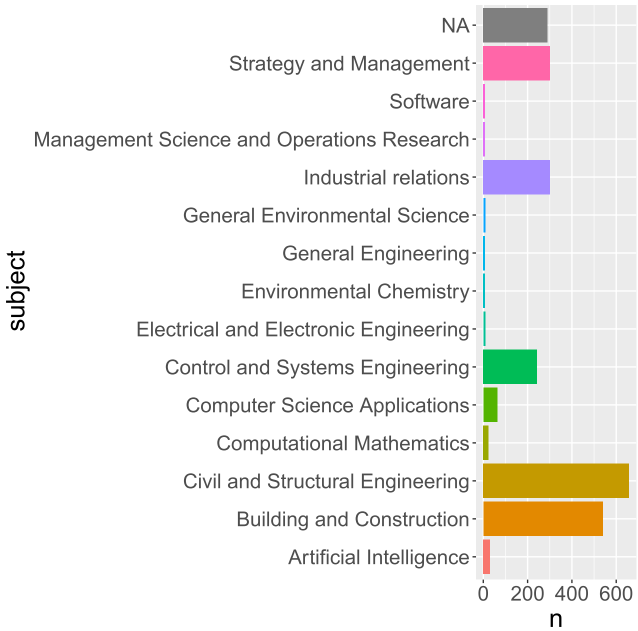
The sample contained articles published from 2014 to 2021, showing a clear correlation between time lapsed from publication and the TC index (figure 13). Although older articles trends to be less represented, probably reflecting some technological caveats.

To obtain more insights on the content of the pool of retrieved references, we performed textual analysis derived from `Title` section and `Abstract` section, figure 13-d and 13-e, respectively.

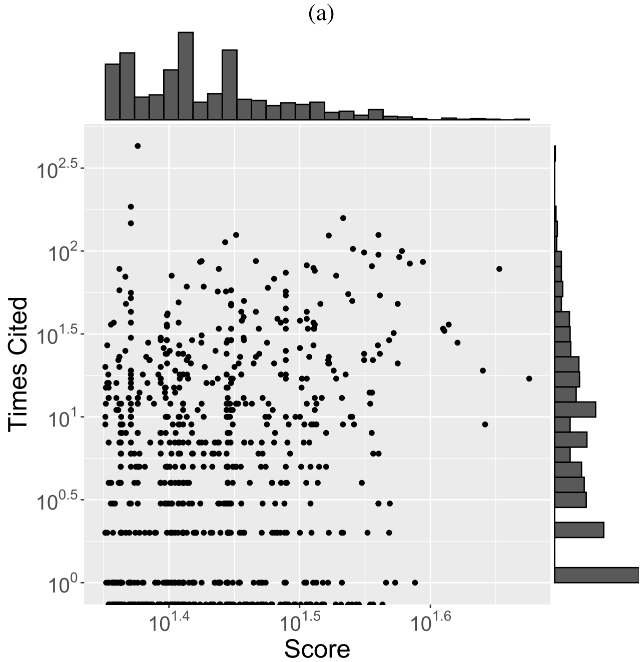
a: `TITLE` *SECTION DERIVED*

The topological themes map for `Title` section derived analysis, shows more defined themes with the predominance of `BIM` and `Construction`, with the former as in the basic domain themes whereas that the latter falling in motor/active domain themes.

BASIC: The next dominant theme in the growth is `framework` co-occurring with: `management`, `systems`, `projects`, `bim-based`, `planning`, `simulation`, `control`, `offshore`, and `developing`. In the case of `modeling`, is associated with `cost`, `knowledge`, `bridge`, `development`, `estimation`, `integration`, `networks`, `accident`, and `applications`. The case of `building` is associated with: `bim`, `design`, `analysis`, `energy`, `engineering`, `modelling`, `structural`, `support`, and `algorithm`. For `construction` we have, `safety`, `collaborative`, `industry`, `monitoring`, `models`, `techniques`, `case`, `integrating`, and `mep`.



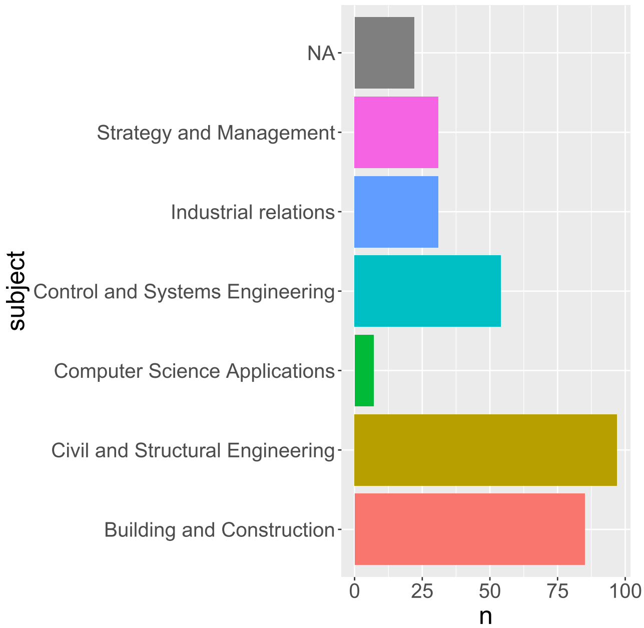
(a)



(b)

FIGURE 11. Disclosed subjects, impact, and relevance metrics for "AEC-AI Industry" and `Optimization` query for the constrained search: `specialized query`. (a) Subjects disclosed for a pool of 1K more relevant articles from CrossRef. (b) Scatter and distribution histogram of relevance score index and times cited index respectively.

NICHE: Whereas that `model` associated with `based`, `fuzzy`, `hybrid`, `Optimization`, `steel`, `artificial`, `concrete`, `prediction`, and `capacity` it turns into specialized development theme (niche). Similarly, `buildings` include: `learning`, `machine`, `algorithms`, `retrofit`, `seismic`, `detecting` and `risk`. Likewise, `system` associated with: `data`, `point`, `automated`, `cloud`, `visual`, `cad`, `clouds`, `dimensional`, and `documentation`. In turn, `learning`, `machine`, `algorithms` and `detecting`. For `model` we have: `based`, `hybrid`, `artificial`, `fuzzy`, `prediction`, `steel`. `capacity`, `concrete`, and `predicting`. The case of `developing` co-occurs with: `offshore`. `agent-based`. `evaluation`. `gas`. `oil`. `ontology`. `platforms`. Whereas `digital` associates with: `exchange`, and `managing`.



(a)

Journal	Num	Mean TC	sd
Automation in Construction	61	29.6	61.13
Journal of Construction Engineering and Management	26	3.5	4.40
Journal of Computing in Civil Engineering	7	7.4	7.96
Engineering Structures	5	15.4	13.90
Mathematics	2	1.5	0.71

(b)

FIGURE 12. Subjects and journal metrics from full-text articles sample retrieved for "AEC-AI Industry" and `Optimization` query for the constrained search `specialized query`. Number of documents (works) with the reported topic/theme (a) and the number (Num) and mean times cited (mean/sd) of documents in the retrieved full-text sample set, grouped by journal, (b).

b: `ABSTRACT` *SECTION DERIVED*

To develop a more in-depth assessment of content in the full-text sample, we analyzed the theme map revealed by the `Abstract` section derived text. The topological themes map shows more defined themes with the predominance of `building` and `construction`, with the former as in the motor domain themes whereas that the latter falling in emergent domain themes, aligned with them appears `proposed`.

MOTOR: The building of `building` is associated to co-occurring with, `building`, `models`, `modeling`, `framework`, `time`, `method`, `industry`, `developed`, and `algorithm`. For `proposed` we have, `proposed`, `cost`, `management`, `approach`, `process`, `case`, `work`, `level`, and `site`.

EMERGENT: Likewise, `construction` associated with `project`, `study`, `safety`, `projects`, `risk`, `knowledge`, `control`, `practices`, and `factors`. In turn, `performance systems`, `factors`, `buildings`, `built`, `machine`, `evaluation`, `potential`, `set`, and `algorithms`.

Interestingly, the `Abstract` section derived themes mapping becomes less explicit in extracting AI-related themes than for the case of `Title` section. From `Title` section derived analysis, clear references to AI and ML uses are allocated in the niche, specialized themes. However, no direct collection is represented in the analysis derived from `Abstract` section.

5) `TITLE` AND `ABSTRACT` SECTIONS' DIMENSIONAL REDUCTION ANALYSIS

Next, we characterized the available sample of full-text articles by dimensional reduction and clustering [67], [77]. Figure 14a–f, shows the low dimensional projection obtained by utilizing CA, MCA, and MDS machine learning methods for dimensionality reduction.

`specialized query` derived a constrained search for the available full-text documents, shows that it contain highly homogeneous textual contents as few clusters are recognized. Residual variance projections display a continuous scattering, making it difficult to extract concepts in clusters forms.

In the case of `specialized query` constrained search, the `Title` section derived analysis shows that up to 4 cluster were obtained by the non-linear MDS method (figure 14-c). In this case, the clusters segregate the concepts-words: `construction, approach`; `bim, analysis`; `management, framework, project`; and `build, system, design, model`. The CA method allows to characterize one cluster of `machine, learning`.

The `Abstract` section derived analysis revealed a similar trend as only two clusters were allocated by the MCA method. A small cluster containing `access, schema, definit, softwa, foundat, repres, import, class, ifc`. The central core cluster seems to accommodate a broad scattering of terms without enough differentiation to achieve additional clustering. The CA method yields two clusters, with a very small one connecting `sleeve, splice`, and a major one with extended projections in the direction of the small one. This projection covers `stress, tensile, confined, bar` towards `length, strength, load, concrete` which may represent design or material compositions. The next axial projection `artificial, linear, air, neural, asphalt, content, network`, may show another material elaboration more related to the properties than to design. MDS method displays similar clustering and low dimensional projection results as the CA method, two orthogonal projections coexist distributing most of the feature variance in the lower dimensional space.

E. RECENT SPECIALIZED CHARACTERIZATION OF `OPTIMIZATION` IN THE AEC-AI INDUSTRY

Following with the design of an adaptive macroscopic approach, we next characterized, for the `specialized recent query` constrained search, the literature pool retrieved from CrossRef as is shown in figure 15– 19.

1) JOURNALS TABLE 1K SAMPLE

We obtained `9822` records associated with AEC-AI Industry and `Optimization` constrained to `specialized recent query`. Then, we retrieved the top 1K more relevant records by CrossrRef's score index (figures 15–16b. We carry out a basic bibliometric analysis summarized in the figure 15 and 16. We observed that for the most influential journals considered in the AEC-AI industry field [2], the last few years (<3 yrs) trend is slightly different for the case of the `Optimization`. We obtained the Journal of Construction Engineering and Management, Mathematics and Automation in Construction as the most prevalent journals. Although that Automation in Construction still is the most influential with the highest impact, another journal appears with more published works, relevant to `Optimization` associated with AEC-AI Industry. From a sample of the top one thousand most relevant documents we found, the top 3 journals account for more than 60% of the records in the sample.

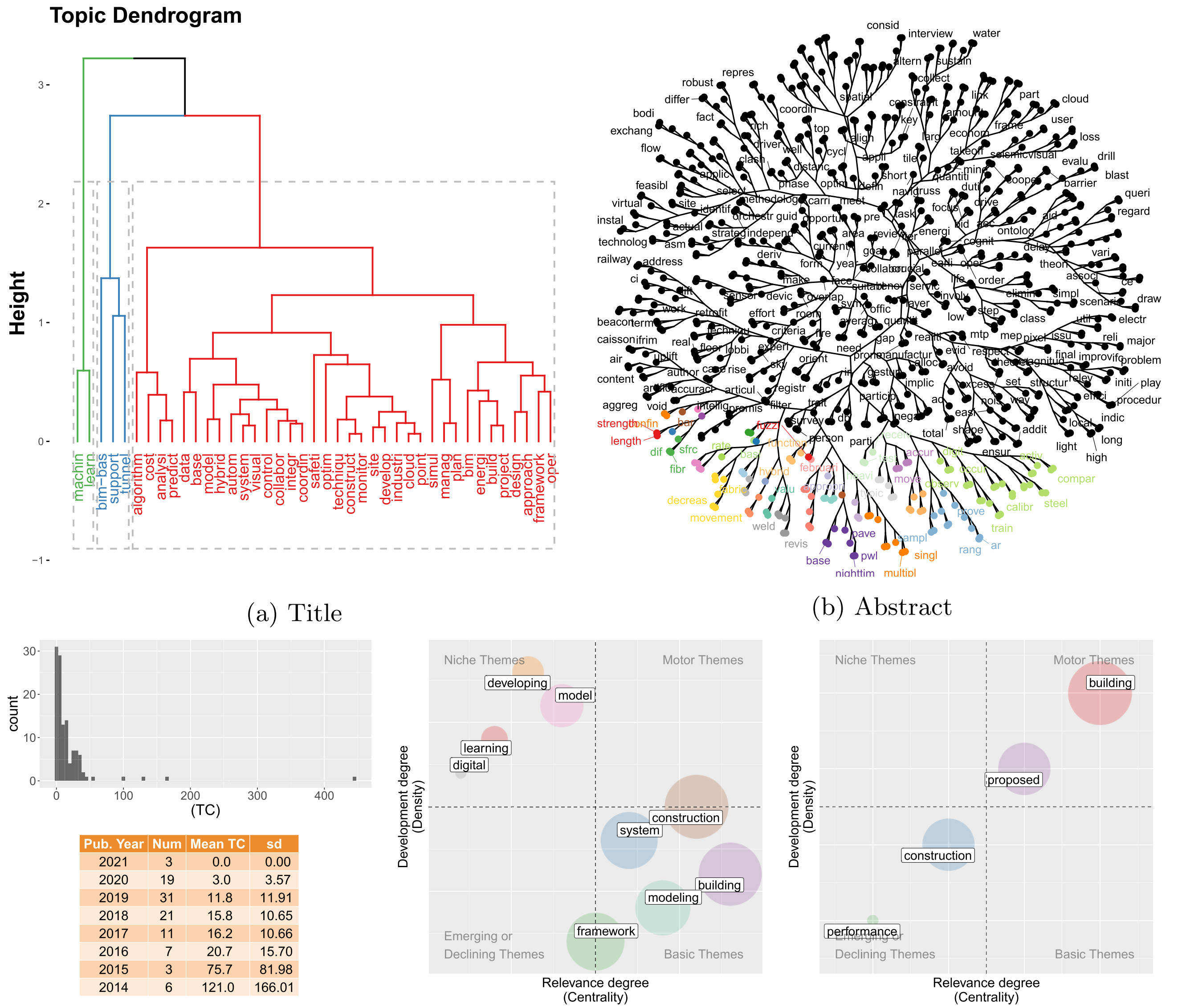
2) SUBJECTS & SCORES 1K SAMPLE

Similar to previous searches, "Civil and Structural Engineering," "Building and Construction" are the most represented subjects. A small increase in prevalence is observed for the subject of "Computer Science Applications." Whereas the distribution of score and times cited index are broadly distributed with a skewed distribution (figure 16b, equivalent to previous search results and literature on the topic [78].

A similar composition of disclosed subjects was observed for the `specialized recent query` constrained search as for other searches presented. As all the previous searches, in this case, we have found also that the majority of retrieved articles did not contain `Abstract (AB)` field, the fraction of articles with abstract was `0.205`. Therefore, we turn to analyze a sample of top-scored and top-times-cited full articles available.

3) SUBJECTS & IMPACTS FOR FULL-TEXT ARTICLES

We have found a more adjusted set of disclosed subjects for the full-text articles analyzed (figure 17a). In the same direction, we found that most of the articles was limited to a more ad-hoc set of journals (figure 17b), derived from `76` available full-text articles.



(a) Title (b) Abstract

Pub. Year	Num	Mean TC	sd
2021	3	0.0	0.00
2020	19	3.0	3.57
2019	31	11.8	11.91
2018	21	15.8	10.65
2017	11	16.2	10.66
2016	7	20.7	15.70
2015	3	75.7	81.98
2014	6	121.0	166.01

(c) Articles Impact (d) Title (e) Abstract

FIGURE 13. Hierarchical clustering and theme mapping derived from textual terms analysis for the full-text sample retrieved for "AEC-AI Industry" and `Optimization` **query using the constrained search with** `specialized query`**. Dendrogram of words in a sample of full articles derived from the most relevant retrieved documents: (a)** `Title` **section (b)** `Abstract` **section. (c) Distributions histograms of times cited TC (upper panel) and articles by year with associated TC stats (lower panel). Documents terms mapping derived from titles words, for** `Title` **section (d) and** `Abstract` **section (e) sections.**

From a sample of the available-full text documents (76 we found: Most of the available full-text articles in the sample are filed in "civil engineering" and "construction management." Automation in construction is the leading journal with highly times cited index by article.

4) TITLE AND ABSTRACT SECTIONS' HIERARCHICAL ANALYSIS

Next, we assessed the similarity of the retrieved records by hierarchical clustering of the `Title` section~ containing words for the full articles sample (figure 18-a). In the figure, a dendrogram of hierarchical clustering of `Title` section derived words groups are shown. A similar analysis is displayed in figure 18-b for `Abstract` section's derived words clustering. To obtain more insights on the content of the pool of retrieved references, we performed a textual analysis on the `Title` and `Abstract` sections, figure 18-d and 18-e, respectively.

First, we evaluated the basic bibliometrics impact to further characterizes the sample of full-text articles. The distribution histogram of citation times by article is displayed in the figure 18-c (upper panel histogram). The average TC index for the articles grouped by journal and number of articles can be found in the table in figure (18-c, lower panel table). The sample contained articles published from 2019 to 2021, showing a clear

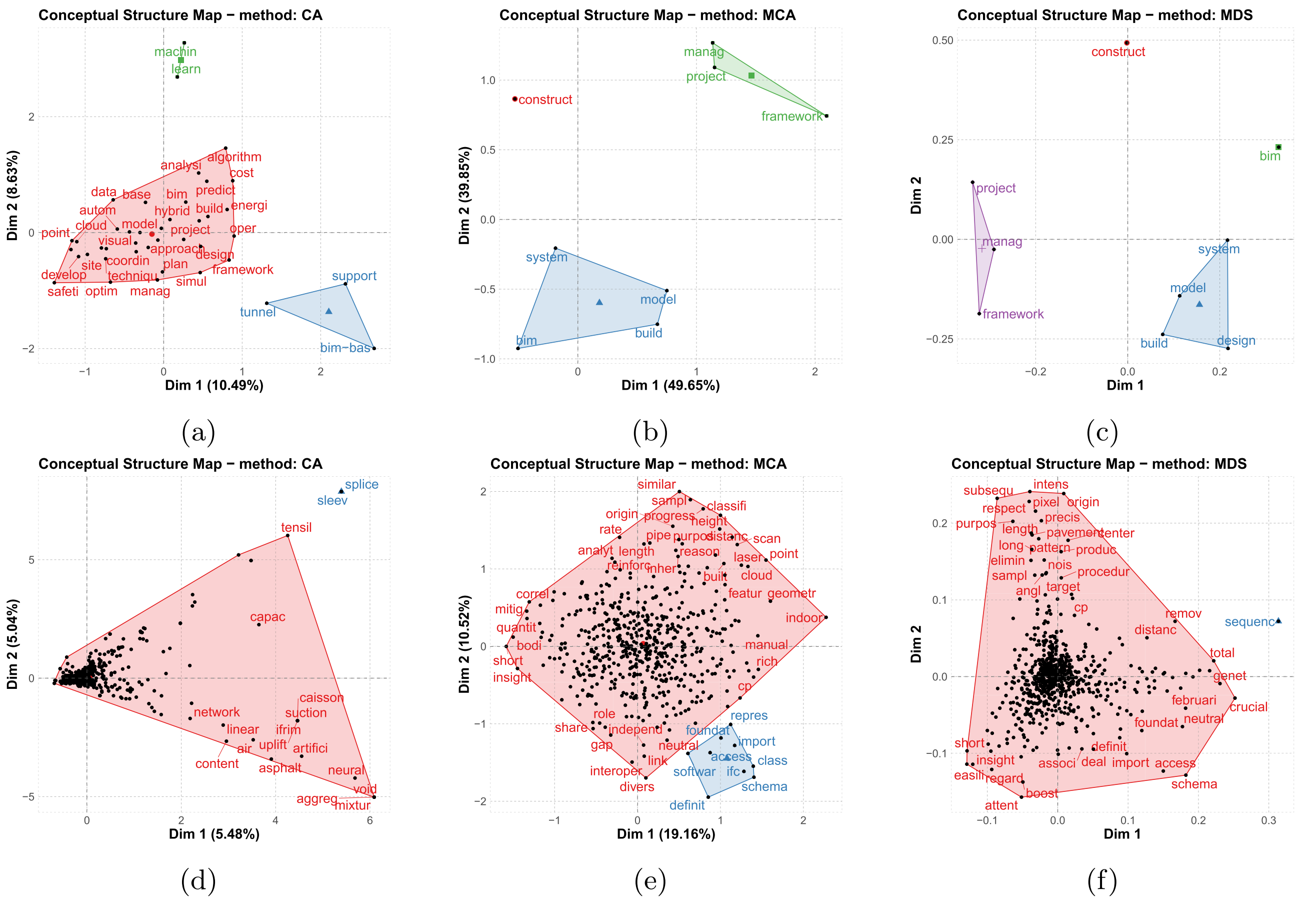


FIGURE 14. Textual clustering analysis derived from `Title` section and `Abstract` sections of full-text articles for "AEC-AI Industry" and `Optimization` and `specialized query` constrained search. Low dimensional projection of concepts derived from the `Title` section by the CA (a), MCA (b) and MDS (c) methods. An equivalent textual and clustering analysis is made from `Abstract` section (d,e,f).

relation between time lapsed from publication and the TC index 18-c.

Correspondence analysis and clustering techniques help to visualize and extract the conceptual structure of the literature corpus, however, it becomes more difficult to visualize as the number of text increases, in the visualization as well as in the computation requirements. Specialized and recently constrained type of search queries allows to better define the conceptual structure of the field research by selecting the most predominant sources and the newest developments.

In the case of `Title` section derived clustering, some of the similarity groups/clusters detected contain the following concepts sets: `base, detect, structure, design, automat`. These probably reflect on the aim of the automatic design of structures and processes. `assess, digit, build, analysis`. A rather basic pattern, building analysis, and digital assessment are commonly brought forwards in BIM related AEC-AI industry. The closest concept to `machine learning` is `concrete`, which reflects on the advances developed in the optimized use of this material in the context of the AEC Industry. `manag, industri, optim, construct, plan, technolog, integr, approach, prefab, project, site`. This variety of terms and concepts are a sample of a target application of `Optimization` and AI technology. Some application-specific in prefabrication is also suggested. `concret, learn, machin`. One of the problems exploded by machine learning from a variety of points of view is concrete, as a material with properties and compositions in construction as well as in structural design, resistance, and life cycle evaluation.

All these concepts match the rich repertory of terms related to the AEC industry. On the other hand, in figure 18-b, an equivalent analysis is performed for the same set of articles, in this case, data analysis and dendrogram are derived from `Abstract`'s section words. In particular, the "standard" dendrogram visualization become more difficult as the word in abstracts increase and then the concepts to cluster, then we projected the clustering in phylogenic form to allow its inspection, unfortunately still is not possible to accommodate the content for evaluation without reducing font size; the clustering still closely reflects the variety of content in the underlying literature sample. However, the detailed concept constructs derivation must be complemented by other techniques. Then, we performed a textual analysis on the `Title` and `Abstract` section derived contents, (figure 18-d and e).

a: TITLE SECTION THEMATIC ANALYSIS

The topological themes map shows more defined themes with a predominance of `analysis`, `learning` and `construction`. We found that the concepts co-occurring in the different quadrant as follows:

BASIC: The next dominant theme in the growth is `analysis` associated with: `building`, `bim`, `planning`, `modeling`, `assessment`, `buildings`, `control`, `engineering`, and `integration`. In the case of `learning`, it associates with `machine`, `concrete`, `risk`, `algorithms`, `data-driven`, `framework`, `surface`, `image`, and `performance`. In the case of `construction`, relates with `approach`, `projects`, `industry`, `digital`, `optimization`, `fabrication`, `factors`, `field`, and `management`.

MOTOR: The case of `design` associates to co-occur with, `collaborative`, `material`, `reality`, `selection`, `support`, `system`, `virtual`, `nil`, and `nil`. For `automatic` we have, `detection`, `deep`, `defects`, `method`, `estimation`, `models`, `networks`, `sequences`, and `sewer`. For the case of `based` we find, `systems`, `bridge`, `safety`, `work`, `algorithm`, `generation`, `heterogeneous`, `mep`, and `point`.

NICHE: Whereas that `materials` associated to `accuracy` which co-occurs with `activity`.

EMERGENT: Likewise, `data` associated to `structural`, `monitoring`.

b: ABSTRACT SECTION THEMATIC ANALYSIS

The themes map for the `Abstract` section derived analysis, shows less defined themes with predominance of `Construction` as basic theme.

BASIC: The next dominant theme in the growth is `construction` co-occurring with: `study`, `project`, `performance`, `safety`, `management`, `time`, `projects`, `process`, and `risk`.

MOTOR: The case of `model` is associated to co-occurring with, `model`, `bim`, `building`, `models`, `approach`, `industry`, `modeling`, `developed`, and `work`.

NICHE: Whereas that `proposed` associates to `method`, `system`, `framework`, `learning`, `systems`, `maintenance`, `machine`, `inspection`, and `surface`

EMERGENT: Likewise, `analysis` associates to `analysis`, `material`, `ar`, `structures`, `walls`, `high`, `materials`, `large`, and `capacity`.

Together these results showed that the constrained `specialized recent query` search captures a literature body containing basic construction, building, and `Optimization` covering main approaches using machine learning and artificial intelligence, including the specialized topics in development. Moreover, the literature sample corresponds to highly cited recent works published in leading journals in the fields of engineering, construction, material science.

5) TITLE AND ABSTRACT SECTIONS' DIMENSIONAL REDUCTION ANALYSIS

We studied the conceptual structure using three methods dimensional reduction and clustering methods for the conceptual content derived from text in the `Title` and `Abstract` sections of the full-text articles figure 19 (a–c) and figure 19 (d–e) respectively.

We found that `Title` section and `Abstract` section's derived low dimensional projections allow cluster formation, however, MCA and MDS methods allow better clustering results, as well as `Abstract` section, produces a higher number of clusters.

From `Title` section derived analysis, we found that one, two, and three clusters are defined by the CA, MCA, and MDS methods respectively. The MCA methods produce a cluster center in `concrete`, where `machine, learning, algorithm, estimate, base, risk`, denote the borders indicating the use in concrete derived problems in AEC-AI Industry.

The non-linear MDS methods, segregate the same cluster as found by MCA, in addition to a cluster delineated by `prefab, optimization, model, system, bim, plan, site`, the third cluster revealed by these methods delineates: `digital, management, data, material, analysis, network, build, integration`.

In the case of `Abstract` section derived analysis, three clusters are projected for CA and MCA methods, whereas the MDS method reveals four groups. This last method delineates a small cluster refereeing to `make, current, tool, lack, foundation`, indicating the state and aim of tools. The next cluster centers in `challenges` and contains `term, focus, address, benefit, fall, product, object`. Other clusters group more difficult to interpret concepts as they become more populated.

Feature detection by low dimensional projection with machine learning methods for dimensional reduction allows a broad sense of the content of a literature body derived from a combined analysis derived from `Title` and `Abstract` sections.

`Title` section derived analysis produces less populated classes, allowing for more specific concept allocations. In some cases, process and/or materials can appear as the case of `prefabrication` and `concrete`.

`Abstract` section derived analysis produces an equivalent amount of clusters, enforcing the notion that the methods used allow extracting low dimensional features with low conceptual resolution.

A low number of clusters (3-4) were systematically produced from `Title` section and `Abstract` section analysis, most probably reflecting the structure of research content with problem definition, main results, and methodology in a compressed sentence structure. With proper recognition

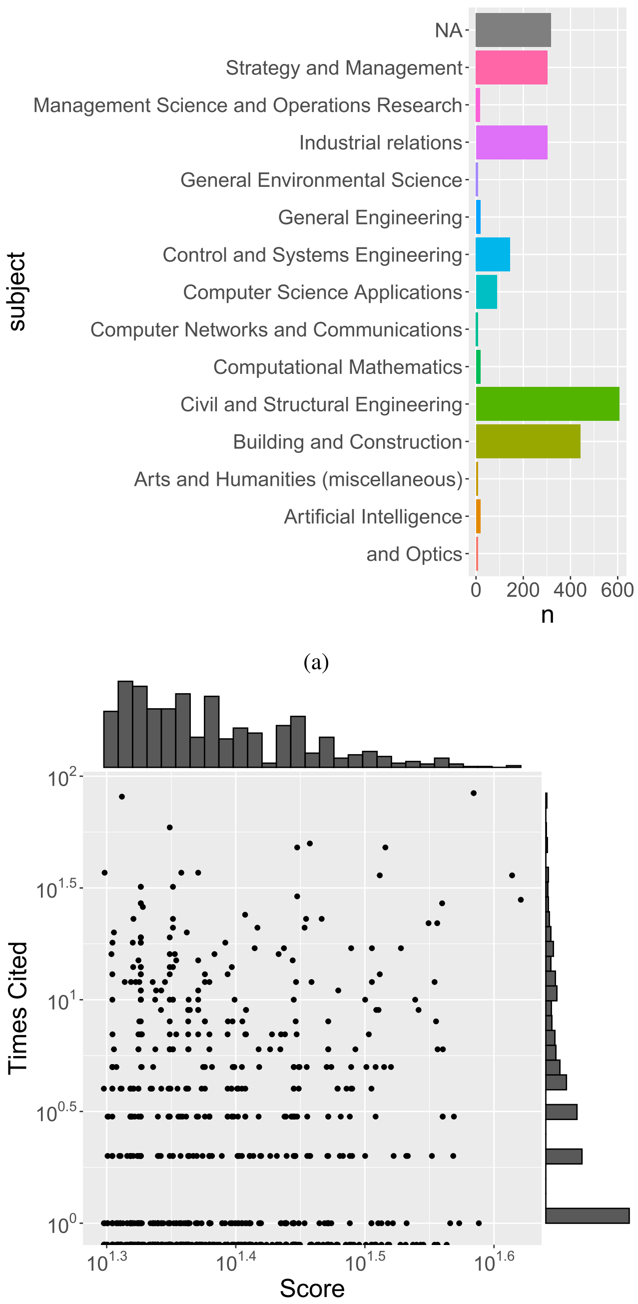
Journal	Num	Mean TC	sd
Journal of Construction Engineering and Management	283	1.20	2.20
Mathematics	185	1.66	3.31
Automation in Construction	160	10.80	12.20
Journal of Legal Affairs and Dispute Resolution in Engineering and Construction	95	0.87	1.45
Journal of Computing in Civil Engineering	86	2.35	4.54
Engineering Structures	66	3.23	4.97
Computational Intelligence	18	0.22	0.43
Journal of Management in Engineering	17	3.41	4.02
IEEE Photonics Journal	9	1.22	1.48
Journal of Optical Communications and Networking	9	5.89	10.66
Practice Periodical on Structural Design and Construction	7	0.57	0.79
Journal of Hydrologic Engineering	6	3.00	3.90
Processes	6	0.33	0.52
Journal of Structural Engineering	5	1.00	1.73
Risks	3	3.33	4.16
Attention, Perception, & Psychophysics	2	1.50	2.12
Computer−Aided Civil and Infrastructure Engineering	2	0.50	0.71
Dermatologic Therapy	2	0.50	0.71
Healthcare	2	40.50	57.28
IEEE Embedded Systems Letters	2	1.00	1.41
Journal of Architectural Engineering	2	1.00	1.41
Journal of Bridge Engineering	2	8.00	11.31
Journal of Energy Engineering	2	1.50	2.12
Journal of Environmental Engineering	2	5.00	7.07
Journal of Professional Issues in Engineering Education and Practice	2	5.50	3.54

FIGURE 15. **Main journals present in the 1K sample retrieved query for "AEC-AI Industry" and** `Optimization` **for the** `specialized recent query`**. The number of articles by journal (Num) and mean citation times (Mean TC) and standard deviations (sd) are shown for top 25 more frequent.**

of their limitations, clustering and dimensional reduction, in combination with constrained search, allows to iterative and progressively define and construct the main components present in the literature corpus for a defined field/topic.

V. DISCUSSION

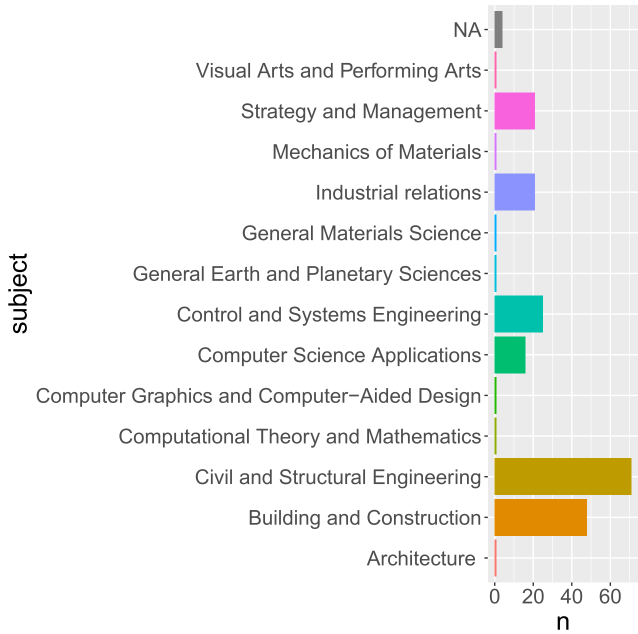
In this work, we presented an adaptive method to analyze the literature corpus of a particular area/discipline. In this instance, we have addressed the Architecture, Engineering, and Construction Industry related to Artificial Intelligence (AEC-AI Industry) and its conceptual corpus in relation to `Optimization`. We think this approach is particularly useful as an adaptive methodology for complex, diversified, and extensive research areas as the AEC Industry combined with the AI field. AEC and AI are different industries, but their combined action would produce an enormous impact. Therefore, there is significant research activity around them and also an enormous challenge for integration, given the complexity of processes involved. This makes its study, research, and development a considerable effort to researchers and teams. With this background, we have started to develop a methodology to examine and review relevant research by a combination of methods, from bibliometric reviews to textual NLP. First, we addressed the issue of the data source by using an open-access multidisciplinary platform. CrossRef, is a growing repository covering about 80% of the major paywall platforms, containing a growing collection of unstructured data on research in science and humanities. Therefore,



(a)

(b)

FIGURE 16. Disclosed subjects, impact, and relevance metrics for "AEC-AI Industry" and `Optimization` query for the constrained search: `specialized recent query`. (a) Subjects disclosed for a set of 1K more relevant articles from CrossRef. (b) Scatter and distribution histogram of relevance score index and times cited index respectively.



(a)

Journal	Num	Mean CT	sd
Automation in Construction	27	14.4	13.6
Journal of Construction Engineering and Management	21	1.9	3.1
Journal of Computing in Civil Engineering	15	1.8	3.8
Engineering Structures	6	3.7	2.9
Computer−Aided Civil and Infrastructure Engineering	1	2.0	NA
GIScience} {\&} Remote Sensing	1	43.0	NA

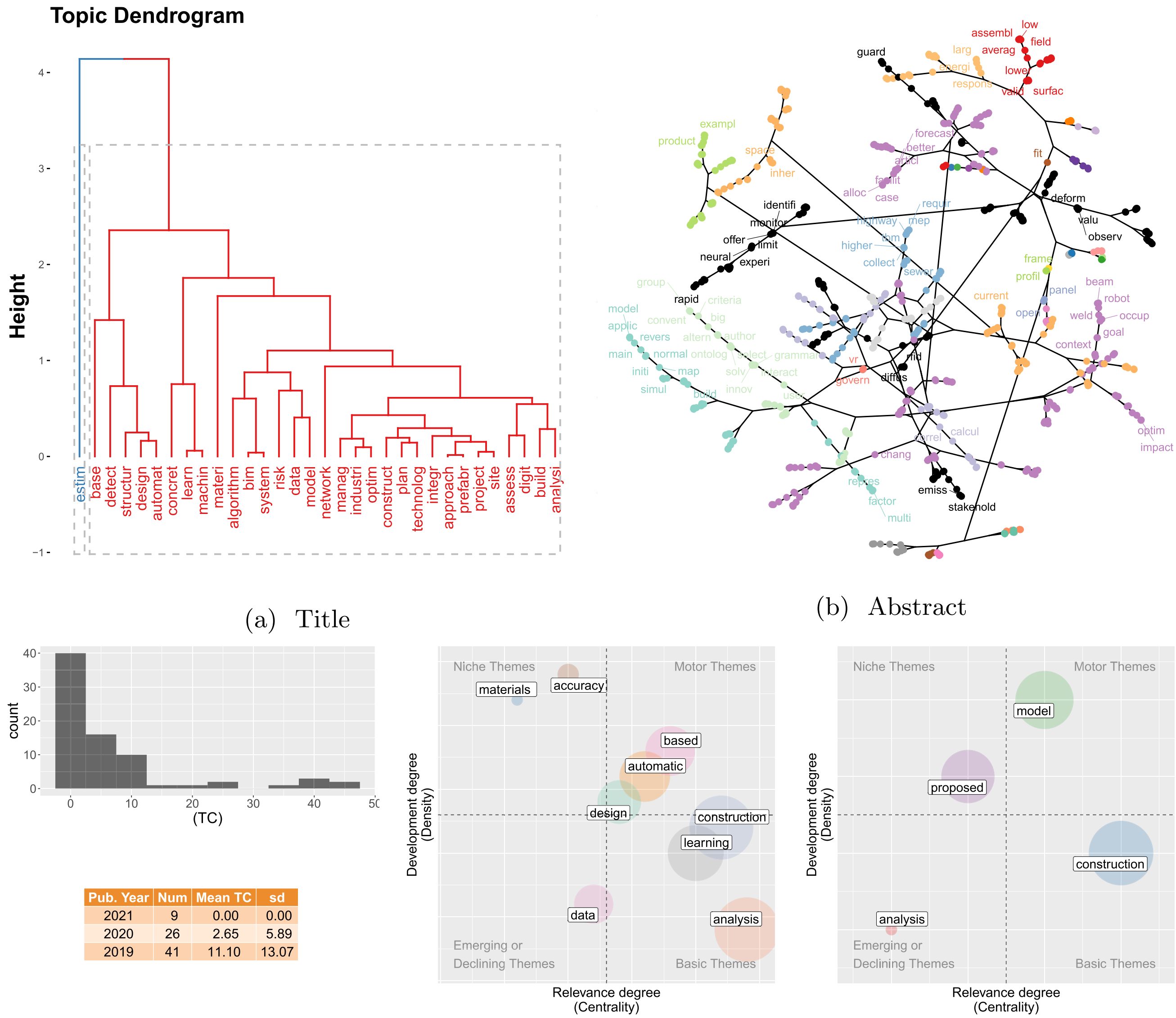
(b)

FIGURE 17. Subjects and journal metrics from a sample of full-text articles sample retrieved for "AEC-AI Industry" and `Optimization` using the constrained search `specialized recent query`. Number of documents (works) with the reported topic/theme (a) and the number (Num) and mean times cited (mean/sd) of documents in the retrieved full-text sample set, grouped by journal, (b).

levering rOpenSci and CrossRef enables a vast horizon of applications, requiring few pieces of assembly. Together, we have mounted an adaptive methodology that allows the progressive incorporation of methods and analysis tools.

We have implemented a methodology leveraging the intrinsic capabilities of AI implemented in the search engine, using a leader multidisciplinary open-access platform as CrossRef. In addition, the methodology by design is adaptable to input methods as well as further AI and analytical methods. At this stage, a macroscopic aim is implemented to map the conceptual and structural components with a combination of bibliometric and machine learning methods. In further stages, we have adapted the method to obtain more granularity, and in this case, to differentiate further the domain and techniques in AI using the `Optimization` (in preparation). However, at this macroscopic level, the adaptive methodology captures a throughout representation of the high-level conceptual knowledge content, as is obtained with classical review methods.

For the particular case in this instance, AEC-AI industry-related query was recursively analyzed to extract complete works amenable to analyze by NLP methodologies. Therefore, we generated three queries; the first was aimed to recover a broad perspective in the related works, only



Pub. Year	Num	Mean TC	sd
2021	9	0.00	0.00
2020	26	2.65	5.89
2019	41	11.10	13.07

(a) Title

(b) Abstract

(c) Articles impact

(d) Title

(e) Abstract

FIGURE 18. `Title` **section derived textual terms analysis for the full-text sample retrieved for "AEC-AI Industry" and** `Optimization` **query** `specialized recent query`**. (a) Dendrogram of hierarchical clustering from** `Title` **section derived words from the sample of full-text articles from the most relevant retrieved documents. (b)** `Abstract` **section derived hierarchical clustering as in (a). (c) Distribution histograms of times cited TC (upper panel) and articles by year with associated TC stats (lower panel). Documents themes mapping derived from a sample of full-text articles, for Articles' derived** `Title` **section (d) and** `Abstract` **section (e).**

constrained to a maximum of 10 years from the publication date, `general query` search. The second, constrained search uses the same criteria as the broad, however also constraint the publisher to a set of related journals. A third one, constrained `specialized recent query`, the search to the same related journal by for the last three years maximum. In this way, the scoring relevance algorithm operates, extracting relevance over different, but related sets of data. Next, we discuss in more detail the results obtained.

Journal sources retrieved for the first general query, and the subjects disclosed for the records, show two main areas of publishing - AEC-Industry and AI/ML. However, is clear that many records are not in the aimed range. We noted, from manual screening many, either only or independently on each specific area, i.e. some articles were purely on AEC-Industry with no content of AI/ML or vice versa. The search results appeared more loosely connected to the aimed search queries, as uses not boolean connections between records. Although that the set of publisher journals display a broad range of editorial scope, most influential ones agree with findings by other independent methods [2].

In order to improve the retrieved concordance with the aimed search, a combination of criteria must be implemented. In the first instance, we implemented a workflow that allows

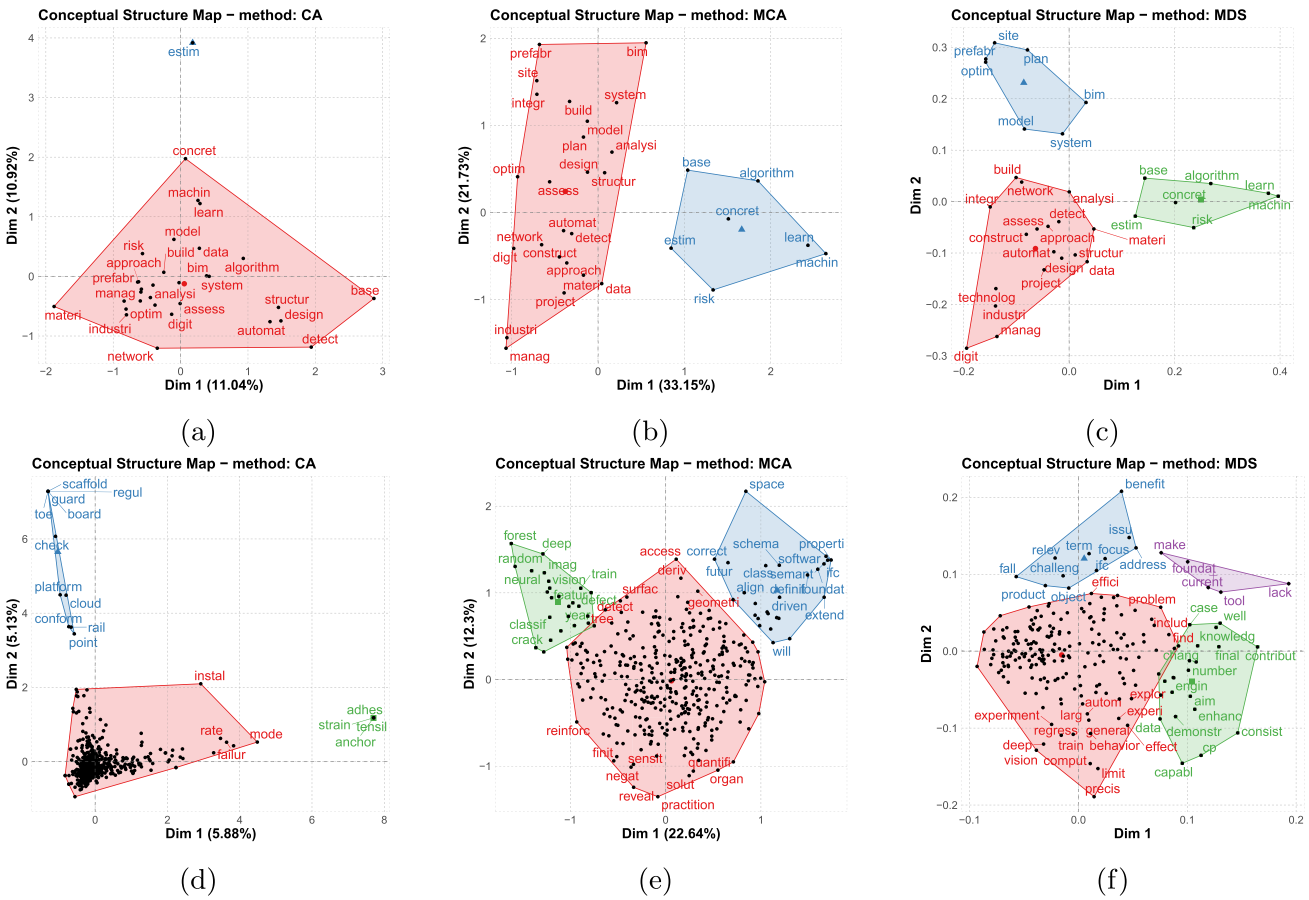


FIGURE 19. `Title` **section and** `Abstract` **section derived clustering for "AEC-AI Industry" and** `Optimization` **query for** `specialized recent query` **retrieved results. Clustering analysis content from the sample of full-text documents. (a) Low dimensional projection of concepts derived by the CA method. (b) Clustering projection for documents contents derived with MCA method. (c) Nonlinear MDS method for derived clustering of contents. The upper row corresponds to** `Title` **section of the articles and the lower row (d-f) corresponds to** `Abstract` **section derived content from the documents.**

constraining the retrieved records in order to decrease the noise in the search and capture more concordant records. A specific methodology should complement the query search results for the CrossRef platform. The underlying algorithm for unstructured data matching used in CrossRef searches offers advantages and disadvantages. The detailed operation of the algorithm assigning the relevance score index by CrossRef is not known. However, fuzzy logic implementation or another scoring algorithm can be coupled on a recurrent basis such as progressive characterization-evaluation cycle allows less biased and general assessments to purely boolean implementations. An adaptive macroscopic methodology is implemented, comprising constrained searches and also evaluating the quality and pertinence of the content retrieved in the search at macroscopic bibliometric levels. The generally constrained query yield `591090` records associated with the AEC-AI Industry and `Optimization`, more than 10 times the number of records characterizing the whole AEC-AI spectrum reported independently [2]. These results indicate that the algorithm scoring relevance produces more broad search results than a boolean search operation. A condition that must be controlled and may also be disadvantageous, as more loosely related items are matched in the search. In some cases, intriguing content was obtained as some records-documents are disclosed in the more faraway subjects than expected for the aimed query. Basic bibliometric analysis indicates that it might be necessary to combine search queries and other techniques to obtain more related search results in line with expected results, in absence of boolean retrieved searches.

Approximately 10-15% of the top one thousand records in the result query were recovered in full-text samples available. The implemented method, adaptive macroscopic characterization, allows to further incorporate those search results and documents to the literature corpus obtained. The method also features adaptive input data and research objectives as well as scalability towards big data size. Furthermore, documents can be segregated and analyzed on a knowledge-content type basis. In this case, full-text documents were separated into Reviews and Articles types and were analyzed separated.

Times cited index for the full-text sample of documents, grouped by journal, follows the trends in content observed in the retrieved 1K sample. `Title` section and `Abstract` section derived clustering more likely represent the grammatical structure along with the combinatorial representation of key conceptual contents in the literature corpus formed by the available full-text sample. Articles types showed more variations in the feature set as is revealed in the hierarchical clustering dendrogram than equivalent clustering for Reviews type of documents. Reviews derived clustering of words from `Title` section reveal a more pure/trimmed tree structure indicating a more homogeneous and compact feature set on the derived concepts. Reviews type of documents show features well in agreement with the type of document aimed to develop a systematic review of the literature with high-level conceptualization. The key term in this representation was `learn*, construction, bim`, which exemplifies the high-level conceptual representation for a set of curated sources.

On the other hand, the results of performing correspondence analysis and dimensional reduction allow systematically approach the content of the sampled articles. From `Title` section derived content analysis, there is a clear suggestion that `design`, and `construction materials` are being addressed using AI/ML methods as well as `BIM` framework. Reviews derived content appear more defined for the analyses corresponding to the `Abstract` section, whereas for the articles type documents a higher variety of content is observed. However, for themes mapping derived from `Abstract` section of Articles, the results display less variety, indicating that some common conceptual mapping is shared. Probably reflecting the structural approach where `model` is central, whereas that other intrinsic commonality of content is the aiming towards `performance`. Therefore, depending on the structure of the literature corpus to analyze, the conceptual structure map can be properly assessed. Articles and reviews type of documents show a differential pattern of structure, with the review reflecting a lower degree of variations in contents.

The constrained search query, in the form of the `specialized query`, yields more specific results associated with the `Optimization` search and the AEC-AI industry. Here `13739` records were obtained in response to the query. Using a constrained search query related to published sources, it was possible to obtain records from highly specialized journals. The constrained search retrieved enriched content associated with selected journals, again it should be noticed that relevance score does not respond to boolean operators and a broad set of related journals are obtained in combination to the `basis` search query and the `Optimization` key-term. Therefore, other sources were also retrieved with no immediate connection to the expected results. The use of constrained search criteria may allow a broad variety of records to be analyzed at the time of also using more selective criteria. Constraining search criteria is possible to retrieve records displaying more tuned results and allowing to compare with other search results and analysis. Using constrained research criteria allows more closely aim at the target without an enormous sacrifice of generality, but reducing noise.

Combining textual analysis content and themes mapping techniques allows to evaluate the development of themes. Notably, for AI-related themes, in the results for the constrained `specialized query` search, the transition from the analysis of `Title` section derived mapping to `Abstract` section derived mapping, a change occurs similar to when comparing articles to reviews type of works. From "specific" to "high level -general" theme mapping type of transition, which agrees that `Title` section reflects in more specific the contents of the document, whereas `Abstract` section further elaborates on the connected concepts and structures. Full-text derived analysis related to `specialized query` constrained research query yielded a literature pool of documents which seems overall more homogeneous. All three conceptual structure map methods used, allow to barely form a single central cluster with quite reduced satelliting ones.

Search queries constrained to specialized and recent works, `specialized recent query`, yielded `9822` records, and maintain or slightly modifies the top frequency of publishers of highest incidence. This "recent specialized query" allows focusing the scope of retrieved results, based on the relevance score, preserving the components of disclosed subjects. In general, the available full-text articles cover a broad spectrum of literature bodies produced for the AEC-AI industry related to `Optimization`. Constrained search is useful to extract and define recent specific conceptual structure content in the literature corpus. Clustering and visualization techniques become more difficult as the text content increases. In the case of `Title` section derived clustering, some of the similarity groups/clusters detected contains the following concepts sets: `base, detect, structure, design, automat`. These probably reflect on the aim of automatic design of structures and processes [79]–[87].

The enclosed cluster by `assess, digit, build, analysis`, a rather basic pattern, building analysis, and digital assessment are commonly brought forwards in the BIM-related AEC-AI industry. The closest concept to `machine learning` is `concrete`, reflects on the advances developed in the optimized use of this material in the context of the AEC Industry. `manag, industri, optim, construct, plan, technolog, integr, approach, prefab, project, site`. This variety of terms and concepts are a sample of a target application of `Optimization` and AI technology. Some application-specific in prefabrication are also suggested [7], [88]–[90]. `concret, learn, machin`. One of the problem exploited by machine learning from a variety of point of views is concrete, as material - with properties and compositions - in construction as well as in structural design, resistance and life cycle evaluation [83], [91]–[99]

Therefore, with proper recognition of their limitations clustering and dimensional reduction, in combination with constrained search, allows to iterative and progressively define and construct the main components present in the literature corpus for a defined field/topic. The results obtained with the "AEC-AI Industry" `basis` query and `Optimization`, show good agreement with other independent characterization of the overall subject [2], however, our implemented method uses a more broad spectrum of relevance to the aimed objective, which allows to include more loosely connected documents. We expected in the future to compare boolean and relevance score comparison in the results. However, the method to obtain good overall results from the open-access multidisciplinary platform conferees advantages over paywall platforms or closed sources.

A. RESULTS SUMMARY & COMPARATIVE

We designed, as an adaptive methodology, a procedure allowing a quick and progressive refinement of subject and areas by a combination of NLP methods and expert knowledge. We analyzed about 450 academics documents related to the AEC-AI industry and optimization methods. In sections III-B and IV-A results are presented in a summary along with the methodological perspective. It has shown useful and accuracy in recovering macroscopic assessment of a literature corpus equivalent to other review methods. Allows to leverage relevance score or other methods implemented by search engines. We have leveraged the CrossRef platform to use/exploit metadata as publisher, dates, and relevance score. Although the methods are extensible to other metadata fields and content.

Three target direction was used to develop an initial landscape of `Optimization` use in the AEC-AI industry.

1) GENERAL SEARCH

Which yield 591090 records. In section IV-C we shown how curated content in the forms of Reviews display simpler structural features of hierarchical and phylogenic thematic content at the time that `construction industry` emerges with: `analysis`, `architecture`, `history`, `practices`, `data`, and `sustainable` as topics; which become tackled by `review` associated with `literature`, `systematic`, `implementation`, `aec`, `adoption`, and `green`. Moreover, *basic motor* theme is `bim`, the technological backbone, associated with `trends`, `infrastructure`, `integration`, and `safety`. Complemented by `engineering` resources of: `applications`, `learning`, `intelligence`, `machine`, and `materials`. These curated Reviews sources are contrasted with more diversity of topics with dominant basic themes include `machine learning` associated with: `concrete`, `prediction`, `strength`, `adaptive`, `asphalt`, and `deep`. Here the MOTOR of development centers `optimization` with a middle degree of developing and higher centrality, which associated with `multi-objective`, `production`, `recognition`, `structures`, `classes`, and `discrete`. These analysis also reveal `properties` associated with `mortar`, `durability`, `engineering`, `geopolymers`, `influence`, and `powder` as specialized niche topics, complemented by `cost`-topic developments associated with `estimation`, `impact`, and `parameters`.

2) SPECIALIZED SEARCH

With 13739 records. Section IV-D detailed result are presented for specialized publisher during a period of the last 10yrs This pool of works contains a greater diversity of topics. However, a basic structure of *emergent* `Construction propositions` ot develop *building* application is the most robust with *performance* as the newly *in-development* concept. Some of the detailed topics in this axis of developing *construction propositions for building* in the AEC-AI industry are: `framework(s)` related with `management`, `bim-based`, `planning`, `simulation`, `control`, `offshore`, which seems to include abroad repertory of macroscopic interest. This effort are combined closely by `modeling – building` which are associated with `cost`, `knowledge`, `bridge`, `estimation`, `integration`, `networks`, `accident`, and `applications`. for the former, whereas than the later includes, `bim`, `design`, `energy`, `structural`, and `algorithm`. The above developments are complemented by more basics topics in `construction` associated with `safety`, `collaborative`, `monitoring`, `case`, `integrating`, and `mep`. Specific areas/topic where `models` been proposed include `fuzzy`, `hybrid`, `steel`, `concrete`, `prediction`, and `capacity`. Here `machine`, `learning`, `algorithms`, `retrofit`, `seismic`, `detecting` and `risk` are developed in connection with `buildings`.

3) RECENT SPECIALIZED SEARCH

Yielding 9822 records. Similarly, in section IV-E, `analysis`, `learning` and `construction` contain the main direction of developments. `bim`, `planning`, `control` and `integration` underlying some of the directions for `analysis`. Whereas that the topics in development associated with `learning`, include an important sample of the main topics raised by other methods. Then, `machine learning` empowers `concrete` probably in the many faces in which this material become key (structural, sustainability, design, cost). Similarly `risk`, also is present in several dimensions amenable to optimization algorithms. `surface` and `image`, probably share a link in the use of algorithm for image analysis assessment of structural health. In the case of `construction`, relates with `projects`, `digital`, `fabrication`, and `management` express the main topics of development Overall, the results obtained offer a macroscopic landscape of themes where a rich texture and diversity topics are been leveraged with AI methods and optimization. Nevertheless, similar as the limitation with traditional methods of literature revision, the detailed structure of developments is not grasped and mesoscopic and microscopic methods become need. We have performed intrinsic validation against curated literature corpus in the form of reviews.

Overall, the method allows a macroscopic assessment of the use of `Optimization` in the AEC-AI industry. There are several challenges to overcome for this purpose:

- AEC-AI industry research field is large, complex, spanning many types of problems, processes and activities.
- Combining AEC-industry alone is a plethora of components in a complex life-cycle.
- AI is in a new era of theoretical and practical developments, reaching all spectrum of activities.
- AEC and AI industries are in course of incorporating an increasingly set of advanced CIT and industrial technologies in the form of Industry 4.0.
- Synergism of all areas is expected to occur to explode literature and research directions.
- Advancing analytical tools to integrate development and research in these areas can greatly improve and advance the field.

Here we have initiated an adaptive methodology for assessment and development of the AEC-AI industry which incorporates open-access platforms, domain-specific expert knowledge, and curated content. The methods can be extended to:

- Incorporate automation.
- More and new methods in AI and Statistics, including scientometric and bibliometric methods.
- Use in other research fields.

VI. CONCLUSION

`Optimization` is a multi-faced concept, especially in the area of AI and ML. `Optimization` can be used to indicate efficiency or improvement, or in more technical jargon extremization of functions, turning into an essential component of ML algorithms. Moreover, specialized areas of research are focused on the element in the search for solution spaces and optimal solutions as metaheuristics. On the other hand, the architecture, engineering, and construction AEC industry include an enormous amount of interdependent processes which is a common desire that requires optimal interaction and continuity. Therefore, a complex mapping is expected around the concept of `Optimization`. For these reasons, we designed an adaptive methodology that may allows characterizing the `Optimization` in its all richness of conceptual structure. In this work, we present the initial stage where high-level macroscopic conceptual mapping occurs. As we are interested in the Knowledge mapping and structure associated with the `Optimization` we did not characterize the networking part of the classical bibliometric counterpart. However, by only analyzing `Title` and `Abstract` sections derived text with machine learning methods, we were able to recover a concordant and coherent representation of the knowledge structure at this level. Moreover, the methods use the available metadata and platform from the free open-access platform, implying a little gain in relation to pay, closed, and structured data sources. The method incorporates the availability of full-text documents to extract and analyze directly from research documents the knowledge structure. Therefore, only biased by the availability of documents. Overall, the proposed procedure allows us to map and extract the conceptual structure with an internally curated representation as a point of reference. Our results indicate that material science investigation uses machine learning methods to optimize the design as well as material properties of concrete. Similarly, other material forms in the construction industry and infrastructure are also involved in the optimization methods/algorithm, as for example asphalt. Building Information Modeling (BIM) is also a central concept to the use of `Optimization`, the whole life cycle of building is subject to approaches requiring `Optimization`.

Overall, we conclude that the development of methods and procedures allows us to efficiently characterize the conceptual structure in adaptive, from macroscopic to microscopic, which can allow us to form a more detailed landscape of use, deficiencies, and requirements to advance in an integrated platform for the AEC-AI industry 4.0.

We are in the process of continually adapting the method to incorporate more NLP and machine learning methods. Similarly, we are extending the analysis to other domain-specific areas to integrate results and to improve biases in search and coverage of related terms.

AUTHOR CONTRIBUTIONS

CM and JG designed the research, CM wrote code, performed the research, and wrote the manuscript. All authors commented and edited and improved the manuscript up to the final version. All authors have read and agreed to the published version of the manuscript.

REFERENCES

[1] J. Woo, S. Shin, A. T. Asutosh, J. Li, and C. J. Kibert, "An overview of state-of-the-art technologies for data-driven construction," in *Proc. Int. Conf. Comput. Civil Building Eng.*, in Lecture Notes in Civil Engineering, vol. 98. Cham, Switzerland: Springer, 2021, pp. 1323–1334,.

[2] A. Darko, A. P. C. Chan, M. A. Adabre, D. J. Edwards, M. R. Hosseini, and E. E. Ameyaw, "Artificial intelligence in the AEC industry: Scientometric analysis and visualization of research activities," *Autom. Construct.*, vol. 112, Apr. 2020, Art. no. 103081.

[3] L. Ceccon and D. Villa, "AI-BIM interdisciplinary spill-overs: Prospected interplay of AI and BIM development paradigms," in *From Building Information Modelling to Mixed Reality* (Springer Tracts in Civil Engineering). Cham, Switzerland: Springer, 2021, pp. 195–217.

[4] Z. Wang, B. He, Y. Yang, C. Shen, and F. Peña-Mora, "Building a next generation AI platform for AEC: A review and research challenges," in *Proc. 37th Int. Conf. (CIB)*, São Paulo, Brazil, Aug. 2020, pp. 27–45. [Online]. Available: http://itc.scix.net/paper/w78-2020-paper-003

[5] L. M. Kipper, S. Iepsen, A. J. Dal Forno, R. Frozza, L. Furstenau, J. Agnes, and D. Cossul, "Scientific mapping to identify competencies required by industry 4.0," *Technol. Soc.*, vol. 64, Feb. 2021, Art. no. 101454.

[6] H.-Y. Chong, C.-Y. Lee, and X. Wang, "A mixed review of the adoption of building information modelling (BIM) for sustainability," *J. Cleaner Prod.*, vol. 142, pp. 4114–4126, Jan. 2017.

[7] X. Li, G. Q. Shen, P. Wu, and T. Yue, "Integrating building information modeling and prefabrication housing production," *Autom. Construct.*, vol. 100, pp. 46–60, Apr. 2019.

[8] M. Tang, H. Liao, V. Yepes, A. Laurinavičius, and L. Tupėnaitė, "Quantifying and mapping the evolution of a leader journal in the field of civil engineering," *J. Civil Eng. Manage.*, vol. 27, no. 2, pp. 100–116, Feb. 2021.

[9] B. Sherafat, C. R. Ahn, R. Akhavian, A. H. Behzadan, M. Golparvar-Fard, H. Kim, Y.-C. Lee, A. Rashidi, and E. R. Azar, "Automated methods for activity recognition of construction workers and equipment: State-of-the-art review," *J. Construct. Eng. Manage.*, vol. 146, no. 6, Jun. 2020, Art. no. 03120002.

[10] G. B. Ozturk and M. Tunca, "Artificial intelligence in building information modeling research: Country and document-based citation and bibliographic coupling analysis," *Celal Bayar Üniv. Fen Bilim. Derg.*, vol. 16, pp. 269–279, Sep. 2020.

[11] E. T. Santos and S. Scheer, *Proceedings of the 18th International Conference on Computing in Civil and Building Engineering* (Lecture Notes in Civil Engineering), vol. 98. Cham, Switzerland: Springer, 2021.

[12] D. Luckey, H. Fritz, D. Legatiuk, K. Dragos, and K. Smarsly, "Artificial intelligence techniques for smart city applications," in *Proc. Int. Conf. Comput. Civil Building Eng.*, in Lecture Notes in Civil Engineering, vol. 98, Aug. 2021, pp. 3–15.

[13] F. Parisi, A. M. Mangini, and M. P. Fanti, "Enabling technologies for smart construction engineering: A review," in *Proc. IEEE 16th Int. Conf. Automat. Sci. Eng. (CASE)*, Aug. 2020, pp. 1546–1551.

[14] T. Slaton, C. Hernandez, and R. Akhavian, "Construction activity recognition with convolutional recurrent networks," *Autom. Construct.*, vol. 113, May 2020, Art. no. 103138.

[15] P. K. Y. Wong, H. Luo, M. Wang, and J. C. Cheng, "Enriched and discriminative human features for person re-identification based on explainable behaviors of convolutional neural networks," in *Proc. Int. Conf. Comput. Civil Building Eng.*, in Lecture Notes in Civil Engineering, vol. 98, 2021, pp. 41–53.

[16] J. K. Chow, Z. Su, J. Wu, Z. Li, P. S. Tan, K.-F. Liu, X. Mao, and Y.-H. Wang, "Artificial intelligence-empowered pipeline for image-based inspection of concrete structures," *Autom. Construct.*, vol. 120, Dec. 2020, Art. no. 103372.

[17] E. Karan, M. Safa, and M. J. Suh, "Use of artificial intelligence in a regulated design environment—A beam design example," in *Proc. Int. Conf. Comput. Civil Building Eng.*, in Lecture Notes in Civil Engineering, vol. 98, 2021, pp. 16–25.

[18] J. Singh and J. C. Cheng, "Automating the generation of 3D multiple pipe layout design using BIM and heuristic search methods," in *Proc. Int. Conf. Comput. Civil Building Eng.*, in Lecture Notes in Civil Engineering, vol. 98, 2021, pp. 54–72.

[19] J. A. Abdalla and R. A. Hawileh, "Assessment of effect of strain amplitude and strain ratio on energy dissipation using machine learning," in *Proc. Int. Conf. Comput. Civil Building Eng.*, in Lecture Notes in Civil Engineering, vol. 98, 2021, pp. 98–108.

[20] D. Shin, "How do people judge the credibility of algorithmic sources?" *AI Soc.*, 2021, doi: 10.1007/s00146-021-01158-4.

[21] D. Shin, "The effects of explainability and causability on perception, trust, and acceptance: Implications for explainable AI," *Int. J. Hum.-Comput. Stud.*, vol. 146, Feb. 2021, Art. no. 102551.

[22] D. Shin, "Embodying algorithms, enactive artificial intelligence and the extended cognition: You can see as much as you know about algorithm," *J. Inf. Sci.*, Jan. 2021, doi: 10.1177/0165551520985495.

[23] N. N. Barros and R. C. Ruschel, "Machine learning for whole-building life cycle assessment: A systematic literature review," in *Proc. Int. Conf. Comput. Civil Building Eng.*, in Lecture Notes in Civil Engineering, vol. 98. Cham, Switzerland: Springer, 2021, pp. 109–122.

[24] M. Khanzadi, A. Kaveh, M. R. Moghaddam, and S. M. Pourbagheri, "Optimization of building components with sustainability aspects in BIM environment," *Periodica Polytechnica Civil Eng.*, vol. 900, pp. 289–308, Nov. 2018.

[25] B. Abbasnejad, M. P. Nepal, A. Ahankoob, A. Nasirian, and R. Drogemuller, "Building information modelling (BIM) adoption and implementation enablers in AEC firms: A systematic literature review," *Archit. Eng. Des. Manage.*, 2020, doi: 10.1080/17452007.2020.1793721.

[26] B. Guo and T. Feng, "Mapping knowledge domains of integration in BIM-based construction networks: A systematic mixed-method review," *Adv. Civil Eng.*, vol. 2019, pp. 1–12, May 2019.

[27] M. D. Martínez-Aires, M. López-Alonso, and M. Martínez-Rojas, "Building information modeling and safety management: A systematic review," *Saf. Sci.*, vol. 101, pp. 11–18, Jan. 2018.

[28] B. Zhong, H. Wu, H. Li, S. Sepasgozar, H. Luo, and L. He, "A scientometric analysis and critical review of construction related ontology research," *Autom. Construct.*, vol. 101, pp. 17–31, May 2019.

[29] C. Newman, D. Edwards, I. Martek, J. Lai, W. D. Thwala, and I. Rillie, "Industry 4.0 deployment in the construction industry: A bibliometric literature review and UK-based case study," *Smart Sustain. Built Environ.*, 2020, doi: 10.1108/SASBE-02-2020-0016.

[30] K. Xue, M. U. Hossain, M. Liu, M. Ma, Y. Zhang, M. Hu, X. Chen, and G. Cao, "BIM integrated LCA for promoting circular economy towards sustainable construction: An analytical review," *Sustainability*, vol. 13, no. 3, p. 1310, Jan. 2021.

[31] Z. A. Shukra and Y. Zhou, "Holistic green BIM: A scientometrics and mixed review," *Eng., Construct. Archit. Manage.*, 2020, doi: 10.1108/ECAM-05-2020-0377.

[32] Z. Liu, Y. Lu, and L. C. Peh, "A review and scientometric analysis of global building information modeling (BIM) research in the architecture, engineering and construction (AEC) industry," *Buildings*, vol. 9, no. 10, p. 210, Sep. 2019.

[33] S. L. M. Correa and E. T. Santos, "BIM support in the tendering phase of infrastructure projects," in *Proc. Int. Conf. Comput. Civil Building Eng.*, in Lecture Notes in Civil Engineering, vol. 98, 2021, pp. 383–392.

[34] J. M. F. Correia, G. C. de Oliveira Neto, R. R. Leite, and D. da Silva, "Plan to overcome barriers to reverse logistics in construction and demolition waste: Survey of the construction industry," *J. Construct. Eng. Manage.*, vol. 147, no. 2, Feb. 2021, Art. no. 04020172.

[35] C. M. Herr and T. Fischer, "BIM adoption across the Chinese AEC industries: An extended BIM adoption model," *J. Comput. Des. Eng.*, vol. 6, no. 2, pp. 173–178, Apr. 2019.

[36] D. Calvetti, P. Mêda, M. C. Gonçalves, and H. Sousa, "Worker 4.0: The future of sensored construction sites," *Buildings*, vol. 10, no. 10, p. 169, Sep. 2020.

[37] L. Wang, M. Huang, X. Zhang, R. Jin, and T. Yang, "Review of BIM adoption in the higher education of AEC disciplines," *J. Civil Eng. Educ.*, vol. 146, no. 3, Jul. 2020, Art. no. 06020001.

[38] D. Fürstenberg, "Information management in AEC projects: A study of applied research approaches," in *Proc. Int. Conf. Comput. Civil Building Eng.*, in Lecture Notes in Civil Engineering, vol. 98, 2021, pp. 272–284.

[39] A. Adibfar, A. Costin, and R. R. A. Issa, "Design copyright in architecture, engineering, and construction industry: Review of history, pitfalls, and lessons learned," *J. Legal Affairs Dispute Resolution Eng. Construct.*, vol. 12, no. 3, 2020, Art. no. 4520032.

[40] M. Abdulkarem, K. Samsudin, F. Z. Rokhani, and M. F. A. Rasid, "Wireless sensor network for structural health monitoring: A contemporary review of technologies, challenges, and future direction," *Struct. Health Monit.*, vol. 19, no. 3, pp. 693–735, 2020.

[41] S. A. Biancardo, N. Viscione, A. Cerbone, and E. Dessì, "BIM-based design for road infrastructure: A critical focus on modeling guardrails and retaining walls," *Infrastructures*, vol. 5, no. 7, p. 59, Jul. 2020.

[42] X. W. Ye, Y. H. Su, and J. P. Han, "A state-of-the-art review on fatigue life assessment of steel bridges," *Math. Problems Eng.*, vol. 2014, pp. 1–13, Jan. 2014.

[43] S. Alizadehsalehi, A. Hadavi, and J. C. Huang, "From BIM to extended reality in AEC industry," *Autom. Construct.*, vol. 116, Aug. 2020, Art. no. 103254.

[44] A. Assadzadeh, M. Arashpour, A. Rashidi, A. Bab-Hadiashar, and S. Fayezi, "A review of data-driven accident prevention systems: Integrating real-time safety management in the civil infrastructure context," in *Proc. 36th Int. Symp. Automat. Robot. Construct. (ISARC)*, May 2019, pp. 289–296.

[45] D. Mocrii, Y. Chen, and P. Musilek, "IoT-based smart homes: A review of system architecture, software, communications, privacy and security," *Internet Things*, vols. 1–2, pp. 81–98, Sep. 2018.

[46] M. Bilal, L. O. Oyedele, J. Qadir, K. Munir, S. O. Ajayi, O. O. Akinade, H. A. Owolabi, H. A. Alaka, and M. Pasha, "Big data in the construction industry: A review of present status, opportunities, and future trends," *Adv. Eng. Inform.*, vol. 30, no. 3, pp. 500–521, Aug. 2016.

[47] Y. Yashiro, R. Ueda, F. Hatori, and N. Yabuki, "An interview-based method for extracting knowledge of skilled workers at construction sites using photographs and deep learning," in *Proc. Int. Conf. Comput. Civil Building Eng.*, in Lecture Notes in Civil Engineering, vol. 98, 2021, pp. 26–40.

[48] M. Jia, A. Komeily, Y. Wang, and R. S. Srinivasan, "Adopting Internet of Things for the development of smart buildings: A review of enabling technologies and applications," *Autom. Construct.*, vol. 101, pp. 111–126, May 2019.

[49] S. Tang, D. R. Shelden, C. M. Eastman, P. Pishdad-Bozorgi, and X. Gao, "A review of building information modeling (BIM) and the Internet of Things (IoT) devices integration: Present status and future trends," *Autom. Construct.*, vol. 101, pp. 127–139, May 2019.

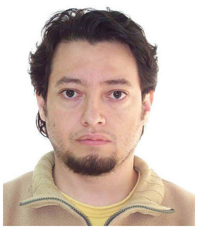
[50] P. Dallasega, E. Rauch, and C. Linder, "Industry 4.0 as an enabler of proximity for construction supply chains: A systematic literature review," *Comput. Ind.*, vol. 99, pp. 205–225, Aug. 2018.
[51] M. Thibaud, H. Chi, W. Zhou, and S. Piramuthu, "Internet of Things (IoT) in high-risk environment, health and safety (EHS) industries: A comprehensive review," *Decis. Support Syst.*, vol. 108, pp. 79–95, Apr. 2018.
[52] J. Li, D. Greenwood, and M. Kassem, "Blockchain in the built environment and construction industry: A systematic review, conceptual models and practical use cases," *Autom. Construct.*, vol. 102, pp. 288–307, Jun. 2019.
[53] N. O. Nawari and S. Ravindran, "Blockchain technologies in BIM workflow environment," in *Proc. ASCE Int. Conf. Comput. Civil Eng. Vis. Inf. Modeling Simulation*. Reston, VA, USA: American Society of Civil Engineers, Jun. 2019, pp. 343–352.
[54] M. Arashpour, T. Ngo, and H. Li, "Scene understanding in construction and buildings using image processing methods: A comprehensive review and a case study," *J. Building Eng.*, vol. 33, Jan. 2021, Art. no. 101672.
[55] J. Wen and M. Gheisari, "Using virtual reality to facilitate communication in the AEC domain: A systematic review," *Construct. Innov.*, vol. 20, no. 3, pp. 509–542, Jun. 2020.
[56] Q. Wang and M.-K. Kim, "Applications of 3D point cloud data in the construction industry: A fifteen-year review from 2004 to 2018," *Adv. Eng. Informat.*, vol. 39, pp. 306–319, Jan. 2019.
[57] X. Liu, X. Wang, G. Wright, J. Cheng, X. Li, and R. Liu, "A state-of-the-art review on the integration of building information modeling (BIM) and geographic information system (GIS)," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 2, p. 53, Feb. 2017.
[58] Y. Song, X. Wang, Y. Tan, P. Wu, M. Sutrisna, J. Cheng, and K. Hampson, "Trends and opportunities of BIM-GIS integration in the architecture, engineering and construction industry: A review from a spatio-temporal statistical perspective," *ISPRS Int. J. Geo-Inf.*, vol. 6, no. 12, p. 397, Dec. 2017.
[59] A. Shah, H. Nasir, M. Fayaz, A. Lajis, and A. Shah, "A review on energy consumption optimization techniques in IoT based smart building environments," *Information*, vol. 10, no. 3, p. 108, Mar. 2019.
[60] S. Eleftheriadis, D. Mumovic, and P. Greening, "Life cycle energy efficiency in building structures: A review of current developments and future outlooks based on BIM capabilities," *Renew. Sustain. Energy Rev.*, vol. 67, pp. 811–825, Jan. 2017.
[61] O. Babalola, E. O. Ibem, and I. C. Ezema, "Implementation of lean practices in the construction industry: A systematic review," *Building Environ.*, vol. 148, pp. 34–43, Jan. 2019.
[62] M. F. Antwi-Afari, H. Li, J. K.-W. Wong, O. T. Oladinrin, J. X. Ge, J. Seo, and A. Y. L. Wong, "Sensing and warning-based technology applications to improve occupational health and safety in the construction industry: A literature review," *Eng., Construct. Archit. Manage.*, vol. 26, no. 8, pp. 1534–1552, Sep. 2019.
[63] E. Valero, A. Sivanathan, F. Bosché, and M. Abdel-Wahab, "Musculoskeletal disorders in construction: A review and a novel system for activity tracking with body area network," *Appl. Ergonom.*, vol. 54, pp. 120–130, May 2016.
[64] N. J. Van Eck, L. Waltman, V. Larivière, and C. Sugimoto, "Crossref as a new source of citation data: A comparison with web of science and Scopus. CWTS blog," Tech. Rep., 2018. [Online]. Available: https://www.cwts.nl/blog?article=n-r2s234
[65] K. Welbers, W. Van Atteveldt, and K. Benoit, "Text analysis in R," *Commun. Methods Meas.*, vol. 11, no. 4, pp. 245–265, 2017.
[66] C. Boettiger, S. Chamberlain, E. Hart, and K. Ram, "Building software, building community: Lessons from the rOpenSci project," *J. Open Res. Softw.*, vol. 3, no. 1, p. e8, 2015.
[67] M. Aria and C. Cuccurullo, "Bibliometrix: An R-tool for comprehensive science mapping analysis," *J. Informetrics*, vol. 11, no. 4, pp. 959–975, Nov. 2017.
[68] K. Benoit, K. Watanabe, H. Wang, P. Nulty, A. Obeng, S. Müller, and A. Matsuo, "Quanteda: An R package for the quantitative analysis of textual data," *J. Open Source Softw.*, vol. 3, no. 30, p. 774, 2018.
[69] S. Chamberlain, "Fulltext: Full text of 'scholarly' articles across many data sources," R Package Version 1.7.0, 2021. [Online]. Available: https://CRAN.R-project.org/package=fulltext
[70] S. Chamberlain, H. Zhu, N. Jahn, C. Boettiger, and K. Ram, "Rcrossref: Client for various 'CrossRef' 'APIs,'" R Package Version 1.1.0, 2020. [Online]. Available: https://CRAN.R-project.org/package=rcrossref
[71] H. Wickham *et al.*, "Welcome to the {tidyverse}," *J. Open Source Softw.*, vol. 4, no. 43, p. 1686, 2019.
[72] M. E. Roberts, B. M. Stewart, and D. Tingley, "{stm}: An {R} package for structural topic models," *J. Stat. Softw.*, vol. 91, no. 2, pp. 1–40, 2019.
[73] B. Grün and K. Hornik, "{topicmodels}: An {R} package for fitting topic models," *J. Stat. Softw.*, vol. 40, no. 13, pp. 1–30, 2011.
[74] RStudio. (2018). *Index @*. [Online]. Available: https://www.Rstudio.Com
[75] *R: A Language and Environment for Statistical Computing*, R Found. Stat. Comput., Vienna, Austria, 2021.
[76] A. Kassambara, *Practical Guide to Cluster Analysis in R: Unsupervised Machine Learning* (Multivariate Analysis), vol. 1. STHDA, 2017, pp. 1–187. [Online]. Available: https://books.google.cl/books?id=plEyDwAAQBAJ
[77] M. J. Cobo, A. G. López-Herrera, E. Herrera-Viedma, and F. Herrera, "Science mapping software tools: Review, analysis, and cooperative study among tools," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 62, no. 7, pp. 1382–1402, Jul. 2011.
[78] S. Fortunato, C. T. Bergstrom, K. Börner, J. A. Evans, D. Helbing, S. Milojević, A. M. Petersen, F. Radicchi, R. Sinatra, B. Uzzi, A. Vespignani, L. Waltman, D. Wang, and A. L. Barabási, "Science of science," *Science*, vol. 359, no. 6379, Mar. 2018, Art. no. eaao0185, doi: 10.1126/science.aao0185.
[79] P. Tavares, C. M. Costa, L. Rocha, P. Malaca, P. Costa, A. P. Moreira, A. Sousa, and G. Veiga, "Collaborative welding system using BIM for robotic reprogramming and spatial augmented reality," *Autom. Construct.*, vol. 106, Oct. 2019, Art. no. 102825.
[80] A. Chatterjee and Y.-C.-Tsai, "Training and testing of smartphone-based pavement condition estimation models using 3D pavement data," *J. Comput. Civil Eng.*, vol. 34, no. 6, Nov. 2020, Art. no. 04020043.
[81] L. Wang, L. Zhuang, and Z. Zhang, "Automatic detection of rail surface cracks with a superpixel-based data-driven framework," *J. Comput. Civil Eng.*, vol. 33, no. 1, Jan. 2019, Art. no. 04018053.
[82] T. N. Khilji, L. L. A. Loures, and E. R. Azar, "Distress recognition in unpaved roads using unmanned aerial systems and deep learning segmentation," *J. Comput. Civil Eng.*, vol. 35, no. 2, Mar. 2021, Art. no. 04020061.
[83] A. Valikhani, A. J. Jahromi, S. Pouyanfar, I. M. Mantawy, and A. Azizinamini, "Machine learning and image processing approaches for estimating concrete surface roughness using basic cameras," *Comput.-Aided Civil Infrastruct. Eng.*, vol. 36, no. 2, pp. 213–226, Feb. 2021.
[84] W. T. Calderon, D. Roberts, and M. Golparvar-Fard, "Synthesizing pose sequences from 3D assets for vision-based activity analysis," *J. Comput. Civil Eng.*, vol. 35, no. 1, Jan. 2021, Art. no. 04020052.
[85] S. S. Kumar, M. Wang, D. M. Abraham, M. R. Jahanshahi, T. Iseley, and J. C. P. Cheng, "Deep learning–based automated detection of sewer defects in CCTV videos," *J. Comput. Civil Eng.*, vol. 34, no. 1, Jan. 2020, Art. no. 04019047.
[86] Z. Xu, R. Kang, and R. Lu, "3D reconstruction and measurement of surface defects in prefabricated elements using point clouds," *J. Comput. Civil Eng.*, vol. 34, no. 5, Sep. 2020, Art. no. 04020033.
[87] A. M. Abdi, "Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data," *GISci. Remote Sens.*, vol. 57, no. 1, pp. 1–20, Jan. 2020.
[88] Y. Al-Qaryouti, K. Baber, and J. M. Gattas, "Computational design and digital fabrication of folded timber sandwich structures," *Autom. Construct.*, vol. 102, pp. 27–44, Jun. 2019.
[89] R. Bortolini, C. T. Formoso, and D. D. Viana, "Site logistics planning and control for engineer-to-order prefabricated building systems using BIM 4D modeling," *Autom. Construct.*, vol. 98, pp. 248–264, Feb. 2019.
[90] A. Nasirian, M. Arashpour, B. Abbasi, E. K. Zavadskas, and A. Akbarnezhad, "Skill set configuration in prefabricated construction: Hybrid optimization and multicriteria decision-making approach," *J. Construct. Eng. Manage.*, vol. 145, no. 9, 2019, Art. no. 04019050.
[91] M. Z. Naser, "Observational analysis of fire-induced spalling of concrete through ensemble machine learning and surrogate modeling," *J. Mater. Civil Eng.*, vol. 33, no. 1, Jan. 2021, Art. no. 04020428.
[92] P. Hüthwohl, R. Lu, and I. Brilakis, "Multi-classifier for reinforced concrete bridge defects," *Autom. Construct.*, vol. 105, Sep. 2019, Art. no. 102824.
[93] C. Sabau, C. Popescu, N. Bagge, G. Sas, T. Blanksvärd, and B. Täljsten, "Local and global behavior of walls with cut-out openings in multi-story reinforced concrete buildings," *Eng. Struct.*, vol. 187, pp. 57–72, May 2019.
[94] E. Gottsäter, M. Johansson, M. Plos, and O. L. Ivanov, "Crack widths in base restrained walls subjected to restraint loading," *Eng. Struct.*, vol. 189, pp. 272–285, Jun. 2019.
[95] M. Poliotti and J.-M. Bairán, "A new concrete plastic-damage model with an evolutive dilatancy parameter," *Eng. Struct.*, vol. 189, pp. 541–549, Jun. 2019.

[96] H. Xie, W. Shi, R. R. A. Issa, X. Guo, Y. Shi, and X. Liu, "Machine learning of concrete temperature development for quality control of field curing," *J. Comput. Civil Eng.*, vol. 34, Sep. 2020, Art. no. 4020031.

[97] J. R. Edwards, A. E. C. Ruiz, J. C. Bastos, Y. Liang, and M. S. Dersch, "Use of field flexural demand data for reliability-based analysis and design of concrete railroad sleepers," *Frontiers Built Environ.*, vol. 6, Oct. 2020, Art. no. 582948.

[98] X. Luo, H. Li, D. Cao, F. Dai, J. Seo, and S. Lee, "Recognizing diverse construction activities in site images via relevance networks of construction-related objects detected by convolutional neural networks," *J. Comput. Civil Eng.*, vol. 32, May 2018, Art. no. 4018012.

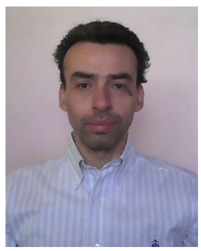
[99] H. Son, C. Kim, and C. Kim, "Automated color model–based concrete detection in construction-site images by using machine learning algorithms," *J. Comput. Civil Eng.*, vol. 26, no. 3, pp. 421–433, May 2012.



CARLOS MAUREIRA received the Ph.D. degree in neuroscience from the University of Chile. He pursuits a broad range of interest, from basic science to technological applications. Inspired by the Autopoietic and Cognition Theory (ACT), he formulated the Biological Organization Principles (BOP) to describe biological systems as we know them in a formal theory. He has performed experimental work in the fields of synaptic communication and plasticity, neuronal biophysics and network excitability in model insects and the mammalian forebrain. His research interests include neuroscience, theoretical biology, and applications of biological cognition, bio-computing, and machine learning to complex problems.

HERNÁN PINTO received the Ph.D. degree from the University of Cantabria. He got a postdoctoral experience at the Structural Engineering Mechanics Group, Civil and Environmental Engineering Department, University of Massachusetts Amherst, USA. He is Construction Engineer from the Pontificia Universidad Católica de Valparaiso. Since 2011, he has been an Associate Professor at the Construction Engineering School, Pontificia Universidad Católica de Valparaíso. From 2020, he serves as Head of the Construction Engineering School. He has participated in several research projects related to the structural mechanics of materials, resilience infrastructures, modeling, and optimization, among others. He has more than 30 published articles and has participated in several international conferences in his field.

VÍCTOR YEPES received the Ph.D. degree in civil engineering. He has been the Academic Director of the M.S. studies in concrete materials and structures, since 2007, and a member of the Concrete Science and Technology Institute (ICITECH). He is currently a Full Professor (tenure) with the Department of Construction Engineering, Universitat Politècnica de València, València, Spain. He is involved in several projects related to the optimization and life-cycle assessment of concrete structures as well as optimization models for infrastructure asset management. He has authored more than 250 journals and conference papers, including 100 published in journal quoted in *JCR*. He serves as an Associate Editor for the *Structure and Infrastructure Engineering* and *Advances in Civil Engineering*.



JOSÉ GARCÍA was born in Santiago, Chile. He received the Ph.D. degree in mathematics from the University of Chile, in 2014, and the Ph.D. degree in computer engineering from the Pontificia Universidad Católica de Valparaíso, in 2018. He graduated as a biochemist and a mathematician. He has served as a consultant in the area of data science and operations research in different industries, such as copper mining, telecommunications, banks, government, and among others. He currently works as an Associate Professor with the Pontificia Universidad Católica de Valparaíso. His lines of research interests include the application of machine learning and heuristics algorithms to different engineering problems.