www.itcon.org - Journal of Information Technology in Construction - ISSN 1874-4753

OPERATIONAL DIGITAL TWINS IN THE BUILT ENVIRONMENT: A SYSTEMATIC REVIEW OF SEMANTIC INFRASTRUCTURE AND DEPLOYMENT PRACTICE

SUBMITTED: April 2026
PUBLISHED: September 2026
EDITOR: Robert Amor
DOI: 10.36680/j.itcon.2026.043

Chady Elias, PhD
Rinker School of Construction Management, University of Florida, Gainesville, FL, USA
chadyelias@ufl.edu

Isabella Hutchins, Undergraduate Student
Department of Civil and Coastal Engineering, University of Florida, Gainesville, FL, USA
hutchins.i@ufl.edu

Raja R. A. Issa, Distinguished Professor
Rinker School of Construction Management, University of Florida, Gainesville, FL, USA
https://orcid.org/0000-0001-5193-3802
raymond-issa@ufl.edu

SUMMARY: *Digital twins (DTs) are increasingly positioned as integration infrastructure for building analytics, automation, and emerging AI-enabled workflows. However, evidence remains limited regarding whether documented deployments provide the information foundations required for scalable adoption across diverse buildings and facility portfolios. This study presents a systematic review of operational digital twins in the built environment, examining semantic integration, operational connectivity, and implications for portable analytics and agent-enabled interaction. Following PRISMA guidelines, records were collected from Scopus, Web of Science, IEEE Xplore, and the ASCE Library. A total of 2,839 records were retrieved, of which 46 studies met the eligibility criteria for full analysis. Indoor environmental quality and energy were the most prevalent application areas, accounting for 34 of the 46 studies, with much of the literature addressing HVAC and building automation. Formal semantic representations were documented in roughly half of the reviewed DTs, but fewer than a third used semantic constructs to shape analytic or decision processes more directly. DTs with stronger semantic integration more often support diagnostic outputs requiring contextual interpretation across equipment and spatial relationships, whereas actions extending to direct control or automated system responses are more commonly realized through fixed, site-specific configurations. Operational outputs also remain largely confined to the platform's own interface, with limited evidence of delivery into maintenance systems, control sequences, or other governed external processes. The findings describe a field that has demonstrated meaningful operational capabilities but has not yet assembled them into coherent architectures that transfer across sites. The review indicates that machine-readable building context and lifecycle-governed metadata remain important constraints on portability, workflow integration, and the structured access needed for emerging agent-enabled interaction. Semantic infrastructure is accordingly framed as a lifecycle information requirement, one that should be established early in the building lifecycle and maintained as buildings evolve, to support durable, scalable, and AI-ready building operations.*

KEYWORDS: *digital twin, semantic interoperability, knowledge graphs, facility management, systematic literature review, AI agents.*

REFERENCE: *Elias, C., Hutchins, I., & Issa, R. R. A. (2026). Operational digital twins in the built environment: A systematic review of semantic infrastructure and deployment practice. Journal of Information Technology in Construction (ITcon), 31, 1015-1041. https://doi.org/10.36680/j.itcon.2026.043*

1. INTRODUCTION

Building operations and maintenance (O&M) increasingly depend on a heterogeneous ecosystem of digital information systems, including building automation systems (BAS), computerized maintenance management systems (CMMS), asset registries, and building information models (BIM). These systems are often adopted at different lifecycle stages to meet distinct operational needs, resulting in divergent data schemas, naming conventions, and vendor-specific structures. Despite this density of digital infrastructure, routine operations often involve manual reconciliation of fragmented records and informal knowledge held by facility personnel (Pishdad-Bozorgi et al., 2018). Substantial effort is devoted to locating authoritative asset records, confirming sensor identity, and assembling contextual history to support informed decisions.

Digital twins (DTs) are widely defined as digital representations of physical systems that are linked through continuous data exchange and synchronization. In this review, operational DTs refer to digital representations of buildings that incorporate operational data in support of building operations and maintenance. In facility operations, DTs serve as an integration infrastructure that unifies operational context across fragmented systems, supporting decision-making through structured, contextualized information retrieval. While industry researchers have extensively studied these systems by developing numerous taxonomies to differentiate integration depth and operational capability (Kritzinger et al., 2018), substantial definitional variation persists. Recent analyses indicate that "digital twin" frequently functions as an umbrella label with an uneven scope, often applied through the relabeling of existing concepts (Abdelrahman et al., 2025). This ambiguity complicates comparisons across studies and obscures the specific level of operational integration evidenced in documented deployments. Prior reviews of built environment DTs report strong emphasis on connected BIM, monitoring, and prototype demonstrations, accompanied by limited evidence of sustained operational integration in day-to-day practice (Deng et al., 2021; Pärn et al., 2017).

Such inconsistencies have become increasingly consequential as DTs are positioned as the foundational context infrastructure for operational analytics and artificial intelligence (AI). For portfolio-level deployment, the decisive constraint is typically the availability of dependable, machine-interpretable context across diverse assets and facilities. Recent cross-domain systematic reviews of DT ontologies indicate that knowledge graphs are frequently used as structured metadata stores, with relatively few documented cases demonstrating semantics embedded directly in runtime decision logic (Karabulut et al., 2024). This gap becomes more consequential as digital platforms and enterprise applications increasingly adopt architectures in which AI agents discover available capabilities through structured, self-describing interfaces and interact with systems through governed tool calls (Yao et al., 2022; Schick et al., 2023). In these tool-augmented architectures, large language models (LLMs) function as reasoning engines that retrieve data, invoke analyses, and initiate actions through external interfaces, reshaping expectations for how information systems present context and support structured requests for retrieval, analysis, and action. For building operations, this pattern positions the DT as the structured interface through which agents would access operational context, when that context is semantically organized and programmatically accessible. Within this framing, AI readiness is treated as an infrastructure property reflecting whether a DT provides machine-readable context and structured access mechanisms that support portable analytics and tool-mediated workflows. Semantic constructs, including ontologies and knowledge graphs, occupy a central role in this assessment, as they determine whether operational context is machine-discoverable through structured representations or remains scattered across local conventions and proprietary data sources.

Standards and schema efforts, including Brick, Project Haystack, and the proposed ASHRAE Standard 223P, reflect growing alignment around machine-interpretable representations intended to support analytics and automation across facilities (Open223, 2026; Pritoni et al., 2024; Project Haystack, 2025). Yet evidence remains limited regarding whether documented built-environment DT deployments provide structured context and operational connectivity consistent with scalable analytics and emerging tool-mediated workflows. This study conducts an implementation-centered systematic review to characterize deployment practices for operational DTs in the built environment, with emphasis on semantic infrastructure and operational integration. The review evaluates documented deployments against the infrastructure conditions needed to support portable analytics and tool-augmented AI workflows, identifying where current practice falls short and what gaps must be addressed for scalable, AI-ready operation. The review is guided by four research questions:

RQ1 (Corpus Characteristics): What are the descriptive characteristics of operational digital twin deployments in the built environment?

RQ2 (Semantic Infrastructure): To what extent do deployments adopt semantic models, and to what extent do semantics drive analytic design and decision logic in deployed systems?

RQ3 (Operational Outcomes): What is the relationship between semantic infrastructure and operational outcomes?

RQ4 (Deployment Barriers): What deployment barriers and infrastructure gaps are evident across documented cases?

The paper is organized as follows. Section 2 provides background on DT concepts in the built environment, semantic context for building operations, and the review literature motivating this study. Section 3 presents the systematic review methodology, followed by results in Section 4 and discussion in Section 5.

2. BACKGROUND

2.1 Digital twin in the built environment

The concept of a digital twin, first articulated in product lifecycle management and subsequently advanced for aerospace applications, generally comprises three elements: a physical asset or system, a digital representation, and data connections that maintain alignment between them (Jones et al., 2020; Grieves, 2014). Taxonomic distinctions differentiate deployments by the direction and automation of data exchange: digital models lack automated data exchange, digital shadows support one-way synchronization from physical to digital, and digital twins exhibit automated bidirectional exchange, as illustrated in Figure 1 (Kritzinger et al., 2018).

Application of DTs to buildings and facilities has expanded alongside advances in sensing, building automation systems (BAS), and building information modeling (BIM). BIM provides a structured geometric and asset representation typically delivered at project handover. These models encode spatial geometry, equipment schedules, and attribute data in federated digital form, supporting design coordination and construction workflows (Eastman et al., 2011). Operational conditions evolve through equipment replacement, sensor reconfiguration, and space changes, while BIM deliverables are not consistently maintained to reflect these changes (Pärn et al., 2017). Accordingly, DT applications in building operations can be distinguished from static BIM by whether the representation is actively linked to data streams and whether that linkage supports ongoing operational functions.

DT deployments in the AECO industry commonly integrate multiple data sources within a common data environment, consolidating BIM, telemetry, and maintenance records into a unified platform (Elias & Issa, 2025; Boje et al., 2020). The operational value of this integration depends on whether sufficient context can be discovered and assembled programmatically, and whether it can be delivered through interfaces and workflow integrations that support routine decision-making by personnel with varying expertise.

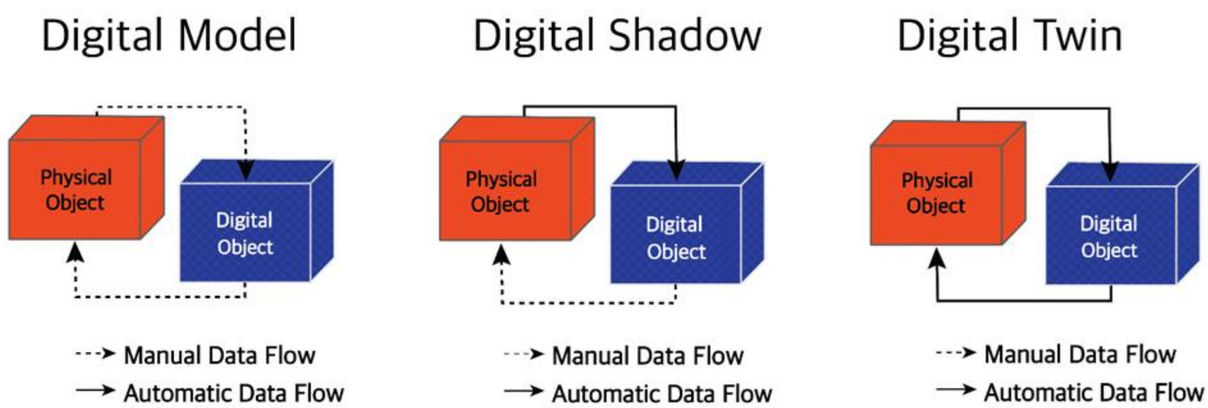

Figure 1: Taxonomy of DTs by data flow directionality (Adapted from Kritzinger et al., 2018).

2.2 Operational digital twin scope

To define a scope aligned with building O&M, the term operational DT is used in this study to denote deployments that satisfy three conditions:

1. Instantiated digital representation: A digital representation of the facility is established (e.g., BIM-derived model, asset graph, equipment registry, or spatial representation) that explicitly represents building entities rather than describing them only at an abstract or conceptual level.
2. Operational data linkage: Operational data (e.g., telemetry, events, status, alarms, or histories) is reliably associated with identifiable entities such that data can be retrieved, interpreted, and compared at the entity level over time.
3. Demonstrated operational use: The linked representation is used to support at least one operational function during the study period, such as monitoring, fault detection and diagnostics, maintenance support, control, or optimization.

This scope emphasizes usability in practice, determined by whether data remain consistently linked to managed assets and spaces in ways that support reliable task execution over time. Kritzinger et al. (2018) distinguish digital models, digital shadows, and digital twins by the direction and automation of information exchange between the physical and digital systems. The operational DT scope remains centered on the use of the digital representation and associated operational data to support building operations and maintenance.

2.3 Semantic context for building operations

A semantic model is a structured representation that assigns explicit meaning to data by defining entities, attributes, and relationships in a shared, machine-interpretable vocabulary. In information systems, semantic models provide a contextual layer between raw data and the applications that rely on it, allowing queries and analyses to operate on defined concepts rather than arbitrary identifiers (Poggi et al., 2008). In building operations, this role is particularly important because data generated by building systems is often interpretable only through local conventions, including vendor-specific metadata, site-specific point naming, and informal knowledge held by facility personnel. A typical BAS point name may encode building abbreviations, floor numbers, equipment tags, and measured variables in a concatenated string, yet both the structure and the meaning of these names vary across sites and vendors (Balaji et al., 2016; Pritoni et al., 2021). This heterogeneity in naming and metadata conventions is widely recognized as a major barrier to scaling building applications, because each new deployment requires substantial manual effort to interpret, align, and remap points and equipment context across sites.

Addressing this barrier requires formalizing the implicit knowledge that experienced operators use to interpret building data. In this study, semantic context refers to machine-interpretable representations of what building entities are, what they measure or control, and how they relate to one another. Two complementary dimensions of semantic context are relevant: entity context and relational context (Xu et al., 2024), as illustrated in Figure 2.

Entity context denotes the set of machine-readable descriptors that specify a building entity's identity, semantic type, functional role, measured or controlled quantity, units, value interpretation, and links to operational data streams and configuration artifacts. Semantic typing enables applications to recognize, for instance, that a given point is a discharge air temperature sensor and not an arbitrary numeric value, or that a piece of equipment functions as a variable air volume (VAV) terminal unit. Several schema and ontology efforts provide standardized vocabularies for expressing such descriptors. Industry Foundation Classes (IFC) provides a comprehensive data model for architectural and mechanical components (Laakso & Kiviniemi, 2012). However, since IFC was developed primarily for design and construction exchange, it provides limited vocabulary for sensor functional roles and real-time data integration (Balaji et al., 2018). Project Haystack defines a controlled vocabulary of tags for annotating points and equipment; its flexible, extensible tag model supports normalization across vendor ecosystems (Project Haystack, 2018a). The Brick schema provides a more explicit ontological class and relationship model for equipment, points, and locations, enabling queries based on semantic type (Brick Consortium, n.d.). ASHRAE Standard 223P proposes concepts for representing BAS information in machine-readable form (Open223, 2026; Pritoni et al., 2024).

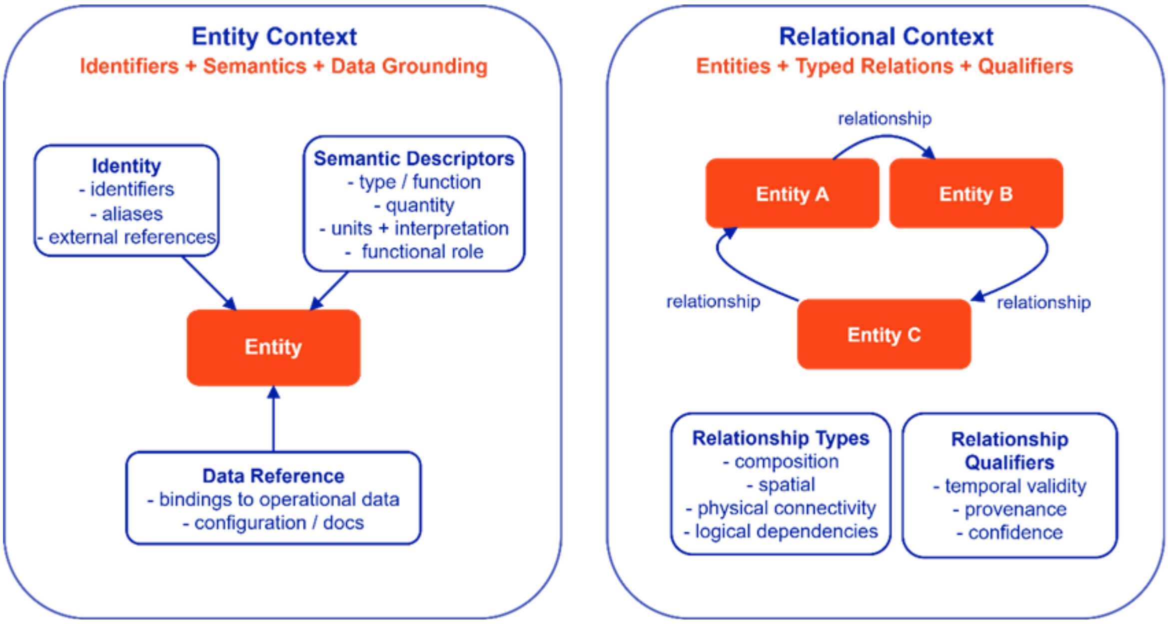

Figure 2: Schematic of entity context and relational context.

Relational context refers to the explicit, machine-interpretable representation of how entities are connected, encompassing both relationship types and relationship qualifiers. Relationship types capture structural and functional linkages such as composition (e.g., which sensors belong to which equipment), spatial associations (e.g., which equipment serves which zones), physical connectivity (e.g., shared distribution infrastructure), and logical dependencies (e.g., coordination implied by control sequences). Relationship qualifiers annotate these links with additional context, such as temporal validity (the period during which a relationship holds), provenance (the source or basis for the asserted relationship), and confidence (the reliability of inferred or extracted relationships).

Standards and ontologies encode these dimensions with varying emphasis. Brick provides relationship predicates such as isPartOf, hasPoint, feeds, and serves to represent equipment composition, point affiliation, and distribution pathways (Brick Consortium, n.d.). ASHRAE Standard 223P emphasizes connection semantics, modeling how equipment interfaces, connection points, and distribution media are linked within building systems (Open223, 2026; Pritoni et al., 2024). Project Haystack encodes containment and distribution relationships through typed reference tags, whereas Brick and ASHRAE 223P formalize equivalent semantics through explicit predicates and connection topology. Formal collaboration among ASHRAE, Project Haystack, and Brick Schema reflects growing alignment on interoperable vocabularies that support programmatic context discovery across these complementary representations (ASHRAE, 2018; Project Haystack, 2018b).

Some relational context can be inferred from BIM geometry and connectivity. In practice, reliable inference remains difficult because BIM data are often fragmented across multiple files, contain modeling errors, and lack the precision needed for consistent system-level reconstruction (Lilis et al., 2025). Prior work has proposed semi-automated approaches for extracting HVAC topology and enriching models through geometric relation checking and knowledge graph representations (Mavrokapnidis et al., 2023; Wang et al., 2024; Lilis et al., 2025). These efforts indicate that relational context can be partially recovered from geometric data, yet consistent system-level reconstruction still requires manual validation and domain-informed correction.

2.4 Analytic portability and semantic sufficiency

Applications can bind to required inputs dynamically when semantic context is represented in a form that supports structured querying. This capability, termed analytic portability, describes the extent to which an application can be deployed across multiple buildings with limited reconfiguration. Balaji et al. (2018) illustrated this by deploying energy efficiency applications across heterogeneous buildings using semantic queries against Brick models.

Recent work extends this logic to control and optimization contexts. Chamari et al. (2025) proposed a portability framework for model predictive control in which control logic is decomposed into modular services with formalized metadata requirements, enabling semantically driven querying, validation, and controller configuration across buildings. Despite achieving partial automation, broader plug-and-play portability remained constrained by heterogeneity and variation in metadata modeling practice.

These examples point toward an application-driven approach to semantic modeling, in which coverage requirements are defined relative to the needs of prioritized applications and expanded incrementally as additional workflows justify extension. Applications declare the entity types, relationships, and descriptors required for execution; the system validates available context against these declarations and uses identified gaps to guide targeted enrichment. Fierro et al. (2022) formalized this principle as semantic sufficiency, demonstrating that applications can specify their metadata requirements in machine-readable form and that modeling effort can be aligned with demonstrated operational value. As analytics and operations increasingly incorporate tool-mediated workflows, in which agents discover required inputs and invoke analyses through governed interfaces, the importance of clearly defined and machine-checkable semantic requirements increases, making application-driven semantic investment both a practical deployment strategy and an architectural consideration.

2.5 Agentic architectures and infrastructure implications

Large language models have progressed from text-generation systems toward reasoning agents that coordinate external tools and data sources to complete multi-step tasks. Early tool-augmented paradigms such as ReAct (Yao et al., 2022) established the core pattern in which the model interleaves reasoning with action, retrieves external information, evaluates returned evidence, and iterates until a request is sufficiently grounded. More recent work has strengthened this pattern through reinforcement-trained reasoning models that support more deliberate intermediate reasoning and self-correction (OpenAI, 2024; DeepSeek-AI, 2025), and through standardized integration frameworks such as the Model Context Protocol (MCP) (Anthropic, 2024), which allow agents to discover available tools, interpret execution requirements, and invoke external functions through a common schema. Collectively, these developments reframe the language model as an orchestration layer whose effectiveness depends less on parametric knowledge than on the structure and governance of the interfaces it interacts with.

This capability is particularly relevant to facility operations, where routine questions are rarely resolved through a single data source. Determining whether a complaint about a "warm" room reflects an actual fault requires interpreting that report against the expected conditions for the specific space, time of use, occupancy pattern, and serving equipment. In practice, answering such questions relies heavily on tacit knowledge accumulated over years on site and, when the available records prove insufficient, on physical investigation to trace duct runs, confirm equipment service areas, or reconcile discrepancies between documentation and as-built conditions. The process can be effective but is inherently bounded by individual expertise, institutional memory, and the time available to assemble context manually. Figure 3 illustrates how human operators and AI agents assemble the context needed to act on such a request.

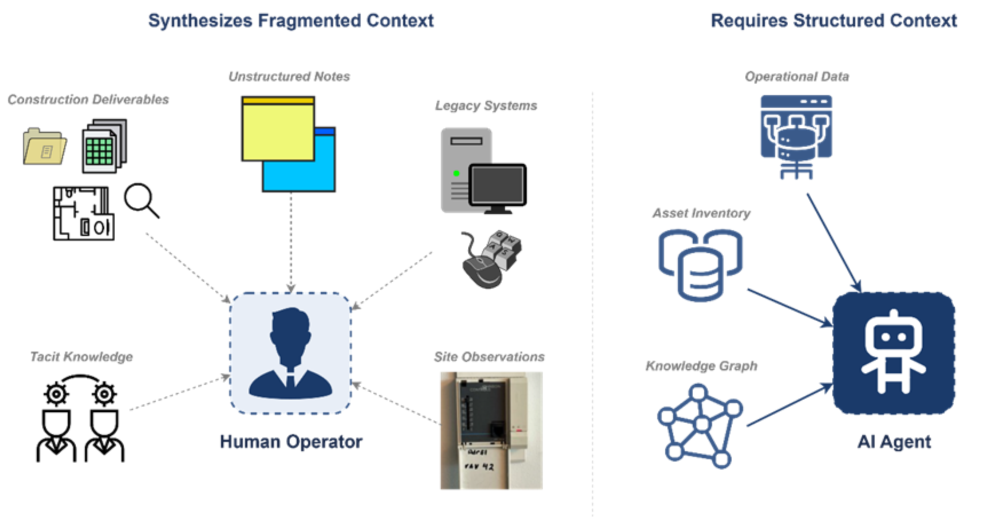

Figure 3: Context assembly by human operators compared with AI agents.

Governance considerations become increasingly important once agents are permitted to act on operational systems. Tool-protocol architectures such as MCP support scoped permissions, separation of read and write operations, and auditable invocation records, properties that are particularly important in cyber-physical domains where erroneous or unauthorized actions can propagate beyond the software layer into physical systems. Early building-domain implementations are already emerging. Li et al. (2025) developed an MCP server for EnergyPlus providing 35 tools for model inspection, editing, and simulation, demonstrating that domain-specific building software can be made accessible to LLM-based agents through structured tool calling. These developments place renewed weight on the persistent challenge of metadata heterogeneity across building systems (Balaji et al., 2016; Pritoni et al., 2021), since standards-aligned ontologies, knowledge graphs, and tagging schemas shape how reliably agents can navigate building context and act on it through structured interfaces. Realizing this potential ultimately depends on whether that context is represented in machine-readable form, rather than scattered across local conventions and proprietary stores.

2.6 Review literature

A substantial body of review literature has emerged on DTs in the built environment, collectively organizing an otherwise diffuse field into recognizable thematic and conceptual clusters. Several recurring themes are evident: definitional and conceptual clarification, mapping of DT application areas, synthesis of analytics and maintenance-oriented methods, and attention to data management and semantic enrichment as enabling infrastructure.

Several reviews highlight the conceptual fragmentation and definitional variation of DT research in the built environment. Deng et al. (2021) synthesized built environment DT research through a staged taxonomy that traces progression from BIM-linked representations toward DT concepts capable of closed-loop, decision-making feedback and autonomous control. Rashidi et al. (2024) analyzed the emerging BIM-DT research landscape through a combined scientometric and systematic synthesis, identifying major thematic clusters and highlighting challenges around standardization, interoperability, data integration, and cybersecurity. Abdelrahman et al. (2025) documented definitional variation and its implications for DT adoption, comparing studies and interpreting reported capabilities.

A second group focuses on facility operations and enabling technologies. Hakimi et al. (2024) provided a bibliometric synthesis of DT-for-facility management research, identifying thematic areas including predictive maintenance, cyber-physical integration, and semantic interoperability. Shahzad et al. (2025) reviewed DT and mixed reality for IoT visualization and reported heterogeneous architectures and limited standardization.

A third group treats DTs as platforms for analytics, with emphasis on fault detection, diagnostics, and AI-enabled maintenance. Hosamo et al. (2022a) compared DT-oriented workflows with traditional practices and highlighted integration challenges across BIM, BAS, and CMMS. Kreuzer et al. (2024) reviewed AI-integrated DT studies and reported limited evidence of bidirectional data exchange between physical and digital systems.

A fourth group emphasizes data management and semantic enrichment as prerequisites for scaling deployments. Correia et al. (2023) synthesized data-management solutions in DT and digital shadow contexts. Benfer and Müller (2024) focused on AI-based metadata inference for semantic DT creation and noted that time-series data alone is often insufficient to recover point meaning and system structure. They also observed that standards and ontologies are most often applied for type labeling, with far less progress on relation inference or extracting operational semantics. Karabulut et al. (2024) reviewed ontology-enabled cross-domain DTs and reported that semantic technologies primarily formalize asset and data meaning and support interoperability, while ontology reuse and richer relationship semantics remain less developed.

Across these reviews, the built environment DT literature is well organized in terms of definitions, taxonomies, application landscapes, analytics approaches, and barriers. Less attention has been given to assessing deployment readiness in ways that distinguish proposed capabilities from demonstrated operational behavior. Although prior research outside DT platforms has shown that semantic input binding can support analytic portability (Balaji et al., 2018; Chamari et al., 2025), it remains unclear whether DT deployments documented in the literature provide comparable infrastructure. As tool-augmented agent architectures mature and the demand for structured, machine-accessible building context grows, characterizing the current state of this infrastructure becomes a practical priority. Section 3 describes the systematic review methodology used to address this question.

3. METHODOLOGY

This section describes the systematic literature review (SLR) methodology used to examine DT deployments in the AECO industry. The review follows established guidelines for systematic reviews in engineering and computing research (Kitchenham & Charters, 2007) and reports findings in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework (Page et al., 2021).

3.1 Review objectives

The review examines deployment characteristics of DT case studies in the built environment, addressing the research questions presented in Section 1. Specifically, the analysis assesses how deeply semantic models are integrated into live system processes and whether delivery pathways exist to connect analytic outputs to operational workflows. In this review, a semantic model refers to any explicit, machine-interpretable representation of entities, properties, and relationships, ranging from lightweight tagging and classification schemas to schema-based information models and formal ontologies or knowledge graphs.

3.2 Search Strategy

A broad preliminary search in Scopus was conducted to identify the terminology associated with DT applications in building operations. Keyword co-occurrence analysis of the retrieved literature produced the network shown in Figure 4, revealing four thematic clusters: BIM and lifecycle terms linked to facility management, built-environment and energy/HVAC terminology, analytics and maintenance terminology, and Industry 4.0 and cyber-physical systems terms. Terms identified through this analysis informed the subsequent formal search strategy and were organized into three concept groups: (i) DT terminology, (ii) built-environment and facility operations context, and (iii) operational and analytic functionality. Terms within each concept group were combined using Boolean OR, and the three groups were combined using Boolean AND. The resulting search expression was adapted to the syntax and indexing conventions of each database while preserving the same conceptual structure and coverage.

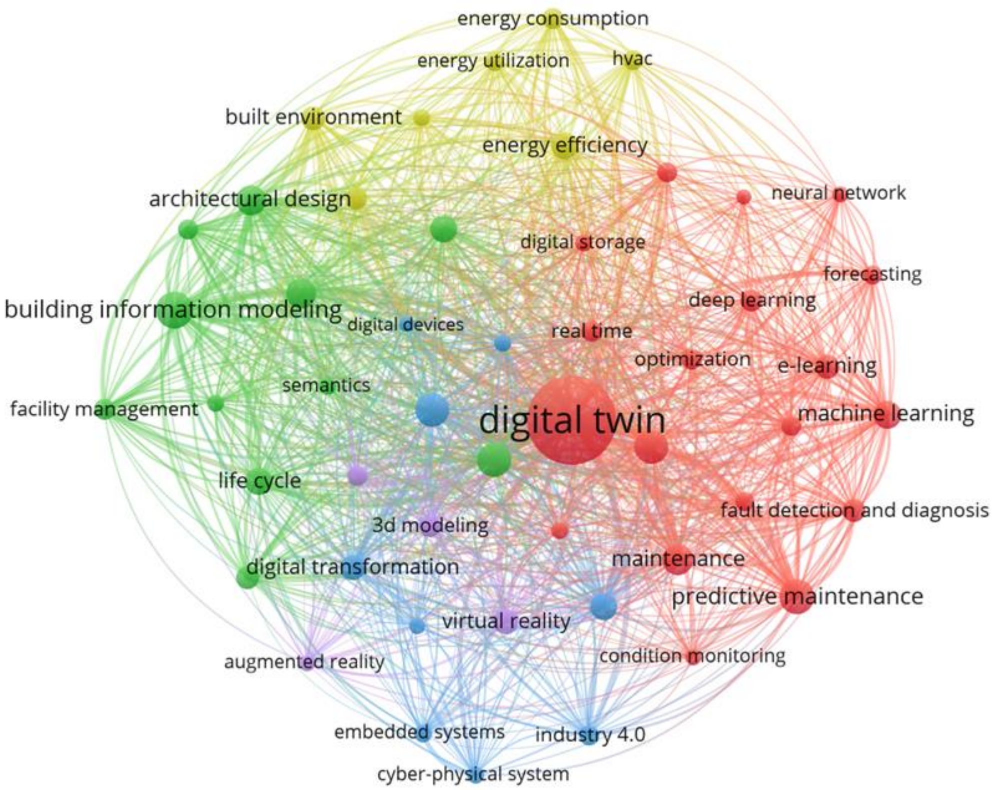

Figure 4: Keyword co-occurrence network from scoping search corpus.

Searches were limited to English-language peer-reviewed journal articles and full conference papers published between January 2015 and August 2025. Four databases were searched: Scopus, Web of Science, IEEE Xplore, and ASCE Library. Table 1 summarizes the databases and initial counts. Database-specific query syntax was adapted for each platform while maintaining equivalent conceptual coverage.

Table 1: Databases searched and records retrieved.

Database	**Coverage**	**Records (n)**
Scopus	Multidisciplinary engineering and science coverage	1,391
Web of Science	Indexed journals and conference proceedings	533
IEEE Xplore	Computing, automation, and smart building venues	805
ASCE Library	Civil engineering and construction informatics	110
Total		**2,839**

3.3 Study selection

Figure 5 presents the PRISMA flow diagram for the study selection process. Records were exported from each database and imported into Rayyan (Ouzzani et al., 2016) for duplicate identification and title-abstract screening. A total of 701 duplicate records were removed, along with 2 retracted records and 76 book chapters or edited volumes, leaving 2,060 unique records for screening.

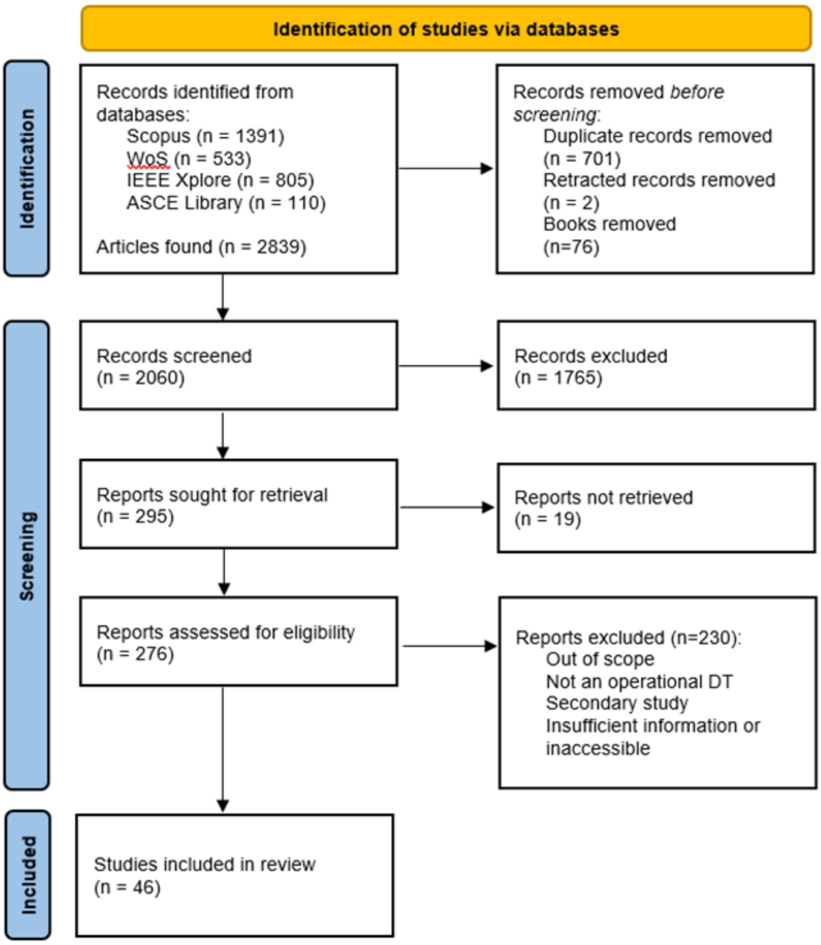

Figure 5: PRISMA flow diagram for systematic review.

Title and abstract screening assessed alignment with the search components. Studies were excluded if they addressed DTs in domains outside the built environment (e.g., manufacturing, aerospace, oil and gas, medicine, wind turbines), presented only conceptual proposals without implementation, or were secondary studies such as reviews and surveys. This screening excluded 1,765 studies, leaving 295 papers for full-text retrieval.

Reports retained for full-text review were managed in Zotero to organize PDFs and metadata. Of the 295 reports sought, 19 could not be retrieved, leaving 276 reports assessed for eligibility. Full-text assessment applied the eligibility criteria presented in Table 2. The DT scope criteria reflect the three conditions defined in Section 2.2: an instantiated representation, linkage to live data sources, and demonstrated use in operations. A total of 230 reports were excluded at this stage, leaving 46 studies in the final corpus.

Table 2: Study eligibility criteria.

Criterion	Requirement
Publication Type	Peer-reviewed journal article or full conference paper; primary research with methods and results
Time and language	January 2015 – August 2025; English
Full text access	Full text retrievable for eligibility assessment
Domain	Building or facility O&M context
Operational DT scope	Satisfies operational digital twin definition (Section 2.2)

3.4 Data extraction

To assess operational capability, two classifications were developed for this review: actionability, representing the highest level of operational output demonstrated, and workflow integration, representing the extent to which those outputs were incorporated into operational processes. Each included study was assessed using a structured extraction form developed iteratively through a pilot phase applied to an initial subset of studies. The schema was drafted based on the research questions and refined after extracting approximately ten studies to resolve ambiguous classifications, calibrate category boundaries, and incorporate fields that emerged as analytically relevant during early extraction. Table 3 presents the final extraction schema.

Table 3: Extracted variables and coding descriptions.

Category	Field	Description
Study context	Year	Publication year
	Country	Study location
	Application area	Primary application domain
	Building function	Primary functional use of the building studied
Semantic infrastructure	Has semantic model	Presence of an explicit semantic representation
	Semantic representation standard	Standard or schema used
	Semantics used in system logic	Runtime use of semantic constructs in system operations
	Semantic tier	Derived classification (see below)
Analytics	Algorithm family	Analytic method categories employed
Operational capability	Workflow integration	Extent to which outputs are delivered beyond the DT interface
	Actionability	Highest level of actionable output demonstrated
	Interface modalities	User-facing delivery methods reported

The primary analytical variable, semantic tier, classifies deployments into three levels. A study was coded as "No Semantic Model" when no formal semantic representation was documented. "Described Only" was assigned when a semantic representation was documented but not demonstrated as active in system behavior. "Used in System Logic" was assigned when semantic constructs were shown to actively participate in system execution, such as scoping queries, resolving point identities, or assembling analytic inputs. This three-level classification separates organizational or documentary uses of semantic models from demonstrated integration into runtime system execution.

3.5 Synthesis and analysis

Descriptive statistics summarize coded variables to characterize the corpus across study context, semantic infrastructure, analytics, and operational capability. Cross-tabulations examine associations between semantic tier and operational variables, including actionability and workflow integration, with row percentages reported to facilitate comparison across tiers. Visualizations were generated using Python (pandas, matplotlib).

4. RESULTS

This section presents findings from the systematic review of 46 operational DT deployments. The analysis covers corpus characteristics, semantic infrastructure patterns, associations between semantic tier and operational outcomes, and analytic methods and delivery modalities. A complete index of included studies is provided in Appendix A.

4.1 Corpus characteristics

Descriptive characteristics for the reviewed corpus (N = 46) are summarized in Figures 6 through 9. Geographic contributions spanned 18 countries, led by China (n=11), the United Kingdom (n=7), and Norway (n=5). Italy contributed three studies, while six countries contributed two each: Sweden, Germany, Singapore, Canada, the United States, and Thailand. Figure 6 presents the geographic distribution.

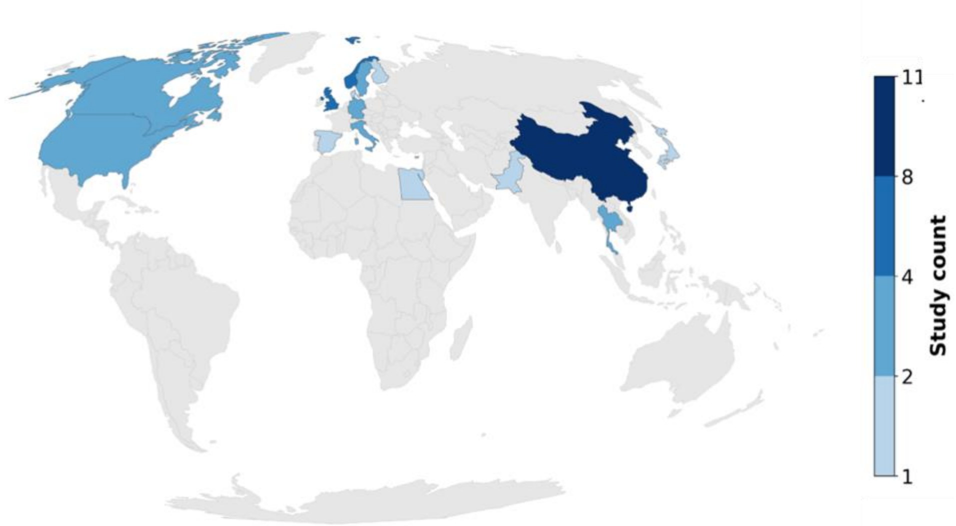

Figure 6: Geographic distribution of the reviewed corpus.

Publication activity concentrated between 2022 and 2025, accounting for 40 of 46 studies, with the largest annual share in 2025 (n=11) and equal contributions from 2024 and 2023 (n=10 each); no eligible studies from 2021 were captured in the final corpus. Figure 7 presents the year distribution.

Educational settings were the most frequently studied building type (n=14), followed by office buildings (n=10) and mixed-use facilities (n=6). Heritage buildings, research facilities, and specialized facility types each accounted for four studies. Healthcare and residential contexts were each represented by two studies. Figure 8 summarizes the building function distribution.

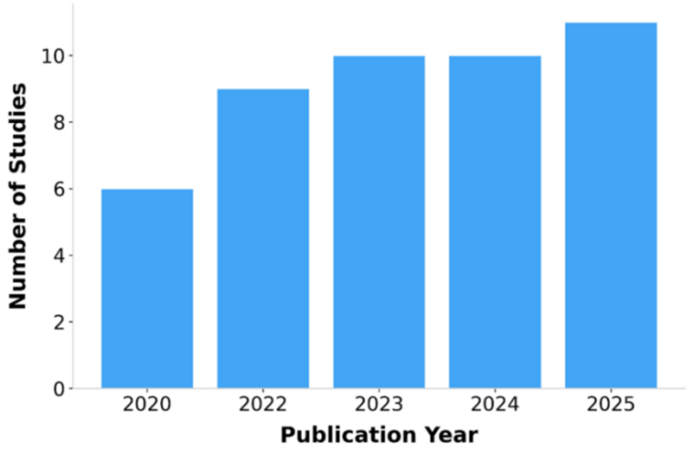

Figure 7: Publication year distribution for the reviewed corpus.

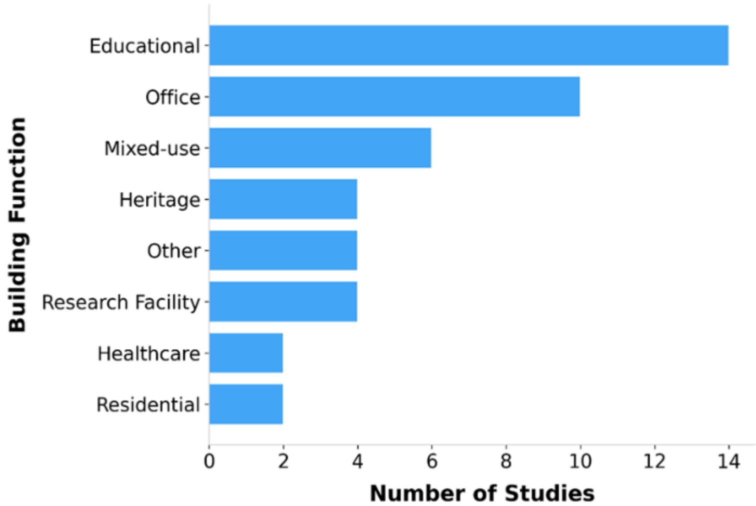

Figure 8: Primary building function by study.

Indoor environmental quality (IEQ) was the most frequently reported application area (n=19), followed by energy (n=15) and maintenance (n=11). Where studies addressed multiple application areas, the primary target was identified based on the stated research objective. Figure 9 presents the application area distribution.

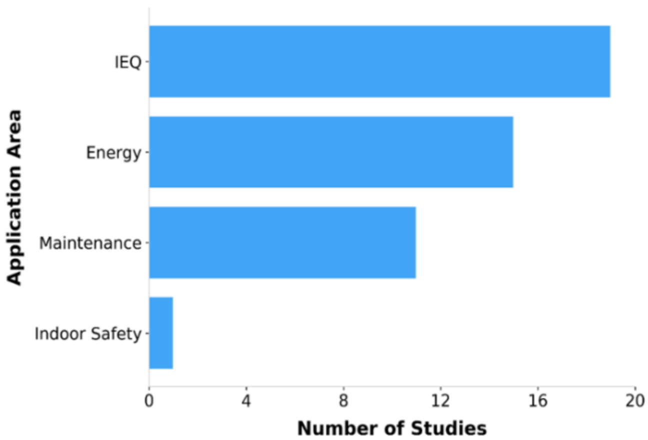

Figure 9: Primary application area distribution for the reviewed corpus.

4.2 Semantic infrastructure

Formal semantic models were documented in 25 of 46 studies. Among the 25 that reported a semantic representation, the semantic tier classification defined in Section 3.4 distinguished 13 studies that described a representation without demonstrating its use in system behavior from 12 that demonstrated semantic constructs actively driving runtime logic. Figure 10 presents the distribution.

Several deployments relied on building structure and identifiers to bind data streams to model elements without a formal representation of typed entities and relationships. Kim et al. (2023) linked parametric BIM with occupancy and lighting sensing for a smart-campus energy prediction twin, passing predefined parameters into components through fixed assignments. A sensor-network twin developed by Gao et al. (2024) streamed indoor environmental measurements to an Unreal Engine visualization, maintaining data associations through device identifiers and model geometry. Ma et al. (2025) integrated BAS and IoT data into a Unity-based mixed-reality interface for a high-rise office floor, binding data streams to model components through COBie-extracted identifiers stored in a relational database and resolved at runtime through component-name matching. Hu and Assaad (2024) spatially registered a mobile robot's sensor readings and UWB-derived position estimates to model elements through manual element-ID lookup and coordinate mapping. Across these deployments, analytic inputs were specified using fixed parameter assignments, device identifiers, and geometric references, with building context encoded in configuration.

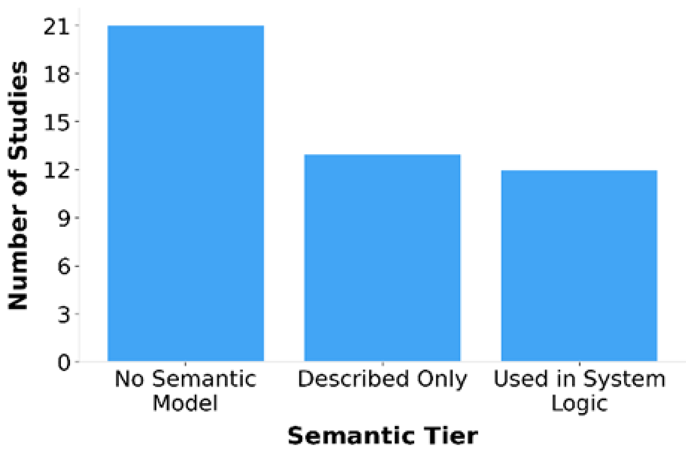

Figure 10: Semantic tier distribution across the corpus.

Studies that documented a semantic model without demonstrating its use in system logic typically described a model, standard, or knowledge graph, while runtime behavior continued to rely on fixed identifier mappings or procedural data handling. Arsiwala et al. (2023), for example, adopted RealEstateCore-aligned DTDL models in Azure Digital Twins for an IEQ-oriented residential twin, specifying entities and relationships such as serves and isCapabilityOf, while runtime updates relied on explicit mappings between physical devices and their corresponding twin entities. Similarly, Hosamo et al. (2023b) proposed a knowledge graph in GraphDB with explicit property mappings linking BIM, BMS, and CMMS data, while operational processing relied on BIM element lookups and predefined data connections. Lu et al. (2020a) likewise integrated BMS, asset management, and sensor data by exporting BIM to IFC and linking IFC objects to operational records via a manually maintained GUID-to-asset-ID matching table stored in DynamoDB; runtime queries resolved these associations using a key-based lookup. Across these deployments, these representations strengthened data organization and exchange, while runtime behavior remained governed by fixed mappings and pre-specified processing steps.

Studies demonstrating semantics in system logic showed that these constructs actively shape runtime behavior through entity discovery and contextual scoping. Merino et al. (2023) implemented a federated data-model approach in which pipelines transformed IFC and Brick into an in-memory hierarchical representation, enabling API operations that retrieved spaces, equipment, and sensors by type and supported traversal across typed entities for HVAC fault detection. The cloud-based twin developed by Ni et al. (2022) was defined as interconnected typed entities linked by defined relationships using DTDL; this allowed queries to combine entity type with position in the network when retrieving data from relevant asset subsets at runtime. Bjørnskov et al. (2025) implemented a

SAREF-based ontology graph in which a translator resolved typed entity classes and their explicit relationships to automatically determine how simulation component models should be connected, so that both the type of each entity and its position in the system topology influenced the assembled simulation at runtime. Collectively, these deployments treated machine-readable context as a resource governing how systems locate relevant entities, bind analytic inputs, and scope reasoning.

Within the 25 studies documenting a semantic model, the representation standard was also recorded. Figure 11 presents the distribution. Brick appeared in 9 studies and IFC in 7, together accounting for the majority of reported standards. NGSI-LD and RealEstateCore each appeared in 2 studies, while custom models accounted for 3. Standards occurring fewer than two times were grouped as "Other." Cross-tabulation indicated that semantic use in system logic was more common among Brick-based deployments (66.7%) than IFC-based deployments (28.6%).

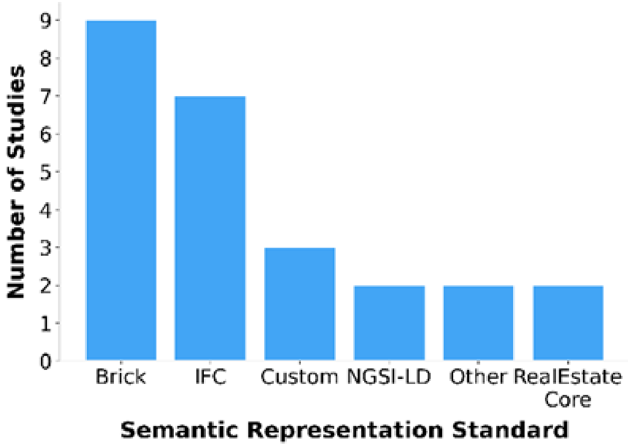

Figure 11: Semantic representation standard distribution across the corpus.

4.3 Actionability and delivery pathways

Two variables characterize the operational dimension of each deployment: actionability, capturing what the DT produces, and workflow integration, capturing how outputs reach operational systems. Actionability was classified across five levels: displaying measured or derived values without condition-triggered warnings (Monitor); flagging conditions via thresholds and producing warnings or notifications (Alert); identifying fault type, cause, or source (Diagnose); recommending actions, setpoints, or schedules (Prescribe); and initiating actions that change external state or create operational artifacts (Execute). Workflow integration was classified across three levels: outputs remaining within the DT interface (No Integration); delivery through generated reports, notifications, or pending approvals supporting operator action (Partial); and automatic external state change that is executed and evidenced (Full).

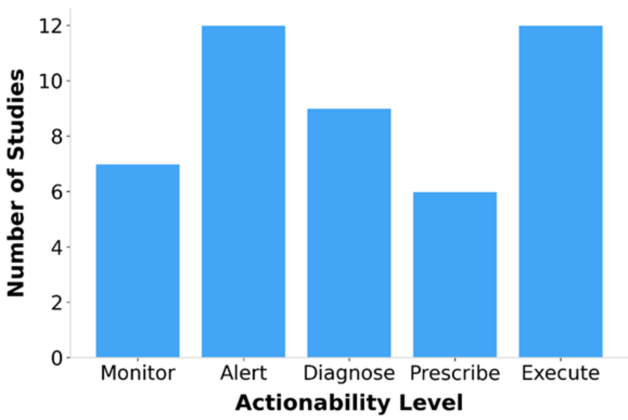

Figure 12: Actionability level distribution across the corpus.

Alert-level and execute-level outputs were the most frequently observed actionability categories (each n=12, 26.1%). Figure 12 presents the full actionability distribution. Outputs remained confined to the DT interface in the majority of studies (n=28, 60.9%), while automatic control actions (n=11, 23.9%) and operator-supported delivery pathways (n=7, 15.2%) were less common. Figure 13 presents the workflow integration distribution.

Among deployments demonstrating automatic control actions, Tan et al. (2022) linked a BIM model with video-derived pedestrian detection and ambient brightness sensing to automatically dispatch commands to corridor light controllers. Gray et al. (2020) integrated location, equipment context, and complaint submissions in an OWL knowledge graph, with SPARQL rule execution automatically modifying BAS setpoints for blinds, temperature, and air volume and converting hardware complaints into issue tickets. Li et al. (2023) facilitated remote control of HVAC, lighting, and shading through sensed states and UWB-based positioning in an office-building energy twin.

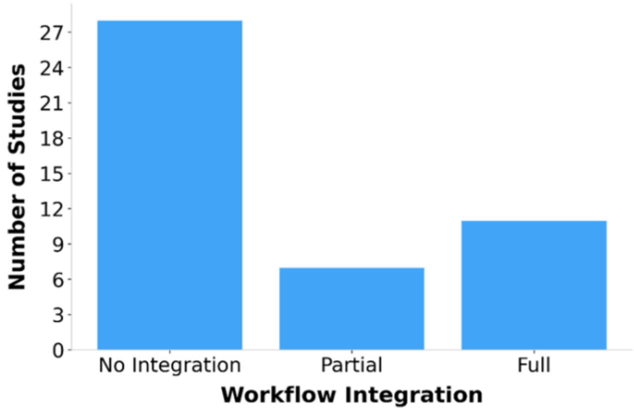

Figure 13: Workflow integration level distribution across the corpus.

Some deployments routed outputs through operator-supported pathways that extended results beyond the interface while retaining human oversight. Rattanatamrong et al. (2022) implemented an edge-cloud DT that emailed facility staff when occupancy exceeded specified thresholds. Merino et al. (2023) delivered diagnosed HVAC faults through email and a 3D visualization interface, with interventions dependent on operator action. The mixed-reality twin developed by Ma et al. (2025) enabled VAV setpoint adjustments from within a HoloLens environment through bidirectional BAS communication via MQTT, with each adjustment subject to manual verification.

Outputs classified as having no workflow integration remained confined to dashboards or 3D viewers. For instance, Biagini et al. (2024) connected IFC-based building models to IoT streams by manually associating BIM elements with device data feeds in a dashboard interface. Similarly, Pruvost et al. (2023) generated operational recommendations and air-filter wear alerts using ontology-driven inference and delivered them to a 3D web interface.

4.4 Operational outcomes across semantic tiers

Figures 14 and 15 present actionability and workflow integration cross-tabulated by semantic tier, with row percentages reported for each tier. Execute-level outputs occurred most frequently among deployments without a documented semantic model (42.9%), whereas alert-level outputs predominated among deployments with a documented semantic model (53.8%). Diagnostic outputs were most common when semantics were used in system logic (50.0%).

Deployments that achieved execute-level actionability without a documented semantic model relied on fixed ID mappings between system entities and their associated control points. This pattern is evident in Han et al. (2024), who connected a 3D model to lighting and air-conditioning controls through predefined endpoint mappings, and in Liu et al. (2020), who triggered physical alarms from SVM-based danger classification over LoRa sensor streams using terminal identifiers. Qian et al. (2024b) similarly activated air-circulation equipment based on occupancy and health-threshold rules. Several additional deployments reached execution through similar application-specific integrations, including Tan et al. (2022), Li et al. (2023), Liu et al. (2025), Englezos et al. (2022), and Abtahi et al. (2025).

Among deployments with a documented semantic model, alert-level outputs were most common, typically generated through threshold-band monitoring, statistical anomaly detection, or ML-based classification. For example, Roda-Sanchez et al. (2024) represented a smart-campus twin as an NGSI-LD knowledge graph using ML-based classification to generate alerts for anomalous occupancy values. Similarly, Chen et al. (2024) developed a scan-to-BIM ACMV twin that flagged conditions through both rule-based and ML-based components, with results imported into Tableau and BIM schedules. Arsiwala et al. (2023) likewise generated alerts when eCO2 concentrations exceeded specified thresholds. Although the underlying RealEstateCore ontology defined typed relationships, these were not used in the alerting logic. Wang et al. (2022) extended IFC with SensorML-derived metadata and defined environmental entity types, with warnings for elevated dust levels generated through regular-expression parsing and inverse distance weighting (IDW) spatial interpolation. Across these deployments, the underlying models primarily supported data organization, while alerts were generated through thresholds, anomaly detection, or classification applied directly to sensor data.

Figure 14: Actionability by semantic tier.

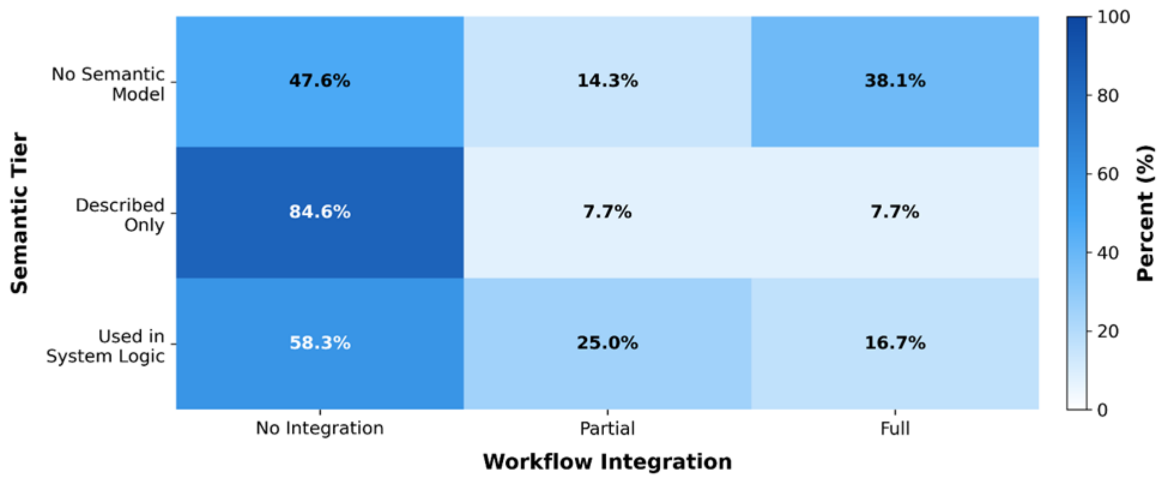

Figure 15: Workflow integration by semantic tier.

DT deployments demonstrating semantics in system logic most frequently produced diagnostic outputs. Xie et al. (2023) developed a Brick-based FDD workflow that semantically bound live sensor streams to typed equipment using fault tags and invoked diagnostics through those bindings. The cognitive DT developed by El Mokhtari et al. (2022) used a BOT/Brick-inspired ontology to support queries over asset, location, and point relationships, with CNN-based work-order classification and anomaly detection retrieving associated time-series data through semantic queries. Ni et al. (2025) developed a Brick-extended conservation twin in which SPARQL queries connected room and sensor entities to support diagnostic interpretation of monitored conditions.

Figure 15 presents workflow integration by semantic tier. Closed-loop actuation occurred most frequently among deployments without a documented semantic model (38.1%), whereas deployments with a documented semantic model were concentrated at the interface level (84.6%). Deployments in which semantics were used in system logic also remained primarily at the interface level (58.3%), although outputs delivered to operators were more common in this tier (25.0%) than among deployments with a documented semantic model alone. Among deployments combining structured metadata with automatic control actions, Gray et al. (2020) employed SPARQL rules that traversed spatial and equipment associations to scope complaint-relevant BAS setpoint modifications. ElArwady et al. (2024) iterated through defined IFC entity types (IfcSensor, IfcSpace, IfcUnitaryEquipment), extracting tag-name properties from a custom property set to create database records that associated real-time IoT data with BIM model elements, supporting remote AC setpoint control via decoded IR signals.

4.5 Analytical approaches and user interfaces

The algorithm family was treated as a multi-label variable, since individual studies often used more than one analytical approach. Figure 16 presents the distribution. Statistical methods were the most frequently reported family (n=25, 54.3%), followed by deep learning (n=20, 43.5%) and classical machine learning (n=14, 30.4%). Simulation, rule-based methods, and optimization were reported less frequently.

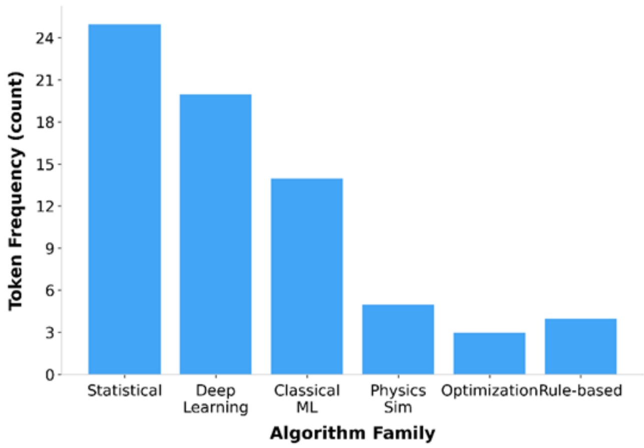

Figure 16: Analytic method family distribution across the corpus (multi-label).

These approaches were deployed across operational contexts. Xie et al. (2020) applied statistical methods including moving average, CUSUM, and binary-segmentation change-point detection to BMS and environmental signals, combining them with fault-tree reasoning to link temperature anomalies to candidate failed assets. Tan et al. (2022) applied deep learning to occupancy-responsive lighting using YOLOv4 pedestrian detection, while Liang et al. (2025) developed a Transformer-based autoencoder for virtual sensing with a mean absolute percentage error of approximately 2%. Peng et al. (2020) combined k-means anomaly detection with LSTM-based fault prediction and extended this workflow to mobile work-order assignment. Kim et al. (2023) trained a support vector regression model for energy and carbon prediction on a smart campus. Hybrid approaches were also prominent: Hosamo et al. (2023a) integrated APAR expert rules with a Bayesian network and classifiers including ANN, SVM, and decision trees, achieving 97% accuracy, while Hosamo et al. (2023b) combined Bayesian network and APAR methods with XGBoost classification and reported a 35% reduction in fault-resolution time. Hosamo et al. (2023c) paired an ANN-based energy and comfort predictor with a multi-objective genetic algorithm to identify setpoint trade-offs. Desogus et al. (2023) used rule-based computation to calculate PMV/PPD indices in accordance with ISO 7730 through Dynamo scripting mapped to BIM elements, while Englezos et al. (2022) employed EnergyPlus co-simulation for thermal validation and battery self-consumption coordination.

User-facing delivery modalities were also recorded as a multi-label variable. Figure 17 presents the distribution.

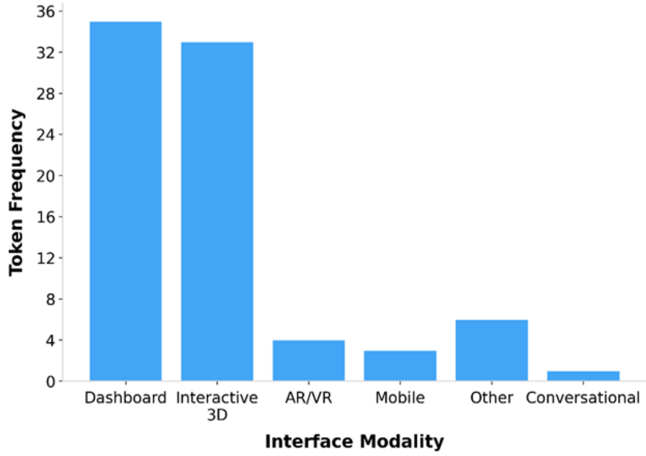

Figure 17: Interface modality distribution across the corpus (multi-label).

Dashboards were the most common modality (n=35, 76.1%), closely followed by interactive 3D environments (n=33, 71.7%). AR/VR interfaces appeared in four studies, mobile interfaces in three, and conversational interfaces in one. Ma et al. (2025) demonstrated mixed-reality interaction using a HoloLens-based interface that supports on-site inspection and setpoint adjustment. Peng et al. (2020) extended delivery to mobile interfaces supporting task assignment alongside 3D visual management. Conversational interfaces represent an emerging modality in the corpus, with Yoon et al. (2025) providing an early demonstration of an ontology-enabled tool-calling agent for HVAC fault detection and reporting.

5. DISCUSSION

Section 4 indicates that the reviewed literature has demonstrated meaningful components of operational DT capability, although these components are seldom integrated into systems that support maintainable and extensible deployment across buildings over time. The main implication is not simply a lack of functional demonstration of analytic capabilities, but the limited development of the underlying knowledge infrastructure needed to connect these components into durable operational systems at scale.

5.1 Field maturity and deployment trajectories

The reviewed studies describe a field that has not yet consolidated around a repeatable model of operational digital twin deployment. Monitoring, diagnostics, visualization, control, and limited workflow delivery have all been demonstrated, but usually as locally assembled capabilities rather than as components of a transferable operational architecture. Some deployments pursued immediate functional outcomes through tightly configured integrations among known equipment, known points, and predefined actions within a bounded setting. Other systems invested more heavily in ontologies and standards-aligned semantic structure, yet stop short of leveraging it to orchestrate workflows or system responses at scale. This pattern is consistent with an early stage of deployment maturity in which metadata structuring and operational functionality have both been demonstrated, but rarely in forms that reduce per-building reconfiguration effort, accommodate heterogeneous equipment inventories and naming conventions, or demonstrate sustained operation across multiple buildings over extended periods.

5.2 Ontologies as operational infrastructure

The association between semantic tier and diagnostic capability documented in Section 4.4 reflects what fault diagnosis demands as a reasoning task. Generating an alert may require only that a sensor value be linked to an expected range, a rule, or another detection criterion. Diagnosis requires situating a detected deviation within the surrounding system context, relating the anomaly to the equipment involved, the spaces or downstream components affected, and the system relationships that help narrow plausible causes. That reasoning process depends on relational context that is not encoded when individual points are mapped only to static thresholds and fixed identifiers. The reviewed deployments that achieved diagnostic depth did so by making system relationships available to the analytic process at runtime. Xie et al. (2023) used semantic annotations to determine which sensors were relevant to a given diagnostic routine, El Mokhtari et al. (2022) used ontology-driven queries to assemble contextually related assets around detected complaints, and Merino et al. (2023) used traversal of equipment hierarchies to delimit fault investigation to the relevant system boundary, with diagnosed results delivered beyond the visualization layer. Ni et al. (2025) pointed in the same direction in a monitoring-oriented setting, where semantic contextual structure improved interpretation by linking environmental readings to spaces and building fabric through structured queries. In each case, the distinguishing factor was not the analytic technique applied to the data, but the availability of structured system context that allows the analytical process to move from signal-level detection to system-level interpretation.

5.3 Accessible interfaces and agent-enabled workflows

Dashboards and interactive three-dimensional environments were the dominant delivery modalities in the reviewed corpus, a pattern that follows from the geometric origins of most operational DT platforms. Deployments grounded in BIM or laser-scan geometry produce spatial data structures whose most direct interface expression is three-dimensional visualization, and many of these systems inherit the interface conventions of the design and modeling environments from which those representations originate. This helps account for the prevalence of visual dashboards and 3D viewers in the literature, while also clarifying why operational accessibility remains uneven. Realizing the benefits of these platforms often still requires familiarity with BIM-like model structures, specialized viewers, or specialized interface conventions. For facility organizations operating under aging infrastructure, workforce constraints, and large multi-building portfolios (Shohet et al., 2025; Pampana et al., 2022), this creates a practical mismatch between the building intelligence available in the platform and the expertise required to extract value from it. Recent labor-market analysis finds that actual AI deployment across professional domains, including architecture, engineering, construction, and management, remains a small fraction of what is theoretically feasible (Massenkoff & McCrory, 2026), pointing to substantial unrealized potential in settings where access to building intelligence requires specialized interfaces. Conversational interfaces offer a complementary access layer by allowing personnel to request building context, asset information, diagnostic summaries, or workflow outputs through natural-language interaction, without requiring the same degree of fluency in model navigation or platform-specific conventions. Prior work has shown the promise of conversational retrieval for building information access (Chen & Tsai, 2021), while mobile and mixed-reality research underscores the importance of field-facing delivery for inspection, verification, maintenance documentation, and coordination between site and office personnel (El Ammari & Hammad, 2019). Yet these modalities remained peripheral in the reviewed literature, appearing more often as bounded interface experiments than as operational delivery channels tested across multi-building service environments.

Yoon et al. (2025) provide an early indication of how accessible interaction may intersect with structured digital twin infrastructure. Their framework used Brick to structure building metadata and expose domain-specific applications for fault detection, energy management, and virtual sensing as callable tools for an LLM agent. Its broader relevance lies in the workflow pattern it demonstrates: a user request can be translated into semantically grounded tool selection and executed against building-specific context to produce diagnostics or reports. The demonstration, however, remained deliberately bounded. It operated on preprocessed historical data, within a single building and a single HVAC scope, using preconfigured applications and agent behavior constrained by engineer-defined task parameters. These constraints clarify both the contribution of the study and the limits of what has so far been demonstrated. Extending this pattern into broader operational settings would require broader asset representation, richer cross-system relationships, and more reliable translation between informal operator language and formal building semantics. Figure 18 depicts a tool-enabled interaction flow in which an LLM agent translates operator requests into calls to DT data, context, diagnostic, and action services.

As interaction extends from information access to workflow execution, requirements for oversight become more important. Retrieving asset status or requesting a diagnostic summary poses a different kind of risk from changing a setpoint, creating a work order, or altering a control schedule. Actions that affect physical conditions or operational records thus require clearer constraints than those that only return information. In practice, this points to the need for scoped access, traceable records of tool use, and approval paths for higher-impact actions. Tool-oriented interface frameworks discussed in Section 2.5, including the Model Context Protocol, are relevant here because they provide structured ways to describe available tools, govern how they are invoked, and trace their use. Early peer-reviewed work in this direction includes EnergyPlus MCP integration (Li et al., 2025), and industry developments such as MCP-based conversational access to BIM data (Martins et al., 2025) suggest growing platform-level interest. However, peer-reviewed evidence demonstrating these patterns in live, multi-system building operations remains limited, and their practical benefits at scale have yet to be validated.

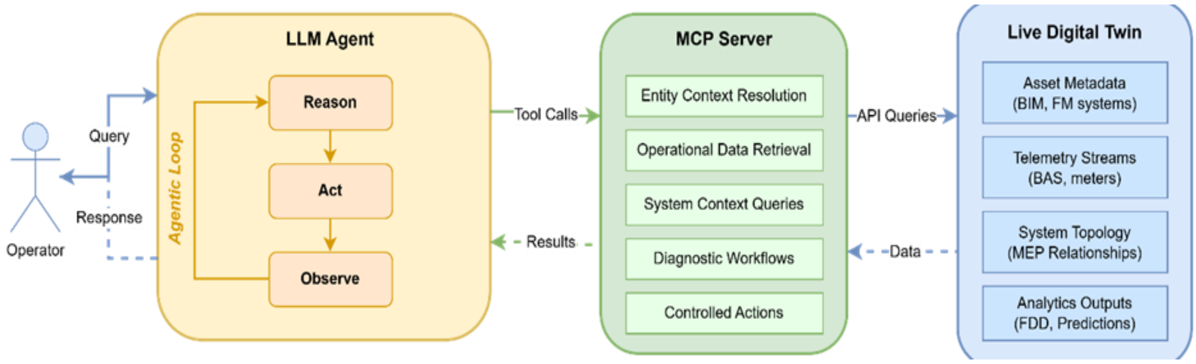

Figure 18: Conceptual agent-based workflow using MCP with live DT infrastructure.

5.4 Organizational and lifecycle barriers to scalable deployment

The barriers to scalable deployment identified across the reviewed studies extend beyond platform design and analytic method. BIM, BAS, and CMMS are typically procured by different teams, at different project stages, through separate contracts and vendor relationships. The fragmentation visible in operational DT deployments therefore reflects more than technical incompatibility between systems. It also reflects organizational separation between the groups responsible for design models, control systems, and maintenance records, often with limited coordination around data governance, metadata quality, and cross-system information requirements. The ontology-based approaches discussed in Section 5.2 provide a technical route toward reducing per-building reconfiguration by allowing applications to reference entity types, system relationships, and contextual descriptors rather than site-specific identifiers alone. Their value across multi-building portfolios, however, depends on whether those metadata foundations are established, validated, and maintained across the procurement and organizational boundaries through which building systems are delivered and operated.

The stage in the building lifecycle at which semantic metadata is first structured is therefore a practical deployment consideration. A substantial portion of the metadata needed for scalable operation is available during design and construction, when equipment classifications, intended system relationships, and naming conventions are still explicit in project documentation and the parties who understand system intent remain engaged. Deferring metadata structuring until operations increases the cost of reconstruction, as project teams disperse, informal knowledge of design rationale is lost, and the building itself changes through equipment replacement, space reconfiguration, and control-system modification. Benfer and Müller (2024) underscore this difficulty by showing that time-series data alone is often insufficient to recover point meaning and system structure, with current inference methods proving more effective for entity classification than for reconstructing relationships or operational semantics. Structured metadata is therefore more appropriately treated as a handover deliverable than as something to be reconstructed later once the information environment has become more fragmented and more costly to interpret. Automated inference methods are better understood in this context as complementary tools for legacy buildings where construction-phase records are unavailable, and for validating delivered metadata against as-built conditions, rather than as the primary mechanism through which scalable knowledge infrastructure is created.

A further limitation of the current literature is the relative absence of evidence on sustained operational use. Most reviewed deployments remain pilots or bounded demonstrations, and comparatively little is reported about what happens after initial implementation. Questions of model upkeep, adaptation to changes in control systems, integration with existing maintenance and operational workflows, and the training needed for staff to use new interfaces and tools remain only lightly addressed. These issues are central to practical deployment, as they shape whether technically promising systems can be integrated into routine operations rather than remaining research-stage implementations. Stronger evidence in this area will require studies that examine not only initial deployment, but also how these systems are maintained, updated, and used over time within real organizational settings.

Taken together, these findings suggest that scalable operational DT deployment depends on machine-readable building context that is structured early in the building lifecycle, governed consistently across organizational boundaries, and maintained as the building and its systems evolve. Semantic infrastructure is therefore better understood as a lifecycle information requirement than as an optional enrichment layer. Investment in that infrastructure simultaneously serves the analytic portability discussed in Section 5.2 and the accessible, tool-mediated interaction examined in Section 5.3, as both depend on the same foundational condition: a structured representation of what building entities are, what they measure or control, and how they relate to one another.

6. CONCLUSION

This study conducted a systematic review of 46 operational DTs in the built environment to examine how building context is represented in documented implementations, how analytic outputs are connected to operational workflows, and what these patterns imply for scalable deployment. The reviewed deployments show that core operational functions such as monitoring, diagnostics, control, and visualization have been realized, though predominantly through configurations assembled around site-specific identifiers and locally defined mappings. Formal semantic models were documented in roughly half of the reviewed DTs, and fewer than a third used semantic constructs to actively shape analytic or decision processes. Where semantic structure informed system behavior more directly, the reviewed DTs more often supported diagnostic interpretation across equipment and spatial relationships. By contrast, execute-level outputs were most frequent among studies with no documented semantic model (42.9%), where execution generally relied on predefined mappings between system entities and control points within tightly configured, site-specific integrations with limited evidence of transferability across sites. Across the corpus, outputs also remained largely confined to dashboards and 3D environments, with relatively little evidence of routine integration into maintenance systems, control workflows, or other governed external processes.

These patterns describe a field progressing from isolated capability demonstrations toward the architectural challenge of connecting those capabilities into durable, transferable systems that can be maintained and extended across buildings over time. A central factor in that transition is the availability of machine-readable building context, specifically whether the entities, relationships, and functional roles within a facility are represented in forms that analytics and applications can discover and use without manual reconfiguration. This requirement becomes more consequential as enterprise and building-domain platforms adopt agent-oriented interaction patterns in which language models retrieve context, invoke diagnostics, and initiate actions through structured interfaces. Conversational and agent-enabled access may lower expertise barriers for facility personnel, but only when underlying entities, relationships, and action pathways are represented with sufficient structure to support safe retrieval, diagnostic reasoning, and governed execution. The review therefore frames AI readiness in building operations not as a capability to be added at the application layer, but as a condition of the underlying information infrastructure.

The strongest practical implication is that semantic metadata should be treated as a lifecycle information requirement, structured during design and construction, delivered at handover, and maintained as buildings evolve, instead of being reconstructed from fragmented operational records once project teams have dispersed. Future research should move beyond bounded pilot demonstrations toward longitudinal and multi-building studies that address several questions the current literature has not yet resolved. These include (i) whether semantic models remain accurate and maintainable as buildings change, (ii) whether metadata quality can be governed consistently across the organizational boundaries through which building systems are procured and operated, (iii) whether diagnostic and analytic outputs can be reliably delivered into channels such as work order systems and control sequences, and (iv) whether adequate approval and traceability mechanisms can be established for actions that affect physical building conditions.

This study is subject to several limitations. The search was conducted across four databases, and relevant studies using different terminology may not have been captured. The review is limited to peer-reviewed journal articles and full conference papers. As a result, DT implementations reported through industry, commercial, or proprietary sources may be underrepresented. The reviewed literature is also concentrated in indoor environmental quality and energy applications, which may limit the extent to which the observed patterns generalize to other operational domains. Additionally, title and abstract screening were performed independently by two reviewers, with disagreements resolved through discussion, though it is possible that studies whose titles and abstracts did not clearly indicate operational deployment were not identified during this stage despite containing relevant content. The semantic tier classification involved interpretive judgment during extraction, and although the coding schema was iteratively refined, some degree of subjectivity in categorization remains inherent to this type of analysis. The strict operational scope excluded conceptual frameworks and early-stage prototypes that may mature into operational deployments. The review is also limited by its search period through August 2025, and findings reflect the state of the published literature at that time. Advancing scalable and AI-ready deployment at portfolio scale will require the field to consolidate the semantic and organizational foundations identified in this review, alongside the analytic capabilities that have received primary attention to date.

REFERENCES

Abdelrahman, M., Macatulad, E., Lei, B., Quintana, M., Miller, C., & Biljecki, F. (2025). What is a digital twin anyway? Deriving the definition for the built environment from over 15,000 scientific publications. Building and Environment, 274, 112748. https://doi.org/10.1016/j.buildenv.2025.112748

Abtahi, M., Rueda, L., Delcroix, B., & Athienitis, A. (2025). Semantic digital twinning for cost-optimal HVAC operation: Real-time application to a house with smart thermostats and PV/battery under a time-of-use tariff. Energy and Buildings, 343, 115938. https://doi.org/10.1016/j.enbuild.2025.115938

Anthropic. (2024). Introducing the model context protocol. https://www.anthropic.com/news/model-context-protocol

Arsiwala, A., Elghaish, F., & Zoher, M. (2023). Digital twin with machine learning for predictive monitoring of CO2 equivalent from existing buildings. Energy and Buildings, 284, 112851. https://doi.org/10.1016/j.enbuild.2023.112851

ASHRAE. (2018, March). ASHRAE BACnet committee works with other organizations on new standard. eSociety. Retrieved January 18, 2026, from https://www.ashrae.org/news/esociety/ashrae-bacnet-committee-works-with-other-organizations-on-new-standard

Balaji, B., Bhattacharya, A., Fierro, G., Gao, J., Gluck, J., Hong, D., Johansen, A., Koh, J., Ploennigs, J., Agarwal, Y., Bergés, M., Culler, D., Gupta, R. K., Kjærgaard, M. B., Srivastava, M., & Whitehouse, K. (2016). Brick: Towards a unified metadata schema for buildings. In Proceedings of the 3rd ACM International Conference on Systems for Energy-Efficient Built Environments (BuildSys '16) (pp. 41–50). ACM. https://doi.org/10.1145/2993422.2993577

Balaji, B., Bhattacharya, A., Fierro, G., Gao, J., Gluck, N., Hong, D., Johansen, A., Koh, J., Ploennigs, J., Agarwal, Y., Bergés, M., Culler, D., Gupta, R. K., Kjærgaard, M. B., Srivastava, M., & Whitehouse, K. (2018). Brick: Metadata schema for portable smart building applications. Applied Energy, 226, 1273–1292. https://doi.org/10.1016/j.apenergy.2018.02.091

Benfer, R., & Müller, J. (2024). Semantic digital twin creation of building systems through time series based metadata inference—A review. Energy and Buildings, 321, 114637. https://doi.org/10.1016/j.enbuild.2024.114637

Biagini, C., Bongini, A., & Marzi, L. (2024). From BIM to digital twin: IoT data integration in asset management platform. Journal of Information Technology in Construction, 29, 1103–1127. https://doi.org/10.36680/j.itcon.2024.049

Bjørnskov, J., Badhwar, A., Singh, D. S., Sehgal, M., Åkesson, R., & Jradi, M. (2025). Development and demonstration of a digital twin platform leveraging ontologies and data-driven simulation models. Journal of Building Performance Simulation, 1–13. https://doi.org/10.1080/19401493.2025.2504005

Boje, C., Guerriero, A., Kubicki, S., & Rezgui, Y. (2020). Towards a semantic construction digital twin: Directions for future research. Automation in Construction, 114, 103179. https://doi.org/10.1016/j.autcon.2020.103179

Brick Consortium. (n.d.). Brick schema. Retrieved January 18, 2026, from https://brickschema.org/

Chamari, L., Walker, S., Petrova, E., & Pauwels, P. (2025). Towards portable model predictive control-based applications for demand side management in buildings. Energy and Buildings, 347, 116257. https://doi.org/10.1016/j.enbuild.2025.116257

Chen, K.-L., & Tsai, M.-H. (2021). Conversation-based information delivery method for facility management. Sensors, 21(14), 4771. https://doi.org/10.3390/s21144771

Chen, X., Pan, Y., Gan, V. J. L., & Yan, K. (2024). 3D reconstruction of semantic-rich digital twins for ACMV monitoring and anomaly detection via scan-to-BIM and time-series data integration. Developments in the Built Environment, 19, 100503. https://doi.org/10.1016/j.dibe.2024.100503

Correia, J. B., Abel, M., & Becker, K. (2023). Data management in digital twins: A systematic literature review. Knowledge and Information Systems, 65, 3165–3196. https://doi.org/10.1007/s10115-023-01870-1

DeepSeek-AI. (2025). DeepSeek-R1: Incentivizing reasoning capability in LLMs via reinforcement learning. arXiv:2501.12948. https://arxiv.org/abs/2501.12948

Deng, M., Menassa, C. C., & Kamat, V. R. (2021). From BIM to digital twins: A systematic review of the evolution of intelligent building representations in the AEC-FM industry. Journal of Information Technology in Construction, 26, 58–83. https://doi.org/10.36680/J.ITCON.2021.005

Desogus, G., Frau, C., Quaquero, E., & Rubiu, G. (2023). From building information model to digital twin: A framework for building thermal comfort monitoring, visualizing, and assessment. Buildings, 13(8), 1971. https://doi.org/10.3390/buildings13081971

Eastman, C., Teicholz, P., Sacks, R., & Liston, K. (2011). BIM handbook: A guide to building information modeling for owners, managers, designers, engineers, and contractors (2nd ed.). John Wiley & Sons.

El Ammari, K., & Hammad, A. (2019). Remote interactive collaboration in facilities management using BIM-based mixed reality. Automation in Construction, 107, 102940. https://doi.org/10.1016/j.autcon.2019.102940

El Mokhtari, K., Panushev, I., & McArthur, J. J. (2022). Development of a cognitive digital twin for building management and operations. Frontiers in Built Environment, 8, 856873. https://doi.org/10.3389/fbuil.2022.856873

ElArwady, Z., Kandil, A., Afiffy, M., & Marzouk, M. (2024). Modeling indoor thermal comfort in buildings using digital twin and machine learning. Developments in the Built Environment, 19, 100480. https://doi.org/10.1016/j.dibe.2024.100480

Elias, C., & Issa, R. R. A. (2025). AI-driven digital twins for predictive maintenance. In Proceedings of the 2025 European Conference on Computing in Construction (EC³), co-located with the 42nd International CIB W78 Conference on IT in Construction, Porto, Portugal, July 14–17, 2025. https://doi.org/10.35490/EC3.2025.369

Englezos, D., Hadjidemetriou, L., Papadopoulos, P., Timotheou, S., Polycarpou, M., & Panayiotou, C. (2022). A digital twin architecture for smart buildings. In 2022 IEEE International Smart Cities Conference (ISC2) (pp. 1–7). IEEE.

Fierro, G., Saha, A., Shapinsky, T., Steen, M., & Eslinger, H. (2022). Application-driven creation of building metadata models with semantic sufficiency. In Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation (BuildSys '22) (pp. 228–237). ACM. https://doi.org/10.1145/3563357.3564083

Gao, Q., Chu, Y., Peng, Z., Jin, Y., Ji, X., Ji, S., Ping, S., Xie, X., Zhu, X., & Yue, Y. (2024). Smart building management system based on digital twin: A case study on real-time environmental monitoring and thermal comfort prediction. In 2024 IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA) (pp. 2157–2163). IEEE.

Gray, F. M., Dibowski, H., Gall, J., & Braun, S. (2020). Occupant feedback and context awareness: On the application of building information modeling and semantic technologies for improved complaint management in commercial buildings. In 2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA) (pp. 101–108). IEEE.

Grieves, M. (2014). Digital twin: Manufacturing excellence through virtual factory replication [White paper]. Michael W. Grieves, LLC.

Hakimi, O., Liu, H., & Abudayyeh, O. (2024). Digital twin for facility management—Bibliometric analysis. Frontiers of Engineering Management, 11(1), 32–49. https://doi.org/10.1007/s42524-023-0254-4

Han, F., Du, F., Jiao, S., & Zou, K. (2024). Predictive analysis of a building's power consumption based on digital twin platforms. Energies, 17(15), 3692. https://doi.org/10.3390/en17153692

Hosamo, H. H., Nielsen, H. K., Alnmr, A. N., Svennevig, P. R., & Svidt, K. (2022a). A review of the digital twin technology for fault detection in buildings. Frontiers in Built Environment, 8, 1013196. https://doi.org/10.3389/fbuil.2022.1013196

Hosamo, H. H., Nielsen, H. K., Kraniotis, D., Svennevig, P. R., & Svidt, K. (2023a). Digital twin framework for automated fault source detection and prediction for comfort performance evaluation of existing non-residential Norwegian buildings. Energy and Buildings, 281, 112732. https://doi.org/10.1016/j.enbuild.2022.112732

Hosamo, H. H., Nielsen, H. K., Kraniotis, D., Svennevig, P. R., & Svidt, K. (2023b). Improving building occupant comfort through a digital twin approach: A Bayesian network model and predictive maintenance method. Energy and Buildings, 288, 112992. https://doi.org/10.1016/j.enbuild.2023.112992

Hosamo, H., Hosamo, M. H., Nielsen, H. K., Svennevig, P. R., & Svidt, K. (2023c). Digital twin of HVAC system (HVACDT) for multiobjective optimization of energy consumption and thermal comfort based on BIM framework with ANN-MOGA. Advances in Building Energy Research, 17(2), 125–171. https://doi.org/10.1080/17512549.2022.2136240

Hu, X., & Assaad, R. H. (2024). A BIM-enabled digital twin framework for real-time indoor environment monitoring and visualization by integrating autonomous robotics, LiDAR-based 3D mobile mapping, IoT sensing, and indoor positioning technologies. Journal of Building Engineering, 86, 108901. https://doi.org/10.1016/j.jobe.2024.108901

Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterising the digital twin: A systematic literature review. CIRP Journal of Manufacturing Science and Technology, 29, 36–52. https://doi.org/10.1016/j.cirpj.2020.02.002

Karabulut, E., Pileggi, S. F., Groth, P., & Degeler, V. (2024). Ontologies in digital twins: A systematic literature review. Future Generation Computer Systems, 153, 442–456. https://doi.org/10.1016/j.future.2023.12.013

Kim, J. B., Wang, F., Khanna, S., Balakrishnan, B., Uddin, M., Aman, J., & Thipparthi, V. V. R. (2023). Digital twin framework for smart campus to reduce greenhouse gas emission. In 2023 IEEE Smart World Congress (SWC) (pp. 1–8). IEEE.

Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering (Tech. Rep. EBSE-2007-01, Version 2.3). Keele University & University of Durham. Retrieved January 18, 2026, from https://legacyfileshare.elsevier.com/promis_misc/525444systematicreviewsguide.pdf

Kreuzer, T., Papapetrou, P., & Zdravkovic, J. (2024). Artificial intelligence in digital twins—A systematic literature review. Data & Knowledge Engineering, 151, 102304. https://doi.org/10.1016/j.datak.2024.102304

Kritzinger, W., Karner, M., Traar, G., Henjes, J., & Sihn, W. (2018). Digital twin in manufacturing: A categorical literature review and classification. IFAC-PapersOnLine, 51(11), 1016–1022. https://doi.org/10.1016/j.ifacol.2018.08.474

Laakso, M., & Kiviniemi, A. (2012). The IFC standard: A review of history, development, and standardization. Journal of Information Technology in Construction, 17, 134–161.

Li, C., Lu, P., Zhu, W., Zhu, H., & Zhang, X. (2023). Intelligent monitoring platform and application for building energy using information based on digital twin. Energies, 16(19), 6839. https://doi.org/10.3390/en16196839

Li, H., Xu, Y., & Hong, T. (2025). EnergyPlus-MCP: A model-context-protocol server for AI-driven building energy modeling. SoftwareX, 32, 102367. https://doi.org/10.1016/j.softx.2025.102367

Liang, Z., Singh, J., & Jin, Y. (2025). Data-driven smart sensing and real-time high-fidelity digital twin for building management. IEEE Journal of Selected Areas in Sensors, 2, 232–246. https://doi.org/10.1109/JSAS.2025.3588827

Lilis, G. N., Wang, M., Katsigarakis, K., Mavrokapnidis, D., Korolija, I., & Rovas, D. (2025). BIM-based semantic enrichment and knowledge graph generation via geometric relation checking. Automation in Construction, 173, Article 106081. https://doi.org/10.1016/j.autcon.2025.106081

Liu, Z., Zhang, A., & Wang, W. (2020). A framework for an indoor safety management system based on digital twin. Sensors, 20(20), 5771. https://doi.org/10.3390/s20205771

Liu, Z., Li, M., & Ji, W. (2025). Development and application of a digital twin model for net zero energy building operation and maintenance utilizing BIM-IoT integration. Energy and Buildings, 328, 115170. https://doi.org/10.1016/j.enbuild.2024.115170

Lu, Q., Parlikad, A. K., Woodall, P., Don Ranasinghe, G., Xie, X., Liang, Z., Konstantinou, E., Heaton, J., & Schooling, J. (2020a). Developing a digital twin at building and city levels: Case study of West Cambridge Campus. Journal of Management in Engineering, 36(3), 05020004. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000763

Ma, T., Xiao, F., Zhang, C., Zhang, J., Zhang, H., Xu, K., & Luo, X. (2025). Digital twin for 3D interactive building operations: Integrating BIM, IoT-enabled building automation systems, AI, and mixed reality. Automation in Construction, 176, 106277. https://doi.org/10.1016/j.autcon.2025.106277

Martins, J., Broz, P., & Beale, M. (2025). Talk to your BIM: Exploring the AEC data model with MCP server + Claude. Autodesk Platform Services Blog. https://aps.autodesk.com/blog/talk-your-bim-exploring-aec-data-model-mcp-server-claude

Massenkoff, M., & McCrory, P. (2026). Labor market impacts of AI: A new measure and early evidence. Anthropic. https://www.anthropic.com/research/labor-market-impacts

Mavrokapnidis, D., Lilis, G. N., Katsigarakis, K., Korolija, I., & Rovas, D. (2023). Semi-automated extraction of HVAC system topology from imperfect building information models. DigiBuild Project.

Merino, J., Xie, X., Moretti, N., Chang, J. Y., & Parlikad, A. (2023). Data integration for digital twins in the built environment based on federated data models. Proceedings of the Institution of Civil Engineers – Smart Infrastructure and Construction, 176(4), 194–211. https://doi.org/10.1680/jsmic.23.00002

Ni, Z., Liu, Y., Karlsson, M., & Gong, S. (2022). Enabling preventive conservation of historic buildings through cloud-based digital twins: A case study in the City Theatre, Norrköping. IEEE Access, 10, 90924–90939. https://doi.org/10.1109/ACCESS.2022.3202181

Ni, Z., Hupkes, J., Eriksson, P., Leijonhufvud, G., Karlsson, M., & Gong, S. (2025). Parametric digital twins for preserving historic buildings: A case study at Löfstad Castle in Östergötland, Sweden. IEEE Access, 13, 3371–3389. https://doi.org/10.1109/ACCESS.2024.3525442

Open223. (2026). 223 overview: ASHRAE Standard 223P user documentation. Retrieved January 13, 2026, from https://docs.open223.info/explanation/223_overview.html

OpenAI. (2024). OpenAI o1 system card. arXiv:2412.16720. https://arxiv.org/abs/2412.16720

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—A web and mobile app for systematic reviews. Systematic Reviews, 5(1), 210. https://doi.org/10.1186/s13643-016-0384-4

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Pampana, A. K., Jeon, J., Yoon, S., Weidner, T. J., & Hastak, M. (2022). Data-driven analysis for facility management in higher education institution. Buildings, 12(12), 2094. https://doi.org/10.3390/buildings12122094

Pärn, E. A., Edwards, D. J., & Sing, M. C. P. (2017). The building information modelling trajectory in facilities management: A review. Automation in Construction, 75, 45–55. https://doi.org/10.1016/j.autcon.2016.12.003

Peng, Y., Zhang, M., Yu, F., Xu, J., & Gao, S. (2020). Digital twin hospital buildings: An exemplary case study through continuous lifecycle integration. Advances in Civil Engineering, 2020, 8846667. https://doi.org/10.1155/2020/8846667

Pishdad-Bozorgi, P., Gao, X., Eastman, C., & Self, A. P. (2018). Planning and developing facility management-enabled building information model (FM-enabled BIM). Automation in Construction, 87, 22–38. https://doi.org/10.1016/j.autcon.2017.12.004

Poggi, A., Lembo, D., Calvanese, D., De Giacomo, G., Lenzerini, M., & Rosati, R. (2008). Linking data to ontologies. In S. Spaccapietra (Ed.), Journal on data semantics X (pp. 133–173). Springer. https://doi.org/10.1007/978-3-540-77688-8_5

Pritoni, M., Paine, D., Fierro, G., Mosiman, C., Poplawski, M., Saha, A., Bender, J., & Granderson, J. (2021). Metadata schemas and ontologies for building energy applications: A critical review and use case analysis. Energies, 14(7), 2024. https://doi.org/10.3390/en14072024

Pritoni, M., Wetter, M., Paul, L., Prakash, A., Huang, W., Bushby, S., Delgoshaei, P., Poplawski, M., Saha, A., Fierro, G., Steen, M., Bender, J., & Ehrlich, P. (2024). Digital and interoperable: The future of building automation is on the horizon. What's in it for me? In Proceedings of the 2024 ACEEE Summer Study on Energy Efficiency in Buildings. American Council for an Energy-Efficient Economy. https://www.aceee.org/sites/default/files/proceedings/ssb24/assets/attachments/20240722163102218_468528b1-734f-45a6-9981-effee9f80658.pdf

Project Haystack. (2018a). Implementing Project Haystack: Applying Haystack tagging for a sample building [White paper]. Retrieved January 18, 2026, from https://project-haystack.org/file/28/Reference-Implementation--Applying-Haystack-Tagging-for-a-Sample-Building.pdf

Project Haystack. (2018b). ASHRAE's BACnet Committee, Project Haystack and Brick Schema collaborating to provide unified data semantic modeling solution. Retrieved January 18, 2026, from https://marketing.project-haystack.org/project-haystack-media/press-releases/ashraes-bacnet-committee-project-haystack-and-brick-schema-collaborating-to-provide-unified-data-semantic-modeling-solution

Project Haystack. (2025, May 13). Project Haystack unveils Haystack 5 with Xeto schema language and RDF integration at Haystack Connect 2025 [Press release]. Retrieved January 18, 2026, from https://marketing.project-haystack.org/project-haystack-media/press-releases/project-haystack-unveils-haystack-5-with-xeto-schema-language-and-rdf-integration-at-haystack-connect-2025

Pruvost, H., Forns-Samso, F., Gnepper, O., & Enge-Rosenblatt, O. (2023). Integrating energy system monitoring and maintenance services into a BIM-based digital twin. In IECON 2023—49th Annual Conference of the IEEE Industrial Electronics Society (pp. 1–6). IEEE.

Qian, Y., Leng, J., Zhou, K., & Liu, Y. (2024b). How to measure and control indoor air quality based on intelligent digital twin platforms: A case study in China. Building and Environment, 253, 111349. https://doi.org/10.1016/j.buildenv.2024.111349

Rashidi, A., Sarvari, H., Chan, D. W. M., Olawumi, T. O., & Edwards, D. J. (2024). A systematic taxonomic review of the application of BIM and digital twins technologies in the construction industry. Engineering, Construction and Architectural Management. Advance online publication. https://doi.org/10.1108/ECAM-02-2024-0239

Rattanatamrong, P., Srisawat, J., Boonchoo, T., & Haga, J. (2022). Development of a digital twin for smart building over edge-cloud continuum. In 2022 IEEE 8th World Forum on Internet of Things (WF-IoT) (pp. 1–6). IEEE.

Roda-Sanchez, L., Cirillo, F., Solmaz, G., Jacobs, T., Garrido-Hidalgo, C., Olivares, T., & Kovacs, E. (2024). Building a smart campus digital twin: System, analytics, and lessons learned from a real-world project. IEEE Internet of Things Journal, 11(3), 4614–4627. https://doi.org/10.1109/JIOT.2023.3300447

Schick, T., Dwivedi-Yu, J., Dessì, R., Raileanu, R., Lomeli, M., Zettlemoyer, L., Cancedda, N., & Scialom, T. (2023). Toolformer: Language models can teach themselves to use tools. arXiv:2302.04761. https://doi.org/10.48550/arXiv.2302.04761

Shahzad, M., Tah, J. H. M., Younas, M., & Almukhtar, A. (2025). Technologies and techniques in digital twins for real-time data visualisation in building maintenance: A state-of-the-art review. Journal of Infrastructure Intelligence and Resilience, 4(4), 100185. https://doi.org/10.1016/j.iintel.2025.100185

Shohet, I. M., Levi, S., Zeibak-Shini, R., & Shahin, F. (2025). Performance-based maintenance and operation of multi-campus critical infrastructure facilities using supply chain multi-choice goal programming. Applied Sciences, 15(20), 11161. https://doi.org/10.3390/app152011161

Tan, Y., Chen, P., Shou, W., & Sadick, A.-M. (2022). Digital twin-driven approach to improving energy efficiency of indoor lighting based on computer vision and dynamic BIM. Energy and Buildings, 270, 112271. https://doi.org/10.1016/j.enbuild.2022.112271

Wang, M., Lilis, G. N., Mavrokapnidis, D., Katsigarakis, K., Korolija, I., & Rovas, D. (2024). A knowledge graph-based framework to automate the generation of building energy models using geometric relation checking and HVAC topology establishment. Energy and Buildings, 325, Article 115035. https://doi.org/10.1016/j.enbuild.2024.115035

Wang, T., Gan, V. J. L., Hu, D., & Liu, H. (2022). Digital twin-enabled built environment sensing and monitoring through semantic enrichment of BIM with SensorML. Automation in Construction, 144, 104625. https://doi.org/10.1016/j.autcon.2022.104625

Xie, X., Merino, J., Moretti, N., Pauwels, P., Chang, J. Y., & Parlikad, A. (2023). Digital twin enabled fault detection and diagnosis process for building HVAC systems. Automation in Construction, 146, 104695. https://doi.org/10.1016/j.autcon.2022.104695

Xie, X., Lu, Q., Rodenas-Herraiz, D., Parlikad, A. K., & Schooling, J. M. (2020). Visualised inspection system for monitoring environmental anomalies during daily operation and maintenance. Engineering, Construction and Architectural Management, 27(8), 1835–1852. https://doi.org/10.1108/ECAM-11-2019-0640

Xu, C., Li, M., Yang, C., Jiang, X., Tang, L., Qi, Y., & Guo, J. (2024). Context graph. arXiv:2406.11160. https://doi.org/10.48550/arXiv.2406.11160

Yao, S., Zhao, J., Yu, D., Du, N., Shafran, I., Narasimhan, K., & Cao, Y. (2022). ReAct: Synergizing reasoning and acting in language models. arXiv:2210.03629. https://doi.org/10.48550/arXiv.2210.03629

Yoon, S., Song, J., & Li, J. (2025). Ontology-enabled AI agent-driven intelligent digital twins for building operations and maintenance. Journal of Building Engineering, 108, 112802. https://doi.org/10.1016/j.jobe.2025.112802

APPENDIX A: FINAL CORPUS (N=46)

Study	Title	Reference
1	Optimising Energy Performance of buildings through Digital Twins and Machine Learning: Lessons learnt and future directions	Renganayagalu et al., 2024
2	Digital Twin Framework for Smart Campus to Reduce Greenhouse Gas Emission	Kim et al., 2023
3	Developing a Digital Twin at Building and City Levels: Case Study of West Cambridge Campus	Lu et al., 2020a
4	Digital Twin-driven approach to improving energy efficiency of indoor lighting based on computer vision and dynamic BIM	Tan et al., 2022
5	Data-Driven Smart Sensing and Real-Time High-Fidelity Digital Twin for Building Management	Liang et al., 2025
6	Parametric Digital Twins for Preserving Historic Buildings: A Case Study at Löfstad Castle in Östergötland, Sweden	Ni et al., 2025
7	Building a Smart Campus Digital Twin: System, Analytics, and Lessons Learned From a Real-World Project	Roda-Sanchez et al., 2024
8	3D reconstruction of semantic-rich digital twins for ACMV monitoring and anomaly detection via scan-to-BIM and time-series data integration	Chen et al., 2024
9	Digital Twin-Enabled Building Information Modeling–Internet of Things (BIM-IoT) Framework for Optimizing Indoor Thermal Comfort Using Machine Learning	Iqbal & Mirzabeigi, 2025
10	Development and demonstration of a digital twin platform leveraging ontologies and data-driven simulation models	Bjørnskov et al., 2025
11	Data Integration for Digital Twins in the built environment based on federated data models	Merino et al., 2023
12	Enabling Preventive Conservation of Historic Buildings Through Cloud-Based Digital Twins: A Case Study in the City Theatre, Norrköping	Ni et al., 2022
13	Development and application of a digital twin model for Net zero energy building operation and maintenance utilizing BIM-IoT integration	Liu et al., 2025
14	Ontology-enabled AI agent-driven intelligent digital twins for building operations and maintenance	Yoon et al., 2025
15	Improving building occupant comfort through a digital twin approach: A Bayesian network model and predictive maintenance method	Hosamo et al., 2023b
16	Energy-data-related digital twin for office building and data centre complex	Kannari et al., 2022
17	A Digital Twin Architecture for Smart Buildings	Englezos et al., 2022
18	A Digital Twin predictive maintenance framework of air handling units based on automatic fault detection and diagnostics	Hosamo et al., 2022b
19	From Building Information Model to Digital Twin: A Framework for Building Thermal Comfort Monitoring, Visualizing, and Assessment	Desogus et al., 2023
20	Digital twin for 3D interactive building operations: Integrating BIM, IoT-enabled building automation systems, AI, and mixed reality	Ma et al., 2025
21	DT-BEMS: Digital twin-enabled building energy management system for information fusion and energy efficiency	Hwang et al., 2025

22	FROM BIM TO DIGITAL TWIN. IOT DATA INTEGRATION IN ASSET MANAGEMENT PLATFORM	Biagini et al., 2024
23	A Digital Twin-Enabled Anomaly Detection System for Asset Monitoring in Operation and Maintenance	Lu et al., 2020b
24	Visualised inspection system for monitoring environmental anomalies during daily operation and maintenance	Xie et al., 2020
25	Towards Digital Twin Data Center Using Building Information Modeling and Real-Time Data Sensing	Pinmas et al., 2022
26	Digital Twin framework for automated fault source detection and prediction for comfort performance evaluation of existing non-residential Norwegian buildings	Hosamo et al., 2023a
27	Digital twin with Machine learning for predictive monitoring of CO2 equivalent from existing buildings	Arsiwala et al., 2023
28	Development of a Digital Twin for Smart Building over Edge-Cloud Continuum	Rattanatamrong et al., 2022
29	Evaluating carbon emissions from the operation of historic dwellings in cities based on an intelligent management platform	Qian et al., 2024a
30	Modeling indoor thermal comfort in buildings using digital twin and machine learning	ElArwady et al., 2024
31	Intelligent Monitoring Platform and Application for Building Energy Using Information Based on Digital Twin	Li et al., 2023
32	Digital Twin of HVAC system (HVACDT) for multiobjective optimization of energy consumption and thermal comfort based on BIM framework with ANN-MOGA	Hosamo et al., 2023c
33	A BIM-enabled digital twin framework for real-time indoor environment monitoring and visualization by integrating autonomous robotics, LiDAR-based 3D mobile mapping, IoT sensing, and indoor positioning technologies	Hu & Assaad, 2024
34	Development of a Cognitive Digital Twin for Building Management and Operations	El Mokhtari et al., 2022
35	Occupant Feedback and Context Awareness: On the Application of Building Information Modeling and Semantic Technologies for Improved Complaint Management in Commercial Buildings	Gray et al., 2020
36	Predictive Analysis of a Building's Power Consumption Based on Digital Twin Platforms	Han et al., 2024
37	Digital Twin Hospital Buildings: An Exemplary Case Study through Continuous Lifecycle Integration	Peng et al., 2020
38	HVAC System Performance in Educational Facilities: A Case Study on the Integration of Digital Twin Technology and IoT Sensors for Predictive Maintenance	Salzano et al., 2025
39	Integrating Energy System Monitoring and Maintenance Services into a BIM-based Digital Twin	Pruvost et al., 2023
40	A Framework for an Indoor Safety Management System Based on Digital Twin	Liu et al., 2020
41	Digital twin enabled fault detection and diagnosis process for building HVAC systems	Xie et al., 2023
42	Smart Building Management System based on Digital Twin: A Case Study on Real-Time Environmental Monitoring and Thermal Comfort Prediction	Gao et al., 2024

43	How to measure and control indoor air quality based on intelligent digital twin platforms: A case study in China	Qian et al., 2024b
44	Enhancing Facility Management with a BIM and IoT Integration Tool and Framework in an Open Standard Environment	Chatsuwan et al., 2025
45	Digital twin-enabled built environment sensing and monitoring through semantic enrichment of BIM with SensorML	Wang et al., 2022
46	Semantic Digital Twinning for Cost-Optimal HVAC Operation: Real-Time Application to a House with Smart Thermostats and PV/Battery under a Time-of-Use Tariff	Abtahi et al., 2025