Journal of Building Engineering 114 (2025) 114214

Contents lists available at ScienceDirect

Journal of Building Engineering

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

Data-driven and physics-based modeling approaches and their integration in building digital twins: A systematic review

Jifar M. Hunde [a], Tesfatsyon S. Ochono [b], Damitha Senevirathne [a], Dagimawi D. Eneyew [a],*, Girma T. Bitsuamlak [b], Miriam A.M. Capretz [a], Katarina Grolinger [a]

[a] Department of Electrical and Computer Engineering, Western University, London, N6A 5B9, ON, Canada
[b] Department of Civil and Environmental Engineering, Western University, London, N6A 3K7, ON, Canada

ARTICLE INFO

Keywords:
Building digital twin
Data-driven modeling
Physics-based modeling
Model integration
Hybrid modeling

ABSTRACT

Interest in digital twin technology has grown significantly within the building sector as part of the broader digital transformation in the architecture, engineering, and construction industry. A building digital twin is a virtual replica that captures a building's static and dynamic behavior through data, information, and models. Digital twin models can be developed using data-driven or physics-based approaches, each with distinct advantages and limitations. Data-driven models can learn complex behaviors from data and scale well, but they require large datasets and often lack interpretability. In contrast, physics-based models offer interpretability and generalizability through fundamental principles but can be computationally demanding. Consequently, building digital twins can benefit greatly from integrating both approaches through hybrid modeling. However, the literature lacks a comprehensive analysis of integration strategies within building digital twins. This study addresses that gap by reviewing advances in data-driven and physics-based modeling and analyzing various integration levels. The results show that most studies rely on siloed models, using either approach independently without leveraging their complementary strengths. Some adopted sequential integration, where one model informs the other but lacks real-time or iterative feedback. A few achieved coupled integration, involving active data exchange and collaboration between models. Only three studies explored fusion integration, where both approaches are fully unified into a single model. Based on this review, a method is proposed for selecting the appropriate level of integration, considering factors such as data availability, interpretability, generalizability, and domain knowledge. Finally, key research gaps and future directions are identified to guide further work.

Contents

* Corresponding author.
E-mail address: deneyew@uwo.ca (D.D. Eneyew).

https://doi.org/10.1016/j.jobe.2025.114214
Received 17 June 2025; Received in revised form 15 August 2025; Accepted 27 September 2025

Available online 3 October 2025

1. Introduction

Buildings account for 34% of global energy consumption and contribute 37% of global CO_2 emissions, according to the 2022 United Nations Environment Program report [1]. Reducing this substantial environmental impact is essential for achieving global net-zero emission targets. Advances in digital transformation present promising strategies for mitigating carbon emissions in the building sector. Among these, Digital Twin (DT) technology has become an innovative tool for optimizing building performance, enhancing energy efficiency, and improving sustainability throughout a building's lifecycle. DT is conceptualized as a dynamic, virtual representation of physical entity [2–5] that integrates real-time or on-demand data [6,7], enables simulation and analytics [8,9], with bidirectional information flow [3,7] to enhance decision-making.

To achieve this level of replication and interaction, both physics-based and data-driven modeling approaches have been employed to create models that closely approximate the dynamic behavior of assets. Physics-Based Modeling (PBM) methods use fundamental physical principles to model the dynamic behavior of the assets. These models are based on governing physical equations, and since these equations are explicitly represented, the results produced are highly interpretable. Unlike data-driven models, physics-based models generalize well to new problems with similar underlying physics [10]. However, physics-based models, especially for complex systems, typically require substantial computational resources during inference, making them unsuitable for real-time DT applications.

Data-Driven Modeling (DDM) methods use measured or simulated data to create, calibrate, and refine models that represent the asset's dynamic behavior. These methods rely on Machine Learning (ML) algorithms and assume that the training data captures known and unknown physics of a given process [10]. A key advantage of data-driven models is their ability to learn from historical data, enabling them to represent complex system dynamics, even those not explicitly defined by physics-based equations. However, their reliance on training data limits generalizability to unseen scenarios, as they inherit biases from the dataset [11]. Additionally, data-driven models, especially those based on Deep Learning (DL), often lack interpretability due to their complex, high-dimensional, and non-linear nature.

Existing review papers have discussed the role of physics-based, data-driven, and hybrid modeling approaches in Building Digital Twins (BDTs) [12–24]. However, most of these review papers [12–20,25–33] focus solely on the role of data-driven methods in specific BDT applications such as indoor environmental performance [13,33], safety risk assessment [14], and facility management [15]. Although a few review papers have considered both data-driven and physics-based modeling approaches, they remain limited to specific applications of BDTs, such as fault detection and diagnosis [21,22] and energy modeling [23,24]. Furthermore, despite the benefits of integrating these approaches, such as combining the interpretability and generalization capabilities of physics-based models with the data-driven models' ability to learn from data and computational efficiency at inference, existing reviews fall short of examining how different levels of integration between the two modeling paradigms can be achieved. Existing reviews also lack a thorough analysis of the associated trade-offs, including accuracy, adaptability, interpretability, and efficiency, that influence the selection of an appropriate integration level across different BDT modeling contexts. Moreover, these reviews lack an application-agnostic synthesis of prior work from the perspective of integration levels between the two modeling approaches, which is essential for identifying application-specific challenges in adopting integrated modeling.

This review bridges this gap by systematically analyzing physics-based and data-driven modeling approaches in BDTs. Through a comprehensive analysis, the integration of physics-based and data-driven models is categorized into four levels: Level 0, siloed modeling; Level 1, sequential (one-way data exchange); Level 2, coupled (continuous data exchange); and Level 3, fusion (a unified model). Each level reflects a progressively deeper integration, ranging from independent operation to full model fusion. Following

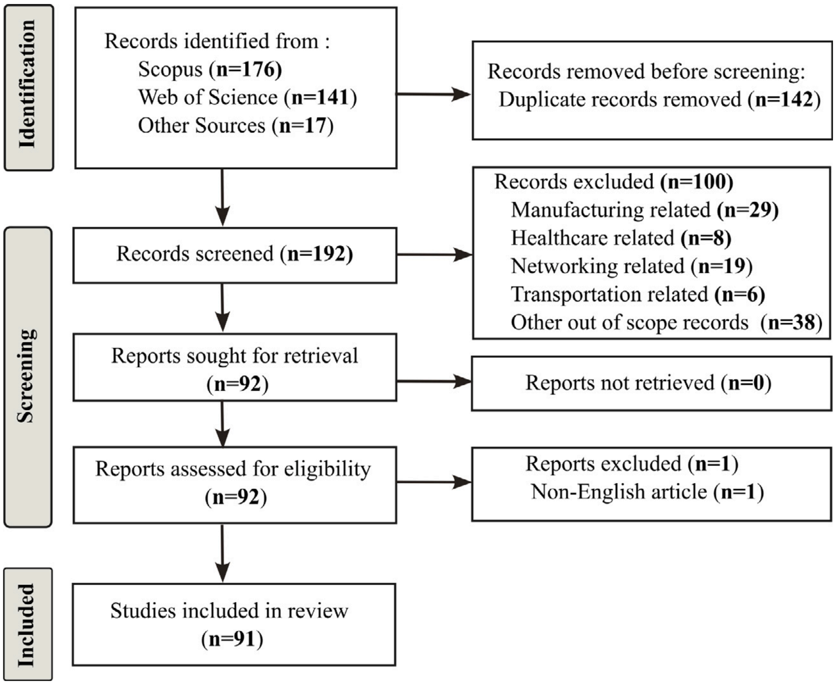

Fig. 1. Protocol for search and selection of studies: PRISMA flow diagram.

this categorization, a decision guideline is proposed for selecting the appropriate integration level for a given DT modeling problem based on data availability, interpretability, and generalization requirements. Finally, the review provides an in-depth analysis of key research gaps and proposes future research directions.

The remainder of this paper is organized as follows: Section 2 outlines the research methodology. Section 3 presents the conceptual architecture and implementation roadmap for building digital twins. Section 4 discusses DT modeling approaches and their integration levels. Section 5 discusses the trade-offs associated with DT modeling approaches and outlines the guidelines for choosing the level of integration. Section 6 highlights key research challenges and future directions. Finally, Section 7 concludes the paper.

2. Methodology

This review employs a systematic review approach to explore recent advances in PBM and DDM, investigate their varying levels of integration, and establish criteria for selecting the appropriate level of integration in DT modeling within the building sector. Guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) framework [34], our methodology ensures a rigorous process for identifying, screening, and selecting studies. Fig. 1 illustrates the systematic approach for identifying relevant studies, applying defined selection criteria to address the objectives of this review.

In order to gather relevant literature, comprehensive searches were conducted in both Scopus and Web of Science databases, as shown in Table 1. Both databases were queried for peer-reviewed articles published between 2019 and 2025. The search queries included explicit keywords: “digital twin”, “building”, and “built environment” to capture building sector-specific studies. To refine the scope, these terms were combined with keywords representing review focus areas: “artificial intelligence”, “machine learning”, “deep learning”, “white box model”, “black box model”, “grey box model”, “data-driven model”, “surrogate model”, and “physics-based model”. Filters restricted results to English-language publications. The Scopus search yielded 176 studies, while the Web of Science search returned 141. Additionally, 17 relevant studies were added based on the authors’ familiarity with the field, resulting in a total of 334 studies.

After removing 142 duplicates, 192 records were screened for relevance. During the screening process, studies outside the BDT domain, focusing on healthcare, manufacturing, networking, transportation, and other areas, were excluded after reviewing the title, abstract, and keywords, resulting in the removal of 100 papers. Additionally, one non-English paper was excluded, as its English title and abstract led to its inclusion in the initial search results, despite the full text being in another language. The remaining 91 papers were included in the review.

Table 1
Selected databases and corresponding search queries.

Source	Search queries
SCOPUS	TITLE-ABS-KEY(((‘‘digital twin*’’) AND (‘‘building*’’ OR ‘‘built environment’’)) AND (‘‘artificial intelligence’’ OR ‘‘machine learning’’ OR ‘‘deep learning’’ OR ‘‘white box model*’’ OR ‘‘black box model*’’ OR ‘‘grey box model*’’ OR ‘‘data-driven model*’’ OR ‘‘surrogate model*’’ OR ‘‘physics-based model*’’)) AND PUBYEAR > 2018 AND PUBYEAR < 2026 AND (LIMIT-TO(SRCTYPE, ‘‘j’’)) AND (LIMIT-TO(DOCTYPE, ‘‘ar’’)) AND (LIMIT-TO(LANGUAGE, ‘‘English’’)) AND (LIMIT-TO(SUBJAREA, ‘‘ENGI’’) OR LIMIT-TO(SUBJAREA, ‘‘COMP’’) OR LIMIT-TO(SUBJAREA, ‘‘ENER’’) OR LIMIT-TO(SUBJAREA, ‘‘MATH’’) OR LIMIT-TO(SUBJAREA, ‘‘ENVI’’))
Web of Science	TS = (((‘‘digital twin*’’) AND (‘‘building*’’ OR ‘‘built environment’’)) AND (‘‘artificial intelligence’’ OR ‘‘machine learning’’ OR ‘‘deep learning’’ OR ‘‘white box model*’’ OR ‘‘black box model*’’ OR ‘‘grey box model*’’ OR ‘‘data-driven model*’’ OR ‘‘surrogate model*’’ OR ‘‘physics-based model*’’)) AND PY = (2019--2025) AND DT = (Article) AND LA = (English) AND SU = (Engineering OR ‘‘Computer Science’’ OR Energy OR Mathematics OR ‘‘Environmental Sciences’’)

Note: Document types are limited to journal articles (LIMIT-TO(DOCTYPE, "ar") in Scopus; DT = (Article) in Web of Science). Subject areas are refined to engineering, computer science, energy, mathematics, or environmental sciences (LIMIT-TO(SUBJAREA, "ENGI"), COMP, ENER, MATH, ENVI in Scopus; SU = (Engineering OR "Computer Science" OR Energy OR Mathematics OR "Environmental Sciences") in Web of Science).

3. BDT conceptual architecture and implementation roadmap

Before the detailed analysis of the reviewed studies, this section discusses the conceptual architecture and components constituting a building digital twin. Following the conceptual architecture, a roadmap for constructing building digital twins is explained.

A digital twin comprises three core components: the physical entity, the virtual representation, and the communication link between them [35]. In the context of building digital twins, the physical entity refers to the building itself with all its systems and components; the virtual component consists of computational models developed using static and dynamic data collected from the building; while the bidirectional connectivity enables data exchange between the physical and virtual entities, enabling real-time synchronization and control.

3.1. Conceptual architecture

Existing studies have proposed various digital twin conceptual architectures to systematically organize core components and their sub-components. For example, Qi et al. [36] introduced a five-dimensional digital twin architecture, where the physical entity, virtual model, digital twin data, and service components interact through a central communication component. Eneyew et al. [37] proposed a layered architecture comprising five layers, specifically addressing the challenge of data interoperability between static and dynamic building data, including the physical layer, data storage layer, data access layer, data integration layer, and application layer. Piras et al. [27] extended this architecture by introducing a digital model layer between the data integration layer and the application layer, incorporating mathematical models that simulate the physical behavior of the building.

Based on the common identified patterns in literature, a high-level four-layer conceptual architecture is illustrated in Fig. 2. It should be noted that this is a high-level conceptual representation, and each layer requires in-depth detailed design of its sub-components for practical implementation.

The physical layer in the given conceptual architecture serves as the foundation of the digital twin, encompassing the actual building infrastructure including sensors that enable data collection [36,37], such as environmental sensors for monitoring temperature, humidity, CO_2 levels, and occupancy, as well as actuators that respond based on information received from the building digital twin [38,39]. The data layer manages the complete data lifecycle, from data acquisition through sensor data streams and external data sources to storage and data integration components that fuse heterogeneous data sources, ensure semantic interoperability between different data formats, and provide unified data access for modeling and application layers. The model layer represents the computational core where physics-based, data-driven, and hybrid modeling approaches create virtual representations of building behavior, including geometric models, energy models, thermal comfort models, and HVAC system models [27]. Finally, the service/application layer provides various building digital twin applications based on data and models from lower-level layers, with examples including indoor environmental quality monitoring, energy use optimization, and fault detection and diagnosis.

3.2. BDT implementation roadmap

A practical implementation roadmap for building digital twins can follow a bottom-up approach, progressing from physical layer components to the application layer. By designating one phase for each layer, a four-phase implementation roadmap can be

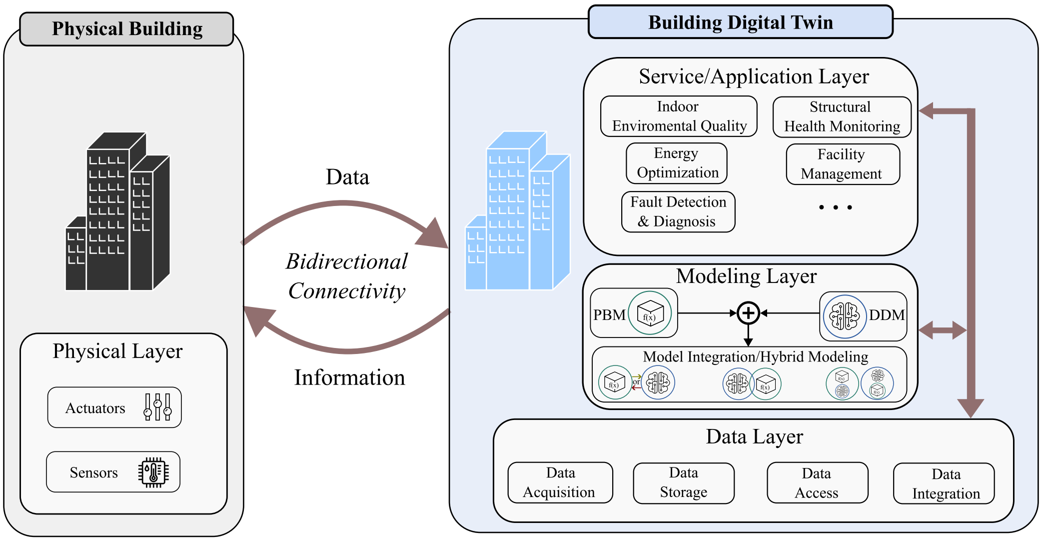

Fig. 2. Conceptual architecture of a building digital twin.

established as outlined below.

Phase 1: Physical Layer Implementation - This phase focuses on establishing the foundational sensing and actuator infrastructure. Key activities include deploying sensors throughout the building, integrating existing building automation systems, and implementing reliable data communication links between the physical building and digital twin. Critical implementation considerations include optimal sensor placement to ensure sufficient coverage, determination of appropriate sensor data reading frequencies, and selection of suitable data communication protocols.

Phase 2: Data Layer Implementation - This phase focuses on deploying a comprehensive data management infrastructure. Key activities involve implementing data processing pipelines for cleaning, filtering, and aggregation operations, developing data integration components that fuse heterogeneous sources while ensuring semantic interoperability, and establishing unified data access mechanisms for seamless retrieval across all sources. Critical implementation considerations include database selection based on data type requirements, robust data quality management through validation and automated checks, effective handling of missing or corrupted data, and comprehensive data governance policies including access control.

Phase 3: Modeling Layer Implementation - This phase represents the core of the building digital twin, involving deployment of PBM, DDM, and hybrid approaches to create virtual representations of building behavior. PBM implementation follows structured steps, including defining modeling objectives, identifying relevant physical principles and governing equations, selecting appropriate solvers and simulation engines, developing models using chosen methodologies, and calibrating with measured data for accuracy [10,40]. DDM implementation involves defining objectives, collecting and preprocessing data, selecting or designing suitable model architectures, and conducting training and evaluation processes. As the modeling layer matures, hybrid approaches are progressively introduced to address individual method limitations, combining the strengths of both paradigms while mitigating their respective weaknesses. Critical implementation considerations in this phase include comprehensive model validation processes, seamless integration mechanisms between modeling paradigms, adequate computational resource allocation, and establishment of frameworks for continuous model updating and maintenance as new data becomes available. Further details of the hybrid modeling approaches and implementation details with examples are discussed in Section 4.

Phase 4: Service/Application Layer Implementation - This phase involves implementing building digital twin applications based on available models and data. Leveraging the unified data access from the data layer and various models in the modeling layer, several applications can be developed based on specific needs, including real-time monitoring dashboards, predictive analytics for energy optimization, fault detection and diagnosis, thermal comfort management, and predictive maintenance scheduling. Critical implementation considerations include user interface design tailored to different stakeholder needs, API development for system integration, service-oriented architecture design to support scalable applications, security protocols for application access, and comprehensive testing frameworks to ensure application reliability and accuracy.

4. DT modeling approaches and levels of integration

Based on a detailed review of studies, different levels of integration between the data-driven and physics-based modeling approaches have been identified. Fig. 3 illustrates the decision methodology used to determine the level of integration of a given study. The levels of integration are categorized into four levels, namely Siloed modeling, Sequential integration, Coupled integration, and Fusion integration, as shown in Fig. 4.

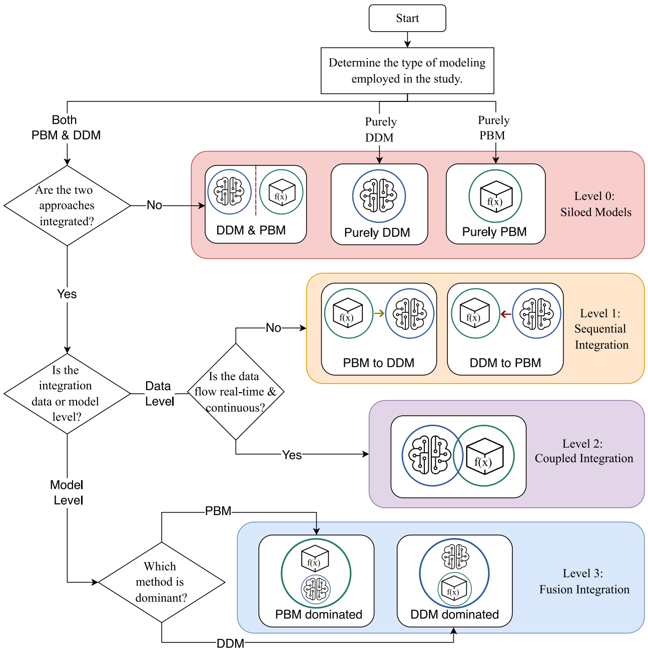

Fig. 3. Conceptual framework for categorizing the level of integration in reviewed studies.

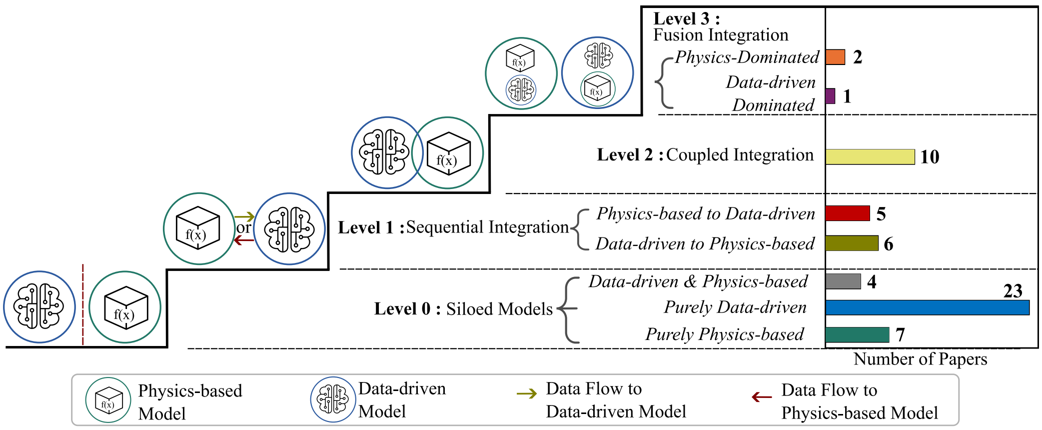

Fig. 4. Digital twin modeling approaches, Integration levels, and the number of studies at each level.

Level 0 (Siloed Modeling): At this level, physics-based and data-driven models operate independently with no interaction or data exchange. They are isolated, working separately without influencing each other. This level is divided into three categories, as shown in Fig. 3: the first category includes studies using purely DDM; the second consists of studies that rely exclusively on physics-based methods; and the third category involves both approaches, though without interaction between them. Within this level, DDM-only approaches rely on a generic model structure where structure and parameters are learned from data, such as in ML, DL, or statistical learning. Conversely, PBM-only approaches base their model structure and parameters on known physical principles, such as energy or mass balance and fluid flow equations.

Level 1 (Sequential Integration): This level of integration occurs at the data level and follows a sequential process. One model informs the other without real-time or iterative interaction. Two sub-categories of integration at this level are identified based on data flow direction: (1) from physics-based to data-driven models and (2) from data-driven to physics-based models, as shown in Fig. 3. At this level of integration, there is a unidirectional data flow from DDM to PBM or vice versa, but this flow is neither real-time nor continuous. An example includes using data from a physics-based simulation model during the design phase to train a surrogate ML model.

Level 2 (Coupled Integration): It represents models that actively exchange data and work collaboratively towards a common goal, where one model's output serves as the input for the other. Although the integration is still at the data level, it involves more dynamic and iterative exchanges than Level 1. Level 2 occurs when there is a continuous and real-time data exchange between DDM and PBM. For instance, an ML model estimating the occupant count parameter in real time for a PBM-based energy simulation.

Level 3 (Fusion Integration): It represents the highest integration level, where PBM and DDM advantages are merged into a single cohesive model. This level integrates both interpretability and generalization (from PBM) and computational efficiency (from DDM) into one unified model. The integration is at the model level, such as directly embedding physical laws within ML models, which combines the strengths of both modeling approaches. This level of integration is further sub-categorized into PBM-dominated or DDM-dominated, depending on which modeling paradigm is dominant, as illustrated in Fig. 3.

The distribution of research efforts across different levels of model integration is shown in Fig. 4. A clear pattern observed is that the number of studies declines as the integration between PBM and DDM increases. The majority of studies adopted Level 0, which involves siloed modeling approaches with a strong emphasis on purely data-driven methods. This distribution reflects the growing availability of sensor data in buildings, which many studies are now leveraging. Another factor could be the relative ease of implementation compared to PBM, which typically requires substantial time, effort, and domain expertise to develop. Level 1 (sequential integration) and Level 2 (coupled integration), which represent data-level integration, have received significantly less attention than Level 0. At Level 1, studies explore data flows in both directions, from physics-based to data-driven models and vice versa, with a relatively balanced focus. Level 3 [41–43] is the least explored among all integration levels. Its limited adoption can be attributed to challenges in aligning physics-based and data-driven representations, managing training instability when enforcing physical constraints, preserving interpretability within unified models, and the lack of mature tools and frameworks for seamless model fusion. The following subsections present a detailed discussion of studies at each integration level.

4.1. Level 0: Siloed models

A summary of the studies at this integration level is provided in Table 2, organized by sub-category, main focus, and specific data-driven or physics-based model used. Most studies exclusively employed purely data-driven models [44–55]. A smaller set of studies applied purely physics-based models, mainly using energy balance equations and heat and mass transfer laws [38,39,56–59]. Some studies employed both data-driven and physics-based models for different tasks; however, these models operated independently without any integration [40,60–62]. The following subsections provide a detailed discussion of each subcategory within Level 0.

4.1.1. Purely data-driven approaches

The reviewed studies applied these methods in various DT modeling domains, including geometry modeling, energy optimization, thermal comfort prediction, fault detection and diagnostics, building monitoring, HVAC modeling, and predictive maintenance.

Building geometry modeling: Purely DDM-based approaches have been applied to building geometry modeling in various studies. For instance, Mehranfar et al. [44] employed a DTree classifier in conjunction with a Density-Based Spatial Clustering of Applications with Noise (DBSCAN)-based semantic segmentation technique. This method processes dense point clouds to produce semantic digital models. Similarly, Pan et al. [45] introduced a DL method, Kernel Point Segmentation and Global Feature Aggregation (KP-SG), to segment semantic point clouds into Building Information Modeling (BIM) elements. The neural network architecture proposed for semantic segmentation is based on an encoder–decoder architecture. A DDM has also been used for 3D modeling in city-wide DTs. For instance, Adreani et al. [46] utilized Support Vector Machines (SVM) and Random Forests (RF) for building detection and rooftop segmentation as part of the Snap4City platform for smart city solutions. They also employed DL models, such as Mask R-CNN and U-Net architectures, to detect and segment rooftops from images.

In addition to generating 3D geometric BIM models, data-driven methods are also applied to BIM models of buildings to extract spatial and structural health information. To obtain spatial information, Abdelrahman et al. [63] utilized a building's BIM model to create a multidimensional vector derived from the BIM graph. The prediction model employs an RF classifier to categorize thermal comfort labels. For structural health information, Luo and Wang [64] applied data-driven methods to masonry-timber heritage buildings, focusing specifically on crack and damage detection. They used a Mask R-CNN model to detect crack areas and a Fully Convolutional Network (FCN) model to identify and analyze individual cracks.

Energy optimization & management: DDM has been a cornerstone in numerous studies focused on building energy optimization and management [53–55,65]. These studies predominantly utilize ML models such as Multilayer Perceptron (MLP), K-Nearest Neighbors (KNN), SVM, DT, and RF. For instance, Tariq et al. [53] employed an MLP model to simulate the behavior of double-skin building facades. The authors preferred the MLP-based approach over PBM due to the high computational cost associated with the latter in optimization tasks. To improve the interpretability of the MLP model, Explainable Artificial Intelligence (XAI) techniques such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) were incorporated. Sayed et al. [54] extended energy optimization by considering user comfort and energy reduction, employing a range of ML models, including KNN, Decision Trees (DTree), RF, SVM, logistic regression, and Naive Bayes. Similarly, Jradi et al. [55] proposed a DT solution for optimizing energy retrofits in existing non-residential buildings, using data-driven models such as MLPs and DTree. DDM requires training datasets that comprehensively represent all relevant operational scenarios. However, rare events are often underrepresented or absent in real-world operational data, posing a significant challenge for model generalization. To overcome this limitation, Choi and Yoon [65] proposed a data augmentation approach based on autoencoders to improve the accuracy of in-situ models for building energy systems. Their method maps real operational data into a latent space to generate synthetic data, thereby addressing the scarcity and low informativeness of real-world data.

Thermal comfort prediction: Targeting the improvement of indoor thermal comfort, ElArwady et al. [67] developed a predictive model using Facebook Prophet and NeuralProphet algorithms within a platform integrating DT, Internet of Things (IoT), and BIM

Table 2
Summary of studies with their focus and modeling technique under Level 0: Siloed modeling.

Category	Focus	DDM	PBM	Ref.
Purely DDM	Building geometry modeling	DTree & DBSCAN	–	[44]
		KP-SG	–	[45]
		SVM, RF, R-CNN, U-Net	–	[46]
		RF	–	[63]
		Mask R-CNN & FCN	–	[64]
	Energy optimization & management	MLP	–	[53]
		KNN, MLP, SVM, DTree, RF, Naive Bayes	–	[54]
		ANN, DTree	–	[55]
		Autoencoders	–	[65]
	Thermal comfort prediction	RF, MLP, LSTM, Convolutional Neural Networks (CNN)	–	[66]
		Facebook Prophet & Neural Prophet	–	[67]
		Bayesian NN, XGB, MLP, SVM, KNN, RF, Naive Bayes, DTree	–	[68]
	Fault detection & diagnosis	Not specified	–	[69]
		CNN	–	[70]
	Facility management & monitoring	Pre-trained Language Model	–	[47]
		GPT-based Language Model	–	[71]
		GPT-4o, Retrieval Augmented Generation	–	[72]
		Not specified	–	[48]
		LSTM	–	[49]
		SVM, RF, MLP	–	[50]
	HVAC modeling	ANN & SVM	–	[51]
	Predictive maintenance	ANN & SVM	–	[73]
		ANN, DTree, SVM	–	[52]
Purely PBM	Energy modeling & optimization	–	Electrical network modeling	[56]
		–	Heat & mass transfer	[57]
	Virtual sensor modeling	–	Energy balance equations	[39]
		–	Energy balance equations	[58]
	Crowd simulation	–	–	[74]
	Temperature & humidity prediction	–	Heat and mass transfer	[59]
	HVAC modeling	–	Thermodynamic and fluid dynamic equations	[38]
PBM & DDM	Indoor environmental quality	Gradient Boosting and LSTM	Sun position algorithm	[40]
		ANN	Photogrammetry	[62]
	Virtual sensor modeling	MLP	Energy balance equations	[60]
		MLP	Energy balance equations	[61]

technologies. Similarly, Hosamo et al. [68] combined BIM with real-time sensor data and a Bayesian network model, evaluating nine ML algorithms for occupant comfort prediction and predictive maintenance. Their research found that Extreme Gradient Boosting (XGB) was the most accurate among the tested algorithms. Further, Norouzi et al. [66] compared five algorithms for indoor temperature prediction, such as Tree-based models (RF and Extra Trees) and DL algorithms (MLP, Long Short-Term Memory networks (LSTM), CNN), aiming to develop DT for HVAC systems.

Fault detection & diagnosis: Several studies [69,70] employed solely data-driven approaches to fault detection and diagnostics tasks. The study by Xie et al. [69] proposed a data-driven approach for fault detection and diagnosis in HVAC systems, addressing the issue of information saturation in building-related data. The method preserved the associated ad-hoc temporal knowledge identified during the data analysis phase, aided by symbolic artificial intelligence techniques and statistical analysis methods, including bag-of-words, symbolic aggregate approximation, the chi-square test, and Kullback–Leibler divergence. Another study by Mokhtari et al. [70] included two proof-of-concept case studies to address two different aspects of fault detection: reactive fault detection and proactive fault detection. In their case studies, the authors emphasize the use of natural language processing and CNN models and refer to Bayesian networks as a possible future direction. Furthermore, this study aims to develop a cognitive DT for building management and operations.

Facility management & monitoring: In the context of facility management and building monitoring applications, several studies [47–50,71,72] have adopted solely data-driven methods. Both et al. [47] proposed a technique that enables the automated generation of technical monitoring applications for existing buildings, leveraging DL and natural language processing. This process was associated with German pre-trained language models. Similarly, Bao and Bu [71] utilized a GPT-based Large Language Model (LLM) to generate operation and maintenance guidelines based on predefined requirements. Building on this direction, Yoon et al. [72] proposed an AI agent-driven intelligent DT framework that leverages LLMs to perform tasks such as virtual in-situ model generation, fault detection and diagnosis model creation, and report summarization in response to user prompts. A Brick schema-integrated ontology was developed to represent the virtual building model, enabling the AI agent to comprehend the system's structure and functionality. In addition to the ontology, the AI agent accessed sensor data from the physical building to answer user queries related to building operations and to support the automated generation of models. Antonino et al. [48] proposed an occupancy monitoring system for office buildings using image recognition sensors integrated with BIM to optimize facility management operations. The study employs an image recognition system based on artificial intelligence for occupancy detection and people tracking. However, the paper does not specify the exact architecture or algorithms used in their approach. Occupant activity recognition is another aspect of facility monitoring where DDM is applied. For instance, the study by Bouchabou et al. [49] applied an LSTM-based ML model for occupant activity recognition.

HVAC modeling: Data-driven modeling also extends to HVAC systems. Park et al. [51] applied ML techniques for HVAC modeling within virtual power plants, comparing non-linear regression, Artificial Neural Networks (ANN), and SVM.

Predictive maintenance: The study conducted by Hosamo et al. [52] focuses on a DT predictive maintenance framework for air handling units in buildings, leveraging BIM, IoT, semantic technologies, and ML, followed by a case study of an educational building. ANNs, SVMs, and DTrees were employed for fault detection tasks to predict and assess critical faults in air handling units. Additionally, Zhao et al. [73] developed an ANN-based method to predict cable forces in a spoke-type steel frame, illustrating how DTs can support building operation and maintenance.

4.1.2. Purely physics-based approaches

In this category, physics-based methods are the primary modeling approach. The reviewed studies applied these methods across various DT modeling domains, including energy modeling, virtual sensor modeling, crowd simulation, and temperature and humidity prediction. The following paragraphs provide an overview of the studies in each of these areas.

Energy modeling & optimization: Pure PBM has been adopted in building energy modeling as the primary methodology for energy optimization, system analysis, and simulation. Bayer et al. [56] applied fundamental electrical network principles to model the energy grids of local residential buildings, incorporating components such as photovoltaics, battery storage systems, and heat pumps within a DT framework. The model enabled the authors to evaluate the profitability of photovoltaic and battery storage systems and to develop effective control strategies for battery charging and discharging. In the context of building energy optimization, Borodinecs et al. [57] proposed a framework for operational optimization using continuous dynamic energy simulation within the OpenStudio platform, where EnergyPlus served as the underlying simulation engine.

Virtual sensor modeling: Physics-based methods have been used as the primary approach for virtual sensor modeling [39,58]. Virtual sensors generate real-time measurements either to replace failed physical sensors, where they function as backup virtual sensors, or to estimate physical variables that are difficult or costly to measure directly, thus serving as observation virtual sensors. For observation virtual sensor modeling, Choi et al. [39] and Yoon et al. [58] employed energy balance equations and utilized Bayesian inference to calibrate the original model. Due to the unavailability of corresponding physical sensors in the case of observation virtual sensors, operational data for DDM cannot be collected, making it challenging to develop a data-driven model. Therefore, PBM remains the most viable approach for modeling observation virtual sensors.

Crowd simulation: Meschini et al. [74] developed a BIM-integrated geographic information system platform for campus asset management, incorporating a physics-based crowd simulation model to assess building safety in emergency scenarios, particularly for fire evacuation.

Temperature & humidity prediction: Ward et al. [59] used a particle filter method based on sequential Bayesian inference for the continuous calibration of a physics-based heat and mass exchange model. This method allows for the calculation of temperature and relative humidity in an underground farm.

HVAC modeling: Walther et al. [38] examine the role of DTs in HVAC controls and propose a framework to close the control-related performance gap between building designs and actual operations. This study discusses two approaches: replicating existing building controls in digital twins (current practice) and using digital twins to specify new HVAC controls (proposed approach). The authors used IDA ICE, a physics-based building performance simulation software, to build the DT prototype.

4.1.3. Data-driven and physics-based

At this level, both types of models work in isolation, and there is no interaction between the two models. Studies in this category focused on specific DT modeling domains, including virtual sensors and indoor environmental quality. An overview of the research in each of these areas is presented in the following paragraphs.

Indoor environmental quality: Elfarri et al. [40] introduced two modeling approaches for predictive DTs. The authors used DDM, based on gradient boosting and LSTM models, to predict indoor temperatures. In contrast, PBM applies to scenarios where the physics is well understood. The authors utilized PBM to predict sun positions. The two modeling approaches remain separate, with no interaction between them. Similarly, Arsiwala et al. [62] applied this approach for monitoring indoor air quality. The authors used LiDAR technology and photogrammetry to produce high-resolution scans of indoor spaces, which were then incorporated into the 3D digital model layer of a multi-layered DT architecture. This architecture was designed to support both monitoring and analysis

Table 3
Summary of studies with their integration objectives and modeling technique under Level 1: Sequential integration.

Sub-category	Objective of integration	PBM	DDM	Ref.
PBM to DDM	Surrogate modeling	Building energy model (with Energyplus)	RF	[75]
		Energy balance equations	MLP	[76]
		CFD based numerical model	AutoDecoder LSTM	[77]
	Training data enhancement	Radiometric intensity correction	RandLA-Net	[78]
		Wind tunnel experiments	SVD, POD	[79]
DDM to PBM	DDM-based input data generation	Seed point and void-growing	KPConv model	[80]
		Void-growing	KPConv model	[81]
		Homography transformation & KLT feature tracking	TernausNet	[82]
		Discrete-element model	CNN	[83]
		Least squares kinematic algorithm	CNN	[84]
		Euclidean distance calculation	YOLOv3	[85]

of carbon emissions. A separate data analytics layer, which included an ANN-based CO_2 predictive component, was used to forecast future emissions.

Virtual sensor modeling: Studies [60,61] within this category employed both PBM and DDM independently for virtual sensor modeling. Since the physical sensor and the training data collected from it were available, the authors utilized DDM for modeling backup virtual sensor using MLP. On the other hand, the authors employed PBM to create observation virtual sensors due to the unavailability of corresponding physical sensors. Specifically, they used an energy balance equation-based PBM for observation virtual sensor modeling.

4.2. Level 1: Sequential integration

This integration is at the data level and sequential in nature, where one model informs the other model, but no real-time or iterative interaction occurs between the models. As listed in Table 3, the reviewed literature contains two types of integration at this level based on their direction of data flow. The first is where the data flows from a physics-based model to a data-driven model. This type of integration is mainly introduced for surrogate modeling, where the output data from a white-box model is used to create a black-box surrogate model. The second type is where data flows from the data-driven model to the physics-based model. This type of integration is used for several reasons: to enhance the ML model and as an input for the physics-based model. Fig. 5 presents different approaches for achieving sequential integration of PBM and DDM. The following subsections discuss the papers that fall into this level of integration and its subcategories, and outline the generalized step-by-step processes (technical loops) required to implement them.

4.2.1. Direction: Physics-based to data-driven

In this type of integration, data flows from a physics-based to a data-driven modeling approach. This approach is used primarily for surrogate modeling and training data enhancement.

Surrogate modeling: Surrogate modeling involves scenarios in which data-driven models approximate complex physics-based models. A general procedure for developing a surrogate model, as illustrated in Fig. 5A, involves generating simulated data from the target physics-based model and using it to train a computationally efficient data-driven model capable of replacing the original, computationally intensive model. Following this procedure, Tahmasebinia et al. [75] utilized data from a physics-based model to train data-driven models for predicting energy consumption. By leveraging DDM-based surrogate modeling, the integration significantly reduces the computational demands of physics-based models by substituting them with efficient surrogate models. Data from EnergyPlus simulations were used to train several ML models, including RF, sigmoid regression, and linear regression, for predicting energy consumption. The interaction between the physics-based EnergyPlus model and the ML-based models was sequential, with simulation data supporting the development of an energy prediction surrogate model. Xie et al. [77] adopted a similar approach by developing an AutoDecoder LSTM surrogate model that transforms discrete sensor-array data into high-dimensional spatiotemporal temperature fields in real time. This enables the forecasting of future fire development and the identification of hazardous regions. Due to the model's fast inference time, the authors were able to predict fire progression within 60 s, a crucial factor for timely fire response. Choi et al. [76] extended surrogate modeling techniques for virtual sensor modeling. They used training data generated by the initial physics-based virtual sensor model, which was based on the energy balance equation. The purpose of introducing the surrogate model was to reduce the initial model's uncertainty and thereby improve its accuracy.

Training data enhancement: In addition to surrogate modeling, sequential integration is important to improve the accuracy of DDM by enhancing the training dataset. This is done by either augmenting the set with simulated data (Fig. 5B) or applying physics-based enhancements to existing data (Fig. 5C). The latter approach involves a pre-processing stage where raw data is refined using a physics-based constraint before model training. For instance, Maru et al. [78] applied this method using a physics-based radiometric correction on raw LiDAR data, which improved the subsequent building facade segmentation by the RandLA-Net model.

In the wind engineering domain, Luo et al. [79] proposed an innovative approach for optimal sensor placement to reconstruct wind pressure fields around tall buildings. Their method leverages wind tunnel experiments that generate pressure data simulating wind structure interactions and vortex dynamics. This experimental data serves as the training set for a DDM framework, which

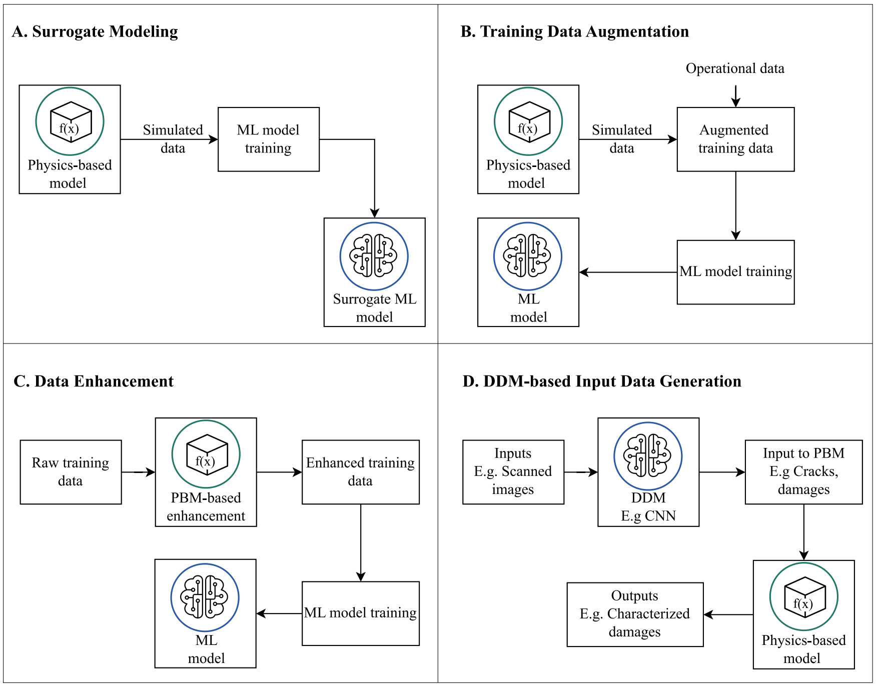

Fig. 5. Approaches for achieving sequential integration of PBM and DDM. (A) Training surrogate ML models using PBM-generated data [75–77]; (B) Generating ML training data using PBM simulations [79]; (C) Enhancing ML training data using PBM [78]; (D) Creating inputs for PBM using DDM [80–84].

employs compressed sensing, Singular Value Decomposition (SVD), or Proper Orthogonal Decomposition (POD) for tailored basis construction, QR pivoting for sensor selection, and convex optimization for sparse reconstruction. The integration between physics-based and data-driven models is sequential, where the physical measurements directly inform the data-driven processes, ultimately enhancing reconstruction accuracy and sensor selection.

4.2.2. Direction: Data-driven to physics-based

In this type of integration, data flows from data-driven to physics-based models. Studies in this subcategory typically use DDM to generate input data for a physics-based model. The integration process begins by training a data-driven model or utilizing an existing pre-trained model. The predicted output of this model, such as class labels, image segments, or semantically segmented point cloud data, is then fed into a physics-based model for further processing. Due to the processing done by the data-driven model, the accuracy of the physics-based model is improved. For example, CNN-based image segmentation results in better damage characterization by a discrete element structural analysis model [83]. The generalized technique is presented in Fig. 5D. The following subsection provides a detailed analysis of a common application for this approach: data utilization.

DDM-based input data generation: Some studies have applied this approach to generate 3D models from point-cloud data [80–82]. For example, Pantoja-Rosero et al. [82] integrated rule-based computer vision methods such as 2D-to-3D triangulation, homography transformations, and Kanade-Lucas-Tomasi (KLT) tracking alongside the TernausNet ML model. The TernausNet ML model is used for segmenting openings, corners, and facades in 2D images. The resulting segments are then transformed into 3D using homography transformations and KLT tracking. Similarly, Pan et al. [80,81] introduced a “void-growing” approach for generating geometric DTs of indoor environments. This method detects void spaces in rooms and integrates geometric and semantic data to handle complex room shapes, varying ceiling heights, and high occlusion. These works highlight the interaction between ML and rule-based models, with data flowing from data-driven to rule-based methods.

Other studies have applied DDM-based input data generation for physics-based models in structural health analysis [83,84], and to detect individuals [85]. Loverdos et al. [83] developed models to detect cracks in aging masonry structures using scanned images. A CNN extracts the micro-geometry and cracks from the pictures, then the discrete-element method processes the resulting data into a structural analysis model. Pantoja-Rosero et al. [84] built upon their earlier work [82] by integrating a rule-based least-squares kinematic algorithm with a CNN model. Beyond 3D modeling, they incorporated damage analysis by leveraging CNN for damage detection and using the kinematic algorithm to characterize the identified damage. Mukhopadhyay et al. [85] used this approach to automatically detect individuals for enforcing social distancing during the COVID-19 pandemic. The authors integrated CNN and YOLOv3 with a rule-based Euclidean distance calculation to identify people in camera footage. The output of the data-driven model served as input for physics-based distance calculations.

Table 4
Summary of studies with their integration objectives and modeling technique under Level 2: Coupled integration.

Objective of integration	PBM	DDM	Ref.
Continuous calibration of physics-based models	Building energy model (with Energyplus)	DECI-Net	[86]
	Energy balance equations	MLP	[58]
	Energy balance equations	MLP	[87]
	Energy balance equations	MLP	[88]
Real-time temporal variable estimation	Building energy model	LSTM	[89]
	Model Predictive Control (MPC)	MLP	[90]
	TRNSYS simulator	GenOpt	[91]
	Building energy model	K-means & Naïve Bayes	[92]
	Hue, saturation, and intensity color model	YOLOv4	[93]
Simulation environment for RL	Energy balance equations (EnergyPlus)	PPO, SAC & TD3	[94]

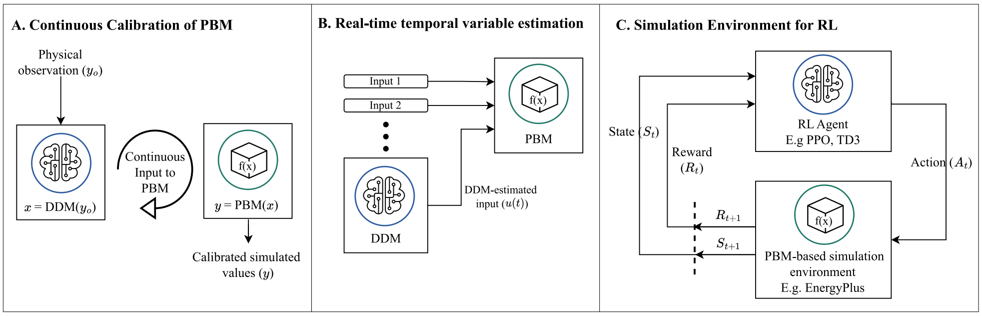

Fig. 6. Approaches for achieving coupled integration of PBM and DDM: (A) Continuous calibration of PBM using DDM as a calibrator [86]; (B) Estimating analytically complex inputs for a physics-based model using DDM [90]; (C) Employing PBM as a Reinforcement Learning (RL) environment to train RL agents [94].

4.3. Level 2: Coupled integration

This integration occurs at the data level and is iterative, whereby one model transmits data to the other in real time or continuously. Three primary application areas of coupled integration are identified in the reviewed studies, as summarized in Table 4. The first involves continuously calibrating a physics-based model using outputs from a data-driven model. The second focuses on generating real-time estimates of temporal variables using DDM, which are then used as dynamic inputs for subsequent physics-based model simulations. The third involves constructing a virtual simulation environment using PBM to support the training and evaluation of Reinforcement Learning (RL) agents. Fig. 6 presents approaches for achieving coupled integration of PBM and DDM in relation to these three areas. The following subsections examine these three areas in detail, reviewing the studies conducted under them and outlining the generalized, step-by-step processes required for their implementation.

Continuous calibration of physics-based models: Eneyew et al. [86] proposed a generative model–based framework for calibrating a physics-based building energy model. The general procedure for calibrating a physics-based model using a data-driven model is illustrated in Fig. 6A. In this approach, the data-driven model continuously supplies calibration inputs (x) to the physics-based model. The data-driven model itself is a generative model that leverages measured sensor values (y_o) as conditioning inputs. In the framework of Eneyew et al. [86], the generative calibrator is implemented using a conditional invertible neural network, coupled with a denoising autoencoder serving as the conditioning network. This calibrator model continuously generates optimal input parameters (x) for the physics-based model based on real-time measurements from physical sensors (y_o). By introducing a data-driven, generative calibrator model, the framework eliminates the need for the computationally intensive Markov Chain Monte Carlo (MCMC)-based Bayesian inference typically used in virtual sensor model calibration [58,87].

Similarly, other studies such as Yoon et al. [87], Yoon and Koo [58], and Li et al. [88] employed DDM for PBM-based virtual sensor calibration. In these studies, the PBM is formulated based on energy balance equations, and the DDM-based calibration relies on MLPs and data-driven MCMC methods to perform Bayesian inference. In the first study [87], the physics-based model provides initial predictions, which are then validated and corrected using DDM. These corrected outputs are fed back into the system to iteratively and continuously refine the models. In the second study [58], both the physics-based and the data-driven models are coupled, and the physics-based model is continuously calibrated through the residuals generated by the data-driven model.

Real-time temporal variable estimation: Inputs to a physics-based model that are analytically complex and difficult to measure directly pose significant challenges for accurate implementation. Such inputs are often time-varying and can exhibit significant uncertainty due to occupant behavior, operational variability, or environmental fluctuations. Examples include occupancy schedules and real-time energy demand profiles. To address this challenge, data-driven models can be employed to estimate these inputs, leveraging sensor measurements and historical data. The process for this approach is illustrated in Fig. 6B. The DDM-estimated variable ($u(t)$) is supplied as input to the physics-based model.

Table 5
Summary of studies with their integration objectives and modeling technique under Level 3: Fusion integration.

Sub-category	Objective of integration	PBM	DDM	Ref.
PBM dominated	To estimate the parameters of PBM correction terms	Energy balance model	Regression	[42]
	To calibrate parameters, quantify uncertainty	RC model (2R2C) + heater sub-model	MLE and NSGA-II	[43]
DDM dominated	To improve interpretability and reduce data needs	Navier–Stokes equations, boundary & initial conditions	PiNN	[41]

As an example of this approach, O'Dwyer et al. [90] integrated data-driven and physics-based methods into the MPC framework for an energy management tool designed to optimize the control, scheduling, forecasting, and coordination of energy assets across a district. The authors used different data-driven models to estimate input variables with analytically complex physical dynamics ($u(t)$), including forecasting electric vehicle charging requests with K-means clustering and predicting heat demand (for room heating and cooling) using algorithms like ANN, RF, and gradient boosting. This study highlights the integration between PBM and DDM, where the ML model provides predicted values for the next time step to inform the physics-based MPC process. The MPC process relies on prediction outputs from the DDM, demonstrating a dynamic and iterative interaction between the two modeling approaches. Additionally, both models function as part of a unified system under the MPC optimization method.

Tan et al. [93] adopted a coupled integration approach for a real-time adaptive lighting control system. In this framework, a PBM-based ambient illuminance detection algorithm and the Bhattacharyya distance algorithm are used to analyze key frames in the video stream, enabling the estimation of brightness and similarity without the need for physical sensors. The output of these algorithms is then used to trigger a DDM-based YOLOv4 pedestrian detection algorithm. Through this iterative and real-time coupled integration, the authors developed an effective real-time adaptive lighting control system, achieving a 79% reduction in energy consumption.

In the area of Building Energy Modeling (BEM), studies such as those by Borja-Conde et al. [91], Agostinelli et al. [92], and Bjørnskov and Jradi [89] demonstrated the effectiveness of coupled integration, each presenting innovative methodologies to enhance energy efficiency and system performance. Borja-Conde et al. [91] combine a physics-based simulator with a data-driven parameter identification process (Generic Optimization (GenOpt)) to optimize HVAC performance, using an iterative interaction where the data-driven model refines the parameters of the physics-based model. Likewise, Agostinelli et al. [92] integrate physics-based methods (Computational Fluid Dynamics (CFD) and BEM) with ML models in a DT framework, where CFD analysis provides detailed insights into airflow and thermal distributions within building spaces, and BEM outputs are optimized using ML techniques for better energy utilization. Similarly, Bjørnskov and Jradi [89] present an ontology-based energy modeling framework that tightly integrates physics-based and data-driven models. These studies illustrate coupled integration by demonstrating active data exchange and cooperation between different modeling approaches. In each case, the output of one model serves as input for another, creating a dynamic and iterative interaction that enhances overall system performance and accuracy.

Simulation environment for RL: Integrating physics-based models with ML enables advanced approaches for dynamic system control and optimization, particularly through RL. A key component of RL is the environment, which defines the system dynamics and feedback the agent relies on for learning. Modeling this environment through PBM is a promising approach, since it provides realistic, physics-informed simulations that enhance training effectiveness. Fig. 6C illustrates using a PBM as the simulation environment for RL training. In this framework, the RL agent sends actions (A_t) to the PBM environment, which then returns the next time step's state (S_{t+1}) and reward (R_{t+1}) to the agent. This process follows the standard RL training loop, with the PBM replacing the conventional environment to enable physics-informed and realistic agent training.

As an example of this approach, Campoy-Nieves et al. [94] implemented a virtual testbed based on EnergyPlus as a simulation environment to support RL-driven energy optimization. Various RL agents, including Soft Actor-Critic (SAC), Twin Delayed Deep Deterministic Policy Gradient (TD3), and Proximal Policy Optimization (PPO), were tested. These agents interact with the EnergyPlus-based environment iteratively to receive updated states and rewards. However, since running EnergyPlus simulations iteratively requires significant computational resources and RL training is not sample-efficient, the overall optimization process is computationally expensive.

4.4. Level 3: Fusion integration

Fusion integration is the highest level of integration and occurs at the model level. It involves directly embedding physical laws within ML models (e.g., Physics-Informed Neural Networks (PiNN) [95], physics-embedded neural networks and neural operators [96]), which combines the strengths of both modeling approaches. Depending on which modeling method dominates the final fused model, fusion integration is categorized as PBM-dominated or DDM-dominated, as listed in Table 5.

4.4.1. PBM dominated

Fig. 7A illustrates the general process of PBM-dominated fusion integration, where the physics-based model serves as the foundational component. In this approach, the PBM first models the core system dynamics based on physical principles. Subsequently,

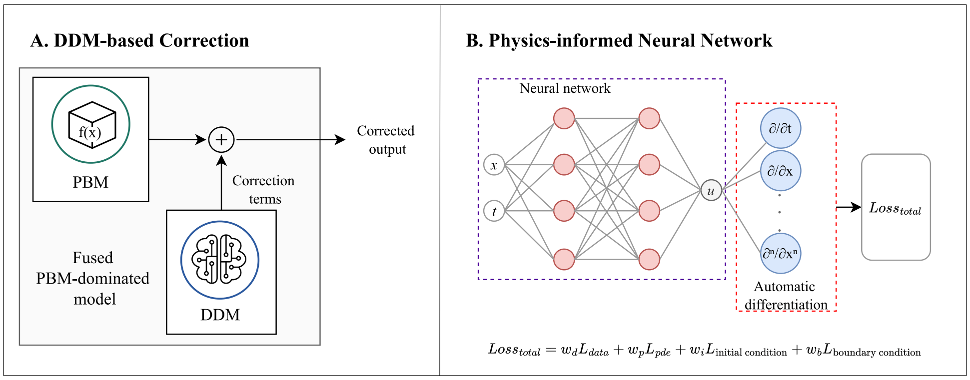

Fig. 7. Approaches for achieving fusion integration of PBM and DDM: (A) Correcting a PBM by learning missing physics from data [42,43]; (B) Incorporating physics-constrained loss terms into the ML training loss function [41].

correction terms are incorporated to improve model fidelity by addressing simplifications and unmodeled effects. These correction terms include parameters that are learned through a DDM process. An example of this approach is presented by Choi et al. [42]. The authors enhanced an energy-balance-based model by adding polynomial correction terms parameterized via polynomial regression within a DDM framework. This method enabled the creation of backup virtual sensors capable of replacing faulty physical sensors by combining PBM with data-driven corrections.

Similarly, Yang and Jardi [43] implemented PBM-dominated fusion to evaluate energy retrofit strategies in existing buildings. Confronted with limited input data for detailed white-box models, they estimated thermal resistances and capacities from common measurements, such as indoor/outdoor temperatures, solar radiation, and heating data, using maximum likelihood estimation (MLE). Their approach integrates a Resistor–Capacitor (RC) model with stochastic terms calibrated via MLE, and the resulting model is embedded within a data-driven optimization loop based on the Non-dominated Sorting Genetic Algorithm (NSGA-II). While this type of fusion maintains the interpretability of PBM through reliance on physical principles, it offers less flexibility than DDM-dominated fusion integration.

4.4.2. DDM dominated

In this category, the data-driven model dominates the physics-based model, where physics-based constraints guide the training of the data-driven model. Approaches that fall into this category include PiNN, physics-embedded neural networks, and neural operators. Physics-embedded neural networks explicitly embed the physics directly in the neural network architecture, whereas PiNNs incorporate the physics into the training process through loss functions without encoding it into the architecture. Since the architecture in PiNN is generic, they are more flexible. For instance, Yang et al. [41] proposed PiNN for solving the Navier–Stokes equation for fluid dynamics applications where the Navier–Stokes equations and their boundary and initial conditions are supplied as additional loss terms on top of the usual loss to train the feed-forward neural network. Fig. 7B illustrates this approach. The physics constrained losses: partial deferential loss (L_{pde}), initial condition loss ($L_{\text{initial condition}}$) and boundary condition ($L_{\text{boundary condition}}$) are weighted before added to the usual data loss (L_{data}). Here, the fusion model is dominated by a data-driven model, and the physics is incorporated as part of the loss formulation. Incorporating the physics helped the model to be more interpretable, as the resulting prediction from the neural network is constrained to respect the underlying physics.

5. Trade-offs and guidelines for choosing level of integration

5.1. DT modeling trade-offs

Based on further analysis of the four levels of integration, the trade-offs involved in selecting a specific level are assessed in terms of interpretability, training data requirements, generalization, and computational demands, as summarized in Table 6.

Level 0: Data-driven models are generally less interpretable, especially when DL models are used, and require significant amounts of training data [44,45,47,66]. They struggle with out-of-distribution data and have high computational costs during training, though inference costs are lower. In contrast, pure physics-based models are highly interpretable as white-box models, require no training data, and are highly generalizable when the underlying physics is accurate. However, complex physics-based models, such as energy simulation [57,86] and CFD models [92], have high inference costs, despite the low computational cost of model creation compared to pure DDM. Moreover, pure physics-based models typically require substantial time and manual effort to develop a model for a new building, due to the need to formulate appropriate physical relationships and perform thorough model validation. In contrast, data-driven models can often be adapted more rapidly by retraining existing models with data from the new building.

Level 1: When data flows from a physics-based to a data-driven model, interpretability is moderate due to the use of physics-generated training data. This physics-driven data augmentation also lowers data requirements [75]. These models generalize better

Table 6
Evaluation of integration levels focusing on interpretability, training data requirement, generalization, and computational requirements.

Integration level	Sub-category	Interpretability	Training data requirement	Generalization	Computational requirements.
Level 0	Purely DDM	Often low due to black-box model (especially for DL based models)	High (especially for DL based models)	Struggle with out-of-distribution data	Often high for training and low for inference
	Purely PBM	High (white-box model)	Not required	High as long as the underlying physics covers the process well	Often low for model creation but high for inference
Level 1	PBM to DDM	Moderate due to inclusion of physics generated training data (e.g. surrogate modeling, data augmentation)	Low (can work with less data due to inclusion of physics generated data)	Better than the pure DDM	Often high for training and low for inference
	DDM to PBM	Moderate (DDM generated data + white box model)	Moderate (It is required for the DDM part)	Moderate (DDM part still suffers from out of distribution data)	Often high (training time for DDM and inference time for PBM)
Level 2	Coupled	Moderate (The inputs of the physics-based white-box model are continuously generated by the black-box data-driven model)	Moderate (It is required for the DDM part)	Moderate (DDM part still suffers from out-of-distribution data)	Often high (training time for DDM and inference time for PBM)
Level 3	PBM-dominated	More interpretable due to physical constraints	Low	High as long as the underlying physics covers the process well	Low for training and high for inference
	DDM-dominated	Interpretable due to physics-based constraints within the DDM	Low (can work with less data due to inclusion of physics constraints within the DDM)	High (due to inclusion of physics constraints within the DDM)	Moderate for training and low for inference

than pure data-driven models, as the inclusion of physics-generated data helps capture the underlying data distribution. Moreover, since the final model is DDM-based, inference costs are lower. On the other hand, when data flows from data-driven to physics-based models, interpretability remains moderate, as incorporating DDM-generated data introduces a black-box element to the otherwise white-box model [80,83]. The data-driven component still requires training data, keeping overall data requirements at a moderate level [82–84]. While the final physics-based model generalizes well, the initial data-driven component struggles with out-of-distribution data. Additionally, since DDM training is computationally intensive and PBM inference also demands significant resources, the overall computational cost remains high.

Level 2: This level involves the continuous integration of data-driven and physics-based models into a coupled model, where data flows dynamically between the two. In cases where the black-box model continuously supplies input parameters to the white-box model (e.g., for control in MPC [90]), interpretability remains moderate, as incorporating DDM-generated data adds a black-box aspect to the otherwise interpretable PBM. The data-driven component still requires training data, keeping overall data requirements at a moderate level [86,87,90,91]. Additionally, the data-driven part continues to struggle with out-of-distribution data. The overall computational requirements remain high since DDM training is computationally intensive, and PBM inference also demands significant resources.

Level 3: Here, models are either PBM-dominated or DDM-dominated. PBM-dominated models are interpretable, require low data, and generalize well when the underlying physics is accurate due to the incorporation of physics into the model [42]. DDM-dominated models incorporate physics-based constraints in various ways, including the loss function formulation (e.g., PiNN [95]) and model architecture. In the loss function, physics-based terms are added to penalize violations of known physical laws. The model architecture can be designed to embed physical laws directly, such as by using neural network layers that solve differential equations or ensure conservation of mass, energy, or momentum. Since these physics constraints are embedded in the loss function, architecture, and other stages of the model, the final results adhere to physical principles, making these models interpretable [41]. Additionally, due to the extra information from the constraints, the amount of training data required to achieve acceptable performance is lower compared to pure DDM. This also results in reduced computational requirements during training. Furthermore, the inclusion of physics constraints enhances generalization performance, as out-of-distribution data is better captured through these constraints.

In summary, as the integration levels increase from Level 0 to Level 3, DDM becomes more interpretable, generalizable, and efficient, with reduced computational and data requirements.

5.2. Guidelines for choosing level of integration

The decision to select a specific integration level in the DT modeling for specific tasks requires balancing competing objectives across interpretability, computational demands, generalization, and other factors. The flowchart illustrating the decision

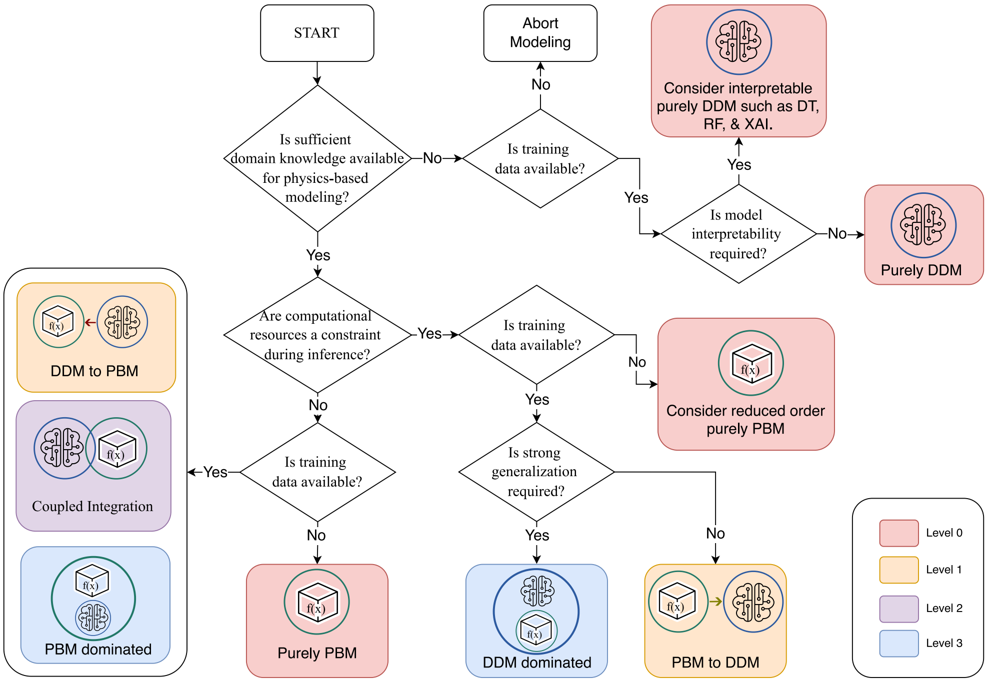

Fig. 8. Practical guideline for selecting the level of model integration.

methodology that balances these competing objectives is shown in Fig. 8. It presents the appropriate integration level and corresponding sub-category based on the availability of domain knowledge, interpretability requirements, data availability, computational constraints, and generalization needs.

The modeling approach selection process begins by assessing the availability of domain knowledge. When there is insufficient domain knowledge to develop a physics-based model but ample data is available to train a data-driven model, a purely DDM is the most appropriate. However, when model interpretability is important, various strategies should be considered to enhance the transparency of the data-driven models. These include selecting inherently interpretable ML models such as DTrees and RFs, as well as incorporating explainability techniques through XAI methods such as SHAP and LIME. In contrast, a pure physics-based model is suitable when sufficient domain knowledge is available, training data is scarce, and computational resources are not a constraint during inference. When computational efficiency is a concern, reduced-order physics-based models such as lumped parameter models and resistance-capacitance networks [43] can be considered, provided they are suitable for the problem at hand.

When both sufficient domain knowledge and adequate training data are available, two modeling strategies can be recommended, depending on inference-time computational constraints. In cases where inference-time constraints are not critical, the model can be developed using a PBM-centric approach enhanced by DDM. This enhancement can be achieved through sequential DDM-to-PBM data flow, continuous data flow via coupled integration, or a full fusion of both modeling paradigms, resulting in a hybrid PBM-dominated model. The choice among these integration strategies depends on the nature of the data flow, whether it is a one-time sequential transfer or a continuous real-time stream, as well as the desired level of model fusion within the PBM-dominated framework.

Conversely, when inference-time computational constraints are critical, a DDM-centric model is preferable due to its lower inference time. In this context, two integration strategies are suitable: (1) sequential integration from PBM to DDM and (2) DDM-dominated fusion integration. The first strategy, PBM-to-DDM, is effective for generating computationally efficient surrogate models that replace complex physics-based models. This approach also enables the incorporation of domain knowledge to enhance the generalization capability of the DDM, although the improvement may be less pronounced compared to the second strategy. The second strategy, DDM-dominated fusion, focuses on developing physics-constrained data-driven models that are not only efficient but also benefit substantially from embedded domain knowledge. This leads to improved interpretability and enhanced generalization performance. By incorporating key strengths of PBM, such as interpretability and generalization, into the DDM framework, this approach achieves a compelling balance between computational efficiency and interpretability.

6. Challenges and future directions

This section examines the challenges and limitations of the current DDM and PBM techniques, as well as their integration for building DTs.

Table 7
Categories of digital twin applications and the number of studies. **PBM-D:** PBM dominated, **DDM-D:** DDM dominated.

Application category	Number of studies	Conceptual	Level 0			Level 1		Level 2	Level 3	
			Pure PBM	Pure DDM	Both	*PBM → DDM*	*DDM → PBM*		PBM-D	DDM-D
Energy optimization and management	28	[5,29]	[39,56–58]	[47,51,54,55,65]	[60,61]	[75,76]		[58,86–94]	[42,43]	[41]
Facility management	19	[97–103]	[38,59,74]	[48,50,70–73]		[77,79]	[85]			
Construction management	10	[6,104–112]								
Building sustainability	10	[8,113–120]		[53]						
Building geometry modeling	7			[44,45]		[78]	[80–82,84]			
Indoor environmental quality	6	[13]		[63,66–68]	[62]					
Smart home and City management	6	[121–123]		[46,49]	[40]					
Fault detection and Diagnosis	3	[124]		[52,69]						
Structural health monitoring	2			[64]			[83]			

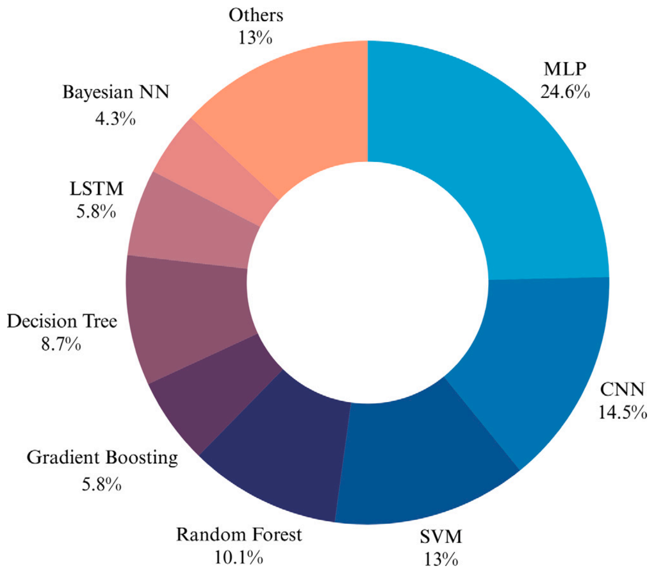

Fig. 9. Different ML models and the number of studies using them.

Narrow Application Focus: The reviewed literature reveals a concentration of DT applications in the building sector around a few dominant areas, particularly energy optimization and management, and facility management, with other domains such as smart home and city management, fault detection, and structural health monitoring receiving comparatively limited attention (Table 7). Future research is needed to expand the application of DT in these underexplored domains. Specifically, further studies could examine DT-enabled automated fault detection and predictive maintenance in residential and commercial buildings [52,69], and advance structural health monitoring for building safety through DT-driven DDM-based approaches [83].

Underexplored Integration Levels: A similar pattern in study distribution is observed for levels of integration, consistent with the distribution across application domains (Table 7). Most studies remain at lower levels or adopt siloed modeling approaches, while integration at higher levels, especially Levels 2 and 3, is relatively rare. This highlights the need for more research on the integration of PBM and DDM in various areas of DT application, such as fault detection and diagnosis, indoor environmental quality, structural health monitoring, and construction management.

Dominance of Low-interpretability Models: Most of the reviewed studies rely heavily on the models that are difficult to interpret, with MLP [50,53,54,58,60,61,66,68,76,87,90], CNN [46,64,66,70,83,84], and LSTM [40,49,66,77,89] together accounting for nearly half of the studies as shown in Fig. 9. Moderately interpretable models such as SVM [46,50–52,54,68,73] and gradient boosting methods [40,68] are also common. In contrast, more interpretable models such as DTree [44,52,54,55,68] are relatively infrequent. The widespread reliance on less interpretable deep learning models presents a significant challenge. While these models deliver strong predictive performance, their lack of transparency limits their applicability in contexts where understanding model behavior is essential, particularly in safety–critical building DT applications. Despite this, concerns about interpretability are rarely addressed. Future research should be conducted to develop various methods for integrating interpretability into DT modeling, such as employing ML explainability techniques (e.g., SHAP and LIME) or incorporating physical principles into DDM through sequential and fusion integrations.

Data Scarcity and the Overlooked Role of Physics-Based Models: Data-driven models require large, diverse datasets to effectively learn the underlying data distribution of a given problem. However, collecting data that fully captures this distribution is often difficult, especially for new or recently retrofitted buildings where operational data is scarce or unavailable. A potential solution to this challenge is to augment existing training data or generate synthetic data for new buildings using physics-based models to represent under-sampled or unseen conditions. This approach enables the creation of initial models for new buildings by generating synthetic data when real data is unavailable, and also enhances model robustness in existing buildings by augmenting training data with rare or extreme scenarios. Despite its potential, there is limited research utilizing physics-based models for data augmentation, especially in generating data not typically captured by sensor systems. This includes anomalies for fault detection, edge cases for safety–critical scenarios, and rare operational states such as HVAC system failures. These gaps underscore the need to investigate further how physics-based models can enhance data-driven methods in contexts where real-world data is scarce, unavailable, or unsafe to obtain.

Challenges in Modeling Complex PBM Components: PBM aims to represent the relationships between system components using physical equations derived from fundamental principles. However, in practice, not all model components may be fully characterized by physics, and simplifications or approximations are often used. DDM offers a promising solution by estimating these difficult-to-model components from data. For example, O'Dwyer et al. [90] demonstrated that modeling the dynamic energy demand component within the mathematical system dynamics for model predictive control is challenging using physics-based state-space models. The authors employed an artificial neural network to model energy demand from historical data, incorporating it into the overall system dynamics model. Despite these applications, there is a lack of studies in the literature that leverage this approach to model complex components within physics-based models. Future research should focus on advancing the integration of data-driven and physics-based modeling methods to more accurately represent complex components in physics-based models.

Challenges in Fusion Integration Modeling: Fusion integration-based models leverage the strengths of both PBM and DDM. These models combine the interpretability of physics-based models with the efficiency of data-driven models. Integrating physics into data-driven models enhances their ability to generalize to unseen scenarios during training. Although the fusion integration of data-driven and physics-based modeling offers the benefits discussed above, there is a lack of studies that utilize fusion integration-level models for DT modeling in the reviewed articles. More research is needed to address key challenges in aligning physics-based and data-driven representations through novel model architectures that prevent training instability when integrating data-driven components while enforcing physical consistency, preserving interpretability within unified models while achieving the inference speed offered by data-driven approaches, and developing comprehensive frameworks and tools that support the model fusion process.

Challenges in Multimodal Data Integration: The development of comprehensive building digital twins requires the integration of heterogeneous data sources, including imagery, semantic information, point clouds, sensor measurements, and building information models (BIM), to enhance completeness and accuracy. However, effectively fusing these multimodal data streams while preserving each modality's unique characteristics and information content presents significant challenges. Future research should focus on developing robust data fusion algorithms that can handle varying data quality, resolution, and temporal alignment across different modalities.

Modeling Challenges and the Role of Large Language Models: Recent advancements in Large Language Models (LLMs) present promising opportunities for enhancing integration between physics-based and data-driven modeling paradigms. LLMs can potentially be used for automating the generation of physics-based model through natural language processing of domain knowledge and improving reasoning and explainability by providing semantic interpretations of complex model behaviors. Future research should explore how LLMs can be systematically integrated into BDT modeling workflows, potentially revolutionizing how domain expertise is encoded and utilized in building digital twin development.

7. Conclusion

This review examined recent advancements in building digital twin (DT) modeling, with a focus on data-driven, physics-based approaches and their integration (hybrid modeling). It introduced a four-level classification of data-driven and physics-based integration by analyzing the degree and nature of integration across existing studies, as follows: Level 0 (siloed models), Level 1 (sequential integration), Level 2 (coupled integration), and Level 3 (fusion integration), representing progressively deeper integration. In addition, the review proposed a decision-making guideline for selecting an appropriate integration level based on key trade-offs, including interpretability, training data requirements, computational demands, domain knowledge availability, and generalization. These trade-offs shape the suitability of each integration level for various building DT application contexts. For example, physics-based models excel in interpretability and generalization, while data-driven models offer strong predictive performance and computational efficiency at inference time. Combining these complementary strengths through hybrid modeling is essential, with the benefits becoming more pronounced as integration progresses from data-level (sequential and coupled) to model-level (fusion) approaches.

Despite the potential of hybrid modeling, the review uncovered that most existing studies remain at the siloed or sequential levels, with limited exploration of more advanced coupled or fusion-level approaches. Integration efforts are also heavily concentrated in a few dominant application areas, such as energy optimization and facility management. Moreover, widespread reliance on low-interpretability data-driven models limits their applicability in safety–critical DT scenarios where model transparency is essential for decision-making. Data scarcity, particularly in newly constructed or under-instrumented buildings with limited operational history, further constrains model development and validation. This scarcity impedes the training of robust data-driven models, especially for rare or anomalous conditions that are critical in practice. Addressing these limitations requires future research to advance hybrid modeling approaches for building digital twins that are interpretable, computationally efficient, and generalizable.

CRediT authorship contribution statement

Jifar M. Hunde: Writing – review & editing, Writing – original draft, Visualization, Methodology, Conceptualization. **Tesfatsyon S. Ochono:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Conceptualization. **Damitha Senevirathne:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Conceptualization. **Dagimawi D. Eneyew:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Conceptualization. **Girma T. Bitsuamlak:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Miriam A.M. Capretz:** Writing – review & editing, Supervision, Methodology, Funding acquisition, Conceptualization. **Katarina Grolinger:** Writing – review & editing, Supervision, Resources, Methodology, Funding acquisition, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT in order to improve the readability and language of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the Climate Action and Awareness Fund (CAAF) from Environment and Climate Change Canada, Canada, and by the Carbon Solutions grant at Western University, Canada.

Data availability

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

References

[1] U.N.E. Programme, 2022 Global Status Report for Buildings and Construction: Towards a Zero-Emission, Efficient and Resilient Buildings and Construction Sector, Tech. Rep., United Nations Environment Programme, Nairobi, 2022.
[2] M. Grieves, J. Vickers, Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems, in: Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches, 2017, pp. 85–113.
[3] H. Omrany, K.M. Al-Obaidi, A. Husain, A. Ghaffarianhoseini, Digital twins in the construction industry: A comprehensive review of current implementations, enabling technologies, and future directions, Sustain. 15 (14) (2023) 10908, http://dx.doi.org/10.3390/su151410908.
[4] Z. Ghorbani, J. Messner, D. Basurto, R. Dicher, M. Escobar, M. Farquhar, A. Flak, J. Fortune, M. Goldman, R. Grant, P. Lane-Nott, A. Mabrich, F. Maier, S. McClure, I. McGaw, B. Meinert, K. Onuma, N. Reyes, N. Saleeb, J. Walter, Digital Twins for the Built Environment: A Position Paper on Integrating Bim and Digital Twin, Tech. Rep., National Institute of Building Sciences, Digital Technology Council, 2024.
[5] S. Yoon, Building digital twinning: Data, information, and models, J. Build. Eng. 76 (2023) 107021, http://dx.doi.org/10.1016/j.jobe.2023.107021.
[6] M. Shahzad, M.T. Shafiq, D. Douglas, M. Kassem, Digital twins in built environments: An investigation of the characteristics, applications, and challenges, Buildings 12 (2) (2022) http://dx.doi.org/10.3390/buildings12020120.
[7] S.M.E. Sepasgozar, Differentiating digital twin from digital shadow: Elucidating a paradigm shift to expedite a smart, sustainable built environment, Buildings 11 (4) (2021) http://dx.doi.org/10.3390/buildings11040151.
[8] I. Yitmen, S. Alizadehsalehi, I. Akiner, M.E. Akiner, An adapted model of cognitive digital twins for building lifecycle management, Appl. Sci.-Basel 11 (9) (2021) http://dx.doi.org/10.3390/app11094276.
[9] Q. Qi, F. Tao, Digital twin and big data towards smart manufacturing and industry 4.0: 360 degree comparison, Ieee Access 6 (2018) 3585–3593.
[10] A. Rasheed, O. San, T. Kvamsdal, Digital twin: values, challenges and enablers from a modeling perspective, IEEE Access 8 (2020) 21980–22012, http://dx.doi.org/10.1109/ACCESS.2020.2970143.
[11] A. Thelen, X. Zhang, O. Fink, Y. Lu, S. Ghosh, B.D. Youn, M.D. Todd, S. Mahadevan, C. Hu, Z. Hu, A comprehensive review of digital twin – part 1: Modeling and twinning enabling technologies, 2022, http://dx.doi.org/10.48550/arXiv.2208.14197, arXiv:2208.14197[cs].
[12] V. Drobnyi, Z. Hu, Y. Fathy, I. Brilakis, Construction and maintenance of building geometric digital twins: State of the art review, Sensors 23 (9) (2023) http://dx.doi.org/10.3390/s23094382.
[13] S. Karatzas, G. Papageorgiou, V. Lazari, S. Bersimis, A. Fousteris, P. Economou, A. Chassiakos, A text analytic framework for gaining insights on the integration of digital twins and machine learning for optimizing indoor building environmental performance, Dev. Built Environ. 18 (2024) http://dx.doi.org/10.1016/j.dibe.2024.100386.
[14] Y. Junjia, A.H. Alias, N.A. Haron, N.A. Bakar, A bibliometrics-based systematic review of safety risk assessment for IBS hoisting construction, Buildings 13 (7) (2023) http://dx.doi.org/10.3390/buildings13071853.
[15] S. Siccardi, V. Villa, Trends in adopting BIM, IoT and DT for facility management: A scientometric analysis and keyword co-occurrence network review, Buildings 13 (1) (2023) http://dx.doi.org/10.3390/buildings13010015.
[16] H. Sadri, I. Yitmen, L.C. Tagliabue, F. Westphal, A. Tezel, A. Taheri, G. Sibenik, Integration of blockchain and digital twins in the smart built environment adopting disruptive technologies—a systematic review, Sustain. (Switz.) 15 (4) (2023) http://dx.doi.org/10.3390/su15043713.
[17] F.N. Abdeen, S. Shirowzhan, S.M.E. Sepasgozar, Citizen-centric digital twin development with machine learning and interfaces for maintaining urban infrastructure, Telemat. Inform. 84 (2023) http://dx.doi.org/10.1016/j.tele.2023.102032.

[18] A. Almusaed, I. Yitmen, A. Almssad, Enhancing smart home design with AI models: A case study of living spaces implementation review, Energies 16 (6) (2023) http://dx.doi.org/10.3390/en16062636.

[19] T.I. Zohdi, A digital-twin and machine-learning framework for ventilation system optimization for capturing infectious disease respiratory emissions, Arch. Comput. Methods Eng. 28 (6, SI) (2021) 4317–4329, http://dx.doi.org/10.1007/s11831-021-09609-3.

[20] H.H. Hosamo, M.H. Hosamo, Digital twin technology for bridge maintenance using 3D laser scanning: A review, Adv. Civ. Eng. 2022 (1) (2022) 2194949, http://dx.doi.org/10.1155/2022/2194949.

[21] F. Hodavand, I.J. Ramaji, N. Sadeghi, Digital twin for fault detection and diagnosis of building operations: A systematic review, Buildings 13 (6) (2023) http://dx.doi.org/10.3390/buildings13061426.

[22] H.H. Hosamo, H.K. Nielsen, A.N. Alnmr, P.R. Svennevig, K. Svidt, A review of the digital twin technology for fault detection in buildings, Front. Built Environ. 8 (2022) http://dx.doi.org/10.3389/fbuil.2022.1013196.

[23] M. Manfren, K.M. Gonzalez-Carreon, P.A.B. James, Interpretable data-driven methods for building energy modelling—a review of critical connections and gaps, Energies 17 (4) (2024) http://dx.doi.org/10.3390/en17040881.

[24] F. Tahmasebinia, L. Lin, S. Wu, Y. Kang, S. Sepasgozar, Exploring the benefits and limitations of digital twin technology in building energy, Appl. Sci. (Switz.) 13 (15) (2023) http://dx.doi.org/10.3390/app13158814.

[25] V.A. Arowoiya, R.C. Moehler, Y. Fang, Digital twin technology for thermal comfort and energy efficiency in buildings: A state-of-the-art and future directions, Energy Built Environ. 5 (5) (2024) 641–656, http://dx.doi.org/10.1016/j.enbenv.2023.05.004.

[26] O. Hakimi, H. Liu, O. Abudayyeh, Digital twin-enabled smart facility management: A bibliometric review, Front. Eng. Manag. 11 (1) (2024) 32–49, http://dx.doi.org/10.1007/s42524-023-0254-4.

[27] G. Piras, S. Agostinelli, F. Muzi, Digital twin framework for built environment: A review of key enablers, Energies 17 (2) (2024) http://dx.doi.org/10.3390/en17020436.

[28] M.B. Shishehgarkhaneh, A. Keivani, R.C. Moehler, N. Jelodari, S.R. Laleh, Internet of things (IoT), building information modeling (BIM), and digital twin (DT) in construction industry: A review, bibliometric, and network analysis, Buildings 12 (10) (2022) http://dx.doi.org/10.3390/buildings12101503.

[29] K. Alanne, S. Sierla, An overview of machine learning applications for smart buildings, Sustain. Cities Soc. 76 (2022) http://dx.doi.org/10.1016/j.scs.2021.103445.

[30] C. Boje, A. Guerriero, S. Kubicki, Y. Rezgui, Towards a semantic construction digital twin: Directions for future research, Autom. Constr. 114 (2020) http://dx.doi.org/10.1016/j.autcon.2020.103179.

[31] Z. Zhang, Z. Wei, S. Court, L. Yang, S. Wang, A. Thirunavukarasu, Y. Zhao, A review of digital twin technologies for enhanced sustainability in the construction industry, Buildings 14 (4) (2024) http://dx.doi.org/10.3390/buildings14041113.

[32] L. Kannari, N. Wessberg, S. Hirvonen, J. Kantorovitch, S. Paiho, Reinforcement learning for control and optimization of real buildings: Identifying and addressing implementation hurdles, J. Build. Eng. 104 (2025) http://dx.doi.org/10.1016/j.jobe.2025.112283.

[33] A.C. Olivas, J.-J. Yee, Towards the use of data-driven methods for indoor airflow field reconstruction: A systematic review, J. Build. Eng. 106 (2025) 112581, http://dx.doi.org/10.1016/j.jobe.2025.112581, URL https://www.sciencedirect.com/science/article/pii/S2352710225008186.

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

[35] M. Grieves, Digital Twin: Manufacturing Excellence Through Virtual Factory Replication, White paper, Florida Institute of Technology, 2014.

[36] Q. Qi, F. Tao, T. Hu, N. Anwer, A. Liu, Y. Wei, L. Wang, A.Y.C. Nee, Enabling technologies and tools for digital twin, J. Manuf. Syst. 58 (2021) 3–21, http://dx.doi.org/10.1016/j.jmsy.2019.10.001, URL https://www.sciencedirect.com/science/article/pii/S027861251930086X.

[37] D.D. Eneyew, M.A.M. Capretz, G.T. Bitsuamlak, Toward smart-building digital twins: BIM and IoT data integration, IEEE Access 10 (2022) 130487–130506, http://dx.doi.org/10.1109/ACCESS.2022.3229370.

[38] K. Walther, M. Molinari, K. Voss, Controls of hvac systems in digital twins–comparative framework and case study on the performance gap, J. Build. Perform. Simul. (2025) http://dx.doi.org/10.1080/19401493.2024.2446517.

[39] Y. Choi, S. Yoon, In-situ observation virtual sensor in building systems toward virtual sensing-enabled digital twins, Energy Build. 281 (2023) 112766, http://dx.doi.org/10.1016/j.enbuild.2022.112766.

[40] E.M. Elfarri, A. Rasheed, O. San, Artificial intelligence-driven digital twin of a modern house demonstrated in virtual reality, IEEE Access Pr. Innov. Open Solut. 11 (2023) 35035–35058, http://dx.doi.org/10.1109/ACCESS.2023.3265191.

[41] S. Yang, H. Kim, Y. Hong, K. Yee, R. Maulik, N. Kang, Data-driven physics-informed neural networks: A digital twin perspective, Comput. Methods Appl. Mech. Engrg. 428 (2024) 117075, http://dx.doi.org/10.1016/j.cma.2024.117075.

[42] Y. Choi, S. Yoon, In-situ backup virtual sensor application in building automation systems toward virtual sensing-enabled digital twins, Case Stud. Therm. Eng. 66 (2025) 105792, http://dx.doi.org/10.1016/j.csite.2025.105792.

[43] Y. Yang, M. Jradi, Adaptive grey-box modelling for energy-efficient building retrofits: Case studies in denmark, Sustain. (Switz.) 17 (2025) http://dx.doi.org/10.3390/su17041702.

[44] M. Mehranfar, A. Braun, A. Borrmann, From dense point clouds to semantic digital models: End-to-end AI-based automation procedure for manhattan-world structures, Autom. Constr. 162 (2024) http://dx.doi.org/10.1016/j.autcon.2024.105392.

[45] X. Pan, Q. Lin, S. Ye, L. Li, L. Guo, B. Harmon, Deep learning based approaches from semantic point clouds to semantic BIM models for heritage digital twin, Herit. Sci. 12 (1) (2024) http://dx.doi.org/10.1186/s40494-024-01179-4.

[46] L. Adreani, P. Bellini, C. Colombo, M. Fanfani, P. Nesi, G. Pantaleo, R. Pisanu, Implementing integrated digital twin modelling and representation into the Snap4City platform for smart city solutions, Multimedia Tools Appl. (2023) http://dx.doi.org/10.1007/s11042-023-16838-0.

[47] M. Both, B. Kaemper, A. Cartus, J. Beermann, T. Fessler, J. Mueller, C. Diedrich, Automated monitoring applications for existing buildings through natural language processing based semantic mapping of operational data and creation of digital twins, Energy Build. 300 (2023) http://dx.doi.org/10.1016/j.enbuild.2023.113635.

[48] M. Antonino, M. Nicola, D.M. Claudio, B. Luciano, R.C. Fulvio, Office building occupancy monitoring through image recognition sensors, Int. J. Saf. Secur. Eng. 9 (4) (2019) 371–380, http://dx.doi.org/10.2495/SAFE-V9-N4-371-380.

[49] D. Bouchabou, J. Grosset, S.M. Nguyen, C. Lohr, X. Puig, A smart home digital twin to support the recognition of activities of daily living, Sensors 23 (17) (2023) http://dx.doi.org/10.3390/s23177586.

[50] A.M. Abdelalim, A. Essawy, A. Sherif, M. Salem, M. Al-Adwani, M.S. Abdullah, Optimizing facilities management through artificial intelligence and digital twin technology in mega-facilities, Sustain. (Switz.) 17 (2025) http://dx.doi.org/10.3390/su17051826.

[51] H.-A. Park, G. Byeon, W. Son, J. Kim, S. Kim, Data-driven modeling of HVAC systems for operation of virtual power plants using a digital twin, Energies 16 (20) (2023) http://dx.doi.org/10.3390/en16207032.

[52] H.H. Hosamo, P.R. Svennevig, K. Svidt, D. Han, H.K. Nielsen, A digital twin predictive maintenance framework of air handling units based on automatic fault detection and diagnostics, Energy Build. 261 (2022) http://dx.doi.org/10.1016/j.enbuild.2022.111988.

[53] R. Tariq, J.A. Recio-Garcia, A.J. Cetina-Quiñones, M.G. Orozco-Del-Castillo, A. Bassam, Explainable artificial intelligence twin for metaheuristic optimization: Double-skin facade with energy storage in buildings, J. Comput. Des. Eng. 12 (2025) 16–35, http://dx.doi.org/10.1093/jcde/qwaf015.

[54] A.N. Sayed, F. Bensaali, Y. Himeur, G. Dimitrakopoulos, I. Varlamis, Enhancing building sustainability: A digital twin approach to energy efficiency and occupancy monitoring, Energy Build. 328 (2025) http://dx.doi.org/10.1016/j.enbuild.2024.115151.
[55] M. Jradi, B.E. Madsen, J.H. Kaiser, DanRETwin: A digital twin solution for optimal energy retrofit decision-making and decarbonization of the danish building stock, Appl. Sci. (Switz.) 13 (17) (2023) http://dx.doi.org/10.3390/app13179778.
[56] D. Bayer, M. Pruckner, A digital twin of a local energy system based on real smart meter data, Energy Inform. 6 (1) (2023) http://dx.doi.org/10.1186/s42162-023-00263-6.
[57] A. Borodinecs, A. Palcikovskis, A. Krumins, K. Lebedeva, Optimizing office building operations: A framework for continuous dynamic energy simulations in decision-making for efficiency, Front. Built Environ. 10 (2024) http://dx.doi.org/10.3389/fbuil.2024.1405182.
[58] S. Yoon, J. Koo, In situ model fusion for building digital twinning, Build. Environ. 243 (2023) 110652, http://dx.doi.org/10.1016/j.buildenv.2023.110652.
[59] R. Ward, R. Choudhary, A. Gregory, M. Jans-Singh, M. Girolami, Continuous calibration of a digital twin: Comparison of particle filter and Bayesian calibration approaches, Data-Centric Eng. 2 (2021) http://dx.doi.org/10.1017/dce.2021.12.
[60] S. Yoon, Y. Choi, J. Koo, In situ virtual sensors in building digital twins: Framework and methodology, J. Ind. Inf. Integr. 36 (2023) 100532, http://dx.doi.org/10.1016/j.jii.2023.100532.
[61] J. Koo, S. Yoon, Simultaneous in-situ calibration for physical and virtual sensors towards digital twin-enabled building operations, Adv. Eng. Inform. 59 (2024) 102239, http://dx.doi.org/10.1016/j.aei.2023.102239.
[62] A. Arsiwala, F. Elghaish, M. Zoher, Digital twin with machine learning for predictive monitoring of CO2 equivalent from existing buildings, Energy Build. 284 (2023) http://dx.doi.org/10.1016/j.enbuild.2023.112851.
[63] M.M. Abdelrahman, A. Chong, C. Miller, Personal thermal comfort models using digital twins: Preference prediction with BIM-extracted spatial–temporal proximity data from Build2Vec, Build. Environ. 207 (B) (2022) http://dx.doi.org/10.1016/j.buildenv.2021.108532.
[64] S. Luo, H. Wang, Digital twin research on masonry–timber architectural heritage pathology cracks using 3D laser scanning and deep learning model, Buildings 14 (4) (2024) http://dx.doi.org/10.3390/buildings14041129.
[65] Y. Choi, S. Yoon, Enhancing in-situ model accuracy in operational building systems with augmentation-based synthetic operational data, J. Build. Eng. 101 (2025) 111623, http://dx.doi.org/10.1016/j.jobe.2024.111623.
[66] P. Norouzi, S. Maalej, R. Mora, Applicability of deep learning algorithms for predicting indoor temperatures: Towards the development of digital twin HVAC systems, Buildings 13 (6) (2023) http://dx.doi.org/10.3390/buildings13061542.
[67] Z. ElArwady, A. Kandil, M. Afiffy, M. Marzouk, Modeling indoor thermal comfort in buildings using digital twin and machine learning, Dev. Built Environ. 19 (2024) http://dx.doi.org/10.1016/j.dibe.2024.100480.
[68] H.H. Hosamo, H.K. Nielsen, D. Kraniotis, P.R. Svennevig, K. Svidt, Improving building occupant comfort through a digital twin approach: A Bayesian network model and predictive maintenance method, Energy Build. 288 (2023) http://dx.doi.org/10.1016/j.enbuild.2023.112992.
[69] X. Xie, J. Merino, N. Moretti, P. Pauwels, J.Y. Chang, A. Parlikad, Digital twin enabled fault detection and diagnosis process for building HVAC systems, Autom. Constr. 146 (2023) http://dx.doi.org/10.1016/j.autcon.2022.104695.
[70] K.E. Mokhtari, I. Panushev, J.J. McArthur, Development of a cognitive digital twin for building management and operations, Front. Built Environ. 8 (2022) http://dx.doi.org/10.3389/fbuil.2022.856873.
[71] S. Bao, H. Bu, Defining and generating operation and maintenance management requirements in digital twin applications using the DT-GPT framework, J. Build. Eng. 104 (2025) 112356, http://dx.doi.org/10.1016/j.jobe.2025.112356, URL https://www.sciencedirect.com/science/article/pii/S2352710225005935.
[72] S. Yoon, J. Song, J. Li, Ontology-enabled ai agent-driven intelligent digital twins for building operations and maintenance, J. Build. Eng. 108 (2025) 112802, URL https://www.sciencedirect.com/science/article/pii/S2352710225010393.
[73] Y. Zhao, N. Wang, Z. Liu, E. Mu, Construction theory for a building intelligent operation and maintenance system based on digital twins and machine learning, Buildings 12 (2) (2022) http://dx.doi.org/10.3390/buildings12020087.
[74] S. Meschini, L. Pellegrini, M. Locatelli, D. Accardo, L.C. Tagliabue, G.M.D. Giuda, M. Avena, Toward cognitive digital twins using a BIM-GIS asset management system for a diffused university, Front. Built Environ. 8 (2022) http://dx.doi.org/10.3389/fbuil.2022.959475.
[75] F. Tahmasebinia, L. Lin, S. Wu, Y. Kang, S. Sepesgozar, Advanced energy performance modelling: Case study of an engineering and technology precinct, Buildings 14 (6) (2024) http://dx.doi.org/10.3390/buildings14061774.
[76] Y. Choi, S. Yoon, Surrogate-assisted high-accuracy observation modeling in building digital twins: In situ nonintrusive modeling without sensor observation (Y), Build. Environ. 242 (2023) 110584, http://dx.doi.org/10.1016/j.buildenv.2023.110584.
[77] W. Xie, Y. Zeng, X. Zhang, H.Y. Wong, T. Zhang, Z. Wang, X. Wu, J. Shi, X. Huang, F. Xiao, A. Usmani, Aiot-powered building digital twin for smart firefighting and super real-time fire forecast, Adv. Eng. Inform. 65 (2025) http://dx.doi.org/10.1016/j.aei.2025.103117.
[78] M.B. Maru, Y. Wang, H. Kim, H. Yoon, S. Park, Improved building facade segmentation through digital twin-enabled RandLA-net with empirical intensity correction model, J. Build. Eng. 78 (2023) http://dx.doi.org/10.1016/j.jobe.2023.107520.
[79] X. Luo, A. Kareem, S. Yoo, Optimal sensor placement for reconstructing wind pressure field around buildings using compressed sensing, J. Build. Eng. 75 (2023) http://dx.doi.org/10.1016/j.jobe.2023.106855.
[80] Y. Pan, A. Braun, A. Borrmann, I. Brilakis, 3D deep-learning-enhanced void-growing approach in creating geometric digital twins of buildings, Proc. Inst. Civ. Eng. Smart Infrastruct. Constr. 176 (1) (2022) 24–40, http://dx.doi.org/10.1680/jsmic.21.00035.
[81] Y. Pan, A. Braun, I. Brilakis, A. Borrmann, Enriching geometric digital twins of buildings with small objects by fusing laser scanning and AI-based image recognition, Autom. Constr. 140 (2022) http://dx.doi.org/10.1016/j.autcon.2022.104375.
[82] B.G. Pantoja-Rosero, R. Achanta, M. Kozinski, P. Fua, F. Perez-Cruz, K. Beyer, Generating LOD3 building models from structure-from-motion and semantic segmentation, Autom. Constr. 141 (2022) http://dx.doi.org/10.1016/j.autcon.2022.104430.
[83] D. Loverdos, V. Sarhosis, Image2DEM: A geometrical digital twin generator for the detailed structural analysis of existing masonry infrastructure stock, SoftwareX 22 (2023) http://dx.doi.org/10.1016/j.softx.2023.101323.
[84] B.G. Pantoja-Rosero, R. Achanta, K. Beyer, Damage-augmented digital twins towards the automated inspection of buildings, Autom. Constr. 150 (2023) http://dx.doi.org/10.1016/j.autcon.2023.104842.
[85] A. Mukhopadhyay, G.S.R. Reddy, K.S. Saluja, S. Ghosh, A. Peña-Rios, G. Gopal, P. Biswas, Virtual-reality-based digital twin of office spaces with social distance measurement feature, Virtual Real. Intell. Hardw. 4 (1) (2022) 55–75, http://dx.doi.org/10.1016/j.vrih.2022.01.004.
[86] D.D. Eneyew, M.A.M. Capretz, G.T. Bitsuamlak, Continuous model calibration framework for smart-building digital twin: A generative model-based approach, Appl. Energy 375 (2024) 124080, http://dx.doi.org/10.1016/j.apenergy.2024.124080.
[87] S. Yoon, J. Lee, J. Li, P. Wang, Virtual in-situ modeling between digital twin and BIM for advanced building operations and maintenance, Autom. Constr. 168 (A) (2024) http://dx.doi.org/10.1016/j.autcon.2024.105823.
[88] Y. Li, J. Lee, J. Li, P. Wang, S. Yoon, Nonintrusive in-situ modeling for unobserved virtual models in digital twin-enabled building HVAC systems: A one-year comparison of data-driven and physics-based approaches in a living laboratory, J. Build. Eng. 101 (2025) 111811, http://dx.doi.org/10.1016/j.jobe.2025.111811.
[89] J. Bjørnskov, M. Jradi, An ontology-based innovative energy modeling framework for scalable and adaptable building digital twins, Energy Build. 292 (2023) http://dx.doi.org/10.1016/j.enbuild.2023.113146.

[90] E. O'Dwyer, I. Pan, R. Charlesworth, S. Butler, N. Shah, Integration of an energy management tool and digital twin for coordination and control of multi-vector smart energy systems, Sustain. Cities Soc. 62 (2020) http://dx.doi.org/10.1016/j.scs.2020.102412.

[91] J.A. Borja-Conde, K. Witheephanich, J.F. Coronel, D. Limon, Thermal modeling of existing buildings in high-fidelity simulators: A novel, Pr. Methodol. Energy Build. 292 (2023) http://dx.doi.org/10.1016/j.enbuild.2023.113127.

[92] S. Agostinelli, F. Cumo, G. Guidi, C. Tomazzoli, Cyber-physical systems improving building energy management: Digital twin and artificial intelligence, Energies 14 (8) (2021) http://dx.doi.org/10.3390/en14082338.

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

[94] A. Campoy-Nieves, A. Manjavacas, J. Jiménez-Raboso, M. Molina-Solana, J. Gómez-Romero, Sinergym – a virtual testbed for building energy optimization with reinforcement learning, Energy Build. 327 (2025) http://dx.doi.org/10.1016/j.enbuild.2024.115075.

[95] M. Raissi, P. Perdikaris, G.E. Karniadakis, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, J. Comput. Phys. 378 (2019) 686–707, http://dx.doi.org/10.1016/j.jcp.2018.10.045.

[96] S.A. Faroughi, N.M. Pawar, C. Fernandes, M. Raissi, S. Das, N.K. Kalantari, S. Kourosh Mahjour, Physics-guided, physics-informed, and physics-encoded neural networks and operators in scientific computing: Fluid and solid mechanics, J. Comput. Inf. Sci. Eng. 24 (040802) (2024) http://dx.doi.org/10.1115/1.4064449.

[97] L. Rampini, F.R. Cecconi, Artificial intelligence in construction asset management: a review of present status, challenges and future opportunities, J. Inf. Technol. Constr. 27 (2022) http://dx.doi.org/10.36680/j.itcon.2022.043.

[98] A. Bujari, A. Calvio, L. Foschini, A. Sabbioni, A. Corradi, A digital twin decision support system for the urban facility management process, Sensors 21 (24) (2021) http://dx.doi.org/10.3390/s21248460.

[99] K. Bäcklund, P. Lundqvist, M. Molinari, Showcasing a digital twin for higher educational buildings: developing the concept toward human centricity, Front. Built Environ. 10 (2024) http://dx.doi.org/10.3389/fbuil.2024.1347451.

[100] Q. Lu, X. Xie, A.K. Parlikad, J.M. Schooling, E. Konstantinou, Moving from building information models to digital twins for operation and maintenance, Proc. Inst. Civ. Eng. Smart Infrastruct. Constr. 174 (2) (2020) 46–56, http://dx.doi.org/10.1680/jsmic.19.00011.

[101] M. Almatared, H. Liu, O. Abudayyeh, O. Hakim, M. Sulaiman, Digital-twin-based fire safety management framework for smart buildings, Buildings 14 (1) (2024) http://dx.doi.org/10.3390/buildings14010004.

[102] Z. Liu, S. Gong, Z. Tan, P. Demian, Immersive technologies-driven building information modeling (BIM) in the context of metaverse, Buildings 13 (6) (2023) http://dx.doi.org/10.3390/buildings13061559.

[103] H. Olimat, H. Liu, O. Abudayyeh, Enabling technologies and recent advancements of smart facility management, Buildings 13 (6) (2023) http://dx.doi.org/10.3390/buildings13061488.

[104] R. Sacks, I. Brilakis, E. Pikas, H.S. Xie, M. Girolami, Construction with digital twin information systems, Data-Centric Eng. 1 (6) (2020) http://dx.doi.org/10.1017/dce.2020.16.

[105] S. Jaryani, I. Yitmen, H. Sadri, S. Alizadehsalehi, Exploring the fusion of knowledge graphs into cognitive modular production, Buildings 13 (9) (2023) http://dx.doi.org/10.3390/buildings13092306.

[106] C.J. Turner, J. Oyekan, L. Stergioulas, D. Griffin, Utilizing industry 4.0 on the construction site: Challenges and opportunities, IEEE Trans. Ind. Inform. 17 (2) (2021) 746–756, http://dx.doi.org/10.1109/TII.2020.3002197.

[107] T. Salem, M. Dragomir, Options for and challenges of employing digital twins in construction management, Appl. Sci. (Switz.) 12 (6) (2022) http://dx.doi.org/10.3390/app12062928.

[108] Z. You, L. Feng, Integration of industry 4.0 related technologies in construction industry: A framework of cyber–physical system, IEEE Access Pr. Innov. Open Solut. 8 (2020) 122908–122922, http://dx.doi.org/10.1109/ACCESS.2020.3007206.

[109] M. Kor, I. Yitmen, S. Alizadehsalehi, An investigation for integration of deep learning and digital twins towards construction 4.0, Smart Sustain. Built Environ. 12 (3) (2023) 461–487, http://dx.doi.org/10.1108/SASBE-08-2021-0148.

[110] A. Pal, J.J. Lin, S.-H. Hsieh, M. Golparvar-Fard, Automated vision-based construction progress monitoring in built environment through digital twin, Dev. Built Environ. 16 (2023) http://dx.doi.org/10.1016/j.dibe.2023.100247.

[111] H.N. Rafsanjani, A.H. Nabizadeh, Towards digital architecture, engineering, and construction (AEC) industry through virtual design and construction (VDC) and digital twin, Energy Built Environ. 4 (2) (2023) 169–178, http://dx.doi.org/10.1016/j.enbenv.2021.10.004.

[112] E. Redondo, Research applied to the field of architectural graphic expression through digital means. Reference framework and roadmap; [investigación aplicada al ámbito de la expresión gráfica arquitectónica por medios digitales. marco de referencia y hoja de ruta], VLC Arquit. 10 (2) (2023) 203–224, http://dx.doi.org/10.4995/vlc.2023.19787.

[113] A. Talla, S. McIlwaine, Industry 4.0 and the circular economy: Using design-stage digital technology to reduce construction waste, Smart Sustain. Built Environ. 13 (1) (2024) 179–198, http://dx.doi.org/10.1108/SASBE-03-2022-0050.

[114] Z. Chen, Y. He, Artificial intelligence and environmental protection of buildings; [sztuczna inteligencja i ochrona środowiska budynków], Probl. Ekorozw. 18 (2) (2023) 254–262, http://dx.doi.org/10.35784/preko.4039.

[115] S. Çetin, C.D. Wolf, N. Bocken, Circular digital built environment: An emerging framework, Sustain. (Switz.) 13 (11) (2021) http://dx.doi.org/10.3390/su13116348.

[116] P. Bosch-Sijtsema, C. Claeson-Jonsson, M. Johansson, M. Roupe, The hype factor of digital technologies in AEC, Constr. Innov. 21 (4) (2021) 899–916, http://dx.doi.org/10.1108/CI-01-2020-0002.

[117] S.A. Argyroudis, S.A. Mitoulis, E. Chatzi, J.W. Baker, I. Brilakis, K. Gkoumas, M. Vousdoukas, W. Hynes, S. Carluccio, O. Keou, D.M. Frangopol, I. Linkov, Digital technologies can enhance climate resilience of critical infrastructure, Clim. Risk Manag. 35 (2022) http://dx.doi.org/10.1016/j.crm.2021.100387.

[118] S. Çetin, V. Gruis, A. Straub, Digitalization for a circular economy in the building industry: Multiple-case study of Dutch social housing organizations, Resour. Conserv. Recycl. Adv. 15 (2022) http://dx.doi.org/10.1016/j.rcradv.2022.200110.

[119] E. Lucchi, Digital twins for the automation of the heritage construction sector, Autom. Constr. 156 (2023) http://dx.doi.org/10.1016/j.autcon.2023.105073.

[120] M.I. El-Belkasy, A.M. Shehata, Framework for digital application in sustainable heritage management, Civ. Eng. Archit. 12 (3A) (2024) 2326–2341, http://dx.doi.org/10.13189/cea.2024.121327.

[121] A. Almusaed, I. Yitmen, Architectural reply for smart building design concepts based on artificial intelligence simulation models and digital twins, Sustain. (Switz.) 15 (6) (2023) http://dx.doi.org/10.3390/su15064955.

[122] S. Goncharov, A. Nechesov, AI-driven digital twins for smart cities †, Eng. Proc. 58 (1) (2023) http://dx.doi.org/10.3390/ecsa-10-16223.

[123] H. Sadri, Ai-driven integration of digital twins and blockchain for smart building management systems: A multi-stage empirical study, J. Build. Eng. 105 (2025) http://dx.doi.org/10.1016/j.jobe.2025.112439.

[124] F. Flammini, Digital twins as run-time predictive models for the resilience of cyber–physical systems: A conceptual framework, Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci. 379 (2207) (2021) http://dx.doi.org/10.1098/rsta.2020.0369.