From BIM to Cognitive Digital Twin

J. J. McArthur, jennifer.mcarthur@torontomu.ca
Faculty of Engineering and Architectural Science, Toronto Metropolitan University, Canada

Karim El Mokhtari, elmkarim@torontomu.ca
Faculty of Engineering and Architectural Science, Toronto Metropolitan University, Canada
FuseForward Solutions Group, Vancouver, BC, Canada

Abstract

Cognitive Digital Twins (CDTs) are well-established for manufacturing but have had very little implementation in the buildings sector. This paper presents the development towards one such implementation: a CDT for a mixed-use (academic/ residential) building developed from the construction BIM. Lessons learned regarding five key elements of CDT development are presented: the integrated data model and supporting ontology; building automation system data acquisition and streaming; data lake; event detection algorithms; and integration and visualization. Insights regarding approach selection, implementation considerations, limitations, and alternatives are presented for each to guide the remaining steps (learning, cognition) in the CDT development.

Keywords: BIM, Digital Twin, Cognitive Digital Twin (CDT), data streaming, case study

1 Introduction

The third industrial revolution introduced computers and manufacturing into the construction industry, incorporating computer-aided design (CAD) and manufacturing (CAM) into construction. Building Information Modelling (BIM) was introduced late in this period, providing new opportunities for collaboration and digitization. Industry 4.0 extends beyond this automation to integrate cyber-physical systems and digital technologies (Sawhney, et al., 2020). Within the AEC industry, this is manifest through Construction 4.0, which exploits increasing digitization to improve project delivery efficiency and outcomes (Klinc & Turk, 2019). BIM has been recognized as both a central element of Construction 4.0 and a key facilitator in its transition (FEIC) along with digital twins (DTs) (Al Faruque, et al., 2021). DT research has increased significantly in recent years (Caramia, et al., 2021), evolving from 'digital shadows' that are simply representations (Sepasgozar, 2021; Fuller, et al., 2020) to responsive Cognitive Digital Twins (CDTs) (Al Faruque, et al., 2021; Fuller, et al., 2020) incorporating the ability to "*detect complex and unpredictable actions and reason about dynamic process optimization strategies to support decision-making in building lifecycle management*" (Yitmen, et al., 2021). Of the three types of cognition that have been explored in the literature – semantic permitting the inference of data relationships, perceptual computing to determine the context of the data, and cognitive supporting the analysis of large volumes of data to create and test hypotheses (Sheth, 2016) – this definition focuses primarily on the latter definition of cognition.

While DTs and Cyber-Physical Systems (CPS) are somewhat established, having demonstrated their value across multiple sectors such as construction, health care, transportation, aviation, education, agriculture, and manufacturing (Garcia & Roofigari-Esfahan, 2020), CDTs are newly-emerging. The most significant body of research on CDTs is contextualized within the manufacturing sector, for example (Al Faruque, et al., 2021; Zheng, et al., 2022). Significant potential applications for CDTs in the AEC community have also been identified, spanning design, construction, and operations & maintenance phases; these are summarized in recent reviews (Ghansah & Lu, 2023; Zheng, et al., 2022). However, there have been limited field studies regarding CDT implementation for building operations, many of which are theoretical in focus; for example, developing a framework for CDTs (Yitmen, et al., 2021) or analyzing relevant processes, opportunities, and challenges (Su, et al., 2023)

This paper addresses this gap by presenting the lessons learned from the first four years of CDT development on the Smart Campus Integration Platform (SCIP). SCIP is a six-year research project to create an integrated, multi-domain {buildings, infrastructure, transportation systems} Cognitive Digital Twin for a university campus. The buildings domain work has focused on the pilot development of a CDT for an academic building topped with a residential tower, which would be scaled to the full campus. This paper presents the key insights from the first half of this project, which focused on the development of an integrated data model and supporting ontology, data acquisition and streaming from the building automation system (BAS) and integration with batch-processed facility management data, the development of a series of algorithms for event detection, and the integration of these research streams into the digital twin. Ongoing research is now focused on enabling cognitive computing, developing and supporting operational efficiency through a suite of supervised, unsupervised, and reinforcement learning algorithms to support autonomous learning. By providing an overview of the lessons learned and outcomes of this project to date, this paper provides a valuable case study to DT scholars.

2 Background

The notion of cognition in computing has been around since Alan Turing and developments in machine learning over the latter half of the 20th century developed early natural language processing, semantic structures, and early machine learning algorithms. Early cognitive computing applications relied on rule-based inferences generated based on expert input, whether provided in the form of the written rules or created within the supporting ontology (Eastman, et al., 2009). With advances in artificial intelligence, machine learning has evolved substantially, evidenced by IBM's Deep Blue and Watson machines, extending beyond mere inference to prediction and optimization capabilities. The ability for algorithms to learn – whether through deep learning, reinforcement, or transfer learning from similar systems – further extends computational cognition.

While no known existing building CDT can claim cognition, there has been significant research developed to develop their semantic, cognitive, and perceptual functionalities. This literature review summarizes this literature mapped to the six core functionalities identified by Al Faruque et al. (2021) that are required for CDT development: perception (data acquisition and structuring), attention (of data warranting attention), data storage and visualization, reasoning, problem solving, and learning. Together, these functionalities support "*computer-driven systems (to) monitor physical processes, create a virtual copy of the physical world and make decentralized decisions based on self-organization mechanisms*" (Smit, et al., 2016).

2.1 Perception

Two elements are required for a CDT to correctly *perceive* data: an appropriate data model and a data acquisition system. The latter provides access to the data itself, while the former permits it to be appropriately sorted and related to other CDT data.

An enormous volume of research has focused on data model and ontology development for the construction sector, see (Pedral Sampaio, et al., 2022). To enable BIM's integration of geometric, topological, and semantic information, a host of supporting data structures and ontologies, most notably Industry Foundation Classes (buildingSMART, 2013), CoBiE (East, 2007), and their integration with other ontologies through OWL (Pauwels & Terkaj, 2016; Rasmussen, et al., 2017). Other ontologies relevant for digital twins are the IoT-focused BRICK ontology (Balaji, et al., 2018) and Project Haystack (Project Haystack Corporation, 2014).

Relational (SQL), unstructured (NoSQL), and graph databases have all been used to populate DTs, relating BIM with IoT data. A recent review (Tang, et al., 2019) found that using an API to link the BIM (a relational database) with a relational time-series database is the most commonlyused approach, but is limited to BIM shared parameters. Another approach relates BIM and time-series data using a new schema (Motamedi, et al., 2014), offering flexibility at the cost of increased expertise in implementation (Tang, et al., 2019). Other approaches noted in this review are developing new query languages, semantic web technologies such as Resource Description Framework (RDF), and semantic web-SQL hybrid approaches using SPARQL (Tang, et al., 2019). NoSQL databases have also been used, both in conjunction with semantic web technologies (Chevallier, et al., 2020), and linked data approaches (Quinn, et al., 2020); this research continues this latter avenue of contribution. Regarding structure, knowledge graphs are seeing increasing adoption for DTs, complementing OWL and RDF technologies as the most common current DT structure in a recent review (D'Amico, et al., 2022). Graph databases have been noted as one of the most promising technologies for DTs (Sacks, et al., 2020) as they are particularly valuable for semantic linked-data approaches because of their conceptual clarity (Boje, et al., 2020). Data streaming to populate these databases frequently combines buffer and streaming services, for example Kafka and Spark (Quinn, et al., 2020). This research leverages linked data approaches underpinned by the BRICK ontology with graph database data structure.

2.2 Attention

Data aggregation can offer significant value in reducing data processing for common queries such as daily average or extreme sensor values (Quinn, et al., 2020). More significant to *attention*, however, is outlier detection, which uses unsupervised machine learning to detect anomalous events such as faults (Alimohammadi & Chen, 2022).

2.3 Data Storage and Visualization

Heterogeneity of data in a CDT requires the use of various data stores. Data lakes provide an efficient solution where structured and unstructured data can co-exist. Building data is stored separately for each instance and accessed through a common data management system (Quinn, et al., 2020). For BIM 3D models storage and delivery, a variety of commercial solutions exist today, however there is an increasing interest towards the use of open-source formats for that end. glTF (Graphics Language Transmission Format or GL Transmission Format) is a popular standard file format for 3D scenes and models. It features the integration of both descriptive structured data and binary geometric data to efficiently encode 3D models. Its integration in WebGL enabled web browsers facilitates streaming of large 3D models from the data lake and visualization platforms (Schilling, et al., 2016). A web-based platform is often preferred for an increased compatibility with end-users' devices, making glTF an efficient format for 3D data exchange.

2.4 Reasoning, Problem Solving, and Learning

An enormous volume of research has explored cognition. Rule-based expert systems, physics-based model predictive control, data-driven algorithms (supervised, semi-supervised, unsupervised, or using reinforcement learning), and hybrid approaches.

Rule-based expert systems are a means to capture and code domain expertise, integrating a knowledge base with an inference engine incorporating descriptive rules created by a domain expert to apply heuristics, forward or backward reasoning and representation rules, and can explain system behaviour under uncertainty (Akram et al., 2014). These have been applied to building applications such as fault detection (Peña et al., 2016).

Physics-based models have been widely used to simulate and optimize system behaviour. Within the buildings domain, these can be grouped into two types: standalone equipment emulators and those incorporating building response. Equipment emulators use equations to incorporate thermodynamic, fluid dynamic, and heat transfer mechanisms (Taheri, et al., 2022; Mariano-Hernández, et al., 2021) while model predictive control incorporating building performance can use resistance-capacitance networks (Dong & Lam, 2014); building energy simulation software such as EnergyPlus (Gunay, et al., 2020); or infer unmeasurable elements using building data (Gilani, et al., 2019). Optimization can be achieved either numerically, for example, (Carrascal, et al., 2016), but increasingly machine learning is integrated to guide optimization, for example (Ascione, et al., 2016).

More recently, data-driven methods have come to dominate the literature, allowing predictive models to be created rapidly and with reduced effort. This literature is too extensive to include here, so the reader is directed to recent literature reviews summarizing such developments as they apply to energy optimization (Aliero, et al., 2022) and fault detection strategies (Nassif, et al., 2021; Melgaard, et al., 2022).

3 Cognitive Digital Twin Development

The CDT was initially created using Autodesk Revit from the as-built set of five design models (2x architectural plus structural, mechanical, and electrical). The federated model was enormous, containing large amounts of unnecessary detail for CDT operation and requiring significant computational resources. In order to create a simplified model, the two architectural models, containing the podium and tower, were imported into a single file and the detailed architectural elements were replaced by simplified versions. Rather than importing the structural, electrical, and mechanical models, a survey of their content was completed and those elements valuable for the planned CDT use cases were identified. The desired use cases, established with input from the Facility Engineer, were: HVAC equipment fault detection; HVAC sensor fault and drift detection; HVAC system controls optimization to maximize energy efficiency; maintenance of indoor comfort conditions; and the detection of events impacting thermal comfort, indoor air quality, and energy consumption that could not otherwise be detected from BAS data. To support these use cases, data was collected from the BAS (control and measurement points), supplemental BTU and electrical submeters, and additional sensors for monitoring branch airflows. Manufacturer equipment models for these elements were replaced with simplified HVAC families incorporating all semantic and relational parameters and an approximate overall geometry. This permitted a lightweight version of the model capable of hosting all relational and semantic data but a size reduction of 99% (1GB to 10MB).

3.1 Ontology and Data Model

The multi-domain data model was initially created in SQL using a simplified structure to support multi-domain data, leveraging the BRICK Ontology to define relationships such as those between assets, their measurement and control points, and locations. BIM data (asset locations and semantic data, location and asset hierarchies) were exported as schedules, which were readily translated into SQL tables to populate the SQL database. However, during CDT development three significant limitations of the SQL approach became apparent: the high level of effort to check for completeness and accuracy of defined elements; limitations on defining new attributes needed for machine learning; and the challenge of defining complex queries. As noted in the literature review, graphDBs offer significant value to overcome these gaps. The visual display of graphDBs permits intuitive database checking and exploration (Fig. 1), greatly simplifying quality control. Neo4j was used as the graphDB using BRICK relationships because of the various features it provides such as being open-source, running with a native graph database and library, and supported by one of the largest networks in the area of graph databases. We note that while RDF/OWL would have offered the benefits of integration with the W3C ontology and inference, the complexity of queries (a simpler one shown in Fig. 2) to sample the time-series data to support use cases made the mapping of this data cumbersome. Further, the inference capability of RDF was just as readily provided by Neo4j. Figures, tables and equations.

Data migration from SQL into Neo4j, was achieved using script-based batches to create nodes and relationships in their order of hierarchy in the ontology: {building, levels, rooms, assets, sub-assets, command points, and measurement points}, which was initially mapped from the BIM as described above. Each batch generates a list of Cypher (Neo4j's query language) commands executed on the Neo4j instance in the data lake. It is worth noting that the Neo4j's community version does not support multiple users (administrator access only), and as most of No-SQL databases, referential integrity is not managed by the DB engine itself but handled at the business logic side. Queries in this new format are much more powerful, permitting highly complex queries with minimal programming, as shown in Figure 2. Beyond the complex queries facilitated by the graphDB, additional benefits to this project have been the ease of database expansion and checking.

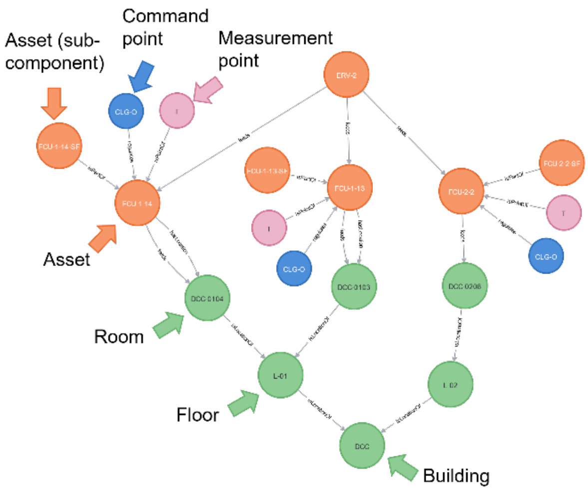
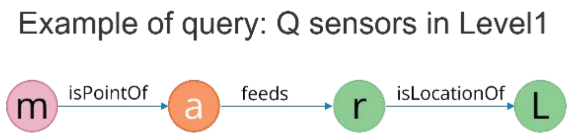


Figure 1. Extract from the GraphDB (Neo4j) Implementation




```
MATCH (m:MeasurementPoint {shortName:'Q'})
-[:isPointOf]->(a:Asset)-[:feeds]->
(r:Location)-[:isLocationOf]->
(L:Location {shortName:'L-01'})
RETURN m,a,r,L
```

Figure 2: Sample Complex Query

3.2 Data Acquisition and Streaming

The primary CDT data source is the BAS, which was programmed to output a change of value at prescribed thresholds. This approach was initially selected to avoid creating traffic on the campus critical network. However, it proved problematic from a machine learning perspective since variable readings are asynchronous. To address this, pre-processing is introduced to develop 15-minute interval data suitable for model training.

Data acquisition was implemented as-follows. First, a sanity check is performed on each variable by querying it once a day to ensure the related sensor is online. Second, an aggregation process is run hourly and daily to calculate and store key statistics on each sensor (min, max, average, standard deviation, records count), furthermore, a shadow table is updated in real-time to reflect the last state of each variable for fast querying. Data can be queried from all these data sources depending on the granularity required. For machine learning algorithms that require the highest time scale granularity, a window bound time-series data is retrieved from the time series database training job, then resampled to the desired time-frame (15 minutes is generally suited for most HVAC-based ML algorithms). The temporary dataset is then destroyed at the end of the training job.

Data streaming was initially performed using a custom software provided by the BAS vendor as part of their contract, locally buffered, and sent to an ElasticNet database using HTTP packets (Quinn, et al., 2020). However, this method proved to have robustness issues, with significant data losses due to new incoming signals from the BAS being ignored during buffering and streaming. A revised streaming strategy (Fig. 3) replaced the API with one developed by the research team and data stream management, data ingestion, and ETL transformation services were implemented in the data lake. The ElasticNet database was also replaced with a time-series database, which proved to be much more robust, dramatically reducing the missing data. However, it revealed one additional issue with the change-of-value approach: oversubscription. When attempting to subscribe to the various devices, several indicated they were oversubscribed and could not accommodate an additional signal. Those devices were primarily served by a particular network controller. For these devices, the revised streaming strategy queries at prescribed time intervals that were determined using data mining on the historical data manually extracted from the BAS.

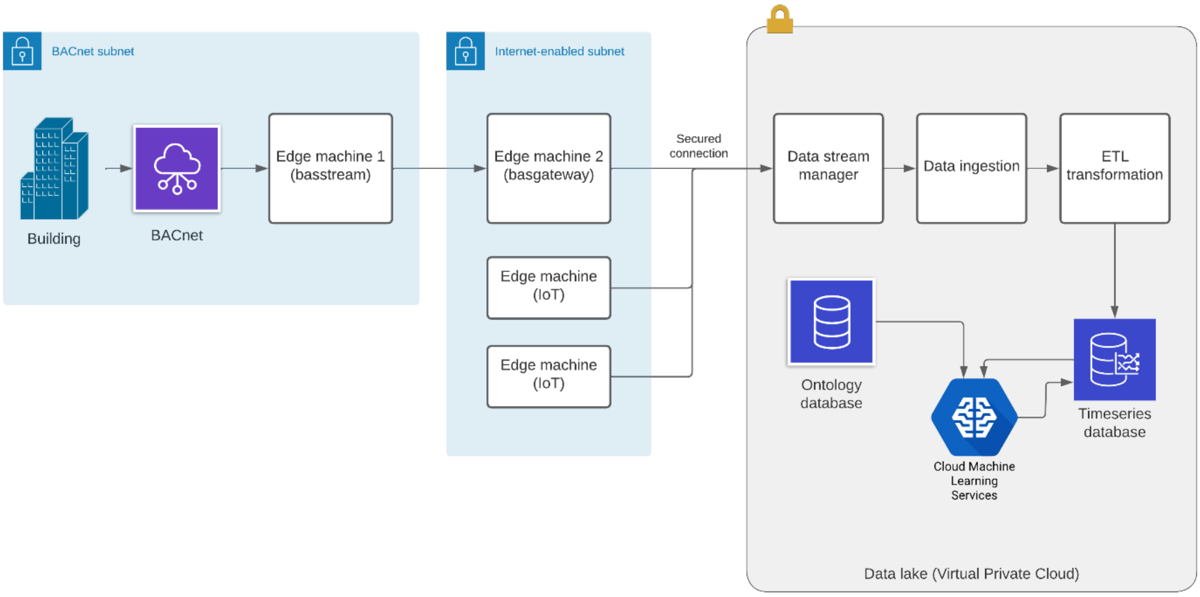


Figure 3: Revised Streaming Approach

3.3 Data Storage

A data lake was established to support this work, consisting of the federated data model and a set of time-series databases for each dynamic data source, for example cameras, IoT devices, and sensor networks. The federated data model was kept as a single multi-domain model to enable extra-building relationships, for example the interaction between a building and the broader utility grids, or the number of pedestrians entering the building.

3.4 Developing Cognition: Reasoning and Problem Solving

This research is currently in-progress and thus this section summarizes the strategies used to develop reasoning and problem-solving and summarizes the preliminary results.

To enable the cognition needed to support the operational use cases, a suite of supervised, unsupervised, and reinforcement learning algorithms are being developed, with increasing focus on the latter to support autonomous learning. Supervised methods have used LSTM and CNN methods to classify data based on its time-series characteristics and relationship with other data points and use mismatch between actual and assigned data for event detection; however, these were found to be computationally too costly to implement at scale (El Mokhtari & McArthur, 2021). Supervised methods are also limited in practice due to the paucity of labeled building data. To address this, unsupervised methods such as Gaussian Mixture Models, clustering, and Bayesian Networks have been explored and show promise in identifying both individual unusual equipment measurements and atypical relationships between measurements on a system level, indicating a faulty sensor or non-BAS-detectable event. Once identified, these can be reviewed by an expert who can then label these cases to support future implementation of semi-supervised learning. By cross-referencing semantic relationships, the context, extent, and severity of such events can also be determined and this will be further investigated in future research.

While rule-based methods have been widely used for reasoning and problem-solving, these have not been considered in this research because this would be redundant with the BAS. At present, such systems are prescriptive, running the facility and raising alarms based on a set of pre-determined rules developed by an expert (the facility design engineer). While the outputs of the supervised, semi-supervised, and unsupervised learning may guide the establishment of new data-informed rules, this is future research and thus is not discussed here.

To implement cognition, unsupervised methods have been developed to monitor data streaming patterns, automatically detect faults, and apply transfer learning with reinforcement learning (Genkin & McArthur, 2024).

3.5 Digital Twin Integration and Visualization

The CDT has been implemented on multiple platforms: Autodesk Forge, AWS TwinMaker, Unreal, and – for the full campus - Cesium. Of these, Forge is immediately compatible with OpenBIM as it can read .ifc files to create the models. For the others, the BIM had to be translated into .gltfs (TwinMaker), while Unreal required the use of TwinMotion to translate the .ifc model. To implement the CDT in Forge, a simplified FM-BIM was uploaded and a series of APIs were developed to support ontology model (knowledge graph) queries to translate user interactions into a list of sensors to be queried and import them from the time series database. Figure 4 shows a sample view of the result of such a query.

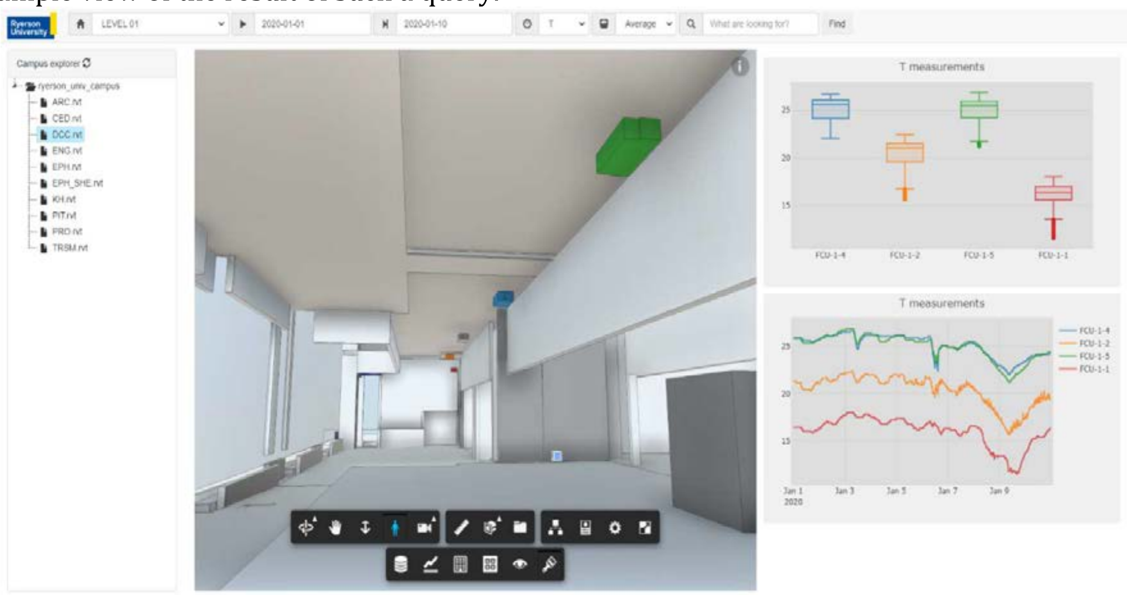


Figure 4: CDT Forge Interface

The Forge interface can be used as-is or integrated into a web browser using React.js or other tools to create a custom dashboard to highlight critical parameters to the user.

4 Conclusions

The research completed to date has resulted in six insights valuable to guide scholars also seeking to create a building CDT from an existing BIM.

1. **They must have the *right* data**: The initial BIM must be complete and accurate with respect to the data required by the CDT. The BIM need not contain all the data, for example streamed data is much more effectively maintained in a time-series database, but must include all elements to which this data will be mapped. The BIM must also be developed using an appropriate ontology containing all necessary elements. In this paper, a linked-data approach was developed to permit the BIM data to be integrated with data from other domains (infrastructure and transportation) with a federated data model defined

using BRICK relationships and simplified classes. The use of graphDBs is recommended for this model to support creation, checking, and complex queries.

2. **Simplicity is key:** There is substantial value in simplifying the BIM, particularly equipment families, as existing CDT visualization tools are better able to process smaller models. At minimum, rooms/spaces and key equipment must be included in a simplified model, with sensors defined for each. Elements not required for CDT functionalities may also be omitted from this simplified model.
3. **Network Traffic must be managed:** Data acquisition and streaming from a BAS requires significant care to balance adequate data granularity without unduly increasing BAS network traffic. Where possible, push data at a prescribed change of value as its network impact is minimal. Subscribing to change of value data is similarly effective, however some devices are prone to oversubscription; in such cases, data mining to determine appropriate query intervals of no more than 15 minutes to support energy audits and building simulation comparisons should be defined.
4. **Pre-process data at the edge:** Data pre-processing and the storage of aggregated measurements is effective to reduce computational costs for the CDT. These can be stored in a separate time-series database within the data lake.
5. **Unsupervised learning supports scaling:** To reduce computational cost while permitting scaling to other buildings, unsupervised machine learning methods are valuable for event detection. Reinforcement and transfer learning have also proven valuable in initial explorations for reducing the learning period necessary for such models.
6. **Leverage existing tools:** There exists a suite of visualization platforms and tools that can readily operationalize BIM-based DTs and support their integration with cloud-hosted databases. The most appropriate tool will be determined based on model size and complexity (Forge proved unstable with very large BIMs while Cesium proved to be more efficient in rendering the full campus view), degree of data integration required (supported by both TM and Forge with custom programming), and the desired degree of realism in the completed model (both TM and Forge have limited rendering capability). Video game engines offer significant value to provide a complex, immersive, and open-source platform for CDT operationalization.

CDTs offer significant value for the architecture, engineering, construction, and operations (AECO) sector as they offer the ability to leverage existing sensor networks and apply AI to improve equipment and asset performance. When integrated with machine-to-machine communication, CDTs support autonomous or semi-autonomous (human-in-the-loop or human-on-the-loop) asset operation. Potential applications for CDTs to enable Construction 4.0 include: autonomous construction equipment; adaptive building controls; online equipment optimization; automatic fault detection and diagnosis for equipment and systems; predictive maintenance; and tracking occupant satisfaction and automatically resolving reported issues.

The primary limitation of this research is that to date it has only been applied to a single campus. While it has included multiple campus-level applications including maintenance request and occupant complaint mapping and utility monitoring and building-level fault detection, and online optimization, there is significant value in applying the approaches discussed herein to other applications. Future research will implement the methods shown in residential, institutional, and industrial contexts to confirm the findings and further refine the approach.

Acknowledgements

This research was funded by Canada's Natural Science and Engineering Research Council (NSERC) [ALLRP/544569-2019] and FuseForward Solutions Group.

5 References

Al Faruque, M., Muthirayan, D., Yu, S. & Khargonekar, P., 2021. Cognitive Digital Twin for Manufacturing Systems. *Design, Automation & Test in Europe Conference & Exhibition.* IEEE, pp. 440-445.

Aliero, M. et al., 2022. Systematic Review Analysis on Smart Building: Challenges and Opportunities. *Sustainability,* 14(5), p. 3009.

Alimohammadi, H. & Chen, S., 2022. Performance evaluation of outlier detection techniques in production timeseries: A systematic review and meta-analysis. *Expert Systems with Applications,* 191( ), p. 116371.

Ascione, F. et al., 2016. Simulation-based model predictive control by the multi-objective optimization of building energy performance and thermal comfort. *Energy and Buildings,* Volume 111, pp. 131-144.

Balaji, B. et al., 2018. Brick : Metadata schema for portable smart building applications. *Applied Energy,* Volume 226, pp. 1273-1292.

Boje, C., Guerriero, A., Kubicki, S. & Rezgui, Y., 2020. Towards a semantic Construction Digital Twin: Directions for future research. *Automation in Construction,* 114( ), p. 103179.

buildingSMART, 2013. *IFC4: The New BuildingSMART Standard,* Northamptonshire: buildingSMART International.

Caramia, G., Corallo, A. & Mangialardi, G., 2021. *The Digital Twin in the AEC/FM Industry: a literature review.* Luxembourg, CIB W78, pp. 11-15.

Carrascal, E., Garrido, I., Garrido, A. J. & Sala, J. M., 2016. Optimization of the heating system use in aged public buildings via Model Predictive Control. *Energies,* 9(4), pp. 1-21.

Chevallier, Z., Finance, B. & Boulakia, B., 2020. *A Reference Architecture for Smart Building Digital Twin.* ESWC.

D'Amico, R., Erkoyuncu, J., Addepalli, S. & Penver, S., 2022. Cognitive digital twin: An approach to improve the maintenance management. *CIRP Journal of Manufacturing Science and Technology,* 38( ), pp. 613-630.

Dong, B. & Lam, K., 2014. A real-time model predictive control for building heating and cooling systems based on the occupancy behavior pattern detection and local weather forecasting. *Building Simulation,* 7(1), pp. 89-106.

Eastman, C., Lee, J., Jeong, Y. & Lee, J., 2009. Automatic rule-based checking of building designs. *Automation in construction,* 18(8), pp. 1011-1033.

East, W. E., 2007. *Construction operations building information exchange (COBie).* , Champaign, IL : Engineer Research and Development Center Construction Engineering Research Lab.

El Mokhtari, K. & McArthur, J., 2021. *A data-driven approach to automatically label BAS points.* Luxembourg, CIB W78, pp. 70-79.

Fuller, A., Fan, Z., Day, C. & Barlow, C., 2020. Digital twin: Enabling technologies, challenges and open research. *IEEE access,* Volume 8, pp. 108952-108971.

Garcia, D. & Roofigari-Esfahan, N., 2020. Technology requirements for CPS implementation in construction. In: C. Anuba & N. Roofigari-Esfahan, eds. *Cyber-Physical Systems in the Built Environment.* Cham: Springer, pp. 15-30.

Genkin, M. & McArthur, J., 2024. A transfer learning approach to minimize reinforcement learning risks in energy optimization for automated and smart buildings. *Energy and Buildings,* 303( ), p. 113760.

Ghansah, F. & Lu, W., 2023. Cyber-physical systems and digital twins for "cognitive building" in the construction industry. *Construction Innovation,* Volume (ahead of print), p. (ahead of print).

Gilani, S. et al., 2019. *Optimizing Energy with Machine Learning Grey-Box Models.* Newcastle-upon-Tyne, pp. 137-147.

Gunay, H., Newsham, G., Ashouri, A. & Wilton, I., 2020. Deriving sequences of operation for air handling units through building performance optimization. *Journal of Building Performance Simulation* , 13(5), pp. 501-515.

Klinc, R. & Turk, Ž., 2019. Construction 4.0–digital transformation of one of the oldest industries. *Economic and Business Review,* 21(3), pp. 393-410.

Mariano-Hernández, D. et al., 2021. A review of strategies for building energy management system: Model predictive control, demand side management, optimization, and fault detect & diagnosis. *Journal of Building Engineering,* Volume 33, p. 101692.
Melgaard, S. et al., 2022. Fault Detection and Diagnosis Encyclopedia for Building Systems: A Systematic Review. *Energies,* 15(12), p. 4366.
Motamedi, A., Hammad, S. & Asen, Y., 2014. Knowledge-assisted BIM-based visual analytics for failure root cause detection in facilities management.. *Autom. Constr.,* Volume 43, pp. 73-83.
Nassif, A., Talib, M., Nasir, Q. & Dakalbab, F., 2021. Machine learning for anomaly detection: A systematic review. *IEEE Access,* Volume 9, pp. 78658-78700.
Pauwels, P. & Terkaj, W., 2016. EXPRESS to OWL for construction industry: Towards a recommendable and usable ifcOWL ontology. *Automation in construction,* 63( ), pp. 100-133.
Pedral Sampaio, R., Aguiar Costa, A. & Flores-Colen, I., 2022. A Systematic Review of Artificial Intelligence Applied to Facility Management in the Building Information Modeling Context and Future Research Directions. *Buildings,* 12(11), p. 1939.
Project Haystack Corporation, 2014. *Project Haystack.* [Online] Available at: https://project-haystack.org/ [Accessed 08 07 2019].
Quinn, C. et al., 2020. Building automation system-BIM integration using a linked data structure. *Automation in Construction,* 118( ), p. 103257.
Rasmussen, M. H., Pauwels, P., Hviid, C. A. & Karlshoj, J., 2017. *Proposing a Central AEC Ontology that Allows for Domain Specific Extensions.* Heraklion, Lean & Computation in Construction Congress.
Sacks, R., Girolami, M. & Brilakis, I., 2020. Building information modelling, artificial intelligence and construction tech. *Developments in the Built Environment,* 4(2020), p. 100011.
Sawhney, A., Riley, M., Irizarry, J. & Pérez, C., 2020. A proposed framework for Construction 4.0 based on a review of literature. *EPiC Series in Built Environment,* Volume 1, pp. 301-309.
Schilling, A., Bolling, J. & Nagel, C., 2016. *Using glTF for streaming CityGML 3D city models.* , , pp. 109-116.
Sepasgozar, S., 2021. Differentiating Digital Twin from Digital Shadow: Elucidating a Paradigm Shift to Expedite a Smart, Sustainable Built Environment. *Buildings,* 11(151), pp. 1-16.
Sheth, A., 2016. Internet of things to smart iot through semantic, cognitive, and perceptual computing. *IEEE Intelligent Systems,* 31(2), pp. 108-112.
Smit, J., Kreuter, S., Moeller, C. & Carlberg, M., 2016. *Industry 4.0,* Brussels: European Parliament.
Su, S. et al., 2023. Digital twin and its potential applications in construction industry: State-of-art review and a conceptual framework. *Advanced Engineering Informatics,* Volume 57, p. 102030.
Taheri, S., Hosseini, P. & Razban, A., 2022. Model predictive control of heating, ventilation, and air conditioning (HVAC) systems: A state-of-the-art review. *Journal of Building Engineering,* 60( ), p. 105067.
Tang, S. et al., 2019. A review of building information modeling (BIM) and the internet of things (IoT) devices integration: Present status and future trends. *Automation in Construction,* 101( ), pp. 127-139.
Yitmen, I., Alizadehsalehi, S., Akıner, İ. & Akıner, M., 2021. An Adapted Model of Cognitive Digital Twins for Building Lifecycle Management. *Applied Sciences,* 11(9), p. 4276.
Zheng, X., Lu, J. & Kiritsis, D., 2022. International Journal of Production Research. *The emergence of cognitive digital twin: vision, challenges and opportunities,* 60(24), pp. 7610-7632.