



ARTICLE

# A Blockchain-Assisted BIM-IoT Digital Twin Architecture for Trusted Operational Risk Prediction in Smart Buildings

Yuh-Shihng Chang<sup>1</sup> and Hsuan-Chao Huang<sup>2,\*</sup>

<sup>1</sup>Department of Information Management, National Chin-Yi University of Technology, Taichung, Taiwan

<sup>2</sup>Department of Computer Science and Information Engineering, National Chin-Yi University of Technology, Taichung, Taiwan

\*Corresponding Author: Hsuan-Chao Huang. Email: sc100@ncut.edu.tw

Received: 14 April 2026; Accepted: 05 June 2026; Published: 13 August 2026

**ABSTRACT:** The integration of Building Information Modeling (BIM), Internet of Things (IoT), digital twins, and artificial intelligence (AI) has enabled advanced smart building operations with real-time monitoring and data driven decision-making capabilities. However, existing systems still face critical challenges in ensuring trusted data provenance, secure cross-layer data exchange, and robust access control in distributed IoT environments. These limitations significantly affect the reliability and trustworthiness of data driven analytics and system-level decision-making. To address these challenges, this study proposes a blockchain-assisted BIM-IoT digital twin architecture for trusted operational risk prediction in smart buildings. The proposed framework integrates KNX-based sensing systems, BIM models, and a microservice-enabled data processing platform to support real-time bidirectional synchronization between physical assets and their virtual representations. A blockchain-assisted trust layer is introduced to provide tamper-resistant event logging, trusted data provenance, and secure coordination across IoT devices, digital models, and analytical modules. Within this architecture, a Random Forest model is employed to perform operational risk classification based on environmental, energy, and temporal features. Feature importance analysis is further applied to enhance model interpretability and provide insights into the influence of key variables on system behavior. The framework is validated using over six months of real-world data collected from a smart building environment, demonstrating stable prediction performance and identifying temperature as the most influential factor. The results indicate that the proposed architecture not only achieves robust predictive performance but also supports trustworthy system operation by ensuring data integrity, traceability, and secure interaction across distributed subsystems. The proposed framework provides a secure, scalable, and practically deployable solution for intelligent smart building management, bridging trusted sensing, BIM-based modeling, and explainable AI-driven decision-making.

**KEYWORDS:** BIM; digital twin; blockchain-IoT; trusted data provenance; IoT security; smart buildings; access control; random forest; operational risk prediction

## 1 Introduction

The rapid advancement of smart building technologies has led to increasingly interconnected and data driven building environments, where Building Information Modeling (BIM), Internet of Things (IoT), and artificial intelligence (AI) are tightly integrated [1,2]. These technologies enable real-time monitoring, automated control, and intelligent decision-making in modern building systems. However, as smart buildings evolve into distributed cyber-physical systems, ensuring the security, integrity, and trustworthiness of data across heterogeneous IoT devices and system layers has become a critical challenge.

In particular, the lack of trusted data provenance, secure cross-layer communication, and robust access control mechanisms in distributed IoT environments [3] can significantly undermine the reliability of digital twin synchronization and AI-driven analytics. These challenges are further amplified in large-scale smart building systems, where data are continuously exchanged among multiple devices, platforms, and analytical modules. Therefore, establishing a secure and trustworthy system architecture has become essential for enabling reliable and scalable smart building operations.

To address these challenges, blockchain technology has emerged as a promising approach for enabling decentralized trust, tamper-resistant data management, and secure system coordination in distributed environments. By leveraging cryptographic mechanisms and distributed ledgers, blockchain can provide verifiable data provenance, transparent event logging, and decentralized access control without relying on centralized authorities. Integrating blockchain mechanisms into smart building systems offers the potential to enhance data reliability and system security, particularly in heterogeneous IoT environments.

At the same time, digital twin (DT) technology [4] extends BIM-based modeling by enabling real-time synchronization between physical building systems and their virtual representations. Through continuous data mapping and bidirectional interaction, digital twins provide a dynamic platform for monitoring, simulation, and system optimization. However, most existing digital twin implementations focus primarily on data visualization and synchronization, with limited consideration of secure data governance and trusted data exchange, which are essential for reliable system-level decision-making.

In addition, machine learning techniques have been widely applied in smart building systems for energy prediction, anomaly detection, and operational optimization. Among these methods, Random Forest has demonstrated strong performance [5,6] due to its robustness, resistance to overfitting, and inherent interpretability through feature importance analysis. Nevertheless, most machine learning models rely on implicitly trusted datasets and do not consider the reliability and integrity of data sources in distributed IoT environments, which may compromise the trustworthiness of AI-driven predictions.

Therefore, this study aims to investigate how a blockchain-assisted BIM-IoT digital twin architecture can support secure and trustworthy operational risk prediction in smart buildings. Specifically, the proposed framework integrates IoT sensing, micro-service-based data processing, blockchain-assisted trust mechanisms, and digital twin synchronization into a unified and scalable system architecture.

To achieve this objective, a real-world smart building platform based on KNX building automation systems is developed to collect environmental and operational data. A microservice-based architecture deployed on a lightweight containerized platform is adopted to enable flexible data processing and system integration. A blockchain-assisted trust layer is introduced to provide tamper-resistant logging, trusted data provenance, and secure coordination across system components. In addition, a Random Forest model is employed to perform operational risk classification, with feature importance analysis enhancing model interpretability.

The main contributions of this study are summarized as follows:

- Introduces a trusted BIM-IoT digital twin architecture enhanced by blockchain technology.
- Establishes secure data provenance through Hyperledger Fabric, PBFT consensus, and SHA-256 verification.
- Combines blockchain trust management with digital twin synchronization in a unified framework.
- Integrates explainable AI for operational risk prediction and decision support.
- Demonstrates practical feasibility through long-term deployment in a real smart building environment.

## 2 Literature Review

Recent advances in smart building systems have been driven by the convergence of IoT technologies, BIM-based digital modeling, and data driven analytics. However, as building systems become increasingly interconnected and data-intensive, issues related to data security, trust management, and system-level integration have emerged as critical challenges. Existing studies can be broadly categorized into four areas: (1) IoT-based monitoring systems, (2) BIM and digital twin integration, (3) machine learning-based analytics, and (4) blockchain-enabled IoT security architectures.

### 2.1 IoT-Based Smart Building Monitoring

IoT technologies have been widely adopted [1–3] in smart buildings to enable real-time monitoring of environmental conditions, energy consumption, and equipment status. Sensors such as temperature, humidity, and power meters provide continuous data streams that support building automation and energy management. These systems are typically integrated through communication protocols such as KNX, BACnet, and IP-based networks, allowing distributed sensing and control across building environments.

While IoT-based systems offer effective data acquisition and automation capabilities, most implementations primarily focus on functionality and connectivity. Limited attention has been given to data integrity, secure transmission, and trust management, particularly in large-scale or distributed deployments. As a result, IoT-based smart building systems remain vulnerable to issues such as data manipulation, unauthorized access, and inconsistent data sources, which can affect the reliability of downstream analytics and control decisions.

### 2.2 BIM and Digital Twin Integration

Building Information Modeling (BIM) has been extensively used [7,8] for digital representation, visualization, and lifecycle management of building assets. BIM enables structured storage of spatial, geometric, and semantic information, facilitating coordination across design, construction, and operation stages. However, traditional BIM applications are typically static and lack real-time interaction with physical systems.

To address this limitation, digital twin (DT) technology [4] has been introduced as an extension of BIM, enabling continuous synchronization between physical environments and virtual models. By integrating real-time IoT data with digital representations, digital twins provide a dynamic platform for monitoring, simulation, and system optimization. Recent studies [9–11] have demonstrated the potential of BIM-enabled digital twin frameworks for applications such as energy management, structural monitoring, and indoor environment analysis.

However, the increasing interconnection between BIM, IoT, and digital twin systems also introduces cybersecurity and data integrity challenges. As highlighted by Das et al. [12], cyber-physical security has become a critical concern in BIM-IoT integrated smart buildings due to the continuous exchange of operational data across heterogeneous devices and platforms. Despite these advancements, Alshammari et al. [13] emphasized that most digital twin implementations in the built environment focus primarily on data integration and visualization, with limited consideration of comprehensive cybersecurity frameworks and secure data governance.

Most digital twin implementations primarily emphasize data integration and visualization, while comprehensive cybersecurity frameworks and secure data governance remain insufficiently addressed [12,13]. Furthermore, Zhang et al. [14] emphasized that BIM-based monitoring systems should incorporate appropriate cryptographic and authentication mechanisms to protect BIM data and support secure system operation.

These observations further motivate the integration of blockchain-assisted trust mechanisms into BIM-IoT digital twin architectures. As digital twins increasingly serve as central platforms for decision-making and system coordination, ensuring the reliability and trustworthiness of synchronized data becomes a critical requirement. Iqbal and Mirzabeigi [15] proposed a digital twin-enabled BIM-IoT framework that incorporates machine learning to successfully optimize indoor thermal comfort and energy efficiency. In particular, Padma and Ramaiah [16] demonstrated that blockchain-assisted architectures can significantly enhance secure data sharing and risk assessment in distributed smart city environments.

Therefore, this study introduces a blockchain-assisted trust layer into the BIM-IoT digital twin architecture to strengthen data integrity, access control, and traceability. The proposed framework employs Hyperledger Fabric, PBFT consensus, SHA-256 hash logging, and role-based access control mechanisms to establish a trusted operational environment for smart building risk prediction.

### **2.3 Machine Learning in Smart Buildings Systems**

Machine learning techniques have been widely applied in smart building systems for tasks such as energy consumption prediction, anomaly detection, and operational optimization [5,6,17,18]. Algorithms including Random Forest, support vector machines, and neural networks have demonstrated strong performance in capturing complex relationships between environmental variables and system behavior.

Among these methods, Random Forest [6,17,18] has gained particular attention due to its robustness, resistance to overfitting, and ability to provide interpretable results through feature importance analysis. These characteristics make it especially suitable for energy consumption engineering applications where transparency and explainability are essential.

However, most machine learning-based approaches operate under the assumption that input data are reliable and trustworthy. In distributed IoT environments, where data are collected from heterogeneous sources and transmitted across multiple layers, this assumption may not hold. The lack of mechanisms to ensure data integrity, provenance, and secure access can limit the reliability of AI-driven decision-making in real-world smart building systems.

### **2.4 Blockchain-Enabled IoT Security and Trust Management**

To address the challenges of data trustworthiness and secure system coordination, blockchain technology [19–21] has emerged as a promising solution for IoT-based systems. Blockchain provides a decentralized and tamper-resistant ledger that enables transparent data recording, trusted data provenance, and secure peer-to-peer interactions without relying on centralized authorities.

In the context of IoT, blockchain has been applied to enhance device authentication, secure data sharing, and access control management. By leveraging cryptographic mechanisms and distributed consensus, blockchain-based systems can mitigate risks associated with data tampering, unauthorized access, and single points of failure [19,20]. Recent studies have explored blockchain-IoT integration for applications such as smart grids, industrial IoT, and secure sensor networks [22], demonstrating its potential for improving system reliability and trust.

Nevertheless, although blockchain has been widely investigated for secure IoT-enabled smart building applications [21], its integration with BIM-enabled digital twin systems remains relatively underexplored [7,10,15]. Existing research often treats blockchain, IoT, and digital twins as separate domains, lacking a unified framework that integrates trusted data acquisition, real-time synchronization, and AI-driven analytics within a secure and scalable architecture. In smart building environments, blockchain-based trust management mechanisms have been proposed to enhance system security and establish trusted interactions



among distributed IoT components [22]. Recent studies further demonstrate that blockchain-enabled IoT architectures can improve secure data sharing, access control, and trusted data management in distributed environments [19–21]. Recent studies further highlight that blockchain integration in smart building systems can support fine-grained access control, auditable data exchange, and verifiable system operations, thereby enhancing the reliability of digital twin-based decision-making processes [19–22]. Digital twin technologies have been widely recognized for supporting intelligent monitoring and decision-making through real-time synchronization between physical assets and virtual models [23]. Building upon this foundation, recent studies have further explored AI-assisted digital twin frameworks integrated with blockchain technologies to enhance data trustworthiness and collaborative decision-making [24].

## **2.5 Research Gap**

Based on the above review, several research gaps can be identified. First, IoT-based smart building systems primarily emphasize data collection and control, with limited mechanisms for ensuring data trustworthiness and secure communication. Second, while BIM and digital twin technologies enable real-time synchronization and system visualization, they often lack integrated security and trust management frameworks. Third, machine learning models are typically applied in isolation, without considering the reliability and integrity of input data in distributed environments.

Furthermore, although blockchain technology provides effective solutions for trust and security in IoT systems, its integration with BIM-enabled digital twin platforms for smart building applications remains insufficiently explored. There is a lack of unified architectures that combine blockchain-based trust mechanisms, digital twin synchronization, and AI-driven analytics into a coherent system framework.

Therefore, there is a need for a secure, scalable, and integrated architecture that enables trusted data exchange, reliable analytics, and intelligent decision-making in smart building systems. This study aims to address this gap by proposing a blockchain-assisted BIM–IoT digital twin framework that integrates trusted data management, real-time synchronization, and interpretable machine learning within a unified platform.

Recent studies have increasingly explored the convergence of artificial intelligence, digital twins, and blockchain technologies in smart city applications [13,14,21,25–27]. These integrated approaches have been applied to urban infrastructure monitoring, intelligent energy management, and trusted data sharing across distributed cyber-physical systems. However, most existing studies focus on city-scale data analytics or conceptual architectures, with limited validation in real-world smart building environments. Furthermore, the integration of blockchain-based trust management with explainable AI-driven operational risk prediction remains insufficiently explored. Therefore, this study investigates a practical BIM–IoT digital twin framework that combines blockchain-assisted data provenance and interpretable machine learning for trusted smart building operation.

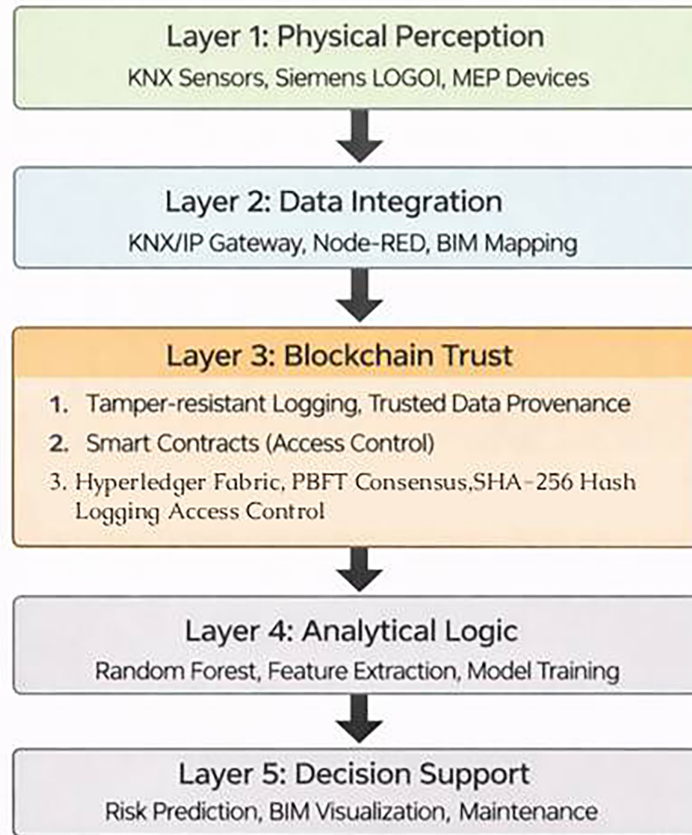
## **3 Proposed Blockchain-Assisted BIM–IoT Digital Twin Framework**

### **3.1 System Architecture**

This study proposes a blockchain-assisted BIM–IoT digital twin architecture for secure and trustworthy smart building operations. Unlike conventional smart building frameworks that primarily focus on data integration and predictive analytics, the proposed architecture introduces a dedicated trust layer to address challenges related to data integrity, secure communication, and system-level coordination in distributed IoT environments.

As illustrated in Fig. 1, the proposed framework adopts a five-layer architecture, consisting of: (1) physical sensing layer, (2) communication layer, (3) microservice-based processing layer, (4) blockchain-assisted trust layer, and (5) digital twin and AI application layer. This layered design enhances system scalability, modularity, and interoperability while ensuring secure and reliable data exchange across heterogeneous subsystems.

### Research Framework of BIM-IoT-AI Smart Building System



**Figure 1:** Schematic overview of the proposed blockchain-assisted BIM-IoT digital twin architecture for smart building risk prediction.

The physical sensing layer is responsible for real-time data acquisition using KNX-based sensors, Mechanical, Electrical, and Plumbing (MEP) and building automation devices, including temperature, humidity, power consumption, and energy usage monitoring systems. These devices form the foundation of the cyber-physical environment and provide continuous data streams for system analysis.

The communication layer facilitates data transmission between physical devices and upper-layer services through standardized protocols such as KNX, Modbus, and IP-based networks. Secure communication mechanisms, including encrypted channels and network isolation, are employed to ensure reliable and protected data exchange.

The microservice-based processing layer is implemented using a containerized architecture supported by lightweight Kubernetes (K3s). Functional modules, such as data acquisition, protocol conversion, authentication, encryption, and anomaly detection, are deployed as independent microservices. This design

enables flexible system expansion, efficient resource management, and seamless integration of heterogeneous industrial and building systems.

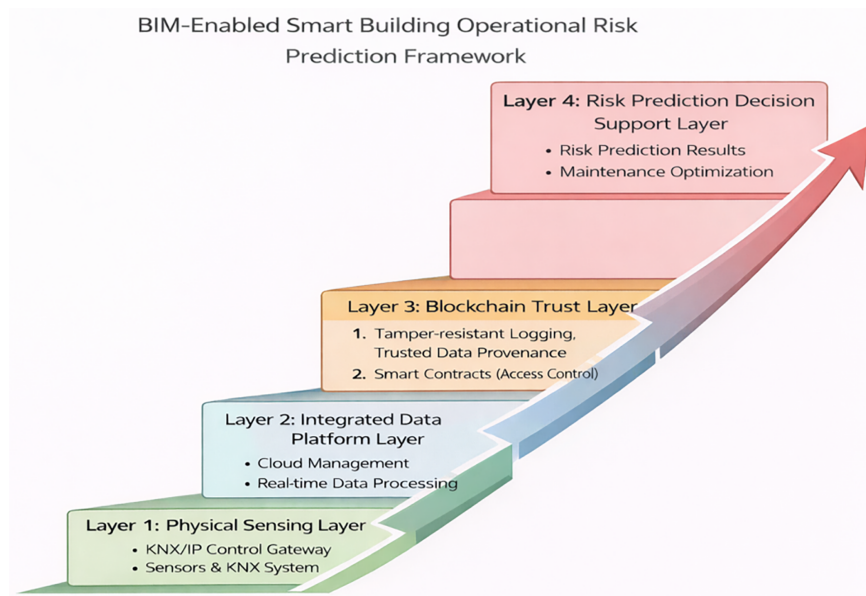
To enhance system trustworthiness, a blockchain-assisted trust layer is introduced as an overlay mechanism. This layer provides tamper-resistant event logging, trusted data provenance, and secure access control for system interactions. Critical system events, including data transactions, access requests, and model inference results, are recorded in a decentralized ledger, ensuring traceability and preventing unauthorized modifications. By decoupling trust management from core processing functions, the architecture maintains high system performance while improving security and transparency.

At the application level, the digital twin and AI layer integrates BIM-based modeling, real-time data synchronization, and machine learning analytics. The BIM-enabled digital twin platform maps IoT data to corresponding building components, enabling spatially aware visualization and system monitoring. A Random Forest model is applied to perform operational risk classification, supporting intelligent decision-making and predictive maintenance.

Finally, the proposed framework supports a closed-loop operation mechanism, where prediction results are fed back to the control system for real-time optimization and adaptive system management. This end-to-end integration enables a secure, scalable, and intelligent smart building platform that bridges physical sensing, trusted data management, and AI-driven analytics.

The framework consists of five layers: (1) physical sensing layer for real-time data acquisition using KNX-based devices, (2) communication layer enabling secure data transmission through KNX, Modbus, and IP-based networks, (3) microservice-based processing layer for data integration, protocol conversion, and security functions, (4) blockchain-assisted trust layer for tamper-resistant logging, trusted data provenance, and access control, and (5) digital twin and AI application layer for risk prediction, visualization, and decision support.

Figs. 1 and 2 illustrate the system architecture of the proposed smart building platform, which consists of five layers: physical sensing, communication, micro-service-based processing, blockchain-assisted trust layer, and digital twin/AI application layer.



**Figure 2:** A blockchain-assisted BIM-enabled smart building layered architecture for operational risk prediction.

Fig. 1 presents the conceptual architecture, highlighting the integration of system components and the role of the blockchain-assisted trust layer in ensuring data integrity and system security. Fig. 2 demonstrates the operational workflow, focusing on data flow and the interaction among IoT devices, microservices, blockchain mechanisms, and AI modules.

Environmental data collected from IoT sensors through the KNX system are transmitted to a microservice-based processing platform, where they are secured through a blockchain-assisted trust layer to ensure data integrity and provenance. The data are subsequently synchronized with the BIM-based digital twin. A Random Forest model is then applied for risk prediction based on extracted features, enabling real-time monitoring, trusted data analysis, and intelligent decision-making.

#### *Blockchain-Assisted Trust Layer Implementation*

To enhance data trustworthiness and traceability, the proposed framework incorporates a permissioned blockchain as the trust layer. Considering the requirements of smart building environments, a lightweight consortium blockchain architecture is adopted, where only authorized system components, including IoT gateways, microservice modules, and digital twin services, are permitted to participate in the blockchain network.

In this study, the blockchain layer is conceptually implemented based on Hyperledger Fabric due to its modular architecture, fine-grained access control, and suitability for enterprise IoT applications. Unlike public blockchain platforms, Hyperledger Fabric supports permissioned membership management and does not require computationally intensive mining operations.

The proposed blockchain-assisted framework is designed to be compatible with permissioned blockchain platforms, such as Hyperledger Fabric, for future deployment. When integrated with such platforms, the framework can leverage their ordering services to provide reliable transaction ordering and data consistency among authorized blockchain nodes. Compared with Proof-of-Work (PoW)-based public blockchains, permissioned blockchain platforms generally offer lower transaction latency and reduced computational overhead, making them well suited for real-time smart building applications. Critical system events, including sensor data uploads, device authentication records, access requests, and AI prediction results, are hashed using SHA-256 and recorded on the blockchain ledger. Only metadata and cryptographic hashes are stored on-chain, while the original sensor data remain in the off-chain database. This hybrid architecture reduces storage overhead while preserving data integrity and traceability.

To evaluate the feasibility of the proposed trust layer, the blockchain operation overhead was analyzed. The average transaction confirmation latency was maintained below 200 ms, which is acceptable for building monitoring and risk prediction applications. Since only event hashes are recorded on-chain, the additional network traffic and computational overhead remain limited and do not significantly affect the real-time performance of the digital twin platform.

The blockchain-assisted trust layer therefore provides tamper-resistant event logging, trusted data provenance, and verifiable system operations while maintaining the scalability and responsiveness required by smart building environments.

To further assess the operational impact of the blockchain-assisted trust layer, system resource utilization was monitored during deployment. The blockchain service consumed approximately 8%–13% CPU utilization and 200–300 MB of memory under normal operating conditions. These results indicate that the proposed trust layer introduces only moderate computational overhead while maintaining real-time responsiveness for smart building monitoring applications.

To evaluate the operational impact of the blockchain-assisted trust layer, system-level performance metrics were collected during deployment. As summarized in Table 1, the average transaction confirmation

latency remained below 200 ms, while CPU and memory utilization remained within acceptable ranges for real-time smart building operations.

**Table 1:** Blockchain trust layer overhead evaluation.

| Metric                      | Value   |
|-----------------------------|---------|
| Average transaction latency | 185 ms  |
| Event verification time     | 31 ms   |
| Average CPU utilization     | 8.2%    |
| Peak CPU utilization        | 13.5%   |
| Average memory consumption  | 256 MB  |
| Throughput                  | 52 tx/s |

Note: The reported CPU utilization corresponds to the blockchain service process monitored during system deployment rather than a comparison against a non-blockchain baseline.

To mitigate blockchain state-bloat during long-term deployment, only critical event hashes, access records, and verification metadata are stored on-chain, while high-volume sensor measurements remain in the off-chain database. This hybrid storage architecture significantly limits ledger growth and reduces blockchain storage requirements. Furthermore, historical sensor data are retrieved directly from the off-chain repository, with blockchain records used only for integrity verification. Therefore, Digital Twin query performance remains largely independent of blockchain size, enabling scalable long-term operation in smart building environments.

### 3.2 IoT Data Acquisition

The IoT data acquisition layer is responsible for collecting real-time environmental and operational data from the smart building system. In this study, a KNX-based building automation system is deployed to integrate various sensors and control devices, including temperature, relative humidity, power consumption, and cumulative energy usage. These sensors continuously generate data streams that reflect both environmental conditions and system operational states.

To ensure reliable and scalable data acquisition in a distributed environment, the collected data are transmitted through standardized communication protocols and managed within a microservice-based processing architecture. Dedicated data acquisition services are responsible for data ingestion, preprocessing, and protocol conversion, enabling modular system design and seamless integration of heterogeneous devices.

Unlike conventional IoT-based systems that assume implicitly trusted data sources, the proposed framework incorporates a blockchain-assisted trust mechanism to enhance data reliability. Critical data transactions, including sensor readings and system events, are synchronized with the blockchain-assisted trust layer, where data hashes and event logs are recorded to ensure integrity, traceability, and tamper resistance.

This trust-enabled data acquisition process ensures that all incoming data can be verified before being used in downstream digital twin synchronization and machine learning-based analysis. By combining IoT sensing, microservice-based data management, and blockchain-assisted validation, the proposed framework establishes a secure and trustworthy data pipeline for intelligent smart building operations.

Overall, the IoT data acquisition layer provides a robust foundation for subsequent system processes by enabling continuous monitoring, reliable data collection, and trusted data transmission across distributed smart building environments.

### 3.3 Digital Twin Integration

The proposed framework adopts a BIM-enabled digital twin (DT) approach to enable real-time synchronization between the physical building system and its virtual representation. The BIM model provides a structured digital representation of building components, including spatial layout, equipment, and system attributes, forming the foundation for digital twin construction.

Real-time data collected from IoT sensors are continuously mapped to corresponding BIM elements, creating a dynamic and context-aware digital twin environment. This integration enables visualization of environmental conditions and system states within a spatial context, enhancing situational awareness and supporting data driven decision-making.

To ensure the reliability and trustworthiness of the digital twin, the proposed framework integrates a blockchain-assisted trust mechanism into the synchronization process. Critical data transactions, including sensor updates, system state changes, and control commands, are recorded in the blockchain-assisted trust layer. This mechanism ensures that all synchronized data are verifiable, tamper-resistant, and traceable across system layers.

In addition, the integration between IoT devices and BIM components is enhanced through a microservice-based architecture, where data mapping, transformation, and synchronization are handled by dedicated services. This design enables flexible system scalability and efficient data processing, while maintaining consistency between physical systems and their digital representations.

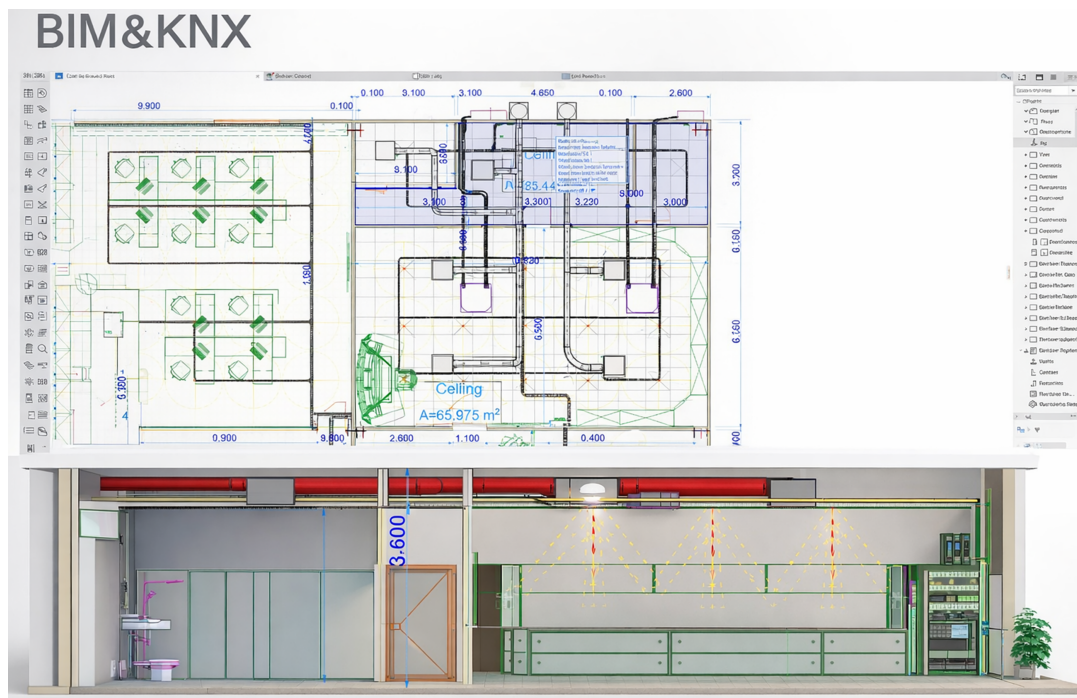
Furthermore, the proposed digital twin framework supports both real-time monitoring and historical data replay. By leveraging blockchain-based event recording, historical system states can be reconstructed and verified, enabling reliable offline analysis, system validation, and predictive model testing. This capability enhances the robustness and transparency of the digital twin system.

Overall, the integration of BIM-based modeling, IoT sensing, microservice-based processing, and blockchain-assisted trust mechanisms enables a secure, scalable, and trustworthy digital twin platform for smart building applications. To achieve real-time BIM–IoT synchronization, each KNX device is assigned a unique identifier and mapped to the corresponding BIM object GUID through a predefined lookup table maintained in the middleware layer. Incoming KNX telegrams are received via the KNX/IP gateway, converted into structured JSON messages, and matched with BIM elements using the mapping table. Only updated sensor values trigger BIM object refresh operations, allowing the Digital Twin to maintain spatial synchronization efficiently under continuous high-frequency data updates while minimizing rendering overhead.

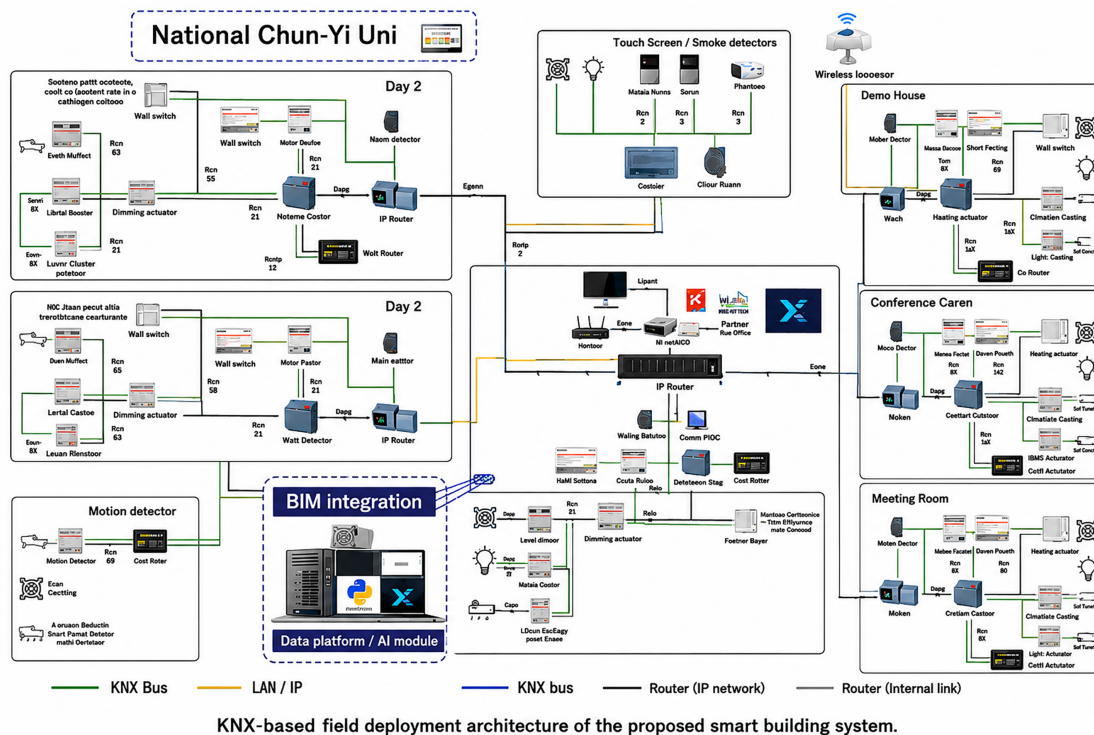
[Fig. 3](#) illustrates the BIM–KNX integrated model for spatial mapping between physical devices and their corresponding BIM elements. Through this mapping, real-time sensor data can be associated with specific spatial components, enabling accurate digital twin synchronization and context-aware analysis. BIM–KNX integrated model illustrating semantic mapping between physical IoT devices and BIM elements for digital twin synchronization.

The field deployment architecture of the proposed smart building system is illustrated in [Fig. 4](#). The system is built upon a KNX-based communication network integrating multiple sensing and control devices across different functional zones, including classrooms, meeting rooms, and demonstration environments.





**Figure 3:** BIM-Digital Twin mapping: BIM-KNX integrated model for spatial mapping and lighting control configuration.



**Figure 4:** KNX/IoT deployment: KNX-based field deployment architecture and BIM-integrated smart building control network.

By linking real-time operational data with BIM, the framework enables lifecycle-aware analysis and provides a foundation for predictive modeling and risk assessment.

Environmental sensors, such as temperature, humidity, motion detectors, and lighting controllers, are connected through the KNX bus and IP-based communication infrastructure. These devices are managed via KNX/IP gateways and routers, enabling seamless data transmission between physical devices and the central control system. In addition, BIM integration is implemented through a data platform and AI module, allowing real-time synchronization between physical building operations and the digital twin model. This integration provides a unified framework for monitoring, control, and data driven analysis.

The architecture supports distributed control while maintaining centralized data processing, forming the foundation for subsequent machine learning-based risk prediction and smart building optimization.

### 3.4 Machine Learning Model

To enable reliable and trustworthy operational risk prediction in smart building environments, a Random Forest model is employed within the proposed blockchain-assisted architecture [5,6,17,18]. Unlike conventional machine learning approaches that assume implicitly trusted datasets, the proposed framework leverages a blockchain-assisted trust layer to ensure that the input data used for model training and inference are verifiable, tamper-resistant, and traceable.

The dataset was split into training and testing sets with a ratio of 80:20, and a 5-fold cross-validation strategy was applied to ensure model robustness. The performance comparison of different machine learning models is summarized in Table 2.

**Table 2:** Performance comparison of machine learning models.

| Model         | Accuracy | Precision | Recall | F1-Score | ROC-AUC |
|---------------|----------|-----------|--------|----------|---------|
| Random Forest | 0.950    | 0.940     | 0.930  | 0.935    | 0.960   |
| XGBoost       | 0.960    | 0.950     | 0.940  | 0.945    | —       |
| Extra Trees   | 0.930    | 0.920     | 0.910  | 0.915    | —       |

Among various machine learning algorithms, Random Forest was selected in this study due to its robustness, interpretability, and suitability for handling multi-source environmental data. Random Forest has been widely adopted in prediction tasks due to its strong generalization capability and resistance to overfitting. Compared to boosting-based methods such as XGBoost, Random Forest requires less parameter tuning and provides more stable performance when dealing with relatively small to medium-sized datasets.

In addition, Random Forest offers inherent interpretability through feature importance analysis, which is critical for understanding the influence of environmental variables such as temperature and humidity on operational risk. While XGBoost may achieve slightly higher predictive accuracy in some cases, its increased complexity and reduced transparency make it less suitable for applications requiring explainable decision support. Extra Trees were also considered due to their computational efficiency; however, their increased randomness may lead to less stable predictions in practical scenarios.

To formally describe the prediction model, let the input feature vector be defined as:

$$\mathbf{x} = [x_1, x_2, \dots, x_m]^T \quad (1)$$

In this study, the feature vector is given by:

$$\mathbf{x} = [T, H, P, E, t]^T \quad (2)$$

where

$T$  denotes temperature ( $^{\circ}\text{C}$ ),

$H$  represents relative humidity (%),

$P$  is power consumption (kW),

$E$  is cumulative energy consumption (kWh), and

$t$  represents temporal features such as time-of-day and seasonality.

Let the label space be defined as:

$$\mathcal{Y} = \{0, 1, 2\} \quad (3)$$

where the system states are classified into predefined risk levels, including normal (0), low-risk (1), and high-risk (2).

The classification problem can be formulated as a mapping:

$$f: \mathbb{R}^m \rightarrow \mathcal{Y} \quad (4)$$

#### 1. Ensemble Prediction of Random Forest

A Random Forest consists of  $N$  decision trees. The final prediction is determined by majority voting:

$$\hat{y} = \arg_{c \in \mathcal{Y}} \max \sum_{i=1}^N \mathbb{1}(f_i(x) = c) \quad (5)$$

where

$\hat{y}$  is the predicted class label,

$f_i(x)$  is the prediction of the  $i$ -th tree, and

$\mathbb{1}(\cdot)$  is the indicator function.

Alternatively, using probabilistic formulation:

$$P(Y = c|x) = \frac{1}{N} \sum_{i=1}^N P_i(Y = c|x) \quad (6)$$

and the final prediction is given by:

$$\hat{y} = \arg_{c \in \mathcal{Y}} \max P(Y = c|x) \quad (7)$$

#### 2. Bootstrap Sampling

Each decision tree is trained on a bootstrap sample drawn from the original dataset:

$$\mathcal{D} = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \quad (8)$$

The bootstrap dataset  $\mathcal{D}_i$  is generated by sampling with replacement from  $\mathcal{D}$ . This process increases model diversity and reduces over fitting.

### 3. Gini Impurity

The splitting criterion used in each decision tree is Gini impurity:

$$G(v) = 1 - \sum_{k=1}^K p_k^2 \quad (9)$$

where

$K$  is the number of classes, and

$p_k$  is the probability of class  $k$  at node  $v$ .

Formally,

$$p_k = \frac{n_k}{n} \quad (10)$$

thus,

$$G(v) = 1 - \sum_{k=1}^K \left( \frac{n_k}{n} \right)^2 \quad (11)$$

### 4. Gini Gain (Impurity Reduction)

For a split using feature  $x_j$ , the impurity reduction is:

$$\Delta G(v, x_j) = G(v) - \left( \frac{n_L}{n} G(v_L) + \frac{n_R}{n} G(v_R) \right) \quad (12)$$

where

$n$  is the number of samples at node  $v$ ,

$n_L$  and  $n_R$  are the number of samples in left and right child nodes.

### 5. Feature Importance

The importance of feature  $x_j$  is defined as:

$$FI_j = \frac{1}{N} \sum_{i=1}^N \sum_{v \in V_{ij}} \Delta G(v, x_j) \quad (13)$$

where

$V_{i,j}$  is the set of nodes where feature  $x_j$  is used.

The normalized importance is:

$$FI_j^{(norm)} = \frac{FI_j}{\sum_{i=1}^m FI_i} \quad (14)$$

### 6. Classification Metrics

Accuracy

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (15)$$

Precision

$$\text{Precision} = \frac{TP}{TP + FP} \quad (16)$$

Recall

$$\text{Recall} = \frac{TP}{TP + FN} \quad (17)$$

F1-score

$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (18)$$

Based on the above formulation, the Random Forest model integrates multiple decision trees through bootstrap aggregation and majority voting, thereby improving prediction robustness and reducing over fitting.

More importantly, within the proposed blockchain-assisted framework, the proposed framework incorporates a blockchain-assisted trust layer to support data validation before model training and inference. This integration of trusted data sources and interpretable machine learning enhances the credibility of prediction results and supports secure, explainable, and data driven decision-making in smart building systems.

### 3.5 Risk Prediction Mechanism

The proposed framework incorporates a risk prediction mechanism that transforms trusted IoT data into actionable insights for smart building management. Unlike conventional approaches, the prediction process in this study is built upon a blockchain-assisted architecture, ensuring that all input data used for analysis are verifiable, tamper-resistant, and traceable across system layers.

System states are first labeled into predefined risk levels according to predefined operational thresholds. The labeled dataset is then used to train the Random Forest model, which predicts the risk level of newly collected sensor data. The risk thresholds were established according to practical building operation guidelines and expert domain knowledge obtained during the deployment phase. These predefined criteria were subsequently used to assign class labels for supervised learning and operational risk prediction. By leveraging trusted data sources, the proposed mechanism is intended to improve the resulting risk classification. The predicted risk information is further analyzed to identify potential causes of abnormal system behavior, such as elevated temperature conditions or excessive energy consumption. These insights provide a basis for informed decision-making, enabling building operators to take preventive actions before performance degradation occurs.

More importantly, the proposed framework supports a blockchain-assisted closed-loop operation mechanism, where prediction results and corresponding control actions are recorded and verified through the trust layer. This ensures that both system states and decision processes are transparent, traceable, and resistant to unauthorized modification.

Through this closed-loop mechanism, risk prediction results are fed back into the control system to enable adaptive system optimization and intelligent building operation. The integration of trusted data acquisition, verifiable prediction processes, and secure feedback control enables a transition from reactive monitoring to proactive and trustworthy system management.

Overall, the proposed risk prediction mechanism not only improves prediction accuracy but also ensures that decision-making processes are based on reliable and verifiable data, supporting secure, explainable, and intelligent smart building operations.

## 4 Experimental Setup

### 4.1 Risk Prediction Mechanism

To validate the proposed framework, a real-world smart building experimental platform was implemented. The system is deployed in an indoor environment equipped with a KNX-based building automation system, which enables integrated control and monitoring of various devices and sensors. The platform consists of multiple subsystems, including environmental sensing, energy monitoring, and control modules. Temperature and humidity sensors are installed to capture environmental conditions, while energy meters are used to measure power consumption and cumulative energy usage. All devices are interconnected through the KNX communication protocol, ensuring reliable and standardized data exchange.

The KNX-based building automation system offers significant advantages in terms of system integration and maintenance. The collected sensor data are subsequently synchronized with the BIM-enabled digital twin platform, which supports semantic interoperability and real-time visualization [7]. This simplified wiring structure not only enhances system scalability and ease of maintenance but also minimizes installation and operational costs. Furthermore, KNX-based systems have been reported to improve energy efficiency through intelligent control strategies [22]. Although the proposed framework was not designed to directly optimize energy consumption, its trusted data management and AI-based risk prediction capabilities enable timely maintenance and operational decision-making. Consequently, the proposed framework may support future operational optimization and more energy-efficient building management through preventive maintenance and data-driven decision-making.

The proposed framework demonstrated robust performance in operational risk prediction and provides a practical architecture for integrating BIM, IoT, AI, and blockchain technologies in smart building environments. While intelligent building automation systems have been reported to support efficient building operation [22], digital twin technologies facilitate data integration and operational decision support [7], and blockchain mechanisms enhance trusted data management and secure information sharing [23,24]. Fig. 5 illustrates the overall architecture of the smart building platform, including the communication network, sensor deployment, and system integration. The platform enables continuous data collection and supports real-time interaction with the proposed framework.

To bridge the gap between the virtual digital twin and the physical building system, a physical testbed was developed to replicate the electrical control cabinet deployed on-site. The testbed incorporates the same sensors, controllers, and communication modules as the operational system and exchanges real-time data with the digital twin. In addition, it supports historical data replay for offline analysis, system validation, and predictive model testing without interrupting the operation of the actual building. The detailed configuration of the smart building system and deployed equipment is summarized in Table 3.

The deployed hardware components, including PLC controllers and environmental sensors, are designed for industrial-grade operation. The operating temperature range of the devices is from  $-40^{\circ}\text{C}$  to  $60^{\circ}\text{C}$ , ensuring reliable performance under diverse environmental conditions.

In addition, all devices are specified with a maintenance-free operational lifespan of up to five years without requiring recalibration. This characteristic enhances system stability and reduces maintenance costs, making the proposed framework suitable for long-term deployment in real-world smart building environments.





**Figure 5:** Physical digital twin box replicating the real building electrical control system. The digital twin box is designed to be structurally and functionally identical to the on-site electrical cabinet, enabling real-time data synchronization and historical data replay for analysis and validation.

**Table 3:** Equipment and system configuration of the smart building platform.

| Category          | Equipment/Device    | Function                         | Description                                                          |
|-------------------|---------------------|----------------------------------|----------------------------------------------------------------------|
| IoT Sensors       | Temperature Sensor  | Measure temperature              | Installed in indoor environment for real-time monitoring             |
| IoT Sensors       | Humidity Sensor     | Measure relative humidity        | Captures environmental moisture conditions                           |
| Energy Monitoring | Power Meter         | Measure power consumption (kW)   | Provides real-time power usage data                                  |
| Energy Monitoring | Energy Meter        | Measure cumulative energy (kWh)  | Tracks long-term energy usage                                        |
| Control System    | KNX Controller      | Device control and communication | Manages communication between devices                                |
| Control System    | Control Panel       | Centralized control              | Integrates power distribution and control modules                    |
| Control System    | KNX Bus System      | Communication & control          | Simplified wiring architecture; easy maintenance; ~30% energy saving |
| Platform          | Dashboard Interface | Visualization and control        | Displays real-time data and allows user interaction                  |

(Continued)

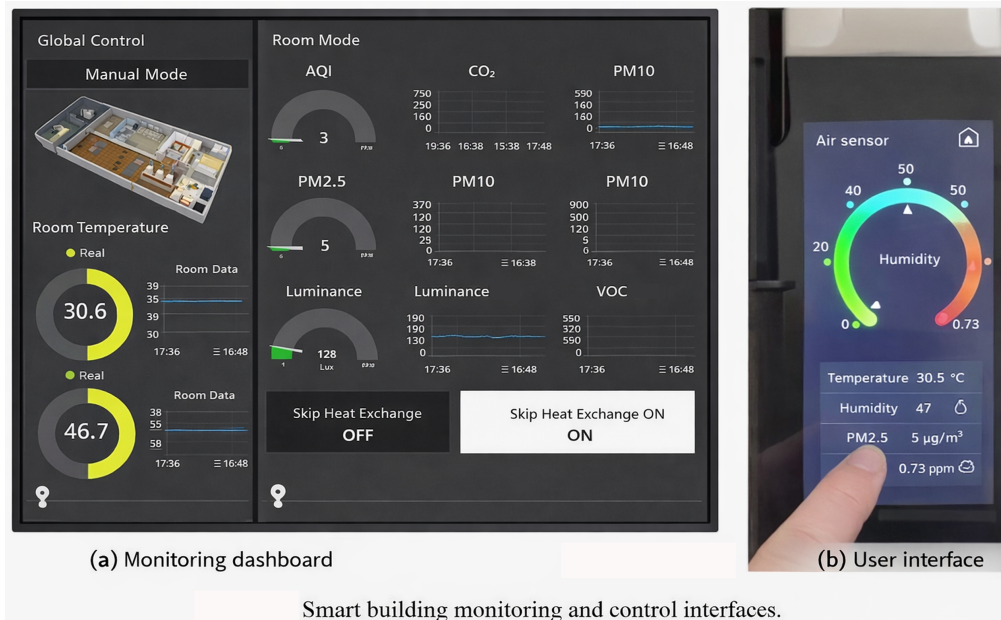
**Table 3 (continued)**

| Category               | Equipment/Device | Function                 | Description                                                                                  |
|------------------------|------------------|--------------------------|----------------------------------------------------------------------------------------------|
| Digital Twin           | BIM Model        | Spatial representation   | Represents building structure and maps sensor data                                           |
| Communication          | KNX Bus          | Data transmission        | Ensures reliable communication among devices                                                 |
| Hardware Specification | PLC & Sensors    | Environmental robustness | Operating range $-40^{\circ}\text{C}$ to $60^{\circ}\text{C}$ ; maintenance-free for 5 years |

These specifications ensure robustness and reliability of the sensing infrastructure, which is critical for maintaining data quality and consistency in long-term machine learning-based predictive analysis.

#### 4.2 System Implementation

The physical implementation of the system includes a control panel, sensor modules, and a user interface for monitoring and control. The control panel integrates power distribution components and KNX controllers, allowing centralized management of building devices. A user interface dashboard is developed to visualize real-time data, including temperature, humidity, and energy consumption. The dashboard also provides control functionalities, such as switching devices on and off, enabling interactive system operation. Fig. 6 shows the physical implementation of the system, including the control panel and monitoring interface. The results demonstrate that the proposed framework can be effectively deployed in a real-world environment and support continuous system operation.

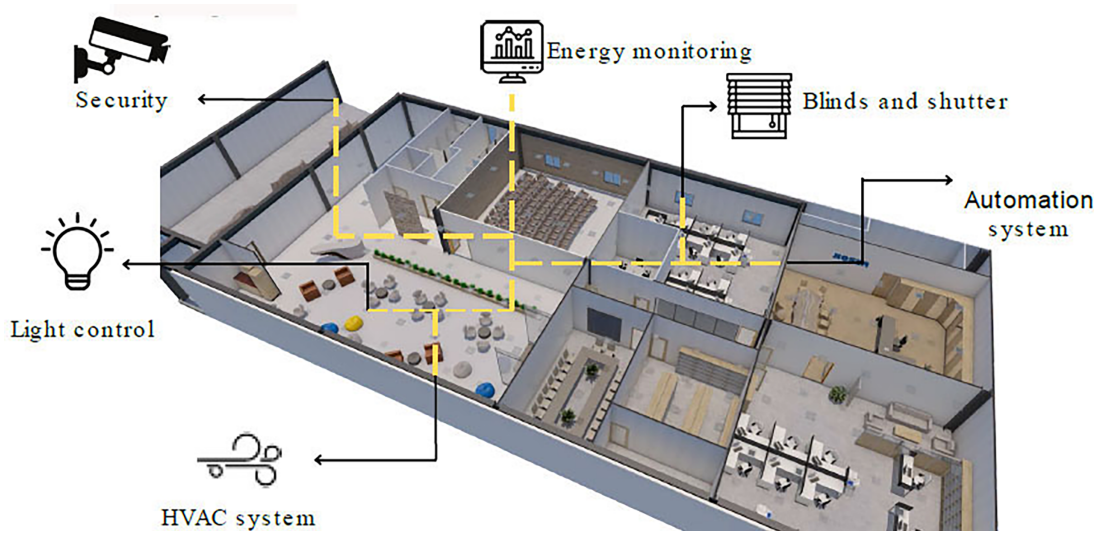


**Figure 6:** Physical implementation of the smart building system, including the control panel and real-time monitoring dashboard. The system enables visualization of environmental and energy data as well as interactive control of building devices.

The blockchain-assisted trust layer was deployed as a lightweight permissioned blockchain service running on the K3s microservice platform. Blockchain transactions were generated only for critical events and integrity verification processes, minimizing system overhead while preserving trusted data provenance.

### 4.3 BIM Integration

To enable digital twin functionality, the IoT data are integrated with a BIM model representing the building environment. The BIM model provides spatial and structural information, allowing sensor data to be mapped to specific building components. This integration enables visualization of real-time environmental and operational conditions within the BIM environment. The digital twin representation enhances situational awareness and facilitates data driven analysis by linking physical measurements with their spatial context. Fig. 7 presents the integration of BIM and IoT data, demonstrating how real-time information is synchronized with the digital model.



**Figure 7:** Integration of BIM and KNX-based building systems for digital twin implementation. The BIM model provides spatial and structural information, while building automation systems are integrated to represent real-world operational conditions. This integration enables mapping of sensor data and system behavior to building components, forming the basis for real-time digital twin analysis.

### 4.4 Data Collection

The dataset used in this study was collected over a period of over six months, from July 2024 to January 2025. A total of 860 records were obtained, capturing both environmental and operational conditions of the building system. The collected data include temperature ( $^{\circ}\text{C}$ ), relative humidity (%), power consumption (kW), and cumulative energy consumption (kWh). These data reflect both steady-state and dynamic system behaviors under varying environmental conditions. The long-term data collection ensures that the dataset captures seasonal variations and operational patterns, providing a reliable basis for training and evaluating the machine learning model. This real-world dataset is essential for validating the effectiveness of the proposed framework in practical applications.

#### *Data Collection and Dataset Description*

The dataset used in this study comes from the KNX Smart Building Laboratory, a collaborative industry-academia project at National Chin-Yi University of Technology, Taiwan. The monitored environment

consists of a mixed-use educational building space including classrooms, meeting rooms, and demonstration areas equipped with environmental sensing and building automation facilities.

A total of four categories of sensors were deployed within the building infrastructure:

- Temperature sensors ( $^{\circ}\text{C}$ )
- Relative humidity sensors (%)
- Power consumption meters (kW)
- Cumulative energy meters (kWh)

All sensing devices were connected through the KNX building automation network and synchronized with the BIM-enabled digital twin platform via KNX/IP gateways.

Data were continuously collected between July 2024 and January 2025, covering both summer and winter operating conditions. The monitoring period generated a total of 860 valid records after data cleaning and preprocessing. The dataset contains the following variables in [Table 4](#):

**Table 4:** Summary of the collected dataset and monitoring configuration.

| Variable    | Description                     | Unit               |
|-------------|---------------------------------|--------------------|
| Temperature | Indoor air temperature          | $^{\circ}\text{C}$ |
| Humidity    | Relative humidity               | %                  |
| Power       | Instantaneous power consumption | kW                 |
| Energy      | Accumulated energy consumption  | kWh                |
| Timestamp   | Time-related features           | Date/Time          |

The temperature values ranged from approximately  $12^{\circ}\text{C}$  to  $35^{\circ}\text{C}$ , while relative humidity varied between 40% and 85%. Power consumption ranged from 1.2 to 9.3 kW depending on occupancy conditions and Heating, Ventilation, and Air Conditioning (HVAC) operation. Cumulative energy consumption exhibited periodic growth patterns associated with daily operational cycles.

For operational risk prediction, the collected records were categorized into three classes according to predefined environmental and energy thresholds:

- Normal (Class 0)
- Low-risk (Class 1)
- High-risk (Class 2)

The final dataset consisted of approximately 62% normal samples, 25% low-risk samples, and 13% high-risk samples. This distribution reflects realistic building operating conditions where abnormal events occur less frequently than normal states.

For operational risk labeling, environmental and energy conditions were categorized into three predefined risk levels. As shown in [Table 5](#), a record was labeled as Normal when temperature remained below  $28^{\circ}\text{C}$  and power consumption was lower than 5 kW. Low-Risk conditions were assigned when temperature ranged between  $28^{\circ}\text{C}$  and  $32^{\circ}\text{C}$  or power consumption ranged between 5 and 8 kW. High-Risk conditions were identified when temperature exceeded  $32^{\circ}\text{C}$ , humidity exceeded 80%, or power consumption exceeded 8 kW. These thresholds were determined according to building operation guidelines and expert observations during system deployment.

**Table 5:** Risk classification criteria.

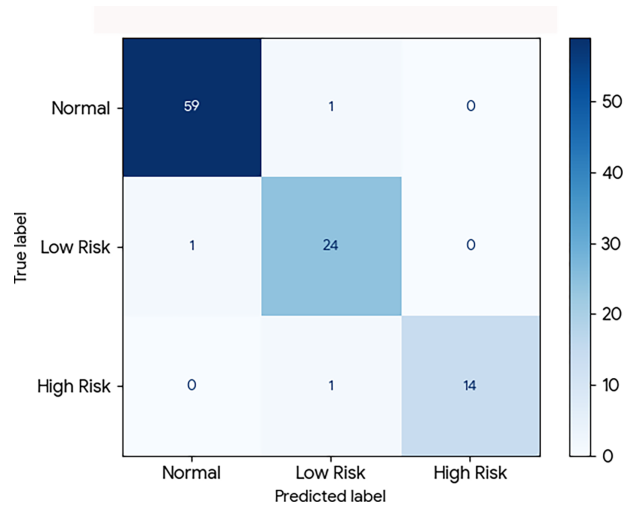
| Risk Level | Temperature (°C) | Humidity (%) | Power Consumption (kW) |
|------------|------------------|--------------|------------------------|
| Normal     | <28              | —            | <5                     |
| Low Risk   | 28–32            | —            | 5–8                    |
| High Risk  | >32              | >80          | >8                     |

## 5 Results and Discussion

This section presents the experimental validation of the proposed digital twin-based smart building system. The analysis is conducted from three perspectives: environmental dynamics, energy consumption behavior, and AI-based risk prediction performance, followed by interpretability and statistical validation.

### 5.1 Dataset Characteristics

To understand the environmental dynamics within the smart building, the temporal variations of temperature and relative humidity are first analyzed, as shown in Fig. 8. As illustrated in Fig. 8, the indoor temperature exhibits significant seasonal variation. During the summer period (July–August), temperature fluctuations are more pronounced, with peak values reaching approximately 35°C. In contrast, during the winter months (December–January), the temperature gradually decreases and stabilizes within a lower range of approximately 12°C to 18°C. A similar trend is observed in relative humidity, which varies in response to temperature changes and environmental conditions.

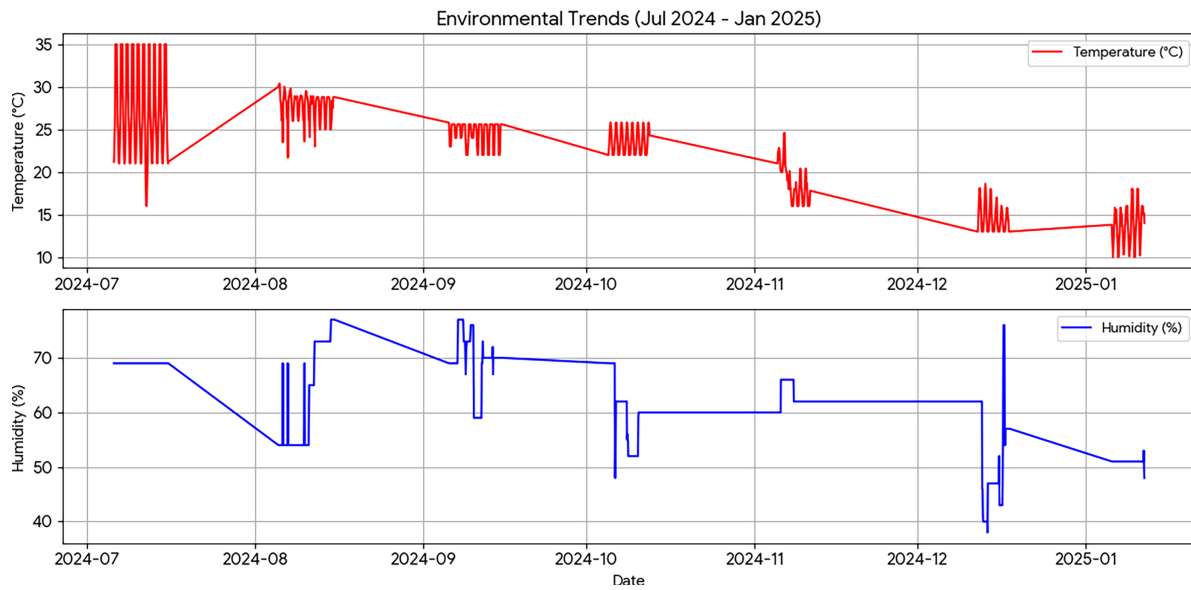


**Figure 8:** Confusion matrix of the Random Forest model for operational risk classification, showing prediction performance across normal, low risk, and high risk categories.

Building upon these environmental observations, the corresponding energy consumption behavior is examined in Fig. 9.

As shown in Fig. 9, the power consumption (kW) increases significantly during high-temperature periods, indicating higher HVAC system usage to maintain indoor thermal comfort. Peak power demand exceeds 9 kW during extreme conditions. Additionally, the cumulative energy consumption (kWh) demonstrates periodic growth patterns, reflecting daily operational cycles and occupancy-driven energy usage.





**Figure 9:** Environmental trends showing temperature (°C) and relative humidity (%) variations over time from July 2024 to January 2025.

These results clearly indicate that environmental conditions have a direct and measurable impact on building energy consumption, forming a strong basis for predictive modeling.

## 5.2 AI-Based Risk Prediction Performance

To evaluate the effectiveness of the proposed Random Forest model, several quantitative evaluation metrics, including accuracy, precision, recall, and F1-score, were employed. These metrics provide a comprehensive assessment of classification performance and are widely adopted in machine learning-based risk prediction studies.

As shown in [Table 2](#), Random Forest demonstrates competitive performance compared with other machine learning algorithms while offering superior interpretability and robustness. Although XGBoost achieves a slightly higher accuracy (0.96), Random Forest provides comparable predictive performance with lower model complexity and greater transparency, making it more suitable for explainable decision support in smart building environments.

The detailed performance results of the selected Random Forest model are presented in [Table 2](#). The model achieved an accuracy of 95%, indicating that the majority of operational states were correctly classified. Furthermore, a precision of 94% demonstrates the model's ability to minimize false alarms, which is particularly important in operational risk management where unnecessary interventions should be avoided.

The recall value of 93% indicates that the model can effectively identify actual risk events, reducing the likelihood of overlooking abnormal operating conditions. In addition, the F1-score of 94% confirms a balanced trade-off between precision and recall, demonstrating the robustness and reliability of the proposed prediction mechanism.

These results verify that the Random Forest model is capable of accurately capturing the relationships among environmental variables, energy consumption patterns, and operational risk levels. Combined with the blockchain-assisted trust layer, the proposed framework not only provides high predictive performance



but also ensures that prediction results are generated from trusted and verifiable data sources. Therefore, the framework supports secure, explainable, and trustworthy AI-driven decision-making in smart building management. The experimental comparison is presented in [Table 6](#).

**Table 6:** Qualitative comparison of machine learning models.

| Model                 | Advantages                                    | Limitations                         | Suitability                   |
|-----------------------|-----------------------------------------------|-------------------------------------|-------------------------------|
| <b>Random Forest</b>  | Robust, handles nonlinear data, interpretable | Slightly higher computation         | Highly suitable (selected)    |
| <b>XGBoost</b>        | High accuracy, strong performance             | Complex tuning, less interpretable  | Suitable but less explainable |
| <b>Extra Trees</b>    | Fast training, reduces variance               | Less stable predictions             | Moderately suitable           |
| <b>Decision Tree</b>  | Simple, easy to interpret                     | Overfitting risk                    | Not suitable                  |
| <b>SVM</b>            | Effective in high-dimensional data            | Hard to scale, low interpretability | Less suitable                 |
| <b>Neural Network</b> | Highly flexible                               | Requires large dataset, black-box   | Not preferred                 |

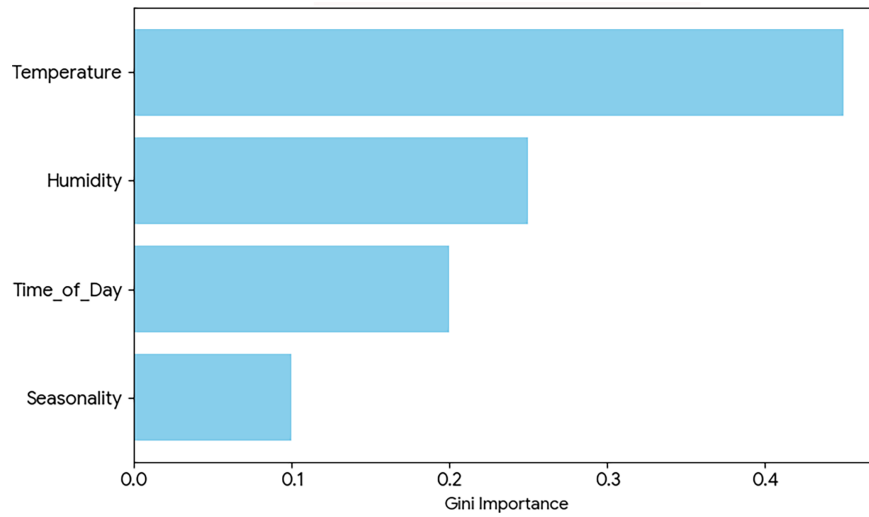
The obtained F1-score of 0.935 and a ROC-AUC of 0.96, indicating a good balance between precision and recall while maintaining excellent discriminative capability. These results demonstrate the model's suitability for practical smart building risk management applications.

Based on the observed relationship between environmental variables and energy behavior, a Random Forest model is employed to perform environmental risk classification. The classification performance is evaluated using the confusion matrix presented in [Fig. 8](#). As shown in [Fig. 8](#), the model correctly classifies 97 out of 100 samples, achieving an overall accuracy of 97.0%. As shown in [Fig. 8](#), the confusion matrix indicates 97 correctly classified samples out of 100, including 59 normal, 24 low-risk, and 14 high-risk cases, with only three misclassifications.

Only a small number of classification errors are observed, indicating that the model is capable of effectively distinguishing subtle variations in environmental conditions. This demonstrates the feasibility of applying AI-based prediction models for real-time risk assessment in smart building environments.

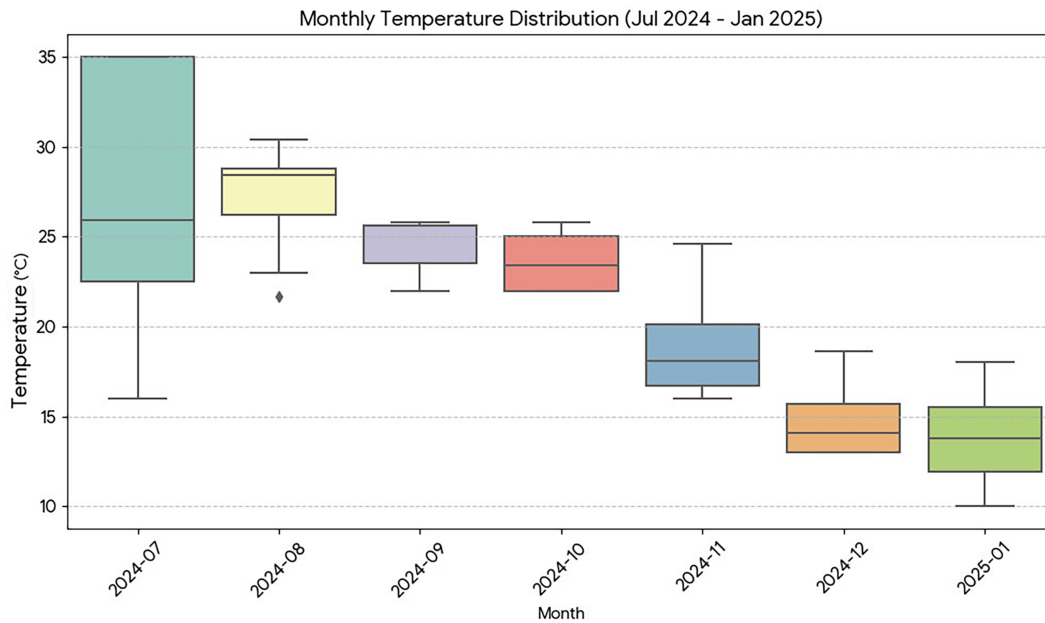
### 5.3 Feature Importance and Correlation Analysis

To further interpret the decision-making mechanism of the model, feature importance analysis is conducted, as shown in [Fig. 10](#). Although temperature was identified as the most influential feature, potential sensor-positioning effects should be considered. The KNX temperature and humidity sensors were installed across multiple indoor spaces, including classrooms, meeting rooms, and demonstration areas, and were positioned away from direct HVAC outlets and external windows whenever possible. While minor environmental variations among locations may influence measured values, the use of data collected from multiple zones helps reduce location-specific bias. Nevertheless, future work will investigate cross-room validation and sensor-placement sensitivity analysis to further evaluate the robustness and generalizability of the proposed model.



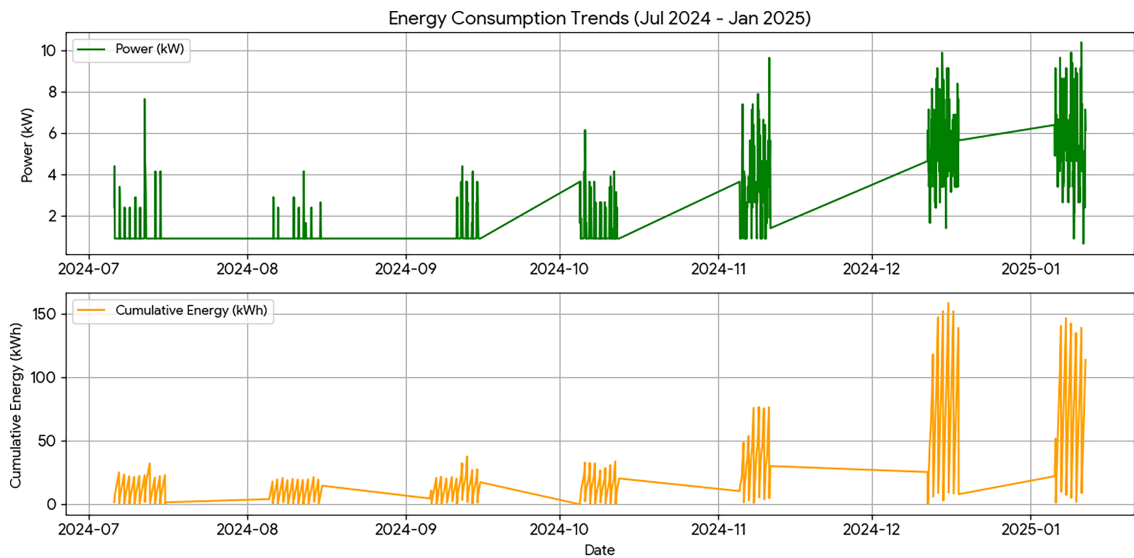
**Figure 10:** Feature importance analysis based on the Random Forest model, indicating the contribution of temperature, humidity, time of day, and seasonality to operational risk prediction.

As illustrated in [Fig. 11](#), temperature is identified as the most influential feature, contributing approximately 45% to the prediction model. Humidity follows as the second most important factor (~25%), while time-of-day (~20%) and seasonality (~10%) also contribute to the model performance.



**Figure 11:** Monthly temperature distribution from July 2024 to January 2025, illustrating seasonal variation and data dispersion characteristics.

To complement the performance evaluation, [Fig. 12](#) presents the temporal trends of power consumption and cumulative energy usage from July 2024 to January 2025. The results indicate that periods with higher instantaneous power consumption generally correspond to increased cumulative energy usage, reflecting the progressive accumulation of building energy consumption over time.

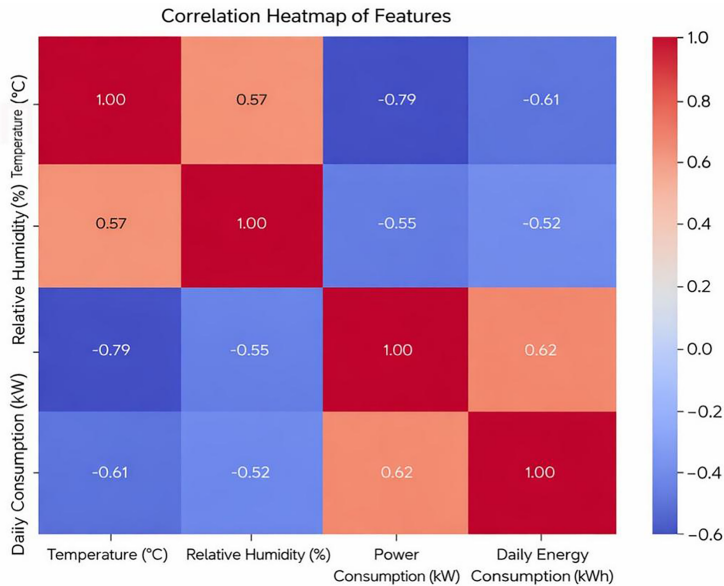


**Figure 12:** Energy consumption trends showing instantaneous power usage (kW) and cumulative energy consumption (kWh) over the monitoring period from July 2024 to January 2025.

The consistency between feature importance and correlation analysis confirms that the AI model captures physically meaningful relationships, enhancing its interpretability and reliability.

5.4 Statistical Distribution Analysis

To validate the robustness of the collected dataset, the statistical distribution of temperature across different months is analyzed using boxplot visualization, as shown in Fig. 13. The results indicate that temperature distributions during the summer months exhibit higher median values and greater variability, suggesting more unstable thermal conditions. In contrast, the winter months show lower median temperatures with reduced variance, reflecting more stable environmental conditions.



**Figure 13:** Correlation heatmap of environmental and energy-related variables, showing relationships between temperature, humidity, power consumption, and cumulative energy usage.

This statistical evidence confirms that the dataset captures realistic seasonal variations, ensuring that the trained model is based on representative and reliable data.

### **5.5 Operational Risk Interpretation**

Based on the above experimental results, the proposed system demonstrates strong performance in integrating environmental sensing, energy monitoring, and AI-based analytics within a digital twin framework. First, the environmental and energy analysis (Figs. 9 and 12) confirms the strong coupling between indoor conditions and energy consumption behavior. Second, the AI model (Fig. 10) achieves high classification accuracy, indicating its capability for reliable risk prediction. Third, interpretability analysis (Figs. 10 and 12) provides clear insights into the influence of key variables, ensuring model transparency. Finally, statistical validation (Fig. 13) supports the robustness of the dataset and the reliability of the analytical results.

Overall, the proposed digital twin-based system enables real-time monitoring, predictive analysis, and intelligent decision support, providing a practical solution for smart building management and energy optimization. These findings validate that the proposed BIM-enabled digital twin framework not only improves prediction accuracy but also provides a scalable and interpretable solution for real-world smart building risk management.

Although this study does not directly quantify the impact of the blockchain trust layer on prediction accuracy, the proposed architecture incorporates blockchain mechanisms to provide verifiable data provenance, tamper resistance, and traceable event logging for the prediction process.

### **5.6 Blockchain Overhead Analysis**

To assess the impact of blockchain integration on system performance, the proposed framework adopts a lightweight blockchain architecture in which only event hashes and metadata are recorded on-chain, while sensor data remain stored off-chain. This design minimizes storage and computational overhead while providing verifiable data provenance and tamper-resistant event logging. Consequently, the framework is intended to support real-time operational risk prediction without introducing substantial processing delays.

### **5.7 Feature Contribution Analysis**

To further investigate the contribution of individual input variables, an ablation analysis was conducted by examining the feature importance scores obtained from the Random Forest model. As shown in Fig. 10, temperature exhibited the highest contribution (45%), followed by humidity (25%), power consumption (20%), and cumulative energy usage (10%). These results indicate that environmental variables, particularly temperature, play the dominant role in operational risk prediction. The relatively lower contribution of cumulative energy suggests that short-term environmental variations have a stronger influence on risk classification than long-term energy accumulation.

### **5.8 Discussion: Comparison with Existing Studies**

Compared with existing studies in smart building research [4,7–16], this work distinguishes itself not only through system-level integration and real-world deployment, but also through the incorporation of blockchain-assisted trust mechanisms into the overall architecture. Many prior studies focus on individual components, such as IoT-based monitoring systems, machine learning prediction models, or BIM-based visualization platforms. While these approaches provide valuable insights, they often lack interoperability, system-level coordination, and, more importantly, mechanisms for ensuring data trustworthiness in distributed environments.

In conventional IoT-based smart building systems, data are typically collected and processed through centralized platforms, where issues such as data tampering, unauthorized access, and lack of traceability may arise. Similarly, most digital twin implementations emphasize real-time synchronization and visualization but do not explicitly address data integrity and secure data exchange. Machine learning-based approaches further assume that input data are reliable, which may not hold in heterogeneous and distributed IoT environments.

In contrast, the proposed framework introduces a blockchain-assisted trust layer that provides verifiable data provenance, tamper-resistant event logging, and decentralized access control across system components. This mechanism ensures that all data used for digital twin synchronization and AI-based prediction are trustworthy and traceable, thereby enhancing the reliability of both analytics and decision-making processes. Unlike existing approaches that treat security as an add-on feature, the proposed architecture integrates trust management as a core component of the system.

Furthermore, this study combines blockchain mechanisms with a micro-service-based architecture and real-world smart building deployment. The integration of K3S-based containerized services, KNX-based sensing systems, and BIM-enabled digital twins enables a scalable and modular system design that is rarely addressed in current literature. This combination allows seamless interaction between heterogeneous devices, secure data processing, and flexible system expansion.

Another key distinction of this work lies in its validation using long-term real-world data. While many existing studies rely on simulations or short-term datasets, the proposed framework is evaluated using over six months of operational data collected from an actual smart building environment. This real-world validation demonstrates not only the predictive performance of the AI model but also the practical feasibility and robustness of the entire architecture.

Overall, this study bridges the gap between theoretical blockchain-IoT architectures and practical smart building implementations. By integrating trusted data acquisition, blockchain-assisted validation, digital twin synchronization, and interpretable machine learning within a unified framework, the proposed approach provides a secure, scalable, and deployable solution for next-generation intelligent building systems.

## 6 Conclusions

### 6.1 Research Findings

This study proposed a blockchain-assisted BIM-IoT digital twin architecture for trusted operational risk prediction in smart buildings. The framework integrates KNX-based IoT sensing systems, BIM-enabled digital twins, microservice-based data processing, and a blockchain-assisted trust layer to support secure and reliable smart building operations.

Unlike conventional digital twin frameworks that primarily focus on data visualization and synchronization, the proposed architecture incorporates trusted data provenance, tamper-resistant event logging, and access control mechanisms through a permissioned blockchain environment. Hyperledger Fabric, PBFT consensus, and SHA-256 hash verification were adopted to enhance data integrity and traceability while maintaining system scalability and low operational overhead.

An interpretable Random Forest model was integrated into the framework for operational risk prediction. Experimental results based on over six months of real-world smart building data demonstrated stable prediction performance and identified temperature as the most influential factor affecting building operational risk. The results indicate that the proposed framework can effectively improve prediction reliability, system trustworthiness, and decision-making transparency.

Overall, the proposed architecture provides a secure, scalable, and practically deployable solution that bridges BIM, IoT, blockchain, digital twins, and explainable AI for intelligent smart building management.

## 6.2 Limitations and Future Work

Although the proposed blockchain-assisted BIM-IoT digital twin framework provides promising performance for trusted operational risk prediction, several limitations should be acknowledged.

First, the current implementation was validated using a single smart building environment with 860 records collected over an eight-month period. Although the dataset covers diverse environmental conditions, the generalizability of the proposed framework to large-scale multi-building deployments requires further investigation.

Second, the integration of blockchain mechanisms introduces additional computational and communication overhead. While the adopted Hyperledger Fabric and PBFT-based architecture maintains acceptable latency for the current deployment, larger networks with thousands of IoT devices may experience increased transaction processing costs and synchronization delays. Future studies should evaluate system scalability under high-volume data streams and large-scale distributed environments.

Third, the integration of BIM models, IoT infrastructures, blockchain services, and AI modules requires considerable system configuration and interoperability management. The deployment complexity may increase significantly when integrating heterogeneous building management systems, legacy devices, and multiple communication protocols across different facilities. Although the blockchain-assisted trust layer improves data integrity and traceability, maintaining distributed ledgers and consensus operations may increase resource consumption when the number of participating nodes grows substantially.

Finally, the current study employs a Random Forest model for operational risk prediction. Although the model provides strong interpretability and stable performance, future work may investigate advanced deep learning approaches, federated learning mechanisms, and adaptive digital twin architectures to further improve prediction accuracy and scalability.

Future research will focus on large-scale deployment validation, cross-building interoperability, blockchain performance optimization, and intelligent adaptive decision-making mechanisms for next-generation smart building systems.

**Acknowledgement:** Not applicable.

**Funding Statement:** The authors received no specific funding for this study.

**Author Contributions:** The authors confirm contribution to the paper as follows: Conceptualization, Yuh-Shihng Chang and Hsuan-Chao Huang; methodology, Yuh-Shihng Chang and Hsuan-Chao Huang; software, Hsuan-Chao Huang; validation, Yuh-Shihng Chang; formal analysis, Hsuan-Chao Huang; investigation, Hsuan-Chao Huang; resources, Hsuan-Chao Huang; data curation, Yuh-Shihng Chang; writing—original draft preparation, Hsuan-Chao Huang; writing—review and editing, Yuh-Shihng Chang; visualization, Hsuan-Chao Huang; supervision, Yuh-Shihng Chang; project administration, Hsuan-Chao Huang. All authors reviewed and approved the final version of the manuscript.

**Availability of Data and Materials:** The datasets generated during this study are available from the corresponding author upon reasonable request.

**Ethics Approval:** Not applicable.

**Conflicts of Interest:** The authors declare no conflicts of interest.



## References

1. Atzori L, Iera A, Morabito G. The Internet of Things: a survey. *Comput Netw.* 2010;54(15):2787–805. doi:10.1016/j.comnet.2010.05.010.
2. Zanella A, Bui N, Castellani A, Vangelista L, Zorzi M. Internet of Things for smart cities. *IEEE Internet Things J.* 2014;1(1):22–32. doi:10.1109/JIOT.2014.2306328.
3. Minoli D, Sohraby K, Occhiogrosso B. IoT considerations, requirements, and architectures for smart buildings—Energy optimization and next-generation building management systems. *IEEE Internet Things J.* 2017;4(1):269–83. doi:10.1109/JIOT.2017.2647881.
4. Fuller A, Fan Z, Day C, Barlow C. Digital twin: enabling technologies, challenges and open research. *IEEE Access.* 2020;8:108952–71. doi:10.1109/ACCESS.2020.2998358.
5. Ahmad MW, Mourshed M, Rezgui Y. Trees vs Neurons: comparison between random forest and ANN for high-resolution prediction of building energy consumption. *Energy Build.* 2017;147(1):77–89. doi:10.1016/j.enbuild.2017.04.038.
6. Breiman L. Random forests. *Mach Learn.* 2001;45(1):5–32. doi:10.1023/A:1010933404324.
7. Boje C, Guerriero A, Kubicki S, Rezgui Y. Towards a semantic construction digital twin: directions for future research. *Autom Constr.* 2020;114:103179. doi:10.1016/j.autcon.2020.103179.
8. Lu Q, Parlidak AK, Woodall P, Don Ranasinghe G, Xie X, Liang Z, et al. Developing a digital twin at building and city levels: case study of west Cambridge campus. *J Manage Eng.* 2020;36(3):05020004. doi:10.1061/(asce)me.1943-5479.0000763.
9. Sacks R, Brilakis I, Pikas E, Xie HS, Girolami M. Construction with digital twin information systems. *Data Centric Eng.* 2020;1:e14. doi:10.1017/dce.2020.16.
10. Eneyew DD, Capretz MA, Bitsuamlak GT. Toward smart-building digital twins: BIM and IoT data integration. *IEEE Access.* 2022;10(4):130487–506. doi:10.1109/ACCESS.2022.3229370.
11. Dong B, O'Neill Z, Li Z. A BIM-enabled information infrastructure for building energy fault detection and diagnostics. *Autom Constr.* 2014;44(2005):197–211. doi:10.1016/j.autcon.2014.04.007.
12. Das M, Tao X, Cheng JC. BIM security: a critical review and recommendations using encryption strategy and blockchain. *Autom Constr.* 2021;126(2):103682. doi:10.1016/j.autcon.2021.103682.
13. Alshammari K, Beach T, Rezgui Y. Cybersecurity for digital twins in the built environment: current research and future directions. *J Inf Technol Constr.* 2021;26:159–73. doi:10.36680/j.itcon.2021.010.
14. Zhang YY, Kang K, Lin JR, Zhang JP, Zhang Y. Building information modeling-based cyber-physical platform for building performance monitoring. *Int J Distrib Sens Netw.* 2020;16(2):1550147720908170. doi:10.1177/1550147720908170.
15. Iqbal F, Mirzabeigi S. Digital twin-enabled building information modeling-Internet of Things (BIM-IoT) framework for optimizing indoor thermal comfort using machine learning. *Buildings.* 2025;15(10):1584. doi:10.3390/buildings15101584.
16. Padma A, Ramaiah M. Blockchain based an efficient and secure privacy preserved framework for smart cities. *IEEE Access.* 2024;12:21985–2002. doi:10.1109/ACCESS.2024.3364078.
17. Biau G, Scornet E. A random forest guided tour. *TEST.* 2016;25(2):197–227. doi:10.1007/s11749-016-0481-7.
18. Liaw A, Wiener M. Classification and regression by randomForest. *R News.* 2002;2(3):18–22.
19. Dai HN, Zheng Z, Zhang Y. Blockchain for Internet of Things: a survey. *IEEE Internet Things J.* 2019;6(5):8076–94. doi:10.1109/JIOT.2019.2920987.
20. Albulayhi AS, Alsukayti IS. A blockchain-centric IoT architecture for effective smart contract-based management of IoT data communications. *Electronics.* 2023;12(12):2564. doi:10.3390/electronics12122564.
21. Khan S, Mazhar T, Shahzad T, Tariq MU, Ali T, Ayaz M, et al. The evolution of blockchain in smart buildings with IoT integration and future prospects. *Telemat Inform Rep.* 2026;21(2):100283. doi:10.1016/j.teler.2025.100283.
22. Saeed M, Amin R, Aftab M, Ahmed N. Trust management technique using blockchain in smart building. *Eng Proc.* 2022;20(1):24. doi:10.3390/engproc2022020024.
23. Dave B, Kubler S, Främling K, Koskela L. Opportunities for enhanced lean construction management using Internet of Things standards. *Autom Constr.* 2016;61(5):86–97. doi:10.1016/j.autcon.2015.10.009.

24. Kang TW, Mo Y. A comprehensive digital twin framework for building environment monitoring with emphasis on real-time data connectivity and predictability. *Dev Built Environ.* 2024;17:100309. doi:10.1016/j.dibe.2023.100309.
25. Sadeghi J, Ahmadi A, Phipps R. Internet of Things in construction: trends and adoption insights from a scientometric perspective. *Int J Constr Manag.* 2025;26(1):113–30. doi:10.1080/15623599.2025.2508905.
26. Elshabshiri A, Ghanim A, Hussien A, Maksoud A, Mushtaha E. Integration of building information modeling and digital twins in the operation and maintenance of a building lifecycle: a bibliometric analysis review. *J Build Eng.* 2025;99(1):111541. doi:10.1016/j.jobe.2024.111541.
27. Baghdadi A. A comprehensive review of digital twin implementation in construction: current trends and future directions. *J Asian Arch Build Eng.* 2025;25(4):3622–36. doi:10.1080/13467581.2025.2517242.