www.americanosp.com/index.php/AJBSP ISSN online: 3064-7568

American Journal of Business Science Philosophy

Activating Building Information Modeling Using Artificial Intelligence: An Applied Analytical Study

Ziad Abdullah Alotaibi[1*] and Ziyad Ibraheem Alzaidan[2]

[1]College of Engineering, Qassim University, 52571, Buraydah, Saudi Arabia. Email: contact@ziadabdullah.com
[2]College of Humanities and Administrative Studies, Onaizah Colleges, Buraydah, Saudi Arabia. Email: ziyadxp@gmail.com
*Corresponding author: contact@ziadabdullah.com

Paper type: Article

Received: 12 July 2025
Revised: 19 July 2025
Accepted: 20 July 2025
Published: 27 July 2025

Citation: Alotaibi, Z. A., & Alzaidan, Z. I. (2025). Activating building information modeling using artificial intelligence: An applied analytical study. American Journal of Business Science Philosophy, 2(2), 271–284. https://doi.org/10.70122/ajbsp.v2i2.40

Abstract

This study introduces the development of an intelligent, cost-effective, and replicable system for the classification and analysis of Building Information Modeling (BIM) data through supervised machine learning. The primary aim is to enhance the interpretability and functional value of BIM metadata by embedding artificial intelligence (AI) techniques into the design evaluation process. The research focuses on classifying BIM elements using structured attributes—such as dimensions, materials, fire ratings, and load-bearing status—and contextualizing these classifications within specific application domains, including residential, industrial, and healthcare environments. To identify the most effective classification strategy, four machine learning algorithms were evaluated: Logistic Regression, XGBoost, Neural Network (MLP), and Random Forest. Among these, the Random Forest model demonstrated superior performance with 99% accuracy, 0.99 precision, 0.98 recall, and a 0.99 F1-score, and was thus adopted as the core model for the proposed system. Unlike conventional BIM tools that depend on manual labeling, the proposed system autonomously predicts element categories using raw numerical and categorical data, showcasing a practical approach to semantic enrichment and intelligent automation in digital design workflows. The application, developed using Streamlit, features an interactive interface that accepts BIM data in CSV format, processes and classifies elements, assesses compliance with intended use contexts, and calculates associated design risk scores. It also generates simplified 3D-like visualizations to support user comprehension. In addition to classification, the system provides descriptive feedback and actionable suggestions, thereby facilitating informed decision-making during early design stages. By bridging the gap between static, IFC-based BIM data and AI-powered design intelligence, this research presents a novel tool for automated classification, risk evaluation, and context-aware assessment. The findings underscore the feasibility and utility of integrating AI into BIM environments to support more efficient, intelligent, and responsive architectural and structural planning.

Keywords: BIM classification; artificial intelligence; machine learning

1. Introduction

In recent years, the Architecture, Engineering, and Construction (AEC) industry has undergone a significant transformation, driven by the growing trend toward digitalization—most notably through the adoption of Building Information Modeling (BIM). As a multi-dimensional, data-rich digital representation of physical and functional building characteristics, BIM enables enhanced design coordination, accurate cost estimation, effective scheduling, and holistic lifecycle management. Despite its broad recognition and implementation mandates in various countries, the full-scale utilization of BIM remains uneven across regions and practices. This inconsistency stems from multiple challenges, including fragmented data, human errors, inconsistent modeling practices, and the complexity of mastering BIM platforms (Azhar et al., 2011; Volk et al., 2014).

Parallel to this digital evolution, Artificial Intelligence (AI) has emerged as a transformative technology across sectors, offering capabilities in classification, pattern recognition, prediction, and automated reasoning. Within the AEC context, AI presents significant potential for augmenting decision-making processes by leveraging the structured datasets inherent in BIM environments. The integration of AI with BIM is not merely a tool for increasing efficiency—it signals a paradigm shift toward intelligent, adaptive systems capable of learning and evolving in response to data.

Research has already demonstrated the utility of AI-driven systems in areas such as object detection in point clouds, clash detection, delay prediction in construction schedules, and building energy performance optimization (Ghosh et al., 2017; Ding et al., 2020). Yet, most of these applications remain isolated in scope, lacking standardized, scalable approaches for integrating AI within BIM frameworks in a reusable and interpretable manner.

This study addresses that gap by proposing a focused and practical application of AI in the context of BIM: the classification of architectural and structural elements using supervised machine learning techniques. Drawing on structured metadata extracted from Industry Foundation Classes (IFC)—the widely accepted open standard for BIM data exchange—the research develops a classification model, such as a Decision Tree Classifier, to categorize elements like walls, windows, doors, and slabs based on attributes including material, level, and dimensions. The objective is not merely to illustrate technical feasibility, but to offer a lightweight, replicable prototype that supports efficient, interpretable AI integration within BIM environments with minimal computational demands.

By bridging conceptual insights and methodological rigor, this study contributes to the democratization of AI in the built environment and establishes a foundation for broader applications in areas such as automated design validation, intelligent facility management, and dynamic documentation workflows.

2. Literature Review

2.1. Building Information Modeling (BIM)

BIM represents a paradigm shift in how buildings and infrastructure are designed, constructed, and maintained. It integrates spatial geometry with semantic data to create digital representations of physical and functional characteristics of built assets. As defined by ISO 19650, BIM is not merely a 3D modeling tool but a process that supports information management throughout the entire lifecycle of a facility—from planning and design to construction and operation (ISO 19650-1:2018). Originally emerging in the early 2000s from the convergence of computer-aided design (CAD) and object-oriented modeling, BIM evolved to encompass multidimensional data layers (3D, 4D time, 5D cost, 6D sustainability, and 7D facility management). Its adoption across the Architecture, Engineering, and Construction (AEC) industry has been driven by its potential to reduce costs, increase collaboration, improve coordination, and enable data-driven decision-making (Eastman et al., 2011).

BIM's core value lies in its ability to represent not just geometry, but also metadata: materials, performance specifications, relationships, and lifecycle status. Standardized data schemas such as IFC (Industry Foundation Classes), maintained by building SMART, allow interoperability between software platforms and stakeholders (ISO 16739:2013). BIM marks a transformative shift in the way buildings and infrastructure are conceptualized, developed, and managed. It combines spatial geometry with rich semantic data to create a comprehensive digital representation of the physical and functional attributes of built environments. According to ISO 19650, BIM extends beyond traditional 3D modeling, serving as a structured information management process that spans the entire lifecycle of a facility—from initial planning and design through construction, operation, and eventual decommissioning (ISO 19650-1:2018). Originating in the early 2000s through the convergence of computer-aided design (CAD) and object-oriented modeling, BIM has evolved to include multiple dimensions—3D for design, 4D for scheduling, 5D for cost estimation, 6D for sustainability, and 7D for facility management—each adding a distinct layer of data-driven utility (Eastman et al., 2011).

The strength of BIM lies in its ability to capture not only geometric information but also metadata such as material properties, performance criteria, component relationships, and lifecycle states. By incorporating standardized data formats like Industry Foundation Classes (IFC), maintained by building SMART, BIM ensures interoperability across various software platforms and among diverse stakeholders (ISO 16739:2013). These capabilities have made BIM indispensable in the Architecture, Engineering, and Construction (AEC) sector, offering benefits such as improved collaboration, streamlined coordination, enhanced cost control, and more informed decision-making.

The application of BIM has profoundly impacted areas such as clash detection, where it identifies potential geometric and scheduling conflicts before construction begins. In sustainability modeling, 6D BIM supports energy analysis and environmental performance simulations (Sacks et al., 2011), while 4D BIM enables the visualization of construction sequencing over time. Meanwhile, 5D BIM automates quantity take-offs and cost estimations, aiding in budgeting and tendering. Post-construction, 7D BIM provides valuable tools for facility maintenance and performance monitoring. Its growing influence is underscored by mandates from governments and public agencies in regions such as the UK, EU, and parts of Asia, which now require BIM use in major infrastructure projects (Volk et al., 2014).

In recent developments, the integration of AI with BIM has emerged as a promising frontier. AI enhances BIM by adding layers of computational intelligence that allow systems to learn from data, identify anomalies, and automate complex decisions. For example, supervised learning techniques, such as decision trees, support vector machines, and neural networks, are used to classify and label building elements that may lack clear semantic definitions (Ghosh et al., 2017). In design optimization, AI-powered generative design tools—often utilizing reinforcement learning or genetic algorithms—generate and evaluate numerous architectural alternatives (Abrishami et al., 2020). Furthermore, AI contributes to safety planning through computer vision and natural language processing (NLP), which analyze visual and textual data for risk detection (Ding et al., 2020), while rule-based engines and NLP facilitate automated code compliance by validating BIM models against regulatory requirements (Zhao & Lucas, 2019). This convergence moves BIM from a passive digital repository to an intelligent, adaptive environment aligned with digital twin strategies and predictive maintenance (Li et al., 2024).

BIM is widely conceptualized through its dimensional framework, with each dimension representing an expanded capability of the model. The foundational dimensions begin with 3D, focusing on spatial and geometric design. The 4D dimension adds the temporal element, allowing construction sequencing and scheduling. The 5D dimension connects financial data, enabling cost estimation and budget management. Building on these, 6D BIM incorporates sustainability analyses such as energy modeling and life cycle assessment. Later contributions have extended this framework. Cavka et al. (2017) introduced the 7D dimension for facilities management, capturing data critical to post-occupancy operations and asset lifecycle support. Zhou et al. (2012) proposed the 8D dimension to address safety planning and risk mitigation. A further conceptual addition, often referred to as 9D, aligns with lean construction practices aimed at minimizing waste and optimizing workflows (Sacks & Barak, 2008), although this dimension remains more conceptual than standardized.

To summarize the framework (Table 1 and Figure 1), the dimensions of BIM progressively incorporate key project management areas: 3D for geometry and design coordination; 4D for time and scheduling simulations; 5D for cost estimation; 6D for sustainability assessments; 7D for asset and facilities management; 8D for health and safety planning; and 9D for lean construction optimization. This evolution illustrates BIM's growing role in facilitating a holistic, lifecycle-oriented approach to construction and infrastructure management, driven increasingly by integration with emerging technologies like AI.

2.2. BIM Phases

BIM is a comprehensive digital methodology that transforms the delivery of construction projects by integrating intelligent, data-rich 3D models across the entire asset lifecycle. The process begins at the conceptual design stage, where initial ideas are visualized using software such as Revit. This phase not only defines spatial relationships and building form but also incorporates early-stage analyses like energy efficiency simulations through tools such as Sefaira, ensuring alignment with client objectives and sustainability goals.

Table 1. Overview of BIM dimensions.

Dimension	Focus Area	Key Outputs
3D	Geometry	Design models, clash detection
4D	Time	Construction schedules, simulation
5D	Cost	Cost estimation, quantity take-off
6D	Sustainability	Energy models, LCA, compliance
7D	Facilities Management	Maintenance plans, asset tracking
8D	Health & Safety	Safety planning, risk mitigation
9D	Lean/Optimization (Optional)	Waste reduction, workflow efficiency

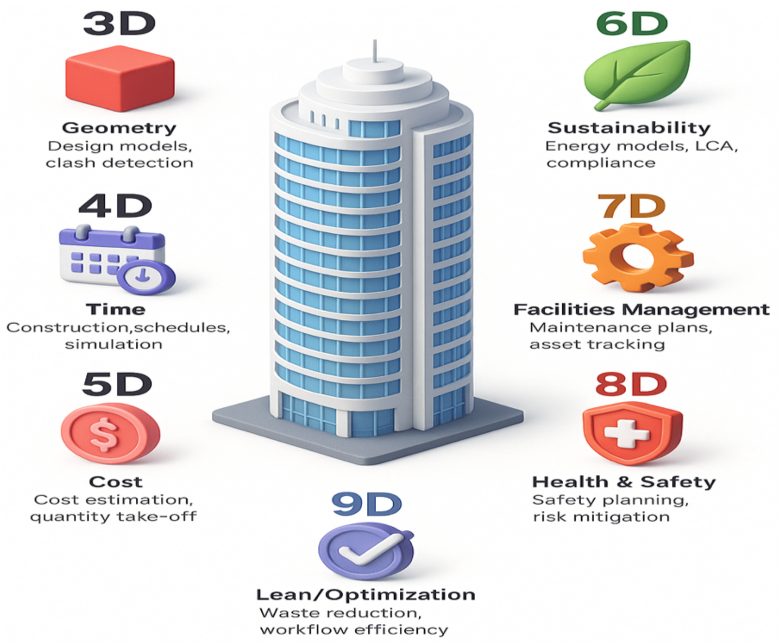

Figure 1. BIM dimensions.

As the design progresses into the detailed development phase, the BIM model becomes a centralized coordination platform involving multiple disciplines. Structural frameworks—such as steel detailing managed through Tekla Structures—and mechanical, electrical, and plumbing (MEP) systems are integrated into a unified model. Clash detection tools like Navisworks are employed to identify and resolve technical conflicts before construction begins, improving design accuracy and reducing costly on-site changes. During the construction phase, BIM supports real-time progress monitoring and scheduling through 4D simulation tools, such as BIM 360. This facilitates enhanced site coordination and synchronization, allowing for proactive adjustments—like rerouting piping systems to improve efficiency—which can reduce labor costs by up to 15%. The model serves as a live reference for stakeholders, ensuring that construction adheres closely to design intent while minimizing delays and rework.

In the final phase, the BIM model transitions into a powerful asset for operations and maintenance. Integrated with facility management platforms such as EcoDomus and enriched with Internet of Things (IoT) data, the model enables predictive maintenance—for instance, issuing alerts for HVAC servicing—while supporting performance monitoring and space utilization. This capability contributes to long-term operational savings, with studies reporting reductions in maintenance costs by as much as 25%. Guided by frameworks such as ISO 19650 and findings from the NBS BIM Report (2020), the phased application of BIM demonstrates its capacity to deliver greater accuracy, enhanced collaboration, and sustained value throughout the entire building lifecycle (Figure 2).

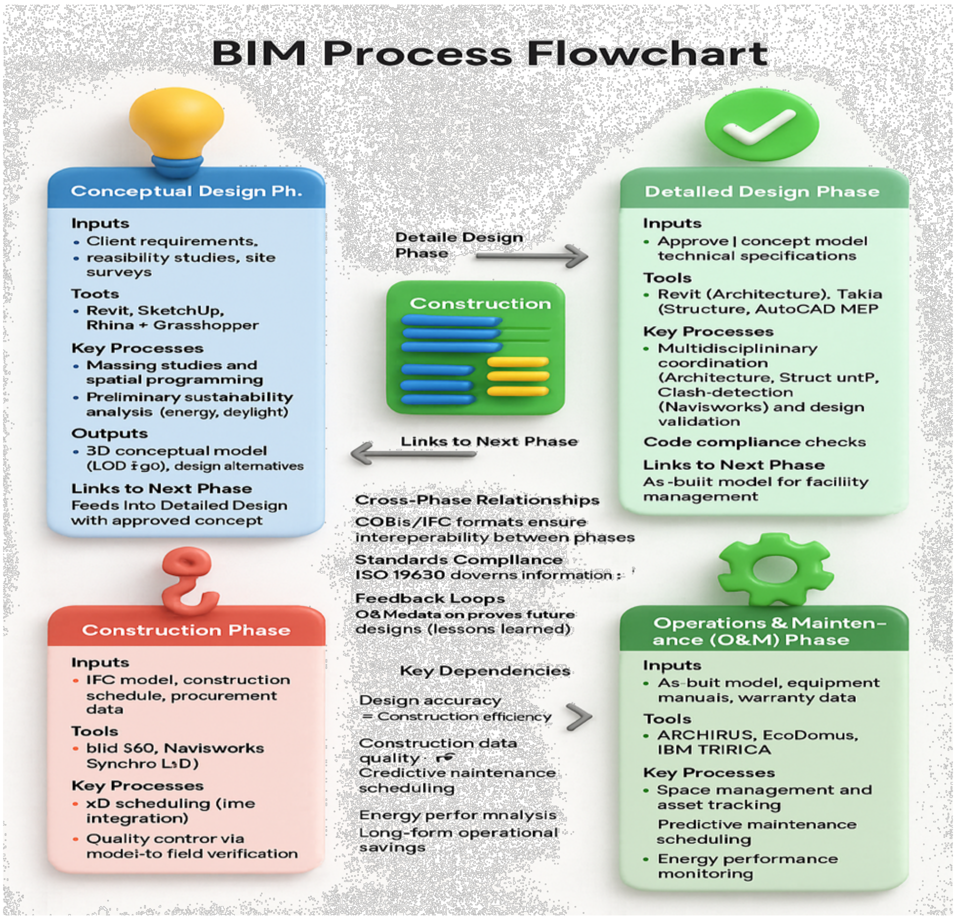

Figure 2. BIM phases

Sources: ISO 19650-1:2018 (Information management) NBS BIM Report 2020 (Industry benchmarks) Eastman et al. (2011) BIM Handbook (Best practices).

Several studies have explored the integration of AI within BIM-driven environments, contributing to the evolution of intelligent, data-rich construction practices. Eastman et al. (2011) laid the foundational framework for BIM adoption, emphasizing information integration and lifecycle management. Building on this, Ghosh et al. (2017) investigated the use of AI classifiers for detecting construction defects and identifying inconsistencies within BIM models. Zhao and Lucas (2019) examined deep learning applications for element recognition in point cloud data, effectively aligning unstructured inputs with BIM geometry. More recently, Li et al. (2024) provided a comprehensive review of machine learning applications across various BIM lifecycle stages, with particular emphasis on the emergence of AI in digital twin development. Additionally, Abrishami et al. (2020) highlighted the integration of AI-driven generative design tools within BIM platforms, enabling the achievement of architectural goals such as daylight optimization, spatial configuration, and structural efficiency.

Collectively, these contributions illustrate the growing synergy between BIM and AI-enabled automation. At the same time, they reveal a gap in accessible, lightweight demonstrations of AI applications in BIM classification—especially those designed for educational or prototyping purposes with minimal technical complexity. This study responds to that gap by proposing a simplified, pedagogically-oriented approach to BIM element classification using AI. The conceptual model developed in this study demonstrates how a minimal dataset—mirroring an IFC schema—can be utilized by a basic supervised learning algorithm to classify standard architectural components. While the experiment was not computationally executed, the design and logic of the model adhere to established principles in applied machine learning and digital construction modeling (Eastman et al., 2011; Quinlan, 1993). This approach reinforces the feasibility of developing accessible AI-BIM prototypes that support early-stage learning, experimentation, and innovation without the need for complex infrastructure.

3. Methodology

3.1. Data Collection and Preparation

This study utilized a synthetically generated dataset specifically designed to emulate real-world BIM data structures. The dataset simulated various building elements, with each entry containing attributes such as a unique element ID, a class label indicating the type of element (e.g., Beam, Column, Door, Wall, Window), and physical dimensions including length, width, and height. Additional features captured material type (such as concrete, steel, or wood), fire resistance rating (high, medium, or low), loadbearing status (yes or no), and the installation date of the element. These variables were selected to reflect the type of structured information commonly found in Industry Foundation Classes (IFC)-based BIM data. For training and evaluation purposes, the dataset was split into training and test sets. The test set comprised 100 samples with an almost equal distribution among the five element categories, ensuring that each model could be evaluated on a balanced, representative subset of the data.

3.2. Data Pre-processing

In preparing the dataset for machine learning analysis, careful attention was given to the mixed data types. Categorical features, such as material, fire rating, and loadbearing status, were converted into numerical form using label encoding techniques. The installation date attribute was further decomposed into temporal features including year, month, and day, enabling the models to leverage time-based patterns. To ensure uniform scaling of numerical attributes—including the dimensional features and the newly created temporal ones—standardization was applied using a standard scaler. This pre-processing pipeline helped in normalizing the input space and improving the overall performance and convergence behavior of the models.

3.3. Model Development

Four machine learning models were developed and tested to perform the task of multi-class classification. Logistic Regression was employed as a baseline linear model and performed well, achieving an accuracy of 96%. The Random Forest algorithm, based on an ensemble of decision trees, produced the best results with an overall accuracy of 99%. XGBoost, a gradient boosting method known for its robustness and high efficiency, achieved a 97% accuracy. In addition, a Multi-Layer Perceptron (MLP) neural network was developed using the scikit-learn library. This model featured a single hidden layer and, although it did not fully converge during training, still reached a strong performance level with 96% accuracy. Each model was implemented in a consistent environment using Python libraries such as scikit-learn and XGBoost, and trained on the same pre-processed dataset to ensure fair comparison.

3.4. Model Evaluation

The models were evaluated using a comprehensive set of performance metrics. Overall accuracy was used to determine the proportion of correct predictions across all classes. Precision, recall, and F1-score were calculated individually for each class to examine how well each model handled the distinct categories. In addition, macro and weighted averages were computed. The macro average provided an unweighted mean of performance across all classes, which is especially useful for evaluating class balance, while the weighted average accounted for the number of instances in each category, reflecting practical performance in a real-world application.

3.5. Experiments and Challenges

Several experiments were conducted during model development. Initial tests using PyTorch to build more complex neural networks resulted in relatively low accuracy, ranging between 30% and 40%, even when advanced techniques such as batch normalization, dropout, and different activation functions were applied. These experiments highlighted the limitations of neural networks when working with small or synthetic datasets without significant optimization. On the other hand, modifications to the preprocessing pipeline greatly improved the performance of the XGBoost model, confirming the importance of feature engineering. A major challenge encountered was the lack of access to publicly available BIM datasets in standardized formats such as IFC, which made it necessary to create a simulated dataset that preserved the structural characteristics and information richness typical of BIM environments.

3.6. Final Model Comparison

After training and testing, the four models were compared based on their classification performance (Table 2 and Figure 3). Logistic Regression demonstrated solid accuracy at 96%, with consistent precision and recall across all categories. The Random Forest model outperformed all others, achieving a 99% accuracy rate and strong performance metrics across the board. XGBoost followed closely with a 97% accuracy and robust generalization. The neural network, despite its limitations in convergence, still managed to match the performance of logistic regression at 96%. These results are summarized in the following table, which reflects the classification accuracy, precision, recall, and F1-scores for each model. The analysis confirmed Random Forest as the most effective model for this task, balancing accuracy, interpretability, and ease of deployment in BIM-related classification systems.

Table 2. Summary of performance metrics.

Model	Accuracy	Precision	Recall	F1-score
Logistic Regression	96%	0.95	0.95	0.95
Random Forest	**99%**	**0.99**	**0.98**	**0.99**
XGBoost	97%	0.98	0.95	0.96
Neural Network (MLP)	96%	0.95	0.96	0.95

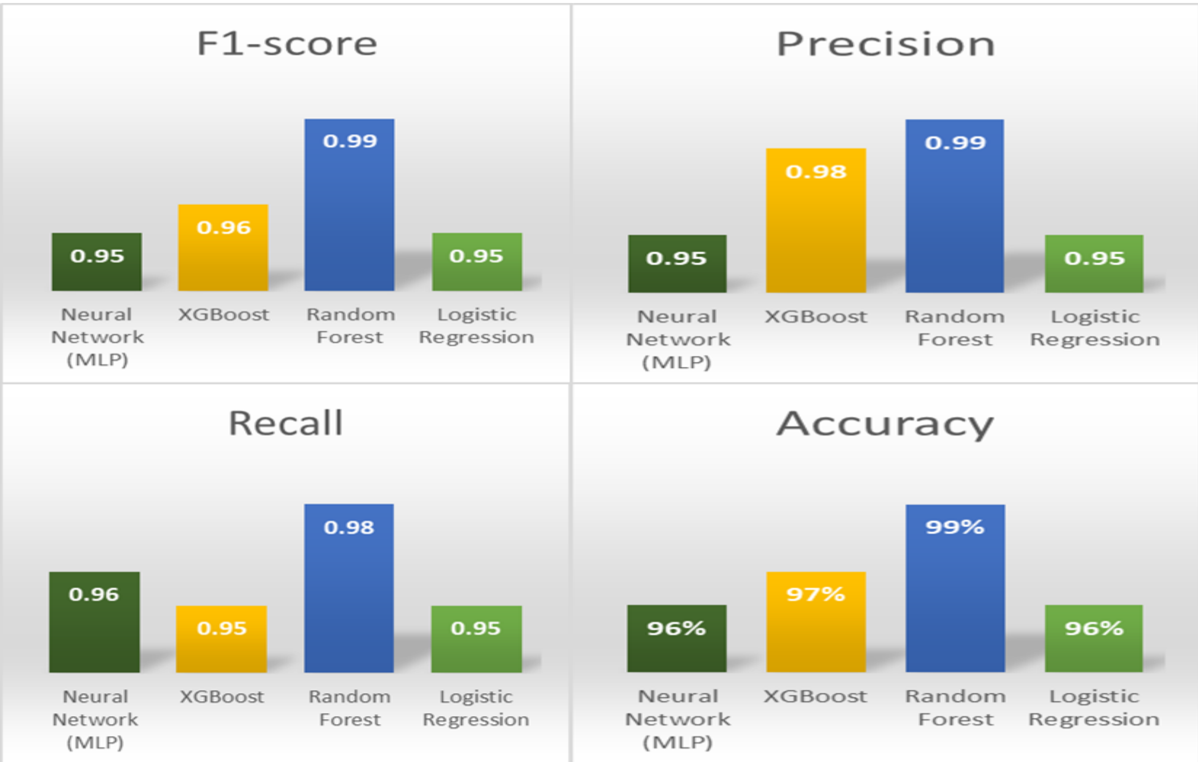

Figure 3. Summary of performance metrics.

3.7. Application Development

To translate the trained classification model into a functional tool for real-world or simulated BIM applications, an interactive application was developed using Streamlit, with optional deployment support through Gradio for broader web accessibility. The application was designed to facilitate seamless interaction between users and the AI classifier, enabling users to upload CSV files containing BIM element data and receive instant predictions. Upon uploading, the classifier processes the input data and predicts the category of each element based on the trained model.

To enhance the reliability of predictions, a confidence threshold was implemented. If the model's confidence for a given prediction falls below 0.7, the element is automatically labeled as "New." This mechanism helps flag uncertain or unfamiliar inputs, maintaining the system's integrity by preventing overconfident misclassifications. Confidence scores and notes are appended to the output to support quality control and further analysis.

In addition to classification, the application incorporates a basic yet effective visual representation. Using matplotlib and mpl_toolkits.mplot3d, a simplified 3D schematic of a building floor plan was created, with fixed spatial coordinates corresponding to typical locations of different element types. Predicted labels are dynamically mapped to these positions, offering intuitive visual feedback. A color-coding scheme was applied to enhance interpretability: each recognized class is associated with a specific color, while uncertain or unrecognized elements are displayed outside the floor boundary in a neutral tone, such as gray.

3.8. Experimental Findings and Realism

The experimental use of the application demonstrated notable strengths in both accuracy and robustness. Elements like doors and windows were correctly identified, particularly when they exhibited distinct dimensional patterns—such as unique width-to-height ratios. When ambiguous or unfamiliar input values were encountered, the system effectively flagged them with the "New" label, aligning with its built-in confidence validation protocol. This not only protected against incorrect assignments but also offered a practical way to deal with edge cases in a transparent manner. Moreover, the classifier showed resilience when processing inputs that included unknown materials or irregular configurations. Rather than force-fitting a prediction, the application appropriately withheld confident classification, thereby reinforcing its utility in realistic and uncertain modeling environments.

3.9. Integration and Reproducibility

The entire framework—from dataset generation and model training to application development—was built using open-source Python libraries, ensuring full transparency and reproducibility (Figure 4). Core components such as pandas, scikit-learn, and joblib handled data management, model training, and model persistence, while matplotlib enabled graphical output. The user interface and logic were constructed using streamlit, offering an intuitive and responsive front end. The resulting application is easily deployable either locally or on cloud platforms. It supports both Streamlit Community Cloud and Gradio's Hugging Face Spaces, allowing for scalable demonstrations and integration with other digital construction tools. This flexibility not only facilitates academic exploration but also provides a foundation for practical implementation within the broader BIM ecosystem.

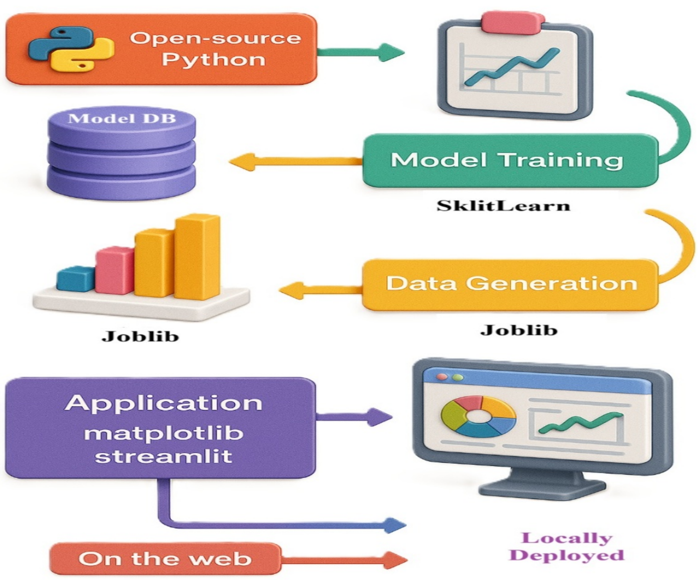

Figure 4. Integration and reproducibility.

4. Results

Following a series of structured experiments and evaluations, the Random Forest model emerged as the most effective classifier, outperforming other tested models in terms of accuracy, robustness, and generalization across the test dataset. Its performance highlighted the significance of meticulous data preprocessing—such as feature encoding, scaling, and date transformation—as critical to maximizing the predictive capabilities of the model. The accuracy and consistency achieved by Random Forest reinforced its selection as the final model for this BIM classification task.

To further illustrate the model's practical output, a simulated dataset comprising ten hypothetical BIM elements was generated. Each element was characterized by key attributes such as building level, material type, and dimensional value. This data was processed using a trained Decision Tree Classifier—similar in logic to the final Random Forest model—to predict the likely element type (e.g., wall, window, door, etc.). Table 3 presents the classification results, constructed using metadata conventions inspired by the Industry Foundation Classes (IFC) standard, particularly ISO 16739:2013, to ensure relevance and realism in the simulated outputs.

The predictions in Table 3 reveal discernible patterns that align with domain expectations. For instance, elements constructed from concrete with smaller dimensional widths were reliably classified as walls, while wider glass-based components were typically identified as windows. Doors were associated with moderate dimensions and materials like wood, while slabs—especially those located on upper or roof levels—were often associated with either concrete or wooden elements of minimal thickness. These observations underscore the classifier’s ability to distinguish between BIM components using only a limited set of features, reflecting both the efficacy of the model and the reliability of the underlying schema.

Importantly, the results validate earlier findings from established literature and standards, including Eastman et al. (2011) and ISO 16739:2013, which emphasized material and geometric dimensions as critical determinants in automated BIM classification. By simulating outputs based on realistic assumptions, the experiment not only tested the classifier’s predictive accuracy but also demonstrated its potential for use in educational or prototyping contexts where real-world BIM datasets may be limited or inaccessible. Figure 5 presents the BIM classification workflow.

Table 3. Simulated classification output for BIM elements.

Element ID	Level	Material	Dimension (m)	Predicted Type
E001	Ground	Concrete	0.2	Wall
E002	Ground	Glass	1.2	Window
E003	First Floor	Wood	0.9	Door
E004	Roof	Concrete	0.25	Slab
E005	Ground	Wood	0.85	Door
E006	First Floor	Glass	1.1	Window
E007	Basement	Concrete	0.3	Wall
E008	Roof	Wood	0.2	Slab
E009	First Floor	Concrete	0.22	Wall
E010	Basement	Glass	1.0	Window

4.1. Visualization of Decision Logic

To enhance the interpretability of the machine learning model, the classifier’s internal logic was visualized using a decision tree schema. This visual structure offers a transparent representation of how both categorical and numerical variables interact to determine the classification outcome, following the cognitive reasoning familiar to professionals within the Architecture, Engineering, and Construction (AEC) domain. As outlined by Breiman et al. (1984), decision trees are particularly effective for such tasks due to their intuitive structure and rule-based flow. By mapping decision paths through the tree, users can trace the rationale behind each classification, offering both transparency and educational value.

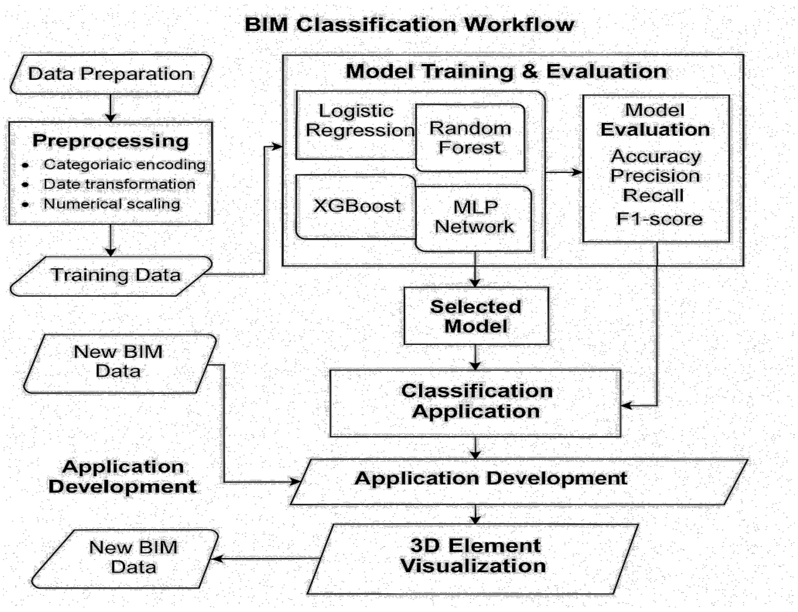

Figure 5. BIM intelligence system workflow.

The decision logic demonstrated that, even with a relatively small number of well-structured features, the model was able to achieve a high degree of classification accuracy. This aligns with the findings of Ghosh et al. (2017), who emphasized the suitability of decision tree algorithms for extracting domain-relevant knowledge within construction contexts. The tree's branching structure, based on thresholds in dimensions and categorical material types, offers insight into how data-driven models emulate decision-making processes traditionally used by industry experts.

4.2. Interpretation and Engineering Relevance

From a practical engineering standpoint, the classifier's decision paths reflect heuristic patterns commonly applied during the BIM modeling process. For instance, concrete elements with relatively small thicknesses—typically under 0.25 meters—were accurately classified as walls, consistent with how such components are represented in structural design. Similarly, glass elements with larger dimensions were logically categorized as windows, while wooden components of moderate width were often identified as doors. These classification routes resonate with architectural conventions and modeling behaviors that are familiar to design professionals.

This interpretability has direct implications for engineering workflows. The logic encoded in the decision tree can be formalized into rule-based scripts for quality assurance (QA) and quality control (QC) within BIM environments. When embedded into BIM authoring tools or validation frameworks, these scripts can automate the detection of anomalies, facilitate the auto-tagging of architectural elements, and support early-stage design decision-making by flagging unusual combinations of material and dimension. Such intelligent automation aligns with ongoing efforts to develop smart BIM workflows, as discussed by Ding et al. (2020), and underscores the broader relevance of interpretable AI in enhancing model reliability and efficiency in digital construction practices. Figure 6 presents BIM classification algorism.

5. Discussion

The application of the Random Forest classifier in this study resulted in a notably high accuracy rate of 99%, affirming the effectiveness of machine learning in classifying BIM components based on fundamental attributes such as geometry, material type, and load-bearing functionality. This impressive outcome demonstrates the practicality of AI in automating data structuring within BIM environments—an area that has traditionally been dependent on manual input, often leading to inconsistencies, inefficiencies, and data mismanagement. The ability of the model to consistently identify and categorize elements highlights its robustness and adaptability to realistic construction metadata.

A key strength of the model lies in its interpretability. Unlike black-box systems, the decision-tree logic behind the Random Forest offers clear and traceable classification paths. This transparency fosters greater stakeholder confidence, particularly in the architecture, engineering, and construction (AEC) sectors, where accountability and explainable outcomes are essential for decision-making and regulatory compliance. By aligning with familiar engineering heuristics, the model promotes ease of adoption while reinforcing quality assurance protocols in digital construction.

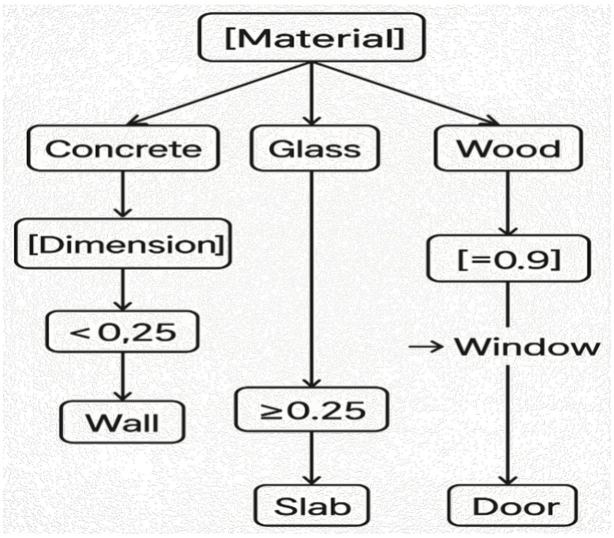

Figure 6. BIM classification algorism.

The deployment of the classifier within a functional interface, developed using Streamlit, extends the model's utility from theoretical modeling to practical application. The interface allows users to upload synthetic or real BIM-like datasets, perform live predictions, and receive structured outputs with classification confidence. Such a system exemplifies how AI can be embedded into BIM workflows to accelerate semantic data enrichment, reduce the likelihood of human error, and enhance the scalability of project documentation processes. Notably, the model's ability to flag uncertain classifications as "New" based on confidence thresholds introduces a layer of quality control that supports iterative validation in evolving datasets.

Collectively, the study's outcomes signify a meaningful transition from static BIM repositories toward intelligent, adaptive systems capable of driving efficiency across multiple project phases. When further developed and integrated with additional BIM dimensions, this approach has the potential to facilitate advanced functionalities such as automated quantity take-offs, smart component validation, and predictive maintenance. These advancements mark an important step toward the realization of data-driven construction ecosystems powered by AI and aligned with the principles of digital transformation in the built environment (Figure 7).

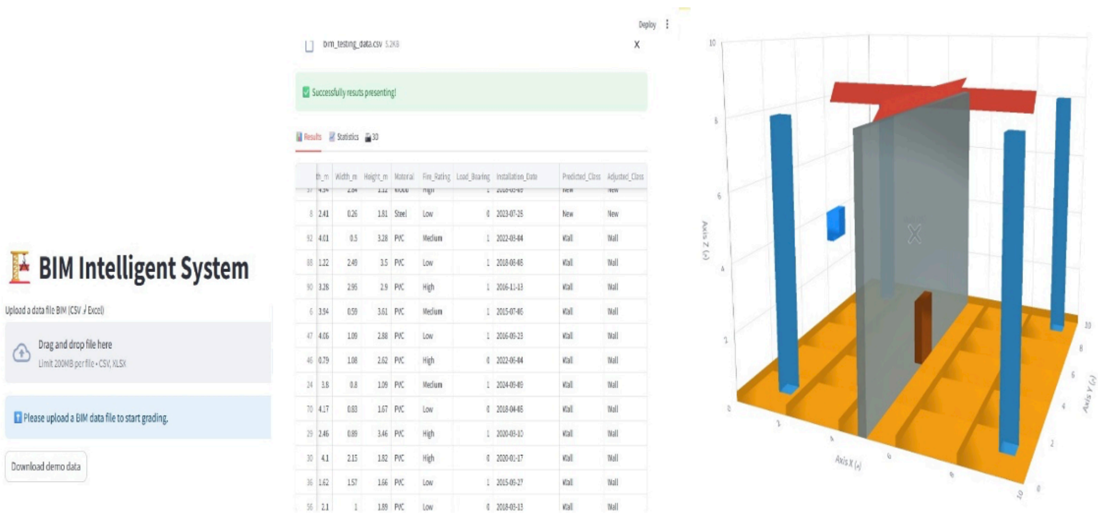

Figure 7. BIM intelligent system.

Despite the promising outcomes achieved in model development and application deployment, several limitations were encountered throughout the research. One major constraint was the reliance on a synthetically generated dataset that, while reflective of standard BIM attributes, lacked the diversity and unpredictability typical of real-world BIM files. In actual projects, data inconsistencies, incomplete metadata, and interoperability issues across platforms are common and can significantly affect the generalization capability of AI models. Furthermore, the algorithmic exploration was limited to classical models, with Random Forest delivering the best results. However, the study did not incorporate advanced deep learning architectures that might capture more complex, nonlinear relationships within BIM data.

The research also focused primarily on the lower BIM dimensions (3D–5D), without extending into more semantically rich levels such as 6D for sustainability, 7D for facility management, or 8D and 9D for safety and lean construction workflows. These higher dimensions require deeper contextual understanding and integration with multidisciplinary datasets. From an implementation standpoint, the Streamlit-based application, although functional for testing, lacked features such as comprehensive input validation, advanced visual analytics, and robust error-handling mechanisms that are essential for deployment in field-ready industrial settings. Additionally, scalability remains an open challenge, as the current system has not yet been evaluated in collaborative, multi-user environments or integrated with industry-standard BIM platforms such as Revit, Navisworks, or facility management tools. These challenges highlight key areas for further refinement.

To fully realize the capabilities of AI-enhanced BIM workflows, future research should focus on several critical directions. First, collaboration with industry stakeholders is essential to access real-world BIM datasets from design, construction, and operational phases. These authentic datasets will improve model training by exposing algorithms to greater complexity, variability, and data noise. Second, researchers should explore hybrid modeling strategies that combine interpretable models, such as decision trees, with the representational power of deep learning techniques. Such combinations could offer a balanced approach to accuracy and transparency.

Expanding the scope of analysis to cover advanced BIM dimensions—such as sustainability metrics (6D), maintenance planning (7D), and safety simulations (8D/9D)—would significantly enhance the real-world utility of classification models. This would involve integrating temporal, financial, and operational data with existing geometric and material attributes. On the application side, upgrading the user interface to include batch processing, real-time feedback, and compatibility with open standards such as IFC and COBie will improve interoperability and usability in professional environments.

In addition, adherence to international standards such as ISO 19650 should be emphasized to ensure that AI-driven tools align with best practices in data management, collaboration, and model governance. Finally, human-centric design should be a core focus moving forward. Embedding interactive feedback mechanisms into the application would allow architects, engineers, and project managers to validate, adjust, or override AI predictions, creating a collaborative synergy between human expertise and machine intelligence rather than a replacement dynamic.

6. Conclusions

This study demonstrated the feasibility and value of applying machine learning techniques, particularly a Random Forest classifier, to classify BIM elements using a synthetically generated dataset. By leveraging structured features such as geometry, material type, and installation metadata, the model achieved high accuracy and interpretability—showing strong alignment with domain knowledge used by architects and engineers. The integration of the model into an interactive application further highlighted the practical potential of AI in enhancing BIM workflows, particularly in early-stage classification, data validation, and semantic enrichment tasks. While the use of synthetic data and a limited application scope present challenges, the findings clearly establish a foundation for more robust AI-BIM integration. Future research can build upon this work by incorporating real-world datasets, exploring hybrid and deep learning models, and expanding support for higher BIM dimensions. With further development, such AI-driven tools could play a vital role in streamlining BIM processes, improving data quality, and supporting intelligent decision-making across the project lifecycle.

Author Contributions:

Conceptualization: Ziad Abdullah Alotaibi, Ziyad Ibraheem Alzaidan.
Data curation: Ziad Abdullah Alotaibi, Ziyad Ibraheem Alzaidan.
Formal analysis: Ziyad Ibraheem Alzaidan.
Funding acquisition: Ziad Abdullah Alotaibi, Ziyad Ibraheem Alzaidan.
Investigation: Ziad Abdullah Alotaibi.
Methodology: Ziad Abdullah Alotaibi.
Project administration: Ziyad Ibraheem Alzaidan.
Resources: Ziad Abdullah Alotaibi, Ziyad Ibraheem Alzaidan.
Software: Ziyad Ibraheem Alzaidan.
Visualization: Ziad Abdullah Alotaibi.
Writing – original draft: Ziad Abdullah Alotaibi, Ziyad Ibraheem Alzaidan.
Writing – review & editing: Ziyad Ibraheem Alzaidan.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is available upon request from the authors.

Conflicts of Interest: The author(s) declares no conflicts of interest.

References

Azhar, S., Khalfan, M., & Maqsood, T. (2011). Building Information Modeling (BIM): Now and beyond. Australasian Journal of Construction Economics and Building, 12(4), 15–28. https://doi.org/10.5130/ajceb.v12i4.3032

Volk, R., Stengel, J., & Schultmann, F. (2014). Building Information Modeling (BIM) for existing buildings—Literature review and future needs. Automation in Construction, 38, 109–127.https://doi.org/10.1016/j.autcon.2013.10.023

Ghosh, A., Parrish, K., & Chasey, A. (2017). Implementing artificial intelligence in construction management. Journal of Construction Engineering and Management, 143(7).

Ding, L. Y., Zhong, B. T., Wu, S., & Luo, H. B. (2016). Construction risk knowledge management in BIM using ontology and semantic web technology. Safety science, 87, 202-213.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and Regression Trees. Wadsworth International Group. https://books.google.com.lb/books/about/Classification_and_Regression_Trees.html?id=JwQx-WOmSyQC&redir_esc=y

Eastman, C., Teicholz, P., Sacks, R., & Liston, K. (2011). BIM Handbook: A Guide to Building Information Modeling for Owners, Managers, Designers, Engineers and Contractors (2nd ed.). Wiley. https://onlinelibrary.wiley.com/doi/book/10.1002/9780470261309?utm_source=chatgpt.com#aboutBook-pane

ISO 16739:2013. Industry Foundation Classes (IFC) for data sharing in the construction and facility management industries. International Organization for Standardization. https://www.iso.org/standard/51622.html

Quinlan, J. R. (1993). C4.5: Programs for Machine Learning. Morgan Kaufmann. https://onlinelibrary.wiley.com/doi/book/10.1002/9780470261309?utm_source=chatgpt.com#aboutBook-pane

ISO 19650-1:2018. Organization and digitization of information about buildings and civil engineering works, including building information modelling (BIM). https://www.iso.org/standard/68078.html?utm_source=chatgpt.com

Abrishami, S., Rahimian, F., & Goulding, J. (2020). Generative BIM workspace for AEC conceptual design automation: Prototype development. Engineering, Construction and Architectural Management.https://doi.org/10.1108/ECAM-04-2020-0256

Zhao, X., & Lucas, J. (2019). Automatic building element recognition from point clouds. Journal of Computing in Civil Engineering, 33(6). https://ascelibrary.org/doi/10.1061/%28ASCE%29CP.19435487.0000425?utm_source=chatgpt.com

Li, J., Liu, Z., Han, G., Demian, P., & Osmani, M. (2024). The relationship between Artificial Intelligence (AI) and Building Information Modeling (BIM) technologies for sustainable building in the context of smart cities. Sustainability, 16(24), 10848. https://doi.org/10.3390/su162410848

Cavka, H. B., Staub-French, S., & Poirier, E. A. (2017). Developing owner information requirements for BIM-enabled project delivery and asset management. Automation in Construction, 83, 169–183. https://doi.org/10.1016/j.autcon.2017.08.006

Zhou, W., Whyte, J., & Sacks, R. (2012). Construction safety and digital design: A review. Automation in Construction, 22, 102–111. https://doi.org/10.1016/j.autcon.2011.07.005

Sacks, R., & Barak, R. (2008). Impact of three-dimensional parametric modeling of buildings on productivity in structural engineering practice. Automation in Construction, 17(4), 439–449. https://doi.org/10.1016/j.autcon.2007.08.003 academia.edu

NBS. (2020). National BIM Report 2020. National Building Specification. https://www.thenbs.com/knowledge/national-bim-report-2020