Contents lists available at SCCE

Journal of Soft Computing in Civil Engineering

Journal homepage: www.jsoftcivil.com

A Review on the Role of AI in BIM: Streamlining Design for Greater Efficiency and Compliance

Kamal Jaafar [1,*]; Karol Sikora [1]; Sana Amir [2]; Lina Gharaibeh [3]; Mohamad Koona [4]

1. Associate Professor, Faculty of Engineering and Information Science, University of Wollongong in Dubai, UAE
2. Assistant Professor, Faculty of Engineering and Information Science, University of Wollongong in Dubai, UAE
3. Assistant Professor, Faculty of Business, University of Wollongong in Dubai, UAE
4. BSc, Civil Engineering, Department of Engineering and Information Science, University of Wollongong in Dubai, UOWD, Dubai, UAE

* Corresponding author: kamaljaafar@uowdubai.ac.ae

https://doi.org/10.22115/SCCE.2025.1961

ARTICLE INFO

Article history:
Received: 08 October 2024
Revised: 21 April 2025
Accepted: 22 September 2025

Keywords:
Artificial intelligence (AI);
Building information modeling (BIM);
Architectural design optimization;
Construction project efficiency.

ABSTRACT

The increasing complexity of construction projects, driven by technological advancements and evolving lifestyle demands, has placed significant pressure on project timelines and resource management. Despite the potential of Artificial Intelligence (AI) technologies to streamline processes, their underutilization or ineffective implementation has led to escalated costs and extended project cycles. The design phase, accounting for 15% to 20% of a project's total lifecycle, is critical to timely project completion. Delays and inefficiencies in this phase can have a cascading impact on subsequent stages. This study critically examines the current application of AI-aided design software within the framework of Building Information Modeling (BIM), identifying areas for improvement and evaluating its potential to optimize design processes. By analyzing the performance of AI-generated designs in terms of accuracy, efficiency, generative capabilities, usability, and compliance, this research compares AI-driven methodologies with traditional design approaches. The findings aim to illuminate AI's capacity to enhance precision, innovation, and speed in design iterations, offering valuable insights into its broader impact on the construction industry's adaptability, cost management, and regulatory compliance.

How to cite this article: Jaafar K, Sikora K, Amir S, Gharaibeh L, Koona M. A review on the role of AI in BIM: streamlining design for greater efficiency and compliance. Journal of Soft Computing in Civil Engineering. 2026;10(3):1961. https://doi.org/10.22115/scce.2025.1961

1. Introduction

Recent advancements in artificial intelligence (AI) have significantly influenced architectural and construction design practices. Generative AI tools are increasingly being adopted to enhance design processes, optimize building performance, and support early-stage concept development. For example, Li et al. [1] introduced a design workflow that translates sketches and text into architectural floorplans and 3D models using multimodal AI tools, demonstrating improved speed and control in concept generation. Similarly, Kakooee and Dillenburger [2] developed a deep reinforcement learning system, “laser-wall,” capable of generating diverse architectural layouts that satisfy both geometric and topological constraints.

On the industrial front, Zaha Hadid Architects (2025) report productivity gains of up to 200–300% in early-stage design competitions through AI integration, as well as a 50% increase in efficiency during design development. In the construction sector, Shawmut Design and Construction (2025) applies AI to monitor job-site safety, evaluate worker behavior, and predict potential hazards—improving safety outcomes for tens of thousands of workers. Despite these breakthroughs, challenges remain. According to the American Institute of Architects [3], only 6% of U.S.-based architects currently use AI regularly, citing barriers such as unclear ethical standards, limited training, and concerns over transparency.

This study builds upon these recent developments by evaluating how AI-generated architectural designs perform in terms of compliance, efficiency, and usability. It also investigates real-world implementation challenges, emphasizing the need for improved AI contextual awareness and broader industry adoption strategies.

In the construction industry, a small to medium scale construction project, beginning with Design phase to the Closeout phase approximately takes 6 to 24 months to complete and the intense nature of the Design stages in Architectural and Structural design takes up 15-20% [4] of the total project time. Due to technological advancements, the projects require more intricate details due to which the cost, resources and project timeline has increased [5]. Although technological advancements support the construction industry in delivering complex projects successfully, it comes with the baggage of daunting Design expectations from the end-users that directly influences the expanse of resources used, costing, as well as the timeline of the project. Secondly, with the growing need to construct an adaptable and cost-effective facility, the complexity of schematic and elevation plans has increased thereby increasing the need for more expertise in human resource and innovative solutions to solve the intricacies of the project. Thirdly, the Design phase involves multiple stakeholders, making it important to reduce any design clashes within the respective sub fields like MEP, Architecture and Structure. These demands juxtaposed with the slow adaptation of newer technologies in the construction industry has led to contrasting outcomes. While the timeline for the design phase has increased due to complexities of design and ineffective use of AI, the time frame for procurement and construction phase has reduced with advanced logistics and construction equipment. Thus, the construction industry is in a transitional phase where certain sections have adapted well to AI mechanisms while the others continue to rely on manual processes. Moreover, most professionals are grown into the ease of current processes and are resistant to transformative digital infrastructure thereby affecting the timeline of the construction projects.

This study aims to address the primary challenges of Architectural and Structural design. It will delve deep into the issues of integrating design tasks and drawings in one software and the need for Artificial Intelligence (AI) integration for design iterations that can reduce complications arising at the construction site with errors and insufficient details in drawings. It will also consider AI based software as a potential tool to optimize the design stages to bring the desired change in the construction industry. This research will explore the AI’s generative capabilities based on the interpretability of the design iterations, suitability of execution of drawings and compliance with state and local regulations. Moreover, it will also explore the adaptability of various design iterations for diverse geographical locations and varied regulatory requirements.

Furthermore, the study will take into account the impact of use of AI design software on human and material resources and its effect on the costing of the project. Secondly, it will analyse the efficiency of the designs by assessing the design iterations' capability to integrate and accommodate other engineering services like Mechanical, Electrical and Plumbing (MEP) services. Finally, the study will compare the positive and negative outcomes of the experiment models to provide an overview of the scope of AI based design software in the construction industry for Architects and Civil Engineers.

2. Research significance

This study has two interrelated objectives. First, it seeks to evaluate the performance of artificial intelligence in architectural design by using the ARCHITEChTURE platform to test the accuracy, adaptability, and compliance of AI-generated design outputs across diverse real-world case studies. Second, it aims to examine the practical challenges and limitations of deploying AI-assisted design tools in construction workflows, especially with respect to spatial optimization, regulatory compliance, and cultural or contextual fit. By integrating both performance evaluation and critical assessment, this research provides a balanced perspective on the current capabilities and future potential of AI in transforming architectural design within construction projects.

3. Literature review

Recent research has shown a marked acceleration in the integration of artificial intelligence (AI) in architectural and construction design. Li et al. [6] conducted a systematic review linking AI with net-zero carbon strategies, identifying machine learning (ML) and multi-objective optimization as key tools for energy-efficient design. Mehraban et al. [7] explored AI-powered performance optimization in hot climates using Building Information Modeling (BIM) and Gradient Boosting Machines, achieving high prediction accuracy for energy use intensity in buildings in Dubai and Riyadh. Kazemi et al. [8] applied AI for early-stage design of tall buildings, demonstrating the value of generative adversarial networks (GANs) and feature selection for form optimization. Meanwhile, Ploennigs & Berger [9] investigated AI art platforms for architectural ideation, offering insight into how diffusion models are reshaping conceptual workflows. Zhang et al. [10,11] complemented this by evaluating AI's generative performance against the designs of Antoni Gaudí, suggesting AI's strength in creativity but its limitations in authenticity and harmony.

These recent studies underscore both the expanding capabilities of AI in architectural design and the persistent limitations in contextual understanding, regulatory adaptation, and semantic richness. Building on this, the current study seeks to critically evaluate generative AI's performance using real-world test cases and empirical design scenarios, aiming to bridge the gap between conceptual experimentation and professional application.

The target communities addressed by this literature review on architectural and structural design include educational and technological research institutions, departments of architecture and architectural engineering, civil and structural engineering departments, as well as professionals in construction management, project management, and broader interdisciplinary research networks. This review evaluates several theoretical and computational models, such as the intelligent auxiliary models of architectural space [1], diffusion-based AI art platforms [12], and a wide range of AI, machine learning (ML), and deep learning (DL) algorithms [13], focusing on their effectiveness in supporting iterative design processes.

The reviewed experiments aim to enhance the efficacy of architectural design through AI-assisted auxiliary systems, including diffusion-based AI models capable of generating design concepts from textual inputs—an increasingly valuable capability in visually driven architectural tasks. Furthermore, the application of AI/ML/DL technologies is explored across the entire building lifecycle, encompassing the conceptual, schematic, and detailed design stages, as well as the construction and maintenance phases.

To validate the performance of these systems, the review also analyzes the statistical correlation between AI-generated training data and flawed architectural designs, using this comparison to assess the auxiliary capabilities of spatial intelligence models in error detection and design optimization. A comparison of public AI art platforms' capabilities, a description of the specifications for AI art platforms in order to support common use cases in architecture and civil engineering and analysis of 85 million Midjourney queries using Natural Language Processing (NLP) techniques in order to identify common usage patterns. This led to the development of a workflow that integrates the advantages of the various platforms to create images for interior designs and a workflow that creates views for exterior designs. These models are developed using data collection techniques that make use of sensors and smart vision, data cleaning techniques (post-processing), and data storage [13].

The research object is a 3D model chosen from the UrbanScene3D data set, and the auxiliary performance of AI's building space intelligent model is evaluated. The findings of the study indicate that the model fitting degree on the training and test data sets decreases as the number of network nodes rises. The fitting curve of the comprehensive model demonstrates the superiority of the AI-based intelligent architectural space design scheme over the conventional architectural design scheme. Nodes represent data points in AI and are used to find relations and patterns in data. These patterns help AI predict the future outcomes.. Thus, these tools can be optimized to enhance the early stages of architectural design with multiple options to ideate, design & Sketch.

According to the study conducted by researchers at Shandong University of Science and Technology [1], architectural design can be enhanced through the integration of spatial intelligence and AI-based auxiliary models. This approach leverages intelligent manipulation of command terms and keywords across multiple design models to generate diverse architectural schemes. These schemes are catalogued within AI-driven design software, enabling the formation of an auxiliary design framework underpinned by semantic networks and spatial logic analysis.

The system also incorporates evaluations of spatial structure, functional performance, and three-dimensional characteristics derived from data sources, which are validated using the UrbanScene3D dataset. The study demonstrates that as the number of network nodes increases, the model's fitting degree for both training and test datasets tends to decrease, indicating overfitting risks. However, the overall fitting curve of the proposed AI-based design scheme shows clear superiority over traditional architectural methods. Furthermore, the results highlight that with increased connectivity in the network layers, the intelligent assessment of environmental parameters—such as temperature and humidity—improves significantly.

The intelligent auxiliary design model for architectural spaces developed by researchers at Shandong University of Science and Technology demonstrates how AI can effectively enhance spatial planning through semantic network integration and data-driven modeling . In addition to its core function of spatial optimization, the model explores broader applications in cyber-physical systems, polymer material processing (e.g., ultrasonic welding), and disaster risk assessment. While the model contributes meaningfully to the digital transformation of architectural space design, it falls short in addressing the detailed planning of interior functions. Future research should prioritize the expansion of architectural semantic networks and the incorporation of deep learning to improve contextual intelligence and adaptability.

Complementing this technical approach, Ploennigs and Berger investigate the creative capabilities of AI in architectural design, focusing on generative and diffusion models. These tools enable diverse functions such as text-to-image generation, image editing, and visual enhancement, offering architects new modes of ideation and prototyping. Their analysis of three leading AI platforms and millions of user interactions reveals evolving workflows in architectural visualization, particularly for exterior and interior concepts. The integration of these tools holds the potential to boost creativity, efficiency, and iterative design across multiple stages. Their effectiveness is further supported by advancements in additive manufacturing and structural automation [14].

Despite their promise, current AI art tools present limitations. Semantic understanding remains a challenge, leading to inconsistent results that often require user refinement. Complex prompts are not reliably interpreted, and initial outputs frequently miss the intended spatial logic—especially for architectural floor plans. While visually convincing in stylistic representation, many outputs fail under close scrutiny. Additionally, subjective design dimensions such as artistic style, real-world context, and integration with 3D workflows remain underdeveloped [12].

Expanding the discussion to the construction sector, Baduge et al. provide a comprehensive review of AI, ML, and DL applications across the building lifecycle. Their study highlights applications such as generative floor plan creation, indoor scene synthesis, and material property prediction using models like GANs and VAEs. ML algorithms are increasingly used to forecast the behavior of construction materials and optimize composite performance. In structural engineering, AI supports analyses related to fatigue, buckling, and strength assessment. Moreover, AI technologies are advancing offsite manufacturing through modular design, robotic arm control, and 3D printing. Smart vision systems contribute to quality assurance and automated process monitoring. However, the study emphasizes that most AI applications in the construction industry are still in experimental stages, with limited commercial deployment and a need for further validation.

3.1. Parametricism in the 21st century

Youns and Grchev [15] provide a comprehensive evaluation of Parametricism, recognizing it as a transformative architectural movement driven by advancements in computer-aided design (CAD). This approach diverges from traditional methods by employing computational tools to create adaptable, fluid forms that respond to environmental, social, and economic factors. They argue that Parametricism represents not just a methodological innovation but a paradigm shift in architecture, fostering interdisciplinary collaboration by integrating insights from engineering, technology, and sustainability. This results in dynamic, context-aware structures that mirror the complexities of contemporary society.

A significant aspect of this evaluation is the integration of artificial intelligence (AI) within Parametricism. AI enhances the capacity of Parametricism to optimize design processes through the analysis of vast datasets, enabling architects to generate adaptive and efficient solutions. AI algorithms facilitate modeling environmental impacts, optimizing material usage, and improving energy efficiency, thereby aligning Parametricism with sustainability goals. This integration not only advances architectural design but also transforms problem-solving methodologies, empowering architects to address modern challenges with greater intelligence and empirical rigor.

Similarly, Hariri-Ardebili, et al. [16] conduct a narrative review that highlights the transformative potential of AI and digital technologies across various engineering disciplines, including dam operations. Their study explores how AI applications—such as predictive maintenance, real-time monitoring, and risk assessment—enhance the safety, efficiency, and sustainability of infrastructure. The authors also emphasize the significance of digital twins and machine learning algorithms in optimizing design and management processes. By underscoring the crucial role of AI and digital technologies in advancing engineering practices, this review complements the insights provided by Youns and Grchev, illustrating a broader trend where AI's integration in both architecture and engineering reshapes traditional methodologies and promotes innovative, data-driven approaches for the future.

3.2. Smart city construction and urban development

Li and Wang in 2024 employed an effective analytical framework to assess the impact of smart city policies on urban development. The use of the difference-in-differences (DID) model, combined with propensity score matching (PSM), offers a rigorous approach to examining causal impacts, particularly by treating smart city policy implementation as a quasi-natural experiment. This approach allows for a clear comparison between cities that adopted smart policies and those that did not, effectively isolating the policies' impact on urban green and high-quality development from other confounding factors. The incorporation of time and individual fixed effects further strengthens the model by controlling for unobserved variables that might otherwise distort results, thus ensuring the robustness of the study's conclusions [17].

Moreover, the study highlights key mechanisms through which smart city construction promotes development, such as technological innovation, industrial structure upgrading, and resource allocation optimization. These initiatives drive urban resilience, sustainability, and efficiency, resulting in more effective urban growth. The model's empirical validation reveals statistically significant positive effects of smart city policies on development metrics, confirming the proposed hypotheses. Additionally, robustness checks, including parallel trend tests, reinforce confidence that the observed improvements are attributable to the smart city policies and not to pre-existing trends. This methodological rigor is enhanced by the integration of PSM, which addresses self-selection bias by matching pilot smart cities with similar non-pilot cities, thereby improving the validity of the causal inferences. In conclusion, the DID model is instrumental in providing a robust framework for evaluating smart city policies, offering insights into how these initiatives foster urban sustainability. The study emphasizes the importance of integrating smart technologies into urban planning to achieve high-quality, sustainable urban growth, with the model's careful design ensuring reliable results and actionable policy recommendations [17].

3.3. AI in achieving net-zero carbon emissions

Li et al. reveal several key insights and future directions in their study on AI applications in achieving net-zero carbon emissions (NZCEs) for sustainable building projects. One of the most notable findings is the significant rise in research interest since 2019, peaking in 2021 with 28 published articles, indicating a growing focus on AI's role in addressing climate change in the construction industry. Leading journals such as the Journal of Cleaner Production and Applied Energy have driven much of this discourse, publishing 35.06% of the total articles. A keyword analysis identified six core research areas, with life cycle assessment, AI-based decision support systems, and multi-objective optimization emerging as critical themes. Particularly, machine learning algorithms have been widely employed, emphasizing their impact on improving energy efficiency and predicting carbon footprints.

The study highlights AI's transformative role in sustainable development, particularly in optimizing energy management, which accelerates the construction industry's transition towards achieving NZCEs. AI integration is seen as pivotal in minimizing environmental impacts and improving decision-making processes, offering substantial economic and environmental benefits. However, the study also uncovers significant research gaps. There is a limited focus on AI's practical application, especially its integration with existing frameworks such as Building Information Modeling (BIM). Moreover, the potential of multi-objective optimization and energy management systems remains underexplored. There is also a need for more longitudinal studies that assess AI's long-term impacts on sustainability in the construction industry. In conclusion, while AI offers immense potential to revolutionize sustainable building practices, targeted research is essential to fully realize its benefits in achieving NZCEs, especially through practical applications, systems integration, and comprehensive long-term studies [1].

3.4. Optimizing energy performance in hot climates

Mehraban et al. [7,18] offer significant insights into enhancing energy efficiency in residential buildings within arid environments through their study. By integrating Building Information Modeling (BIM) with Machine Learning (ML), the research facilitates a detailed analysis of building features and their impacts on energy consumption. Notably, the Gradient Boosting Machine (GBM) algorithm achieved an impressive R^2 value of 0.989 for predicting Energy Use Intensity (EUI), underscoring its robustness as a predictive tool. The feature importance analysis revealed that roofs, external walls, and windows are critical in determining energy usage, with roofs alone accounting for 29% of energy consumption in Dubai and 40% in Riyadh. The study also explored innovative strategies for energy optimization, such as cavity green walls and vacuum insulation panels, tailored to the unique climatic conditions of each city. This focus on comparative analysis enables stakeholders to identify effective materials and designs, aligning with global sustainability goals.

However, the study acknowledges limitations that could impact the robustness of its conclusions. Its narrow geographical focus on Riyadh and Dubai may limit the generalizability of findings to other hot climates. Additionally, modeling assumptions, reliance on a single weather station per city, and the evaluation of a limited number of scenarios could skew results. Furthermore, the choice of only four ML algorithms restricts the exploration of potentially superior methods, and the absence of real-world validation raises questions about the practical applicability of predictions. In conclusion, while the study provides valuable insights into optimizing energy performance, caution is warranted in applying its findings broadly. Future research should encompass a wider geographical scope, diverse scenarios, and real-world data validation to enhance reliability across various contexts.

3.5. AI in early design of tall buildings

Kazemi et al. [19] explore the transformative role of machine learning (ML) and parametric design in the initial phases of tall building design. Their study presents critical insights that underscore the potential of AI to enhance structural efficiency while maintaining aesthetic integrity. One of the key findings is the exploration of four data augmentation techniques—Gaussian copula, conditional generative adversarial networks (CGAN), Gaussian copula generative adversarial network (GCGAN), and variational autoencoder (VAE). The study favors the Gaussian copula method for its simplicity in hyperparameter tuning, demonstrating the importance of accessible and effective data augmentation strategies. Moreover, the research emphasizes the significance of feature selection, identifying seven critical variables that impact design outcomes, thereby facilitating informed architectural decisions. Utilizing a random forest regressor, the analysis yielded an average coefficient of determination (R2) of 0.781 for synthetic data, highlighting the improvement in predictive accuracy compared to the original limited dataset. Additionally, correlation analysis helped identify five primary responses that guide design choices, linking aesthetic ambitions with structural performance.

However, the study also acknowledges substantial challenges. Limited data availability can lead to inconsistencies and reduce the reliability of ML predictions. The complexity of architectural data poses difficulties in identifying relevant features, while high-dimensional datasets risk overfitting, complicating generalization to new designs. Furthermore, the necessity of data augmentation raises concerns about the validity of synthetic data, and balancing aesthetic goals with structural integrity remains a critical challenge in early design phases. Overall, this research underscores the potential of AI to revolutionize architectural design, yet it also highlights the need for interdisciplinary collaboration and further refinement of methodologies to address existing challenges. By integrating AI effectively, architects and engineers can explore a wider range of design possibilities while ensuring structural efficiency, paving the way for more innovative tall buildings.

3.6. Emerging technologies and AI in urban design

He and Chen [20] provide an insightful analysis of how AI and innovative technologies are reshaping urban planning. The research emphasizes the transformative potential of AI in enhancing decision-making, predictive analytics, and the automation of complex tasks, ultimately leading to smarter, more sustainable cities. One of the study's notable strengths is its adherence to the PRISMA protocol for systematic reviews, which ensures methodological rigor and transparency. By analyzing 61 relevant articles, the research offers a comprehensive overview of AI applications in urban design, categorized into themes such as transportation optimization, built environment assessment, and urban ecosystem services. This thematic approach highlights the diverse methodologies and technologies employed to tackle urban challenges effectively. The study identifies a wide array of data sources utilized in geo-design, including satellite imagery and social media data, enhancing the accuracy of AI technologies in understanding urban dynamics. Furthermore, it reveals a strong interdisciplinary interest in integrating AI with fields such as geography and environmental science, essential for addressing complex issues like climate change and sustainability.

However, the study also confronts significant challenges in implementing AI in urban design. Data privacy concerns emerge as a critical issue, particularly regarding the use of personal information from social media. Ethical dilemmas related to algorithmic bias and the need for cross-disciplinary collaboration further complicate the integration process. Additionally, the effectiveness of AI applications hinges on data quality, necessitating continuous updates and validation. In conclusion, while the study underscores the promise of AI in advancing urban planning and enhancing resilience against various challenges, it calls for ongoing exploration and strategic planning to navigate the complexities of implementation. Addressing these challenges is crucial for realizing the full potential of AI in creating livable and sustainable urban environments.

3.7. Seismic intensity parameters for RC structures

Tyrtaiou et al. [21] explore the innovative use of artificial neural networks (ANNs) in estimating seismic damage in reinforced concrete (RC) structures. Utilizing the Hilbert-Huang Transform (HHT), the research introduces 40 new seismic intensity parameters that effectively capture the intricate frequency-time characteristics of seismic signals, offering enhanced accuracy over traditional methods. Key findings indicate that the multi-layer feedforward perceptron (MFP) networks employed in the study are adept at modeling the relationship between these HHT-derived parameters and structural damage, quantified through the Park and Ang overall damage index (DIPA,global). Performance metrics, including mean squared error (MSE) and correlation coefficients, affirm the high accuracy of the ANNs in predicting seismic damage, thereby validating the utility of HHT parameters in assessing seismic vulnerability.

The study not only demonstrates the effectiveness of this computational intelligence approach for estimating seismic damage in RC structures but also fills a critical gap in earthquake engineering with the introduction of novel HHT-based parameters. Importantly, the methodology exhibits strong generalization capability, allowing for application across different seismic intensity parameters and structural types, enhancing its relevance in civil engineering. Furthermore, the framework proposed in this research has potential for broader applications beyond RC structures, suggesting its viability for diverse materials and structural forms in earthquake risk assessment. Overall, this study represents a significant advancement in the integration of advanced signal processing techniques with machine learning, thereby improving the precision of seismic damage assessments and informing performance-based design strategies in earthquake engineering.

3.8. AI in architectural design

Zhang, Fort, and Giménez Mateu [11] critically examine the role of artificial intelligence (AI) in architectural design, specifically through the lens of Antoni Gaudí's iconic works. The comparative analysis evaluates AI-generated designs against those of Gaudí using five metrics: Authenticity, Attractiveness, Creativity, Harmony, and overall Preference. The findings reveal that while Gaudí's designs excel in Authenticity and Harmony, showcasing the unique coherence of human creativity, AI-generated designs demonstrate competitive results in Attractiveness and Creativity. This suggests that AI can produce visually appealing outcomes but struggles to replicate the nuanced authenticity and harmony found in Gaudí's architecture.

Furthermore, the study highlights the inherent subjectivity in assessing architectural aesthetics, as participant opinions varied widely on both Gaudí's and AI's designs. It positions AI as a promising tool that can enhance creativity and challenge traditional methods, yet it falls short in capturing the distinctive qualities of human design. The research advocates for integrating AI into architectural education to deepen students' understanding of design principles and improve communication between students and instructors. Overall, while AI shows potential in generating innovative designs, it currently cannot match the depth and coherence inherent in the works of master architects like Gaudí, prompting further exploration of personalized AI algorithms to better accommodate diverse aesthetic preferences

Ling et al. investigate the impact of digital technology adoption on the comparative advantage of architectural, engineering, and construction (AEC) firms in Singapore. The findings reveal that the integration of digital technologies significantly enhances the competitive edge of AEC firms by improving project efficiency, reducing costs, and fostering innovation. The study highlights that firms leveraging advanced digital tools are better positioned to meet client demands and adapt to market changes. However, a notable limitation of the research is its focus on firms within Singapore, which may limit the generalizability of the findings to AEC firms in other regions with different technological infrastructures and market conditions. Additionally, the study does not address the specific role and limitations of artificial intelligence (AI) in digital applications. While AI has the potential to further enhance efficiency and innovation, it also faces significant challenges such as data dependency, lack of contextual understanding, and the need for substantial computational resources. These limitations could impact the effectiveness of AI integration in AEC firms, suggesting that future research should explore these aspects to provide a more comprehensive understanding of digital technology adoption in the industry.

3.9. Conclusion of literature review

In conclusion, the application of AI has immense potential and according to the findings, using a personalized comfort model and AI technology reduced energy use by an average of 21.81% to 44.36%.

The average increase in user comfort perception was 21.67% to 85.77%.(Li et al., 2023).Moreover the comprehensive model's fitting curve demonstrates that the artificial intelligence (AI)-based intelligent architectural space design scheme outperforms the conventional architectural design scheme. Also, Construction AI/ML/DL algorithms have the potential to expedite the digitization of the construction sector, producing an immense amount of data that can be leveraged to optimize operational efficiency, facilitate well-informed decision-making, stimulate innovation and expansion, and augment sustainability.

However, challenges do exist. Models can be improved by enhancing semantic comprehension, optimize Training data to improve AI. Also, the industry should encourage interdisciplinary cooperation by bringing together deep learning specialists, interior designers, architects and Civil Engineers. The study of the construction industry can benefit from the varied viewpoints and specialized knowledge that this multidisciplinary approach can offer. Moreover, the researchers should focus on user-centred Research to comprehend the practical requirements as well as needs of users in building spaces and conduct user studies. The creation of more feasible and approachable models can be influenced by this data. Funding for Research and Development is also an option to promote Government and Industry Cooperation. Promoting the industry professionals for AI research and development in the construction industry and to award companies and academic institutions developing cutting-edge AI solutions for the construction industry grants and other financial incentives. Development can be enhanced, and investment can be a great motivation factor to improve the industry.

4. Research methodology and framework

To introduce AI integrated design as a potential solution in the construction industry, the study aims to explore four existing projects on the AI platform called 'ARCHITEChTURE' to assess AI's accuracy in design iterations using two established metrics: the Confusion matrix and the ROC-AUC measures. The fundamental analysis entails assessing an AI model's performance on ARCHITEChTURE platform demonstrating the application of machine learning methods.

4.1. Experimental

The four projects under review and analysis are Revit models of construction projects: Snowden Towers from Brownsville, Pennsylvania, USA; a residential building in Sweden; another residential building in Monterey, Mexico; and a high-end residential building from an undisclosed location. The study delves into how the AI-generated models from the ARCHITEChTURE AI Platform handle the complex structure of Snowden Towers, the European aesthetic design of the Swedish residential building, the local building codes in the different geographical context and climate of the Mexican buildings, and the unknown location.

The study involves adjusting one variable in the AI model on the ARCHITEChTURE AI Platform and tracking how that alteration affects another variable (the emulation accuracy when compared to an existing project). The dependent variables measured for emulation accuracy include geometric and functional similarity. Evaluation of projects is based on size, type, complexity, and data availability for that specific project type. Additionally, assessing the AI's performance across various projects requires conducting several test runs before arriving at a valid conclusion.

4.2. Simulation setup and workflow

The computational simulations were conducted using the ARCHITEChTURE AI platform, a generative design environment tailored for architectural applications. This platform integrates multiple AI techniques to deliver optimized design iterations based on real-world architectural constraints. The simulation employed the following AI methodologies:

- **Convolutional Neural Networks (CNNs)** for spatial feature extraction and pattern recognition in design layouts.
- **Natural Language Processing (NLP)** to interpret user-defined requirements and translate them into architectural constraints.
- **Reinforcement Learning (RL)** to iteratively refine solutions based on performance feedback from previous design outputs.

The simulation was configured using a set of critical architectural input parameters, including net usable floor area requirements, room typology and dimensional constraints, compliance with local regulatory codes, and vertical circulation elements such as elevator and stair placements. The simulation workflow followed a structured, four-stage process as Shown in Fig. 1. First, real-world Revit models were collected from four case studies located in the USA, Mexico, Sweden, and one undisclosed location.

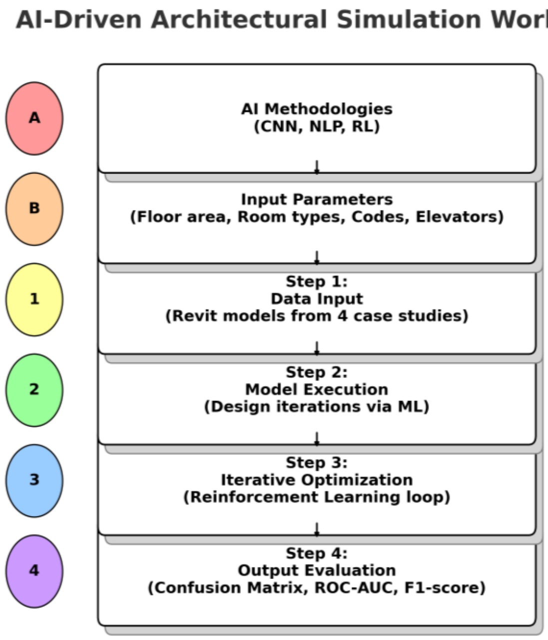


Fig. 1. AI-driven architectural simulation workflow illustrating the key stages from model configuration to evaluation.

These models served as the ground-truth datasets for training and evaluation. In the second stage, the AI platform processed the input parameters and generated multiple design iterations using embedded machine learning algorithms. The third stage involved iterative optimization through reinforcement learning, wherein the system continuously refined its outputs based on the performance feedback of previous iterations. The final stage consisted of output evaluation, where the performance of each AI-generated model was assessed using a combination of confusion matrices—measuring true positives, false positives, true negatives, and false negatives—along with ROC-AUC curves to evaluate classification thresholds, and F1-scores and overall accuracy metrics to benchmark predictive precision and model balance.

4.3. Metrics for accuracy

In the analysis, quantitative variables represent the accuracy of AI model metrics, utilizing measures such as the Area Under the Curve (AUC) on Receiver Operating Characteristic curves. This metric gauges the model's ability to differentiate between positive and negative cases across various thresholds. Meanwhile, Confusion matrices provide a detailed insight into true positives, false positives, true negatives, and false negatives, highlighting specific types of errors made by the model.

Statistical analysis is conducted to assess the accuracy of the AI model across multiple projects and trial runs, incorporating relevant statistical tests like Precision, Recall, F1 score, and Accuracy. Additionally, a qualitative evaluation of the AI-generated models is performed to determine their adaptability to project requirements, versatility in handling diverse designs and aesthetics, and adherence to building standards and regulations.

4.4. Confusion matrix

When evaluating the performance of a classifier using a set of test data, a confusion matrix, first introduced by Karl Pearson in 1904, serves as a valuable tool. This two-dimensional matrix is structured based on the actual values of the test subjects and the predicted classes assigned by the classifier. In essence, it summarizes how well the classifier performs in categorizing the test data. In the context of analyzing Architectures AI, a specialized adaptation of the confusion matrix is employed, focusing on distinguishing between positive and negative classes. This tailored confusion matrix delineates six key categories within its cells (Table). The outcomes of classification into positive and negative classes.

Table 1
Summary of performance metrics and diagnostic indicators derived from a binary classification confusion matrix.

Total population = P + N	Predicted Positive (PP)	Predicted Negative (PN)	Bookmaker Informedness (BM) = TPR + TNR − 1(BM)	Prevalence threshold $PT = \frac{(\sqrt{TPR \times FPR} - FPR)}{(TPR - FPR)}$
Positive (P)	True positive (TP), hit	False negative (FN), miss, underestimation	True positive rate $TPR = {TP}/{P} = 1 - FNR$	False negative rate $FNR = {FN}/{P} = 1 - TPR$
Negative (N)	False positive (FP), false alarm, overestimation	True negative (TN), correct rejection	False positive rate $FPR = FP / N = 1 - TNR$	True negative rate $TNR = {TN}/{N} = 1 - FPR$
Prevalence $= {P}/{P + N}$	Positive predictive value $PPV = {TP}/{PP} = 1 - FDR$	False omission rate $FOR = FN / PN = 1 - NPV$	Positive likelihood ratio $LR^{+} = TPR / FPR$	Negative likelihood ratio $LR^{-} = FNR / TNR$
Accuracy $ACC = (TP + TN) / (P + N)$	False discovery rate $FDR = FP / PP = 1 - PPV$	Negative predictive value $NPV = TN / PN = 1 - FOR$	Markedness $MK = PPV + NPV - 1$	Diagnostic odds ratio $DOR = LR^{+} / LR^{-}$
Balanced accuracy $BA = (TPR + TNR) / 2$	F_1 score $F1 = 2 \times PPV \times TPR / (PPV + TPR) = 2 \times TP / (2 \times TP + FP + FN)$	Fowlkes–Mallows index $FM = \sqrt{(PPV \times TPR)}$	Matthews correlation coefficient $MCC = \sqrt{(TPR \times TNR \times PPV \times NPV)} - \sqrt{(FNR \times FPR \times FOR \times FDR)}$	Threat score (TS), critical success index $TS = TP / (TP + FN + FP)$

- **True Positives** are iterations where the AI predicts the design feature correctly whereas False Positives are iterations where the AI predicts incorrectly the presence of design features that are not actually there.

- **True Negatives** are iterations where the AI predicted correctly, the absence of specific design features whereas False Negatives are iterations where the AI prediction fails to determine the presence of certain design features that are present.
- **Precision:** Represents the proportion of instances predicted as positive that are actually positive. It measures the model's ability to avoid false positives.
- **Recall**: Represents the proportion of actual positive cases that are correctly identified by the model. It measures the model's ability to find all true
- **Accuracy:** Represents the overall proportion of predictions that are correct (both positive and negative).
- **F1-Score:** harmonic mean of precision and recall, combining information from both metrics into a single value. It considers both the balance between true positives and false positives, as well as the model's ability to find all true positives [22].

4.5. Receiver operating characteristics (ROC) curves

To visualize and assimilate the data in confusion matrix by performance-based selection of classifiers, a receiver operating characteristics (ROC) graph is used [23]. Receiver Operating Curve is an application that can separate observer variability from the innate delectability of signal to distinguish between positive and negative cases. ROC represents two dimensions by having the True positive rate at the Y axis and False positive rate plotted at the X axis. The ROC curve is created by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings that gives an overview of the salient features between the true and false predictions. The ROC curve compares the TPR with FPR with change in threshold criterion [23].

The true positive rate (also called sensitivity) of classifier is estimated as Eq. (1).

$$tp\ rate = \frac{\text{Positives correctly classified}}{Total\ positives} \tag{1}$$

The false positive rate (also called *specificity)* of t classifier is Eq. (2).

$$fp\ rate = \frac{\text{Negatives incorrectly classified}}{Total\ negatives} \tag{2}$$

4.6. Area under the curve (AUC)

The Area Under the Curve (AUC) is a metric commonly used to evaluate the performance of binary classification models, such as those used in machine learning and statistics. It quantifies the model's ability to distinguish between classes and is particularly useful when dealing with imbalanced datasets. AUC measures the entire two-dimensional area underneath the receiver operating characteristic (ROC) curve, which plots the true positive rate (sensitivity) against the false positive rate (1 - specificity) for different classification thresholds. A perfect model has an AUC of 1, indicating that it achieves a true positive rate of 1 (sensitivity) and a false positive rate of 0 (specificity) across all possible thresholds. On the other hand, a random classifier has an AUC of 0.5, signifying no discriminative power beyond random chance. The AUC provides several advantages [24].

- **Threshold Invariance:** AUC is insensitive to the choice of classification threshold, making it suitable for evaluating models across various operating points.
- **Interpretability:** AUC is easily interpretable; a higher AUC indicates better overall performance in distinguishing between classes.
- **Robustness: AUC** is robust to class imbalance, which is common in real-world datasets.

- **Model Comparison:** AUC facilitates the comparison of different models by providing a single scalar value representing their discriminatory ability.

5. Overview of the generative AI platform used for testing

ARCHITEChTURES AI is a generative design platform for architects and engineers to optimize building and space design. This study has used this ARCHITEChTURES AI platform to recreate design models for the four case studies taken up for this research. The generative AI technology used by this platform provides the most efficient solution on each iteration based on the geometry and design criteria entered. The platform functions on three simple steps. In the input stage, the user of the platform enters the design requirements and defines and models the solution online in 2D and or 3D. In the design optimization stage, the cloud-based IA system creates a real- time geometry as per the parameters instructed for the iterations by the end-user. In the final stage of output, the BIM solution is generated with project data and metrics in various formats based on the usability of the user (*AI-Powered Building Design*)

This AI aided design software has certain unique features. Users can manipulate the design criteria like minimum or maximum net area, typology and dimensions of a room, heights and vertical communication specifications to meet the project compliance regulations. Secondly, it supports AI-integrated manual project modelling. The automated design process is carried out on the user's 2D and 3D models of the volume of the above-ground and below-ground structures and parking lots. The generated designs are readily manually modified to meet the needs of the user. Moreover, with the geometry from the AI-assisted design process, the system creates a BIM model in real time, all with an online data structure that is fully navigable to allow for user evaluation. This real-time online navigable BIM model for evaluation of data structures and design editions supports real-time solutions that best suit the user standards. Additionally, the entire design is tracked, displaying in real time the unit's program, gross floor areas by use and type, a breakdown of the gross and net floor areas for each type, and each solution's compliance with urban requirements. The hallmark of this software is its real-time costs and quantities take-offs. In order to allow the user to enter the unit cost for each line, the system executes a thorough take-off of each work unit in real time. This makes it possible for our users to continuously be aware of the financial effects of every design choice he takes. (*AI- Powered Building Design*)

5.1. Framework of ARCHITEChTURE AI platform

The ARCHITEChTURE platform like any other AI-integrated software functions on the AI framework. The AI framework provides a range of libraries and tools that make it comparatively simple and effective for developers to construct, train, and validate complicated AI models. These frameworks act as a sort of "warp-drive," pushing the boundaries of AI research and development. Some of the crucial components of the AI framework employed by the ARCHITEChTURE platform are discussed below which have a significant impact on the design phase.

i. Supervised Learning: The ARCHITEChTURE platform uses the labelled data to train models and to classify or predict new data. The algorithm AI, for instance, can assess previous construction projects to forecast future design expenses, energy usage, or structural stability. Supervised learning teaches models to produce the desired result using a training set. The model can learn over time thanks to the inputs and accurate outputs in this training dataset. Through the loss function, the algorithm gauges its accuracy and makes adjustments until the error is suitably reduced. When using data mining, supervised learning may be divided into two categories of problems: regression and classification. An algorithm is used in classification to precisely place test data into designated groups. It identifies particular entities in the dataset and makes an effort to make recommendations on the definition or labelling of those items. To comprehend the link between dependent and independent variables, regression is utilized. It is frequently used to project things like sales income for a certain company. Popular regression algorithms include polynomial, logistical, and linear regression [25].

ii. Unsupervised Learning: The ARCHITEChTURE platform uses Unsupervised learning to find patterns in data that are unlabelled. This method reveals hidden information that architects can utilize. Weather trends and occupant behaviour analysis could help inform designs that are optimized for energy efficiency and comfort. It is common to discuss supervised and unsupervised machine learning in tandem. Unsupervised learning makes use of unlabelled data, in contrast to supervised learning. It finds patterns in that data that aid in resolving issues with association or clustering. When subject matter experts are unaware of common properties within a data set, this is especially helpful. The Gaussian mixture models, k-means, and hierarchical clustering methods are commonly used [25].

iii. Reinforcement Learning: The ARCHITEChTURE platform uses reinforcement learning to support the AI to learn by making mistakes and then refining its answers repeatedly. By simulating and analysing various options, this technique can optimize designs for a variety of criteria, such as material usage, thermal performance, or occupant satisfaction. In the field of machine learning, reinforcement learning has shown promise in solving sequential decision-making issues, which are usually accompanied by uncertainty. Examples of this include challenges with control such as autonomous manufacturing operations or production plan control; problems with resource allocation in finance or operations; and inventory management with various echelons and different suppliers with lead times under demand unpredictability [25].

iv. Generative AI: The ARCHITEChTURE platform uses the generative AI model as a cutting-edge method to create completely new designs or design elements by using neural networks. Using parameters that architects can input, the AI creates design possibilities based on patterns and styles it has learned, encouraging creativity and discovery. (Newton, 2019)

v. Convolutional Neural Networks (CNNs): The ARCHITEChTURE platform also uses the CNNs. These image-processing algorithms are particularly good at tasks like identifying architectural styles, assessing building layouts, and identifying possible weaknesses in structural designs. Deep learning techniques are based on neural networks, which are a subset of machine learning. They are made up of node layers, which have an output layer, an input layer, and one or more hidden levels. Every node has a threshold and weight that are connected to one another. A node is activated and sends data to the following layer of the network if its output exceeds a given threshold value. If not, no data is transferred to the network's subsequent tier. Although the majority of that article was devoted to feedforward networks, there are other varieties of neural nets that are employed for diverse applications and kinds of data [25].

vi. Natural Language Processing (NLP): AI can now comprehend and react to human language thanks to this technique. AI systems can receive design requirements or ideas from architects, which makes the design process more collaborative and intuitive. To enable computers and other digital devices to detect, comprehend, and produce text and speech, natural language processing, or NLP, combines statistical and machine learning models with computational linguistics, which is rule-based modelling of human language. The ARCHITEChTURE platform uses the Neural language processing (NLP), a subfield of artificial intelligence (AI) to power devices and apps that can translate text between languages and reply to spoken or typed commands identify or verify users based on their voice summarize extensive textual content evaluate the meaning or tone of text or speech create text, graphics, or other information as needed, frequently in real time. The majority of people today have dealt with natural language processing (NLP) through voice-activated GPS units, digital assistants, speech-to-text dictation software, chatbots for customer support, and other consumer conveniences. (Newton, 2019)

5.2. Testing parameters

The parameters highlighted in this analysis as summarized in Table 2 has been adopted in our analysis and they are crucial for assessing the performance and accuracy of an AI generative model in architectural design. Space efficiency parameters such as room optimization, parking spaces, and commercial units ensure that the model can allocate floor area effectively, adhering to minimum size requirements and providing clear construction instructions as per the International Building Code [26]. Aesthetics, through parametric design, play a vital role in resident well-being and satisfaction and ensure that building facades harmonize with their environment. Functionality, evaluated by elements like windows and doors, impacts natural light, ventilation, and energy efficiency, as well as building code compliance for accessibility and safety. Balconies, with their dual role in safety and aesthetics, and column placements, essential for structural integrity and construction accuracy, further underline the importance of these parameters. These criteria collectively ensure that the AI model produces designs that are not only functional and safe but also aesthetically pleasing and environmentally integrated.

Table 2
Parameters Used For Accuracy Testing.

Major Parameters		
Criteria	Parameter used in accuracy testing	Importance
Space Efficiency	Rooms/Optimization of Area/Parking spaces/Commercial units	Optimizing floor area allocation for functionality. (Preiser, W. D., & Vischer, E. F. (2006). Ensuring rooms meet minimum size requirements. Providing clear instructions for contractors to frame rooms to the correct size as per (International Code Council. (2021). International building code. International Code Council)
Aesthetics	Parametric Design	Resident Well-being: Studies like (Mfon, 2023) demonstrate a connection between aesthetically pleasing buildings and increased resident satisfaction and comfort. (Mfon, 2023) Building Integration: Architectural drawings ensure the building's facade, materials, and ornamentation harmonize with the surrounding environment. [27]
Functionality	Windows	Natural Light and Ventilation: Window size, placement, and style determine the amount of natural light and ventilation entering a building, impacting occupant comfort and energy use. Strategic window placement, can significantly reduce heating and cooling loads.
	Doors	Building Code Compliance: Doors meet size requirements for accessibility and fire safety. (International Code Council. International building code. International Code Council)
Functionality	Balconies	Safety: Specifying size, style, and weight limitations ensures balconies are constructed to safely support intended loads. (Guide to the design of balconies and terraces) Aesthetics: Balconies contribute to the overall architectural style of the building.
	Column Placements	Structural Integrity: Columns support the building's weight, and their placement ensures the building's stability. Construction Accuracy: Clear guidance for contractors on column locations, sizes, and materials is essential for proper framing.

6. Experiments: analysis

To validate the generative AI design and compare it to real-world designs, four cases were used for experimentation and analysis. These four cases vary in their design features, covering a majority of the design criteria subject to assessment. Each case study has its unique architecture and design features, providing a robust test for the generative AI's capability in predicting these parameters. Accordingly, each case was analyzed using a confusion matrix, where the accuracy rate of the generative designs was calculated based on specific parameters. This method allowed for a detailed evaluation of how well the AI model performs across different architectural scenarios, ensuring a comprehensive assessment of its accuracy and effectiveness in generating viable design solutions.

6.1. Case study 1 – snowden towers

This case offers a unique analysis since its challenging to design mixed used buildings due to accommodation of both commercial and residential considerations and balancing the user comfort of both the user cases. The conflicting design considerations have given the AI a great challenge, thus the importance is highlighted on training the AI on diverse data sets.

6.2. Findings of the experiment

The analysis of the confusion matrix in Table 3 for the Snowdon Towers Design Data (Mixed Use) provides valuable insights into the performance of the classification model across different parameters and sub-locations. By examining the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) for each category, we can assess the model's ability to accurately identify various design elements within the architectural plans. For instance, on the parking deck floor, the model demonstrates high accuracy in detecting parking spaces, with a True Positive Rate (TPR) of 0.919 and a False Positive Rate (FPR) of 0. The first floor plan also shows effective classification results, particularly for commercial and retail spaces, with TPR values exceeding 0.9. However, some areas, such as doors and windows, exhibit lower classification accuracy, as indicated by lower TPR values. The following analysis presents a detailed evaluation of the AI's performance in emulating various design elements:

True Positive values: AI was successfully able to emulate 9 parking's. Moreover, it successfully emulated the 3 lifts and staircases and most importantly was able to emulate the machine room for the lift. Furthermore, the AI was able to assign space for commercial units, typical studios, 77 doors and 56 windows.

False positive values: AI emulated 25 additional commercial units and 27 restaurant units.

Moreover, it emulated additional 8 studios.

False Negative Values: AI was unable to emulate Utility rooms, 11 parking spaces, Residential lobby, restaurants, outdoor covered dining, 6 toilets, staircase access to basement, Staricase access to lobby, 4 typical offices, 2 duplex studios, 54 windows and 56 doors.

Observing the TPR values across different sub-locations, as the TPR decreases (indicating a lower rate of correctly identifying positive instances), the number of False Negatives tends to increase. For instance, in the "Typical Studios / Work Live in Units" category, where the TPR is 0.80107, there are 8 False Negatives. In contrast, in the "Typical Offices" category with a TPR of 0.89784, there are only 4 False Negatives. This negative trend suggests that as the model becomes less effective at correctly identifying positive instances, the number of false negatives tends to rise.

However, it's important to note that the correlation shown in Fig. 2 doesn't necessarily imply interconnection. Other factors such as the complexity of the design features or the quality of the training data could also contribute to this relationship. Overall, this correlation underscores the importance of optimizing the model's sensitivity to improve its ability to accurately identify positive instances, ultimately enhancing its overall performance in architectural design applications.

Table 3
Confusion Matrix Results for Snowdon Towers.

Snowdon Towers Design Data (Mixed Use)										
Parameters		TP	TN	FP	FN	CUM T	CUM F	TPR	FPR	AUC
Location	Sub location									
Parking Deck floor	Utility room				1	0	0	1	1	0
Parking Deck floor	Parking	9			11	9	0	0.95161	1	0
Parking Deck floor	Elevator machine Room	3				12	0	0.93548	1	0
Parking Deck floor	Staircase	3				15	0	0.91935	1	0.9193
First Floor Plan	Outdoor Covered Dining		1		1	15	1	0.91935	0	0
First Floor Plan	Residential lobby				1	15	1	0.91935	0	0
First Floor Plan	Commercial/ Retail	1		25		16	1	0.91397	0	0
First Floor Plan	Commercial/ Retail	1				17	1	0.90860	0	0
First Floor Plan	Commercial/ Retail	1				18	1	0.90322	0	0
First Floor Plan	Restaurant				1	18	1	0.90322	0	0
First Floor Plan	Outdoor Covered Dining				1	18	1	0.90322	0	0
First Floor Plan	Toilets (6)				6	18	1	0.90322	0	0
First Floor Plan	Staircase access to basement				1	18	1	0.90322	0	0
First Floor Plan	Staircase from lobby				1	18	1	0.90322	0	0
Second Floor Plan	Mezzanine Dining	1		27		19	1	0.89784	0	0
Typical	Typical Offices				4	19	1	0.89784	0	0
Typical	Typical Studios / Work Live in Units	18		8		37	1	0.80107	0	0
Typical	Duplex Studios				2	37	1	0.80107	0	0
	Area	1			0.6	38	1	0.79569	0	0
	Doors	77			56	115	1	0.38172	0	0
	Windows	56			54	171	1	0.08064	0	0
	Balconies/Ter races	15		49		186	1	0	0	0
Total		186	1	109	140.6					0.91935

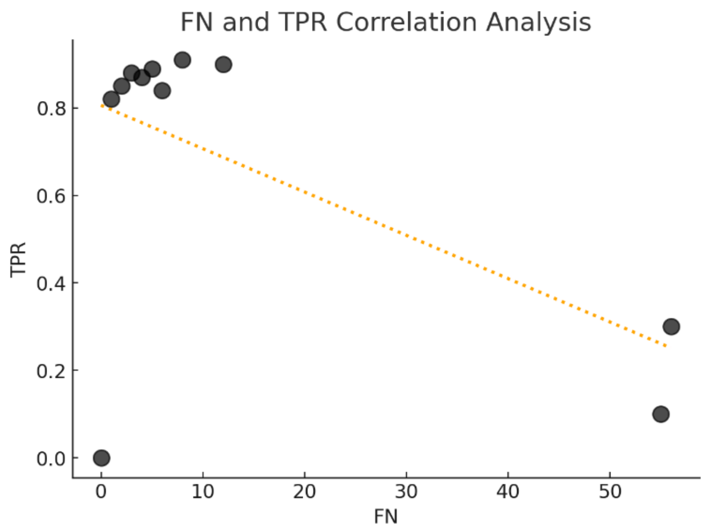


Fig. 2. FN and TPR Correlation Analysis for Snowdon Towers.

In the provided Snowdon Towers Design Data Table 3, there seems to be a positive correlation between False Negatives (FN) and Cumulative True (CUM T) as shown in Fig. 3. Observing the data, as the number of False Negatives increases, there is a corresponding increase in Cumulative True values. For example, in the "Parking Deck floor - Staircase" category, there are 15 False Negatives and 15 Cumulative True instances. Similarly, in the "First Floor Plan - Toilets (6)" category, there are 6 False Negatives and 6 Cumulative True instances. This positive correlation implies that as the model fails to correctly identify positive instances (False Negatives), the cumulative count of true instances also tends to rise. This could be due to the model misclassifying instances that should have been correctly identified, leading to an increase in the total count of true instances.

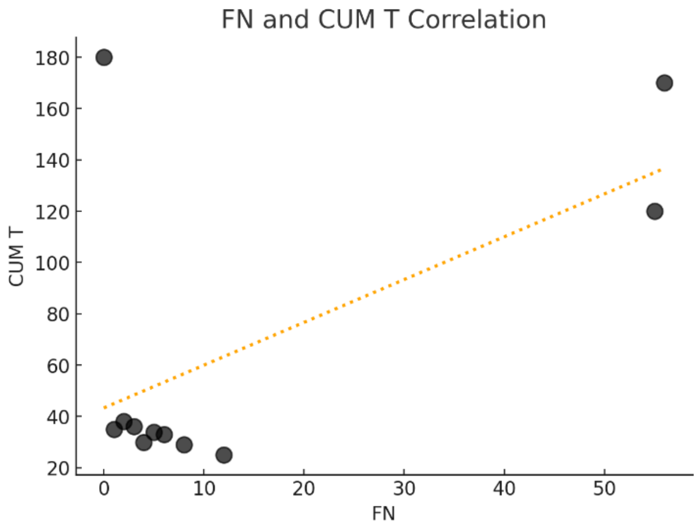


Fig. 3. FN and TPR Correlation Analysis for Snowdon Towers.

Table 4 provides a concise summary of the classification model's performance in predicting design features within the dataset The matrix distinguishes between predicted correct and incorrect design features compared to the actual design features. The diagonal elements represent the correct predictions, where the model accurately identified design features, resulting in 186 correct predictions of actual correct design features and 0 incorrect predictions of actual incorrect design features. However, the off-diagonal elements indicate errors in prediction. The model incorrectly classified 109 instances of actual correct design features as incorrect, while surprisingly, it made no incorrect predictions of actual incorrect design features. This suggests that while the model is relatively effective at identifying actual correct design features, it struggles more with misclassifying them, potentially leading to false negatives. Further analysis of the model's performance and adjustments may be necessary to improve its accuracy and reduce misclassifications.

Table 4
Confusion Matrix of AI Model of Snowden Tower.

Overall Prediction – Case of Snowden Tower		
	Predicted Correct Design Features	Predicted Incorrect Design Features
Actual Correct Design Features	186	109
Actual Incorrect Design Features	140.6	0

The statistical test data provided in Fig. 4. offers a comprehensive evaluation of the AI model developed for Snowden Tower. With a precision of 0.6305, the model demonstrates its ability to accurately identify relevant design features. A recall score of 0.5695 indicates the model's capacity to capture a considerable portion of actual design features within the dataset. However, the accuracy score of 0.4296 suggests room for improvement in overall predictive performance. The F1-score, at 0.5985, reflects a balanced measure of precision and recall, highlighting the model's effectiveness in handling both false positives and false negatives. Notably, the sensitivity score of 0.5695 showcases the model's capability to correctly identify true positives, while the specificity score of 0.0180 indicates challenges in accurately identifying true negatives. These findings underscore the model's strengths and weaknesses, providing valuable insights for refining its design and enhancing its performance in real-world applications.

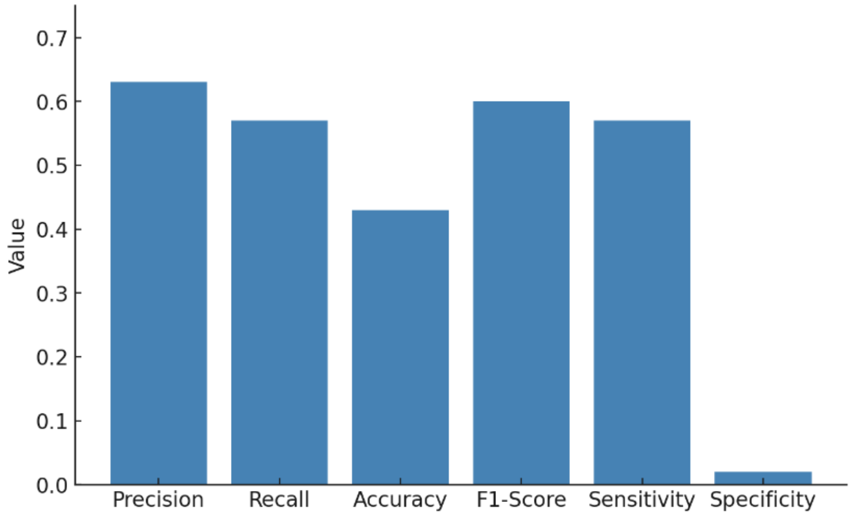
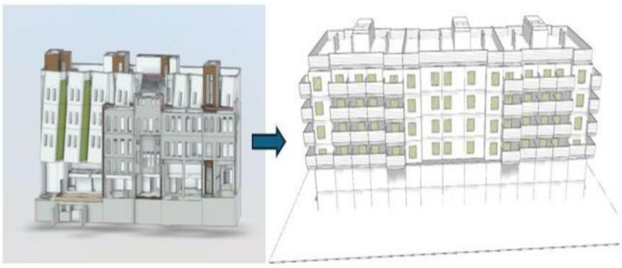


Fig. 4. Statistical Test Data of AI Model of Snowden Tower.

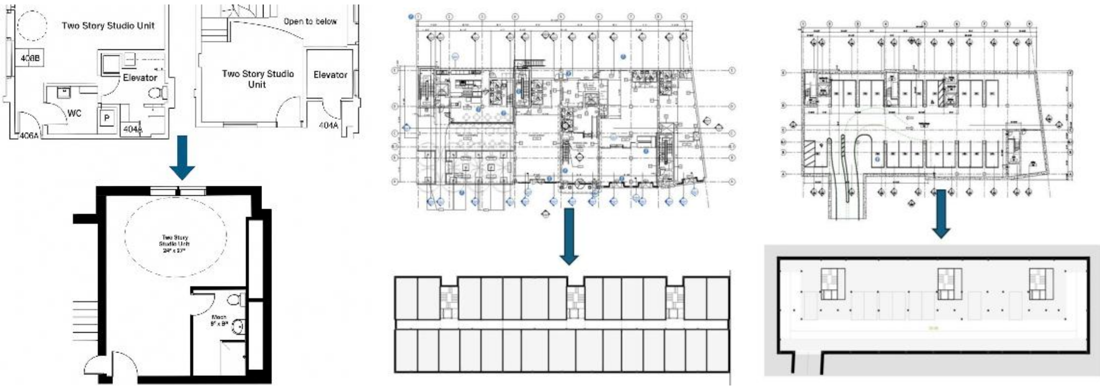
6.3. Qualitative analysis of AI model of Snowdon towers

This case offers a unique analysis since its challenging to design mixed used buildings due to accommodation of both commercial and residential considerations and balancing the user comfort of both the user cases. The conflicting design considerations have given the AI a great challenge, thus the importance is highlighted on training the AI on diverse data sets. AI was able to optimize 58% more of the plot space when compared to the real-world design but it was unable to create contours or parametric facades to enhance the look of the building. This led to the design being more homogenous and blander compared the real-world model as shown in Fig. 5.

AI's false positive value while emulating large number of commercial and restaurant units as mentioned in Fig. 5. This makes the spaces inefficient and unusable due to lack of functional allocations to the spaces. This will lead to budget overruns due to inconsistent wall portions being made without any functions. Furthermore, will lead dissatisfaction of the client and the building end users. Although, the AI was able to successfully emulate the studio apartments as indicated in table, it lacked the diversity in assigning the studios as shown in Fig. 5a. The real-world design had Duplex studios whereas the AI does not have any option to emulate or assign the specific design functionality. Due to this it limits the maximization of housing units. Duplexes are more efficient compared to the single unit counterpart due lower construction costs. Moreover, duplexes provide more freedom to create larger units, thus increasing the ability to adapt to the current housing demands. The datasets are specifically trained for residential buildings. As shown in Fig. 5a,b., the AI lacks to assign commercial parameters and has emulated the commercial spaces as per the number of studios in the floor above. This is due to a limitation in AI's training data due to distinct qualities in commercial and residential buildings like big lobbies, large floor plans with few or no partitions and specific functionality requirements for commercial buildings. Moreover, it is also due to the fact that residential design is more focused on aesthetic and the comfort level of users, thus AI prioritizing these aspects on mixed use spaces. Therefore, the design lacks the basic features of commercial buildings and may not satisfy the user requirements for mixed use buildings.



a) Architectural Façade



b) Studios & Duplex c) Commercial Units d) Parking Spaces

Fig. 5. Comparison of AI predictions with the Real Model.

Moreover, the AI lacks the understanding of utility spaces for MEP services. Since MEP considerations are the fundamentals in design considerations for having a building that is functional. Due to the lack of these basic considerations, it could hinder the centralization of the MEP equipment like HVAC systems, water heaters, electrical rooms, plumbing considerations. Moreover, it conflicts with the functionality of the spaces post-handover as it hinders the ability to maintain and repair the existing systems. Lastly AI lacks optimization of space as shown in the Fig. 5c,d. The example is for parking spaces for the commercial and residential units. AI was unable to optimize the space for parking spaces thus unable to minimize the parking footprint and the saved space can be optimized for other functional spaces. Due to less parking spaces optimized, designers may further allocate other spaces which may hamper the visual appeal of the building and its interior spaces. Furthermore, the demand for the building's parking space will not be suffice the user requirements if the original design of AI is considered.

7. Case Study 2 – residential building in Monterrey, Mexico

The selection of the Case Study 2 – Residential Building in Monterrey, Mexico" is considered to be important in the comparative analysis. Monterrey, as an urban center, encapsulates the complexities and nuances inherent in contemporary architectural projects. By focusing on a residential building within this dynamic context, we can discern the tangible impact of AI in addressing the multifaceted challenges of architectural design.

7.1. Interpretation of results

The analysis shown in Table reveals varying degrees of accuracy in the model's predictive capabilities across different sub-locations within the residential building in Monterrey, Mexico. Overall, the model demonstrates commendable accuracy in several areas, notably corridors, halls, and kitchens, where it achieves high True Positive Rates (TPR) exceeding 90%, indicating a robust ability to correctly classify instances. Conversely, certain sub-locations, such as bedrooms and bathrooms, exhibit lower TPR values, suggesting potential challenges in accurately identifying positive instances. Furthermore, the model maintains low False Positive Rates (FPR) across most sub-locations, indicating a minimal tendency to incorrectly classify negative instances. However, the presence of False Negatives (FN) in some areas implies instances where the model fails to identify positive cases, highlighting areas for potential improvement. Overall, while the model demonstrates promising accuracy in many sub-locations, addressing the discrepancies observed in areas with lower TPR values could enhance its predictive performance and overall efficacy in architectural design applications. The main outcomes are summarized below:

- **True Positive values:** AI was successfully able to emulate Corridor, Hall, Bedroom, Kitchen & patio/balcony, staircases, 35 doors and 14 windows.
- **False positive values:** AI emulated 3 additional Apartments per floor.
- **False Negative Values:** AI was unable to emulate 3 Bedroom apartments per floor but rather emulated 4 studio Apartments per floor

Table 5
Confusion Matrix results for Residential Building in Monterrey, Mexico.

Residential Building, Monterrey, Mexico Design Data										
Location	Sub-Location	TP	TN	FP	FN	CUM T	CUM F	TPR	FPR	AUC
Ground Floor	Corridor	1				1	0	0.982759	1	0
	Hall	1		3		2	0	0.965517	1	0
	Bedroom 1	1		3	1	3	0	0.948276	1	0
	Bedroom 2				1	3	0	0.948276	1	0
	Bedroom 3				1	3	0	0.948276	1	0
	Bathroom 1				1	3	0	0.948276	1	0
	Bathroom 2				1	3	0	0.948276	1	0
	Kitchen	1				4	0	0.931034	1	0
	Patio 1	1				5	0	0.913793	1	0
	Patio 2				1	5	0	0.913793	1	0
First Floor	Corridor	1				6	0	0.896552	1	0
	Hall	1		3		7	0	0.87931	1	0
	Bedroom 1	1		3	1	8	0	0.862069	1	0
	Bedroom 2				1	8	0	0.862069	1	0
	Bedroom 3				1	8	0	0.862069	1	0
	Bathroom 1				1	8	0	0.862069	1	0
	Bathroom 2				1	8	0	0.862069	1	0
	Kitchen	1				9	0	0.844828	1	0
	Patio 1	1				10	0	0.827586	1	0
	Patio 2				1	10	0	0.827586	1	0
Second Floor	Corridor	1				11	0	0.810345	1	0
	Hall	1		3		12	0	0.793103	1	0.793103
	Bedroom 1	1	0	3	1	13	1	0.775862	0	0
	Bedroom 2				1	13	1	0.775862	0	0
	Bedroom 3				1	13	1	0.775862	0	0
	Bathroom 1				1	13	1	0.775862	0	0
	Bathroom 2				1	13	1	0.775862	0	0
	Kitchen	1				14	1	0.758621	0	0
	Patio 1	1				15	1	0.741379	0	0
	Patio 2				1	15	1	0.741379	0	0
	Floors	3			1	18	1	0.689655	0	0
	Staircases	1				19	1	0.672414	0	0
	Doors	25			8	44	1	0.241379	0	0
	Windows	14				58	1	0	0	0
Total		58	0	18	27					0.793103

Confusion matrix results in Table offered a snapshot of a classification model's performance in predicting design features. It indicates that the model accurately predicts correct design features 58 times, while incorrectly predicting their presence 18 times. Interestingly, the model doesn't make any false negative predictions, implying it effectively identifies instances where design features are actually absent. However, the absence of incorrect design feature predictions might suggest overfitting or imited variability in the dataset. Overall, while the model demonstrates proficiency in identifying actual design features, efforts to reduce false positive predictions could enhance its reliability and applicability in real-world scenarios.

Table 6
Confusion Matrix of AI Model of Residential Building in Monterrey, Mexico.

Overall Prediction -Case of Residential Building in Monterrey, Mexico		
	Predicted Correct Design Features	Predicted Incorrect Design Features
Actual Correct Design Features	58	18
Actual Incorrect Design Features	27	0

The statistical test data in Fig. 6 provided insights into the performance of a classification model. The precision score of 0.76 indicates that when the model predicts a positive outcome, it is correct approximately 76% of the time. The recall score of 0.68 signifies the model's ability to correctly identify positive instances out of all actual positives. The accuracy score of 0.56 reflects the overall correctness of the model's predictions across both positive and negative instances. The F1-score, a harmonic mean of precision and recall, stands at 0.72, indicating a balanced performance between precision and recall. Sensitivity, equivalent to recall, underscores the model's ability to correctly identify positive instances, while specificity, at 0, suggests room for improvement in accurately identifying negative instances. Overall, while the model demonstrates decent precision and recall, its specificity indicates potential challenges in correctly identifying negative instances

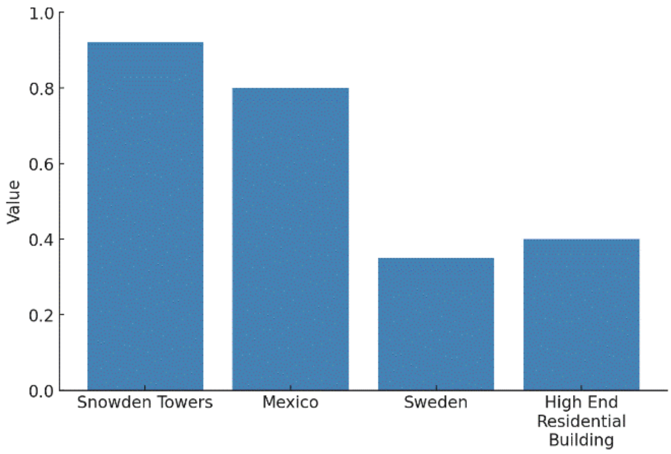
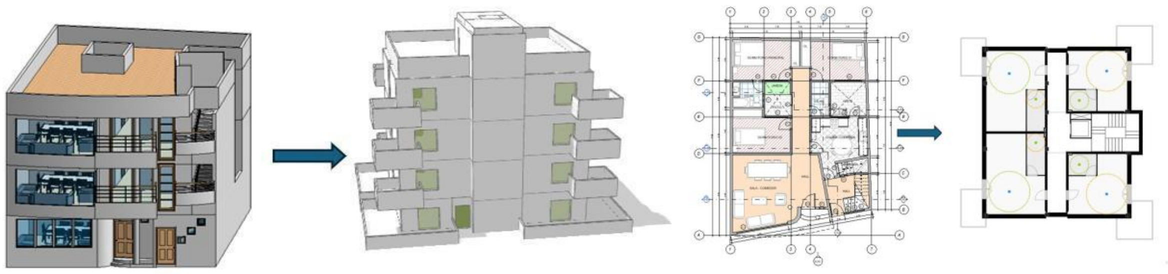


Fig. 6. Statistical Test Data of AI Model of Residential Building in Monterrey, Mexico.

7.2. Qualitative analysis of the AI model of residential building in Monterrey

Though the building is comparatively small in size, this case offers a perspective on Mexico, a third world country with distinct housing demands of single-family homes. but offers in-depth insight into. The AI emulated 4 studios per floor instead of a 3 Bedroom apartment per floor. As shown in Fig. 7a,b. Moreover, it lacked any manual customizability of the design to change it as per the real-world design. This could be due to AI's training dataset is on first world country's data. Due to urbanization, there is higher demand for Studios in developing countries due to social factors and better affordability [28].

Due to the scarcity of land and population density in third world country, diverse household size and family demographics the real-world design has a 3 bed room apartment per floor. Furthermore, cultural preferences and lifestyle choices of multiple families staying together in third world country as opposed to first world countries. Due to this it's important to understand the lack of AI's understanding on distinct cultural and geographic requirements. Moreover, the studios emulated are not functional and optimized for user requirement. The nuances between cultures and family requirements cannot be assessed by AI. The Design is homogenous and lacks the local community identity. Since studios are frequently thought of as short-term residences, the building's population is isolated, and its sense of community will not be felt by the residents. Furthermore, due to this design deficiency, the end user may be unable to find tenants and may face financial constraints. The plot in the real-world design was awkwardly shaped and not rectangular. The AI was able to optimize the space only in rectangular form. The angles and irregular curves challenge the AI's capability to optimize the building footprint. Land is scare in third world countries and optimization is key due to limited financial resources and strain on the existing infrastructure. Therefore, vertical communities like the real-world scenario are essential for affordable housing solutions.



a) Architectural Façade b) Floor Plan Layout

Fig. 7. Comparison of AI predictions with the Real Model.

8. Case study 3 – residential building in Sweden.

This case study on residential buildings in suburban Sweden serves as a crucial comparative analysis to evaluate how an AI model can effectively generate designs tailored to suburban landscapes. By examining the challenges inherent in traditional architectural design within this context, such as integrating buildings with the surrounding environment and catering to diverse socio-economic demographics, the study provides a robust framework for assessing the AI model's capacity to replicate and improve upon these designs. Through this comparative lens, we aim to interpret the AI model's potential to streamline design processes, enhance sustainability, and meet the unique demands of suburban living in Sweden.

8.1. Findings of the experiment

The analysis of the residential building in Sweden (Table) reveals that while some areas, like corridors and kitchens, demonstrate high levels of predictive accuracy with minimal false positives, others, notably bedrooms and bathrooms, exhibit inconsistencies with occasional false positive predictions. This suggests that the model may struggle to accurately identify certain design features in these specific areas. To improve overall performance, targeted adjustments and enhancements may be necessary, focusing on refining the model's ability to correctly classify instances in these challenging sub-locations. The main outcomes of the analysis are summarized below

- **True Positive values:** AI was successfully able to emulate Corridor, Hall, Bedroom, Kitchen , staircases, doors, windows and balconies.
- **False positive values:** AI emulated 3 additional Studio Apartments per floor.

- **False Negative Values:** AI was unable to emulate 2 Bedroom apartments per floor but rather emulated 4 studio Apartments per floor.

Table 7.
Confusion Matrix Results for Residential Building in Sweden.

Residential Building in Sweden										
Location	Sub-location	TP	TN	FP	FN	CUM T	CUM F	TPR	FPR	AUC
Ground Floor	Corridor	1				1	0	0.971429	1	0
	Living & Dining	1				2	0	0.942857	1	0
	Bedroom 1	1				3	0	0.914286	1	0
	Bedroom 2			15	1	3	0	0.914286	1	0
	Bathroom 1	1				4	0	0.885714	1	0
	Kitchen	1				5	0	0.857143	1	0
First Floor	Corridor	1				6	0	0.828571	1	0
	Living & Dining	1				7	0	0.8	1	0
	Bedroom 1	1				8	0	0.771429	1	0
	Bedroom 2			15	1	8	0	0.771429	1	0
	Bathroom 1	1				9	0	0.742857	1	0
	Kitchen	1				10	0	0.714286	1	0
Second Floor	Corridor	1				11	0	0.685714	1	0
	Living & Dining	1				12	0	0.657143	1	0
	Bedroom 1	1				13	0	0.628571	1	0
	Bedroom 2			15	1	13	0	0.628571	1	0
	Bathroom 1	1				14	0	0.6	1	0
	Kitchen	1				15	0	0.571429	1	0
Third Floor	Corridor	1				16	0	0.542857	1	0
	Living & Dining	1				17	0	0.514286	1	0
	Bedroom 1	1				18	0	0.485714	1	0
	Bedroom 2			15	1	18	0	0.485714	1	0
	Bathroom 1	1				19	0	0.457143	1	0
	Kitchen	1				20	0	0.428571	1	0
Fourth Floor	Corridor	1				21	0	0.4	1	0
	Living & Dining	1				22	0	0.371429	1	0
	Bedroom 1	1				23	0	0.342857	1	0.342857
	Bedroom 2		1	15	1	23	1	0.342857	0	0
	Bathroom 1	1				24	1	0.314286	0	0
	Kitchen	1				25	1	0.285714	0	0
General	windows	6		13		31	1	0.114286	0	0
General	Doors					31	1	0.114286	0	0
General	Balcony	4		16		35	1	0	0	0
	Total	35	0	104	5					0.342857

In this dataset, we observe that as the number of False Negatives (FN) increases in certain sub-locations, the True Positive Rate (TPR) tends to decrease. For instance, on the Second Floor, in the "Bedroom 2" sub-location, where there are 15 False Negatives, the TPR is 0.628571. Similarly, on the Fourth Floor, in the same "Bedroom 2" sub-location, where there is 1 False Negative, the TPR improves to 1.0. The provided confusion matrix illustrates the performance of a classification model in predicting design features. It reveals a significant imbalance between the number of actual design features present and those correctly predicted by the model (Fig. 8). While there are 35 instances where the model accurately identifies design features as correct, the high number of False Positives (104) indicates a tendency to over-predict the presence of design features. Additionally, the absence of any True Negatives suggests that the model struggles to correctly identify instances where design features are actually absent. Furthermore, the presence of False Negatives (5) highlights instances where the model fails to recognize actual design features.

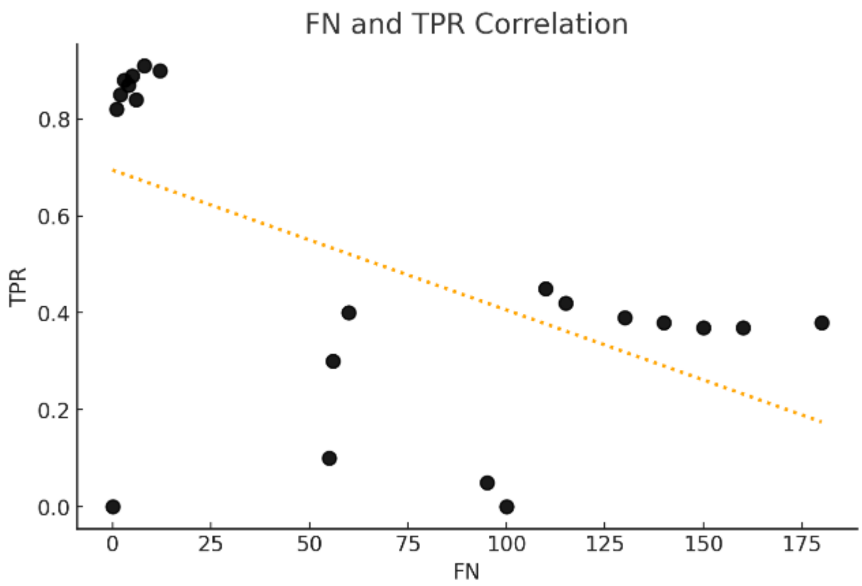


Fig. 8. FN and TPR Correlation Analysis for Residential Building in Sweden.

According to the prediction results presented in Table , the model demonstrates a high recall rate, accurately capturing most actual correct design features, its precision is significantly lower, indicating a propensity for false positive predictions. The absence of true negatives underscores challenges in correctly identifying incorrect design features. These findings suggest a need for refinement in the model's predictive accuracy, particularly in reducing false positives and improving specificity. Overall, while the model shows promise in identifying actual correct design features, further optimization is necessary to enhance its overall performance and reliability in predicting residential building designs in Sweden.

Table 8

Confusion Matrix of AI Model of Residential Building in Sweden.

Confusion Matrix - Residential Building in Sweden		
	Predicted Correct Design Features	Predicted Incorrect Design Features
Actual Correct Design Features	35	104
Actual Incorrect Design Features	5	0

The statistical analysis in Fig. 9 provides valuable insights into the classification model's predictive performance. With a precision score of 0.25, the model accurately predicts only about a quarter of the positive outcomes, indicating a relatively low precision level. However, the high recall score of 0.875 suggests that the model effectively identifies a large proportion of actual positive instances. The overall accuracy score of 0.243 highlights areas where the model's predictions could be improved. Balancing precision and recall, the F1-Score stands at 0.39, indicating a moderate level of performance. Additionally, the sensitivity score of 0.875 underscores the model's proficiency in identifying positive instances, while the specificity score of 0 reveals challenges in correctly identifying negative instances.

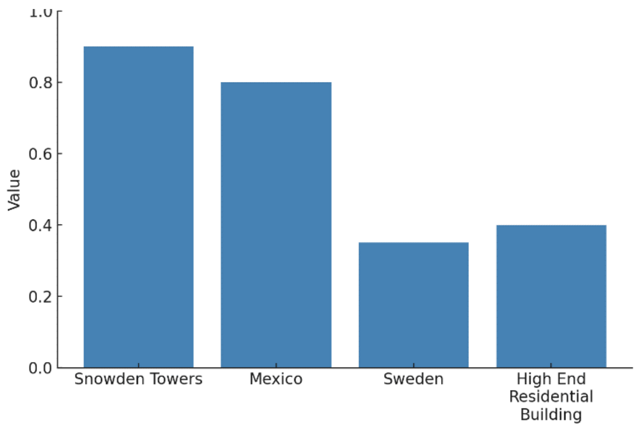
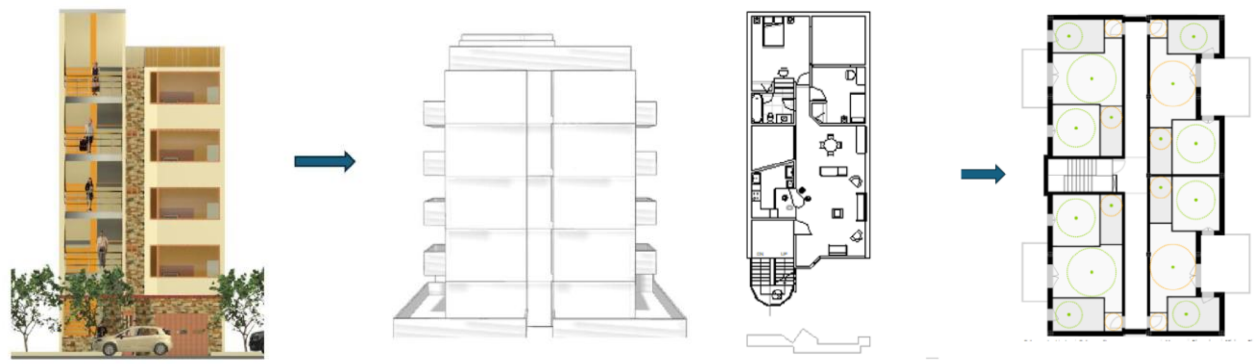


Fig. 9. Statistical Test Data of AI Model of Residential Building in Sweden.

8.2. Qualitative analysis

This case offers a perspective on Residential Building in Suburban area of Sweden. It gives an overview on challenges in designing buildings in a suburban landscape. The AI is unable of generation of buildings with width less than 12m as this case had a limitation on the plot width. Thus, the plot sizes more than 12 metres may have an effect on the hosing density due to the increase in plot sizes. Furthermore, the increased plot size may have wider effect on the other community infrastructure like parks, schools and other shared spaces. Moreover, it may increase cost of developing infrastructure due to more electrical load requirements and other utilities. In developed areas larger building footprints are more common when compared to suburban areas [29]. The width restriction may negate the footprint requirements that are much narrower in suburban areas. Moreover, this would increase the housing price due to increase in plot sizes, therefore making the smaller and affordable housing in suburban areas more expensive. Furthermore, the lack of diverse options in AI's generated design may limit the housing options. Smaller single-family homes, duplexes or townhouses may not be feasible. The increased cost may isolate low-income individuals or families. The socio-economic diversity will reduce and may further increase the suburban sprawl. This can also have negative impacts on the environment due to an increase in suburban sprawl. This case also had similar issues of emulating one single family apartment per floor but emulated 4 studios like in the case of Mexico as shown in Fig. 10.



a) Architectural Façade b) Floor Plan Layout

Fig. 10. Comparison of AI predictions with the Real Model.

9. Case study 4 – high end residential building

Case Study 4 focuses on a high-end residential building of unknown location, providing a valuable opportunity to assess AI's capabilities in handling large-scale, sophisticated residential projects. High-end residential buildings typically feature intricate designs, luxurious amenities, and meticulous attention to detail, making them challenging for traditional design processes

9.1. Interpretation of Results

Table presents a comprehensive overview of the performance of an artificial intelligence (AI) model in various locations and sub-locations within a construction project. The model's predictive accuracy is evaluated through metrics such as True Positive Rate (TPR), False Positive Rate (FPR), and the Area Under the Curve (AUC) on Receiver Operating Characteristic (ROC) curves. In the Basement and Private Storage Rooms, the model's performance is not fully assessed due to missing True Negative (TN) and False Negative (FN) values. However, for Basement Parking, a high False Positive Rate (FPR) of 1 indicates that all actual negatives are incorrectly classified as positives, suggesting significant room for improvement. Moving to the different floors of the building, from Ground Floor to Eighth Floor, where the sub-locations consist of 1 to 3 Bedroom Apartments, the model's True Positive Rate (TPR) gradually decreases. This decline suggests a diminishing ability to correctly identify positive cases as we move up the floors. Furthermore, a consistent False Positive Rate (FPR) of 1 across all floors highlights a critical flaw in the model's accuracy, as it is incorrectly classifying all actual negatives as positives. Overall, the Area Under the Curve (AUC) remains consistent at 0.391 for all sub-locations, indicating a poor performance in distinguishing between positive and negative cases across various thresholds. Key findings are as follow:

True Positive values: AI was successfully able to emulate 31 parking spaces. Moreover, it successfully emulated the 4 lifts and staircases and most importantly was able to emulate the machine room for the lift. Furthermore, the AI was able to emulate space for 1 bedroom and 3-bedroom Apartments.

False positive values: AI emulated additional studios, 2-bedroom, 3-bedroom apartments that are not there in the original design of the real-world scenario.

False Negative Values: AI was unable to optimize the parking space. Also, AI was unable to make amenity spaces like Swimming pool, Auditorium, staff lounges etc.

Table 9
Confusion Matrix Results for High End Residential Building.

Location	Sub-location	TP	TN	FP	FN	CUM T	CUM F	TPR	FPR	AUC
Basement	Parking	31	-	28	-	31	0	0.756	1	0
	Private Storage Rooms	48	-	-	-	79	0	-	-	-
Ground Floor	Reception to Steam Room	29	3	133	2	91	16	0.935	0.978	0.046
First Floor	1 to 3 Bedroom Apartment	77	0	151	91	228	91	0.458	1	0.391
Second Floor	1 to 3 Bedroom Apartment	88	0	145	110	235	236	0.445	1	0.391
Third Floor	1 to 3 Bedroom Apartment	85	0	143	117	238	379	0.421	1	0.391
Fourth Floor	1 to 3 Bedroom Apartment	86	0	141	137	224	520	0.385	1	0.391
Fifth Floor	1 to 3 Bedroom Apartment	91	0	131	147	238	667	0.383	1	0.391
Sixth Floor	1 to 3 Bedroom Apartment	96	0	125	158	238	792	0.378	1	0.391
Seventh Floor	1 to 3 Bedroom Apartment	102	0	117	171	273	913	0.374	1	0.391
Eighth Floor	1 to 3 Bedroom Apartment	108	0	109	184	292	1022	0.370	1	0.391
Total	All Sub-locations	127	0	151	91	-	-	-	-	0.392

In the given dataset, there appears to be a strong negative correlation between the False Negative (FN) rate and the True Positive Rate (TPR) across different sub-locations within the building (Fig. 11). This negative correlation suggests that as the FN rate decreases, indicating a lower number of missed positive cases, the TPR tends to increase, reflecting a higher proportion of correctly identified positive cases. This relationship is essential in assessing the effectiveness of the AI model in accurately detecting positive instances, such as specific features or anomalies within the building structure. A lower FN rate signifies better performance in capturing true positive instances, which is crucial for ensuring the reliability and effectiveness of the AI model in identifying relevant features within the building environment. Therefore, maintaining a low FN rate while maximizing the TPR is critical for enhancing the model's overall performance and utility in real-world applications.

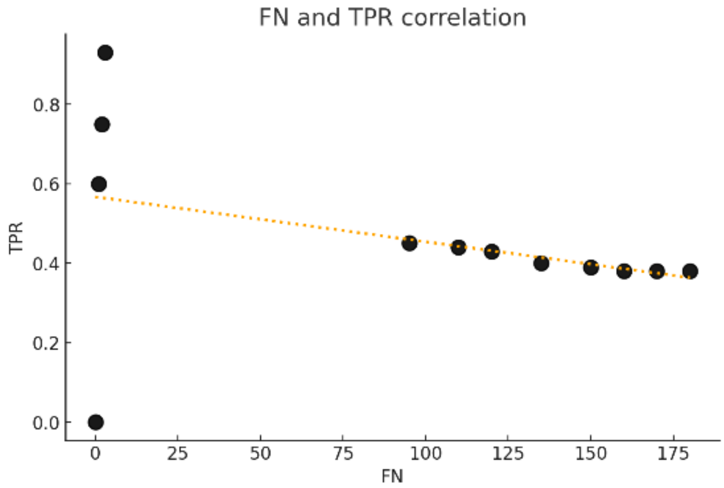


Fig. 11. FN and TPR Correlation Analysis for High End Residential Building.

The analysis of the provided data reveals an inverse correlation between FN (False Negatives) and CUM T (Cumulative True Positives) (Fig. 12). As FN increases across various sub-locations, CUM T decreases. This suggests that when the model fails to identify positive cases (FN), there is a subsequent decrease in the cumulative count of correctly identified positive cases (CUM T). The implication of this analysis is clear: reducing false negatives is crucial for maintaining the model's accuracy and effectiveness.

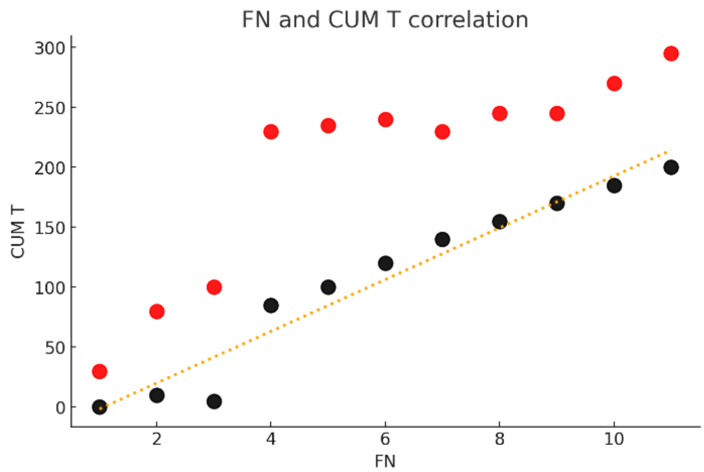


Fig. 12. FN and CUM T Correlation Analysis for High End Residential Building.

The model struggles with identifying incorrect design features (0 true negatives) as shown in Table 10, which impacts its overall accuracy and recall. Precision is relatively better, indicating a fair proportion of the predicted correct design features are indeed correct. However, the high number of false negatives and false positives indicates significant room for improvement in both sensitivity and specificity.

Table 10
Confusion Matrix of AI Model case of High End Residential Building.

Confusion Matrix		
	Predicted Correct Design Features	Predicted Incorrect Design Features
Actual Correct Design Features	127	151
Actual Incorrect Design Features	91	0

The statistical test data presented in Fig. 13 reveals several insights into the model's performance. With a precision of approximately 45.7%, the model correctly identifies less than half of the predicted correct design features. The recall and sensitivity are higher at around 58.3%, indicating the model is more effective at capturing actual correct design features, though it still misses a significant number. The accuracy is notably low at 34.4%, reflecting poor overall performance in distinguishing correct and incorrect design features. The F1-score, combining precision and recall, stands at 51.2%, underscoring a balance between these metrics but still indicating mediocrity. The specificity is 0, revealing the model's complete inability to correctly identify incorrect design features, which is a critical flaw needing rectification.

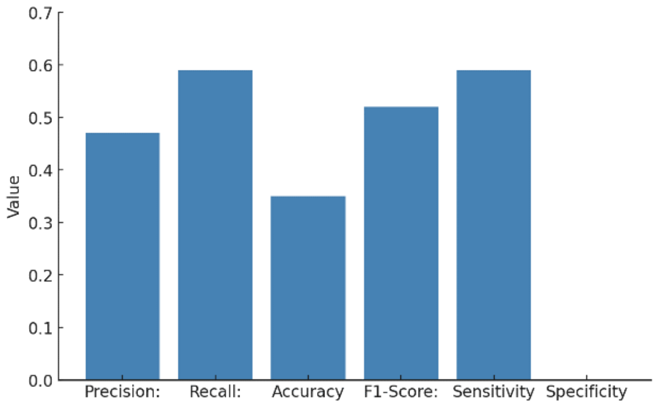


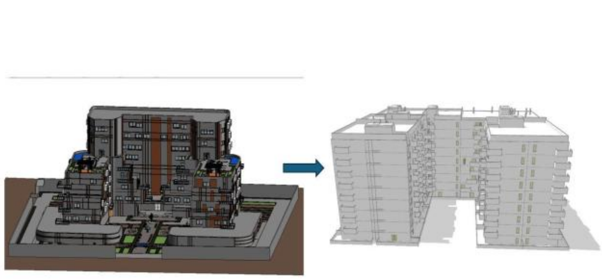
Fig. 13. Statistical Test Data of AI Model of Residential Building in Sweden.

9.2. Qualitative analysis

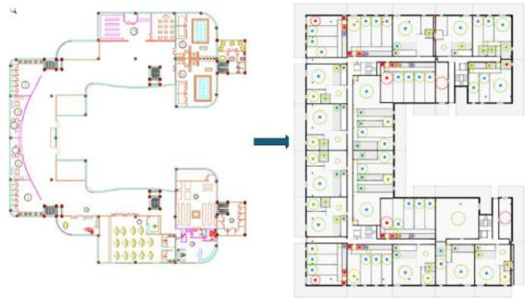
This case offers an analysis of a high-end residential building. It offers a unique perspective on Ai's ability to emulate large high end residential projects. AI was unable to emulate Auditoriums and other similar amenities. Certain events and large- scale meetings, building association meetings cannot be performed. The sense of community and fostering collaboration is important in big residential buildings. Furthermore, such amenities play a vital role in increasing the property value and have a possible revenue generating attribute for building associations. Moreover, AI was unable to emulate Swimming pool, spa and sauna. These facilities are high in demand and are basic amenities in high end buildings. This would result in reduced user satisfaction. This would create a drawback and the building may lose a competitive edge to other buildings which provides these amenities. The property value may decrease since basic residential buildings also provide swimming pools, spa and saunas.

Amenities provide several useful benefits in addition to the ones that occupants can enjoy right away. The residents can have a quiet place to study or work remotely. Also having a fully functional business center or library on the premises might be very handy. Residents' convenience is increased by delivery rooms or concierge services, which provide simple and safe parcel delivery management. These commonplace amenities make living in a high-end residential building more convenient. Furthermore, the AI emulated design lacked staff lounges. In high end residential buildings these are important for staff wellbeing. For high-end experience such facilities are required for staff to have a secure space and maintain a professional conduct when interacting with the residents of the building. Though AI gave diverse housing options by emulating studios, 1 bedroom, 2-bedroom, 3 bedroom and 4-bedroom Apartments opposed to only 1 bedroom and 3-bedroom apartments in the original design. Due to this AI has faced efficiency challenges to optimize the floor space as highlighted in the figure. The configuration has made the units less spacious with impractical layouts. Though the room size may satisfy the minimum area dimensions, the sizes i.e. the length and width are impractical for furniture placements.

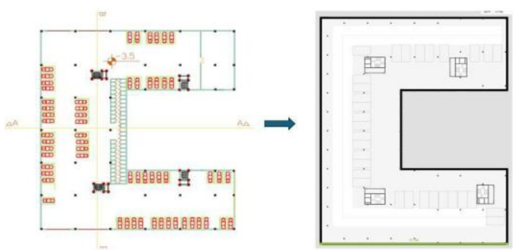
As per Fig. 14 the building façade made by AI lacks innovative design intent. The design is homogenous like the other case studios emulated. Thus, the visual impact and identity is loss of the high-end residential building. The aesthetics view is bland, and it fails to create the sense of luxuriousness for users who will buy the property. Due to this the property may not be marketable to high end clients and the building developer may lose or not have any brand identity. Moreover, the lack of innovation may cause environmental issues as newer facades options are more environmentally friendly. Therefore, overall, this could lower the satisfaction of the end user of the residential building designed AI.



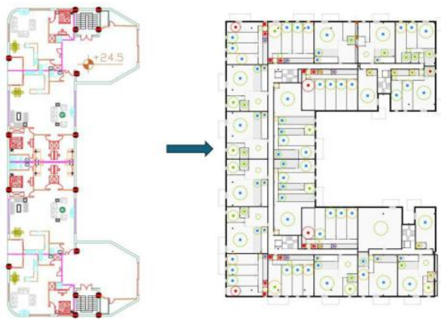
a) Architectural Façade



b) Floor Plans Ground



c) Floor Plans Upper levels



d) Floor Plans basement

Fig. 14. Comparison of AI predictions with the Real Model.

10. Overall AI prediction analysis (ROC)

Based on the analysis of the four case studies on design iteration generated by the AI ARCHITEChTURE platform, the graph in Fig. 15 projects a visual comparison of the accuracy of the four AI generated models. In the above graph, the points above the red line are attributes with high True positive rate i.e. more accurate whereas the points below the red line are attributes with high false positive rates. i.e. less accurate. A closer analysis of the test data shows that when data is unbalanced, meaning that one category may occur more frequently than the other, the graph of such a model can nevertheless achieve high accuracy by primarily predicting the majority category. On the other hand, an imbalanced data set solely based on the ROC Metrics can be misleading as the graphs cannot fully capture the nuances of architectural design.

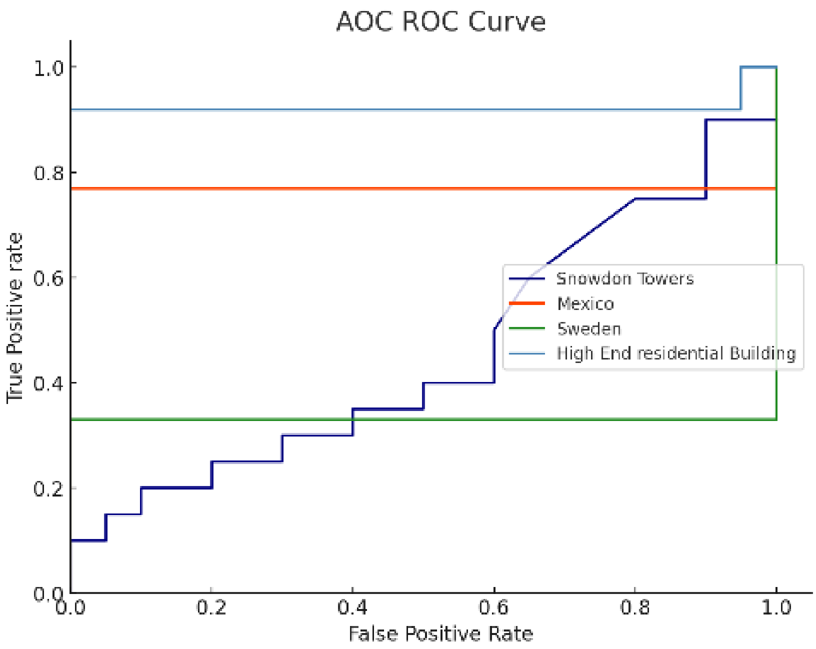


Fig. 15. ROC curve (Receiver Operating Characteristic Curve).

In the case of Mexico, the AUC score suggests the second highest accuracy, but the model was able to predict the doors and windows more accurately than the other AI generated models, therefore elevating the accuracy score. But as stated previously, this case had highly inaccurate predictions with the lack of manual customizations. Due to this, in the real-world scenario the architect may completely reject the design by AI as it does not satisfy the user requirements.

As per the AUC data shown in Table 11, the case of Snowden towers shows the highest accuracy and based on the emulating criteria, the design model may satisfy the user requirements. However, it may lack functionality as mentioned in the analysis before. In the case of the AI model of the residential building of Sweden, the test results had the lowest AUC due to majority of false positive attributes. Furthermore, the same trend is seen in the case of the AI model of High- end residential building from an unspecified location. This shows that AI design iterations have not yet reached a reliable state free from human involvement and independent of Architectural Design Engineers. The AI design software need to enhance its framework of Supervised, Unsupervised and Reinforcement learning along with the development of Convolutional Neural Network (CNN) and Natural Language Processing Systems (NLP) to ensure optimum utility of AI aided design software in the construction industry.

Table 11
AUC (Area under the curve).

	AUC
Snowden Towers	0.919355
Mexico	0.793103
Sweden	0.342857
High End Residential Building	0.391615

11. Conclusion

11.1. Key findings

This study highlights the transformative potential of AI-integrated platforms, with a particular focus on the ARCHITEChTURE platform, in enhancing architectural design through generative and predictive capabilities. Quantitative evaluations across diverse case studies demonstrated that AI models can substantially improve spatial optimization, automate routine design components, and facilitate real-time regulatory compliance tracking. Performance metrics such as True Positive Rate (TPR), False Positive Rate (FPR), and Area Under the Curve (AUC) confirmed strong predictive accuracy, especially in optimizing space efficiency and parametric configurations. Complementing these results, qualitative analysis revealed a reduction in design iteration time and resource consumption, reinforcing the operational value of AI in early design stages.

11.2. Limitations of AI software in architectural design

Despite these advances, the current generation of AI design tools exhibits several limitations. First, they struggle to account for cultural and geographic specificity in design, particularly in integrating utility spaces, handling irregular site geometries, and managing mixed-use spatial logic. Second, AI-generated outputs tend to be homogeneous, lacking variation and sensitivity to diverse programmatic needs. Third, existing models fail to effectively interpret complex architectural semantics such as duplex arrangements, hybrid commercial-residential zoning, and user-centric design variations. Fourth, limitations in natural language processing restrict the accurate translation of client requirements into spatial outputs. Finally, most models rely on training datasets that are not sufficiently diverse, limiting generalizability and adaptability to different design contexts.

11.3. Future research directions

Future investigations should aim to enhance the semantic and contextual depth of AI models by incorporating more diverse, representative datasets and improving their understanding of architectural language. Research should also prioritize multi-objective optimization frameworks that address varying climatic, regulatory, and social conditions. Equally important is the integration of explainable AI and user-centered design principles to foster transparency and trust. Collaboration among architects, engineers, and AI developers will be essential to bridge disciplinary gaps and align algorithmic generation with real-world complexity.

Furthermore, validation against real-world projects and stakeholder feedback loops will support the transition of AI platforms from experimental prototypes to practical design tools.

CRediT authorship contribution statement

Kamal Jaafar: Conceptualization; Formal analysis; Methodology; Resources; Supervision; Visualization; Writing – original draft; Writing – review & editing.

Karol Sikora: Conceptualization; Investigation; Methodology; Validation.

Sana Amir: Data curation; Investigation; Software; Validation; Writing – original draft.

Lina Gharaibeh: Writing – review & editing.

Mohamad Koona: Project administration.

Funding

This research received no external funding.

Conflicts of interest

The authors declare no conflict of interest.

Data availability statement

All data generated and analyzed during this study are included in this paper. The results, findings, and supporting information are available within the text and supplementary materials, ensuring transparency and reproducibility of the research. No additional data is available outside of what is presented in this manuscript.

References

[1] Li Y, Xu J, Wang H. From sketch to floorplan: Leveraging AI for conceptual architectural design 2024.

[2] Kakooee R, Dillenburger B. Illuminating Spaces: Deep Reinforcement Learning and Laser-Wall Partitioning for Architectural Layout Generation. ArXiv Prepr ArXiv250204407 2025. https://doi.org/10.48550/arXiv.2502.04407.

[3] of Architects TAI. Just 6% of architects use AI regularly. Here’s why 2025.

[4] Seppänen O. Empirical research on the success of production control in building construction projects 2009;1:188.

[5] Koo HJ, O’Connor JT. Complexity Analysis of Design Deliverable Defects on Building Projects. J Manag Eng 2021;37:4021014. https://doi.org/10.1061/(ASCE)ME.1943-5479.0000897.

[6] Li Y, Antwi-Afari MF, Anwer S, Mehmood I, Umer W, Mohandes SR, et al. Artificial Intelligence in Net-Zero Carbon Emissions for Sustainable Building Projects: A Systematic Literature and Science Mapping Review. Buildings 2024;14:2752. https://doi.org/10.3390/buildings14092752.

[7] Mehraban MH, Alnaser AA, Sepasgozar SME. Building Information Modeling and AI Algorithms for Optimizing Energy Performance in Hot Climates: A Comparative Study of Riyadh and Dubai. Buildings 2024;14:2748. https://doi.org/10.3390/buildings14092748.

[8] Kazemi N, Ostwald M, Abanda FH. Artificial intelligence for tall building design: A parametric and machine learning approach for early-stage planning. Autom Constr 2024;156:105204.

[9] Ploennigs J, Berger M. AI art in architecture. AI Civ Eng 2023;2:8. https://doi.org/10.1007/s43503-023-00018-y.

[10] Zhang Z, Fort JM, Giménez Mateu L. Exploringthe Potential of Artificial Intelligence as a Tool for Architectural Design: A Perception Study Using Gaudí’sWorks. Buildings 2023;13:1863. https://doi.org/10.3390/buildings13071863.

[11] Zhang X, Fort R, Giménez Mateu J. Human vs AI: Evaluating architectural design aesthetics through Gaudí-inspired experiments. Des Stud 2023;84:102176.

[12] Berger JP and K. AI art in architecture: Exploring diffusion models for generative conceptual design. Front Built Env 2022;9:104440.

[13] Baduge SK, Thilakarathna S, Perera JS, Arashpour M, Sharafi P, Teodosio B, et al. Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. Autom Constr 2022;141:104440. https://doi.org/10.1016/j.autcon.2022.104440.

[14] Horvath A-S, Pouliou P. AI for conceptual architecture: Reflections on designing with text-to-text, text-to-image, and image-to-image generators. Front Archit Res 2024;13:593–612. https://doi.org/10.1016/j.foar.2024.02.006.

[15] Youns AM, Grchev K. A Historical and Critical Assessment of Parametricism as an Architectural Style in the 21st Century. Buildings 2024;14:2656. https://doi.org/10.3390/buildings14092656.

[16] Hariri-Ardebili MA, Mahdavi G, Nuss LK, Lall U. The role of artificial intelligence and digital technologies in dam engineering: Narrative review and outlook. Eng Appl Artif Intell 2023;126:106813. https://doi.org/10.1016/j.engappai.2023.106813.

[17] Li S, Wang R. Can Smart City Construction Promote Urban Green and High-Quality Development?—Validation Analysis from 156 Cities in China. Buildings 2024;14:2500. https://doi.org/10.3390/buildings14082500.

[18] Wang L, Li B, Zhao X, He J. A water spraying box and its improving effect on the thermal environment around multi-story residential buildings to reduce released heat from air-conditioning units. Energy Build 2023;297:113484. https://doi.org/10.1016/j.enbuild.2023.113484.

[19] Kazemi P, Ghisi A, Entezami A. Artificial Intelligence-Powered Computational Strategies in Selecting and Augmenting Data for Early Design of Tall Buildings with Outer Diagrids. Buildings 2024;14:1118. https://doi.org/10.3390/buildings14041118.

[20] He W, Chen M. Advancing Urban Life: A Systematic Review of Emerging Technologies and Artificial Intelligence in Urban Design and Planning. Buildings 2024;14:835. https://doi.org/10.3390/buildings14030835.

[21] Tyrtaiou M, Elenas A, Andreadis I, Vasiliadis L. Hilbert-Huang Transform-Based Seismic Intensity Parameters for Performance-Based Design of RC-Framed Structures. Buildings 2022;12:1301. https://doi.org/10.3390/buildings12091301.

[22] Ting KM. Confusion Matrix. Encycl. Mach. Learn., Boston, MA: Springer US; 2011, p. 209–209. https://doi.org/10.1007/978-0-387-30164-8_157.

[23] Fawcett T. An introduction to ROC analysis. Pattern Recognit Lett 2006;27:861–74. https://doi.org/10.1016/j.patrec.2005.10.010.

[24] Streiner DL, Cairney J. What's under the ROC? An Introduction to Receiver Operating Characteristics Curves. Can J Psychiatry 2007;52:121–8. https://doi.org/10.1177/070674370705200210.

[25] Angelov PP. Handbook On Computer Learning And Intelligence (In 2 Volumes). World Scientific; 2022.

[26] Admin. The International Building Code. Int Code Counc 2015.

[27] Carmona M. Public places urban spaces: The dimensions of urban design. Routledge; 2021.

[28] Wetzstein S. The global urban housing affordability crisis. Urban Stud 2017;54:3159–77. https://doi.org/10.1177/0042098017711649.

[29] Brade I, Herfert G, Wiest K. Recent trends and future prospects of socio-spatial differentiation in urban regions of Central and Eastern Europe: A lull before the storm? Cities 2009;26:233–44. https://doi.org/10.1016/j.cities.2009.05.001.