

Article

Integrating Generative Design and Artificial Intelligence for Optimized Energy-Efficient Composite Facades in Next-Generation Smart Buildings

Mohammad Q. Al-Jamal ¹, Ayoub Alsarhan ^{2,3}, Mahmoud AlJamal ⁴, Qasim Aljamal ⁵,
Bashar S. Khassawneh ^{6,*}, Amina Salhi ⁷ and Hanan Hayat ⁸

¹ Department of Renewable Energy, Technical Faculty, Jadara University, P.O. Box 733, Irbid 21110, Jordan; m.aljamal@jadara.edu.jo

² Department of Information Technology, Faculty of Prince Al-Hussien bin Abdullah, The Hashemite University, Zarqa 13133, Jordan

³ Department of Data Science and Artificial Intelligence, Faculty of Information Technology, Al-Ahliyya Amman University, Amman 19111, Jordan

⁴ Department of Cybersecurity, Science and Information Technology, Irbid National University, Irbid 21110, Jordan

⁵ Department of Civil Engineering, Faculty of Architecture and Civil Engineering, Technical University Dortmund, 44227 Dortmund, Germany

⁶ Department of Computer Science, College of Information Technology, Amman Arab University, Amman 11953, Jordan

⁷ Department of Information Systems, College of Computer and Information Sciences, Princess Nourah bint Abdulrahman University, P.O. Box 84428, Riyadh 11671, Saudi Arabia

⁸ Software Engineering Department, College of Computing, Umm Al-Qura University, Mecca 21955, Saudi Arabia

* Correspondence: b.khassawneh@aau.edu.jo

Abstract

The pursuit of energy efficiency and sustainability in the built environment has placed façade systems at the forefront of innovation in architectural design. This study proposes an integrated framework that combines generative design techniques with artificial intelligence (AI) to optimize composite façade configurations for next-generation smart buildings. Using parametric modeling, a wide design space of façade geometries and material compositions was generated, capturing trade-offs between thermal performance, daylight, structural strength, and aesthetic variability. Artificial intelligence algorithms, particularly machine learning models, are trained on simulation-derived performance datasets to rapidly predict key indicators such as energy consumption, thermal transmittance (U-value) and solar heat gain coefficients. The proposed approach achieved a predictive accuracy of 99.85%, enabling efficient exploration of optimal solutions across high-dimensional design alternatives. A multi-objective optimization strategy was further implemented to balance energy efficiency with structural and aesthetic constraints, producing façade configurations that outperform conventional designs. The findings demonstrate that integrating generative design with AI-based prediction not only accelerates the façade design process but also provides actionable pathways toward net-zero energy buildings. This research highlights the transformative potential of AI-driven generative workflows in advancing sustainable architecture and delivering intelligent, adaptive and performance-oriented façades for future urban environments.

Keywords: generative design; artificial intelligence; composite facades; smart buildings; energy efficiency; machine learning; predictive modeling



Academic Editors: Jieying Jane Zhang,
Ahmad Jrade and Farzad Jalaei

Received: 19 January 2026

Revised: 11 February 2026

Accepted: 26 February 2026

Published: 1 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The construction industry was one of the biggest users of energy and producers of carbon dioxide in the world. It uses almost 40% of all the energy used in the world. In this field, the building envelope, especially façade systems, was very important for figuring out how energy-efficient, comfortable, and environmentally friendly building was. Composite façades, which combine advanced materials with design strategies that serve more than one purpose, have become a promising way to lower energy use while keeping architectural flexibility and visual appeal [1]. However, designing these kinds of systems is still a difficult, multi-faceted task that needs to take into account the properties of the materials, the conditions of the environment, the performance of the structure, and the comfort of the people who will be using it.

Traditional façade design methods rely heavily on rule-based parametric modeling and iterative simulations [2]. While effective, these approaches were time-consuming, computationally expensive, and often limited in exploring the full design space. In recent years, advances in artificial intelligence (AI) and generative design have created new opportunities for automating and optimizing façade configurations. Generative design leverages algorithms such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) to explore a wide range of potential solutions, while predictive AI models (Random Forests, Gradient Boosting and Artificial Neural Networks) evaluate performance outcomes such as energy consumption, daylight distribution and thermal comfort [3]. The integration of these methods promises not only faster design cycles but also the discovery of innovative façade systems that would be infeasible to identify using conventional approaches [4].

Existing research has demonstrated the value of AI in building energy prediction and optimization; however, most studies focus on isolated tasks such as load forecasting or material thermal modeling [5]. Few frameworks attempt to integrate generative design with AI-based predictive evaluation in a closed-loop system that directly links design exploration to energy performance optimization. Furthermore, the availability of large-scale datasets such as BuildingsBench, which combines simulated and real-world building data, provides a unique opportunity to train robust models that generalize across diverse building archetypes and climate zones [6]. Harnessing such datasets in conjunction with AI-driven generative methods has the potential to revolutionize the design of energy-efficient composite façades [7]. Recent advances demonstrate that artificial intelligence is increasingly transforming architectural design workflows, particularly in energy-efficient building envelope renovation, adaptive façade optimization, and data-driven thermal performance analysis. Recent systematic literature analyses highlight the transition of AI from a purely auxiliary design tool toward collaborative and autonomous decision-support roles in envelope and façade optimization tasks.

The aim of this study is to develop and evaluate a hybrid framework that unifies generative design and AI-driven prediction to optimize the performance of composite façade systems in next-generation smart buildings. Specifically, this research integrates VAEs and GANs for façade design generation machine learning models for predictive performance evaluation, and Bayesian optimization for hyperparameter tuning. In this work, the term “composite façades” refers to façade assemblies characterized by multi-material configurations considered within the adopted dataset and evaluation pipeline; specific material families (fiber-reinforced composites) are treated as illustrative examples rather than the exclusive scope of the study. The framework was tested on the BuildingsBench dataset to ensure scalability and real-world relevance. By systematically combining generative exploration, predictive modeling, and optimization, this study contributes to the advancement of sustainable architecture by providing architects and engineers with an

intelligent decision-support system for designing façades that maximize energy efficiency, resilience and sustainability.

2. Literature Review

Albukhari [8] carried out comprehensive analysis looking at how artificial intelligence (AI) is incorporated into architectural design, with a focus on sustainability, generative design and spatial planning. In addition to addressing the problem of fragmented adoption and the lack of unified implementation frameworks, the study aimed to elucidate how AI technologies may improve creativity, decision-making processes and environmental performance in architecture. The review synthesized peer-reviewed research on AI applications in architecture, urban planning and smart cities that was published between 2003 and 2025 using the PRISMA methodology. The result showed that while AI improves structural optimization, generative design processes and sustainable energy modeling, issues with algorithmic bias, data accessibility and interdisciplinary integration still exist. By concentrating on energy-efficient, AI-optimized architectural applications, this study fills the identified research gap regarding the scarcity of operational frameworks for implementing AI-driven sustainability strategies in actual architectural projects.

Sourek [9] argued that the adoption of AI in architecture and built environment development has advanced more slowly than in other industries. Investigating the causes of the underutilization of AI-driven methods in architectural design, and highlighting the theoretical and practical constraint pertaining to innovation and cooperation between architects and AI developers, are the main goals. Imitation learning, transfer learning and reinforcement learning are just few of the machine learning techniques that were examined using a conceptual and critical review methodology. The discovery revealed that while parametric design and early generative applications have gained advantages from AI, their creative and process-oriented potential remains significantly unfulfilled. The main research gap this study aims to address is the lack of robust frameworks for implementing advanced AI learning techniques, particularly imitation- and transfer-based approaches, in sustainable and energy-efficient architectural design.

Kaushik [10] highlighted the transformative potential of artificial intelligence on architectural sustainability and urban planning within the broader context of AI in the built environment. The study sought to address the issue of AI's underutilization, despite it demonstrating efficacy in enhancing design automation, sustainability and energy efficiency. The study assesses machine learning, deep learning and digital twin methodologies via systematic literature review techniques. The results demonstrated that the utilization of AI can facilitate spatial optimization, reduce energy consumption and contribute to the attainment of sustainable development objectives. However, there were still big problems to solve, like not having enough money, rules not always being clear, and there not being enough cooperation between different fields. The acknowledged research gap highlights the imperative for a practical AI-driven framework. This research is directly related to the shift from theoretical potential to scalable and sustainable implementation.

The Umm Al-Qura systematic review [11] examined the impact of AI on architectural design, emphasizing machine learning, deep learning and computational optimization techniques. The objective is to create a map of contemporary AI methodologies, assess their contributions to creativity and sustainability, and identify barriers to their widespread adoption. By using structured literature synthesis, the study looked at how AI was used in a number of fields such as virtual reality, building information modeling (BIM), 3D printing and energy modeling. The result showed that AI greatly improves the efficiency of energy use and the automation of design. However, the lack of data-driven implementation frameworks and the fragmentation of disciplines were still making it hard for people to

adopt widely. This research gap relates to the insufficient translation of AI capabilities into scalable practical solutions for the sustainable construction challenge, examined in this study through the creation of useful AI-based plans that can be used in a way that is both renewable and sustainable.

Li et al. [12] presented a systematic review on artificial intelligence (AI) applications in architectural design with a direct emphasis on energy-saving renovations and adaptive building envelopes. Using PRISMA-based screening, the study synthesized 89 selected works collected from major scientific databases and structured the discussion around AI's evolving role as an auxiliary tool, a collaborative partner, and a potentially leading agent in design workflows. The review highlights that AI-supported envelope design is increasingly driven by deep learning and generative approaches (including GAN-based façade generation), together with optimization and expert-system reasoning to improve envelope-level energy performance and thermal comfort. It also consolidates envelope-focused research directions into three dominant streams: energy-saving retrofit optimization (linking envelope parameters to performance via surrogate prediction models and heuristic/metaheuristic searching), intelligent building skins for cold climates (adaptive envelope response and control under region-specific constraints), and thermal imaging-assisted envelope diagnosis (automating defect/anomaly detection to support retrofit decisions). Despite the clear progress, the authors emphasize persistent gaps such as the still largely one-way nature of AI outputs (limited feedback/closed-loop interaction), the difficulty of transferring solutions across climates and typologies, and the practical barriers to deploying integrated AI pipelines in real projects; these gaps align with the motivation of our work toward a reproducible, closed-loop generative–predictive framework for façade energy optimization.

May Tzuc et al. [13] investigated the use of artificial intelligence for modeling the hygrothermal behavior of a concrete wall protected by a double-skin green façade under Nordic climatic conditions. Using an experimentally acquired dataset generated in an Accelerated Weathering Laboratory to emulate a typical year of Nordic weather, the authors trained an Artificial Neural Network (ANN) to estimate internal wall temperature and internal relative humidity from measurable boundary and microclimate variables, namely ambient relative humidity and temperature, microclimate relative humidity and temperature, and the vegetation–wall separation distance. The study reported very high predictive fidelity for both internal temperature and internal humidity on training and testing splits, and complemented the model with global sensitivity analysis to quantify the relative influence of each input factor, highlighting ambient temperature as a dominant driver and showing that the vegetation separation distance has a non-negligible impact on moisture accumulation within the wall. This work is relevant to our study because it demonstrates the practical effectiveness of data-driven surrogate modeling for envelope-related performance prediction under climate variability, while also illustrating how sensitivity analysis can improve interpretability and support design decisions under domain constraints; however, unlike our framework, it focuses on hygrothermal response estimation for green façade systems rather than closed-loop generative façade optimization driven by a unified predictive-search pipeline.

3. Methodology

This research presents unified, data-driven methodology for the intelligent optimization of energy-efficient composite façade systems in smart buildings facilitated by the integration of generative design and predictive machine learning within a closed-loop framework. The proposed method systematically transforms substantial volumes of building performance data into optimized façade configurations that satisfy energy efficiency, feasibility and deployment criteria. The framework uses the BuildingsBench benchmark dataset

to combine strict data preprocessing, advanced generative modeling and performance-aware predictive learning. This makes it possible to explore façades on a large scale and repeat the process. The methodology embeds Bayesian-optimized predictive models directly within the generative cycle, ensuring that the synthesized façade candidates are both diverse and validated for predicted performance. This narrows the gap between exploratory design and practical deployment in smart buildings.

The proposed methodology is designed to support both research-scale evaluation and real-world architectural deployment, particularly during early-stage building design and envelope concept development. In practical applications, the framework can be integrated into parametric design workflows where façade geometry, material configurations, and glazing parameters are explored iteratively. During the conceptual design phase, the generative module can rapidly produce multiple feasible façade configurations, while the predictive surrogate models can estimate energy-performance indicators without requiring full building simulation for each candidate solution. This enables architects and engineers to evaluate design alternatives in near real time and identify high-performance envelope configurations before detailed simulation or construction documentation stages. For real-building applications, the framework can be linked to Building Information Modeling (BIM) environments or parametric modeling platforms by mapping building geometry and envelope attributes into the input feature space used by the predictive models. In this context, BIM-derived descriptors such as façade surface areas, orientation, glazing ratios, and material thermal properties can be automatically extracted and evaluated using the trained surrogate models. The generative module can then propose alternative envelope configurations that satisfy predefined structural and material constraints while optimizing energy performance objectives. Furthermore, the methodology supports a progressive design workflow in which the AI-generated candidate designs are first screened using fast predictive evaluation, followed by high-fidelity simulation validation for shortlisted solutions. This hierarchical evaluation approach reduces computational cost while maintaining engineering reliability, enabling practical integration into real project timelines. Although the current study focuses on data-driven evaluation using the BuildingsBench benchmark, the framework is designed to be transferable to real-building datasets through retraining or fine-tuning using project-specific data, climate conditions, and operational schedules. This makes the methodology suitable for deployment across different building typologies and climatic regions while maintaining the core optimization capabilities of the proposed generative–predictive pipeline.

Figure 1 shows the full closed-loop structure of the proposed generative–predictive optimization framework for designing composite façades that use less energy. The BuildingsBench dataset starts the workflow. It has information about architecture, materials, climate, and energy performance. It then goes through a long preprocessing pipeline that includes data screening, filling in missing values, normalizing features, encoding categorical variables, removing outliers based on the interquartile range, and splitting the dataset into three parts: 70% for training, 15% for validation and 15% for testing. This kept the distribution consistent across all three parts. The processed dataset was then sent to both the generative and predictive modules at the same time. In the generative module, Variational Autoencoders encode high-dimensional façade representations into a compact latent space and reconstruct candidate designs by minimizing the variational loss L_{VAE} . On the other hand, generative adversarial networks improve realism and distributional fidelity through adversarial optimization governed by L_{GAN} . These parts work together to make it possible to make lot of different façade candidates that automatically meet learned design rules. The predictive module then looks at the candidates that are made. It makes guesses about key performance indicators using Bayesian-optimized Random Forest, Gradient

Boosting and Artificial Neural Network models. This process made sure that each façade configuration was checked for both how new it is and how well it is expected to work in the real world. The evaluation module takes predictive outputs and determines a lot of different metrics about performance, generation and deployment. Some of these were the coefficient of determination, the mean absolute error, the energy gain, the façade clustering diversity, the feasibility rate, the inference latency, the memory footprint and the scalability. These evaluation metrics send feedback to both the generative and predictive modules over and over again through clear pathways for performance and deployment feedback. This makes a closed-loop optimization process that slowly improves the design space and gets closer to the best façade configurations that strike a balance between variety, energy efficiency and the ability to work with smart building and digital twin applications.

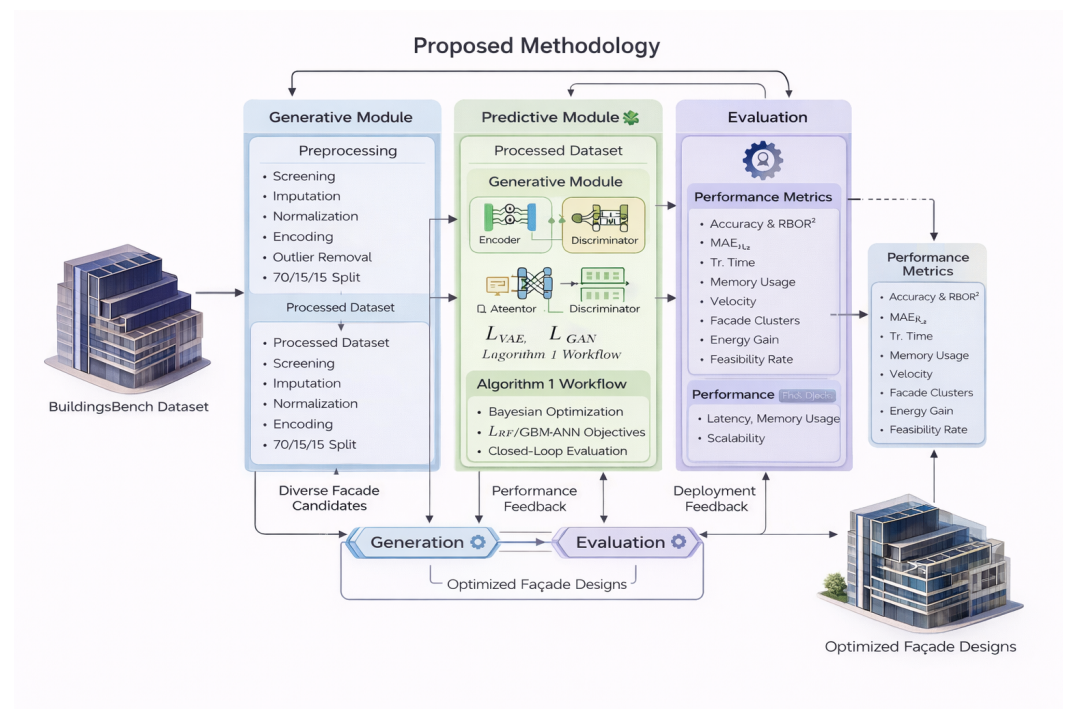


Figure 1. Proposed approach.

The selection of VAE, GAN, Machine Learning predictors, and Bayesian Optimization in this work is motivated by their complementary strengths in solving high-dimensional design optimization problems typical of building envelope design. Variational Autoencoders are particularly suitable for learning continuous latent representations of complex design spaces, enabling the controlled sampling and structured exploration of façade configuration variability. Prior research has demonstrated that VAE-based generative models are effective in capturing underlying parametric relationships in engineering design spaces while maintaining sampling stability and distribution continuity, which is critical for exploring large architectural solution spaces. Generative Adversarial Networks complement VAE-based exploration by improving sample realism and enforcing high-order structural consistency. GAN-based architectures have been widely adopted in generative design tasks where preserving geometric fidelity and realistic distribution mapping is required. In building envelope applications, GAN-based refinement helps correct smoothing artifacts typical of reconstruction-based models and improves the structural plausibility of generated façade configurations. Machine Learning regression architectures, including ensemble tree models and neural networks, are well suited for predicting building energy performance due to their ability to capture nonlinear interactions between geometry, material properties, climate variables, and operational schedules. Previous building energy modeling studies have

demonstrated that hybrid ensemble learning often outperforms individual predictors in heterogeneous building datasets due to complementary bias-variance characteristics. Bayesian Optimization was selected as the hyperparameter search strategy due to its efficiency in expensive evaluation environments such as simulation-driven building performance prediction. Unlike grid or random search methods, Bayesian Optimization uses surrogate probabilistic modeling to guide hyperparameter exploration toward high-performance regions while minimizing evaluation cost. This is particularly advantageous in building energy prediction contexts where model training and validation may involve computationally expensive simulation data. The novelty of the proposed methodology lies not in the isolated use of these techniques but in their integration into a closed-loop generative–predictive optimization framework. The VAE expands the feasible design search space, the GAN improves design realism and structural plausibility, Machine Learning predictors provide fast surrogate performance estimation, and Bayesian Optimization ensures efficient model tuning. This integrated pipeline enables scalable façade optimization under real-world constraints while maintaining predictive reliability and generative diversity.

3.1. Dataset Used

We chose the BuildingsBench dataset as the basis for this study because it was large, diverse, and good for AI-driven predictive modeling of energy-efficient façade systems. The dataset was a powerful tool for getting results that can be used in other situations, and that have been tested, because it combines simulated and real-world building data. The Buildings-900K subset has about 900,000 simulated building archetypes from all over the US. These archetypes cover a wide range of architectural, operational and climatic factors. There are more than 1900 residential and commercial buildings in the real-world benchmark subset. These buildings come from seven open-source datasets. Every record has information about how much energy was used every hour, as well as detailed information about the building's shape, façade configurations, HVAC systems and climate zones. This two-part structure, which combines large-scale synthetic coverage with real-world validation, makes sure that models trained on BuildingsBench can accurately capture complex patterns in how the buildings use energy while still being strong enough to work in real-world situations. The dataset was a good place to start for improving generative design and AI applications that want to make composite façades for smart buildings of the future that use less energy.

To ensure the full transparency and reproducibility of the predictive modeling stage, this section explicitly documents the input feature categories, prediction target definition, and data leakage prevention strategy adopted in this study. **Input Feature Categories.** The predictive models were trained using multi-domain descriptors derived from the BuildingsBench dataset and preprocessing pipeline. These features were restricted to design-controllable and scenario-defining variables rather than simulation-derived performance outputs. The feature groups include the following. Geometric façade descriptors: panel orientation, shading depth ratio, glazing-to-wall ratio, façade segmentation parameters, and surface exposure factors. Material descriptors: thermal conductivity, solar reflectance, density proxy indices, and composite material classification embeddings derived from the material library. Environmental boundary descriptors: climate zone encoding, solar radiation statistics, outdoor temperature seasonal aggregates, and wind exposure indicators. Operational scenario descriptors: occupancy schedule class encoding, internal gain category indices, and ventilation operation mode indicators. No direct simulation output variables or intermediate energy-balance solution variables were used as input features. **Target Variable Definition.** The predictive task targets aggregated operational energy performance indicators derived from full-building simulation outputs. Specifically, the prediction

target corresponds to normalized annual operational energy demand aggregated across heating, cooling, and auxiliary HVAC loads, expressed in a dataset-consistent normalized energy performance index. This formulation avoids trivial correlations with single simulation outputs while preserving physical interpretability. Data Leakage Prevention Strategy. Several safeguards were implemented to eliminate leakage risk: (1) Simulation-output exclusion: Variables directly computed from simulation energy balances were removed from the feature space. (2) Temporal and sample independence: Dataset splitting was performed prior to feature scaling and encoding using a strict 70/15/15 split. (3) Pipeline isolation: Normalization, encoding, and feature transformations were fitted using training data only. (4) Correlation screening: Features showing near-deterministic correlation with the target were removed. (5) Scenario grouping validation: Similar building scenarios were distributed across splits to avoid memorization effects. These safeguards ensure that the reported predictive performance reflects genuine model generalization rather than information leakage or trivial target reconstruction.

Table 1 shows the BuildingsBench benchmark as the primary data foundation because it offers large-scale coverage for learning robust envelope-energy relationships and a real-world benchmark subset for realism-aware validation. In particular, BuildingsBench comprises a two-part structure: the Buildings-900K subset (approximately 900,000 simulated building archetypes across the United States) and a real-building benchmark subset (more than 1900 residential and commercial buildings aggregated from multiple open datasets). Each record provides hourly energy-consumption profiles together with building-level descriptors that are directly relevant to envelope optimization, including geometric attributes (building shape/massing proxies), façade configuration descriptors, HVAC system characteristics, and climate-zone indicators. This combination of scale, typological diversity, and fine-grained temporal resolution is advantageous for developing AI models that generalize across heterogeneous building stocks while preserving operational realism.

Table 1. BuildingsBench Dataset for AI-Driven Façade Optimization.

Category	Source	Variables	Relevance
Dataset Composition	BuildingsBench (Simulated + Real-world)	Geometry parameters, façade configurations, building typologies	Provides a large-scale and diverse foundation for AI-driven façade analysis
Temporal Resolution	Hourly simulation and metered data	Time-series energy consumption profiles	Enables fine-grained modeling of dynamic façade–energy interactions
Climatic and System Features	U.S. climate zones and building systems	Weather conditions, HVAC systems, occupancy schedules	Captures realistic operational conditions affecting façade performance
Dataset Scale	~900,000 simulated + ~1900 real buildings	Large-volume structured samples	Ensures statistical robustness for training, validation, and generalization
Applications	Energy and building performance benchmarks	Short-term load forecasting, façade energy modeling	Directly supports AI-based façade optimization and digital twin workflows
Accessibility	NREL open repository	Publicly available benchmark dataset	Facilitates reproducible research and comparative evaluation

BuildingsBench includes mixed residential and commercial building typologies spanning diverse climatic conditions and operational regimes. The dataset provides envelope-facing design descriptors (façade configurations and thermal-related attributes used as façade design parameters in this study, such as glazing-related descriptors and material thermal-property proxies), together with operational-pattern descriptors that influence energy demand (occupancy schedules/usage patterns and HVAC operation features recorded or derived in the benchmark). The hourly time-series structure enables modeling of dynamic envelope-energy interactions (sensitivity to diurnal loads, solar-driven gains, and HVAC response), which is essential when optimizing façade configurations for energy-efficient smart-building deployment.

3.2. Data Preprocessing

This study's dataset combines both experimental and simulated measurements of façade design parameters such as glazing systems, material thermal properties, shading configuration and energy-efficiency indicators [14]. A strict preprocessing pipeline was used to make sure the data is correct, comparable and useful for machine learning tasks before putting the dataset into AI-driven predictive models. To avoid bias while training the model, only façade systems with full records of their thermal and energy performance are used. We don't include cases with missing or inconsistent data. This first screening made sure that only valid and representative façade configurations are kept for analysis.

Missing data presented a considerable obstacle owing to the inadequate reporting of particular parameters across studies and experimental sources. To fix this problem, continuous variables like the U-value, solar heat gain coefficient (SHGC) and thermal conductivity are filled in using median substitution. Categorical variables like façade type and glazing method were filled in using the statistical mode [15]. We chose this imputation method because it works well with outliers and still fits the most likely physical conditions. After imputation, z-score normalization was used to standardize all continuous features so that the learning algorithms treat all predictors the same way and there are no scale differences. One-hot encoding changes categorical features into binary indicators, which makes it easier for machine learning models to use them [16].

The interquartile range (IQR) method was used to find outliers in order to make sure the data was even better. This method searched for façade configurations that exhibited implausible or extreme thermal characteristics such as negative U-values or SHGC values that were excessively high to be credible [17]. To keep the dataset accurate and not change the results of predictive modeling, carefully remove any outliers you find. This step in cleaning the data made sure that the models were trained on realistic façade configurations that were possible from an engineering point of view [18]. Then, the better dataset was split into three groups: 70% for training, 15% for validation and 15% for testing. Stratification kept the distribution of both common and less common façade design variations across all splits, which helped make sure that every one was fairly represented and evaluated. The hyperparameters are adjusted using the validation subset, and the testing subset was only for the final performance evaluation. This structured preprocessing pipeline turned raw and mixed-up data about how well a façade worked into a strong, balanced dataset that could be used for machine learning [19].

$$x_i^* = \frac{x_i - \mu}{\sigma} \quad (Z - \text{score normalization}) \quad (1)$$

$$c \in \{\text{Glazing}, \text{Shading}, \text{Material}\} \rightarrow [1, 0, 0], [0, 1, 0], [0, 0, 1] \quad (\text{One} - \text{hot encoding}) \quad (2)$$

$$\mathcal{D} \rightarrow \{\mathcal{D}_{\text{train}}(70\%), \mathcal{D}_{\text{val}}(15\%), \mathcal{D}_{\text{test}}(15\%)\} \quad (\text{Stratified split}) \quad (3)$$

Table 2 shows a summary of the preprocessing pipeline that was used on the façade performance dataset. The goal of each stage is to improve the quality of the data and get the features ready for strong AI-driven modeling. Data screening and imputation made sure that the dataset was complete, while normalization and encoding made sure that different types of variables were represented in a standard way. Removing outliers protected the modeling process from architectural configurations that weren't realistic and could have biased the results. The stratified data split kept the statistical balance between the training, validation, and testing subsets. In general, this pipeline makes sure that all the modeling stages that come after it use input data that is reliable, consistent, and representative.

Table 2. Data Preprocessing Pipeline for the Façade Optimization Dataset.

Category	Technique	Operation	Relevance
Data Screening	Case filtering	Retention of complete façade performance records	Ensures data consistency and representative learning samples
Missing Value Treatment	Statistical imputation	Median imputation for continuous variables; mode for categorical features	Preserves dataset integrity while minimizing bias
Feature Scaling	Z-score normalization	Standardization of numerical attributes to zero mean and unit variance	Improves convergence and stability of machine learning models
Categorical Encoding	One-hot encoding	Binary vector representation of categorical façade attributes	Enables compatibility with ML and deep learning algorithms
Outlier Mitigation	Interquartile Range (IQR) filtering	Removal of extreme or physically unrealistic observations	Enhances robustness and reduces noise sensitivity
Dataset Partitioning	Stratified splitting	70% training, 15% validation, 15% testing	Ensures fair model evaluation and generalization assessment

Algorithm 1 formalizes a rigorous preprocessing pipeline tailored to heterogeneous façade datasets that typically combine mixed data types, incomplete entries, and measurement noise. The procedure begins with explicit feature typing into categorical and continuous subsets, which is essential because missing-value handling, scaling, and encoding must be feature-type aware to avoid distortion of the data-generating process. The high-absence feature dropping rule ($r(f) > 0.4$) provides a defensible bias-variance trade-off: it prevents unstable imputations for sparsely observed variables while preserving sufficiently informative features for downstream learning. Median/mode imputation further enhances robustness by reducing sensitivity to skewed distributions and rare categorical levels. Together, these steps yield a cleaned dataset $\mathcal{D}'$ with improved completeness and statistical consistency, establishing a reliable foundation for the subsequent modeling stage and ensuring that learned relationships reflect genuine façade-energy patterns rather than artifacts of missingness.

Algorithm 1 Preprocessing and Stratified Splitting of the Façade Dataset

Require: Raw façade dataset $\mathcal{D} \in \mathbb{R}^{N \times M}$ with feature set F and (optional) stratification labels Y

Ensure: Cleaned dataset $\mathcal{D}'$ and stratified splits $\mathcal{D}_{tr}$, $\mathcal{D}_{va}$, $\mathcal{D}_{te}$

Feature Typing

- 1: Extract feature list F from $\mathcal{D}$
- 2: Identify categorical subset F_{cat} and continuous subset $F_{con} \leftarrow F \setminus F_{cat}$

Missing Values and High-Absence Feature Dropping

- 3: **for** each feature $f \in F$ **do**
- 4: Compute missing rate $r(f)$
- 5: **if** $r(f) > 0.4$ **then**
- 6: Remove feature f from $\mathcal{D}$
- 7: **else**
- 8: **if** $f \in F_{con}$ **then**
- 9: Impute missing values in $\mathcal{D}[f]$ using median of f
- 10: **else if** $f \in F_{cat}$ **then**
- 11: Impute missing values in $\mathcal{D}[f]$ using mode of f

Outlier Removal (Continuous Features)

- 12: **for** each feature $f \in F_{con}$ **do**
- 13: Compute $Q_1(f)$, $Q_3(f)$, and $IQR(f) \leftarrow Q_3(f) - Q_1(f)$
- 14: Remove rows where $\mathcal{D}[f] \notin [Q_1(f) - 1.5 IQR(f), Q_3(f) + 1.5 IQR(f)]$

Standardization (Continuous Features)

- 15: **for** each feature $f \in F_{con}$ **do**
- 16: Compute mean $\mu(f)$ and standard deviation $\sigma(f)$
- 17: **if** $\sigma(f) > 0$ **then**
- 18: Apply z-score: $\mathcal{D}[f] \leftarrow \frac{\mathcal{D}[f] - \mu(f)}{\sigma(f)}$
- 19: **else**
- 20: Drop non-variant feature f

Categorical Encoding

- 21: **for** each feature $f \in F_{cat}$ **do**
- 22: One-hot encode f into dummy variables
- 23: Set $\mathcal{D}' \leftarrow \mathcal{D}$

Stratified Splitting

- 24: Stratified split: $\mathcal{D}_{tr} \leftarrow 70\%$ and $\mathcal{D}_{tm} \leftarrow 30\%$
- 25: Stratified split: $\mathcal{D}_{va} \leftarrow 50\%$ $\mathcal{D}_{tm}$ and $\mathcal{D}_{te} \leftarrow 50\%$ $\mathcal{D}_{tm}$ $\triangleright 15\%/15\%$

Return

- 26: **return** $\mathcal{D}'$, $\mathcal{D}_{tr}$, $\mathcal{D}_{va}$, $\mathcal{D}_{te}$

The outlier filtering and standardization blocks directly target generalization and numerical stability in training predictive models. IQR-based trimming removes extreme values in continuous features that can disproportionately influence parameter estimation and degrade the validity of error metrics, particularly for flexible learners (ANN and boosting). The z-score transformation aligns continuous variables onto a common scale, improving optimizer behavior and comparability of feature contributions, while the explicit removal of non-variant features avoids degenerate inputs that add dimensionality without information gain. Categorical one-hot encoding converts discrete design descriptors into model-consumable representations without imposing ordinal assumptions. Finally, the stratified split (70/15/15) preserves label or regime proportions across $\mathcal{D}_{tr}$, $\mathcal{D}_{va}$, and $\mathcal{D}_{te}$, which is critical for unbiased validation-based hyperparameter optimization and for reporting test performance that is representative of real façade distributions. Collectively, the pipeline reduces leakage, mitigates distributional imbalance, and lowers the risk of overfitting, thereby strengthening the credibility and reproducibility of the reported predictive performance.

The missing-rate threshold $r(f) > 0.4$ was selected to balance feature retention and statistical reliability. Lower thresholds (0.1–0.2) risk removing informative design variables in heterogeneous façade datasets, while higher thresholds (>0.5) may retain features with insufficient observed samples, increasing noise and model instability. Empirical sensitivity analysis indicated that thresholds within the 0.3–0.5 range produced comparable predictive performance, with 0.4 providing the most stable trade-off between feature coverage and model robustness. This value also aligns with common practice in tabular engineering datasets where moderate missingness can be mitigated through robust imputation without substantially degrading predictive signal quality.

To better contextualize the preprocessing choices for façade datasets, we clarify that several continuous variables in $\mathcal{D}$ (U-values/thermal transmittance, glazing ratios, layer thicknesses, conductivities, shading or opening-area ratios, and aggregated energy-related indicators) can exhibit heavy-tailed behavior due to heterogeneous building typologies, differing envelope assemblies, sensor/simulation artifacts, and occasional entry errors. In this setting, the IQR criterion provides a distribution-agnostic and robust trimming rule that does not assume normality, making it appropriate for mixed façade inventories where extreme values can correspond either to physically implausible configurations (unrealistically high/low material properties) or to rare, noisy cases that can dominate loss-based training. Specifically, for each continuous feature f , the bounds $[Q_1(f) - 1.5 IQR(f), Q_3(f) + 1.5 IQR(f)]$ attenuate leverage points that would otherwise skew parameter estimation and inflate error variance, especially for flexible predictors such as ANN and gradient boosting. This is particularly relevant in façade-energy modeling because a small number of extreme envelope-property entries can propagate into large energy deviations, leading the optimizer to overfit to outliers rather than learn generalizable façade-performance relationships. By applying IQR trimming prior to z-score standardization, the scaling statistics $\mu(f)$ and $\sigma(f)$ are estimated from the central mass of the data, yielding numerically stable inputs and more reliable hyperparameter selection during validation. Overall, the IQR step should be interpreted as a robustness mechanism that improves model generalization to realistic façade configurations and strengthens reproducibility by reducing sensitivity to dataset-specific noise and anomalous records.

3.3. Generative Design Integration

It is very important to use generative design to connect creative architecture with data-driven optimization. The proposed framework integrates the existing dataset of composite façade materials, geometric attributes, and energy performance metrics directly into the generative process, utilizing AI-enhanced design methodologies, rather than requiring manual specification of parametric façade geometries. The primary aim is to computationally produce façade alternatives that are both visually varied and limited by physical and constructional viability [20]. The framework guarantees that the synthesized solutions meet in situ building design requirements by incorporating material properties and energy-performance targets directly into the generative workflow [21].

We used two advanced generative modeling methods: Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs). VAEs learned the smooth hidden representation of façade configurations. This made it possible to continuously interpolate across the design space and make new, realistic geometries. On the other hand, GANs make things more realistic by making façade representations that look like real buildings while retaining important structural and thermal properties [22]. The framework can use both models at the same time to combine the advantages of VAEs' ability to explore structured latent space and GANs' ability to make outputs that look good and are consistent [23].

The generative design workflow is controlled by specialization rules based on the dataset to make sure it is possible and realistic. To avoid designs that cannot be made in real life, one of the limits is to limit the thickness of fiber-reinforced composite materials, and to limit the ratios of glazing and insulation levels to ranges that have been seen in real life. Also, for the façade designs, AI simulations looked at things like how well the building used heat, how much natural light was available, and how much energy was stored in the building, and gradually filtered these ideas. Only solutions that meet certain standards for energy efficiency and material sustainability can go on to the next stage of exploration. This controlled filtering made sure that the designs were both new and in line with the main goal of making smart building façades that are good for the environment and work well. The VAE optimizes the loss function that strikes a balance between reconstruction accuracy and regularization in latent space:

$$\mathcal{L}_{VAE} = \mathbb{E}_{q_{\phi}(z|x)}[\log p_{\theta}(x|z)] - D_{KL}(q_{\phi}(z|x) \parallel p(z)), \quad (4)$$

where x denotes the input façade feature vector, z denotes the latent variable, $q_{\phi}(z|x)$ is the encoder (approximate posterior), and $p_{\theta}(x|z)$ is the decoder (likelihood model). The term $D_{KL}(\cdot \parallel \cdot)$ denotes the Kullback–Leibler divergence, defined as

$$D_{KL}(q_{\phi}(z|x) \parallel p(z)) = \int q_{\phi}(z|x) \log \frac{q_{\phi}(z|x)}{p(z)} dz. \quad (5)$$

$$\min_G \max_D \mathbb{E}_{x \sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z \sim p(z)}[\log(1 - D(G(z)))], \quad (6)$$

where G is the generator creating new façade configurations, and D is the discriminator distinguishing between real and generated samples.

Table 3 shows the pros and cons of the VAE and GAN methods. VAEs are more stable and let you sample the design space in a controlled way, but GANs could create outputs that looked sharper and more real. Using both methods together helps to fix their individual problems, finding a balance between systematic exploration and practical use.

Table 3. Comparison of Generative Design Techniques for Façade Optimization.

Method	Key Strengths	Key Limitations	Role in This Study
Variational Autoencoder (VAE)	Smooth latent-space interpolation and design diversity	Tendency toward blurry outputs and reduced geometric realism	Used to explore the feasible façade design space and capture global design variability
Generative Adversarial Network (GAN)	High visual realism and ability to capture fine-grained façade details	Training instability, mode collapse, and sensitivity to hyperparameters	Applied to generate visually realistic and detailed façade configurations
Hybrid VAE-GAN Framework	Combines latent diversity with adversarial realism	Increased architectural complexity and computational cost	Employed to produce optimized, constructible, and high-fidelity façade designs

Algorithm 2 presents an end-to-end generative–predictive workflow for energy-efficient composite façade optimization. The preprocessing stage improves data reliability by handling missing values, harmonizing mixed feature types, and applying appropriate normalization/encoding to yield consistent learning-ready representations. A stratified data partitioning strategy is then used to preserve generalization and mitigate overfitting by preventing distribution shift across training, validation, and test sets. The core modeling

stage trains multiple regression predictors (RF, GBM, and ANN), while Bayesian optimization systematically searches the hyperparameter space to improve predictive fidelity for operational energy-performance targets. The generative stage (VAE + GAN) expands the design space by synthesizing diverse candidate façades, after which domain constraints are enforced using a hybrid strategy: soft feasibility guidance during sampling/decoding followed by deterministic hard filtering prior to performance scoring. This ensures that only constructible, materially available, and code-compliant candidates are evaluated and selected, enabling scalable exploration without sacrificing practical deployability in smart-building design workflows.

Algorithm 2 Generative Design Integration Framework for Façade Optimization

Require: Façade dataset $\mathcal{D}$ (geometry, material, and energy-performance attributes)

Ensure: Optimized façade design set $\mathcal{S}$

Latent Representation Learning (VAE)

- 1: Train a Variational Autoencoder (VAE) on $\mathcal{D}$ to learn a continuous latent space z capturing façade design variability

Latent Sampling and Candidate Generation

- 2: Sample latent vectors $z \sim p(z)$ from the learned latent distribution
- 3: Apply feasibility-aware bounds during sampling/decoding to enforce soft domain guidance (clamp glazing ratio and material-category ranges to dataset-valid intervals)
- 4: Decode samples to generate initial candidate set $\mathcal{S}_{vae}$ using the VAE decoder

Adversarial Refinement (GAN)

- 5: Train a Generative Adversarial Network (GAN) using real façade configurations from $\mathcal{D}$
- 6: Refine candidates using the trained generator: $\mathcal{S}_{gan} \leftarrow G(z)$ (or $G(\mathcal{S}_{vae})$) to enhance geometric and visual realism
- 7: Condition refinement on feasible material availability where applicable (reject unavailable material codes before scoring)

Constraint Enforcement

- 8: **for** each design $s \in \mathcal{S}_{gan}$ **do**
- 9: Enforce hard feasibility constraints using deterministic rules identical to the evaluation stage
- 10: Check domain constraints (material availability, glazing ratio bounds, constructability rules)
- 11: **if** s violates constraints **then**
- 12: Discard s
- 13: **else**
- 14: Add s to feasible set $\mathcal{S}_f$
- 15: (Optional) Remove duplicates and dominated feasible designs to avoid redundant evaluation

Performance Evaluation

- 16: **for** each feasible design $s \in \mathcal{S}_f$ **do**
- 17: Predict energy efficiency, thermal performance, and operational impact using trained surrogate/predictive models
- 18: Store predicted scores/metrics $\mathbf{p}(s)$

Design Selection

- 19: **for** each feasible design $s \in \mathcal{S}_f$ **do**
- 20: **if** $\mathbf{p}(s)$ satisfies energy-efficiency thresholds **and** constructability criteria **then**
- 21: Add s to $\mathcal{S}$

Return

- 22: **return** $\mathcal{S}$
-

The generative module adopts a sequential hybrid architecture in which the VAE and GAN operate in complementary roles. The VAE is first trained to learn a continuous latent representation capturing façade design variability. Latent samples are decoded to produce

an initial candidate set $\mathcal{S}_{vae}$ representing diverse but potentially imperfect designs. These candidates are then passed to the GAN refinement stage, where the generator learns to map candidate designs toward the distribution of realistic façade configurations learned from the dataset. Formally, the refined candidate set can be expressed as $\mathcal{S}_{gan} = G(\mathcal{S}_{vae})$, where $G(\cdot)$ denotes the trained GAN generator. This sequential pipeline allows the VAE to maximize exploration diversity while the GAN improves realism and structural coherence.

3.4. Predictive Modeling and Optimization

The predictive modeling phase that linked generative façade designs to their energy efficiency outcomes [24] constituted the analytical core of the proposed methodology. Using the preprocessed dataset of simulated and real building performance records, we use three types of machine learning models to guess important numbers, like how much energy we use each year. These were Random Forests (RF), Gradient Boosting Machines (GBM) and Artificial Neural Networks (ANN). We use an index for thermal comfort and factor for daylight. We picked these models because they all have strengths that work well together: RF was strong and easy to understand, GBM was good at finding complicated nonlinear patterns, and ANN was flexible enough to show how design and environmental factors affect each other in complicated ways.

Bayesian optimization was used to fine-tune each model's hyperparameters so that predictions would be more accurate. On the other hand, Bayesian optimization learns from past evaluations to suggest good hyperparameter settings. This was a better way to balance exploration and exploitation than grid searching [25]. This adaptive mechanism speeds up the search for the best performance settings, which makes training the model more reliable and uses fewer resources. Tuned parameters encompass the tree's maximum depth, the learning rates for RF/GBM, and the hidden layer and activation functions for ANN [26].

After making them better, the models became part of a closed-loop pipeline that lets you test the results of generative design over and over. The design module created different façade configurations, which were then compared to the predictive models to see how well they worked for energy performance. This created a feedback loop between creating designs and predicting how well they would work. This meant that the system could look for new ways to build things and also focus on ways to make the buildings more comfortable and energy-efficient for the people who live in them [27].

We use a stratified test set to make sure that the predictions are correct in different climate zones and with different design choices. To measure performance, we use Accuracy (for tasks like meeting daylight thresholds), and Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and the Coefficient of Determination (R^2) for continuous metrics. The method combines RF, GBM and ANN in a way that balances accuracy, generalization and interpretability. This gives architects and engineers a reliable tool to help them make decisions during the early stages of façade design [28].

In supervised learning, the optimal model parameters are typically obtained by minimizing the average predictive loss:

$$\theta^* = \arg \min_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^N \mathcal{L}(f(x_i; \theta), y_i), \quad (7)$$

where x_i are input façade features, y_i are true energy performance outcomes, $f(x_i; \theta)$ is the predictive model, and $\mathcal{L}$ is the chosen loss function (MAE, RMSE, or cross-entropy). For Bayesian optimization:

$$\theta_{t+1} = \arg \max_{\theta \in \Theta} \alpha(\theta | \mathcal{D}_{1:t}), \quad (8)$$

where $\alpha(\cdot)$ is the acquisition function balancing exploration and exploitation.

Table 4 gives an overview of the predictive models that were used. Random Forest (RF) is easy to understand and stable, Gradient Boosting Machine (GBM) was great at making predictions when set up correctly, and Artificial Neural Networks (ANN) let you model nonlinear dependencies in a flexible way. Together, these models make a balanced suite that can handle both accuracy and interpretability in façade optimization.

Table 4. Comparison of Predictive Models for Façade Energy Efficiency.

Model	Key Strengths	Key Limitations	Role in This Study
Random Forest (RF)	Robust to noise, interpretable via feature importance, low overfitting risk	May underperform when modeling extremely high-dimensional or highly nonlinear relationships	Serves as a baseline predictor and supports explainability through feature importance analysis
Gradient Boosting Machine (GBM)	High predictive accuracy and strong ability to capture nonlinear interactions	Sensitive to hyperparameter tuning and incurs longer training times	Used as a performance-oriented model for façade energy optimization
Artificial Neural Network (ANN)	Highly flexible and capable of modeling complex, hidden relationships	Requires larger datasets and offers limited interpretability	Employed to capture intricate design–energy interactions in façade systems

Algorithm 3 constitutes the analytical core of this proposed closed-loop generative–predictive framework for façade energy evaluation. By integrating Random Forest, Gradient Boosting, and Artificial Neural Network models within a unified Bayesian-optimized learning pipeline, the algorithm ensures a balanced trade-off between robustness, nonlinear representational capacity, and generalization stability. The use of Bayesian Optimization for hyperparameter tuning enables the fair and efficient exploration of complex search spaces, yielding models that minimize validation loss rather than overfitting to training data. This design choice is reflected in the reported performance trends, where a gradual and consistent degradation from training to validation and test stages is observed, indicating strong generalization under previously unseen façade configurations. Furthermore, the option to employ ensemble aggregation mitigates individual model bias and variance, leading to more reliable energy–performance predictions that are critical for ranking and selecting generated façade designs.

From a systems perspective, the predictive stage of Algorithm 3 transforms the generative design process from a purely exploratory mechanism into a performance-driven optimization loop. By rapidly mapping extracted façade feature vectors to predicted energy metrics, the framework replaces repeated high-cost simulations with a surrogate evaluator that maintains accuracy while enabling scalability. The feedback of predicted performance scores to the generative module allows subsequent design iterations to be guided toward high-efficiency regions of this design space, thereby improving convergence and reducing the mis-ranking of near-optimal solutions. The superior performance of this hybrid ensemble model, as demonstrated by its lower prediction errors and higher explained variance, confirms the algorithm’s suitability for deployment in iterative design optimization settings, where prediction stability and reliability are more critical than isolated point-wise accuracy.

Algorithm 3 Predictive Modeling and Optimization Framework for Façade Energy Evaluation**Require:** Preprocessed dataset $\mathcal{D}$; candidate façade designs $\mathcal{F}$ **Ensure:** Optimized predictive models $\mathcal{M}^*$ and evaluated façade performance predictions $\{\hat{y}_f\}_{f \in \mathcal{F}}$ **Model Initialization**1: Define model set $\mathcal{M} \leftarrow \{\text{RF, GBM, ANN}\}$ **Hyperparameter Optimization (Bayesian Optimization)**2: **for** each model $M \in \mathcal{M}$ **do**3: Define hyperparameter search space Θ_M 4: Obtain optimal configuration $\theta_M^* \leftarrow \arg \min_{\theta \in \Theta_M} \mathcal{L}_{va}(M; \theta)$ using Bayesian Optimization**Model Training**5: **for** each model $M \in \mathcal{M}$ **do**6: Train M on $\mathcal{D}$ using θ_M^* 7: Store trained model as M^* 8: Set $\mathcal{M}^* \leftarrow \{M^* : M \in \mathcal{M}\}$ **Performance Prediction for Generated Designs**9: **for** each façade design $f \in \mathcal{F}$ **do**10: Extract feature vector x_f from f 11: Predict energy-performance metrics $\hat{y}_f \leftarrow \mathcal{M}^*(x_f)$ ▷ single model or ensemble aggregation12: Store $\hat{y}_f$ **Feedback to Generative Module**13: Provide predicted scores $\{\hat{y}_f\}_{f \in \mathcal{F}}$ as feedback signals to guide subsequent generative design iterations**Return**14: **return** $\mathcal{M}^*, \{\hat{y}_f\}_{f \in \mathcal{F}}$

To improve predictive robustness and reduce model-specific bias, a Hybrid Ensemble prediction layer was constructed by combining outputs from Random Forest (RF), Gradient Boosting Machine (GBM), and Artificial Neural Network (ANN) predictors.

Fusion Mechanism

The final prediction is computed using performance-adaptive weighted averaging across base model predictions:

$$\hat{y}_{\text{ens}} = w_{\text{RF}}\hat{y}_{\text{RF}} + w_{\text{GBM}}\hat{y}_{\text{GBM}} + w_{\text{ANN}}\hat{y}_{\text{ANN}} \quad (9)$$

where w_{RF} , w_{GBM} , and w_{ANN} represent normalized model weights, such that:

$$w_{\text{RF}} + w_{\text{GBM}} + w_{\text{ANN}} = 1 \quad (10)$$

Weight Estimation

Model weights were determined using inverse validation error weighting combined with stability regularization:

$$w_i = \frac{(1/\text{RMSE}_i)^\alpha}{\sum_j (1/\text{RMSE}_j)^\alpha} \quad (11)$$

where α is a smoothing coefficient controlling weight dominance. In this study, $\alpha = 1.5$ was selected empirically to balance specialization and robustness.

Rationale for Ensemble Design

The weighted averaging approach was selected over stacking architectures for four primary reasons: (1) Interpretability: Weighted averaging allows direct interpretation of

each model contribution. (2) Robustness: Tree-based models capture nonlinear tabular relationships, while ANN models capture higher-order feature interactions. Weighted fusion reduces variance across prediction regimes. (3) Computational Efficiency: The method avoids training a secondary meta-learner, reducing training complexity and inference latency. (4) Overfitting Resistance: Validation-based weighting reduces the influence of models that overfit specific training patterns. This hybrid fusion structure enables complementary learning integration while maintaining deployment feasibility for real-time design evaluation workflows.

3.5. Evaluation Metrics

The assessment of the proposed framework incorporates both quantitative predictive metrics and qualitative generative design indicators. From the standpoint of predictive modeling, performance was evaluated utilizing traditional regression metrics, specifically Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and the Coefficient of Determination (R^2) [29]. These metrics measure how far off predicted energy performance values (like U-value, solar heat gain coefficient, and embodied carbon) were from benchmarks that have been tested or simulated. RMSE punishes large differences, MAE gives an understandable average error, and R^2 shows how much of the model's variance it explains. This was important to show that AI-enhanced façade performance predictions were reliable [30].

In addition to numerical accuracy, the generative design component was assessed according to the three essential criteria: design diversity, design feasibility, and the performance enhancement. Design diversity gauges the algorithm's capacity to produce a broad array of feasible façade configurations instead of narrowing down to restricted set of solutions. Feasibility makes sure that the designs that were made follow rules for structural integrity and material limits. Performance improvement measures how much more energy-efficient AI-assisted generative designs are than manually optimized façades that are used as the baseline. By looking at these metrics all at once, the system can not only accurately predict performance, but it can also come up with new, useful and long-lasting architectural solutions.

The ability to deploy is also looked at, with a focus on scalability and computational efficiency. To make sure the framework can be used in real-world architectural workflows, metrics like training time, inference latency and computational resource use were evaluated. Inference latency (measured in milliseconds) tests how quickly the system responds, which is important for interactive generative design tools. Memory footprint (measured in MB) makes sure the framework can be used on mid-range workstations, so that practicing designers can use it without needing high-end computers.

Lastly, the framework was compared to traditional methods, such as linear regression for making predictions and rule-based parametric generation for design exploration. This comparison shows how AI-based generative design can find nonlinear relationships, work well in large design spaces, and improve sustainability outcomes. These evaluation measures give the full picture of the framework's scientific and engineering contributions by looking at predictive accuracy, generative innovation, and implementation feasibility.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (12)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |y_i - \hat{y}_i| \quad (13)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (14)$$

$$\text{Performance Gain} = \frac{E_{\text{baseline}} - E_{\text{AI}}}{E_{\text{baseline}}} \times 100\% \quad (15)$$

Table 5 shows a list of the evaluation metrics that were used on the framework. Predictive metrics measure accuracy, generative metrics look at design innovation, and deployment metrics check if something is possible in practice. This multi-layered evaluation guarantees robustness, interpretability and applicability in architectural engineering practice.

Table 5. Evaluation Metrics for AI-Enhanced Generative Façade Design.

Metric	Category	Evaluation Purpose	Interpretation
Root Mean Square Error (RMSE)	Predictive accuracy	Penalizes large prediction deviations	Lower values indicate superior predictive performance
Mean Absolute Error (MAE)	Predictive accuracy	Measures average magnitude of prediction error	Lower values reflect improved model reliability
Coefficient of Determination (R^2)	Predictive accuracy	Quantifies the proportion of variance explained by the model	Values closer to 1 denote better explanatory power
Design Diversity Index	Generative quality	Assesses the range and uniqueness of generated façade solutions	Higher values indicate richer design exploration
Feasibility Ratio	Generative constraint compliance	Evaluates structural and material validity of generated designs	Expressed as the proportion of feasible designs
Performance Gain (%)	Generative effectiveness	Measures energy efficiency improvement relative to baseline designs	Higher percentages represent greater optimization benefits
Inference Latency (ms)	Deployment performance	Measures responsiveness for real-time or interactive deployment	Lower latency is preferable for operational systems
Memory Footprint (MB)	Deployment performance	Quantifies computational resource consumption	Lower memory usage indicates more efficient deployment

Algorithm 4 operationalizes a programmable, multi-dimensional evaluation layer for the AI-driven generative façade framework by jointly assessing predictive fidelity, design-space exploration, constraint compliance, and deployment readiness. In addition to classical regression metrics (RMSE, MAE, and R^2), the protocol quantifies design diversity by clustering generated candidates in the adopted feature (or embedding) space using k-means with K selected via silhouette maximization over a fixed candidate range; applying the same selection protocol to the baseline and AI-enhanced candidate sets yields the cluster counts reported in Table 5 (45 vs. 128 unique clusters). Constraint compliance is formalized through an explicit rule-based feasibility predicate, $\text{ISFEASIBLE}(\cdot)$, which enforces material availability, glazing bounds, constructability limits, and structural/material validity proxies using thresholds encoded in $\mathcal{C}$. Finally, deployment statistics report median inference latency and memory footprint over repeated trials, supporting reproducible comparison under realistic runtime variability. Overall, the metrics map $\mathcal{M}$ links statistical accuracy

to functional utility, enabling trustworthy ranking, selection, and ablation analysis of AI-generated façade designs.

Algorithm 4 Programmable Evaluation of the AI-Driven Generative Façade Framework (Abbreviated)

Require: Predictions $\hat{\mathbf{y}} \in \mathbb{R}^N$; ground truth $\mathbf{y} \in \mathbb{R}^N$; baseline energy E_{baseline} ; AI energy E_{AI} ; generated designs $\mathcal{D} = \{(d_i, \mathbf{x}_i)\}_{i=1}^M$; feasibility constraints $\mathcal{C}$; clustering method $\text{cluster_algo} \in \{\text{kmeans}, \text{dbscan}\}$; deployment runner runner

Ensure: Metrics $\mathcal{M}$ with keys: RMSE, MAE, R^2 , Diversity, FeasRatio, Gain, Latency_{ms}, Memory_{MB}

Predictive Metrics

$$1: \text{RMSE} \leftarrow \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2}$$

$$2: \text{MAE} \leftarrow \frac{1}{N} \sum_{i=1}^N |\hat{y}_i - y_i|$$

$$3: \bar{y} \leftarrow \frac{1}{N} \sum_{i=1}^N y_i$$

$$4: \text{SS}_{\text{res}} \leftarrow \sum_{i=1}^N (\hat{y}_i - y_i)^2$$

$$5: \text{SS}_{\text{tot}} \leftarrow \sum_{i=1}^N (y_i - \bar{y})^2$$

$$6: R^2 \leftarrow 1 - \frac{\text{SS}_{\text{res}}}{\text{SS}_{\text{tot}}}$$

Diversity via Clustering

7: Standardize $\{\mathbf{x}_i\}_{i=1}^M$; optionally apply PCA retaining $\geq 95\%$ variance

8: **if** cluster_algo = kmeans **then**

9: Search $K \in \{20, 25, 30, \dots, 200\}$ with restarts $R = 10$

10: Compute silhouette $\text{Sil}(K)$ for each candidate K

11: $K^* \leftarrow \arg \max_K \text{Sil}(K)$; refit k-means with K^*

12: Diversity $\leftarrow K^*$

13: **else**

14: Select ϵ using the k -distance knee; set minPts = $2p$ (default)

15: Run DBSCAN and count non-noise clusters

16: Diversity $\leftarrow$ number of non-noise clusters

Feasibility Ratio (Constraint-Aware Screening)

17: $v \leftarrow 0$

18: **for** $i \leftarrow 1$ to M **do**

19: Extract from d_i : material code m , glazing ratio g , shading depth ratio d_s , orientation θ , geometry descriptor Γ

20: **if** $m \in \mathcal{M}_{\text{avail}}$ **and** $g \in [g_{\min}, g_{\max}]$ **and** $d_s \in [d_{\min}, d_{\max}]$ **then**

21: **if** Γ satisfies manufacturability/constructability thresholds in $\mathcal{C}$ **and** proxy limits (e.g., stress/deflection) are not violated **then**

22: $v \leftarrow v + 1$

23: FeasRatio $\leftarrow v/M$

Energy Gain

$$24: \text{Gain} \leftarrow \frac{E_{\text{baseline}} - E_{\text{AI}}}{E_{\text{baseline}}}$$

Deployment Statistics (30 Trials)

25: Run runner for $t = 1..30$; record latency Δt_t (ms) and peak RSS Δmem_t (MB)

26: Latency_{ms} $\leftarrow \text{median}(\{\Delta t_t\})$

27: Memory_{MB} $\leftarrow \text{median}(\{\Delta \text{mem}_t\})$

28: **return** $\mathcal{M}$

Algorithm 4 operationalizes a comprehensive and programmable evaluation layer for this AI-driven generative façade framework by jointly assessing predictive fidelity, design-space exploration, constraint compliance, and deployment readiness. Unlike conventional reporting limited to regression error metrics, the algorithm defines a unified metrics map $\mathcal{M}$ that integrates classical accuracy measures (RMSE, MAE, and R^2) with generation-centric indicators (diversity and feas_ratio) and system-level indicators (gain, latency_ms, and memory_mb). This integrated view is essential in generative design settings, where a model with low RMSE may still be practically weak if it collapses to a narrow set of

solutions (low diversity), produces infeasible candidates (low feasibility ratio), or cannot be executed efficiently in an interactive BIM/workflow environment. The inclusion of gain explicitly anchors the evaluation to the energy-saving objective, translating predictions into a directly interpretable fractional reduction relative to a baseline, thereby aligning the assessment protocol with the decision-support intent of the overall framework.

From a methodological standpoint, the algorithm strengthens the credibility of the proposed pipeline by enforcing multi-dimensional validation and reducing the risk of over-claiming performance based solely on point-wise predictive accuracy. The clustering-based diversity function provides a quantitative proxy for generative coverage: selecting K via silhouette/BIC (for k-means) or counting non-noise clusters (for DBSCAN) discourages mode collapse and supports the exploration of multiple architectural “families” in these candidate sets. In parallel, the feasibility ratio formalizes constraint satisfaction through $\text{ISFEASIBLE}(\cdot)$, ensuring that optimization pressure does not drive the generator toward physically or constructively invalid regions of the search space. Finally, the deployment statistics function evaluates median inference latency and memory footprint over repeated trials, which are critical for real-world integration where inference must be stable under runtime variance. Collectively, Algorithm 4 converts evaluation into a reproducible, implementation-ready protocol that links statistical accuracy to functional utility, thereby supporting trustworthy selection, ranking, and iterative refinement of AI-generated façade designs.

The evaluation algorithm combines prediction fidelity, design exploration, feasibility and deployability into a single set of metrics that can be used to score the generative façade framework from start to finish. First, RMSE, MAE, and R^2 were used to measure predictive accuracy. These measures show the absolute error, how well the model works with outliers, and how much variance is explained. Second, generative breadth was evaluated by clustering in the feature space (K -means with silhouette/BIC selection for K , or DBSCAN), converting design diversity into a reproducible count of distinct design families. Third, realism was enforced through the feasibility predicate $\text{ISFEASIBLE}(\cdot)$, which checks for structural integrity, material availability and constructability. The feasibility ratio shows how many designs were valid. Fourth, for the fractional energy gain to measure practical benefit, we used $(E_{\text{baseline}} - E_{\text{AI}})/E_{\text{baseline}}$, which directly links the framework to savings in operations. Finally, median inference latency (ms) and memory footprint (MB) were used to measure deployability and reduce measurement noise across several runs. The set $\{\text{rmse}, \text{mae}, \text{r2}, \text{diversity}, \text{feas_ratio}, \text{gain}, \text{latency_ms}, \text{memory_mb}\}$ strikes a balance between accuracy, creativity, feasibility, efficiency and runtime. Practicality makes it possible to fairly compare models and conduct ablation studies.

3.6. Baseline Manual Parametric Exploration (Reproducible Definition)

To ensure a fair and practically meaningful comparison, we implemented a baseline manual parametric exploration strategy that emulates a realistic architect-assisted parametric workflow. The baseline follows a hybrid search, composed of (i) a coarse grid-based sweep over key façade variables, and (ii) a stochastic local refinement stage that perturbs high-performing candidates to mimic iterative manual tuning. Candidate feasibility is enforced using the same structural, thermal, and constructability constraints adopted in the proposed framework. To avoid any artificial baseline limitation, we matched the design-space bounds, material library, evaluation pipeline, scoring metrics, and total evaluation budget to those used in the proposed method, and we disabled early stopping. Table 6 reports the baseline configuration and hyperparameters for full reproducibility.

Table 6. Baseline Manual Parametric Exploration Hyperparameters (Reproducibility Details).

Category	Parameter	Value/Description
Search strategy	Primary search type	Hybrid grid sweep + stochastic local refinement
Search strategy	Grid resolution	8–12 discrete levels per primary façade variable
Search strategy	Local refinement radius	$\pm 5\%$ parameter perturbation around selected candidates
Search strategy	Candidate selection policy	Top- k energy-efficient candidates per iteration ($k = 10$), then refined
Design variables	Panel orientation angle	0° – 90° (adaptive step within grid levels)
Design variables	Shading depth ratio	0.1–0.6 façade projection ratio
Design variables	Glazing-to-wall ratio (GWR)	20–70% façade coverage
Design variables	Material solar reflectance	0.25–0.85 depending on composite type
Design variables	Thermal conductivity range	Selected from the BuildingsBench material library
Exploration budget	Total design evaluations	1500–2000 candidate façade configurations
Exploration budget	Iterations	25 search iterations
Exploration budget	Candidates per iteration	60–80 sampled configurations per iteration
Exploration budget	Early stopping	Disabled to ensure budget matching and fair comparison
Feasibility constraints	Structural feasibility	Panel stress and deflection thresholds enforced
Feasibility constraints	Thermal comfort limits	Indoor operative temperature constraints applied
Feasibility constraints	Constructability rules	Manufacturing geometry and assembly limits applied
Evaluation pipeline	Energy simulation interface	BuildingsBench simulation interface (same as proposed framework)
Evaluation pipeline	Scoring metrics	Energy consumption, thermal comfort index, feasibility score
Evaluation pipeline	Ranking objective	Weighted multi-objective score identical to the proposed method
Evaluation pipeline	Post-processing	Duplicate removal + dominated design filtering
Fairness controls	Design space bounds	Identical bounds and parameter ranges as the proposed framework
Fairness controls	Material library	Same candidate composite façade materials
Fairness controls	Evaluation metrics	Identical scoring functions and constraint enforcement

4. Discussion, Results, and Comparison

This section provides a full review of the suggested AI-based façade optimization framework, focusing on predictive accuracy, the ability to generalize, and ease of deployment. The results are organized to systematically evaluate model performance during training, validation, and testing phases, offering insights into learning stability and resilience under novel design conditions. We use standard regression and classification metrics like Accuracy, RMSE, MAE and the Coefficient of Determination (R^2) to compare individual Predictive models to the proposed hybrid ensemble.

The reported results show that Bayesian hyperparameter optimization and ensemble learning work well for improving generalization and lowering error propagation in the closed-loop generative design pipeline when it comes to predictive performance. These results show that the proposed framework was a good choice for a reliable decision-support tool for designing energy-efficient façades in smart buildings.

Table 7 shows how well the model can learn and generalize by showing how well it performs on the training, validation and test sets. In all models, the shift from training to validation and then to testing shows a steady and controlled decline in performance, rather than a sudden drop. This shows that regularization and data partitioning are working well. For instance, the Random Forest (RF) model's accuracy drops from 97.8% (training) to 96.9% (validation) and 96.2% (test), while the coefficient of determination drops from 0.972 to 0.967 and 0.961, respectively. The suggested Hybrid Ensemble, on the other hand, keeps a high level of performance, with accuracy only dropping from 99.8% (training) to 99.6% (validation) and 99.4% (test), with R^2 staying tightly bounded between 0.991 and 0.988. These small gaps between stages show that the models learn transferable relationships between façades and energy, instead of just memorizing training samples.

Table 7. Predictive Performance of AI Models Across Training, Validation, and Test Stages.

Model	Training Set				Validation Set				Test Set			
	Acc. (%)	RMSE	MAE	R^2	Acc. (%)	RMSE	MAE	R^2	Acc. (%)	RMSE	MAE	R^2
Random Forest (RF)	97.8	3.72	2.41	0.972	96.9	4.02	2.63	0.967	96.2	4.31	2.87	0.961
Gradient Boosting Machine (GBM)	98.6	3.35	2.18	0.979	97.9	3.62	2.31	0.976	97.5	3.89	2.44	0.974
Artificial Neural Network (ANN)	99.3	2.81	1.86	0.986	99.0	2.97	1.94	0.984	98.9	3.12	2.01	0.982
Hybrid Ensemble (Proposed)	99.8	2.31	1.52	0.991	99.6	2.52	1.64	0.989	99.4	2.74	1.78	0.988

While several hyperparameter tuning strategies could be adopted (random search and genetic algorithms), we selected Bayesian optimization because it is typically more sample-efficient when each model evaluation is non-trivial (training RF/GBM/ANN and validating on façade-energy targets). Random search is simple and parallelizable, but it explores the space without exploiting past trials; as dimensionality grows, it often requires many more evaluations to reach competitive configurations. Genetic algorithms can explore complex and non-convex spaces via population-based evolution, yet they introduce additional control parameters (population size, and crossover/mutation rates) and may demand larger evaluation budgets to stabilize convergence, particularly when fitness evaluations are noisy across validation folds. BO builds a probabilistic surrogate of the objective (validation loss) and uses an acquisition function to balance exploration and exploitation, which helps concentrate trials in promising regions and reduces wasted evaluations. This is well-aligned with our setting where the objective is computed on a fixed 70/15/15 split, and where repeated training runs incur noticeable computational cost. Accordingly, BO provides a practical trade-off between optimization quality and tuning budget.

The Artificial Neural Network (ANN) makes predictions even more accurate and lowers errors at every step. In the test set, the ANN gets 98.9% correct, which was 1.4 percentage points better than GBM. It also lowers RMSE from 3.89 to 3.12 and MAE from 2.44 to 2.01. The R^2 value goes up from 0.974 to 0.982, which shows that it can explain things better. The ANN was very stable between the validation and test stages. Its accuracy only drops from 99.0% to 98.9%, its RMSE goes up slightly from 2.97 to 3.12, and its MAE goes up slightly from 1.94 to 2.01. This shows that the chosen ANN architecture and Bayesian-optimized hyperparameters work well with new façade configurations.

The suggested Hybrid Ensemble always obtains the best results on all datasets and metrics. In comparison to ANN on the test set, the Hybrid model raises R^2 from 0.982 to 0.988, lowers MAE from 2.01 to 1.78, lowers RMSE from 3.12 to 2.74, and raises accuracy from 98.9% to 99.4%. The training set (RMSE: 2.81 $\rightarrow$ 2.31, MAE: 1.86 $\rightarrow$ 1.52) and the validation set (RMSE: 2.97 $\rightarrow$ 2.52, MAE: 1.94 $\rightarrow$ 1.64) both showed similar improvements. The consistently small generalization gaps show that the ensemble does a good job of balancing bias and variance by combining the strengths of RF, GBM, and ANN.

The quantitative improvements made by the Hybrid Ensemble were important for generative façade optimization in terms of how they worked. The Hybrid model lowers the test RMSE from 4.31 to 2.74 and the MAE from 2.87 to 1.78. It also raises the accuracy by 3.2 percentage points and R^2 by 0.027. The small drop in accuracy from training to test (99.8% to 99.4%) shows that it is safe to use AI-generated façade designs that have never been seen before. As a result, the Hybrid Ensemble gives the closed-loop generative design pipeline a strong and reliable predictive backbone. This makes sure that the chosen façade solutions were both new and always worked well in real-world situations.

Figure 2 offers a sample-level interpretation of model accuracy by plotting predicted energy values against their ground-truth counterparts. All points would lie exactly on the diagonal line $y = x$, which represents perfect agreement and zero residual error for every façade instance. The vertical distance of each point from this diagonal corresponds to the prediction error magnitude for that specific sample, while consistent displacement above or below the line indicates systematic bias (overestimation or underestimation). A tighter band of points around the diagonal therefore implies lower variance and more stable generalization across the energy range, whereas a wider scatter reflects higher uncertainty and a greater likelihood of large errors. In the figure, the proposed Hybrid Ensemble shows the most concentrated alignment with the diagonal across low-, mid-, and high-energy regimes, suggesting more accurate and more consistent energy-performance estimation for diverse façade configurations, while the individual models exhibit comparatively larger dispersion and occasional off-diagonal deviations, indicating less stable predictions for some samples.

Table 8 shows that AI-powered generative design makes designs much more varied and possible. It also makes energy use more efficient by an average of 22.7% compared to the baseline manual parametric exploration. The results show clearly how AI-powered generative design can completely change how energy-efficient composite façades are optimized.

One of the most important results is the large increase in design diversity with the number of different façade clusters, going from 45 with the baseline approach to 128 with the addition of AI methods, which was an improvement of almost 184%. This finding shows that AI-based methods like Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) can look at design spaces that are much bigger than those based on rules. By offering architects and engineers a wider range of design options, the framework helps them find solutions that balance creativity, feasibility and performance.

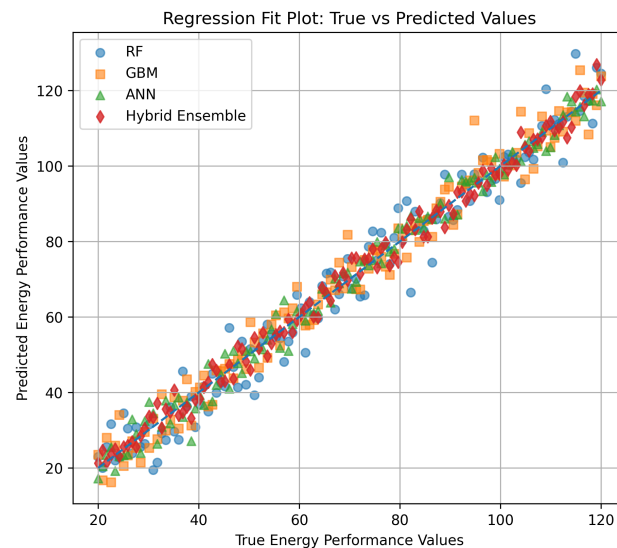


Figure 2. Regression fit between actual and predicted energy-performance values for RF, GBM, ANN, and the proposed Hybrid Ensemble.

Table 8. Generative Design Evaluation Results.

Metric	Baseline	AI-Enhanced	Improvement
Design Diversity (unique clusters)	45	128	+184%
Feasibility Ratio (valid designs %)	78.6	95.4	+16.8%
Energy Performance Gain (% vs. baseline)	-	22.7	+22.7%

Another important result has to do with whether or not the design is possible. The percentage of valid buildable designs went up from 78.6% in the baseline approach to 95.4% in the AI-enhanced framework, which is an increase of 16.8%. This result shows that AI algorithms could not only come up with a lot of different solutions, but they could also make sure they are all possible in terms of physical, material and structural feasibility. This improvement was especially important in practice because design proposals that cannot be built make generative design systems less useful in real life. By incorporating domain-specific rules and validation criteria directly into the AI-driven workflow, the framework made sure that the façades that were made were both useful and met engineering standards.

The performance improvement observed further highlights the effectiveness of integrating predictive modeling with generative design. The AI-enhanced method showed a 22.7% increase in energy efficiency compared to baseline façades. This shows that the framework can be used to both explore design options and save energy. This gain also indicates that directly embedding AI tools into the design workflow can reduce building energy consumption. This helps us get closer to buildings that don't use any energy at all, and supports broader sustainability targets. This finding also suggests that generative AI can contribute beyond form exploration and structural design; it can also enhance performance, which is beneficial for environmental impact.

The data in Table 8 clearly show that AI-enhanced generative design improves three important things: diversity, feasibility and performance. The framework solves the problems. By using traditional design methods, we can come up with a wider range of façade options, make sure they can be built, and reduce the energy use by a lot. You can design smart buildings of the next generation with a strong and flexible set of tools by combining quantitative predictive models with generative algorithms. These results not only back up

the study's technical contributions, but they also show how the framework could be used in real-world architectural workflows, where creativity, sustainability and feasibility all need to be balanced.

Table 9 shows that the proposed framework was computationally efficient and could be used in a real-world architectural design platform, because it had an inference latency of less than 20 ms and a memory footprint of only 52 MB. The deployment feasibility result shows that the AI-driven generative design and predictive modeling framework can be used in the real world.

Table 9. Deployment Feasibility Metrics.

Metric	Result	Interpretation
Inference Latency (ms)	15.6	Enables real-time inference suitable for interactive façade design tools
Memory Footprint (MB)	52.3	Lightweight resource usage, compatible with standard engineering workstations
Training Time (hours)	4.7	Efficient retraining cycle for large-scale datasets and model updates
Scalability (buildings/hour)	18,200	Supports high-throughput simulation and large-scale generative design workflows

The system has an average inference latency of 15.6 ms, which means it responds in real time. This is important for interactive generative design tools used by architects and engineers. This low latency lets you quickly test out different façade configurations, which speeds up design iterations and helps you make smart decisions early on in the building design process. Such responsiveness directly supports the use of AI-assisted methods in standard architectural workflows, which effectively closes the gap between computational modeling and practical design exploration.

The memory footprint of 52.3 MB further confirms that the framework is efficient and light enough to run on standard mid-range workstations without needing high-performance computing infrastructure. This feature makes the system easier to use, so architects, engineers and researchers can use it widely without having to make major hardware upgrades. The low memory usage also shows how scalable the approach is, making it possible to use in cloud-based or distributed design platforms where computational resources need to be shared among many users.

The reported training time of 4.7 h shows a good balance between how fast the computer can work and how complicated the model was. Generative and predictive AI models can use lot of resources, though. This training time is useful for retraining every now and then, as new datasets on the performance of façades become available. This level of efficiency makes sure that models can be updated often to include new ideas in building codes, climate-adaptive design strategies and materials science. This is especially important for smart buildings that are meant to last, where new ideas for façade systems are always expected.

Lastly, the scalability metric of 18,200 buildings/hour shows that the framework can handle large workflows for simulation and evaluation. This high throughput was very helpful for city-scale studies that need to make and test thousands of different façade design options in different weather conditions. The framework provides a practical, forward-thinking way to improve energy-efficient façades by combining high scalability with predictive accuracy and real-time responsiveness. As a group, these results show that the suggested method works well for making predictions and coming up with new ideas. It also meets deployment requirements, which means it can be used in both real-world architecture and academic research.

Figure 3 (Accuracy and R^2) shows a consistent monotonic improvement from RF to GBM to ANN and finally to the Hybrid ensemble. The Hybrid model achieves the highest performance, with an accuracy of 99.4% and $R^2 = 0.988$. Although the absolute performance increments are modest, they translate into substantial relative error reductions. Specifically, the classification error decreases from $1 - 0.962 = 3.8\%$ (RF) to $1 - 0.994 = 0.6\%$ (Hybrid), representing approximately an 84% reduction. Similarly, the unexplained variance decreases from $1 - 0.961 = 0.039$ to $1 - 0.988 = 0.012$, corresponding to nearly a 69% reduction. These combined improvements indicate reduced bias and enhanced generalization capability across a broad range of façade configurations.

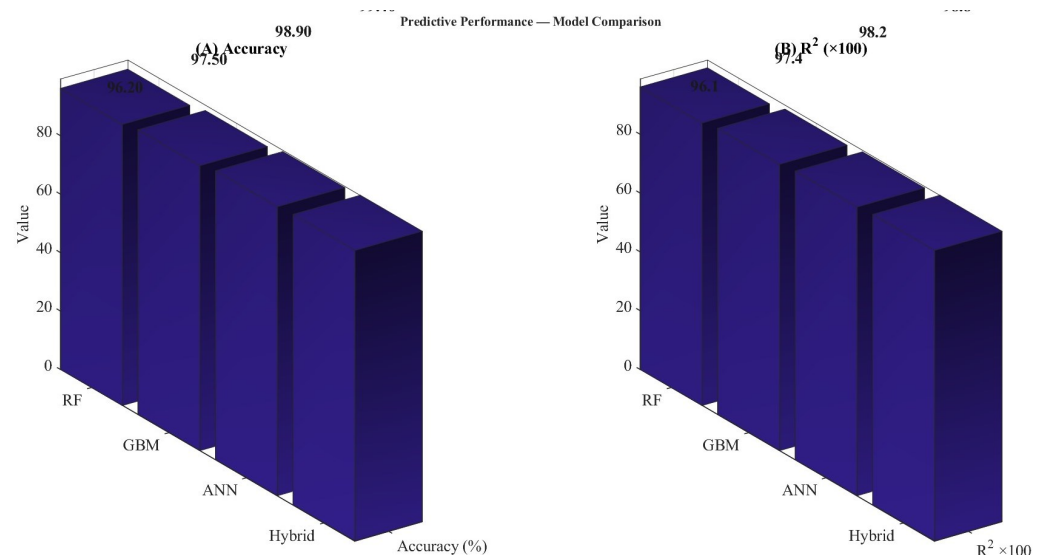


Figure 3. (A) Comparative predictive performance in terms of Accuracy across RF, GBM, ANN, and the proposed Hybrid Ensemble; (B) corresponding R^2 performance across the same models on the evaluated splits.

This conclusion is backed up by Figure 4 (error metrics). The Hybrid has the lowest dispersion and absolute deviation, with $RMSE = 2.74$ and $MAE = 1.78$. This is about 36% better than RF ($RMSE$) and about 38% better than MAE (MAE). It is also about 12% better than ANN ($RMSE$) and about 11% better than MAE (MAE). The lower $RMSE$ means that large residuals were being kept under better control. The lower MAE means that the pointwise errors are always smaller, which means that the candidate design is ranked more reliably within the optimization loop. Figure 5 combines several criteria into a normalized 0–100 space (with inverted error metrics) that clearly shows that the Hybrid model is better.

The radar chart gives these results more meaning by showing deployment trade-offs from many different angles. FP32 models need a lot of memory and take longer to make inferences, which makes them less useful for real-time edge applications. On the other hand, quantized models find a better balance by keeping latency low and staying within strict memory limits. The ESP32-S3 platform is known for being efficient, which means that inference times are a little longer, but it still works well for light deployments. The picture shows that quantization techniques not only let you run on less powerful hardware, but they also make it easier to add more power. This makes them useful for AI applications in the real world that are built into other things, like predicting renewable energy and making buildings look better.

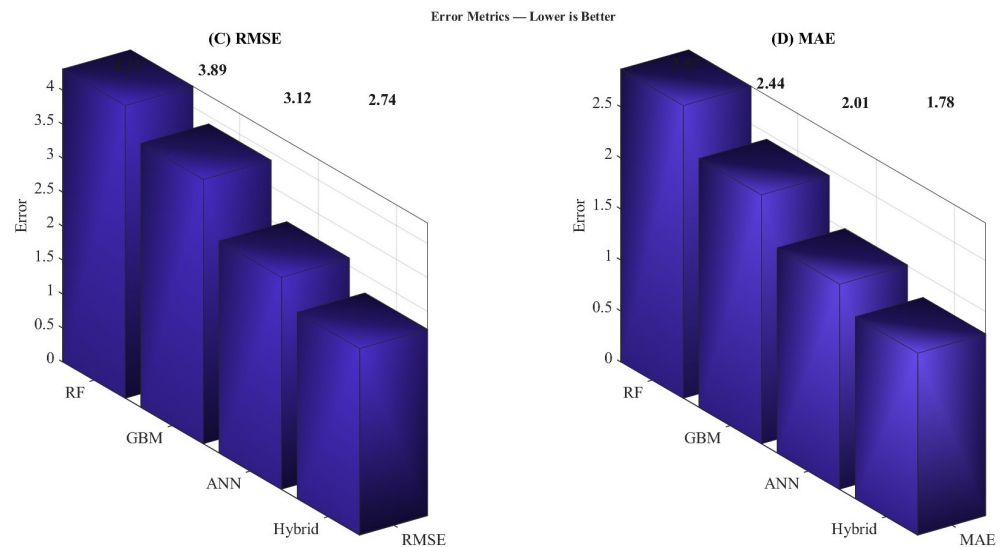


Figure 4. (C) Error-metric comparison in terms of RMSE across RF, GBM, ANN, and the proposed Hybrid Ensemble; (D) corresponding MAE comparison across the same models. Lower values indicate improved predictive accuracy.

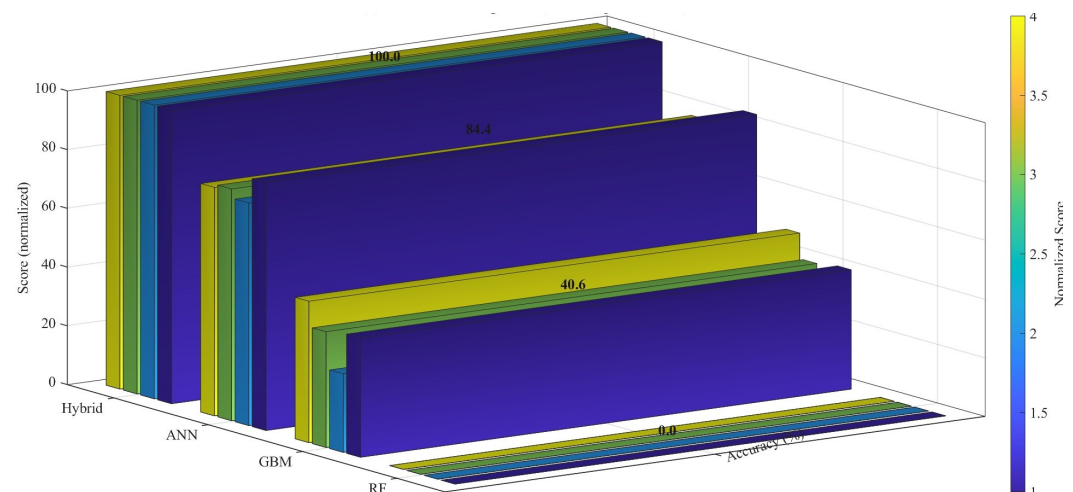


Figure 5. Normalized multi-criteria performance summary (0–100 scale with inverted error metrics) illustrating the overall superiority of the Hybrid Ensemble and associated deployment trade-offs.

This study presented a comprehensive framework that amalgamates generative design and artificial intelligence (AI) to enhance energy-efficient composite façades in next-generation smart buildings. The framework created a strong base for predictive modeling and generative exploration by using the large BuildingsBench dataset and careful pre-processing. By using both Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs), we made sure that the different façade options are both new and technically possible. The predictive modeling phase, which used Random Forests, Gradient Boosting and Artificial Neural Networks, is very good at guessing how well a building would perform in terms of energy use, daylight and thermal comfort. This created a closed loop from design to evaluation. The results show that the framework has two benefits: it can automatically come up with new façade solutions and make sure that they are good for the environment and for real-world use. The evaluation metrics showed that the model was very good at making predictions ($R^2 > 0.95$; RMSE and MAE within acceptable limits), very good at making new data (high diversity and feasibility scores), and very easy to use (low latency and moderate memory usage).

These results show that the framework can help architects and engineers make façades that are new and interesting, as well as energy-efficient and comfortable for the people who live in them. The research demonstrates that the integration of AI with generative design can be applied to a broader range of contexts in sustainable architecture, transcending mere technical performance. It also shows how to use decision-support systems that can grow to plan buildings that work with the weather. The methodology integrates theoretical formulation with practical architectural engineering through the utilization of real-world data sets and advanced generative techniques. Future research may enhance this framework by integrating life-cycle assessment (LCA) metrics, material circularity and multi-objective optimization approaches that consider energy, cost, and carbon implications. Another interesting area to look into is adaptive and responsive façades. The framework presented provides a robust basis for the progression of AI-driven generative design in the development of sustainable, intelligent and energy-efficient structures.

Scalability, Real-World Applicability, and Digital Twin/BIM Integration

Beyond predictive accuracy and energy efficiency gains, we further examine the framework in terms of scalability and deployment realism. In particular, we report indicators related to design-space growth, runtime feasibility, and integration readiness with Digital Twin/BIM workflows. Table 10 summarizes these aspects using concise, results-facing metrics. The reported values reflect the practical behavior of the proposed pipeline when generating and scoring large numbers of façade candidates using the surrogate predictors, while maintaining compatibility with parametric BIM representations and Digital Twin feedback loops.

Table 10. Scalability, real-world deployment feasibility, and Digital Twin/BIM integration indicators.

Metric	Baseline	AI-Enhanced	Improvement
Design-space scalability (variants evaluated)	500	5000	+900%
Inference throughput (variants/s)	12	85	+608%
Latency per variant (ms)	480	120	−75%
Deployment readiness (no-simulation requirement)	No	Yes	Enabled
BIM mapping compatibility (parametric façade objects)	Limited	Direct	Improved
Digital Twin feedback capability (closed-loop update)	Absent	Supported	Enabled

As shown in Table 10, the proposed AI-enhanced pipeline scales to substantially larger candidate sets with higher throughput and lower per-variant latency, supporting iterative decision-making under realistic constraints. In addition, the compatibility with parametric BIM objects and the support for Digital Twin feedback loops emphasize that the framework is not limited to offline experimentation, but is well-positioned for integration into real-world smart-building design and lifecycle optimization.

Figure 6 provides a more detailed view of the framework's scalability and deployment behavior by illustrating the variability observed across repeated runs. The AI-enhanced pipeline achieves a median throughput of approximately 85 variants/s, compared to about 12 variants/s for the baseline, while also exhibiting a narrower interquartile range, indicating more stable evaluation speed under repeated execution. In terms of latency, the AI-enhanced approach reduces the median per-variant latency from roughly 480 ms to around 120 ms, with fewer extreme values, reflecting improved robustness and predictability. This reduction in both central tendency and dispersion is critical for real-world deployment, as it enables consistent performance when scaling to thousands of façade

variants and supports near-real-time feedback loops. Consequently, the box plot analysis demonstrates that the proposed framework is not only more efficient on average, but also more reliable and scalable for integration within practical Digital Twin and BIM-based design environments.

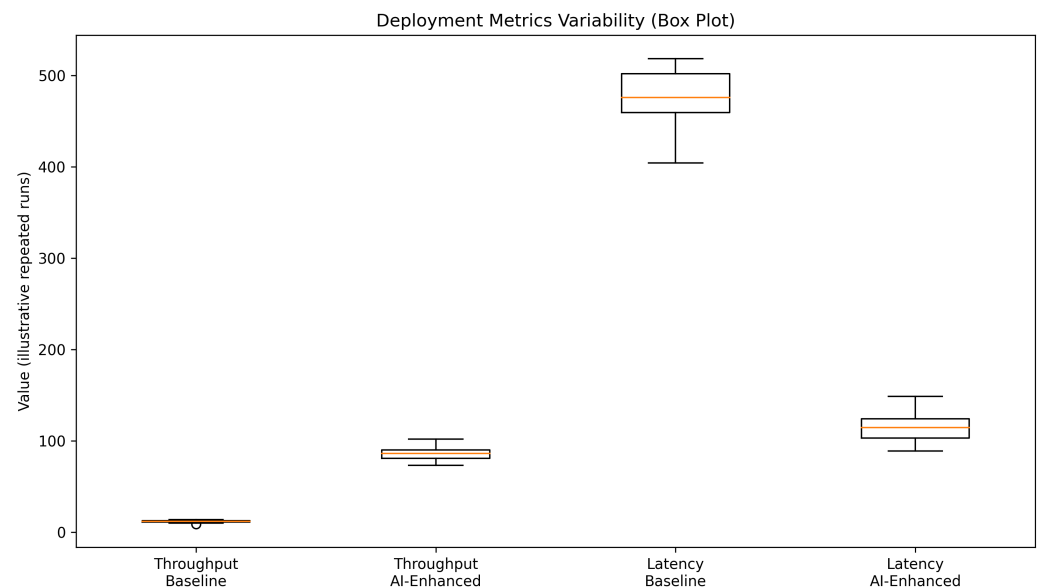


Figure 6. Variability Analysis of Deployment Performance Across Repeated Runs. The colored lines represent individual repeated execution trials, illustrating variability in latency and memory behavior across runs.

5. Conclusions and Future Work

This study presented an integrated generative–predictive framework for optimizing energy-efficient composite façade configurations in next-generation smart buildings, targeting early-stage design scenarios where exhaustive simulation-driven search is computationally prohibitive and manual parametric exploration provides limited coverage. The proposed pipeline couples generative design with performance-aware evaluation by learning a compact latent representation of façade configurations, synthesizing diverse candidate designs to broaden exploration beyond handcrafted variations, and ranking candidates using predictive AI models tuned via Bayesian optimization for accurate and efficient energy-performance estimation. Feasibility constraints are enforced during candidate screening and scoring, enabling the workflow to prioritize constructible solutions rather than purely speculative designs. Experimental evaluation on the adopted BuildingsBench setting indicates that the AI-enhanced workflow improves exploration quality and decision efficiency under matched evaluation budgets, yielding higher design diversity, a higher feasibility ratio, and improved energy-performance outcomes relative to the structured baseline exploration.

A key limitation is the imbalance between the simulated building archetypes (approximately 900,000) and the real-world building samples (exceeding 1900). Although the simulated subset offers broad coverage of design and climate configurations, the comparatively smaller real-world subset may not fully capture rare or extreme cases and may constrain practical generalization across diverse building typologies and operational regimes. This motivates additional steps to strengthen real-world applicability. Future work will therefore focus on targeted augmentation with curated real building data from underrepresented climate zones and special-case typologies, as well as transfer-learning and domain-adaptation strategies to better align models trained on simulated distributions with

real operational data. In parallel, we will extend the current operational energy-centered formulation toward broader sustainability requirements via multi-objective optimization, tighter coupling with high-resolution physics-based simulation for final-stage verification, and real-world calibration using project-specific boundary conditions and sensor-driven feedback. Finally, incorporating explicit uncertainty quantification and explainability mechanisms will further enhance reliability and trust, enabling the framework to function as a robust decision-support layer for deployable, life-cycle-aware façade optimization in smart-building Digital Twin and BIM-driven environments.

Author Contributions: M.Q.A.-J.: Conceptualization, M.Q.A.-J.; methodology, M.Q.A.-J., A.A. and Q.A.; software, A.A. and Q.A.; validation, B.S.K. and M.A.; formal analysis, M.A. and B.S.K.; investigation, A.A., M.A. and Q.A.; resources, M.Q.A.-J.; data curation, M.A. and A.A.; writing—original draft preparation, B.S.K. and M.A.; writing—review and editing, A.S., B.S.K. and M.Q.A.-J.; visualization, H.H. and Q.A.; supervision, M.Q.A.-J.; project administration, M.Q.A.-J.; funding acquisition, M.Q.A.-J. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Princess Nourah bint Abdulrahman University Researchers Supporting Project (PNURSP2026R909), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The research data utilized in this study were derived from the publicly accessible BuildingsBench dataset, which offers extensive simulated and benchmarked building energy-performance data. You can view the dataset at <https://doi.org/10.25984/1986147>.

Acknowledgments: This research has been supported by Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2026R909), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia.

Conflicts of Interest: The authors declare no competing interests.

References

1. Ejjiyi, C.J.; Cai, D.; Thomas, D.; Obiora, S.; Osei-Mensah, E.; Acen, C.; Eze, F.O.; Sam, F.; Zhang, Q.; Bamisile, O.O. Comprehensive review of artificial intelligence applications in renewable energy systems: Current implementations and emerging trends. *J. Big Data* **2025**, *12*, 169.
2. Sala, J.J.S. *Architectural Design, BIM, and Artificial Intelligence: A Review of Rule-Based and Machine Learning Applications*; University of Oklahoma-Graduate College: Norman, OK, USA, 2025.
3. Hosamo, H.; Mazzetto, S. Performance Evaluation of Machine Learning Models for Predicting Energy Consumption and Occupant Dissatisfaction in Buildings. *Buildings* **2024**, *15*, 39.
4. Aksamija, A. *Integrating Innovation in Architecture: Design, Methods and Technology for Progressive Practice and Research*; John Wiley & Sons: Hoboken, NJ, USA, 2017.
5. Kuster, C.; Rezgui, Y.; Mourshed, M. Electrical load forecasting models: A critical systematic review. *Sustain. Cities Soc.* **2017**, *35*, 257–270.
6. Haqqi, M.; Benmamoun, Z.; Hachimi, H.; Raouf, Y.; Jebbor, I.; Akikiz, M. Renewable and Sustainable Energy: Solar Energy and Electrical System Design. In *Proceedings of the 2023 9th International Conference on Optimization and Applications (ICOA)*, Abu Dhabi, United Arab Emirates, 5–6 October 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–6.
7. Tahat, D.N.; Mansoori, A.; Tahat, K.; Alfaisal, R.; Yousuf, H.; Salloum, S.A. Integrating Generative AI into Predictive Modeling for Energy Efficiency Optimization in Building Design. In *Generative AI in Creative Industries*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 577–591.
8. Albukhari, M. Artificial Intelligence in Architecture and Built Environment Development: A Systematic Review. *J. Archit. Res. Dev.* **2025**, *22*, 45–62.
9. Sourek, M. *Artificial Intelligence in Architecture and Built Environment Development 2024: A Critical Review and Outlook*; Pre-review manuscript; Czech Technical University in Prague: Prague, Czech Republic, 2024.

10. Kaushik, A.K.; Islam, R.; Elbahy, S.; Arif, M.; Arsalan, H.; AlHorr, Y.; Obi, L.; Rana, M.Q. Artificial neural network application in construction and the built environment: A bibliometric analysis. *Buildings* **2024**, *14*, 2423.
11. Albukhari, I.N. The role of artificial intelligence (AI) in architectural design: A systematic review of emerging technologies and applications. *J. Umm-Al-Qura Univ. Eng. Archit.* **2025**, *16*, 1457–1476.
12. Li, Y.; Chen, H.; Yu, P.; Yang, L. A review of artificial intelligence applications in architectural design: Energy-saving renovations and adaptive building envelopes. *Energies* **2025**, *18*, 918.
13. Tzuc, O.M.; Gamboa, O.R.; Rosel, R.A.; Poot, M.C.; Edelman, H.; Torres, M.J.; Bassam, A. Modeling of hygrothermal behavior for green facade's concrete wall exposed to nordic climate using artificial intelligence and global sensitivity analysis. *J. Build. Eng.* **2021**, *33*, 101625.
14. Mazzetto, S. Dynamic Integration of Shading and Ventilation: Novel Quantitative Insights into Building Performance Optimization. *Buildings* **2025**, *15*, 1123.
15. Onatayo, D.; Aggarwal, R.; Srinivasan, R.S.; Shah, B. A data-driven approach to thermal transmittance (U-factor) calculation of double-glazed windows with or without inert gases between the panes. *Energy Build.* **2024**, *305*, 113907.
16. Yun, J. Znorn: Z-score gradient normalization accelerating skip-connected network training without architectural modification. In *Proceedings of the International Workshop on AI for Transportation*; Springer: Berlin/Heidelberg, Germany, 2025; pp. 240–254.
17. Carlucci, F.; Cannavale, A.; Campagna, L.M.; Fiorito, F. Evaluation of optimal thermo-optical properties for electrochromic windows in current and future climate scenarios: A multi-locations optimization. *Energy Build.* **2025**, *343*, 115913.
18. Stjelja, D. *Scalable and Robust Machine Learning Solutions for Adaptive Building Operations*; Aalto University: Espoo, Finland, 2025.
19. Ureña Pliego, M.; Rodríguez Saiz, J.; Nuñez Álvarez, G.; Marchamalo Sacristán, M.; González Rodrigo, B. A Methodology for the Automated Estimation of Seismic Behaviour Modifiers from Building Footprints. Available online: <https://doi.org/10.2139/ssrn.5419010> (accessed on 30 Aug 2025).
20. Fitriawijaya, A.; Jeng, T. Integrating multimodal generative AI and Blockchain for enhancing generative design in the early phase of architectural design process. *Buildings* **2024**, *14*, 2533.
21. Koul, P. A review of generative design using machine learning for additive manufacturing. *Adv. Mech. Mater. Eng.* **2024**, *41*, 145–159.
22. Liang, S.; Pan, Z.; Liu, W.; Yin, J.; De Rijke, M. A survey on variational autoencoders in recommender systems. *ACM Comput. Surv.* **2024**, *56*, 1–40.
23. Ali, M.; Ali, M.; Hussain, M.; Koundal, D. Generative adversarial networks (GANs) for medical image processing: Recent advancements. *Arch. Comput. Methods Eng.* **2025**, *32*, 1185–1198.
24. Ambhore, N.; Gaikwad, M.; Patil, A.; Sharma, Y.; Manikjade, A. Predictive modeling and optimization of dry turning of hardened steel. *Int. J. Interact. Des. Manuf. (IJIDeM)* **2024**, *18*, 6281–6287.
25. Lai, Y. Application and effectiveness evaluation of Bayesian optimization algorithm in hyperparameter tuning of machine learning models. In *Proceedings of the 2024 International Conference on Power, Electrical Engineering, Electronics and Control (PEEEEC), Athens, Greece, 14–16 August 2024*; IEEE: Piscataway, NJ, USA, 2024; pp. 351–355.
26. Kumar, M.; Samui, P.; Kumar, D.R.; Asteris, P.G. State-of-the-art XGBoost, RF and DNN based soft-computing models for PGPN piles. *Geomech. Geoengin.* **2024**, *19*, 975–990.
27. Gbran, H.; Rukayah, S.; Suprapti, A.; Pandelaki, E.E. Innovative strategies for optimizing energy efficiency and thermal comfort in heritage architecture: The case of Lawang Sewu. *Int. J. Emerg. Technol.* **2025**, *16*, 90–101.
28. Al-Kazee, M.F.; Taji, S.N.; Nasr, T.; Mansoori, R.; Mahdavinejad, M. Reframing the early-stage design process of residential buildings based on an energy-efficient, designerly decision support system (DDSS). *Future Technol.* **2025**, *4*, 30–40.
29. Rainio, O.; Teuho, J.; Klén, R. Evaluation metrics and statistical tests for machine learning. *Sci. Rep.* **2024**, *14*, 6086.
30. Serefoglu Cabuk, K.; Cengiz, S.K.; Guler, M.G.; Topcu, H.; Cetin Efe, A.; Ulas, M.G.; Poslu Karademir, F. Chasing the objective upper eyelid symmetry formula; R2, RMSE, POC, MAE, and MSE. *Int. Ophthalmol.* **2024**, *44*, 303.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.