An automated and explainable machine learning framework for multi-objective energy performance evaluation and early-stage design support of indoor arenas

Engineering Applications of Artificial Intelligence 176 (2026) 114753

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

An automated and explainable machine learning framework for multi-objective energy performance evaluation and early-stage design support of indoor arenas

Yicheng Wang [a,b], Peijun Lu [c], Yimin Sun [a,d,*], Mauro Berta [b], Hao Wang [a]

[a] School of Architecture, State Key Laboratory of Subtropical Building Science, South China University of Technology, Guangzhou, China
[b] Department of Architecture and Design, Politecnico di Torino, Torino, Italy
[c] School of Architecture, Tsinghua University, Beijing, China
[d] Guangdong Research Center of Sustainable Architecture and Urban Design Engineering Technology, Guangzhou, China

ARTICLE INFO

Keywords:
Energy performance
Indoor arena
Feature importance
Explainable machine learning
Automated performance pipeline

ABSTRACT

Indoor arenas, as large-span public buildings, entail high construction and operational costs, particularly in energy use. The pursuit of iconic forms often overshadows rational, performance-driven design, while their complex geometry and diverse functions make energy performance evaluation challenging at early design stages. This study develops an integrated framework that combines parametric simulation, a stacked-ensemble machine learning model, and SHAP (SHapley Additive exPlanations)-based interpretation to relate geometric configurations to the energy performance of indoor arenas. An automated performance evaluation pipeline integrates multiple evaluation modules, enabling automated data transfer, processing, and large-scale analysis. Using partially stratified and Latin Hypercube sampling, 9667 design scenarios were generated across three representative climate zones. Five indicators in three climate zones were evaluated, including energy use intensity (EUI) for cooling, heating, lighting, and total energy, and energy yield intensity (EYI). The ensemble models achieved high predictive accuracy (average $R^2 > 0.95$), providing a reliable basis for interpretation. SHAP analysis reveals the varying contributions of design parameters to each performance indicator across climates, as well as the nonlinear interaction mechanisms between complex roof morphology and overall energy behavior. These findings establish climate-based strategies for balancing energy consumption and generation, and demonstrate an integrated and interpretable control workflow for performance-informed design and analytical support for performance-oriented design decisions in complex-built systems.

1. Introduction

Indoor arenas, as a representative type of large-span public building alongside airports, exhibition halls, and convention centers, are widely recognized for their high construction costs and substantial operational energy expenditures (Flyvbjerg, 2011) (Zimbalist, 2016) (Alm, 2012). Since the building sector accounts for more than one-third of global energy use and carbon emissions (International Energy Agency (IEA), 2021), improving the energy performance of such facilities represents a significant opportunity for achieving carbon neutrality goals. These buildings are characterized by vast interior volumes, complex curved roof geometries, and variable operational demands ranging from competitive events to daily public fitness activities (Qian et al., 2025) (Qian et al., 2024). The pursuit of iconic architectural expression often overshadows performance-driven design considerations, while geometric complexity makes systematic energy evaluation particularly challenging.

The early design stage offers the greatest potential to influence long-term energy performance (Hassan and El-Rayes, 2024) (Depecker et al., 2001) (Granadeiro et al., 2013). At this phase, fundamental decisions regarding building massing, orientation, roof geometry, and fenestration are made while design flexibility remains high. Once established, these morphological parameters constrain both consumption patterns and renewable energy harvesting capacity for decades of operation. Research has demonstrated that morphological optimization can achieve energy savings of up to 60% in Mediterranean climates (Ciardiello

* Corresponding author. School of Architecture, South China University of Technology, Guangzhou, 510640, China.
E-mail addresses: yicheng.wang@polito.it (Y. Wang), lupj@mail.tsinghua.edu.cn (P. Lu), arymsun@scut.edu.cn (Y. Sun), mauro.berta@polito.it (M. Berta), wh875568319@gmail.com (H. Wang).

https://doi.org/10.1016/j.engappai.2026.114753
Received 27 October 2025; Received in revised form 26 February 2026; Accepted 5 April 2026
Available online 7 April 2026

et al., 2020)–20% across other climate conditions (Fang and Cho, 2019). Despite this potential, early-stage performance assessment remains difficult due to rapid iteration cycles, geometric uncertainty, competing objectives, and fragmented analytical tools (Li et al., 2018) (Hensel, 2013).

Indoor arenas present challenges for early-stage performance-oriented design. Unlike conventional rectangular buildings, arena roof systems involve diverse morphological types including domes, barrel vaults, saddle forms, and multi-slope configurations, along with variable skylight distributions and complex spatial topologies. Previous studies have confirmed strong associations between arena morphology and energy performance through field measurements (Huang et al., 2019), combined field-simulation analysis (Zhang et al., 2025a), and parametric experiments (Dong et al., 2024) (Lyu et al., 2025). These studies report that variations in roof form and envelope configuration correspond to differences in energy use intensity of up to 6% and PV generation potential exceeding 14%. However, traditional simulation approaches, while effective for fixed configurations, are less suited to early-stage exploration where rapid iteration and broad design space coverage are required.

A critical gap remains: architects making morphological decisions about indoor arenas lack access to quantitative, interpretable performance feedback during conceptual design. Physics-based simulation requires substantial modeling effort and computational time, making it impractical for rapid iteration. Machine learning offers faster prediction, but most applications focus on regular building geometries and treat models as opaque predictors without revealing parameter-performance relationships. This opacity limits usefulness for design practice, where architects benefit from understanding not only predicted outcomes but also the patterns connecting parameters to performance.

This study develops an integrated framework combining parametric simulation, ensemble machine learning, and SHAP-based predictive sensitivity framework to inform performance information and support subsequent design decision-making for early-stage arena design. The contributions are threefold:

First, we present an automated evaluation pipeline for geometrically complex buildings, integrating Grasshopper-based parametric modeling with coupled daylight-energy simulation and large-scale data processing. Using this pipeline, 9667 design scenarios were generated across three climate zones, addressing a building type where geometric complexity has previously limited systematic study.

Second, we provide climate-specific feature importance analysis, characterizing how design parameters including skylight ratio, roof redundancy height, roof type, and orientation relate to five energy performance indicators across Guangzhou, Shanghai, and Harbin. The analysis reveals distinct climate-dependent sensitivity patterns and nonlinear response characteristics.

Third, we demonstrate how SHAP-based interpretation supports early-stage design by revealing associative patterns between morphological parameters and predicted performance. This provides quantitative information previously unavailable during conceptual design, while clarifying that these model-derived insights require physics-based validation before application as design guidance.

The remainder of this paper is organized as follows. Section 2 reviews related work on performance-oriented design, machine learning for building energy prediction, and explainable AI. Section 3 presents the methodology. Section 4 reports results and discusses implications for design practice. Section 5 concludes with limitations and future directions.

2. Related works

This section reviews three streams of research relevant to the proposed framework: early-stage design support methods and parametric modeling, energy performance evaluation approaches for building design, and machine learning with explainable artificial intelligence techniques. For each stream, we discuss the strengths and limitations of existing approaches to position the contributions of this study.

2.1. Advances and challenges in early-stage performance design support

The early design stage is widely recognized as the phase with the greatest potential to influence long-term building performance. Fundamental form decisions regarding building massing, orientation, and roof configuration are determined while design flexibility remains high (Raji et al., 2017). Once made, these morphological decisions broadly define the operational energy footprint for decades of use. This recognition has motivated integrated design approaches that couple performance analysis with iterative evaluation and feedback-driven design refinement (Negendahl, 2015) (Li et al., 2022a).

Early attempts relied primarily on Building Performance Simulation (BPS) tools to evaluate design alternatives (Østergård et al., 2016). However, BPS-based workflows present significant barriers for early-stage application: translating design models into analysis-ready representations requires substantial domain expertise and time, each simulation entails considerable computational effort and detailed inputs that are often unavailable during conceptual design, and the rapid iteration of architectural form is therefore poorly aligned with the time demands of performance analysis (Han et al., 2018) (Salata et al., 2024). Whether building performance simulation is integrated directly into the design process or outsourced to building physics specialists, it remains challenging.

Multi-objective optimization (MOO) emerged as an alternative, using algorithms such as NSGA-II to identify Pareto-optimal solutions. However, MOO operates within a relatively closed objective system (Picard and Schiffmann, 2021). Early design is characterized by high divergence: objectives extend beyond quantifiable metrics to encompass aesthetic expression and spatial experience that resist formalization. When objectives exceed three, algorithms struggle to identify complete Pareto fronts (Zhang, 2024) (Guo et al., 2024). Moreover, MOO is largely viewed as a computer-driven automated process that is not conducive to architect participation. Consequently, most MOO-based workflows are better suited to design development phases, with architect involvement limited to selecting among algorithmically generated alternatives. Recognizing these limitations, recent research has shifted toward approaches emphasizing architect interaction and information transparency, such as incorporating decision preferences and machine learning (ML) into the optimization process (United Arab Emirates University of Sharjah et al., 2023) (Lin et al., 2020) (Kesireddy and Medrano, 2024). This suggests that effective early-stage support should provide transparent performance intelligence rather than automate design decisions.

Parametric modeling plays an important role in performance-oriented design (Zhang et al., 2025b) (Li et al., 2024). On one hand, constructing parametric models requires identifying key design variables (W. and S, 2025). The critical variables and modeling focus differ across design stages: during conceptual and early design phases, morphological decisions are central, including building orientation, massing, functional distribution, and opening configurations (Salata et al., 2024). Exploration of construction materials and HVAC systems typically belongs to later design development stages, while these are usually set as default templates during early design (Wang et al., 2024a). Research indicates that in early design stages, comparative performance assessment among alternatives is more important than precise simulation results (Salata et al., 2024). Several studies have noted confusion in identifying and optimizing key design variables appropriate to different design stages (Yu et al., 2022) (Gassar et al., 2021). On the other hand, appropriate simplification, and efficient translation of design models into performance analysis models is necessary. Beyond engineering-driven optimization, recent work in computational design has emphasized the role of geometry as a generative driver of environmental adaptation (United Arab Emirates University of Sharjah et al.,

2023). demonstrates how parametric morphologies can be designed to respond to climate conditions through form-based intelligence, reinforcing the view that geometric decisions carry design agency beyond mere technical performance. Energy simulation typically requires closed, non-curved geometries, while daylighting analysis requires model simplification to reduce computational cost. However, over-simplification is prevalent in parametric model construction. Many studies focus on residential and office buildings, which are characterized by box-like forms (Guo et al., 2024) (Pan et al., 2023). In contrast, parametric models for complex public buildings require more precise and elaborate translation, and are more prone to errors during computational analysis (Li et al., 2021).

2.2. Machine learning approach for building energy performance

As noted in Section 2.1, BPS methods struggle to meet the demands of rapid early-stage design iteration. Most physics-based simulation engines such as Radiance and EnergyPlus entail high computational costs, while computer-driven optimization approaches lack extensible objective frameworks (Tian et al., 2021). In recent years, machine learning (ML) has emerged as a promising complement to physics-based simulation for building performance prediction (Wu et al., 2024). By learning from large datasets generated through either measured data or simulation, ML models can rapidly approximate complex input–output relationships, enabling a broader exploration of the design space with far lower computational costs, which makes them particularly suitable for early design (Razmi et al., 2022). Algorithms such as Random Forests, Gradient Boosting, and Neural Networks have been applied to predict energy loads and photovoltaic yields with encouraging accuracy (Geyer and Singaravel, 2018). Among these, ensemble learning methods have demonstrated advantages for building performance prediction due to their ability to reduce overfitting and capture complex nonlinear relationships (Ali et al., 2023). Ensemble approaches can be broadly categorized into three types: Bagging methods such as Random Forest train multiple models on bootstrap samples in parallel to reduce variance; Boosting methods such as XGBoost and LightGBM train models sequentially, with each iteration focusing on samples that previous models predicted poorly; Stacking methods combine predictions from diverse base learners through a meta-learner, leveraging the complementary strengths of different algorithms (Seyyedzadeh et al., 2025). Recent studies have shown that stacking ensembles often outperform individual models in building energy prediction tasks, as they can integrate the robustness of tree-based methods with the pattern recognition capabilities of other algorithms (Dinmohammadi and Shafiee, 2023) (Wang et al., 2020).

Pan et al. combined self-organizing maps with multilayer perceptron neural networks to explore geometry design spaces in the early stages of arena projects (Dostmohammadi et al., 2024). Wang et al. proposed a hybrid ML model based on factor generators, integrating arena operational monitoring data to optimize facility management (Pan et al., 2020). However, most studies remain project-specific, with limited generalizability, and conventional ML models are often treated as “black boxes,” offering little insight into parameter interactions or trade-offs (Hooshyar and Yang, 2024). This lack of interpretability reduces their value for architects and engineers, who require transparent explanations alongside predictive accuracy (Razmi et al., 2022).

To overcome the limitations of black-box machine learning, explainable artificial intelligence (XAI) techniques have recently been introduced into the field of building performance modeling. Among post-hoc explanation methods, Permutation Feature Importance (PFI) measures the increase in model error when feature values are shuffled (Wang et al., 2023), while Local Interpretable Model-agnostic Explanations (LIME) approximates complex models with locally interpretable surrogates (Molnar et al., 2020). However, these methods have limitations in capturing feature interactions or providing consistent global explanations. Shapley Additive Explanations (SHAP) has attracted particular attention due to its rigorous game-theoretic foundation and ability to provide both global and local interpretability (Song et al., 2023). By quantifying the marginal contribution of each input variable to a model's output, SHAP allows researchers to rank features by importance, identify nonlinear responses and threshold effects, and explain case-specific predictions. In this way, SHAP provides transparent and actionable insights for building performance modeling (Shen and Pan, 2023) (Zhang et al., 2022). In the context of building performance, SHAP has been applied in several emerging studies, for example, to interpret solar radiation prediction models, to analyze the sensitivities of cooling and heating loads, or to evaluate indoor comfort parameters (Liu et al., 2022) (Amini Toosi et al., 2023) (Neubauer et al., 2024) (Neubauer et al., 2025). These applications demonstrate SHAP's capacity to bridge predictive accuracy with design reasoning, offering insights that are more transparent and actionable than those derived from conventional feature importance methods. However, most current applications remain limited in scope: they often address a single performance objective or are disconnected from the architectural design process. Besides, the potential of SHAP to support early-stage, multi-objective design optimization in complex building types such as indoor arenas remain largely unexplored.

2.3. Performance-oriented design research for indoor arenas

Performance-oriented design research for indoor arenas and other large-span sports buildings has expanded substantially in recent years, reflecting increasing awareness of their high energy demand, complex environmental requirements, and iconic architectural forms (Luo et al., 2024).

Among different performance domains, structural performance assessment is the most mature. Due to the prevalence of large-span roof systems and complex load-bearing configurations, finite element analysis has long been embedded in both conceptual validation and detailed engineering stages (Elnour et al., 2022). Parametric structural models coupled with numerical optimization are routinely employed to ensure structural safety, stiffness, and material efficiency (Feng et al., 2024). In contrast, research on indoor environmental comfort and energy performance has largely focused on specific case studies or post-form optimization (Du et al., 2025). Zong et al. conducted numerical simulations to optimize window opening ratios for energy and daylighting performance in university arenas (Turrin et al., 2011). Du et al. compared light-thermal environments and energy consumption in ice sports venues between China and Finland (Stefańska and Rokicki, 2022). Javanmard et al. integrated machine learning and parametric design for energy-efficient building cladding in arid climates (Nazifi Charandabi et al., 2025). Yue et al. proposed a metamodel-based multi-objective optimization approach for campus sports facilities (Junlin and Longwei, 2021). Pan et al. developed a numerical framework integrating functional layout, structural design, and performance evaluation for arenas (Yue et al., 2021). Yang et al. established parametric models for specific case studies and discussed interdisciplinary optimization methods (Pan et al., 2019) (Yang et al., 2018). While these studies demonstrate the effectiveness of multi-objective optimization in balancing competing goals, they are typically applied in later design phases and are not well suited to supporting rapid form exploration during early-stage design (Yang et al., 2020) (Sharma and Kumar, 2022).

More fundamentally, generalized, and reusable performance evaluation models for indoor arenas remain underdeveloped. The large spatial scale, heterogeneous functional zoning, complex roof geometries, and event-driven operation patterns challenge the transferability of case-based simulation models (Luo et al., 2024) (Javanmard et al., 2024). Conventional BPS workflows require analysis-ready inputs and substantial computational time, resulting in a mismatch between the pace of conceptual form iteration and the demands of performance evaluation (Østergård et al., 2016). As a result, performance analysis is often used as verification or optimization tool rather than as a source of

design intelligence during early-stage decision-making. These limitations highlight a critical gap: there is a lack of scalable, interpretable, and design-oriented performance modeling frameworks capable of supporting early-stage, multi-objective exploration for complex building types such as indoor arenas.

3. Methodology

3.1. Research workflow

This study presents an integrated explainable machine learning workflow for identifying the influence of design variables on energy-related performance indicators of indoor arenas. The workflow (Fig. 1) comprises three consecutive phases: 1) dataset generation, 2) predictive model construction, and 3) model explanation and early design support.

In Phase 1, an automated performance evaluation pipeline is developed to generate the dataset. A parametric simulation-description model was established to encode key design variables into simulation-ready geometry models. A partially stratified sampling scheme, combined with local Latin Hypercube Sampling (LHS), was applied to generate diverse scenarios across three representative climate zones and multiple seating capacities. An integrated toolchain was developed that automated data transfer between parametric geometry and simulation engines. Five performance indicators covering energy demand and energy generation. This phase generated a high-dimensional dataset of design–performance spanning multiple climatic contexts (see Section 2.2).

Phase 2 involved data preprocessing (cleaning and encoding). Multiple regression algorithms were trained for each performance indicator, and the best-performing models were integrated via a stacked ensemble to enhance predictive stability and generalization. Five-fold cross-validation was employed to mitigate overfitting. The outcome of this phase was a set of robust surrogate models suitable for subsequent interpretability analysis (see Section 2.3).

Phase 3 addressed model interpretability. Correlation analysis provided a preliminary check of variable relationships. SHapley Additive exPlanations (SHAP) was then used to quantify the marginal contribution of each design parameter at both global and local scales. The interpretability framework is detailed in Section 2.4.

3.2. Feature definition and data collection

3.2.1. Design features and performance indicators

The selection of design features was guided by their strong relevance to both the geometry configuration and the energy performance of indoor arenas. Seven key design features were selected to represent the primary morphological and environmental variables that influence performance: Location, Num_seat, Ori_roof, Types_roof, Rdncy_height, Ratio_SLF, and Dist_SL (Table 1). Among these, Location and Num_seat strongly influence operational load and the scale of the building envelope. This study selects three representative cities from different climate zones in China: Guangzhou (GZ), Shanghai (SH), and Harbin (HB), which cover typical Köppen climate classifications: Guangzhou (Cwa), Shanghai (Cfa), and Harbin (Dwb). Within each location layer, we further stratify by seating capacity. The seating capacity ranges from 4000 to 12,000, with a step size of 1000. This range and step size are chosen based on standard scales of large sports venues in China, ensuring the representativeness of the sampling space. Ori_roof, Types_roof, and Rdncy_height directly affect both the heat transfer surface

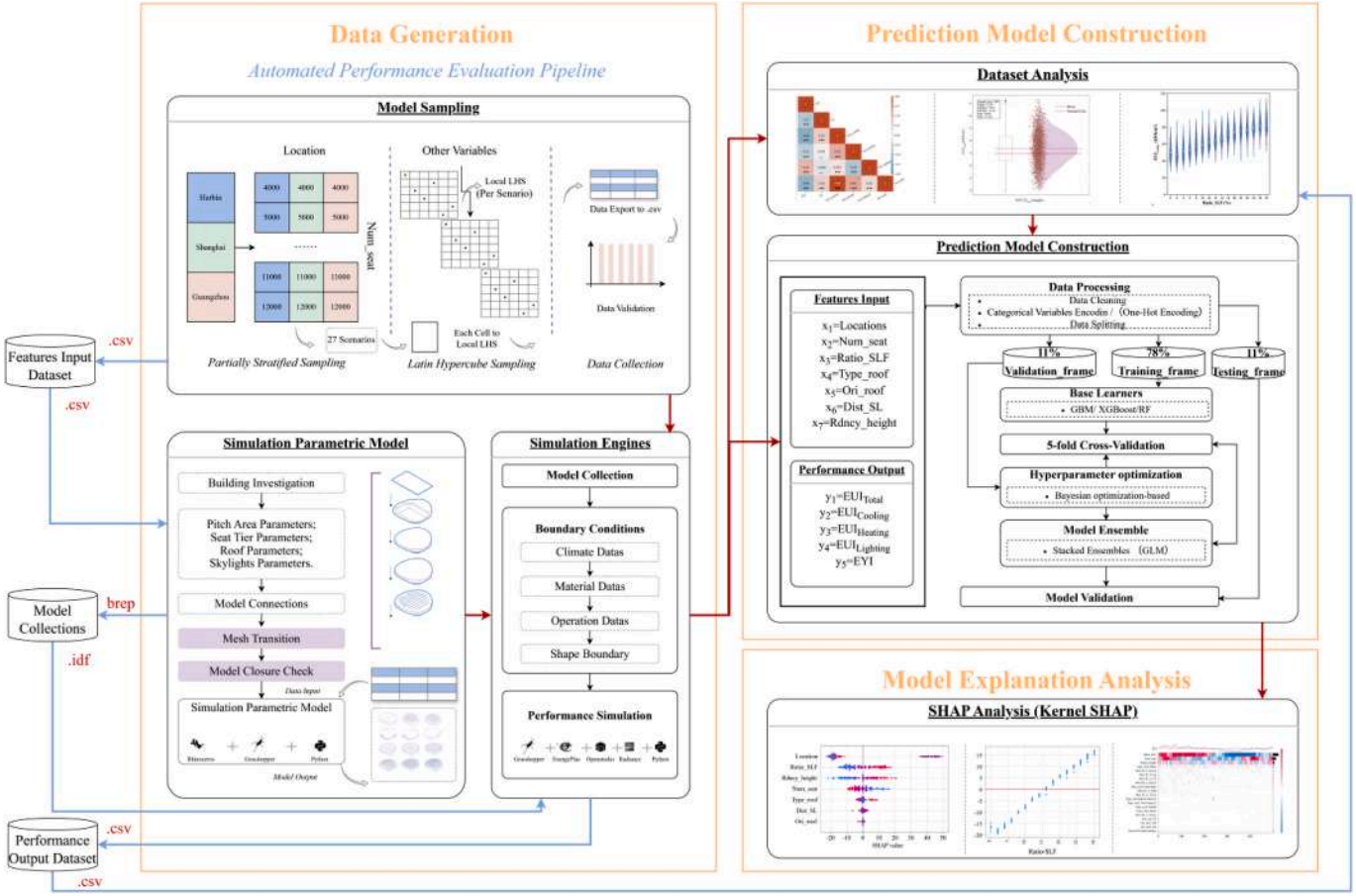

Fig. 1. Workflow of the study.

Table 1
Variable information.

Variable	Variable Description	Variable Type	Variable Attribute	Variable Value Range
Location	Climate Zone	First-level stratified variable	Categorical	'Guangzhou (GZ),' 'Shanghai (SH),' 'Harbin (HB)'
Num_seat	Seating capacity	Second-level stratified variable	Continuous	4000 to 12000 (step size 1000)
Ori_roof	Roof orientation	LHS Variable	Categorical	S, SW,W, NW,N, NE,E, SE
Types_roof	Roof geometry	LHS Variable	Categorical	'Plane,' 'Dome,' 'Four-slope,' 'Barrel Vault (x),' 'Two-slope (x),' 'Saddle'
Rdncy_height	Roof redundancy height	LHS Variable	Continuous	0-20 (m)
Ratio_SLF	Skylight to floor ratio	LHS Variable	Continuous	0-30 (%)
Dist_SL	Skylight distribution type	LHS Variable	Categorical	'A_Cl', 'L_Rad', 'L_Ann', 'L_Eq (x)', 'L_Eq (y)', 'L_Cl (x)', 'L_Cl (y)', 'P_Eq'

area and the efficiency of photovoltaic (PV) generation. Ratio_SLF and Dist_SL are closely related to daylight penetration and solar heat gain. Collectively, these features span three core aspects of indoor arena design: morphology, climate responsiveness, and renewable energy integration.

It should be noted that this study deliberately focuses on geometric and morphological parameters that architects can directly manipulate during the early conceptual design stage. Parameters such as detailed glazing properties (e.g., U-value, SHGC), HVAC system types, and envelope thermal mass were not included as design variables, because these are typically determined at later design stages when architectural form has been established and engineering systems are specified. At the early stage targeted by this study, these parameters remain undefined. By contrast, morphological decisions such as roof geometry, skylight ratio, and building orientation are central to early-stage design exploration and have been shown to exert significant influence on energy performance (Salata et al., 2024). This scope definition aligns with the framework's objective of supporting architects' form-making decisions rather than detailed engineering optimization.

To comprehensively assess the performance of different design configurations, both energy consumption and energy production aspects were evaluated.

Energy Use Intensity (EUI) is a widely recognized metric for quantifying a building's energy performance, defined as the total annual energy consumption per unit of gross floor area. It is a standard benchmark in building energy codes and green building rating systems, enabling direct comparison of the energy-saving potential across different designs. For large-span sports buildings, heating, ventilation, and air conditioning (HVAC) systems and lighting systems typically account for most of the total energy demand and are significantly influenced by the building's morphology. Four EUI-based indicators were adopted: cooling energy use intensity ($EUI_{\_Cooling}$), heating energy use intensity ($EUI_{\_Heating}$), lighting energy use intensity ($EUI_{\_Lighting}$), and total energy use intensity ($EUI_{\_Total}$). The general definition of EUI is:

$$EUI = \frac{E_{annual}}{A_{gross}} \tag{1}$$

where E_{annual} is the total annual energy consumption (kWh/year), aggregated from hourly simulation outputs for cooling, heating, and lighting loads; and A_{gross} is the gross floor area of the competition hall (m^2).

Large-span sports buildings typically feature expansive, unobstructed roof surfaces with high solar exposure, making them ideal platforms for photovoltaic (PV) integration (Wu et al., 2025). Considering these characteristics, this study incorporates PV-related performance indicators to evaluate both the energy consumption and energy production capacity of different roof design configurations, thereby enabling a balanced assessment of their contribution to overall building energy performance. For roof systems incorporating PV installations, energy yield intensity (EYI) is used. Energy yield (EY) represents the total annual electricity generation of the PV system, considering the module rated capacity, system efficiency, solar irradiance, tilt angle, azimuth angle, and shading conditions. Energy yield intensity (EYI) normalizes EY by the gross floor area, allowing direct comparison with EUI:

$$EYI = \frac{EY}{A_{gross}} \tag{2}$$

While EYI reflects the contribution of PV generation to the building's net energy balance. Together, these indicators enable a balanced evaluation of energy consumption reduction and renewable energy production potential across various design configurations.

3.2.2. *Partially stratified and Latin Hypercube Sampling*

To ensure balanced and representative coverage of the design space, a partially stratified sampling (PSS) strategy was adopted to control sampling bias at multiple levels (Li et al., 2022b). First, stratification by climate zone and seating capacity ensures balanced representation across these primary factors that strongly influence both building scale and operational loads. Without stratification, random sampling could over-represent certain climate-capacity combinations. Second, within each stratum, LHS ensures uniform marginal distributions across the remaining five design variables, preventing clustering in any region of the parameter space. The effectiveness of this bias control is verified by: (a) chi-square goodness-of-fit tests confirming uniform distribution across categorical strata ($\chi^2 < 1.0$, p = 1.000), and (b) Pearson correlations among continuous variables ($|r|<0.05$), indicating successful minimization of inter-variable dependencies that could confound subsequent feature importance analysis. As shown in Fig. 1, two categorical variables Location (Harbin, Shanghai, Guangzhou) and Num_seat are used as stratification dimensions because they strongly affect operational loads and envelope scale. Within each stratum, the remaining five variables (Ori_roof, Type_roof, Rdncy_height, Ratio_SLF, and Dist_SL) are sampled by a local Latin Hypercube Sampling (LHS) scheme, which ensures uniform coverage in high-dimensional spaces and reduces the total number of samples required compared to full factorial designs (Fang et al., 2025).

Let k_c be the number of stratified variables $k_c = 2$ and V_j the value set of the j-th stratified variable. All combinations of stratified variables define m scenarios:

$$S = \prod_{j=1}^{k_c} V_j, m = \left| S \right| \tag{4}$$

For each scenario $s_t \in S$ with n_s s amples required, the range of each unstratified variable x_j (j = 1, …,5) is divided into n_s equally probable intervals. One value is randomly drawn from each interval without replacement, and the values across all 5variables are randomly paired to form n_s sample points:

$$X_{s_t} = \left\{ x_i^{(t)} = \left(x_{i1}^{(t)}, x_{i2}^{(t)}, x_{i3}^{(t)}, x_{i4}^{(t)}, x_{i5}^{(t)} \right) \mid i = 1, \ldots, n_s \right\} \tag{5}$$

The final sample set is the union of all scenario-specific LHS samples:

$$X = \bigcup_{t=1}^{m} X_{s_t} \tag{6}$$

This PSS + LHS approach ensures categorical balance across climate zones and seating capacities, while maintaining uniform marginal distributions of continuous and remaining categorical variables in the high-dimensional feature space.

3.2.3. Dataset generation

The dataset generation process was established through an automated simulation workflow that integrates Python, Grasshopper, and multiple simulation engines, enabling seamless data transfer and high-throughput computation. Based on sampled design parameters, the Simulation–Parametric Model (SPM) served as the generative core of this workflow. Each sample configuration was first instantiated through the SPM, which encoded both geometry and physical attributes of the indoor arena design. The instantiated models were then automatically converted into simulation-ready formats using Honeybee, linking EnergyPlus for energy modeling and Radiance for lighting analysis. The daylight simulation results were dynamically fed back into EnergyPlus as boundary conditions to update lighting energy demand and complete the coupled daylight–energy simulation. Batch simulations were executed to generate hourly outputs for cooling, heating, lighting, total energy consumption, and photovoltaic (PV) generation. These were subsequently post-processed in Python to derive five performance indicators. All results were automatically stored and indexed, allowing the system to complete all design cases autonomously and compile a comprehensive dataset. This unified boundary and automation framework ensured consistent coupling between daylight and energy modules, enabling climate-responsive, high-throughput evaluation of design parameters while substantially reducing manual workload during early-stage performance assessment.

The SPM employed in this study was developed and validated in our previous research. The model comprises four main modules (competition field, spectator seating, roof geometry, and skylight configuration) which are encoded as three-dimensional geometries using GH-Python scripting on the Rhinoceros platform. The generated models can be automatically converted into simulation-ready input formats compatible with EnergyPlus and Radiance. This approach enables rapid generation of large-span sports building design schemes while maintaining both geometry adaptability and computational efficiency. Indeed, geometric simplification for energy simulation has been shown to be a well-established practice in performance-based design research; for instance (Shields and Zhang, 2015), demonstrated that heuristic-based simplification of complex curved building envelopes produces valid energy models with minimal impact on simulation accuracy. A complete description of the model construction methodology, geometry rules, and validation procedure can be found in (Santos et al.), and is not repeated here.

Simulation boundary conditions were established based on the sampled design variables, encompassing sky models, construction materials, operational schedules, and photovoltaic (PV) system parameters. Weather data for Harbin, Shanghai, and Guangzhou were sourced from the IWEC database in EPW format and processed through Ladybug.

For energy performance simulation, the spatial topology of the main

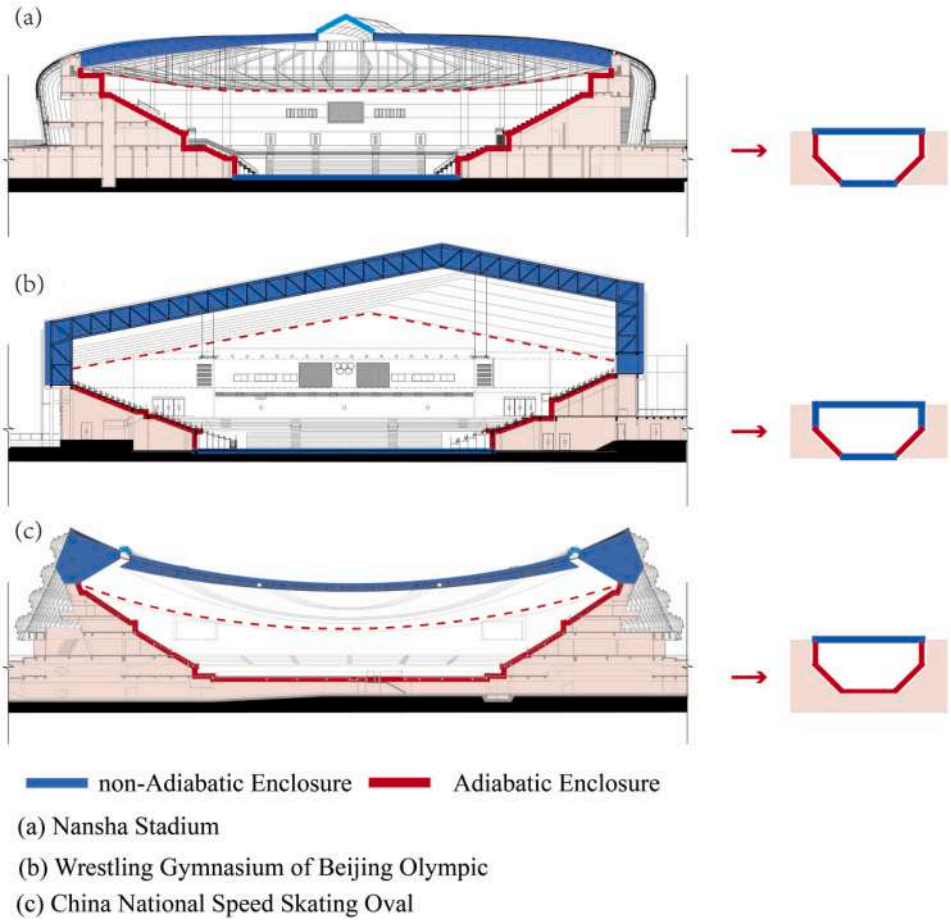

Fig. 2. Typical topological structure and enclosure boundary condition settings for competition hall space and other spaces
(Source: elaboration by the author, technical drawings from (Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2012) (Sun et al., 2012) (Sun, 2009)).

arena and its auxiliary spaces was defined according to typical configurations of large-span sports venues (Fig. 2). Seating areas enclosing rooms with HVAC operation were modeled as adiabatic zones, while roofs and walls directly exposed to ambient air were treated as non-adiabatic. The base model excluded basements or ground-coupled spaces. Thermal resistance (R) and heat transfer coefficients (K) for different envelope components were assigned according to regional codes and verified design cases (Table 2).

Internal loads, occupancy, and control schedules followed sports building items of GB 55015-2021 "General Specification for Building Energy Conservation and Renewable Energy Utilization" (Wang et al., 2024b), adopting a representative operation period from 09:00 to 21:00. Lighting, equipment, and ventilation schedules were defined in three modes—pre-use, active-use, and post-use—to capture realistic variations in internal heat gains. The operational schedule reflects the typical usage pattern of Chinese indoor arenas, which primarily serve daily public fitness, training, and recreational activities rather than exclusive event-based operation. This continuous-use assumption is particularly relevant for the skylight analysis, as daylighting strategies would be unnecessary in purely event-driven venues where artificial lighting dominates.

For this purpose, 170 daylight reference points were uniformly distributed across the main activity area in a 4 m × 4 m grid on a plane 0.75 m above the floor. Optical and thermal material properties were defined in accordance with GB50033-2013 "Standard for Daylighting Design of Buildings" (Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2021) and GB 50176-2016 "Thermal Design Code for Civil Building" (Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2013). The specifications of non-translucent and translucent envelopes, including reflectance, visible transmittance, U-value, and SHGC, were assigned according to standard practices for large-span sports buildings and are summarized in Table 2. In this coupled daylight–energy simulation, daylight analysis was integrated into the energy modeling workflow to dynamically assess the impact of skylight configuration on lighting energy demand. The computed daylight availability determined the switching schedule of electric lighting, thereby affecting the EUI_Lighting during annual simulations.

3.2.4. Data validation

To verify that the dataset achieved balanced coverage and minimized bias, statistical checks were performed on both categorical and continuous variables.

First, chi-square goodness-of-fit tests were applied to the two stratified categorical variables—climate zone (Harbin, Shanghai, Guangzhou) and seating capacity—to confirm uniform sample allocation within each category. The results ($\chi 2<1.0$, $p = 1.000$) indicate that the stratification process successfully eliminated category-level imbalance.

Next, pairwise correlations among performance indicators (y–y) were analyzed using both Spearman rank correlation (r_s) and Pearson correlation (r) to evaluate monotonic and linear relationships, respectively. Spearman's method, being non-parametric and robust to outliers, was employed as the primary metric, while Pearson's correlation was used as a complementary measure to confirm linear associations (Sun, 2022).

The Spearman Rank Correlation Coefficient (SRCC), proposed by Charles Spearman, is a non-parametric statistical method for assessing the monotonic relationship between two variables based on their ranked (ordered) values:

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)} \tag{7}$$

where r_s represents the Spearman rank correlation coefficient, d_i is the difference between the ranks of each pair of observations, and n is the total number of observation pairs.

The Pearson correlation coefficient (PCC), developed by Karl Pearson, measures the strength of the linear relationship between two continuous variables. It is defined as:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \tag{8}$$

Where r represents the Pearson correlation coefficient, x_i and y_i are the observed values of the two variables, and $\bar{x}$ and $\bar{y}$ are the mean values of these observations.

Low absolute correlation values across features ($|r_s|<0.05$ in most cases) confirmed that the PSS + LHS scheme effectively minimized dependencies among variables, thereby reducing the risk of multicollinearity in subsequent machine learning modelling and improving the interpretability of SHAP and permutation feature importance (PFI) results.

Table 2
Boundary condition settings for coupled daylight–energy simulation (Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2021) (Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2013) (Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2016).

Non-transmitting Envelope					
Surface	**Materials (indoor)**		**Reflectance**	**Thermal Resistance R(m^2·K/W)**	
Floor (No Basement)	Light-colored wood floors		0.58	0.35 (GZ); 1.01 (SH); 1.01 (HB)	
Seating tiers (Adiabatic)	Self-leveling mortar		0.32	10 (GZ); 10 (SH); 10 (HB)	
Wall	Perforated aluminum plates		0.45	1.15 (GZ); 1.68 (SH); 3.05 (HB)	
Roof surface	Perforated steel plates		0.72	4.64(GZ); 4.64(SH); 6.34(HB)	
Light-transmitting Envelope					
Locations	**Materials**	**Visible Transmittance**	**Visible Transmittance (considering τw)**	**Coefficient K (W/(m^2·K))**	**Solar Heat Gain Coefficient (SHGC)**
Guangzhou	6C + 12A+6C	0.81	0.61	2.0	0.30
Shanghai	6Low-E+12A+6C	0.68	0.51	2.0	0.30
Harbin	6Low-E+12A+6C	0.68	0.41	2.0	0.55
Other Parameters					
Parameters			**Setting**	**Schedule and Occupation**	
Lighting Power Density (W/m^2)			8.0	Normal Operation (90%), 1 Hour Before/After (20%)	
Electrical Equipment Power Density (W/m^2)			10.0	Normal Operation (85%), 1 Hour Before/After (10%)	
Personnel Density (m^2/person)			4	Normal Operation (90%), 1 Hour Before/After (10%)	
Heat Dissipation per Person (W/person)			61 (Sensible), 73 (Latent)	——	
Fresh Air Volume per Person (m^3/(h·person))			20	Normal Operation (100%), Other Times (0%)	
Heating Room Temperature (°C)			18	Normal Operation (18 °C), 1 Hour Before (15 °C), Other Times (12 °C)	
Air Conditioning Room Temperature (°C)			26	Normal Operation (26 °C), 1 Hour Before (28 °C), Other Times (37 °C)	

3.3. Machine learning model development

The development of the machine learning (ML) model began with feature engineering, where outliers were identified and treated, and categorical variables were encoded using one-hot encoding. Before training, the dataset was randomly divided into 78% for training, 11% for validation, and 11% for testing. The validation set was used for five-fold cross-validation within each base model to optimize hyperparameters and generate out-of-fold (OOF) predictions, while the testing set was reserved exclusively for evaluating the generalization performance of the final stacked ensemble.

Data preprocessing involved four steps: (1) Outlier detection and treatment: Samples with simulation errors (e.g., non-convergent EnergyPlus runs, negative energy values) were identified and removed, resulting in 9667 valid samples from an initial 10000 runs (96.7% success rate). (2) Categorical encoding: Categorical variables (Location, Type_roof, Ori_roof, Dist_SL) were transformed using one-hot encoding to enable gradient-based learning, expanding the feature space dimensions. (3) Feature scaling: Continuous variables were standardized (zero mean, unit variance) for the base learners, though tree-based models are inherently scale-invariant. (4) Data splitting: The dataset was randomly partitioned into training (78%), validation (11%), and testing (11%) sets using stratified sampling to maintain climate zone proportions across splits.

A stacked ensemble learning strategy was adopted. The stacked ensemble was chosen to leverage the complementary strengths of heterogeneous base learners, allowing the meta-learner to capture both linear and complex non-linear relationships in the design–performance mapping, which is particularly valuable in high-dimensional, multi-climate datasets (De Winter et al., 2016) (Pavlyshenko, 2018). As demonstrated in Fig. 7, the ensemble consistently outperforms individual models across different performance indicators, justifying the added complexity. In the first layer, three representative regression models with diverse learning preferences and generalization capabilities were selected as base learners: Extreme Gradient Boosting (XGBoost), Gradient Boosting Machine (GBM), and Random Forest (RF). These models collectively capture different aspects of the complex feature space of architectural design parameters (Lu et al., 2023). Each base model was trained using five-fold cross-validation, generating predictions for each fold. The predictions from all base models were then concatenated to form the meta-feature set. In the second layer, a Generalized Linear Model (GLM) was used as the meta-learner to integrate the outputs from the first layer, producing the final predictions. All models were built with automatic hyperparameter tuning to ensure robust performance across different climate zones and building types. These algorithms were selected due to their proven predictive performance in building energy modelling and their complementary learning biases. XGBoost, GBM, and RF are tree-based ensemble methods capable of handling non-linear relationships and feature interactions (Boutahri and Tilioua, 2024) (Tyagi and Singh, 2020) (Alshboul et al., 2022), while GLM provides a linear baseline (Cho et al., 2020). Hyperparameters of the base learners were optimized using Bayesian optimization. Bayesian optimization is an iterative framework for global optimization of expensive black-box functions, which has been widely adopted for hyperparameter tuning in machine learning (Konstantinov and Utkin, 2020). Unlike grid search or random search, Bayesian optimization builds a probabilistic surrogate model of the objective function and uses an acquisition function to determine the next evaluation point, effectively balancing exploration, and exploitation.

In each iteration, the surrogate model is fitted to all observations made so far. The acquisition function then evaluates the utility of candidate hyperparameter configurations based on the predicted mean and uncertainty from the surrogate model. The Expected Improvement (EI) is a commonly used acquisition function, defined as:

$$\mathrm{E}[\mathrm{I}(\lambda)] = \mathrm{E}[\max(\mathrm{fmin} - \mathrm{y}, 0)] \tag{9}$$

where *fmin* is the best observed value so far and *y* is the predicted value at configuration λ. This formulation encourages sampling in regions where the predicted value is low (exploitation) or the uncertainty is high (exploration).

In this study, Bayesian optimization was implemented using the scikit-optimize library with 20 iterations and 5-fold cross-validation for each of the 15 prediction models. The negative root mean square error (RMSE) was used as the scoring metric. The hyperparameter search space and representative optimal values are reported in Section 3.2.

The predictive performance of the stacked ensemble was evaluated using three commonly accepted regression metrics—coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE). These metrics ensure that the subsequent explainability analysis is based on models with reliable predictive accuracy (Airlangga and Liu, 2025) (Chicco et al., 2021).

1) Coefficient of determination (R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \widehat{y_i})^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{10}$$

where y_i is the observed value, $\widehat{y_i}$ is the predicted value, and $\overline{y}$ is the mean of observed values. R^2 measures the proportion of variance in the dependent variable that is predictable from the independent variables, with 1 indicating perfect prediction.

2) Root Mean Square Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y_i})^2} \tag{11}$$

RMSE penalizes large errors more heavily and is sensitive to outliers, providing an indication of the model's accuracy in absolute terms.

3) Mean Absolute Error (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \widehat{y_i}| \tag{12}$$

MAE measures the average magnitude of errors without considering their direction, offering a more interpretable metric for practical applications.

3.4. Explainable machine learning analysis

This study emphasizes model interpretability to elucidate how architectural design features influence energy performance. After predictive modeling, SHapley Additive exPlanations (SHAP) were applied to quantify the marginal contribution of each input feature to model predictions. Grounded in cooperative game theory, SHAP decomposes the model output into the additive contributions of individual features, ensuring consistency and local accuracy (Chai and Draxler, 2014).

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!\,(|F| - |S| - 1)!}{|F|!} \left[f_{S \cup \{j\}}\left(x_{S \cup \{j\}}\right) - f_S(x_S) \right] \tag{13}$$

where ϕ_j is the SHAP value for feature*j*, *F* is the full set of features, and *S* is a subset of *F* excluding *j*.

Given the complexity of the stacked ensemble architecture, Kernel SHAP was adopted as the explanation module to compute both global and local attributions: Global: Feature ranking plots, summarizing overall importance and direction of effect. Local: Dependence plots and force plots, illustrating the feature impact for individual samples (Baptista et al., 2022).

We defined a black-box prediction function that takes a NumPy array

of predictors (column order fixed) and returns model outputs; the explainer calls this function without accessing model internals. The background distribution comprised 500 samples randomly drawn from the training data (seed = 1). Unless stated otherwise, SHAP values were computed on the same 500 instances, and global results were summarized from these values.

Categorical variables were one-hot encoded for learning. To avoid splitting attribution across many dummy columns, we grouped SHAP values back to their original categorical variables by summing the SHAP values of all dummies belonging to the same variable. Grouping was applied to the categorical features representing roof type, skylight distribution, and roof orientation. Global results were communicated with SHAP summary plots (feature ranking and direction of effect). Local behavior was illustrated with dependence plots.

Finally, SHAP summary plots (global rankings and effect directions), heatmaps, and dependence plots were used to visualize the results. Cross-comparison across climate zones and performance indicators revealed the most sensitive design variables, consistent influence patterns, and key nonlinear trade-offs—providing a transparent and interpretable basis for early-stage design decisions.

It should be noted that SHAP quantifies feature contributions to model predictions rather than establishing physical causation. The resulting attributions reflect learned statistical associations within the dataset and should be interpreted as model-based insights that complement, rather than replace, physics-based reasoning.

4. Results and discussion

4.1. Dataset generation and benchmarking check

To assess the effectiveness and uniformity of the sampling process, both statistical analysis and visual inspection were conducted for continuous and categorical variables. For continuous variables, Pearson correlation coefficients were calculated to examine potential linear dependencies among parameters. As shown in Fig. 3a, the coefficients are generally close to zero, indicating weak pairwise correlations. The scatter plot matrix further illustrates that all variables are evenly distributed within their respective value ranges, without noticeable clustering or gaps, confirming that the Latin Hypercube Sampling (LHS) method effectively minimized inter-variable correlation and achieved uniform coverage across the parameter space. For categorical variables, frequency statistics and chi-square goodness-of-fit tests were performed. The category counts of each variable are approximately balanced, with no significant deviations. The chi-square test results ($p > 0.05$) confirm that the observed distributions do not significantly differ from the expected uniform distributions, demonstrating that the partially stratified sampling (PSS) approach successfully ensured balanced representation across all categorical strata.

Following validation, each sampled configuration was instantiated into a Simulation–Parametric Model (SPM) as described in Section 2.2.3, and subsequently converted into EnergyPlus and Radiance input formats for batch simulation. Hourly energy consumption and photovoltaic generation data were post-processed in Python to derive five performance indicators. After filtering duplicate and invalid records, 9667 valid design–performance samples were obtained, forming a high-dimensional, multi-climate dataset for subsequent surrogate modeling and explainable analysis (Fig. 3b).

Although this study simulated indoor arenas across three representative climate zones: Harbin, Shanghai, and Guangzhou, only Shanghai currently provides an official local benchmark for energy use in large and medium-sized sports buildings, Guangzhou and Harbin classify such facilities under general large-scale public buildings without specific thresholds. Second, Shanghai's mixed climate, requiring both substantial cooling and heating, provides a representative case for evaluating the plausibility of simulated results across multiple load types. Therefore, the Shanghai standard was adopted as a representative reference to validate the overall reasonableness of the simulated dataset. According to the *Guideline of Efficient Energy Conservation for Large and Medium-sized Sport Buildings* (Shanghai Municipal Standard) (Li, 2022), which applies to indoor arenas with seating capacities above 3,000, the equivalent unit electricity consumption for the competition hall area should not exceed 70 kWh/(m^2·a), excluding office and auxiliary spaces. The guideline also defines an excellent energy performance level of 40 kWh/(m^2·a). As shown in Fig. 3c, the simulated EUI_{Total} for Shanghai exhibits a mean of 73.7 kWh/m^2, closely aligned with the benchmark standard. The simulated range (39.9–112.6 kWh/m^2) covers both the excellent and typical levels, confirming that the dataset represents a realistic and regulation-consistent distribution of energy performance. Using the Shanghai benchmark as a representative case thus provides a reasonable basis for assessing the plausibility of the generated samples across different climate zones.

To further verify the internal consistency of the dataset, Pearson and Spearman correlations were computed among the five energy performance indicators (Fig. 4).

Results show that the dominant factors driving total energy use vary by climate. In Guangzhou, cooling energy use intensity exhibits a robust positive correlation with total energy use ($r = 0.99$; $r_s = 0.99$), confirming that cooling demand is the primary driver of total energy consumption, whereas lighting energy use shows negligible correlation. In Shanghai, the correlation between cooling and total energy use remains strong ($r = 0.84$; $r_s = 0.84$), while the correlation with heating energy use increases markedly ($r = 0.63$; $r_s = 0.72$). In Harbin, heating energy

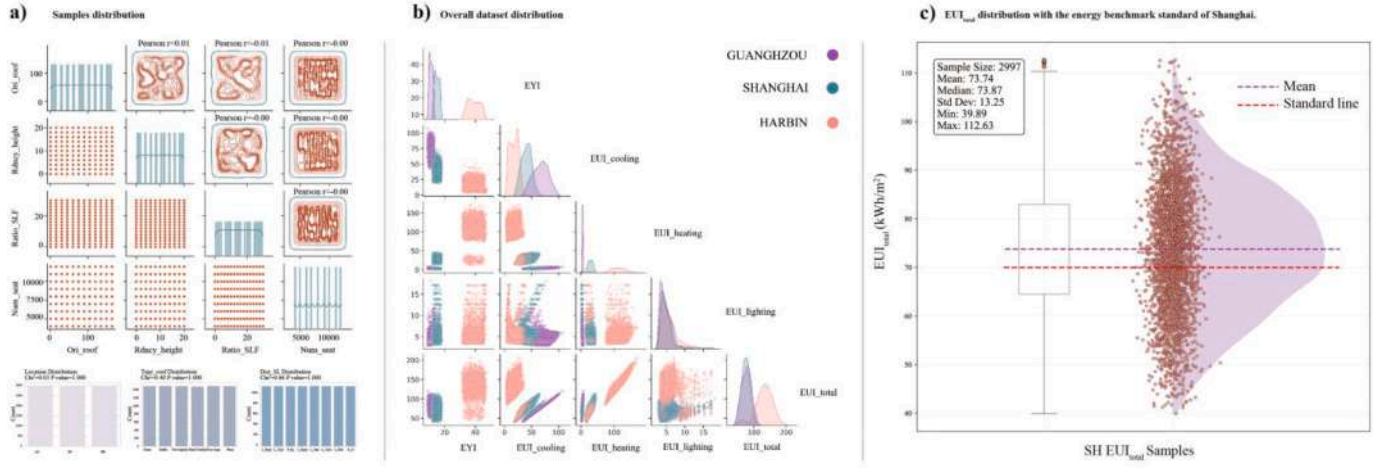

Fig. 3. Sampling dataset distribution.

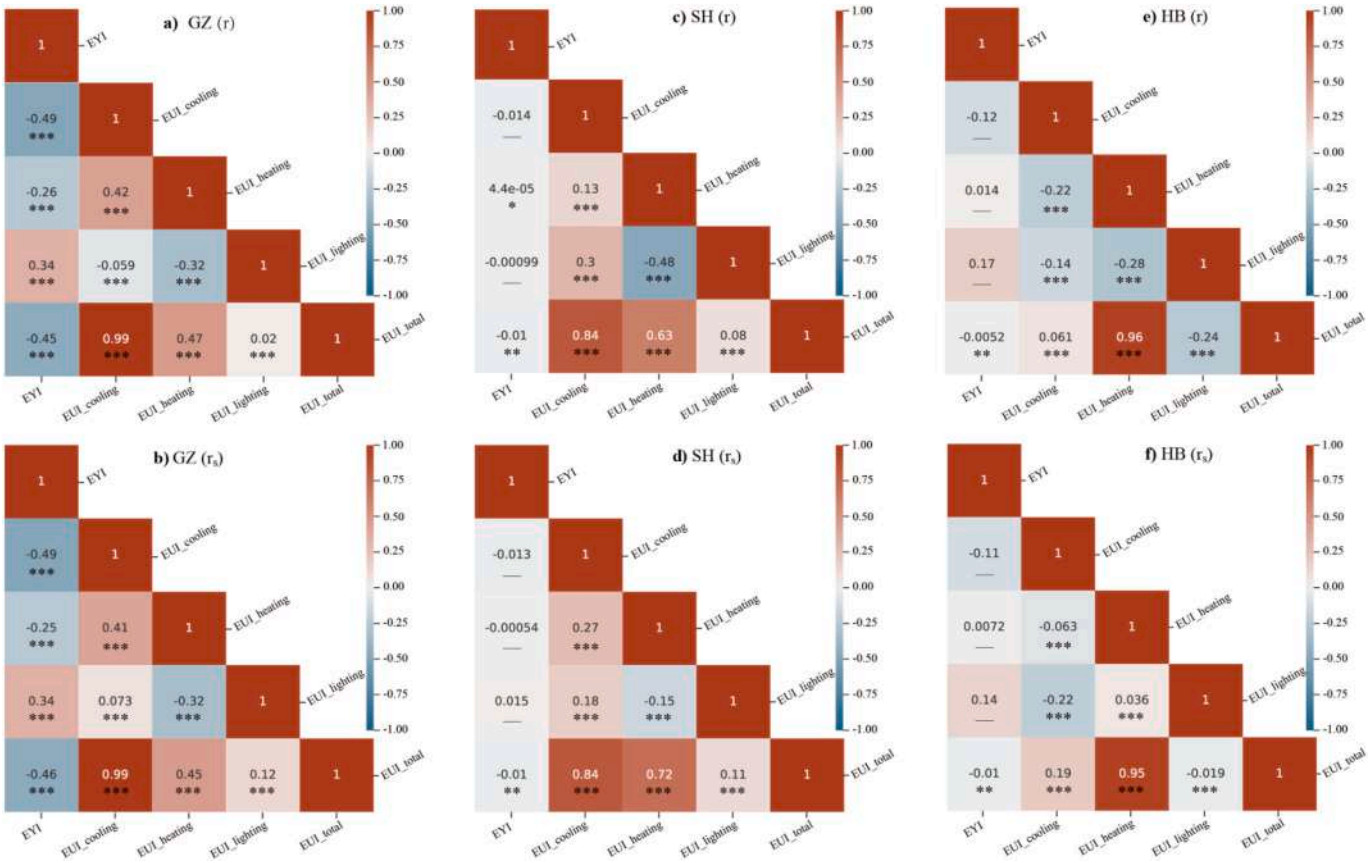

Fig. 4. Correlation analysis among energy performance indicators (*Significance:***p ≤ 0.001, **p ≤ 0.01, *p ≤ 0.05*).

use exhibits the strongest correlation with total energy use ($r = 0.95$; $r_s = 0.96$), while cooling energy use correlation is relatively low.

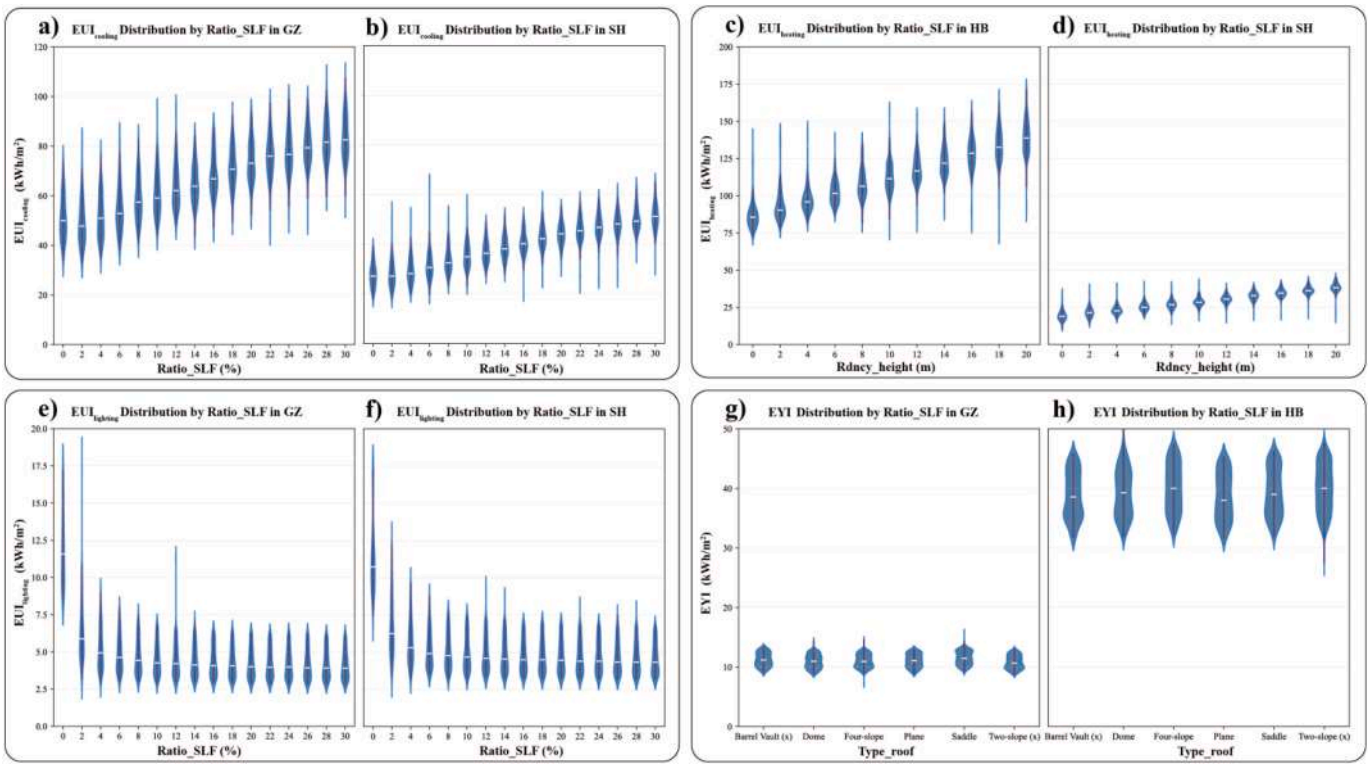

Fig. 5. Performance distribution by different features.

Across all three climate zones, lighting energy use shows no significant correlation with total energy use, suggesting that daylight performance and energy performance contribute to overall building efficiency in a relatively independent manner.

Before the explainable machine learning analysis, a preliminary examination of performance distributions was conducted to verify the dataset quality and to identify general response patterns between design features and energy performance (Fig. 5). These violin plots illustrate the variability and central tendencies of key indicators across design parameters. The dataset exhibits distinct distributional characteristics across climate zones, reflecting the underlying physical differences in energy demand profiles. As shown in Fig. 3b, Guangzhou samples display the highest EUI_cooling values (mean: 66.7 kWh/m^2) with minimal heating demand, while Harbin shows the inverse pattern with dominant EUI_heating (mean: 113.2 kWh/m^2). Shanghai presents intermediate values for both indicators, consistent with its mixed climate requiring both substantial cooling and heating. This climate-dependent structure validates the decision to train separate models for each climate zone rather than a single unified model. A clear non-linear relationship can be observed between the Ratio_SLF and $EUI_{cooling}$ energy demand initially decreases as skylight area increases. This inflection occurs around 2% Ratio_SLF (Fig. 5a and b). In contrast, the $EUI_{lighting}$ consistently decreases with increasing Ratio_SLF, yet its marginal improvement becomes negligible beyond the same threshold, suggesting a point of diminishing returns for daylight contribution (Fig. 5e and f). Regarding geometric factors, redundant height exerts a stronger impact on heating energy use in Harbin, where greater roof cavity volumes lead to higher heat losses under cold conditions, whereas its effect in Shanghai remains comparatively weak due to milder winter temperatures (Fig. 5c and d). For Harbin, the distribution of EYI highlights the influence of roof morphology under high-latitude solar conditions: four-slope and two-slope roofs achieve significantly higher photovoltaic generation than flat roofs, owing to their enhanced solar exposure angles (Fig. 5g and h). Overall, these distribution patterns demonstrate strong internal consistency and realistic physical behavior of the simulated dataset, confirming its validity for subsequent explainable analysis. They also provide an intuitive reference for comparison with the SHAP-based interpretations that follow, where the same relationships and thresholds are further quantified and explained.

4.2. Predictive performance of ML models

Table 3 summarizes the hyperparameter search space for the three base learners, along with representative optimal values obtained for the EUI_cooling model in Shanghai. The optimal hyperparameters varied across different prediction models depending on the specific indicator and climate characteristics, but generally remained within similar ranges, indicating stable convergence of the optimization process.

Fig. 6 presents representative parity plots for selected models, while Table 3 summarizes the predictive performance of all 20 models across three climate zones and five performance indicators. The predicted and actual values align closely along the reference line, confirming that the stacked ensemble models accurately capture the nonlinear relationships between design variables and energy performance outcomes.

Overall, the models achieved high predictive accuracy, with most testing R^2 values exceeding 0.83 and RMSE/MAE remaining within acceptable bounds.

Fig. 7 presents the RMSE comparison of different regression models during the cross-validation phase of the training process. The results represent the validation performance of individual base learners: Extreme Gradient Boosting (XGB), Gradient Boosting Machine (GBM), and Random Forest (RF)—and the stacked ensemble (SE) framework. Across all performance indicators, the stacked ensemble consistently achieves the lowest or near-lowest RMSE values, confirming its superior generalization ability.

To assess model stability, we examined the variance of performance metrics across the five cross-validation folds. The coefficient of variation (CV) of RMSE across folds remained below 15% for all models, indicating consistent performance regardless of data partitioning. Table 4 reports the mean and standard deviation of R^2, RMSE, and MAE across folds for each performance indicator and climate zone. The low standard deviations confirm that the stacked ensemble achieves stable predictions not dependent on specific training samples.

Table 3
Hyperparameter search space and representative optimal values for base learners.

Base Learner	Hyperparameter	Search Range	Optimal Value[a]
XGBoost	learning_rate	[0.01, 0.3]	0.086
	max_depth	[3, 10]	4
	n_estimators	[100, 500]	152
	min_child_weight	[1, 10]	5
	subsample	[0.6, 1.0]	0.99
	colsample_bytree	[0.6, 1.0]	0.67
	reg_alpha	[0.0, 1.0]	0.77
	reg_lambda	[0.5, 5.0]	0.12
GBM	learning_rate	[0.01, 0.3]	0.040
	max_depth	[3, 10]	8
	n_estimators	[100, 500]	151
	min_samples_leaf	[1, 10]	9
	min_samples_split	[2, 20]	10
	subsample	[0.6, 1.0]	0.77
Random Forest	max_depth	[5, 30]	12
	n_estimators	[100, 500]	69
	max_features	['sqrt', 'log2]	sqrt
	min_samples_leaf	[1, 10]	7
	min_samples_split	[2, 20]	13

[a] Note: Optimal values shown are for the EUI_cooling model in Shanghai as a representative example. Hyperparameters were independently optimized for each of the 15 prediction models (5 indicators × 3 climate zones) using the same search space.

For EUI_cooling, testing R^2 ranged from 0.90 (GZ) to 0.98 (ALL), indicating consistently strong model performance across climates. RMSE values (2.1 – 4.1) demonstrate stable prediction accuracy for cooling loads, which dominate total energy use in warm regions. For EUI_heating, accuracy varied across climates, reflecting regional load differences. All three locations performed well despite Harbin's higher absolute RMSE (6.3), which corresponds to larger load magnitudes rather than prediction bias. For EUI_lighting, the models achieved the highest stability and accuracy, with testing R^2 values between 0.95 and 0.99 and extremely low RMSE (<0.4) and MAE (<0.2). This reflects the robustness of the ensemble in capturing lighting energy consumption, which is less sensitive to nonlinearity and climate variation. For EUI_total, the models maintained reliable performance (R^2 = 0.92 – 0.98), though slightly higher RMSE (3.1– 5.9) was expected because the total energy use aggregates errors from multiple subcomponents (cooling, heating, lighting). For EYI, predictive precision was generally high but varied slightly by climate. EYI achieved R^2 values of 0.87– 0.99 with RMSE around 0.3– 2.6.

These results demonstrate that the proposed ensemble framework delivers consistent predictive performance across different climates and energy use indicators, providing a solid basis for the subsequent SHAP-based interpretability analysis.

4.3. SHAP-based interpretation analysis

The SHAP summary plots (Fig. 8) provide a comprehensive interpretation of how design variables influence energy performance across the three representative climate zones. As the preliminary analysis in Fig. 3 already revealed substantial differences among the datasets of different climates, this section reports and discusses SHAP results separately by climate zone. Fig. 9 further presents normalized radar charts summarizing the overall influence of design variables for each zone, allowing a clearer comparison of the relative importance, and

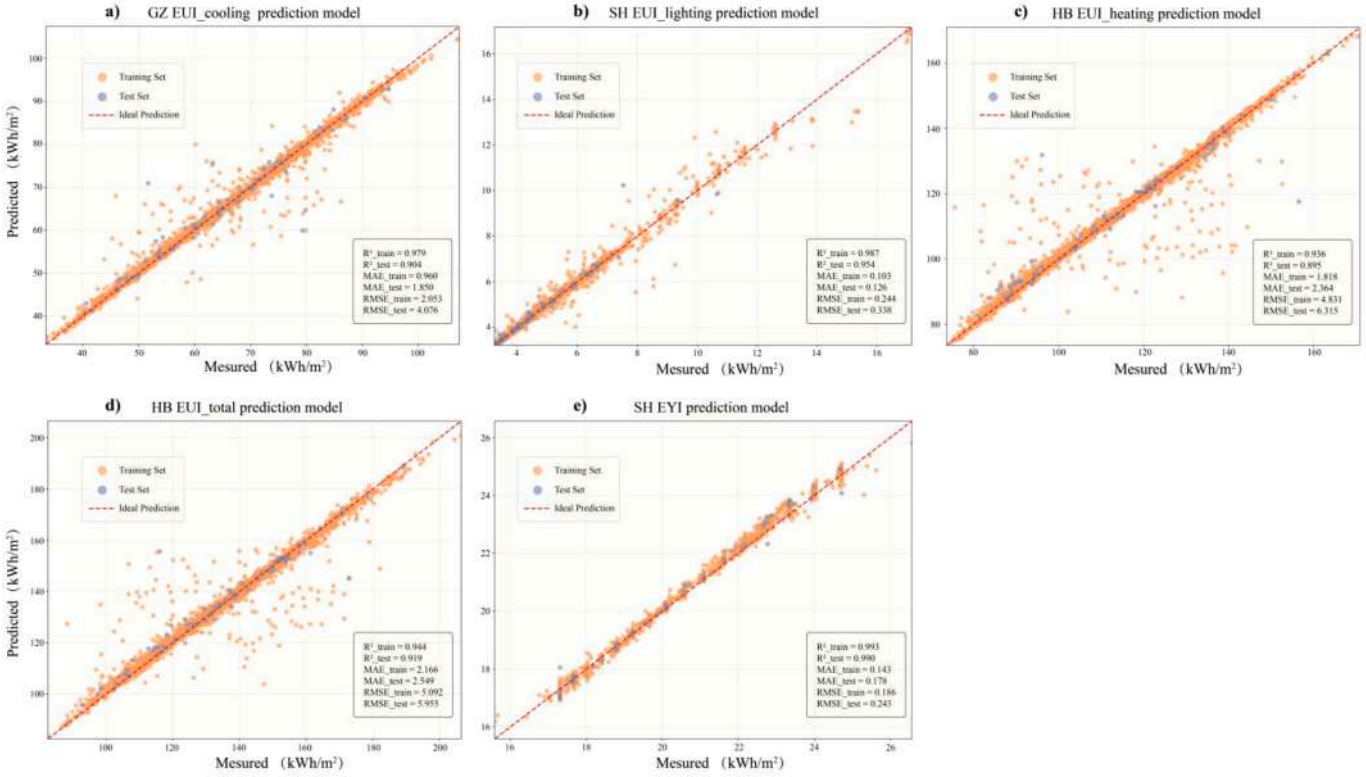

Fig. 6. Parity plots of model predictions versus actual values for five performance indicators.

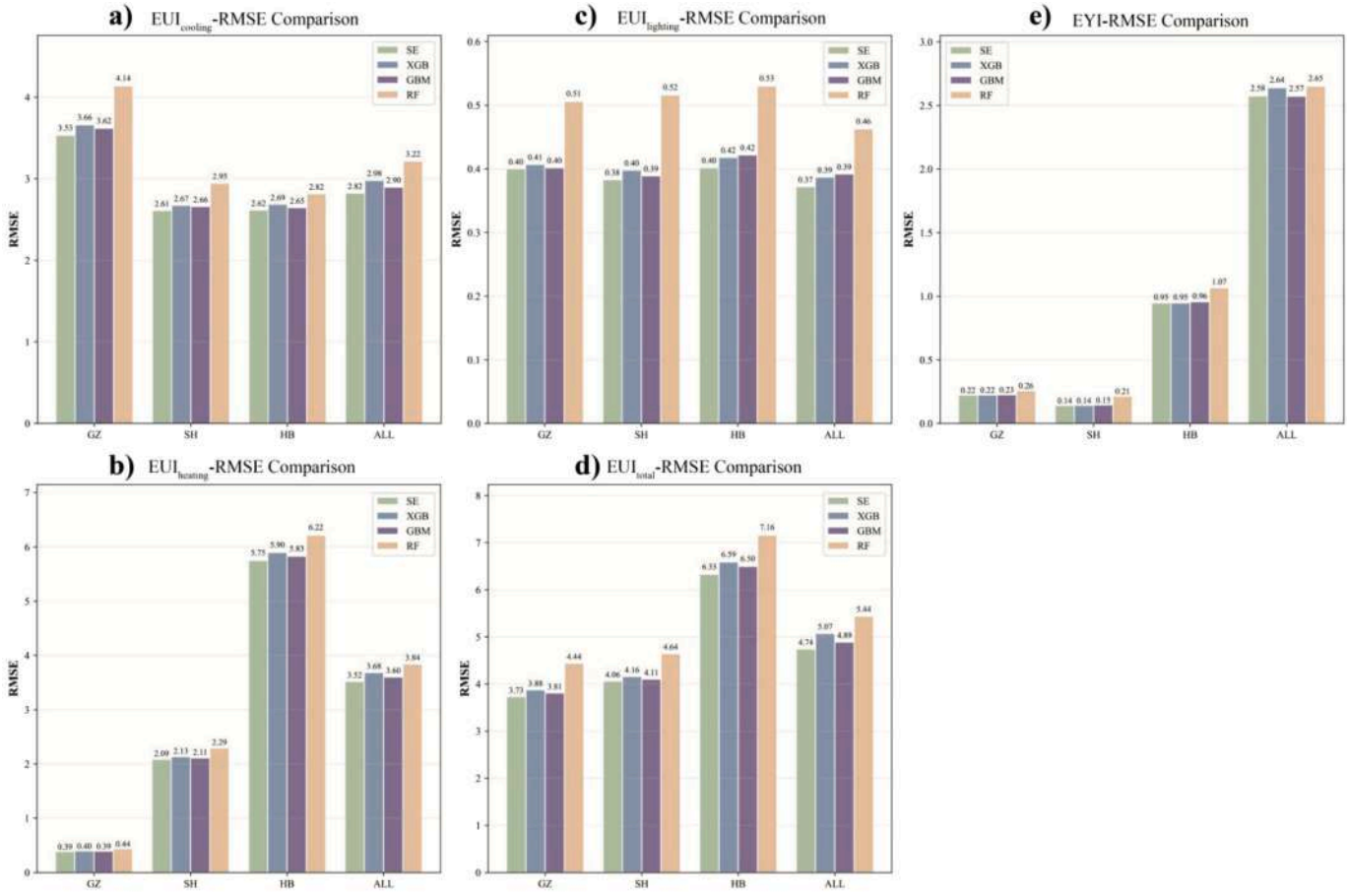

Fig. 7. RMSE Comparison between different models (Validation frame).

Table 4
Performance metrics of all best models across climate zone.

		ALL		GZ		SH		HB	
		Training	Testing	Training	Testing	Training	Testing	Training	Testing
EUI_cooling	RMSE	1.8373	2.8725	2.0531	4.0762	1.8656	2.1961	1.3720	2.0848
	MAE	0.8363	1.2669	0.9599	1.8470	0.8090	1.0658	0.6450	1.1269
	R^2	0.9936	0.9844	0.9794	0.9039	0.9587	0.9455	0.9692	0.9326
EUI_heating	RMSE	2.2742	3.7715	0.2689	0.4147	1.8685	1.5382	4.8313	6.3151
	MAE	0.6980	1.1124	0.1072	0.1570	0.7159	0.6750	1.8179	2.3637
	R^2	0.9978	0.9937	0.9633	0.9130	0.9238	0.9441	0.9356	0.8952
EUI_lighting	RMSE	0.1696	0.3526	0.2724	0.2469	0.2439	0.3379	0.1316	0.4129
	MAE	0.0730	0.1364	0.0919	0.1108	0.1033	0.1262	0.0673	0.1940
	R^2	0.9944	0.9741	0.9861	0.9867	0.9869	0.9538	0.9966	0.9696
EUI_total	RMSE	2.9028	4.7846	2.2496	3.8168	3.3934	3.0640	5.0921	5.9545
	MAE	1.2078	1.9652	1.0757	1.6905	1.4820	1.5033	2.1664	2.5485
	R^2	0.9924	0.9789	0.9767	0.9289	0.9343	0.9474	0.9443	0.9194
EYI	RMSE	2.4701	2.6199	0.1533	0.2115	0.1856	0.2428	0.7694	1.5280
	MAE	1.5696	1.6345	0.0463	0.0656	0.1432	0.1779	0.1937	0.3282
	R^2	0.9961	0.9864	0.9825	0.9654	0.9930	0.9903	0.9652	0.8674

(Units: RMSE and MAE share the units of the target variable. EUI and EYI are reported in kWh/(m^2·yr)).

ranking of morphological features.

Among the geometry variables, the Ratio_SLF exhibits the strongest effect on both EUI_cooling and EYI predictions. Overall, increasing Ratio_SLF,tends to raise cooling energy demand while reducing lighting energy use and photovoltaic yield. The average model contribution of Ratio_SLF ranges from 48 % to 71 % for EUI_cooling and from 68 % to 78 % for EYI.

The Rdncy_height is the dominant predictor for EUI_heating, with an average model contribution of 48 %–59 %. Increasing redundant height consistently elevates heating energy demand across all climate zones. Notably, when considering EUI_total , its effect is most pronounced in Harbin, where the model contribution reaches 41 %, significantly higher than in Shanghai or Guangzhou.

Num_seat, which indirectly determines total volume and occupant density, shows bimodal behavior: SHAP values for EUI_cooling are negative, larger arenas dilute internal gains per unit volume. Whereas for EUI_heating they become positive, as increased envelope area amplifies conductive losses.

The Types_roof shows a moderate yet meaningful influence on the EYI prediction model, with an average contribution ranging from 10 % to 13 %. Roof morphology affects not only the overall solar exposure of the building but also the orientation and tilt configuration of photovoltaic panels, thereby influencing the total potential of on-site energy generation.

The SHAP heatmap and dependence and force plots (Figs. 10–11) provide deeper insights into the nonlinear response mechanisms and cross-variable interactions behind the global SHAP rankings.

In the cooling-dominated Guangzhou case, In the cooling-dominated Guangzhou case, the dependence plot of Ratio_SLF shows a clear nonlinear threshold: a distinct nonlinear threshold. As the skylight ratio increases to approximately 15%, the SHAP value for EUI_cooling shifts from negative to strongly positive, indicating that its average influence on the cooling energy prediction becomes positive. In contrast, the dependence plot for EUI_lighting shows that beyond 5%, the model-predicted impact of Ratio_SLF becomes negative, with a gradually diminishing rate of change. The dependence plot of EUI_total further highlights the critical role of skylights in reducing overall energy consumption through decreased lighting demand, despite the associated cooling penalties at higher skylight ratios.

In indoor arena design, skylights are often one of the most debated elements. During competitive events, they are typically shaded to maintain a stable lighting environment and prevent glare, whereas for everyday public sports and training use, skylights effectively reduce lighting energy demand and improve visual comfort (Shanghai Municipal Bureau of Sports, 2016). For subtropical climates such as Guangzhou, however, preventing overheating due to excessive solar radiation remains a key design challenge. The SHAP analysis captures this nonlinear dual behavior remarkably well, quantifying how the same design parameter can alternately contribute to energy savings or penalties depending on climatic and operational contexts.

Overall, the SHAP-based interpretation reveals that environmental and morphological parameters collectively shape the energy behavior of large indoor arenas through nonlinear and synergistic effects. Adaptive roof geometries—optimizing skylight ratio, orientation, and redundant height under given climatic conditions—can effectively balance the competing objectives of daylight utilization and energy efficiency, providing architects with clear, data-driven guidance for early-stage design optimization. These SHAP-derived patterns represent associative relationships captured by the surrogate model. While the observed thresholds (e.g., the ~15% skylight ratio inflection point) are consistent with known building physics principles, where increased glazing area leads to higher solar heat gain—they should be understood as model-based attributions that warrant further validation through targeted parametric simulations or empirical measurements.

4.4. Implications for design optimization

The results of the correlation analysis and SHAP-based interpretation offer several implications for optimizing design in large indoor arenas. First, the identification of climate-specific dominant factors highlights the need to prioritize different design strategies across regions. In Guangzhou, where cooling demand significantly drives total energy consumption, design optimization should focus on reducing cooling intensity through careful control of skylight ratios and roof morphologies. By contrast, in Harbin, heating demand is the primary driver, and reducing redundant roof height becomes a more effective strategy. Shanghai presents mixed conditions, requiring a balanced consideration of both cooling and heating factors.

Second, the nonlinear dependencies revealed by SHAP dependence plots suggest the existence of threshold effects that can inform more efficient design interventions. For example, increasing the skylight ratio improves daylighting efficiency when the ratio is below approximately 15% in Guangzhou; however, beyond this point, the marginal benefit diminishes, while cooling penalties increase sharply. Recognizing such critical thresholds helps avoid over-design and allows architects to balance daylighting benefits against cooling loads more effectively.

Third, the consistent weak correlation between energy production and energy consumption suggests that energy generation and demand should be optimized independently. This finding highlights the importance of integrating passive energy-saving measures with active energy production strategies, rather than relying solely on PV systems to offset high cooling or heating demands directly.

Fig. 8. SHAP summary plots for energy performance indicators across climate zones.

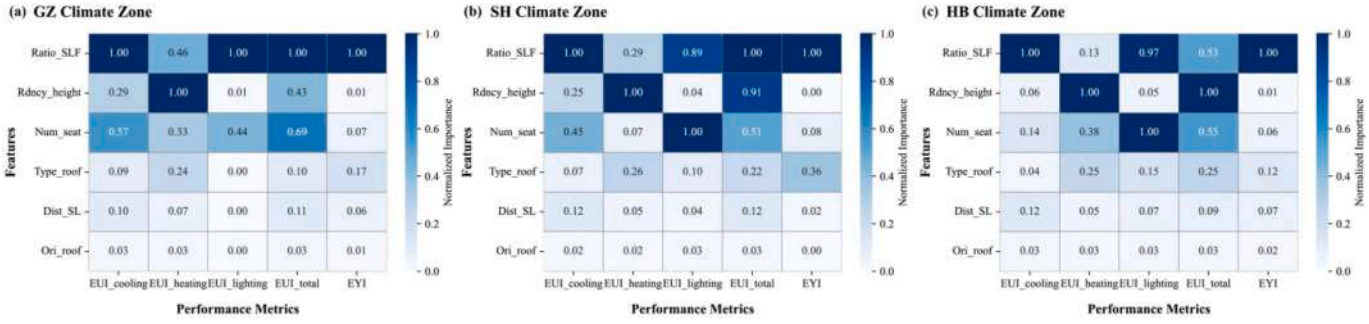

Fig. 9. Feature importance Heatmap comparison cross three climate zones (normalized mean SHAP).

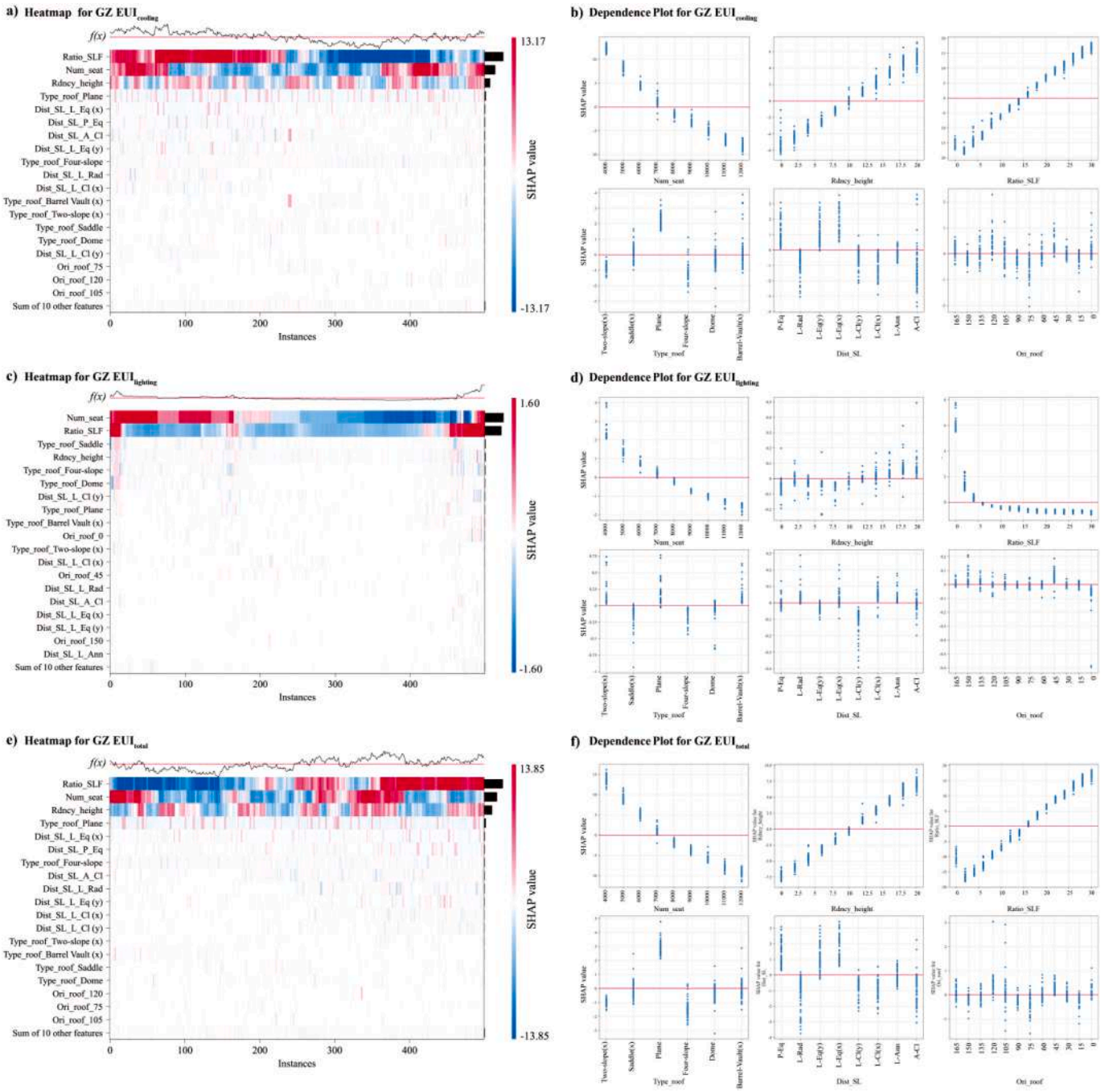

Fig. 10. Typical feature attribution analysis for GZ.

Finally, explainable machine learning provides a practical framework for early-stage design optimization. By quantifying the relative influence of design parameters, the proposed approach allows architects and engineers to narrow down the design space and prioritize high-sensitivity parameters, thereby reducing trial-and-error iterations in simulation-based workflows. This methodological contribution demonstrates how interpretable ML models can be embedded into parametric design processes to support multi-objective optimization and informed decision-making in indoor arena design.

4.5. Evaluation of the proposed framework

This subsection evaluates the adaptability and robustness of the proposed interpretable framework across different climate zones. This subsection evaluates the robustness and scalability of the proposed interpretable framework under different climatic conditions. The framework integrates a Simulation Parametric Model (SPM) with an automated performance analysis toolchain built on Grasshopper and Python, enabling large-scale sampling, batch simulation, and data-driven learning. Through this integration, it supports the generation and interpretation of extensive multi-objective performance datasets across diverse climates and design configurations. The study encompassed three representative cities: Harbin, Shanghai, and Guangzhou, covering typical *Köppen–Geiger* climate types: cold continental (Dwb), hot summer/cold winter (Cfa), and hot summer/warm winter (Cwa). As summarized in Table 3, the stacked ensemble model maintained high predictive accuracy across all zones, with average R^2 values above 0.95 for all performance indicators. Slight deviations were observed in the

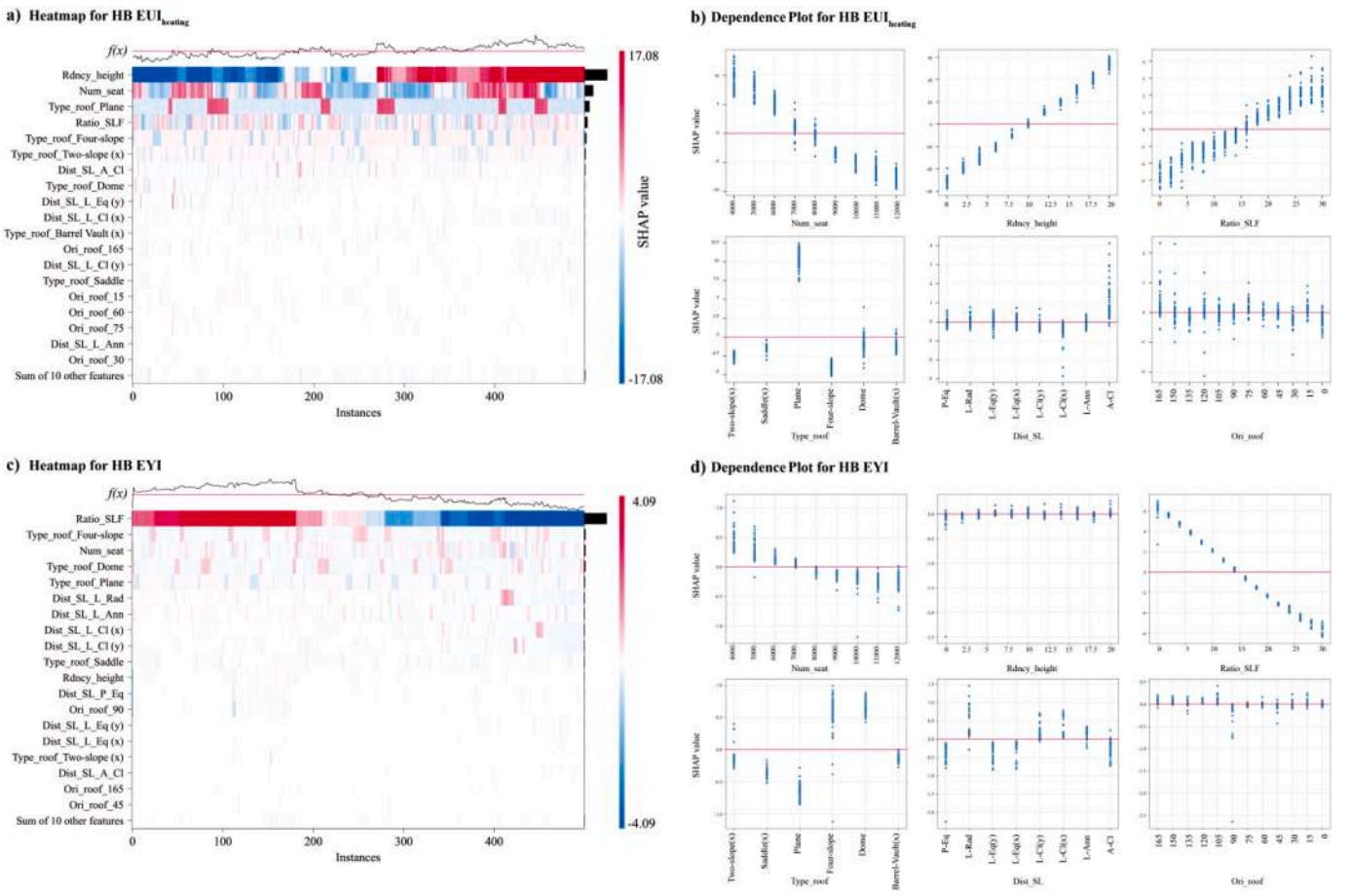

Fig. 11. Typical feature attribution analysis for HB.

cooling-dominated Guangzhou case, where higher solar intensity and envelope sensitivity increased the nonlinearity of skylight-related parameters. Nevertheless, the model consistently captured the main design performance interactions across climates.

Because the framework combines parametric modeling, automated data processing, and explainable prediction, it achieves high efficiency in both data generation and knowledge extraction. Once trained for a specific location, the ensemble model functions as a fast and reusable surrogate, enabling rapid performance prediction and SHAP-based interpretation without rerunning computationally intensive simulations. Overall, the large-scale, climate-based dataset and consistent model accuracy demonstrate the adaptability, scalability, and interpretability of the proposed framework for early-stage, performance-oriented architectural design.

Regarding computational efficiency, the simulation data generation phase required approximately 160 h on a standard workstation (Intel i7-10700, 32 GB RAM), while the complete model training process for all 15 surrogate models took approximately 45 min. Once trained, single-point predictions require less than 0.01 s, enabling near-instantaneous performance feedback during design iteration.

While the trained models are specific to the three Chinese climate zones studied, the framework itself is designed to be transferable. Applying this approach to other climatic or regulatory contexts would require regenerating simulation datasets using local weather data and construction standards, followed by model retraining. The modular architecture ensures that only the data generation and model training steps need to be repeated, while the overall workflow remains unchanged. Future work could explore transfer learning techniques to reduce data requirements when extending the framework to new regions.

5. Conclusions, limitations, and future work

5.1. Conclusions

This study proposes an integrated framework that combines parametric simulation, stacked ensemble modeling, and SHAP-driven interpretability to investigate the performance of design features in indoor arenas. A total of 9667 parametric design scenarios were generated across three representative climate zones: Guangzhou, Shanghai, and Harbin to train and evaluate the models. Unlike conventional simulation-based or black-box machine learning approaches, the proposed framework ensures predictive robustness across diverse climatic contexts while providing transparent, feature-level insights that can directly inform architectural design and control strategies.

The originality of this research lies in three key aspects:

1) Coupling geometric design features with multi-dimensional performance indicators. By encoding skylight ratio, roof redundancy height, orientation, and morphology into a systematic parametric dataset, this study demonstrates how architectural form-making decisions directly translate into quantifiable impacts on cooling, heating, lighting, and photovoltaic generation.
2) Establishing climate-specific mechanisms of influence. The SHAP-based interpretation in this study reveals how the dominance of cooling, heating, or combined loads shifts across Guangzhou, Shanghai, and Harbin, and identifies threshold effects (e.g., skylight ratio saturation around 10% in Guangzhou) that offer precise guidance for design intervention.
3) Embedding explainable machine learning into an automated performance evaluation pipeline. Rather than applying ML purely for

prediction, the study integrates it into a closed-loop workflow that supports sensitivity screening, design-space reduction, and multi-objective trade-off exploration, thus bridging performance modeling and adaptive decision-making.

By consolidating these contributions, the research provides a new paradigm for evaluating and optimizing large-span public buildings. The findings highlight that the climate zone remains the overarching determinant of energy demand. However, design-specific parameters, such as skylight ratio and roof redundancy height, can decisively alter performance outcomes when considered within their nonlinear thresholds. The proposed framework demonstrates how performance-oriented design can be grounded in data-driven evidence while retaining architectural agency, offering a replicable method for other building types and climates.

5.2. Limitations and future work

While this study demonstrates the value of integrating explainable machine learning with parametric simulation for early-stage design support, several limitations should be acknowledged, along with promising directions for future research.

(1) **Scope limitations.** The current framework focuses on energy consumption and photovoltaic generation as performance indicators. However, the framework is intentionally designed to be open and extensible. Future implementations can readily incorporate additional performance metrics such as thermal comfort (overheating risk, PMV/PPD), visual comfort (glare probability), ventilation effectiveness, and life-cycle environmental impacts. This extensibility would enable the framework to provide architects with a more comprehensive set of performance evaluations during early design phases, supporting informed decision-making across energy, comfort, and carbon considerations.

Beyond performance metrics, the parametric model currently concentrates on the geometrically complex roof system and main competition hall, with seating tiers modeled as adiabatic boundaries following typical arena topology. Additionally, the current simulation adopts standardized assumptions for occupancy schedules, HVAC setpoints, and envelope properties based on regional codes; in practice, these factors vary considerably across real arena operations and could modulate the relative sensitivity of energy performance to geometric parameters. Future extensions should incorporate explicit thermal mass modeling and expand the parametric scope to include auxiliary spaces, as well as diverse operational scenarios (e.g., HVAC system types, variable occupancy patterns, and advanced envelope configurations), for more comprehensive whole-building analysis. Furthermore, translating the current analytical framework into practical design workflows would require addressing additional challenges such as stakeholder preference elicitation, regulatory compliance checking, and the integration of non-geometric constraints including material selection, structural feasibility, and cost.

(2) **Interpretability and validation constraints.** SHAP provides associative, model-based attributions rather than establishing physical causation. Several observed patterns require further validation through controlled parametric studies, sensitivity analyses with physics-based simulation, or empirical measurements from real buildings. This is particularly true for the nonlinear threshold effects of skylight ratio and the interaction between roof geometry and EYI. Future work should systematically identify which SHAP-derived insights align with established building physics principles and which represent novel hypotheses requiring independent verification. Furthermore, the current study relies on simulation-generated data. Calibration against measured energy consumption from representative arenas would strengthen the validity of both simulation models and ML surrogates.

(3) **Data efficiency and generalization.** The current study employed nearly 10,000 samples across three climate zones. Systematic analysis of model robustness under reduced data availability, such as learning curve analysis to determine minimum sample requirements, would provide valuable guidance for practitioners with limited computational resources. Moreover, the current approach trains separate models for each climate zone independently. Recent advances in transfer learning (Wijaya et al., 2024) demonstrate the potential for knowledge transfer between related domains, which could improve model performance when extending to new climate zones with limited regional data.

(4) **Emerging methodological directions.** Several recent methodological advances offer promising pathways to address the limitations of the current framework. First, while SHAP provides valuable feature attributions, it does not yield explicit mathematical relationships between design parameters and energy performance. Symbolic regression algorithms and Kolmogorov-Arnold Networks (KAN), as explored in recent studies could derive closed-form expressions that offer more actionable and interpretable design guidance than black-box predictions (Wei et al., 2025). Second, integrating interpretable machine learning with metaheuristic optimization algorithms, as demonstrated by recent work, could potentially transform the current analysis-focused framework into a genuine design optimization tool capable of automatically searching for high-performance design configurations (Wang et al., 2025). In particular, recent computational design studies have demonstrated how multi-criteria reasoning can be embedded into geometry optimization workflows to negotiate competing performance objectives simultaneously (Zhang et al., 2026). These directions represent natural extensions that could significantly enhance the practical utility of the proposed framework.

CRediT authorship contribution statement

Yicheng Wang: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation, Conceptualization. **Peijun Lu:** Visualization, Software, Data curation. **Yimin Sun:** Supervision, Methodology, Funding acquisition, Conceptualization. **Mauro Berta:** Supervision, Investigation, Conceptualization. **Hao Wang:** Validation, Software.

Code and data availability

The automated simulation and data-generation workflow presented in this study was developed by the authors as a research prototype and has been registered as proprietary software. The current implementation relies on a specific software environment developed during the early stages of the research, including GH-Python scripting based on Python 2.7.

As recent versions of Grasshopper now support Python 3 and provide improved interoperability with contemporary simulation tools, the authors are actively refactoring the workflow to ensure compatibility with updated software ecosystems, improved numerical stability, and enhanced usability for architects and engineers. The long-term objective is to release the framework as an open-source, practice-oriented plugin following comprehensive testing, documentation, and interface optimization.

Although the source code is not publicly released at this stage, detailed descriptions of the simulation framework, parameter definitions, hyperparameter configurations, and dataset structure are fully documented in the manuscript and Appendix to support transparency

and reproducibility.

Declaration of competing interest

This work was supported by [National Natural Science Foundation of China, Grant No. 52178016]. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors declare no other competing interests.

Data availability

Data will be made available on request.

References

Airlangga, G., Liu, A., 2025. A hybrid gradient boosting and neural network model for predicting urban happiness: integrating ensemble learning with deep representation for enhanced accuracy. Mach. Learn. Knowl. Extr. 7, 4. https://doi.org/10.3390/make7010004.

Ali, A., Jayaraman, R., Mayyas, A., Alaifan, B., Azar, E., 2023. Machine learning as a surrogate to building performance simulation: predicting energy consumption under different operational settings. Energy Build. https://doi.org/10.1016/j.enbuild.2023.112940.

Alm, J., 2012. World Stadium Index: Stadiums Built for Major Sporting Events — Bright Future or Future Burden? Danish Institute for Sports Studies & Play the Game.

Alshboul, O., Shehadeh, A., Almasabha, G., Almuflih, A., 2022. Extreme gradient boosting-based machine learning approach for green building cost prediction. Sustainability. https://doi.org/10.3390/su14116651.

Amini Toosi, H., Del Pero, C., Leonforte, F., Lavagna, M., Aste, N., 2023. Machine learning for performance prediction in smart buildings: photovoltaic self-consumption and life cycle cost optimization. Appl. Energy 334, 120648. https://doi.org/10.1016/j.apenergy.2023.120648.

Baptista, M., Goebel, K., Henriques, E., 2022. Relation between prognostics predictor evaluation metrics and local interpretability SHAP values. Artif. Intell. 306, 103667. https://doi.org/10.1016/j.artint.2022.103667.

Boutahri, Y., Tilioua, A., 2024. Machine learning-based predictive model for thermal comfort and energy optimization in smart buildings. Results Eng. https://doi.org/10.1016/j.rineng.2024.102148.

Chai, T., Draxler, R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? – arguments against avoiding RMSE in the literature. Geosci. Model Dev. (GMD) 7, 1247–1250. https://doi.org/10.5194/GMD-7-1247-2014.

Chicco, D., Warrens, M., Jurman, G., 2021. The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. PeerJ Comput. Sci. 7. https://doi.org/10.7717/peerj-cs.623.

Cho, H., Kim, Y., Lee, E., Choi, D., Lee, Y., Rhee, W., 2020. Basic enhancement strategies when using bayesian optimization for hyperparameter tuning of deep neural networks. IEEE Access 8, 52588–52608. https://doi.org/10.1109/access.2020.2981072.

Ciardiello, A., Rosso, F., Dell'Olmo, J., Ciancio, V., Ferrero, M., Salata, F., 2020. Multi-objective approach to the optimization of shape and envelope in building energy design. Appl. Energy 280, 115984. https://doi.org/10.1016/j.apenergy.2020.115984.

De Winter, J., Gosling, S., Potter, J., 2016. Comparing the pearson and spearman correlation coefficients across distributions and sample sizes: a tutorial using simulations and empirical data. Psychol. Methods 21 (3), 273–290. https://doi.org/10.1037/met0000079.

Depecker, P., Menezo, C., Virgone, J., Lepers, S., 2001. Design of buildings shape and energetic consumption. Build. Environ. 36 (5), 627–635. https://doi.org/10.1016/S0360-1323(00)00044-5.

Dinmohammadi, F., Shafiee, M., 2023. Predicting energy consumption in residential buildings using advanced machine learning algorithms. Energies. https://doi.org/10.3390/en16093748.

Dong, Y., Duan, H., Li, X., Zhang, R., 2024. Influence of different forms on BIPV gymnasium carbon-saving potential based on energy consumption and solar energy in multi-climate zones. Sustainability 16 (4), 1656. https://doi.org/10.3390/su16041656.

Dostmohammadi, M., Pedram, M., Hoseinzadeh, S., Garcia, D., 2024. A GA-stacking ensemble approach for forecasting energy consumption in a smart household: a comparative study of ensemble methods. J. Environ. Manag. 364, 121264. https://doi.org/10.1016/j.jenvman.2024.121264.

Du, X., Li, Y., Zheng, F., Xu, Y., 2025. Comparison of light-thermal environment and energy consumption in ice sports venues between China and Finland. Energy Build. 349, 116523. https://doi.org/10.1016/j.enbuild.2025.116523.

Elnour, M., Fadli, F., Himeur, Y., Petri, I., Rezgui, Y., Meskin, N., Ahmad, A.M., 2022. Performance and energy optimization of building automation and management systems: towards smart sustainable carbon-neutral sports facilities. Renew. Sustain. Energy Rev. 162, 112401. https://doi.org/10.1016/j.rser.2022.112401.

Fang, Y., Cho, S., 2019. Design optimization of building geometry and fenestration for daylighting and energy performance. Sol. Energy. https://doi.org/10.1016/j.solener.2019.08.039.

Fang, Z., Lei, L., Zheng, R., 2025. A novel building sampling approach leveraging data mining and stratified sampling theory for energy optimization. Energy Build. https://doi.org/10.1016/j.enbuild.2025.115366.

Feng, X., Lou, X., Lv, H., Su, Y., 2024. Study on lightweight design of tensegrity structures with multi-self-stress modes. Structures 65, 106750. https://doi.org/10.1016/j.istruc.2024.106750.

Flyvbjerg, B., 2011. Over budget, over time, over and over again. Inthe Oxford Handbook of Project Management. Oxford University Press, pp. 321–344. https://doi.org/10.1093/oxfordhb/9780199563142.003.0014.

Gassar, A., Koo, C., Kim, T., Cha, S., 2021. Performance optimization studies on heating, cooling and lighting energy systems of buildings during the design stage: a review. Sustainability. https://doi.org/10.3390/su13179815.

Geyer, P., Singaravel, S., 2018. Component-based machine learning for performance prediction in building design. Appl. Energy. https://doi.org/10.1016/J.APENERGY.2018.07.011.

Granadeiro, V., Duarte, J.P., Correia, J.R., Leal, V.M.S., 2013. Building envelope shape design in early stages of the design process: integrating architectural design systems and energy simulation. Autom. ConStruct. 32, 196–209. https://doi.org/10.1016/j.autcon.2012.12.003.

Guo, F., Miao, S., Xu, S., Luo, M., Dong, J., Zhang, H., 2024. Multi-objective optimization design for cold-region office buildings balancing outdoor thermal comfort and building energy consumption. Energies. https://doi.org/10.3390/en18010062.

Han, T., Huang, Q., Zhang, A., Zhang, Q., 2018. Simulation-based decision support tools in the early design stages of a green building—A review. Sustainability. https://doi.org/10.3390/su10103696.

Hassan, A.A., El-Rayes, K., 2024. Optimal use of renewable energy technologies during building schematic design phase. Appl. Energy 353, 122006. https://doi.org/10.1016/j.apenergy.2023.122006.

Hensel, M., 2013. Performance-Oriented Architecture: Rethinking Architectural Design and the Built Environment. Wiley.

Hooshyar, D., Yang, Y., 2024. Problems with SHAP and LIME in interpretable AI for education: a comparative study of post-hoc explanations and neural-symbolic rule extraction. IEEE Access 12, 137472–137490. https://doi.org/10.1109/access.2024.3463948.

Huang, X., Ma, X., Zhang, Q., 2019. Effect of building interface form on thermal comfort in gymnasiums in hot and humid climates. Front. Archit. Res. 8 (1), 32–43. https://doi.org/10.1016/j.foar.2018.11.002.

International Energy Agency (IEA), 2021. World energy outlook 2021. https://www.iea.org/reports/world-energy-outlook-2021.

Javanmard, Z., Davtalab, J., Nikpour, M., Sivandipour, A., 2024. Integrating machine learning and parametric design for energy-efficient building cladding systems in arid climates: sport hall in Kerman. J. Build. Eng. 97, 110693. https://doi.org/10.1016/j.jobe.2024.110693.

Junlin, Z., Longwei, Z., 2021. Multi objective optimization of window to wall ratio of university gymnasium in severe cold area by coupling natural lighting and energy consumption. E3S Web of Conferences 293, 02044. https://doi.org/10.1051/e3sconf/202129302044.

Kesireddy, A., Medrano, F., 2024. Elite multi-criteria decision making - pareto front optimization in multi-objective optimization. Algorithms 17, 206. https://doi.org/10.3390/a17050206.

Konstantinov, A., Utkin, L., 2020. Interpretable machine learning with an ensemble of gradient boosting machines. ArXiv, abs/2010.07388. https://doi.org/10.1016/J.KNOSYS.2021.106993.

Li, Z., 2022. Extracting spatial effects from machine learning model using local interpretation method: an example of SHAP and XGBoost. Comput. Environ. Urban Syst. 96, 101845. https://doi.org/10.1016/j.compenvurbsys.2022.101845.

Li, Z., Chen, H., Lin, B., Zhu, Y., 2018. Fast bidirectional building performance optimization at the early design stage. Build. Simulat. 11, 647–661. https://doi.org/10.1007/S12273-018-0432-1.

Li, Z., Tian, M., Zhao, Y., Zhang, Z., Ying, Y., 2021. Development of an integrated performance design platform for residential buildings based on climate adaptability. Energies 14 (24), 8223. https://doi.org/10.3390/en14248223.

Li, Z., Tian, M., Zhu, X., Xie, S., He, X., 2022a. A review of integrated design process for building climate responsiveness. Energies. https://doi.org/10.3390/en15197133.

Li, S., Tao, Z., Liu, X., Xue, Z., Liu, X., 2022b. Performance investigation of a grid-connected system integrated photovoltaic, battery storage and electric vehicles: a case study for gymnasium building. Energy Build. https://doi.org/10.1016/j.enbuild.2022.112255.

Li, W., Hong, W., Xu, X., Makvandi, M., Zhang, B., Chen, Q., Yuan, P., 2024. A radical shift to probabilistic thinking: integrating energy analysis in performance-driven building morphology generation. Build. Environ. https://doi.org/10.1016/j.buildenv.2024.112443.

Lin, B., Chen, H., Liu, Y., He, Q., Li, Z., 2020. A preference-based multi-objective building performance optimization method for early design stage. Build. Simulat. 14, 477–494. https://doi.org/10.1007/s12273-020-0673-7.

Liu, X., Tang, H., Ding, Y., Yan, D., 2022. Investigating the performance of machine learning models combined with different feature selection methods to estimate the energy consumption of buildings. Energy Build. https://doi.org/10.1016/j.enbuild.2022.112408.

Lu, M., Hou, Q., Qin, S., Zhou, L., Hua, D., Wang, X., Cheng, L., 2023. A stacking ensemble model of various machine learning models for daily runoff forecasting. Water. https://doi.org/10.3390/w15071265.

Luo, Z., Qi, X., Sun, C., Dong, Q., Gu, J., Gao, X., 2024. Investigation of influential variations among variables in daylighting glare metrics using machine learning and SHAP. Build. Environ. 254, 111394. https://doi.org/10.1016/j.buildenv.2024.111394.

Lyu, Y., Long, Z., Zhou, R., Gao, X., 2025. The passive optimization design of Large- and medium-sized gymnasiums in hot summer and cold winter regions oriented on energy saving: a case study of shanghai. Buildings 15 (15), 2745. https://doi.org/10.3390/buildings15152745.

Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2021. GB 55015-2021: General Code for Building Energy Efficiency and Renewable Energy Utilization. China Architecture & Building Press, Beijing.

Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2012. Code for Heating, Ventilation, and Air Conditioning Design of Civil Buildings: GB 50736-2012* [S]. China Architecture & Building Press, Beijing.

Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2013. Building Lighting Design Standard: GB 50033-2013 [S]. China Architecture & Building Press, Beijing.

Ministry of Housing and Urban-Rural Development of the People's Republic of China, 2016. Code for Thermal Design of Civil Buildings: GB 50176-2016 [S]. China Architecture & Building Press, Beijing.

Molnar, C., Konig, G., Bischl, B., Casalicchio, G., 2020. Model-agnostic feature importance and effects with dependent features: a conditional subgroup approach. Data Min. Knowl. Discov. 38, 2903–2941. https://doi.org/10.1007/s10618-022-00901-9.

Nazifi Charandabi, R., De Castro Motta, J., Carpentieri, G., Babilio, E., Bonada, J., Fantuzzi, N., Fraternali, F., 2025. Deployability, mechanical response, and energy harvesting capacity of a novel solar roof for sports stadiums. Dev. Built Environ. 24, 100771. https://doi.org/10.1016/j.dibe.2025.100771.

Negendahl, K., 2015. Building performance simulation in the early design stage: an introduction to integrated dynamic models. Autom. ConStruct. 54, 39–53. https://doi.org/10.1016/j.autcon.2015.03.002.

Neubauer, A., Brandt, S., Kriegel, M., 2024. Relationship between feature importance and building characteristics for heating load predictions. Appl. Energy 359, 122668. https://doi.org/10.1016/j.apenergy.2024.122668.

Neubauer, A., Brandt, S., Kriegel, M., 2025. Explainable multi-step heating load forecasting: using SHAP values and temporal attention mechanisms for enhanced interpretability. Energy AI 20, 100480. https://doi.org/10.1016/j.egyai.2025.100480.

Østergård, T., Jensen, R.L., Maagaard, S.E., 2016. Building simulations supporting decision making in early design – a review. Renew. Sustain. Energy Rev. 61, 187–201. https://doi.org/10.1016/j.rser.2016.03.045.

Pan, W., Turrin, M., Louter, C., Sarıyıldız, S., Sun, Y., 2019. Integrating multi-functional space and long-span structure in the early design stage of indoor sports arenas by using parametric modelling and multi-objective optimization. J. Build. Eng. https://doi.org/10.1016/J.JOBE.2019.01.006.

Pan, W., Sun, Y., Turrin, M., Louter, C., Sarıyıldız, S., 2020. Design exploration of quantitative performance and geometry typology for indoor arena based on self-organizing map and multi-layered perceptron neural network. Autom. ConStruct. https://doi.org/10.1016/j.autcon.2020.103163.

Pan, Y., Zhu, M., Lv, Y., Yang, Y., Liang, Y., Yin, R., Yang, Y., Jia, X., Wang, X., Zeng, F., Huang, S., Hou, D., Xu, L., Yin, R., Yuan, X., 2023. Building energy simulation and its application for building performance optimization: a review of methods, tools, and case studies. Adv. Appl. Energy. https://doi.org/10.1016/j.adapen.2023.100135.

Pavlyshenko, B., 2018. Using stacking approaches for machine learning models. 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), pp. 255–258. https://doi.org/10.1109/DSMP.2018.8478522.

Picard, C., Schiffmann, J., 2021. Realistic constrained multiobjective optimization benchmark problems from design. IEEE Trans. Evol. Comput. 25, 234–246. https://doi.org/10.1109/tevc.2020.3020046.

Qian, F., Shi, Z., Yang, L., 2024. A review of green, low-carbon, and energy-efficient research in sports buildings. Energies 17 (16), 4020. https://doi.org/10.3390/en17164020.

Qian, F., Shi, Z., Yang, L., 2025. AI-Assisted multidimensional optimization of thermal and morphological performance in small-to-medium sports buildings. Appl. Sci. 15 (18), 9912. https://doi.org/10.3390/app15189912.

Raji, B., Tenpierik, M., Dobbelsteen, A., 2017. Early-stage design considerations for the energy-efficiency of high-rise office buildings. Sustainability 9, 623. https://doi.org/10.3390/su9040623.

Razmi, A., Rahbar, M., Bemanian, M., 2022. PCA-ANN integrated NSGA-III framework for dormitory building design optimization: energy efficiency, daylight, and thermal comfort. Appl. Energy 305, 117828. https://doi.org/10.1016/j.apenergy.2021.117828.

Salata, F., Ciardiello, A., Dell'Olmo, J., Ciancio, V., Ferrero, M., Rosso, F., 2024. Geometry optimization in the schematic design phase of low-energy buildings for all European climates through genetic algorithms. Sustain. Cities Soc. 112, 105639. https://doi.org/10.1016/j.scs.2024.105639.

Seyyedzadeh, M., Zeynali, P., Tajdid, A., Ghavifekr, A., Behinfaraz, R., 2025. Comparative evaluation of machine learning methods for predicting energy consumption in buildings. 2025 10th International Conference on Technology and Energy Management (ICTEM), pp. 1–5. https://doi.org/10.1109/ictem66196.2025.11063498.

Shanghai Municipal Bureau of Sports, 2016. Shanghai Municipal Bureau of Quality and Technical Supervision. Guideline of Efficient Energy Conservation for Large and medium-sized Sport Buildings (DB31/T 989-2016). Shanghai Standards Press, Shanghai, China.

Sharma, S., Kumar, V., 2022. A comprehensive review on multi-objective optimization techniques: past, present and future. Arch. Comput. Methods Eng. 29, 5605–5633. https://doi.org/10.1007/s11831-022-09778-9.

Shen, Y., Pan, Y., 2023. BIM-supported automatic energy performance analysis for green building design using explainable machine learning and multi-objective optimization. Appl. Energy. https://doi.org/10.1016/j.apenergy.2022.120575.

Shields, M., Zhang, J., 2015. The generalization of Latin hypercube sampling. Reliab. Eng. Syst. Saf. 148, 96–108. https://doi.org/10.1016/j.ress.2015.12.002.

Song, B., Zhou, R., Ahmed, F., 2023. Multi-modal machine learning in engineering design: a review and future directions. J. Comput. Inf. Sci. Eng. 24. https://doi.org/10.48550/arXiv.2302.10909.

Stefańska, A., Rokicki, W., 2022. Architectural design optimisation in reticulated free-form canopies. Buildings 12 (8), 1068. https://doi.org/10.3390/buildings12081068.

Sun, Yimin, 2009. Beijing olympic wrestling arena. Architectural Creation (12), 36–37.

Sun, Weihua, 2022. National short track speed skating arena. Contemporary Architect. (6), 60–67.

Sun, Yimin, et al., 2012. Chang'an sports hall, dongguan, Guangdong, 2009-2013. Architectural Creation (7), 72–79.

Tian, Y., Si, L., Zhang, X., Cheng, R., He, C., Tan, K., Jin, Y., 2021. Evolutionary large-scale multi-objective optimization: a survey. ACM Comput. Surv. 54, 1–34. https://doi.org/10.1145/3470971.

Turrin, M., Von Buelow, P., Stouffs, R., 2011. Design explorations of performance driven geometry in architectural design using parametric modeling and genetic algorithms. Adv. Eng. Inform. 25 (4), 656–675.

Tyagi, S., Singh, P., 2020. Short term and long term building electricity consumption prediction using extreme gradient boosting. Recent. Adv. Comput. Sci. Commun/. https://doi.org/10.2174/2666255813666201218160223.

United Arab Emirates University of Sharjah, Maksoud, A., Al-Beer, H. B., United Arab Emirates University of Sharjah, Ali Hussien, A., United Arab Emirates University of Sharjah, Dirar, S., United Arab Emirates University of Sharjah, Mushtaha, E., United Arab Emirates University of Sharjah, Yahia, M. W., & United Arab Emirates University of Sharjah, 2023. Computational design for futuristic environmentally adaptive building forms and structures. Archit. Eng. 8 (1), 13–24. https://doi.org/10.23968/2500-0055-2023-8-1-13-24.

W., S., & M., S EVALUATION OF CAD-INTEGRATED BUILDING ENERGY SIMULATION TOOLS FOR THE EARLY DESIGN STAGE, 2025. JoDA Journal of Digital Architecture. https://doi.org/10.24167/joda.v5i1.14216.

Wang, R., Lu, S., Feng, W., 2020. A novel improved model for building energy consumption prediction based on model integration. Appl. Energy. https://doi.org/10.1016/j.apenergy.2020.114561.

Wang, P., Hu, J., Chen, W., 2023. A hybrid machine learning model to optimize thermal comfort and carbon emissions of large-space public buildings. J. Clean. Prod. https://doi.org/10.1016/j.jclepro.2023.136538.

Wang, X., Teigland, R., Hollberg, A., 2024a. Identifying influential architectural design variables for early-stage building sustainability optimization. Build. Environ. 252, 111295. https://doi.org/10.1016/j.buildenv.2024.111295.

Wang, Y., Sun, Y., Lu, P., Berta, M., 2024b. An integrated climate-based daylight performance evaluation framework for indoor arenas' roof system. J. Build. Eng. 90, 109164. https://doi.org/10.1016/j.jobe.2024.109164.

Wang, K., Shen, T., Wei, J., Liu, J., Hu, W., 2025. An intelligent framework for deriving formulas of aerodynamic forces between high-rise buildings under interference effects using symbolic regression algorithms. J. Build. Eng. 99, 111614. https://doi.org/10.1016/j.jobe.2024.111614.

Wei, J., Shen, T., Wang, K., Liu, J., Wang, S., Hu, W., 2025. Transfer learning framework for the wind pressure prediction of high-rise building surfaces using wind tunnel experiments and machine learning. Build. Environ. 271, 112620. https://doi.org/10.1016/j.buildenv.2025.112620.

Wijaya, D., Utami, S., Mangkuto, R., 2024. Multi-objective optimisation of skylight design parameters for a low-rise building in the tropics. Int. J. Technol. https://doi.org/10.14716/ijtech.v15i4.5484.

Wu, C., Pan, H., Luo, Z., Liu, C., Huang, H., 2024. Multi-objective optimization of residential building energy consumption, daylighting, and thermal comfort based on BO-XGBoost-NSGA-II. Build. Environ. 254, 111386. https://doi.org/10.1016/j.buildenv.2024.111386.

Wu, Y., Zhao, L., Sun, J., 2025. Analysis of the emergency shelter function of large-scale international sports event venues. Transp. Res. Interdiscip. Perspect. 34, 101680. https://doi.org/10.1016/j.trip.2025.101680.

Yang, D., Ren, S., Turrin, M., Sariyildiz, S., Sun, Y., 2018. Multi-disciplinary and multi-objective optimization problem re-formulation in computational design exploration: a case of conceptual sports building design. Autom. ConStruct. 92, 242–269. https://doi.org/10.1016/j.autcon.2018.03.023.

Yang, D., Di Stefano, D., Turrin, M., Sariyildiz, S., Sun, Y., 2020. Dynamic and interactive re-formulation of multi-objective optimization problems for conceptual architectural design exploration. Autom. ConStruct. 118, 103251. https://doi.org/10.1016/j.autcon.2020.103251.

Yu, H., Yang, W., Li, Q., Li, J., 2022. Optimizing buildings' life cycle performance while allowing diversity in the early design stage. Sustainability. https://doi.org/10.3390/su14148316.

Yue, N., Li, L., Morandi, A., Zhao, Y., 2021. A metamodel-based multi-objective optimization method to balance thermal comfort and energy efficiency in a campus gymnasium. Energy Build. 253, 1115.

Zhang, Z., 2024. Multi-objective optimization method for building energy-efficient design based on multi-agent-assisted NSGA-II. Energy Inform. 7. https://doi.org/10.1186/s42162-024-00394-4.

Zhang, Y., Teoh, B., Zhang, L., Chen, J., Wu, M., 2022. Data-driven estimation of building energy consumption and GHG emissions using explainable artificial intelligence. Energy. https://doi.org/10.1016/j.energy.2022.125468.

Zhang, W., Li, L., Li, Y., 2025a. Natural ventilation and energy consumption research for dry sports halls within national fitness centers in cold regions—case study of qingdao. Buildings 15 (5), 734. https://doi.org/10.3390/buildings15050734.
Zhang, Y., Peng, J., Wang, Z., Xi, M., Liu, J., Xu, L., 2025b. Machine learning-assisted sustainable mix design of waste glass powder concrete with Strength–Cost–CO2 emissions trade-offs. Buildings 15 (15), 2640. https://doi.org/10.3390/buildings15152640.
Zhang, Y., Wang, Z., Liu, J., Li, Y., Huang, Z., Yu, X., 2026. A metaheuristic-driven categorical boosting framework with interpretability for high-precision prediction of mechanical properties in corroded reinforced concrete beams. Eng. Appl. Artif. Intell. 163, 112804. https://doi.org/10.1016/j.engappai.2025.112804.
Zimbalist, A., 2016. Circus Maximus: the Economic Gamble Behind Hosting the Olympics and the World Cup. Brookings Institution Press.