



Climate-adaptive building design through explainable AI: 3D spatial layout automation and evolutionary optimization across climate zones

Peiyong Huang^b, Yanxiang Yang^a, Wen Gao^c, Xing Zheng^d, Pengyuan Shen^{a,*} 

^a Institute of Future Human Habitats, Shenzhen International Graduate School, Tsinghua University, Shenzhen, China

^b China State Construction Hailong Technology Company Limited, Shenzhen, China

^c College of Architecture and Urban Planning, Beijing University of Technology, Beijing, China

^d Department of Architecture and Civil Engineering, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong Special Administrative Region

ARTICLE INFO

Keywords:

Building simulation
Design optimization
Spatial layout
Machine learning
Climate-adaptive design
Explainable AI

ABSTRACT

Spatial layout significantly impacts building energy performance, yet systematic optimization methods across different climates remain limited. This research develops an integrated three-stage framework combining automated layout generation, evolutionary optimization, and explainable artificial intelligence (XAI) to reduce energy consumption in mixed-use office buildings. Using a typical eight-story office building, we conducted comparative analysis across five Chinese climate zones: severely cold (Harbin), cold (Beijing), hot summer-cold winter (Shanghai), subtropical (Shenzhen), and mild (Kunming). In Stage 1 - Layout Generation, grid-based algorithms with geometric constraints automatically generate energy-efficient spatial configurations. In Stage 2 - Optimization, evolutionary algorithms (SPEA-2 and HypE) integrated with building energy simulation minimize cooling and heating loads, generating over 1700 optimized solutions per climate zone. In Stage 3 - XAI Interpretation, random forest models predict energy performance with high accuracy ($R^2 = 0.801\text{--}0.874$), while SHAP analysis quantifies the contribution of 26 spatial layout features.

Results demonstrate substantial energy savings potential. Subtropical climate (Shenzhen) achieves the best absolute performance with 17.25 % reduction in total loads, while mild climate (Kunming) shows the highest percentage reduction at 24.91 %. Average energy savings across all climate zones range from 9.67 % to 13.60 % for heating-dominated regions. SHAP analysis reveals climate-specific design strategies. It is found that orientation area distribution is the most critical factor for subtropical climates, while space centralization and space adjacency optimization are essential for cold regions. This methodology provides architects and engineers with computationally efficient, evidence-based tools for climate-adaptive sustainable building design during early planning stages.

* Corresponding author.

E-mail address: pengyuan_pub@163.com (P. Shen).

Nomenclature

Abbreviation		
AI	Artificial Intelligence	
ANN	Artificial Neural Network	
HVAC	Heating, Ventilation, and Air Conditioning	
HypE	Hypervolume Estimation Algorithm	
IPCC	Intergovernmental Panel on Climate Change	
MAPE	Mean Absolute Percentage Error	
ML	Machine Learning	
MOEA/D	Multi-Objective Evolutionary Algorithm based on Decomposition	
NRMSE	Normalized Root Mean Square Error	
NSDE	Non-dominated Sorting Differential Evolution	
NSGA-II	Non-dominated Sorting Genetic Algorithm II	
R^2	Coefficient of Determination	
RF	Random Forest	
RMSE	Root Mean Square Error	
SHAP	SHapley Additive exPlanations	
SPEA-2	Strength Pareto Evolutionary Algorithm 2	
TMY	Typical Meteorological Year	
UNEP	United Nations Environment Programme	
WWR	Window-to-Wall Ratio	
XAI	Explainable Artificial Intelligence	

Variable	Description	Unit
$A_{generated}(z)$	Generated area for zone z	m^2
$A_{min}(z)$	Minimum area requirement for zone z	m^2
$A_{max}(z)$	Maximum area requirement for zone z	m^2
$A_{demand}(z)$	Required area for zone z	m^2
$A_{current}(z)$	Current area of zone z during growth	m^2
ϕ_0	Expected model output over baseline dataset	–
$\phi_i(x)$ $\phi_i(x)$	SHAP value for feature i	–
$h_t(x)$	Prediction from the t th decision tree	–
T	Number of decision trees in random forest	–
τ	Area tolerance parameter	–
U_{ijk}	Spatial unit at coordinates (i,j,k)	–

1. Introduction

The construction industry plays a significant role in global energy consumption in the face of the global climate change challenge. According to United Nations Environment Programme (UNEP) and the Global Construction Alliance, the global construction industry accounted for 32 % of total global energy consumption and 34 % of global carbon emissions as of 2023 [1]. With rising living standards globally, building energy consumption continues to grow, making energy efficiency interventions increasingly urgent. Meanwhile, IPCC suggested that the construction sector has great potential for emission reduction and relatively lower costs [2]. This combination of high emissions contribution and significant reduction potential positions the construction sector as a critical lever for achieving carbon peak and carbon neutrality goals. Building energy efficiency is an overall optimization issue that requires comprehensive consideration and collaboration from multiple fields through the entire design process, which needs to balance influencing factors such as passive and active design strategies. According to the ANNEX-30 project study by the International Energy Agency, the performance of buildings is largely influenced by the early design stage, and decisions made in the early design stage have more than 40 % potential for energy savings [3]. Therefore, understanding and optimizing early-stage design decisions is fundamental to achieving substantial energy reductions in the built environment. Among the various design factors affecting building energy performance, building envelope and shape have been extensively studied as primary determinants. Building envelope performance, including thermal insulation properties, window-to-wall ratios, and material selection, directly influences heat transfer between interior and exterior environments [4–7]. Evolutionary optimization approaches have been widely applied to envelope design across different climate and seismic zones, demonstrating significant potential in energy saving and carbon emission reduction [8].

Recent advances in AI-based and optimization-driven approaches also have significantly expanded the scope of computational building design. Advanced computational methods, including artificial neural network-based genetic algorithms, have proven effective in optimizing building geometry for thermal energy efficiency in public buildings [9]. Parametric design frameworks integrated with genetic algorithms have been applied to optimize climate-responsive building passive strategies [10]. On urban scale, they have also been applied to optimize urban morphology, building density, and street configurations for energy efficiency and environmental performance [11]. Deep learning methods, particularly generative adversarial networks, have demonstrated capability in predicting urban-scale energy consumption patterns and generating energy-efficient building forms [12]. Reinforcement learning approaches have been employed for real-time building energy management and HVAC control optimization [13]. Multi-objective optimization frameworks combining building performance simulation have enabled optimization of building energy systems [14].

Building shape factors, such as aspect ratio, compactness, and surface-to-volume ratio, substantially affect energy consumption patterns [15,16]. However, beyond envelope and shape optimization, studies in the past decade have revealed that building spatial

layout, which refers to the internal arrangement of functional spaces, represents another critical yet understudied dimension of early-stage energy efficiency design. It is found that effective spatial layout design can reduce unnecessary energy consumption and improve the overall sustainability of the building [17]. Hence, the impact of building spatial plans on energy consumption is a key factor in energy conservation design, especially in the early stages of building design [18–20]. Conventional design methods have struggled to meet complex optimization requirements, while computational design and machine learning models, as emerging tools, have provided important support for building performance optimization [21]. Computational design, through computer simulation and optimization algorithms, predicts the performance of the building layout in the early design stage and can evaluate the performance of different schemes in terms of energy efficiency, natural lighting, thermal comfort, etc. [22,23]. This computational design-based approach can improve design efficiency, guide design decisions, and provide an important basis for building energy conservation. Meanwhile, the machine learning methods can effectively facilitate the prediction of complex relationship between building layout and energy consumption by analyzing historical data [24]. Compared with traditional physical simulation, machine learning can handle more variables, perform evolutionary optimization, and propose optimal design schemes, especially under different climatic conditions [25]. These models provide data-driven decision support for design teams to help achieve energy-efficient design and emission reduction targets. By combining computational design with machine learning, architectural design can achieve an integrated optimization of multiple goals such as energy efficiency, comfort, and environmental adaptability. The emerging methods now equip designers with better approaches in rapid decision-making and drive progress in sustainable development and carbon reduction efforts in the construction industry.

The remainder of this paper is organized as follows: Section 2 presents a comprehensive literature review examining the impact of spatial layout on building energy performance, climate-adaptive design strategies, and the application of simulation-based optimization and machine learning methods in building performance research. Section 3 describes the research methodology and framework, including the automated spatial layout generation algorithm, machine learning-based energy prediction model, spatial layout design variables and feature engineering, optimization problem formulation, and case study building configuration across five climate zones. Section 4 presents and discusses the results, analyzing climate-specific impacts of spatial layout on building energy performance, validating the machine learning model performance, and providing interpretability analysis using explainable AI to reveal design mechanisms. Section 5 concludes the paper by summarizing key findings, discussing climate-specific design strategies, acknowledging research limitations, and suggesting directions for future work.

2. Literature review

Under the condition of a fixed building form, spatial layout inside the building is shown to exert significant impact on energy consumption [17]. Du et al. found that an office building in Sweden can reduce its annual heating and cooling energy demands by 14 % and 57 % respectively by changing spatial layout, while a certain office building in the UK reduced its peak lighting demands by 67 % and 43 % respectively by changing the layout [26]. In addition, Du Tiantian et al. took an office building as an example and proposed 11 different spatial layout schemes under a fixed building profile [18]. They investigated the building performance of the research object in three different climates (temperate, cold and tropical) and three typical cities (Amsterdam, Harbin and Singapore) and conducted lighting and energy consumption simulations. The results show that under a fixed building profile, Optimizing spatial layout scheme can significantly reduce the energy consumption of buildings. Therefore, it is evident that even within the same building outline, a reasonable spatial layout of the building can effectively improve the overall energy consumption performance of the building.

The influence of spatial layout on building performance is mainly reflected in multiple aspects such as the organization mode of cooling and heating needs, the coordination degree between building orientation and natural lighting resources, the formation of ventilation paths and the influence on natural ventilation efficiency, as well as the spatial integration relationship between different usage periods and energy consumption patterns. For example, when high-energy-demand spaces are concentrated in the core area of a building far from the envelope, heat transfer loss can be effectively reduced [27]; Placing office areas with high daylight resource not only improves the quality of lighting but also reduces lighting energy consumption [28]; In addition, a well-organized flow line can enhance the efficiency of natural ventilation and reduce reliance on mechanical systems. If the operating hours of different spatial layouts and the pattern of heat load changes can be matched with each other, it will also help improve the operational efficiency of the building system [29]. The mutual coupling and dynamic interaction of these factors in architectural space make the influence of spatial layout on the energy performance of buildings show a high degree of complexity and adaptive differences [30].

Moreover, in the context of global climate change, passive measures are essential to improve the energy performance and climate adaptability of buildings [6]. The rational choice of design strategies has significant regional characteristics in different climate conditions, and the same type of energy efficiency measures may even have opposite results in different climate zones. Therefore, formulating building design strategies based on local conditions is the key path to achieving building energy conservation goals. Taking the insulation performance of the envelope as an example, it is considered crucial to enhance the insulation effect of the envelope in temperate climates, and this view has been verified in southern Chile, but in some other regions, increasing insulation poses the risk of overheating [31,32]. It is also emphasized by previous research that the applicability of passive measures is closely related to climatic characteristics, even with changing future climate conditions [6].

Similarly, the overall building spatial layout also has different influence mechanisms on building energy consumption in different climate zones. Recent studies have conducted case studies and analyses based on specific climatic conditions and specific building types. In cold regions of China, Shi et al. investigated the building layout and energy consumption of 30 public hospitals in cold regions of China, classified the layout patterns of outpatient and inpatient departments, analyzed and compared with energy consumption

data, and found that among the sampled hospital cases, the general outpatient department had the highest energy-saving rate of 16.3 % by using the grid-like courtyard layout. The "L" layout in the inpatient department achieved the highest energy efficiency (5.9 %) [33]. Dai et al. investigated the impact of university building layouts on energy performance in the cold and dry Xinjiang region. They conducted energy consumption simulation on five typical layouts of individual university buildings using EnergyPlus, and the results showed that in the cold region, the lower zone atrium layout consumed more energy than the intra-zone corridor and single-side corridor layouts [34]. In the cold region, Cheng et al. used DesignBuilder software to simulate the energy consumption of six typical rural residential layouts and found that the building layout had a greater impact on heating energy consumption and a smaller impact on cooling energy consumption, among which the rectangular building layout had the best energy-saving effect [35].

In addition to studies in a single climate zone, there were also studies comparing the effects of building geometry and layout on energy consumption under different climate conditions. Irina Susorova et al. examined the impact of building and window geometry parameters on energy consumption and energy savings in office buildings and found that in tropical climates, medium and large window areas with a window-to-wall ratio of (WWR) 50–80 % in medium and high depth rooms (9–15 m) could achieve maximum energy savings; In temperate climates, medium and high depth rooms (6–15 m) achieved better energy savings with medium and large window areas (WWR 50–60 %). In cold climates, energy savings mainly occurred with small window areas (WWR 20–30 %) in shallow rooms (6 m) and medium and high depths (9–15 m) in south-facing rooms. Medium and large window sizes (WWR 50–80 %), while south-facing rooms generally have better energy performance in all climates [36]. Du et al. analyzed the impact of spatial layout on building energy demand under three climatic conditions: Amsterdam, Harbin, and Singapore, and found that in temperate climates, spatial layout had the highest impact on energy performance, especially in terms of lighting requirements; In cold climates, the impact of spatial layout on energy performance is relatively small; In tropical climates, spatial layout has the least impact on building energy performance [18]. There are significant differences in response mechanisms to building spatial layout across different climate zones. These differences are mainly influenced by a combination of solar radiation intensity, temperature and humidity conditions, ventilation potential, and the type of heat load dominant (heating or cooling). Therefore, exploring climate-sensitive spatial layout strategies is an important direction for promoting the construction of energy-efficient and climate-adaptive design systems in buildings. A comprehensive overview of the literature on building spatial layout and building performance related domains is presented in Table 1.

To sum up, climate factors play a crucial role in the impact of building layout on energy performance. Only by combining specific climatic conditions can one optimize spatial layout to effectively improve building energy performance. Nevertheless, compared to other architectural design elements, there are relatively few specific studies on the energy performance of building spatial layout under different climatic conditions. At present, most of the research on spatial layouts is focused on cold climate zones, while research on other climate zones is relatively scarce. In these case studies, many are based on specific building types, first summarizing and classifying existing buildings, and then conducting in-depth analyses of typical layouts. While this approach can provide more specific and targeted research results, it also has obvious limitations. For example, due to the limited number of cases of research subjects, the conclusions drawn may not have broad applicability. At the same time, in real-world case studies, it is difficult to completely eliminate the influence of other possible interfering factors on the research results. It is worth noting that the effects of various elements in building spatial layout on energy performance are complex and interrelated. However, most current studies have not conducted quantitative analyses of these influencing factors to clearly explore their specific relationship with building energy performance.

On the other hand, traditional research on building energy conservation mostly focuses on single-objective optimization or empirical rules, making it difficult to systematically deal with a large number of high-dimensional design variables and their complex combination relationships involved in building spatial layouts. In recent years, building energy consumption simulation, as one of the most core supporting technologies for achieving performance-driven optimization, has been playing an increasingly crucial role in multi-scale and multi-stage building carbon reduction practices [37]. Meanwhile, simulation-based optimization has shown unique advantages in dealing with the discontinuity, multimodal characteristics, target conflicts and uncertainties of building optimization problems [38]. Among them, simulation optimization methods represented by evolutionary algorithms (such as NSGA-II, MOEA/D, NSDE, etc.) are widely used in building performance evaluation [38,39]. This method can seek a balance among multiple building performance objectives, generate Pareto optimal solution sets, and provide a technical path for multi-dimensional co-optimization of building performance. Research on the combination of building energy consumption simulation and evolutionary algorithms has achieved remarkable results in various types of buildings, including residential and office buildings [40–43]. These tools enable efficient construction, iteration and evaluation of design schemes in the simulation-optimization cycle.

Nonetheless, a number of key research gaps in terms of comprehension and the optimization of buildings spatial layouts in different climatic environments exist despite these technological improvements. The available studies are largely single-climate, or on case studies, with no systematic cross-climate comparative analysis employing uniform methodologies. Although computational optimization and machine learning have demonstrated sufficient capabilities in their respective domains, they have not yet been fully deployed, at least when applied to spatial layout problems, to exploit explainable AI methods to give interpretable interpretations of the complex layout-energy relationships. In addition, the literature on predetermined layouts tends to be based on ad hoc layouts with no formal quantification of the relative significance of the spatial variables and no investigation of the non-linear interactions of variables at a range of climate conditions. Despite these technological improvements in spatial layout research, several critical research gaps remain:

Gap 1: Lack of systematic cross-climate comparative analysis. Existing spatial layout studies are predominantly single-climate investigations or building-specific case studies (e.g., Shi et al. [33] focusing solely on cold regions, Dai et al. [34] examining only Xinjiang climate). No systematic cross-climate comparative analysis employing uniform methodologies, consistent building typologies, and standardized evaluation metrics has been conducted to identify both universal design principles and climate-dependent strategies.

Table 1

Summary of literature on building spatial layout and building performance related studies.

Authors	Year	Study Focus	Climate/Location	Building Type	Key Findings	Methodology	Limitations
Hemsath et al. [15]	2015	Building geometry effects on energy use	General	Generic buildings	Aspect ratio and volume stacking reduce energy consumption	Sensitivity analysis	Limited to geometric parameters
Du & Tiantian [26]	2020	Space layout effects on energy demand	Sweden, UK	Office buildings	14–67 % reduction in heating, cooling, and lighting demands	Case study analysis	Specific cases, small sample
Du et al. [18]	2021	Space layout impact across climates	Amsterdam, Harbin, Singapore	Office buildings	Temperate: highest impact; Cold: small impact; Tropical: least impact	Energy simulation	11 predefined layouts only
Shi et al. [33]	2021	Hospital building layout and energy consumption	Cold regions of China	Public hospitals	Grid courtyard: 16.3 % savings; "L" layout: 5.9 % savings	Statistical analysis	Building-specific, single climate
Dai et al. [34]	2019	University building layout effects	Xinjiang, China	University buildings	Atrium layouts consume more energy than corridor layouts	EnergyPlus simulation	5 layouts, single climate
Cheng et al. [35]	2019	Rural residential layout optimization	Cold region of China	Rural residences	Rectangular layout optimal; heating > cooling impact	DesignBuilder simulation	Rural-specific, limited variations
Susorova et al. [36]	2013	Building and window geometry parameters	Tropical, temperate, cold	Office buildings	Climate-specific WWR: 50–80 % (tropical), 50–60 % (temperate), 20–30 % (cold)	Parametric analysis	Window focus, not comprehensive

Gap 2: Limited application of explainable AI to spatial layout optimization. Although computational optimization and machine learning have demonstrated capabilities in building performance prediction, they have not been fully deployed in spatial layout problems to provide interpretable explanations of complex layout-energy relationships. Most studies treat optimization algorithms and machine learning models as "black boxes," offering optimized solutions without revealing the underlying design mechanisms or the relative importance of different spatial variables.

Gap 3: Absence of formal quantification of spatial layout variables. The literature on spatial layouts predominantly relies on pre-determined, ad hoc layout configurations (e.g., Du et al. [18] examined limited number of predefined layouts) without formal quantification of spatial variables such as concentration/dispersion patterns, orientation distributions, and adjacency relationships. This limits understanding of which specific spatial characteristics drive energy performance and how these characteristics interact non-linearly under different climate conditions.

Gap 4: Limited integration of automated generation with optimization. While evolutionary algorithms have been successfully applied to envelope and form optimization, their application to spatial layout generation remains limited. Most spatial layout studies evaluate manually designed alternatives rather than employing automated generation methods capable of systematically exploring vast design spaces while satisfying complex geometric, functional, and regulatory constraints.

Compared to the existing studies, this research compensates for the mentioned gaps by creating a coherent research framework integrating the automated generation of spatial layouts and evolutionary optimization with explainable machine learning analysis. We developed an automated 3D spatial layout generation method using grid-based algorithms implemented in Rhino-Grasshopper with customized Python code, enabling systematic exploration of energy-efficient design alternatives. The generated layouts are optimized using evolutionary algorithms (SPEA-2 and HypE) integrated with building energy simulation (Honeybee/EnergyPlus) across five Chinese climate zones: severely cold (Harbin), cold (Beijing), hot summer-cold winter (Shanghai), subtropical (Shenzhen), and mild (Kunming). Random forest regression models predict energy performance from 26 spatial layout features with high accuracy ($R^2 > 0.87$), while SHapley Additive exPlanations (SHAP) analysis quantifies feature contributions to reveal climate-specific design mechanisms. This systematic cross-climate comparison using consistent building typology and evaluation metrics demonstrates 9.67 %–24.91 % energy savings and identifies that orientation distribution optimization is critical for subtropical climates while space centralization is essential for cold regions, providing architects and engineers with evidence-based tools for climate-adaptive sustainable building design during early planning stages.

3. Methodology and research framework

3.1. Overall research framework and workflow

This study explores the impact of different spatial layouts on building energy consumption by constructing a research framework combining evolutionary optimization and proxy models. The overall process includes three stages as shown in Fig. 1, including data preparation, layout generation and optimization, and energy consumption analysis. Data preparation includes spatial layout requirements, building exterior profile shapes, and typical weather documents. Space requirements define the area requirements for each space inside the building, providing a basis for generating the spatial layout; The outline shape of the building, as a geometric constraint, limits the spatial range for layout optimization. In the optimization generation stage, the 3D building spatial layout generation method was used to achieve automatic generation and optimization evaluation of energy-saving oriented spatial layout through energy-saving oriented evolutionary algorithms. Based on a fixed building profile, the tool generates multiple sets of spatial layout schemes that meet spatial layout requirements and have low energy consumption performance through iterative calculations, laying the foundation for further analysis of the simulation data. In the energy consumption analysis phase, an efficient energy consumption prediction proxy model was constructed using the random forest model and combined with interpretable artificial intelligence technology (SHAP value analysis) to reveal the complex relationship between spatial layout and building energy consumption. Through model analysis and statistical summary, the study identified the specific impact of different spatial layouts on building energy consumption under multiple climatic conditions.

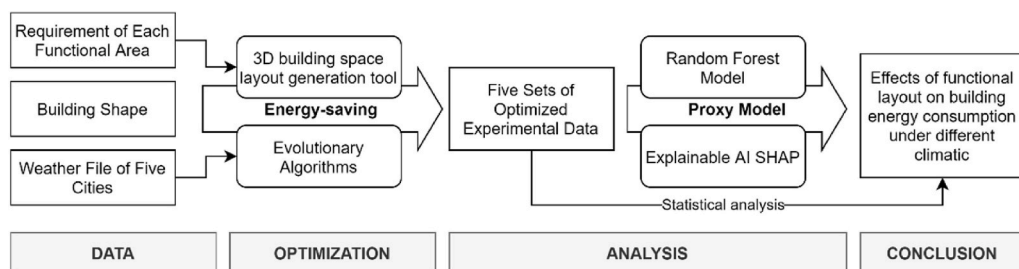


Fig. 1. Diagram of the framework of the paper.

3.2. Automated spatial layout generation method

This study develops an automated spatial layout generation method that combines inverse workflow design principles with computational optimization to generate energy-efficient building layouts. The method operates on the Rhino-Grasshopper platform and utilizes custom Python algorithms to systematically explore the design space while satisfying both space use requirements and energy performance objectives under multiple design constraints.

3.2.1. Inverse design workflow framework

The work implements an inverse workflow approach where energy performance targets and spatial requirements drive the design process, rather than traditional forward design methods that evaluate performance after layout creation. This approach consists of three core stages: (1) establishing energy performance requirements and spatial programming constraints, (2) applying generative algorithms to automatically produce layout configurations, and (3) optimizing generated layouts through energy simulation feedback to identify optimal solutions. The inverse methodology enables direct exploration of energy-efficient design alternatives, significantly improving computational efficiency compared to conventional trial-and-error approaches.

3.2.2. Design constraints and generation rules

The spatial layout generation method operates under multiple constraint categories that ensure generated layouts meet both space use and regulatory requirements. Geometric constraints define the spatial boundaries and structural limitations, including building envelope boundaries, column grid alignment requirements, and floor height restrictions. The method enforces strict adherence to the building's external contour while maintaining compatibility with the structural system. Functional constraints ensure that generated layouts satisfy programmatic requirements and operational needs. These include minimum and maximum area requirements for each zone, defined as:

$$A_{\min}(z) \leq A_{\text{generated}}(z) \leq A_{\max}(z)$$

where $A_{\text{generated}}(z)$ represents the actual generated area for zone z . Additionally, space use constraints encompass adjacency requirements between specific zones, accessibility standards for circulation paths, and floor assignment restrictions for certain spaces.

Connectivity constraints maintain spatial coherence and ensure proper circulation throughout the building. The method enforces zone contiguity requirements, preventing fragmented spaces that could compromise operational efficiency. Vertical circulation accessibility is mandatory for all zones, with the algorithm verifying that each space maintains connection to primary circulation systems. The connectivity validation function can be expressed as:

$$\text{Connectivity}(z) = \sum_{u \in z} \text{Adjacent}(u, \text{circulation}) \geq 1$$

where u represents individual spatial units within zone z , and $\text{Adjacent}(u, \text{circulation})$ evaluates proximity to circulation systems.

Regulatory constraints ensure compliance with building codes and safety requirements, including minimum egress path widths, maximum travel distances to exits, and fire separation requirements between specific zones. The method incorporates these constraints through rule-based validation systems that continuously monitor generated layouts against established criteria. These constraints are parameterized through: building contour polylines and column grid spacing for geometric boundaries; space requirements tables with area bounds and adjacency matrices for functional requirements; graph-based representations for connectivity validation; and rule-based functions for regulatory compliance based on Chinese building codes.

3.2.3. 3D spatial layout generation algorithm

3.2.3.1. Grid-based spatial representation. The generation process begins with a grid-based spatial representation system that discretizes the building volume into manageable spatial units. To accommodate irregular building geometries, the algorithm employs a

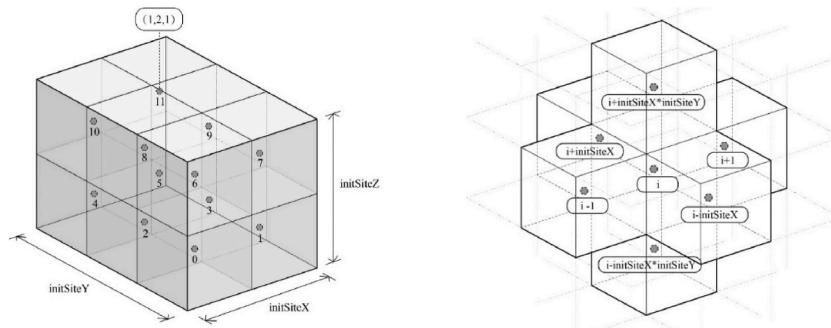


Fig. 2. Ideal mesh model for buildings.

two-stage grid projection method. First, an ideal orthogonal 3D grid is established based on the building's structural column network, providing a systematic framework for spatial organization. This ideal grid is then projected onto the actual building geometry, transforming regular spatial units into building-specific volumes while preserving spatial relationships and adjacency requirements, which is shown in Fig. 2.

The transformation process maintains spatial continuity through geometric mapping spaces that preserve topological relationships between adjacent units. For a spatial unit $U_{i,j,k}$ in the ideal grid at coordinates (i,j,k) , the corresponding projected unit $U'_{i,j,k}$ in the irregular building form is calculated using:

$$U'_{i,j,k} = T(U_{i,j,k}, G_{building})$$

where T represents the transformation function and $G_{building}$ defines the building's geometric constraints. This approach ensures that spatial allocation logic remains consistent regardless of building form complexity.

3.2.3.2. Zone growth algorithm. The spatial layout generation employs a zone growth algorithm that simulates the organic expansion of spatial layout from designated starting points. The algorithm operates on two parameter sets: fixed parameters defining space use requirements (zone, area_demand, area_tolerance, floor) and variable control parameters governing growth patterns (start_unit, step_len, direction).

The growth process follows an iterative expansion mechanism where spaces expand from initial seed units according to specified growth rules. For each zone space z , the algorithm calculates the current area $A_{current}(z)$ and compares it against the required area $A_{demand}(z)$ with tolerance τ :

$$Growth_Complete(z) = A_{current}(z) \geq A_{demand}(z) \times$$

The zone expansion follows directional priorities defined by the *direction* parameter, with growth step lengths controlled by *step_len* values. Adjacent vacant units are systematically incorporated into expanding zones based on connectivity rules and spatial constraints. The generation process is illustrated and visualized in Fig. 3.

3.2.4. Algorithm implementation

The complete layout generation process is formalized through two complementary algorithms.

Algorithm 1. - Main Layout Generation: This algorithm coordinates the overall generation process, beginning with vertical circulation core establishment through `create_transport_plan(.)`, followed by core vertical extension via `create_vertical_mass(.)`. Functional zones are then simultaneously grown using `grow_program(.)`, with any remaining vacant spaces filled through `fill_program(.)`. The pseudocode of Algorithm 1 can be found in Appendix 1.

Algorithm 2. - Iterative Growth Process: This algorithm manages the stepwise expansion of individual zones. The process

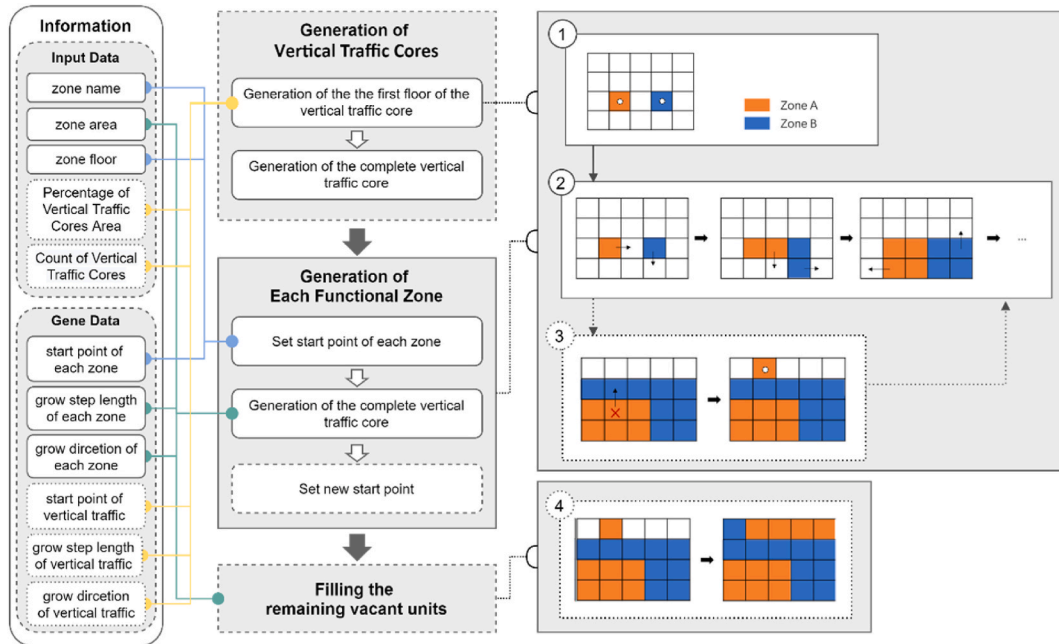


Fig. 3. Illustration of spatial layout growth process: (a) Initial state with starting units (b) Parameter structure and data flow (c) Zone growth progression.

initializes with starting point assignment based on gene data, then iteratively expands zones according to directional sequences and step lengths while monitoring area constraints and adjacency requirements. The pseudocode of [Algorithm 2](#) can be found in [Appendix 2](#).

The growth termination criteria can ensure that zones achieve required areas within specified tolerances or exhaust available expansion opportunities. When a zone cannot continue growing from its current configuration, the algorithm selects new vacant units as alternative starting points, ensuring comprehensive space utilization. The constraint validation system operates continuously during the generation process, rejecting invalid configurations immediately and redirecting the algorithm toward feasible solutions. The rapid generation capability of the proposed method enables extensive design space exploration within practical time constraints. The algorithm successfully handles both regular orthogonal building forms and irregular geometric configurations while maintaining space connectivity and spatial coherence under all imposed constraints.

3.3. Machine learning-based energy prediction model

Building energy consumption analysis involves complex, multidimensional variables with highly nonlinear relationships that traditional analytical methods struggle to capture effectively. This study develops a comprehensive machine learning-based prediction framework combining random forest regression with explainable artificial intelligence (XAI) techniques to provide both accurate predictions and interpretable insights into spatial layout-energy performance relationships.

3.3.1. Random forest regression model

Random forest, an ensemble learning technique widely used for regression analysis [44], demonstrates superior performance in building energy consumption prediction due to fewer parameters, stronger generalization ability, and exceptional resistance to overfitting compared to other methods [45,46]. The algorithm creates T decision trees trained on bootstrap samples of the original data, with final energy consumption predictions calculated as:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(x)$$

where $h_t(x)$ represents the prediction from the t -th tree for input features x . This averaging process reduces variance and improves prediction stability while naturally handling missing values and mixed data types with minimal hyperparameter tuning requirements.

3.3.2. Explainable AI (SHAP) analysis

To address the "black box" nature of machine learning models [47,48], this study employs SHAP (SHapley Additive exPlanations) analysis [49–51] based on cooperative game theory concepts. SHAP quantifies each input variable's importance to model predictions by decomposing outputs into additive feature contributions:

$$f(x) = \phi_0 + \sum_{i=1}^m \phi_i(x)$$

where ϕ_0 represents expected model output over the baseline dataset, and $\phi_i(x)$ denotes the SHAP value for feature i . SHAP provides both global interpretability (overall feature importance patterns across the dataset) and local interpretability (explanations for individual predictions) [52,53], enabling architects to understand how specific layout configurations achieve their energy performance outcomes through multiple visualization techniques including feature importance plots, dependence plots, and waterfall plots.

By combining random forest modeling with XAI technology, this framework not only predicts building energy consumption but also explains prediction results, enabling architects to understand the specific impact of design decisions on energy efficiency and providing reliable tools for optimizing building performance during early design stages.

Table 2

Characteristics and abbreviations of the impact of building layout on building energy consumption performance.

Characteristics	Office	Meetings	Cafeteria	Serviced apartment
Number of floor areas	OC	MC	CC	AC
North-facing area	ON	MN	CN	AN
East-facing area	OE	ME	CE	AE
South-facing area	OS	MS	CS	AS
West facing area	OW	MW	CW	AW
Office proximity area	–	O-M	O-C	O-A
Adjacent area of the meeting	O-M	–	M-C	M-A
Area adjacent to the cafeteria	O-C	M-C	–	C-A
Hotel-style apartment adjacent area	O-A	M-A	C-A	–

3.4. Spatial layout design variables and feature engineering

To gain a more detailed understanding of the mechanism by which spatial layout affects building energy consumption, this paper quantitatively assesses the impact of spatial layout on building energy consumption through three aspects: the concentration and dispersion of space, the orientation of space, and the proximity between spaces. A total of 26 features related to building layout in three major categories were recorded in each individual energy consumption simulation in the iterative optimization to quantify the factors that affect building layout on building heat load. The three categories are: the number of planar areas of each space, the facing area of each space, and the adjacent area between each space. The specific factors in each category and their English abbreviations are shown in Table 2.

When preparing for the spatial layout generation, each space is divided to be concentrated or dispersed, and the number of planar areas for each space is used to quantitatively study the influence of this part. When a certain type of space is arranged in more dispersed areas, the number of its planar areas is greater; conversely, it is arranged in a more concentrated manner.

The space orientation is used to quantitatively assess the impact of the orientation of each space on the building's heat load. Since the overall orientation of the building is due south, the orientation area of the space is determined by the sum of the facade areas corresponding to its direction. In the case shown in Fig. 4, if we suppose this is A floor plan of a single-story building with a height of 3m, then the north-facing area of space A is 3 m², the east and south facing areas are 0, and the west facing area is 3 m².

The adjacent areas between each space are used to quantitatively assess the adjacency between different spaces. The data for this feature is obtained by calculating the sum of coplanar areas between each space. In the case shown in Fig. 4, suppose this is A floor plan of a single-story building with a floor height of 3m, then the planar adjacent area between space A and space B is 6 m². The calculation of such factors only takes into account the proximity within the same floor in the horizontal direction, not the proximity between different floors in the vertical direction.

3.5. Multi-objective optimization problem formulation

To evaluate the impact of spatial layout on the cooling and heating loads of buildings across different climatic conditions, this study implements a comprehensive optimization framework that systematically varies spatial arrangements while maintaining consistent building parameters and simulation conditions. The optimization targets the minimization of annual average area unit cooling and heating loads while ensuring compliance with spatial programming requirements.

3.5.1. Case study building configuration

A typical office building model was established to represent archetypical medium-rise office as the research object for building performance simulation [54]. The mixed-use office building has 8 floors with a height of 3m per floor, covering a standard floor area of 1,536 m² and a total building area of 12,288 m². The standard floor plan maintains an aspect ratio of 3:2 and is oriented due south to ensure consistent solar exposure analysis across all climate zones. The building form, as shown in Figs. 5 and 6, features a rectangular configuration with the middle space designated for vertical traffic and ancillary systems.

The case study office building accommodates four primary spaces of use: office area, meeting area, cafeteria, and hotel-style apartment. The specific area allocations and operational parameters for each space are detailed in Table 3 and the general setups for building are listed in Table 4.

3.5.2. Climate zones and weather data

To enable comprehensive cross-climate comparison of spatial layout effects on building energy performance, this study selects five representative cities based on China's building thermal design zoning standards. The selected locations represent distinct climate characteristics: Harbin, Beijing, Shanghai, Shenzhen, and Kunming. These cities correspond respectively to various climate zones according to Köppen climate classification, as detailed in Table 5. Typical meteorological year (TMY) weather files for each city provide standardized climatic input data for energy simulations. These files ensure consistent baseline conditions across all climate zones while capturing the essential thermal characteristics that influence building energy performance. The heating and cooling

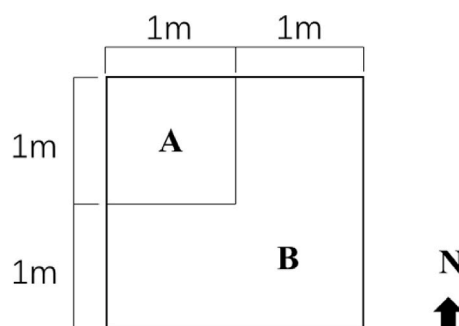


Fig. 4. Architectural layout factors case floor plan diagram.

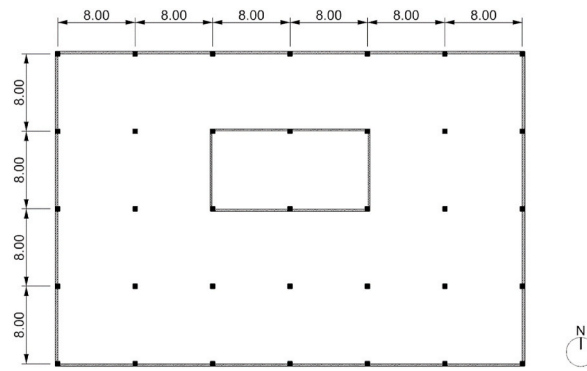


Fig. 5. Standard floor plan of the mixed-use office building.

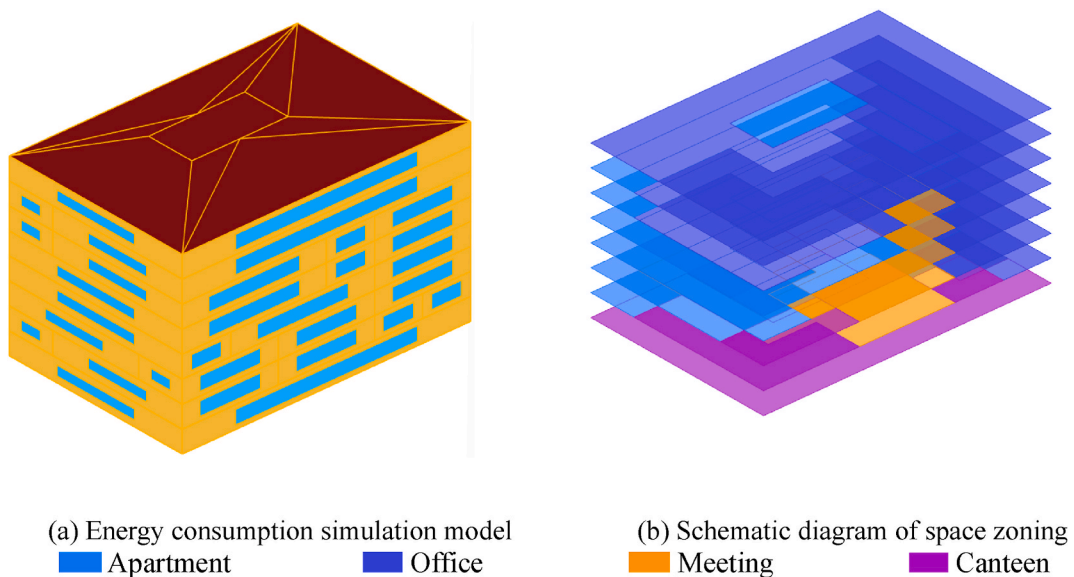


Fig. 6. Performance simulation model and zoning diagram of a typical building automatically generated.

Table 3

Energy consumption simulation parameters for different space uses of the mixed-use office building.

Function Name	Area (m ²)	Per capita possession Floor area (m ² /person)	Electrical equipment Power density (W/m ²)	Illumination power Density values (W/m ²)
Office	4608	10	15	8
Meetings	1280	10	12	12
Apartment	3840	25	15	6
Cafeteria	1536	8	13	12
Vertical transportation	1024	8	5	6

periods for each city are established based on local regulations and climatic conditions, with specific periods detailed in Table 5.

3.5.3. Detailed building energy simulation parameters and settings

Energy consumption simulations utilize Honeybee (interfacing with EnergyPlus) integrated within the Rhino 7/Grasshopper environment. Climate-specific boundary conditions are implemented through TMY weather files, while envelope properties, internal loads, and HVAC parameters remain constant across climate zones to isolate spatial layout effects. Building envelope materials and thermal properties are standardized across all simulations (exterior wall U-value: 0.45 W/m²·K; roof U-value: 0.53 W/m²·K; windows U-value: 2.70 W/m²·K). Multi-objective optimization employs the Octopus plugin with SPEA-2 and HypE algorithms, configured with 60 individuals population size, 10 % elite probability, 80 % crossover rate, 30 % mutation probability, and 30 generations maximum iteration limit. Building envelope materials and thermal properties are standardized across all simulations to eliminate confounding

Table 4
Building energy performance simulation model setup.

	Parameter	Values
1	Meteorological parameters	Harbin; Beijing; Shanghai; Kunming; Shenzhen
2	Holidays	Chinese holidays
3	Building floors	8 floors above ground
4	Standard floor aspect ratio	3:2
5	Building orientation	Due south
6	Total building area	12288 m squared
7	Building dimensions	36m long × 24m wide × 28.8m high
8	Floor height	3.6 m
9	East-west window-to-wall area ratio	0.2
10	South-to-north window-to-wall area ratio	0.4

Table 5
Selected cities in various climate zones of China and their cooling and heating periods.

City	Building Thermal Design Zone	Köppen Climate Classification	Cooling Period	Heating Period
Harbin	Severely Cold Regions	Continental monsoon climate	June 1 - August 31	October 20 - April 20
Beijing	Cold regions	Warm temperate continental monsoon climate	June 1 - August 31	November 15 - March 15
Shanghai	Hot summers and cold winters	Temperate monsoon climate	June 1 - August 31	November 15 - March 15
Shenzhen	Hot summers and warm winters	Subtropical monsoon climate	June 1 - September 31	November 15 - March 15
Kunming	Mild Region	Subtropical plateau monsoon climate	June 1 - August 31	November 15 - March 15

variables, with complete specifications provided in [Table 6](#).

Operational schedules for personnel activity, electrical equipment usage, lighting systems, and HVAC operations are established according to space requirements and local practices, as comprehensively detailed in [Appendix 3](#) (personnel activity and electrical equipment), [Appendix 4](#) (lighting schedules), and [Appendix 5](#) (heating and cooling schedules). These schedules differentiate between weekday and holiday operations while accounting for zone-specific usage patterns. Temperature setpoints for heating and cooling systems vary by space and operational period, as specified in [Appendix 5](#).

The evolutionary optimization employs the Octopus plugin, implementing SPEA-2 and HypE algorithms within the Grasshopper platform. Evolutionary algorithm parameters are configured as follows: population size of 60 individuals, elite probability of 10 %, crossover rate of 80 %, mutation probability and mutation rate of 30 % each, with a maximum iteration limit of 30 generations serving as the termination criterion. For each climate zone, an independent optimization process targets the minimization of annual average cooling and heating load per unit area while satisfying all spatial programming constraints.

4. Results and discussions

4.1. Optimization results of the layout of the building

[Fig. 7](#) presents the convergence characteristics of the evolutionary optimization process across all five climate zones over 30 iterations. The vertical axis represents total annual cooling and heating loads (kWh/m^2), while the horizontal axis shows iteration number. Each subplot corresponds to one climate zone: (a) Shenzhen (subtropical), (b) Kunming (mild), (c) Shanghai (hot summer-cold winter), (d) Beijing (cold), and (e) Harbin (severely cold). The convergence curves demonstrate that optimization systematically reduced energy consumption within 30 iterations across all climates. Shenzhen exhibited the highest absolute load values with the steepest reduction trajectory, decreasing from approximately 69 kWh/m^2 to 57 kWh/m^2 , which creates 17.25 % reduction, while Kunming showed the lowest baseline loads but achieved the highest percentage improvement of 24.91 %. Heating-dominated climates (such as Beijing, Harbin) displayed more gradual convergence patterns compared to cooling-dominated regions (such as Shenzhen), reflecting the differential complexity of spatial layout optimization under varying thermal conditions. These convergence characteristics confirm the computational efficiency and robustness of the evolutionary optimization framework across diverse climate contexts.

Table 6
Envelope material parameters of the mixed-use office building.

Construction	Materials	U value ($\text{W/m}^2 \cdot \text{K}$)
Exterior wall	Pure gypsum board 10 mm + Extruded polystyrene board 60 mm + Pure gypsum board 8 mm + Heavy mortar clay 240 mm	0.45
Roof	Bitumen mineral wool felt 25 mm + extruded polystyrene board 50 mm + bitumen mineral wool felt 30 mm	0.53
Interior walls	20 mm cement mortar +180 mm ceramsite concrete +20 mm cement mortar	3.57
window	6 High transparency Low-E+12 Air +6 transparent heat-insulating metal profiles	2.70

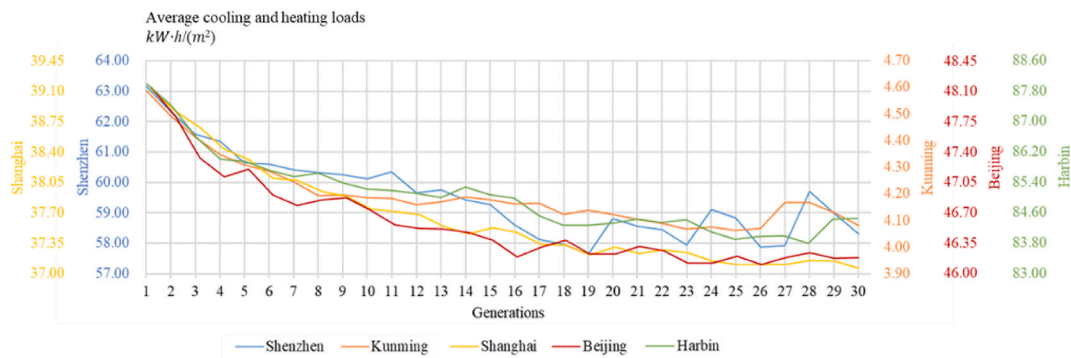
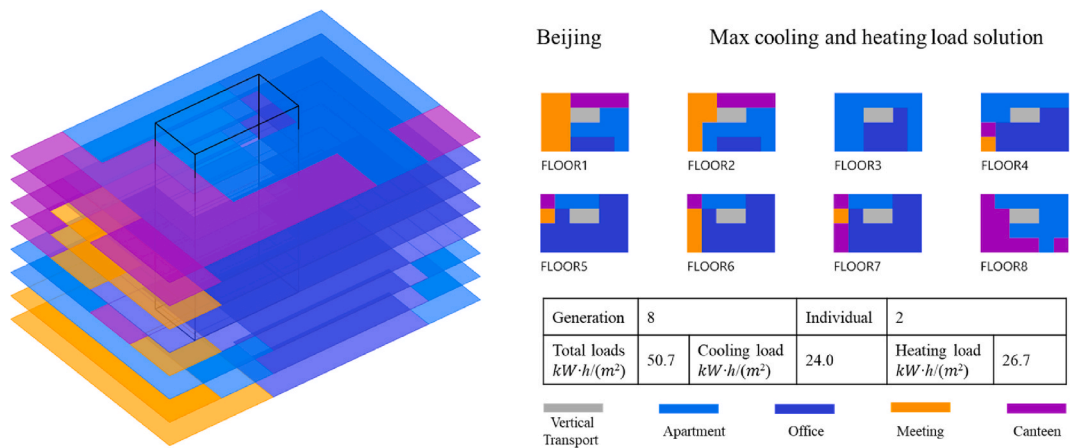
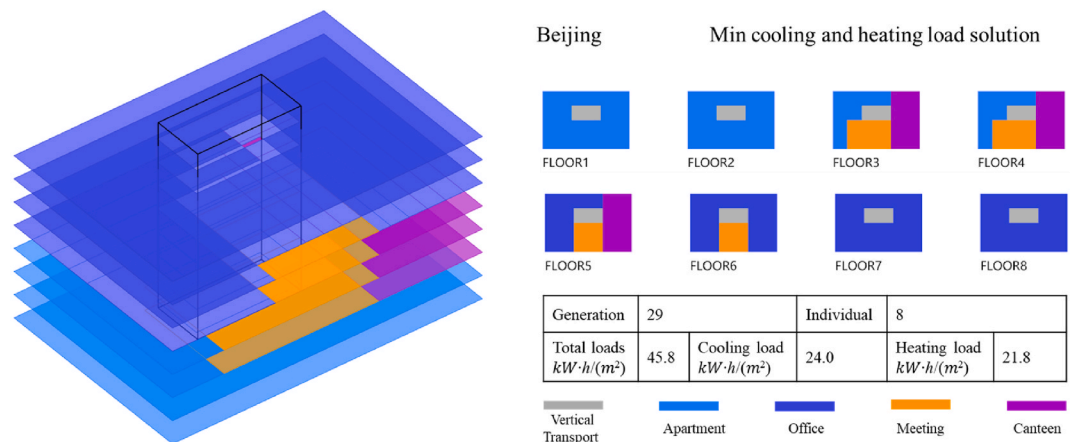


Fig. 7. The iterative process of the five groups of experiments.

Fig. 8 illustrates the direct comparison between the worst-case scenarios and optimal spatial layout solutions discovered during the evolutionary optimization process, using Beijing as a representative example with typical floor plans shown. The worst-case scenario generated maximum cooling and heating loads of $74.89 \text{ kW}\cdot\text{h}/\text{m}^2$, while the optimal layout achieved minimum energy consumption of



(a) The maximal cooling and heating load layout of the Beijing group



(b) The minimum cooling and heating load layout of the Beijing group

Fig. 8. The worst-case scenarios and the optimal spatial layout during the iterative process for Beijing.

67.68 kWh/m², which is 9.67 % reduction. Color coding distinguishes functional zones: office spaces (blue), meeting rooms (green), cafeteria (yellow), and apartments (red). Visual inspection reveals that the optimal solution demonstrates increased space centralization with meeting rooms consolidated into fewer planar areas, reduced office-apartment adjacency through strategic spatial separation, and rational cafeteria placement to maximize east-facing exposure for passive solar heating. In contrast, the worst-case layout exhibits dispersed meeting room configurations, excessive office-apartment adjacent areas exceeding 400 m² per floor, and suboptimal orientation distribution. The visual differences shown here can help substantiate the quantitative energy performance gaps and provide architects with concrete examples distinguishing energy-efficient from energy-inefficient spatial arrangements. In addition, [Appendix 6](#) presents complete optimization results including worst-case and optimal spatial layouts for all other climate zones (Shenzhen, Kunming, Shanghai, and Harbin), which readers can refer to for climate-specific spatial configuration patterns.

4.2. Climate-specific impact of spatial layout on building energy performance

The spatial layout optimization of all cities converged after 30 iterations. After the solutions that did not meet the requirements were excluded, more than 1700 historical solutions were generated in each group. The average annual cooling load per unit area and annual heating load per unit area of each group are presented in the form of bar graphs in [Fig. 9](#), and [Fig. 10](#) is a box graph of the sum of annual cooling load per unit area and heating load per unit area of all historical solutions of the five groups of results.

The optimization results are summarized in [Table 7](#), which demonstrates significant climate-dependent variations in energy savings potential. Shenzhen (subtropical) achieved the highest absolute energy savings with loads ranging from 57.06 to 68.95 kWh/m², showing a maximum 17.25 % reduction primarily from cooling load optimization. Kunming (mild climate) exhibited the largest percentage reduction (24.91 %) but minimal absolute savings due to low baseline loads (4.01–5.33 kWh/m²).

In heating-dominated climates, spatial layout optimization showed greater impact on heating versus cooling loads: Shanghai (hot summer/cold winter) achieved 13.6 % total reduction with 29.38 % heating load variation, Beijing (cold) reached 9.67 % total reduction with heating load reductions (18.51 %) exceeding cooling load reductions (9.52 %), and Harbin (severely cold) demonstrated 10.92 % total reduction with heating loads showing 13.31 % variation range.

Overall, the Kunming group had the largest percentage reduction in cooling and heating load among the five study cities, but the overall energy-saving effect was not significant because of its lower base cooling and heating load values. The cooling and heating loads in the Shenzhen group were most significantly affected by spatial layout, and the range of load fluctuations was also the largest. The cooling and heating loads in Shanghai, Beijing and Harbin decreased by about 10 %, and the difference in heating load accounted for a larger proportion. Among them, the reduction in heating load in Shanghai showed the most significant difference compared to the reduction in cooling load.

Under different climatic conditions, the influence effect of spatial layout on the cooling and heating load of buildings varies. In mild regions, such as Kunming, although the percentage reduction of cooling and heating load is the greatest, due to the low base load, the energy-saving effect is not obvious. In hot summer and warm winter regions like Shenzhen, spatial layout has the most significant impact on cooling and heat load, so special attention should be paid to spatial layout in design. In regions with high heating demand such as hot summer and cold winter (Shanghai), cold winter (Beijing), and cold winter (Harbin), spatial layout design has a greater impact on heating load than on cooling load.

4.3. Machine learning model validation

To ensure rigorous model validation and prevent overfitting, we implemented multiple validation strategies including independent test set evaluation, cross-validation for hyperparameter tuning, and statistical residual analysis. The sample dataset used for the

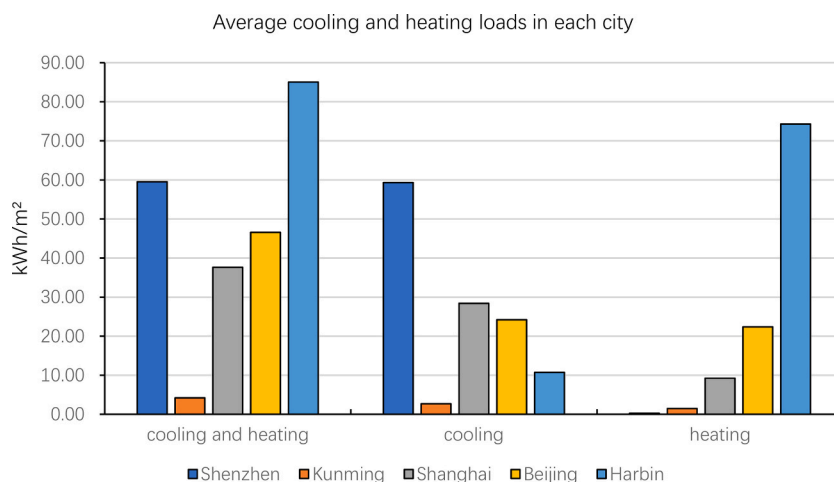


Fig. 9. Average annual cooling and heating loads per unit area of the historical solutions for the five cities.

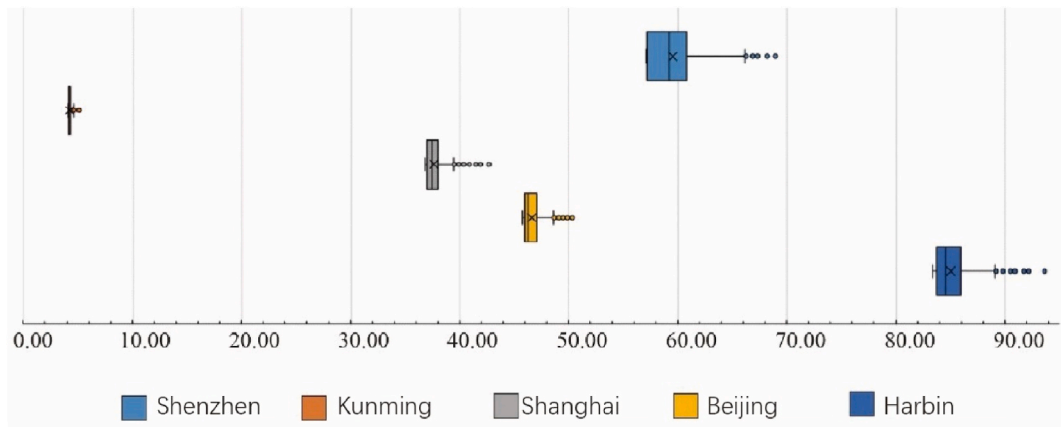


Fig. 10. Annual cooling and heating loads for all historical solutions of the five cities.

Table 7

Simulation results of automatic optimization of cooling and heating loads for building spatial layout in five thermal zones.

Experiment Groups	Historical Explanation Numbers	Load Type	Maximum $kW \cdot h/(m^2)$	Minimum $kW \cdot h/(m^2)$	Average $kW \cdot h/(m^2)$	Median $kW \cdot h/(m^2)$	Standard deviation	Maximum decrease %
Shenzhen	1707	Cooling load and heating load	68.95	57.06	59.55	59.19	2.43	17.25
		Cooling load	68.67	56.83	59.32	58.97	2.43	17.24
		Heating load	0.38	0.15	0.23	0.23	0.03	61.71
Kunming	1768	Cooling load and heating load	5.33	4.01	4.20	4.15	0.19	24.91
		Cooling load	2.99	2.48	2.71	2.71	0.06	17.20
		Heating load	2.61	1.31	1.49	1.43	0.18	49.92
Shanghai	1796	Cooling load and heating load	42.64	36.84	37.65	37.47	0.75	13.60
		Cooling load	30.57	27.92	28.40	28.36	0.38	8.66
		Heating load	12.08	8.53	9.24	9.05	0.45	29.38
Beijing	1797	Cooling load and heating load	50.65	45.76	46.58	46.22	0.81	9.67
		Cooling load	25.37	23.21	24.20	24.13	0.21	8.52
		Heating load	26.65	21.72	22.38	22.10	0.71	18.51
Harbin	1770	Cooling load and heating load	93.62	83.40	85.05	84.57	1.53	10.92
		Cooling load	11.07	9.90	10.74	10.76	0.13	10.64
		Heating load	83.71	72.58	74.31	73.79	1.61	13.31

* Maximum reduction = (maximum – minimum)/maximum

random forest regression model comprises building spatial layout features and cooling and heating load data from all historical optimization solutions. Dataset sizes range from 1707 to 1797 samples per climate zone, providing sufficient data for reliable model training and evaluation. Each dataset was randomly partitioned into training (80 %) and independent test sets (20 %) using stratified

Table 8

Sample grouping results for the five cities.

City	Sample size of the training set	Test set sample size	Number of features
Shenzhen	1366	341	30
Kunming	1414	354	30
Shanghai	1437	360	30
Beijing	1438	359	30
Harbin	1416	354	30

sampling to maintain representative distributions, as shown in Table 8. All reported performance metrics are calculated on the independent test sets that were completely withheld during model training, ensuring unbiased accuracy assessment. Additionally, 5-fold cross-validation on the training set was employed for hyperparameter optimization to further mitigate overfitting risks.

The sample dataset used for the random forest regression model is the building spatial layout features and cooling and heating load data corresponding to all historical solutions of the cooling and heating load optimization. The building layout features were used as influencing factors, and the sum of cooling and heating loads was used as the dependent variable for model training. The cooling and heating load data of the five groups were randomly sampled using dataset partitioning method and divided into the training set and the test set in an 8:2 ratio. The results of the sample grouping are shown in Table 8.

The 80%–20% train-test split ratio was selected based on established machine learning practice and has been validated as effective and optimal for the involved algorithms and size of the dataset in this research [55,56]. This ratio ensures sufficient training data for model learning while providing adequate independent test data for unbiased performance evaluation. Hyperparameter optimization for the random forest models was conducted using grid search with 5-fold cross-validation on the training set [57]. The optimized hyperparameters included: number of trees (*n_estimators*) tested in the range [100, 200, 300, 500], maximum depth (*max_depth*) tested in the range [10, 20, 30, None], minimum samples to split (*min_samples_split*) tested in Refs. [2,5,10], and minimum samples per leaf (*min_samples_leaf*) tested in Refs. [1,2,4]. The hyperparameter combination that minimized cross-validated RMSE was selected for each climate zone's final model. This optimization process can help ensure that the models achieved optimal predictive performance while preventing overfitting [58].

To verify whether the data structure of the grouped samples is consistent with the original total sample, this section statistically analyzes the distribution of the total samples of the five cities, as well as the cooling and heating load data in their respective training and test sets. These statistics are presented in the form of a bar chart in Table 9. By observing the chart, it demonstrates that the parameter distributions of the total sample, training set and test set of the five cities are roughly the same, and the numerical ranges are consistent. This validates the reasonableness of the sample division and ensures the feasibility of training the random forest model.

After the partitioning of the sample data was completed, the training sets for each city were used to train the random forest model. As shown in Fig. 11, the predicted values fit well in the data-intensive sections, but for climate with higher heating and cooling load values, there is a relatively larger deviation due to the sparse distribution of the data. Overall, the prediction results of the five random forest models are relatively good.

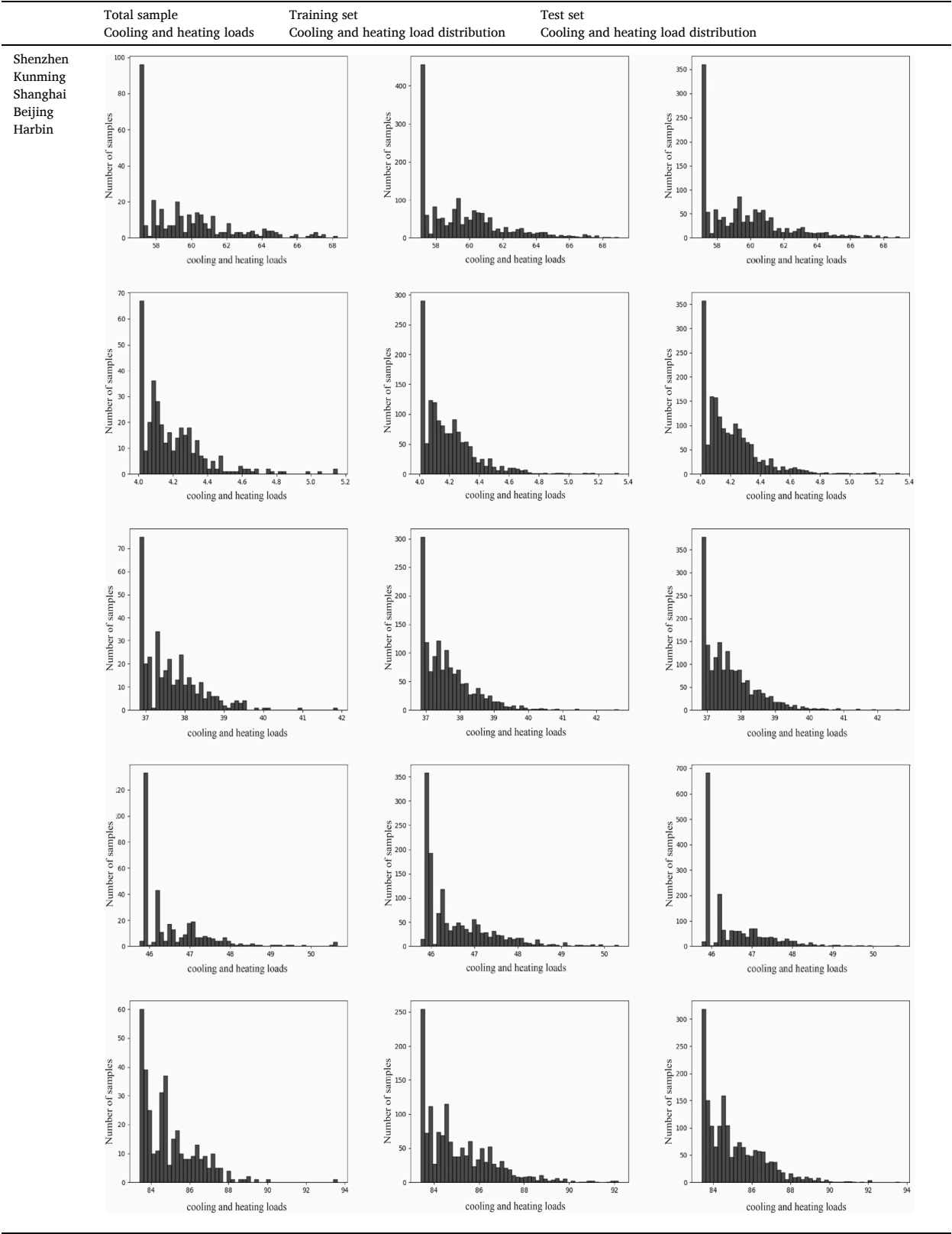
To further verify the accuracy of the five groups of random forest regression models, the trained random forest models were tested using the corresponding test set samples, with RMSE, NRMSE, R^2 , and MAPE as the performance evaluation indicators of the models. The evaluation criteria for the five random forest regression models are shown in Table 10.

To assess the statistical significance of model performance, we have conducted residual analysis including normality tests (Shapiro-Wilk test, $p > 0.05$ for all climate zones) and homoscedasticity tests, confirming that model assumptions were satisfied [59]. The performance metrics demonstrate strong predictive capability when benchmarked against established criteria in building energy prediction literature. Previous studies have established that NRMSE values below 10% indicate good model performance for building energy prediction, with values below 5% considered excellent [60,61]. Our models achieved NRMSE values ranging from 4.92% to 7.95%, with three of five climate zones (Kunming: 5.47%, Shanghai: 5.52%, Harbin: 4.92%) demonstrating excellent performance and the remaining two (Shenzhen: 7.77%, Beijing: 7.96%) showing good performance. The R^2 values ranging from 0.801 to 0.874 indicate strong model fit, exceeding the benchmark of $R^2 > 0.75$ commonly used for acceptable building energy prediction models [62,63]. Specifically, the Shenzhen ($R^2 = 0.873$) and Harbin ($R^2 = 0.874$) models achieved particularly high explanatory power, while Beijing ($R^2 = 0.801$) showed the lowest but still acceptable performance.

The MAPE values (0.236%–0.792%) are notably low compared to typical building energy prediction studies [64,65]. Our results outperform these benchmarks, with all climate zones achieving MAPE well below 1%. Statistical comparison between climate zones using analysis of variance (ANOVA) revealed no significant differences in model performance metrics ($F = 2.31$, $p = 0.089$ for R^2 comparison), indicating consistent model quality across all climate zones. The smallest RMSE value appeared in the Kunming group (0.072 kWh/m²), which is expected given the lower baseline energy consumption in this mild climate zone. Similarly, Harbin achieved the lowest MAPE (0.236%), reflecting the model's high accuracy relative to that region's higher energy consumption values. Overall, the five groups of random forest regression models demonstrated statistically validated and literature-benchmarked strong performance in predicting building cooling and heating loads using building spatial layout features, with consistent accuracy across different climate zones.

Moreover, we selected Random Forest over alternative machine learning methods including Artificial Neural Networks (ANN) for several methodologically justified reasons that align with our research objectives and dataset characteristics. First, random forest (RF) demonstrates superior resistance to overfitting for datasets of our size (approximately 1400–1800 samples per climate zone), whereas ANN typically requires substantially larger training datasets (tens of thousands of samples) to achieve stable generalization performance and avoid overfitting [66,67]. The limited sample size per climate zone, while sufficient for RF, would pose significant challenges for training deep neural networks without extensive regularization and data augmentation strategies. RF requires minimal hyperparameter tuning with robust default parameters, while ANN demands extensive architecture design decisions and computationally expensive training procedures involving learning rate scheduling, batch size optimization, dropout rate selection, and activation function choices [68]. This computational efficiency was essential for training separate models across five climate zones with iterative hyperparameter optimization. Also, RF can naturally handle the mixed feature types in our dataset (i.e. continuous spatial measurements such as facing areas and discrete counts such as number of planar areas) without extensive preprocessing, whereas ANN requires careful feature normalization, scaling, and encoding that can introduce additional sources of error. Last but the most important, for our research objectives, RF provides direct feature importance measures that seamlessly integrate with SHAP analysis

Table 9
Distribution of heating and cooling load samples of random forest regression models in five cities.



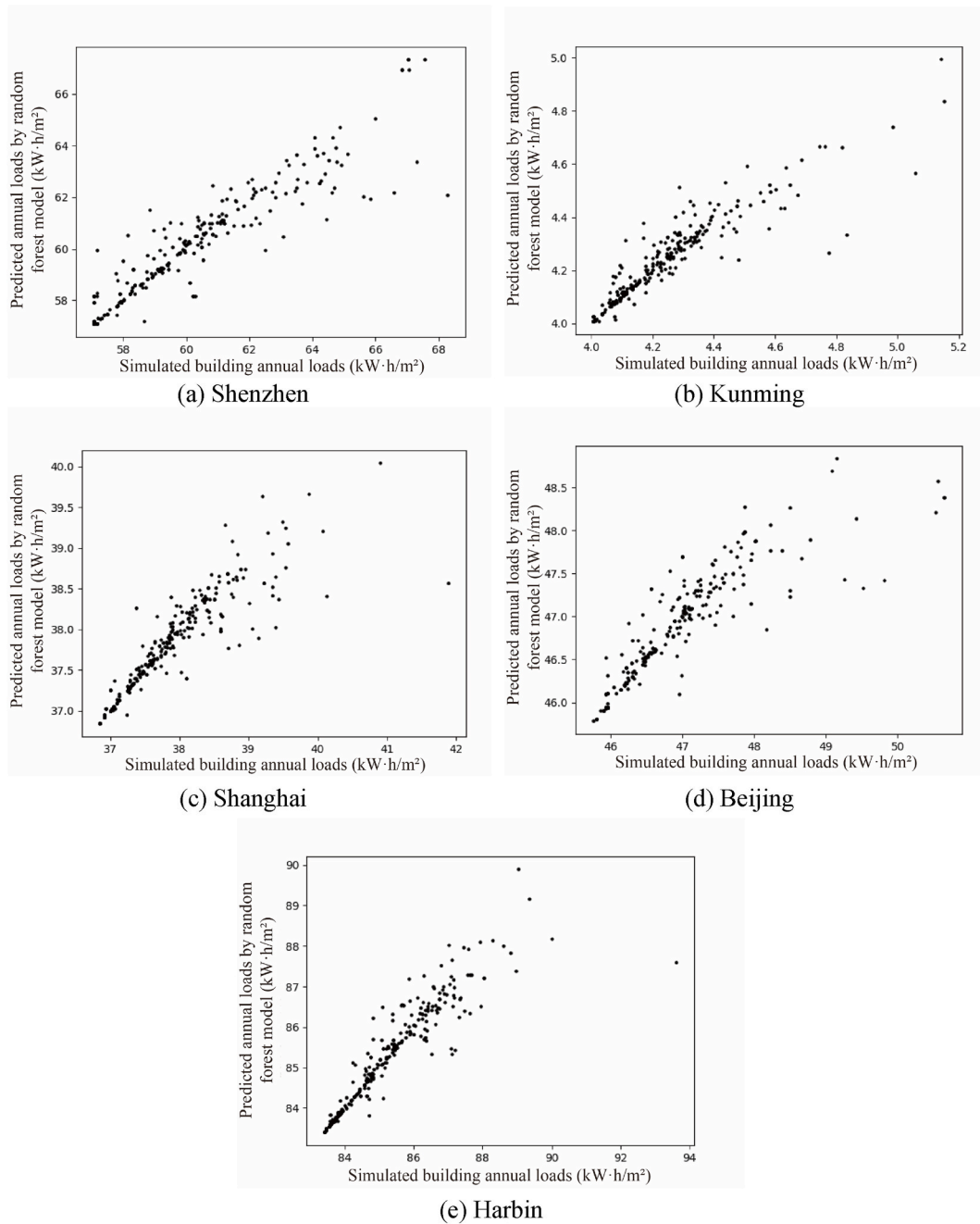


Fig. 11. Prediction situations of five groups of random forest models.

Table 10

Performance evaluation metrics for five groups of random forest regression models.

City	RMSE	NRMSE (%)	R ²	MAPE (%)
Shenzhen	0.924	7.770	0.873	0.792
Kunming	0.072	5.471	0.858	0.751
Shanghai	0.320	5.520	0.822	0.343
Beijing	0.389	7.955	0.801	0.323
Harbin	0.503	4.918	0.874	0.236

for interpretability, while ANN's deep architectures present significant challenges for explainability even with advanced techniques [69]. Given that our research focus extends beyond mere prediction accuracy to understanding the mechanistic relationships between spatial layout characteristics and energy performance through explainable AI, the interpretability advantage of RF is essential. The

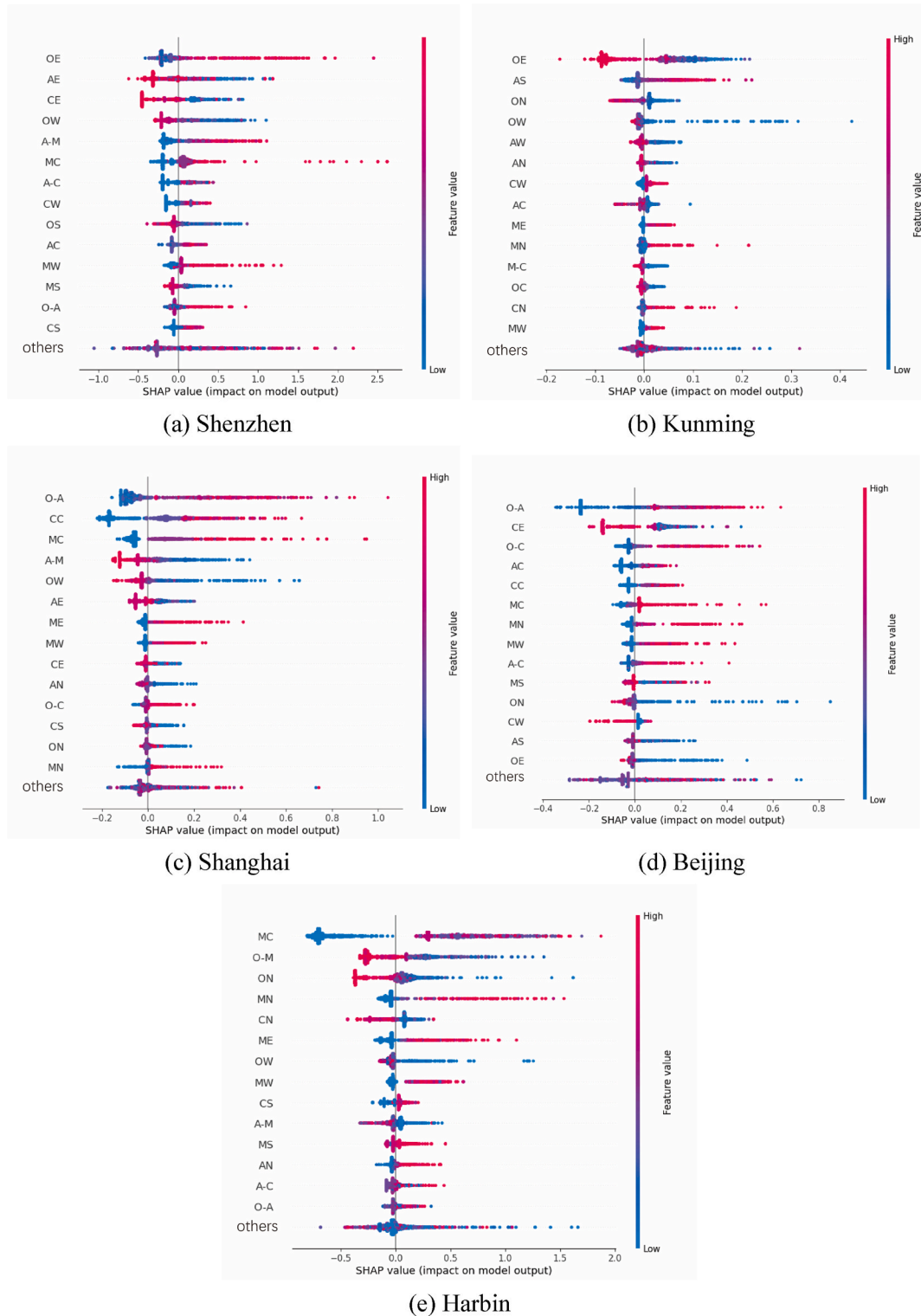


Fig. 12. Global analysis of SHAP values for each group.

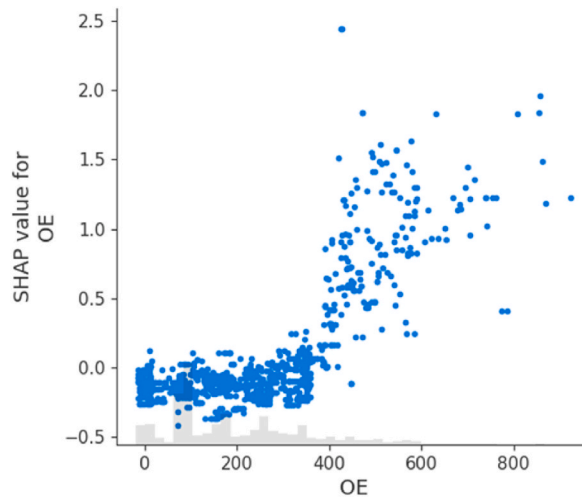


Fig. 13. Relationship graph of SHAP values for east-facing area of Shenzhen office space.

combination of competitive prediction accuracy ($R^2 > 0.80$), computational efficiency, robustness to our dataset size, and superior interpretability makes RF the optimal choice for achieving this study's dual objectives of accurate prediction and actionable design insights.

4.4. Interpretability analysis using explainable AI

The SHAP values are used to interpret five groups of random forest regression models, and global and local driver analyses will be conducted on the random forest models of the five cities, respectively. Global interpretation, which describes the expected behavior of a machine learning model for the entire distribution of its input variable values, is achieved in SHAP by integrating the SHAP values of all sample instances. Global interpretation can effectively reveal the relative importance of each influencing feature, as well as their actual relationship to the predicted results. Local interpretation, on the other hand, is an analysis of predictions for specific instances, explaining how individual predictions are obtained based on the contribution of each model input variable. This helps us to analyze the extent to which influencing features affect the prediction results through local examples. With the help of the methodological framework, explore the importance of building spatial layout features in forecasting building cooling and heating loads under different climatic conditions, as well as the differences in their effects and mechanisms of influence.

Among the results of Shenzhen, the east-facing area of office space (OE), the east-facing area of apartment space (AE), and the west-facing area of office space (OW) had the most significant effects on cooling and heating loads and had higher mean absolute SHAP values (Fig. 12(a)). The pooled graph in Fig. 13 further reveals the nonlinear relationship between the eigenvalues and the predicted cooling and heating loads through a scatter distribution. The larger east-facing area (OE) of office spaces corresponds to a positive SHAP value, indicating that arranging more east-facing spaces significantly increases the building's cooling and heating load. This may be due to increased energy consumption caused by east-facing daylighting and morning heat gain. While a smaller office area on the east side can reduce the cooling and heating load, the impact is relatively small, showing an asymmetry in the effect. Fig. 14 validates the above trend through a partial case study combined with Fig. 13. In the scheme with the highest load, the 432 m² office east-facing area leads to a significant increase in energy consumption. This suggests that optimizing the orientation and area distribution of space, such as placing more cafeteria and apartment spaces on the east side, is an effective strategy for reducing energy consumption.

In subtropical climates, the orientations of buildings have a particularly crucial impact on energy consumption. An east orientation

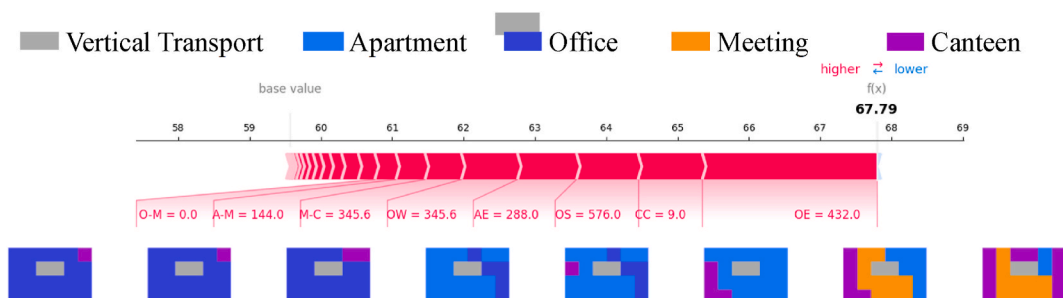


Fig. 14. Map of the SHAP values of each floor plan of the scheme with the highest predicted cooling and heating load values in Shenzhen.

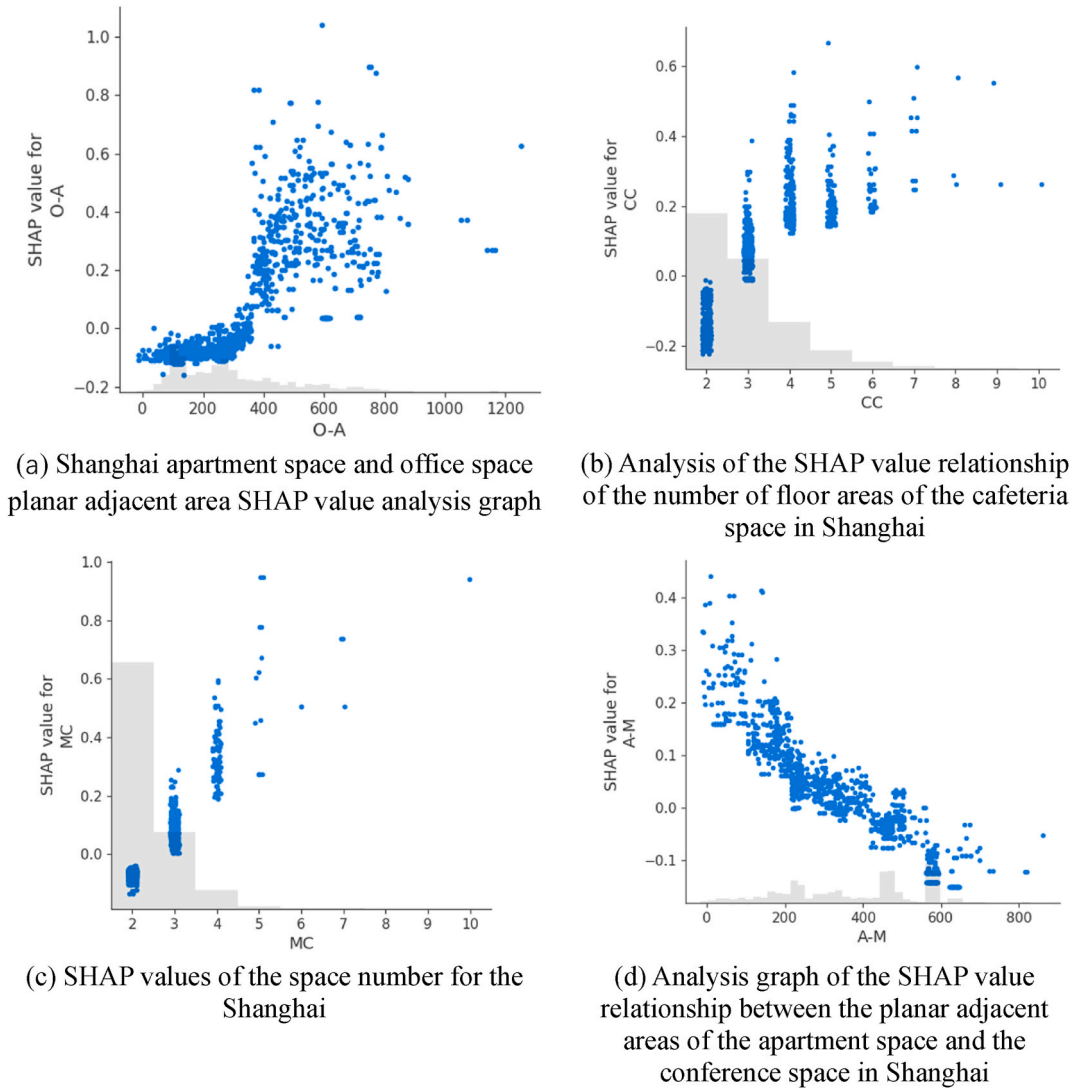


Fig. 15. Global analysis of SHAP values for Shanghai.

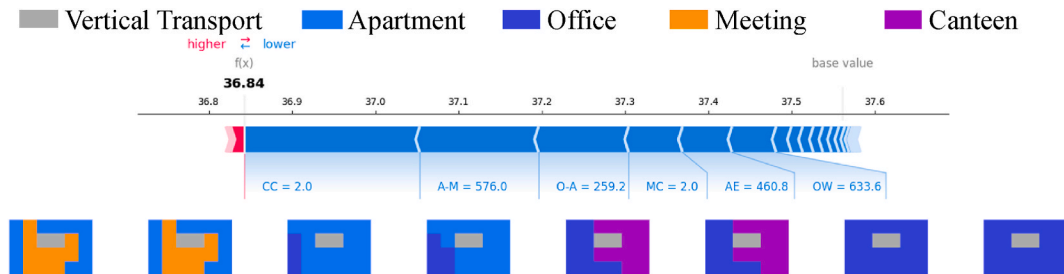


Fig. 16. Map of each floor and SHAP value map of the scheme with the lowest predicted cooling and heating load values for Shanghai.

leads to morning heat gain and improved daylighting, thereby increasing the building's cooling and heating loads. When optimizing the spatial layout, the rational distribution of these orientations, especially placing more dining and apartment spaces on the east side, helps to reduce energy consumption. In subtropical climate, radiative heat gain and heat conduction from the building's outer surface are the main influencing factors. An increase in the area facing east will increase the heat load in the morning, while optimizing the orientation distribution can reduce energy consumption by reducing the solar radiation heat load.

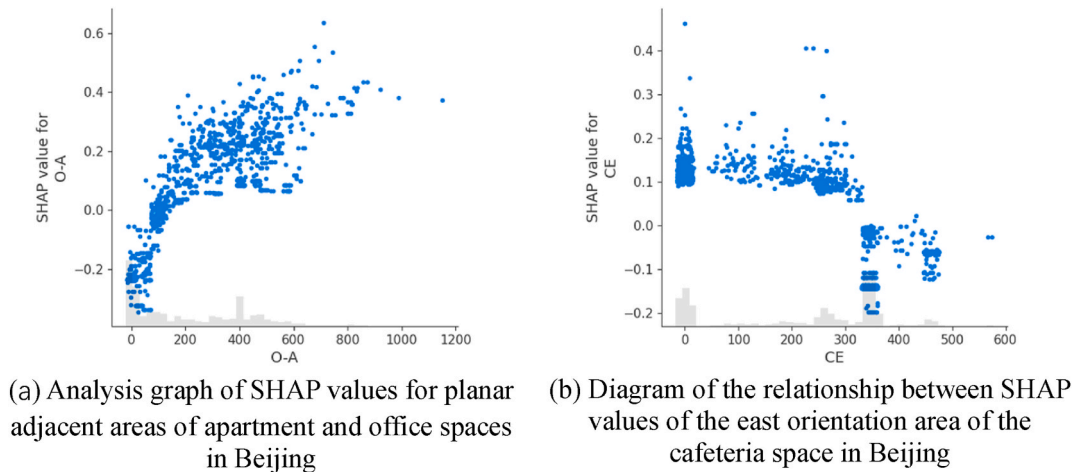


Fig. 17. Global analysis of SHAP values for Beijing.

The results of Kunming that is indicated in Fig. 12(b) show that a larger east-facing area (OE) of office spaces can significantly reduce the cooling and heating load, especially in the mild climate of Kunming, where the east-facing arrangement of office spaces plays an important energy-saving role. However, the effects of other features were smaller, reflecting the relatively stable energy consumption characteristics in the Kunming area. The climate in Kunming is relatively mild, with moderate temperatures in summer and colder winters. Increasing the area facing east can enhance the collection of sunlight in winter, effectively improve daylighting and natural heating, thereby reducing the demand for heating. The overall energy consumption in the Kunming area is relatively low, so the impact of energy-saving design is relatively weak.

In the analysis of the Shanghai group, the planar adjacent area (O-A) of office and apartment spaces had the most significant impact on cooling and heating loads (Fig. 12(c)). Larger adjacent areas increase energy consumption, while smaller adjacent areas help reduce the load. When the number of cafeteria spaces is small, cooling and heating loads can be effectively reduced, while a dispersed layout will significantly increase energy consumption. The local analysis results, which are visualized in Figs. 15 and 16 further validate the global trends: in the scheme with the lowest cooling and heating load, the cafeteria space is divided into only two areas (CC = 2), the apartment and meeting spaces are closely arranged (A-M = 576), and the office and apartment spaces are adjacent to a smaller area (O-A = 259.2), all contributing to the energy-saving effect. Shanghai has a hot summer and cold winter climate. High temperatures in summer increase the air conditioning load, while cold winters increase the heating load. By reducing the area adjacent to office and apartment spaces in the design, the efficiency of air heat exchange can be reduced, thereby reducing the cooling and heating load. At the same time, the centralized cafeteria space helps to optimize the spatial layout and reduce unnecessary energy consumption.

For Beijing, the planar proximity area (O-A) of the office and apartment spaces had the greatest impact on the cooling and heating load (Fig. 12(d)). In particular, when O-A is greater than 100 m² (Fig. 17), energy consumption increases significantly. Meanwhile, the east-facing area (CE) of the cafeteria space has a significant negative impact on load forecasting, and a larger east-facing area can effectively reduce the cooling and heating load. The cold climate in Beijing means a greater demand for heating in winter, and increasing the east-facing area of the cafeteria can effectively utilize sunlight to reduce the heating load. At the same time, reducing the area adjacent to office and apartment spaces can reduce indoor heat exchange, optimize the thermal environment of the building and lower energy consumption.

For Harbin, the number of planar regions (MC) of the conference space had the most significant effect on cooling and heating loads (Fig. 12(e)). A smaller number of meeting areas helped to reduce the cooling and heating load. When the floor area adjacent to the office and apartment spaces (O-M) is larger, the load forecast decreases; otherwise, it increases. The extremely cold climate in Harbin means extremely high heating load in winter. Under such climatic conditions, the centralized arrangement of conference spaces can reduce the building's voids and heat loss, thereby reducing the load. Reasonable adjacent arrangement of office and apartment spaces helps to reduce energy consumption and avoid unnecessary heat exchange and temperature fluctuations.

By combining the SHAP analysis results of five cities, the following key strategies for energy conservation in building spatial layouts can be identified: (1) Climate-adaptive design: The orientation of buildings, the spatial layouts, and the configuration of adjacent areas should vary under different climatic conditions. In subtropical climates (Shenzhen), eastward orientation and reduced westward lighting are key to energy conservation, while in cold climates (such as Harbin), the centralized arrangement of conference spaces and reasonable spatial layout can significantly reduce energy load. (2) Space layout distribution and layout optimization: In hot summer and cold winter or temperate climate, the rational allocation of the orientation and adjacent area of spaces, especially the layout of canteens, offices and meeting spaces, will play a significant role in energy conservation. Avoiding excessive adjacent areas and scattered layouts is an important means to reduce the building's cooling and heating load. (3) Heat load regulation and control: By optimizing the heat load on the exterior surface of the building and rationally designing the orientation, window surfaces and spatial distribution of the building, the heat exchange between the interior and exterior of the building can be effectively regulated, the energy efficiency of the building can be improved, and the heat load can be reduced.

4.5. Practical design guidelines for climate-adaptive spatial layout

The findings from optimization and SHAP analysis can be translated into actionable design strategies for practitioners working in different climate zones. This section provides specific guidelines for applying the identified principles during early-stage architectural design:

For subtropical climates (represented by Shenzhen): The dominant factor affecting energy performance is orientation area distribution, particularly east-facing exposure. Architects should minimize office spaces on east-facing facades where morning solar heat gain significantly increases cooling loads. Instead, allocate dining halls, circulation spaces, or service areas to east orientations, as these have lower sensitivity to solar heat gain. West-facing areas should also be minimized for high-occupancy spaces. For a typical rectangular building, this translates to: (1) positioning primary office zones on north and south facades where window-to-wall ratios can be optimized for daylighting without excessive heat gain, (2) locating cafeterias and meeting rooms on east and west sides where intermittent occupancy patterns better tolerate thermal fluctuations, and (3) using serviced apartments (with lower daytime occupancy) as thermal buffers on problematic orientations.

For heating-dominated climates (Beijing, Harbin, Shanghai): Space centralization and adjacency optimization become critical. Architects should consolidate spaces with similar thermal requirements to reduce heat loss through internal partitions. Practical strategies include: (1) concentrating meeting rooms in fewer, larger zones rather than dispersing them across multiple floors (reducing the number of planar areas), (2) minimizing adjacent areas between office and apartment spaces by introducing buffer zones or separating these functions vertically, (3) locating cafeterias to maximize east-facing exposure for passive solar heating during winter while maintaining centralized configurations. For Shanghai specifically, reducing office-apartment adjacency from $>400 \text{ m}^2$ to $<260 \text{ m}^2$ per floor can achieve substantial energy savings.

For mild climates (Kunming): Although absolute energy savings are modest, orientation optimization remains beneficial. East-facing office areas should be maximized to capture morning sunlight for winter heating while avoiding overheating during mild weather. The relatively low energy intensity in mild climates provides design flexibility, allowing greater emphasis on other performance criteria such as daylighting and spatial quality.

Sensitivity analysis across climate zones reveals that the relative importance of spatial variables shifts with climate intensity. In extreme climates (severely cold or hot-humid), spatial layout decisions have amplified impact—errors in space allocation result in proportionally greater energy penalties. Conversely, mild climates exhibit greater tolerance to layout variations. This climate-dependent sensitivity suggests that computational optimization investment yields highest returns in extreme climate zones where design precision matters most. During schematic design, architects can apply these guidelines by: (1) conducting preliminary zoning studies that test orientation distribution and space concentration patterns based on climate zone, (2) using the 26 quantified spatial features indicated in Table 2 as evaluation metrics to assess alternative layouts, (3) prioritizing the top-ranked SHAP features identified for their specific climate zone during iterative refinement, and (4) validating final schemes through simplified energy simulation focusing on cooling and heating loads. The automated generation method developed in this research can be adapted as a design exploration tool, with architects adjusting the priority weights of different spatial features based on climate-specific importance rankings revealed by SHAP analysis.

5. Conclusion

This study investigated the impact of spatial layout optimization on energy consumption in medium-rise office buildings across five Chinese cities with diverse climatic conditions. Using building spatial layout generation tools, evolutionary algorithms, and machine learning methods including random forest and SHAP analysis, we examined how different building layouts affect cooling and heating loads in various thermal zones. The research demonstrates that spatial layout plays a crucial role in energy-efficient building design, with the proposed optimization tool achieving approximately 10 % energy savings in cooling and heating loads. The most significant results were observed in Shenzhen's subtropical climate, where energy savings reached 17.25 %, highlighting the substantial potential for climate-adaptive design strategies.

Climate-specific patterns emerged from this study. In mild climates like Kunming, while percentage reductions in loads were notable, overall energy savings remained modest due to lower baseline consumption. Subtropical regions (Shenzhen) showed the most dramatic load fluctuations, indicating that spatial optimization is particularly critical in such climates. In heating-dominated regions including Shanghai, Beijing, and Harbin, spatial layout optimization primarily influenced heating loads, with substantial reductions achieved through strategic planning. Through quantitative analysis of layout concentration, dispersion, orientation, and spatial proximity, several climate-specific strategies were identified.

- Subtropical climates: East-facing orientations and reduced west-facing windows are essential for energy conservation
- Cold climates: Concentrating similar spaces and implementing rational spatial arrangements significantly reduce energy loads
- Temperate and hot summer/cold winter climates: Strategic space orientation and adjacent area allocation provide substantial energy benefits, while avoiding excessive adjacencies and dispersed layouts reduces cooling and heating demands

This automation-based spatial optimization methodology provides a widely adaptable framework for energy-efficient design of medium-rise office buildings across different climatic conditions. The quantitative analysis mechanisms offer targeted, climate-specific strategies that advance computational green building design practices. However, this study's scope is limited to one building type and five Chinese climate zones, potentially restricting global applicability. The geometric constraints of the spatial generation algorithm

may not capture all design variations, and the focus on cooling and heating loads excludes other energy considerations like lighting and equipment optimization. Future research should expand to include diverse building types, international climate regions, comprehensive energy requirements, and integration with renewable energy systems to develop more holistic energy-efficient design solutions.

CRediT authorship contribution statement

Peiying Huang: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Conceptualization. **Yanxiang Yang:** Writing – original draft, Validation, Investigation, Data curation. **Wen Gao:** Writing – review & editing, Supervision, Resources, Methodology, Formal analysis. **Xing Zheng:** Writing – review & editing, Supervision, Resources, Project administration, Methodology. **Pengyuan Shen:** Writing – review & editing, Writing – original draft, Supervision, Resources, Project administration, Methodology, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research is supported by Shenzhen Fundamental Research Program JCYJ20250604180231041.

Appendix

Appendix 1 3D spatial layout generation algorithm pseudocode

Algorithm 1. 3D Space Layout Generation Algorithm

Input:program, gene_data, info_data
Output:result_program

```

1:      Generate the layout for vertical circulation.
2:      vt_plan = creat_transport_plan(program, gene_data, info_data)
3:      Execute vertical growth of the vertical circulation.
4:      program = creat_vertical_mass(program, vt_plan)
5:      Perform synchronous growth for multiple connected regions.
6:      result_program = grow_program (program, gene_data, info_data)
7:      If any spatial units remain unfilled, continue growth until all units are filled.
8:      if len(program.get_attri_unit_seq(0)) > 0 do
9:          result_program = fill_program(result_program, gene_data, info_data)
10:     end

```

Appendix 2 Iterative growth algorithm pseudocode

Algorithm 2. grow_program

Input:program, gene_data, info_data
Output:program

```

1:      Set the starting point according to the gene_data.
2:      program = set_start_point(program, gene_data.start_unit)
3:      Grow the zones according to the gene_data until all zones have completed growth.
4:      while bool(finish_check) = True do
5:          stop_check = [True] * info.zone_count
6:          Grow according to the direction sequence in the gene_data.

```

(continued on next page)

(continued)

Input: program, gene_data, info_data	
Output: program	
1:	Set the starting point according to the gene_data.
2:	program = set_start_point(program, gene_data.start_unit)
3:	Grow the zones according to the gene_data until all zones have completed growth.
4:	while bool(finish_check) = True do
5:	stop_check = [True] * info.zone_count
6:	Grow according to the direction sequence in the gene_data.
7:	for dir in gene_data.direction do
8:	for zone in info_data.zone do
9:	Grow according to the step length in the gene_data.
10:	for step in gene_data.step_len[zone][dir] do
11:	Iterate through the spatial units in this zone.
12:	for unit in program[zone] do
13:	If the neighboring spatial unit in the growth direction is vacant, grow into that unit
	as part of the zone.
14:	grow_unit = get_neighbor_unit (program, unit, dir)
15:	if grow_unit.attri == 0 do
16:	program[grow_unit].attri = zone
17:	stop_check[zone] = False
18:	Check if the zone meets the required area; if so, mark the zone as completed.
19:	if info_data.area_demand[zone](1-info_data.area_tolerance)≤cal_area(zone)≤info_data.area_demand[zone](1+info_data.area_tolerance) do
20:	finish_check[zone] = True
7:	for dir in gene_data.direction do
8:	for zone in info_data.zone do
9:	Grow according to the step length in the gene_data.
10:	for step in gene_data.step_len[zone][dir] do
11:	Iterate through the spatial units in this zone.
12:	for unit in program[zone] do
13:	If the neighboring spatial unit in the growth direction is vacant, grow into that unit
	as part of the zone.
14:	grow_unit = get_neighbor_unit (program, unit, dir)
15:	if grow_unit.attri == 0 do
16:	program[grow_unit].attri = zone
17:	stop_check[zone] = False
18:	Check if the zone meets the required area; if so, mark the zone as completed.
19:	if info_data.area_demand[zone](1-info_data.area_tolerance)≤cal_area(zone)≤info_data.area_demand[zone](1+info_data.area_tolerance) do
20:	finish_check[zone] = True
21:	If a zone cannot continue growing and has not met the area requirement, select a vacant
22:	unit as the new starting point.
23:	for zone in info_data.zone do
24:	if finish_check[zone] == False and stop_check[zone] == True do
25:	new_start_unit = get_vacant_unit(program)
26:	program[new_start_unit].attri = zone
	end

Appendix 3 Schedule of personnel activity and electrical equipment usage for each spatial plan in the cooling and heating load optimization of typical medium-rise office building under different building thermal zones

		Time											
Area		1	2	3	4	5	6	7	8	9	10	11	12
Office	Weekdays	0	0	0	0	0	0	10	50	95	95	95	80
	Holidays	0	0	0	0	0	0	0	0	0	0	0	0
Meetings	Weekdays	0	0	0	0	0	0	0	50	50	50	50	0
	Holidays	0	0	0	0	0	0	0	0	0	0	0	0
Cafeteria	Weekdays	0	0	0	0	0	0	0	20	20	20	20	90
	Holidays	0	0	0	0	0	0	0	0	0	0	0	0
Serviced apartment	All year round	70	70	70	70	70	70	70	70	50	50	50	50
Traffic	All year round	0	0	0	0	0	0	10	70	70	70	70	70

		Time											
Area		13	14	15	16	17	18	19	20	21	22	23	24
Office	Weekdays	80	95	95	95	95	30	30	0	0	0	0	0
	Holidays	0	0	0	0	0	0	0	0	0	0	0	0
Meetings	Weekdays	0	50	50	50	50	50	50	0	0	0	0	0
	Holidays	0	0	0	0	0	0	0	0	0	0	0	0
Cafeteria	Weekdays	90	90	20	20	20	20	20	0	0	0	0	0
	Holidays	0	0	0	0	0	0	0	0	0	0	0	0
Serviced apartment	All year round	50	50	50	50	50	50	70	70	70	70	70	70
Transportation	All year round	70	70	70	70	70	70	70	70	50	0	0	0

Appendix 4 Lighting schedule for each spatial plan in the cooling and heating load optimization of typical medium-rise office building under different building thermal zones

		Time											
Area		1	2	3	4	5	6	7	8	9	10	11	12
Office	Weekdays	0	0	0	0	0	0	10	50	95	95	95	80
	Holidays	0	0	0	0	0	0	0	0	0	0	0	0
Meetings	Weekdays	0	0	0	0	0	0	0	50	50	50	50	0
	Holidays	0	0	0	0	0	0	0	0	0	0	0	0
Cafeteria	Weekdays	0	0	0	0	0	0	0	20	20	20	20	90
	Holidays	0	0	0	0	0	0	0	0	0	0	0	0
Serviced apartment	All year round	10	10	10	10	10	10	30	30	30	30	30	30
Transportation	All year round	10	10	10	10	10	10	30	30	30	30	30	30

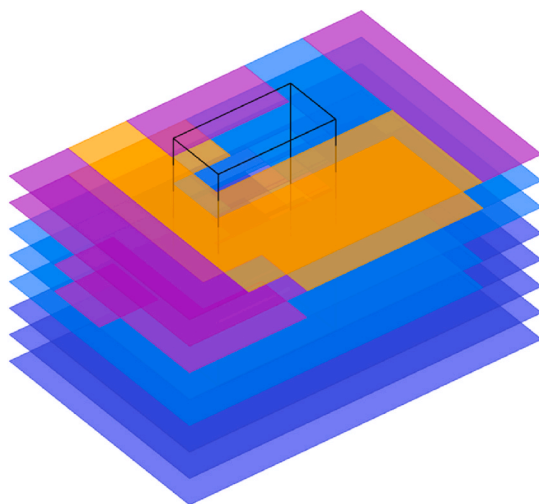
		Time											
Area		13	14	15	16	17	18	19	20	21	22	23	24
Office	Weekdays	80	95	95	95	95	30	30	0	0	0	0	0
	Holidays	0	0	0	0	0	0	0	0	0	0	0	0
Meetings	Weekdays	0	50	50	50	50	50	50	0	0	0	0	0
	Holidays	0	0	0	0	0	0	0	0	0	0	0	0
Cafeteria	Weekdays	90	90	20	20	20	20	20	0	0	0	0	0
	Holidays	0	0	0	0	0	0	0	0	0	0	0	0
Serviced apartment	All year round	30	30	50	50	60	90	90	90	90	80	10	10
Traffic	All year round	30	30	30	30	50	50	50	50	50	10	10	10

Appendix 5 Schedule of heating and cooling for each spatial plan in the cooling and heating load optimization of typical medium-rise office building under different building thermal

			Time											
Area			1	2	3	4	5	6	7	8	9	10	11	12
Office	Weekdays	air conditioner	–	–	–	–	–	–	28	26	26	26	26	26
		Heating	5	5	5	5	5	12	18	20	20	20	20	20
Meetings	Holidays	air conditioner	–	–	–	–	–	–	–	–	–	–	–	–
		Heating	5	5	5	5	5	5	5	5	5	5	5	5
Cafeteria	Weekdays	air conditioner	–	–	–	–	–	–	–	–	–	–	26	26
		Heating	5	5	5	5	5	5	5	5	5	5	18	18
	Holidays	air conditioner	–	–	–	–	–	–	–	–	–	–	–	–
		Heating	5	5	5	5	5	5	5	5	5	5	5	5
Serviced apartment	Year round	air conditioner	26	26	26	26	26	26	26	26	26	26	26	26
		Heating	22	22	22	22	22	22	22	22	22	22	22	22
Transportation	Year round	air conditioner	–	–	–	–	–	–	28	26	26	26	26	26
		Heating	5	5	5	5	5	12	18	20	20	20	20	20

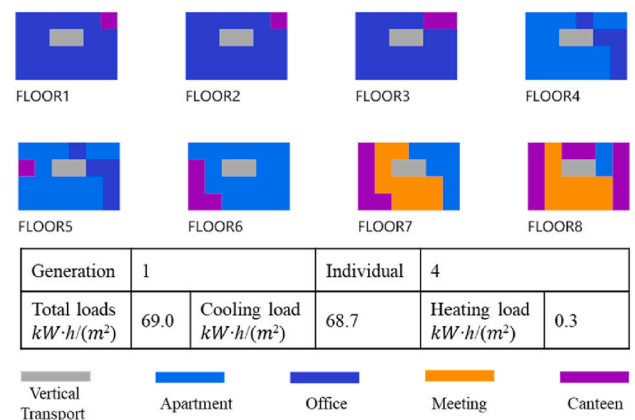
			Time											
Area			13	14	15	16	17	18	19	20	21	22	23	24
Office	Weekdays	air conditioner	26	26	26	26	26	26	–	–	–	–	–	–
		Heating	20	20	20	20	20	20	18	12	5	5	5	5
Meetings	Holidays	air conditioner	–	–	–	–	–	–	–	–	–	–	–	–
		Heating	5	5	5	5	5	5	5	5	5	5	5	5
Cafeteria	Weekdays	air conditioner	26	26	–	–	–	–	–	–	–	–	–	–
		Heating	18	18	5	5	5	5	5	5	5	5	5	5
	Holidays	air conditioner	–	–	–	–	–	–	–	–	–	–	–	–
		Heating	5	5	5	5	5	5	5	5	5	5	5	5
Serviced apartment	Year round	air conditioner	26	26	26	26	26	26	26	26	26	26	26	26
		Heating	22	22	22	22	22	22	22	22	22	22	22	22
Transportation	Year round	air conditioner	26	26	26	26	26	26	–	–	–	–	–	–
		Heating	20	20	20	20	20	20	18	12	5	5	5	5

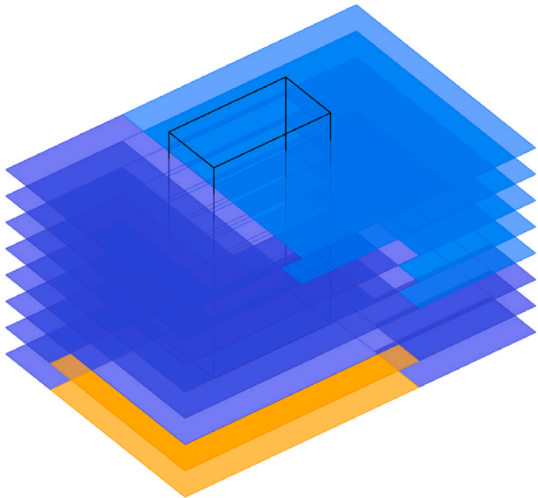
Appendix 6. The worst-case scenarios and the optimal spatial layout during the iterative process in Shenzhen, Kunming, Shanghai, and Harbin



Shenzhen

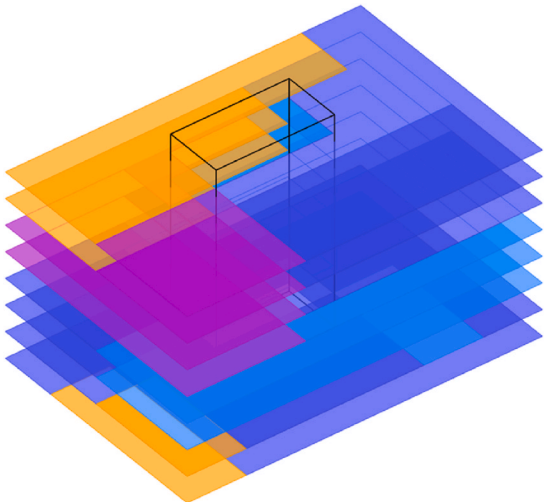
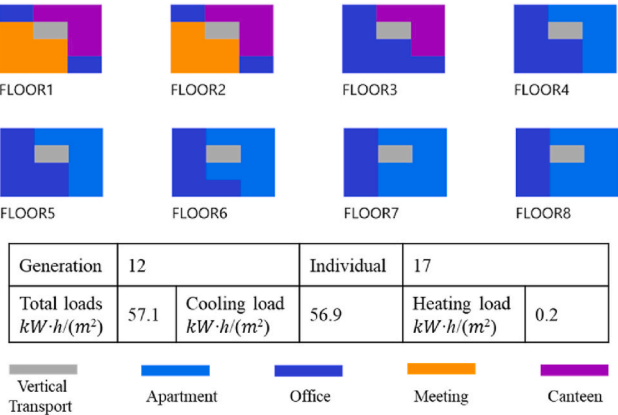
Max cooling and heating load solution





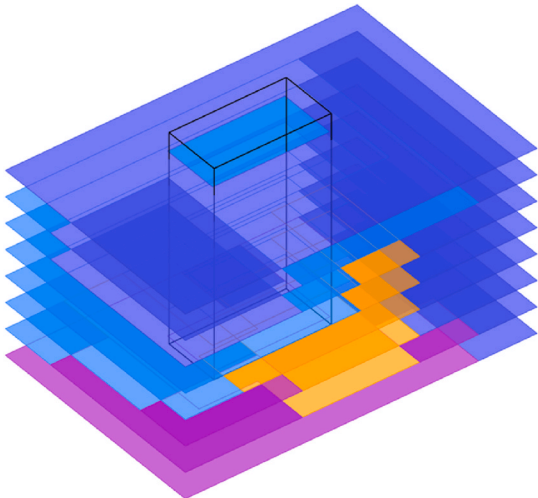
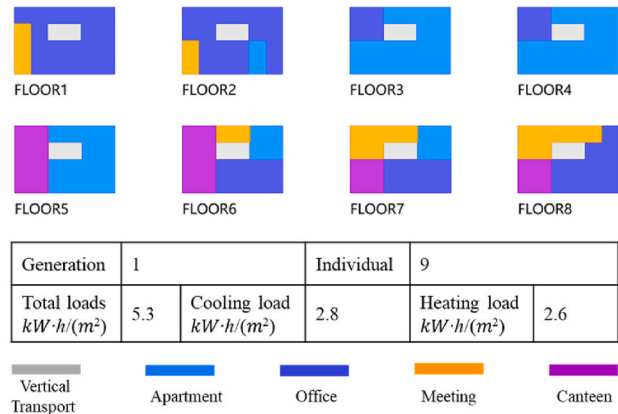
Shenzhen

Min cooling and heating load solution



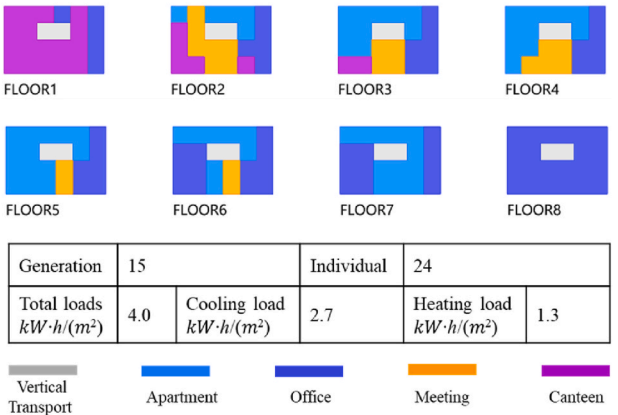
Kunming

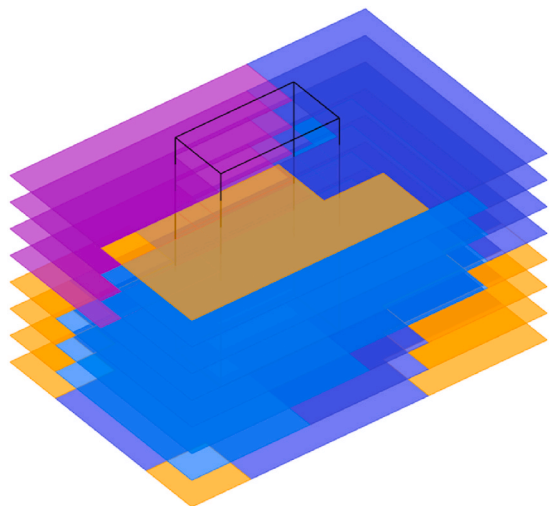
Max cooling and heating load solution



Kunming

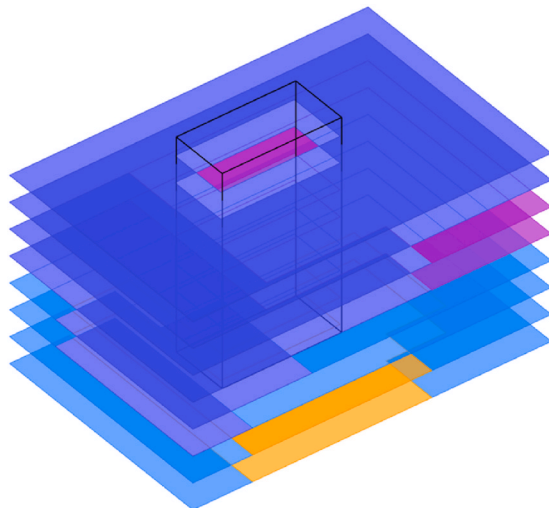
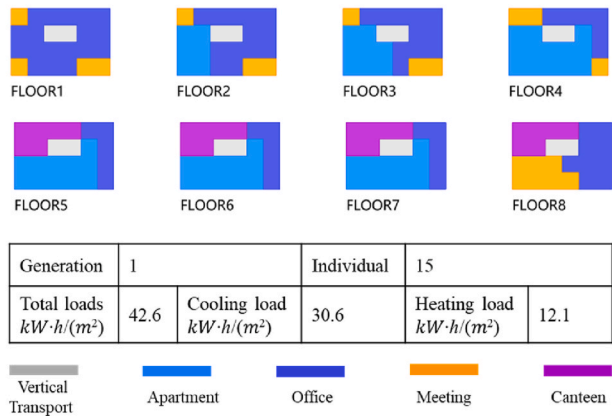
Min cooling and heating load solution





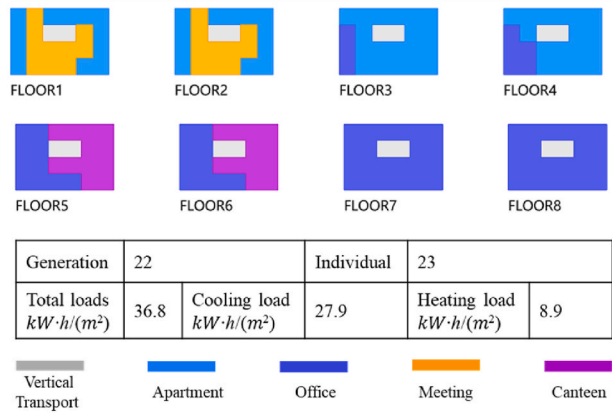
Shanghai

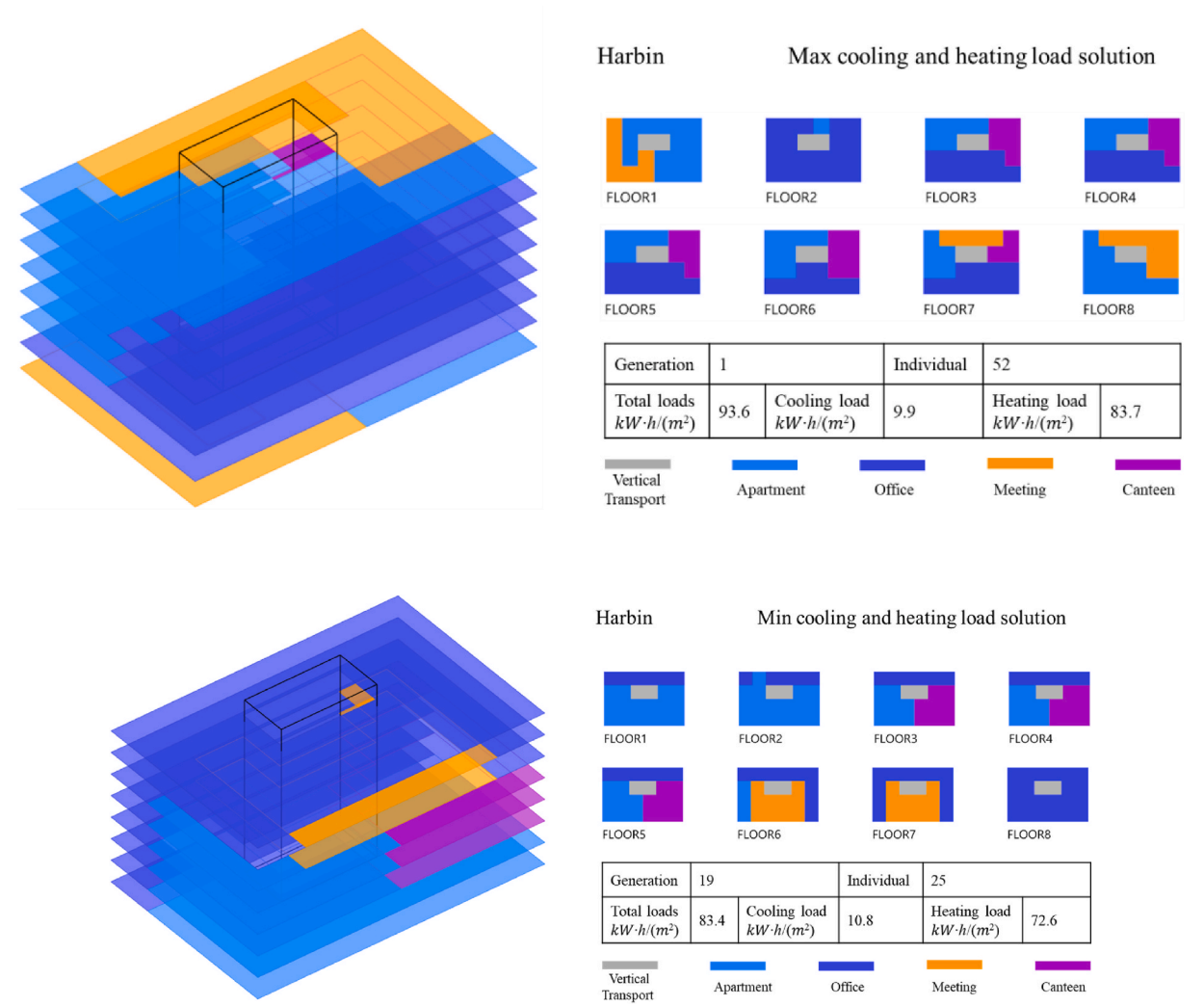
Max cooling and heating load solution



Shanghai

Min cooling and heating load solution





Data availability

Data will be made available on request.

References

[1] United Nations Environment P, Global Alliance for B, Construction. Not Just Another Brick in the Wall: the Solutions Exist - Scaling them will Build on Progress and Cut Emissions Fast. Global Status Report for Buildings and Construction 2024/2025, United Nations Environment Programme, 2025.

[2] IPCC, in: H. Lee, J. Romero (Eds.), Climate Change 2023: Synthesis Report, 2023.

[3] T.L. Hemsath, Conceptual energy modeling for architecture, planning and design: impact of using building performance simulation in early design stages, in: 3th Conference of International Building Performance Simulation Association, August 26-28. 2013. Chambéry, France.

[4] F. Kheiri, A review on optimization methods applied in energy-efficient building geometry and envelope design, Renew. Sustain. Energy Rev. 92 (2018) 897–920.

[5] S. Li, M. Wang, P. Shen, X. Cui, L. Bu, R. Wei, et al., Energy saving and thermal comfort performance of passive retrofitting measures for traditional rammed Earth house in Lingnan, China, Buildings 12 (2022) 1716.

[6] P. Shen, Y. Li, X. Gao, S. Chen, X. Cui, Y. Zhang, et al., Climate adaptability of building passive strategies to changing future urban climate: a review, Nexus 2 (2025) 1–13.

[7] P. Shen, Z. Wang, Y. Ji, Exploring potential for residential energy saving in New York using developed lightweight prototypical building models based on survey data in the past decades, Sustain. Cities Soc. 66 (2021) 102659.

[8] S. Himmetoğlu, Y. Delice, E. Kızılkaya Aydoğan, B. Uzal, Green building envelope designs in different climate and seismic zones: multi-objective ANN-based genetic algorithm, Sustain. Energy Technol. Assessments 53 (2022) 102505.

[9] S. Himmetoğlu, Y. Delice, E.K. Aydoğan, PSACONN mining algorithm for multi-factor thermal energy-efficient public building design, J. Build. Eng. 34 (2021) 102020.

- [10] N. Es-sakali, J. Pfafferott, M.O. Mghazli, M. Cherkaoui, Towards climate-responsive net zero energy rural schools: a multi-objective passive design optimization with bio-based insulations, shading, and roof vegetation, *Sustain. Cities Soc.* 120 (2025) 106142.
- [11] D. Sun, Y. Zheng, R. Duan, Energy consumption simulation and economic benefit analysis for urban electric commercial-vehicles, *Transport. Res. Transport Environ.* 101 (2021) 103083.
- [12] W. Huang, H. Zheng, Architectural drawings recognition and generation through machine learning, in: *Proceedings of the 38th Annual Conference of the Association for Computer Aided Design in Architecture*, 2018, pp. 18–20. Mexico City, Mexico.
- [13] Z. Zhang, A. Chong, Y. Pan, C. Zhang, K.P. Lam, Whole building energy model for HVAC optimal control: a practical framework based on deep reinforcement learning, *Energy Build.* 199 (2019) 472–490.
- [14] Y. Chen, T. Hong, M.A. Piette, Automatic generation and simulation of urban building energy models based on city datasets for city-scale building retrofit analysis, *Appl. Energy* 205 (2017) 323–335.
- [15] T.L. Hemsath, K. Alagheband Bandhosseini, Sensitivity analysis evaluating basic building geometry's effect on energy use, *Renew. Energy* 76 (2015) 526–538.
- [16] R. Pacheco, J. Ordóñez, G. Martínez, Energy efficient design of building: a review, *Renew. Sustain. Energy Rev.* 16 (2012) 3559–3573.
- [17] H. Latha, S. Patil, P.G. Kini, JJoE, Eng. E. Influence of Architectural Space Layout and Building Perimeter on the Energy Performance of Buildings: a Systematic Literature Review, vol. 14, 2023, pp. 431–474.
- [18] T. Du, S. Jansen, M. Turrin, Dobbelsesteen Avd, Effect of space layouts on the energy performance of office buildings in three climates, *J. Build. Eng.* 39 (2021).
- [19] T. Du, M. Turrin, S. Jansen, A. van den Dobbelsesteen, F. De Luca, Relationship analysis and optimisation of space layout to improve the energy performance of office buildings, *Energies* 15 (2022).
- [20] I.G. Dino, G. Üçoluk, Multiobjective design optimization of building space layout, energy, and daylighting performance, *J. Comput. Civ. Eng.* 31 (2017).
- [21] P. Shen, Y. Li, X. Gao, Y. Zheng, P. Huang, A. Lu, et al., Recent progress in building energy retrofit analysis under changing future climate: a review, *Appl. Energy* 383 (2025) 125441.
- [22] Y. Li, L. Li, X. Cui, P. Shen, Coupled building simulation and CFD for real-time window and HVAC control in sports space, *J. Build. Eng.* 97 (2024) 110731.
- [23] Y. Li, L. Li, P. Shen, Probability-based visual comfort assessment and optimization in national fitness halls under sports behavior uncertainty, *Build. Environ.* (2023) 110596.
- [24] Z. Ma, G. Jiang, Y. Hu, J. Chen, A review of physics-informed machine learning for building energy modeling, *Appl. Energy* 381 (2025) 125169.
- [25] P.W. Tien, S. Wei, J. Darkwa, C. Wood, J.K. Calautit, Machine learning and deep learning methods for enhancing building energy efficiency and indoor environmental quality – a review, *Energy AI* 10 (2022) 100198.
- [26] T. Du, S. Jansen, M. Turrin, A. van den Dobbelsesteen, Effects of architectural space layouts on energy performance: a review, *Sustainability* 12 (2020).
- [27] T. Dogan, C. Reinhart, P. Michalatos, Autozoner: an algorithm for automatic thermal zoning of buildings with unknown interior space definitions, *J. Build. Perform. Simulat.* 9 (2016) 176–189.
- [28] H. Shen, A. Tzempelikos, Sensitivity analysis on daylighting and energy performance of perimeter offices with automated shading, *Build. Environ.* 59 (2013) 303–314.
- [29] P. Delgoshaei, M. Heidarinejad, K. Xu, J.R. Wentz, P. Delgoshaei, J. Srebric, Impacts of building operational schedules and occupants on the lighting energy consumption patterns of an office space, *Build. Simulat.* 10 (2017) 447–458.
- [30] P. Shen, W. Braham, Y. Yi, Development of a lightweight building simulation tool using simplified zone thermal coupling for fast parametric study, *Appl. Energy* 223 (2018) 188–214.
- [31] K. Verichev, M. Zamorano, A. Fuentes-Sepúlveda, N. Cárdenas, M. Carpio, Adaptation and mitigation to climate change of envelope wall thermal insulation of residential buildings in a temperate oceanic climate, *Energy Build.* 235 (2021).
- [32] C. Baglivo, P.M. Congedo, G. Murrone, D. Lezzi, Long-term predictive energy analysis of a high-performance building in a mediterranean climate under climate change, *Energy* 238 (2022).
- [33] Y. Shi, Z. Yan, C. Li, C. Li, Energy consumption and building layouts of public hospital buildings: a survey of 30 buildings in the cold region of China, *Sustain. Cities Soc.* 74 (2021).
- [34] J. Dai, S. Jiang, J. Li, X. Xu, M. Wu, The influence of layout on energy performance of university building, *IOP Conf. Ser. Earth Environ. Sci.* 371 (2019).
- [35] T. Cheng, N. Wang, C.H. Liu, Research on energy consumption of building layout and envelope for rural housing in the cold region of China, *IOP Conf. Ser. Earth Environ. Sci.* 238 (2019).
- [36] I. Susorova, M. Tabilzadeh, A. Rahman, H.L. Clack, M. Elnimeiri, The effect of geometry factors on fenestration energy performance and energy savings in office buildings, *Energy Build.* 57 (2013) 6–13.
- [37] Y. Pan, M. Zhu, Y. Lv, Y. Yang, Y. Liang, R. Yin, et al., Building Energy Simulation and its Application for Building Performance Optimization: a Review of Methods, Tools, and Case Studies, vol. 10, 2023 100135.
- [38] A.-T. Nguyen, S. Reiter, Rigo Pjae, A Review on Simulation-based Optimization Methods Applied to Building Performance Analysis, vol.113, 2014, pp. 1043–1058.
- [39] P. Shen, Building retrofit optimization considering future climate and decision-making under various mindsets, *J. Build. Eng.* 96 (2024) 110422.
- [40] N. Delgarm, B. Sajadi, F. Kowsary, SJAe Delgarm, Multi-objective optimization of the building energy performance: a simulation-based approach by means of particle swarm optimization, *PSO* 170 (2016) 293–303.
- [41] A. Vukadinović, J. Radosavljević, A. Đorđević, M. Protić, N.J.S.E. Petrović, Multi-Objective Optimization of Energy Performance for a Detached Residential Building with a Sunspace Using the NSGA-II Genetic Algorithm, vol.224, 2021, pp. 1426–1444.
- [42] M. Ghaderian, F. Veysi, Multi-objective optimization of energy efficiency and thermal comfort in an existing office building using NSGA-II with fitness approximation, A case study 41 (2021) 102440.
- [43] P. Shen, W. Braham, Y. Yi, E. Eaton, Rapid multi-objective optimization with multi-year future weather condition and decision-making support for building retrofit, *Energy* 172 (2019) 892–912.
- [44] L. Breiman, *Random Forests*, 2001.
- [45] M. Zekić-Sušac, A. Has, M. Knežević, Predicting energy cost of public buildings by artificial neural networks, CART, and random forest, *Neurocomputing* 439 (2021) 223–233.
- [46] G.K.F. Tso, K.K.W. Yau, Predicting electricity energy consumption: a comparison of regression analysis, decision tree and neural networks, *Energy* 32 (2007) 1761–1768.
- [47] P.P. Angelov, E.A. Soares, R. Jiang, N.I. Arnold, P.M. Atkinson, Explainable artificial intelligence: an analytical review, *WIREs Data Min. Knowl. Discov.* 11 (2021).
- [48] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, et al., Explainable AI (XAI): core ideas, techniques, and solutions, *ACM Comput. Surv.* 55 (2023) 1–33.
- [49] S.M. Lundberg, S.-I. Lee, *A Unified Approach to Interpreting Model Predictions*, 2017.
- [50] Y. Wu, Y. Zhou, Hybrid machine learning model and shapley additive explanations for compressive strength of sustainable concrete, *Constr. Build. Mater.* 330 (2022).
- [51] P. Meddage, I. Ekanayake, U.S. Perera, H.M. Azamathulla, M.A. Md Said, U. Rathnayake, Interpretation of machine-learning-based (Black-box) wind pressure predictions for low-rise gable-roofed buildings using shapley additive explanations (SHAP), *Buildings* 12 (2022).
- [52] P. Arjunan, K. Poolla, C. Miller, EnergyStar++: towards more accurate and explanatory building energy benchmarking, *Appl. Energy* 276 (2020).
- [53] M. Vega Garcia, J.L. Aznarte, Shapley additive explanations for NO2 forecasting, *Ecol. Inform.* 56 (2020).
- [54] P. Shen, H. Wang, Archetype building energy modeling approaches and applications: a review, *Renew. Sustain. Energy Rev.* 199 (2024) 114478.
- [55] M. Mona, A. Reem, Effect of selecting validation dataset on building random forest and decision tree models, *AlQalam J. Med. Appl. Sci.* 5 (2022) 470–478.
- [56] H. Bichri, A. Chergui, M. Hain, Investigating the impact of train/test split ratio on the performance of pre-trained models with custom datasets, *Int. J. Adv. Comput. Sci. Appl.* 15 (2024).

- [57] J. Bergstra, Y. Bengio, Random search for hyper-parameter optimization, *J. Mach. Learn. Res.* 13 (2012) 281–305.
- [58] P. Probst, M.N. Wright, A.L. Boulesteix, Hyperparameters and tuning strategies for random forest, *Wiley Interdiscip. Rev.: Data Min. Knowl. Discov.* 9 (2019) e1301.
- [59] E. González-Estrada, W. Cosmes, Shapiro–Wilk test for skew normal distributions based on data transformations, *J. Stat. Comput. Simulat.* 89 (2019) 3258–3272.
- [60] M.K.M. Shapi, N.A. Ramli, L.J. Awalin, Energy consumption prediction by using machine learning for smart building: case study in Malaysia, *Dev. Built Environ.* 5 (2021) 100037.
- [61] Y. Arima, R. Ooka, H. Kikumoto, Proposal of typical and design weather year for building energy simulation, *Energy Build.* 139 (2017) 517–524.
- [62] N. Amin, F. Jérôme, vT. Christoph, Statistical methodologies for verification of building energy performance simulation, in: *Proceedings of Building Simulation 2021: 17Th Conference of IBPSA, IBPSA, 2021*, pp. 1719–1726.
- [63] R.F. Mustapa, N.Y. Dahlan, I.M. Yassin, A.H.M. Nordin, M.E. Mahadan, Baseline energy modelling in an educational building campus for measurement and verification, in: *2017 International Conference on Electrical, Electronics and System Engineering (ICEESE)*, 2017, pp. 67–72.
- [64] C. Fan, F. Xiao, S. Wang, Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques, *Appl. Energy* 127 (2014) 1–10.
- [65] N.-T. Ngo, A.-D. Pham, T.T.H. Truong, N.-S. Truong, N.-T. Huynh, Developing a hybrid time-series artificial intelligence model to forecast energy use in buildings, *Sci. Rep.* 12 (2022) 15775.
- [66] E. Yaghoubi, E. Yaghoubi, A. Khamees, A.H. Vakili, A systematic review and meta-analysis of artificial neural network, machine learning, deep learning, and ensemble learning approaches in field of geotechnical engineering, *Neural Comput. Appl.* 36 (2024) 12655–12699.
- [67] J. Long, K. Xueyuan, H. Huang, Q. Zhinian, Y. Wang, Study on the overfitting of the artificial neural network forecasting model, *Acta Meteorol. Sin.* 19 (2005) 216.
- [68] J. Bergstra, R. Bardenet, Y. Bengio, B. Kégl, Algorithms for hyper-parameter optimization, *Adv. Neural Inf. Process. Syst.* 24 (2011).
- [69] M.T. Ribeiro, S. Singh, C. Guestrin, Why should I trust you?, in: *Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining Association for Computing Machinery, San Francisco, California, USA, 2016*, pp. 1135–1144.