A graph-based multi-modal deep learning framework for automated energy performance assessment and retrofit decision support in Chinese residential buildings

Xingjian Zhao

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

The University of Sheffield

School of Architecture and Landscape, Faculty of Social Sciences

Submission Date: 31, August 2025

Declaration

I, Xingjian Zhao, declare that I am the sole author of this thesis and that the research presented within is a result of my own efforts and achievements. Where this is not the case, it has clearly been stated. I confirm that this work has not been submitted for any other degree.

Abstract

Over recent decades, the increase in global energy consumption has led to negative environmental impacts and has become a major concern that hinders social development. Since 2007, the Chinese government has implemented large-scale energy-efficiency retrofit programmes for existing buildings. However, in practical projects, the evaluation and decision-making process for retrofit options is still constrained by incomplete building information and the limited efficiency of current assessment workflows. To support retrofit decision-making, this thesis proposes a data-driven approach for household-level energy prediction and retrofit decision support. The research is organised into four interconnected stages. First, based on existing floor plans of ageing residential buildings, an automated framework for architectural element recognition and semantic retrieval is developed, enabling the automatic generation of baseline Building Energy Models (BEMs). Second, these baseline models are enriched with thermal material properties and occupant behaviour parameters, defined in accordance with the current literature. Large-scale parametric energy simulations are then carried out to produce a comprehensive residential building energy consumption dataset tailored to the Chinese housing context. Third, a multi-modal deep learning model, FusionGNN, is proposed to predict household annual Energy Use Intensity (EUI). The model is first pre-trained on the simulation dataset and then fine-tuned using measured data collected from 120 households. Finally, the calibrated FusionGNN is applied to a case study of four typical residential typologies to rapidly evaluate energy performance under various retrofit scenarios. By integrating predicted energy outcomes with retrofit cost assessments, a multi-objective optimisation is conducted to identify the optimal retrofit strategies. Overall, this research provides a scalable data-driven decision-support framework for residential building retrofits in China, offering evidence to help stakeholders achieve a more appropriate balance between energy savings and economic feasibility.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my parents, Chong Zhao and Dongmei Li, for their unwavering financial and emotional support throughout my PhD journey. Their constant encouragement and belief in me have been invaluable, providing me with the strength to persevere through challenges and achieve my goals.

I would also like to extend my heartfelt appreciation to my supervisors, Tsung-Hsien Wang and Chengzhi Peng, for their invaluable guidance, patience, and encouragement throughout my research. Their expertise, insightful advice, and continuous support have been instrumental in shaping my academic and research development.

Additionally, I am sincerely grateful to my friends, Minglong Li, Zidi Yang, and Mincen Yu, for their companionship and support during my time in the UK. Their kindness, encouragement, and help in both academic and personal aspects of my life have made this journey more meaningful and fulfilling.

Finally, I would like to thank everyone who has supported me in any way during my PhD journey. Your encouragement and guidance have played a significant role in the completion of this thesis, and I am truly grateful.

Contents

Abstract 3

Acknowledgements 4

1. Introduction 11

1.1 Background 11

1.2 Research questions 14

1.3 Thesis structure 15

2. Literature Review 16

2.1 Overview of China's Residential Buildings and Energy Retrofits 16

2.1.1 Evolution of residential buildings 16

2.1.2 Progress and challenges of residential energy retrofits 17

2.2 Automated Residential Floor Plan Analysis 20

2.2.1 Space reconstruction from 2D floor plan 20

2.2.2 Room function classification 23

2.3 Current Building Energy Performance Evaluation Methods 26

2.3.1 Physical energy models 26

2.3.2 Data-driven models 29

2.3.3 Hybrid models 31

2.4 Current Applications of Deep Learning in Residential Building Energy Retrofit 34

2.4.1 Building geometrical feature extraction 35

2.4.2 Occupant behaviour modelling 37

2.4.3 Inter-unit thermal interaction encoding 42

2.5 Research Gaps 47

3. Methodology 50

3.1 Automated Building Energy Modelling 52

3.1.1 Residential building data collection 52

3.1.2 Automated building energy modelling 58

3.2 Residential Building Energy Consumption Dataset Development 69

3.2.1 Preliminary BEM enrichment 70

3.2.2 Large-scale parametric energy simulation 79
3.2.3 Dataset construction 79
3.3 FusionGNN: A Multi-Modal Graph-Based Energy Prediction Model 80
3.3.1 Data preparation and feature construction 81
3.3.2 Model development 85
3.3.3 Model calibration 92
3.4 Case Study 95
3.4.1 Energy performance evaluation under different retrofit scenarios 95
3.4.2 Retrofit cost calculation 101
3.4.3 Multi-objective optimisation 101
4. Results 102
4.1 Results of Automated Building Energy Modelling 102
4.1.1 Architectural element recognition 102
4.1.2 Multi-household reconfiguration 104
4.1.3 Room type classification 107
4.1.4 Automated energy modelling 107
4.2 Results of Large-scale Parametric Energy Simulation 108
4.2.1 Overview 109
4.2.2 Parameter distribution 111
4.2.3 Energy use intensity distribution 115
4.2.4 Cross-scale sensitivity analysis 117
4.3 Results of FusionGNN 121
4.3.1 Training performance 122
4.3.2 Statistics of the real-world energy use dataset 125
4.3.3 Performance gap between simulation and measurements 128
4.3.4 Transfer learning calibration 129
4.4 Results of Case Study 130
4.4.1 Typology A - one stair with two units 131
4.4.2 Typology B - one stair with three units 132
4.4.3 Typology C - one stair with four units 133

4.4.4 Typology D - two stairs with four units 134
5. Discussion 135
5.1 Discussion of Automated Building Energy Modelling 135
5.1.1 Floor plan segmentation to vectorisation 136
5.1.2 Multi-household reconfiguration 137
5.1.3 Room type classification 139
5.1.4 Automated energy modelling 141
5.1.5 Limitations 142
5.2 Discussion of Large-scale Parametric Energy Simulation 143
5.2.1 Dataset representativeness and uncertainty 144
5.2.2 Occupant behaviour modelling and uncertainty 146
5.2.3 Differences in correlations across room, household, and building scales 148
5.3 Discussion of FusionGNN 150
5.3.1 Distribution of high-loss nodes 150
5.3.2 Model transferability and scalability 153
5.3.3 Representativeness of the real-world energy use dataset 154
5.3.4 Simulation to real performance gap 155
5.3.5 Transfer learning 157
5.4 Discussion of Case Study 159
5.4.1 Discussion of typology-dependent optimisation 159
5.4.2 Cost assumption and uncertainty 161
5.4.3 Interpretability and decision-maker-facing outputs 162
6. Conclusion 164
6.1 Answers to the Research Questions 164
6.2 Contributions 166
6.2.1 Methodological contributions 166
6.2.2 Empirical contributions 167
6.2.3 Conceptual or strategic contributions 168
6.3 Limitations 169

6.4 Future work 171
Appendix A 174
Appendix B 181
Appendix C 185
References 186

Fig. 1 Single-household floor plan examples 24
Fig. 2 Methodology framework 51
Fig. 3 Climate zones in China 53
Fig. 4 Statistics of residential communities in Jinan 55
Fig. 5 Example of collected floor plans 56
Fig. 6 Four typical floor plan typologies in 64 ageing residential communities 57
Fig. 7 Annotation examples (a) raw floor plan image (b) floor plan annotation with wall, door, window, and room region (c) visualisation of room cluster annotation (d) visualisation of room type annotation 58
Fig. 8 Floor plan segmentation using the modified DMTN model (a) the input sample floor plan (b) four architectural element segmentation results 60
Fig. 9 Image processing for floor plan vectorisation (a) segmented floor plan (b) room pixels (c) wall, door and window pixels (d) iterative room boundary dilation (e) vectorised floor plan 61
Fig. 10 Multi-household reconfiguration through the CTRC algorithm (a) vectorised floor plan; (b) core nodes highlighted in the spatial connectivity network; (c) core node classification; (d) household cluster identification (e) room node search (f) reconfigured floor plan 63
Fig. 11 Individual household room feature collection and room type prediction using Graph Attention Network 66
Fig. 12 Visualisation of input floor plan image and automatically generated energy model geometry 68
Fig. 13 Urban context reconstruction through building footprint data (a) building footprint (b) reconstructed urban context 71

Fig. 14 Contextual building shading visualisation 72
Fig. 15 Typical Chinese residential occupant behaviours 76
Fig. 16 Example of random schedule perturbation 77
Fig. 17 Five input 2D feature channels and the room index map 83
Fig. 18 Four typical residential floor plan typologies and corresponding energy models 96
Fig. 19 Multi-household reconfiguration result using the CTRC algorithm (left: vectorised floor plan, middle: room cluster ground truth, right: household clustering result through CTRC) 106
Fig. 20 Relative error statistics of Group1 and Group2 against the ground-truth models 108
Fig. 21 Statistics of household area, total floors and household configuration in the simulation dataset 111
Fig. 22 Statistics of shading building coverage ratio and mean height across different orientations 112
Fig. 23 Statistics of room occupancy rate and household inactive months 115
Fig. 24 Statistics of simulated EUIs in the simulation dataset 117
Fig. 25 Learning curve of FusionGNN 123
Fig. 26 Predictions vs. actual labels and residuals at epochs 40, 50, and 60 124
Fig. 27 Household node prediction loss range 125
Fig. 28 Statistics of household area, floor levels, household configuration and occupant behaviour archetype from questionnaire 127
Fig. 29 Measured EUI distribution from questionnaire 128
Fig. 30 Simulated EUI and measured EUI distribution 129
Fig. 31 Predicted and measured EUI before (a) and after (b) transfer learning 130
Fig. 32 Pareto front between retrofit cost and energy-saving of Typology A .. 132
Fig. 33 Pareto front between retrofit cost and energy-saving of Typology B .. 133
Fig. 34 Pareto front between retrofit cost and energy-saving of Typology C . 134
Fig. 35 Pareto front between retrofit cost and energy-saving of Typology D . 135
Fig. 36 Wall errors of architectural element extraction (a) input floor plan (b)

segmented floor plan (c) vectorised floor plan 137
Fig. 37 Examples of core-node misclassification (a) misclassification of communal and function core nodes (b) misclassification of communal and household core nodes 139
Fig. 38 Top 10 features with the highest standardised mean differences for high-loss nodes 151

Table. 1 Example of thermal zone identifier naming scheme 68
Table. 2 Energy simulation parameter configurations 75
Table. 3 Energy simulation parameter assignment in different levels 79
Table. 4 Building energy-related parameter configurations 97
Table. 5 Insulation materials for building energy retrofit 98
Table. 6 Windows for building energy retrofit 99
Table. 7 Retrofit scenario codes 100
Table. 8 Precision, recall and F1 score of architectural element segmentation (modified DMTN) 104
Table. 9 Confusion matrix of the core node classification 104
Table. 10 Confusion matrix of the room type classification 107
Table. 11 Statistics of building parameters in the simulation dataset 114
Table. 12 Top correlations between EUI (Parameter 1) and parameters (Parameter 2) at the room level 119
Table. 13 Top correlations between EUI (Parameter 1) and parameters (Parameter 2) at the household level 120
Table. 14 Top correlations between EUI (Parameter 1) and parameters (Parameter 2) at the building level 121
Table. 15 Room type classification accuracy among different approaches 140

1. Introduction

1.1 Background

Over the past decades, energy consumption has continued to rise (Cristino et al., 2022), causing irreversible impacts on the natural environment (Liu et al., 2022). In 2024, global energy demand grew by about 2.2%, nearly twice the average annual growth rate over the past decade. Among the three major energy-intensive sectors (industry, transportation, and buildings), the building sector is the largest energy consumer; in 2019 it accounted for 36% of final energy consumption and 38% of total carbon emissions (Sun et al., 2022). As the largest developing country in the world, China accounted for approximately 23.6% of global energy consumption in 2018, and buildings are responsible for 31.4% of the total energy consumption (Guo et al., 2020). With continued urbanisation and improvements in living standards, residential building energy use in China has increased rapidly (Wang et al., 2019b). Statistics show that residential buildings account for around 13% of the carbon emissions within the building sector, with the protracted operational phase of these structures representing over 80% of their total life-cycle impacts (Yan et al., 2022).

China possesses a substantial inventory of existing residential structures, with residential floor area constituting around 79% of the total building floor area (Wu and Ying, 2024). Numerous existing buildings were constructed in the past and thus demonstrate inadequate envelope insulation, material deterioration, and inefficient energy systems, resulting in elevated energy usage (Chen et al., 2024). To mitigate rapidly growing energy consumption and associated environmental impact, the Chinese government has implemented the energy-efficient retrofit project of more than 500 million square metres of existing residential buildings (Liang et al., 2017) to enhance the energy performance of existing buildings and reduce greenhouse gas emissions. At present, residential retrofits are concentrated mainly in ageing urban communities in the cold northern regions, with common measures including external wall insulation, replacement of energy-efficient windows and doors, and heating

system upgrades. By the end of 2016, almost 1.3 billion m^2 of existing buildings nationwide had been subjected to energy retrofits, resulting in significant decreases in heating and cooling demand, enhanced thermal comfort, and reduced carbon emissions (Wu et al., 2024).

However, given the large number and diversity of ageing communities, current retrofit schemes remain largely experience-driven and lack quantitative, building-specific assessments of energy-savings (Xu et al., 2021). Many retrofit projects also lack systematic performance evaluation at the design stage, underscoring the need for more rigorous technical guidelines and data-driven planning tools (Xue et al., 2022). Due to the absence of 3D models in ageing neighbourhoods and the prevalence of 2D archival data, engineers are required to manually create computable digital models for energy evaluation and retrofit design (Feng et al., 2022). In residential communities with several buildings and retrofit initiatives, such individualised reconstruction is labour-intensive and time-consuming.

In multi-household residential buildings, different households can exhibit markedly different energy use due to variations in thermostat set points, operating durations, and lifestyle habits. Occupant behaviour is widely considered one of the most influential factors in building energy modelling; behavioural differences alone can lead to roughly 10 – 25% variation in energy use among otherwise similar households (Zhang et al., 2024). This implies that the realised savings of the same retrofit measure may deviate substantially from predictions across households. More importantly, thermal coupling exists within multi-household buildings: adjacent units exchange heat through party walls and floor slabs, so one household's cooling or heating behaviour can affect its neighbours (van den Brom et al., 2018). Such inter-unit heat transfer is regarded as an important contributor to gaps between predicted and measured performance.

Traditional Building Performance Simulation (BPS) remains a classical tool for evaluating energy retrofit schemes. These methods, grounded in physical models and thermodynamic equations, build detailed representations of buildings and perform energy balance calculations (Shu and Zhao, 2023). Common tools include

EnergyPlus, DOE-2, and DeST, which can simulate heating and cooling loads and energy use under pre-retrofit and post-retrofit conditions to support savings estimation and option ranking. However, in practice traditional simulations depend heavily on comprehensive, high-fidelity input data. Reliable analyses require complete information on envelope thermal performance, the type and efficiency of the heating, ventilation, and air conditioning (HVAC) system, as well as occupancy and equipment schedules. For ageing communities, design and construction records are often incomplete, and data on heterogeneous household behaviours and equipment usage are difficult to collect (Dino et al., 2020). Moreover, the computational complexity and long runtimes of physical models pose pronounced efficiency bottlenecks for large-scale, multi-scenario retrofit assessments.

In recent years, with the rise of big data and machine learning, data-driven building energy prediction has developed rapidly. Unlike physics-based simulations, data-driven models learn hidden patterns between building features and energy use from large samples and then predict outcomes for new cases (Deb and Schlueter, 2021). Common approaches include machine learning algorithms (e.g., random forests, support vector machines) and deep learning models (e.g., Recurrent Neural Networks and Convolutional Neural Networks). Compared with physical models, data-driven methods are more flexible in terms of inputs and can deliver performance predictions without requiring complete building detail (Ali et al., 2024). More importantly, they are computationally efficient and can provide near real-time estimates, enabling rapid, large-batch evaluation of numerous retrofit scenarios.

Despite their advantages, data-driven models face several limitations and challenges. First, they depend on high-quality building energy datasets for training and validation; in practice, many ageing communities lack comprehensive energy monitoring systems, making actual consumption data difficult to obtain. Second, residential energy use is jointly driven by building geometry and physical attributes, climate, and occupant behaviour. Effectively integrating heterogeneous data types, such as geometry, physical properties, and behavioural habits, into a single model remains a key research challenge. Differences in scale and complex correlations across

modalities mean that naive feature concatenation can introduce noise or lead to over-fitting; this calls for more robust feature engineering and representation learning to convert multi-modal information into model-usable embeddings. In addition, the complex thermal coupling among households in residential buildings necessitates features that capture inter-unit relationships and unit positions. In multi-household buildings, internal heat exchange can affect both unit-level and whole-building energy use; treating each household as an independent sample often fails to capture neighbourhood dependencies.

1.2 Research questions

Building on the above, this study investigates five research questions:

1. How to establish an automated building energy modelling framework for multi-household residential buildings that maintains engineering accuracy while reducing manual modelling time and cost?
2. How to construct a high-quality energy dataset that integrates urban context, building geometry, physical attributes, occupant behaviour, adjacency relationships, and energy use data to support deep learning model development?
3. Across the room, household, and building scales, are the impacts and contributions of fine-grained features to energy use consistent, and how should retrofit strategies be prioritised across at different scales?
4. How to develop a data-driven model that fuses multi-modal and multi-scale features and explicitly encodes intra-building thermal coupling for energy-use prediction?
5. For four typical legacy residential estate layouts, how do the emphases of retrofit strategies differ, and how can multi-objective optimality between energy-savings and cost be achieved?

1.3 Thesis structure

This thesis is organised as follows: Chapter 1 introduces the research background and motivation, formulates the research questions, and outlines the methodological pipeline and overall structure. Chapter 2 reviews the development of residential energy retrofits in China, automatic residential floor plan analysis, current building energy performance evaluation methods (physics-based, data-driven, and hybrid), and applications of deep learning in residential retrofits, and identifies the research gaps. Chapter 3 presents the methodology, including an automated building energy modelling framework for multi-household housing, construction of a China multi-household residential energy dataset, development and calibration of the energy prediction model FusionGNN, and a case study that optimises retrofit measures for four typical residential typologies. Chapter 4 reports the results, covering the stepwise accuracy of the automated pipeline, parameter and energy use distributions of the dataset, FusionGNN training accuracy and field calibration, and Pareto front outcomes for layout specific retrofit strategies. Chapter 5 presents the discussion, examining error sources in the automated energy-modelling workflow; conducting correlation analyses between inputs and energy use at the room, household, and building scales; analysing the distribution of high-loss nodes in FusionGNN and the distribution of calibration samples; and identifying retrofit priorities for different unit-layout typologies. Chapter 6 concludes by summarising contributions, noting limitations, and outlining directions for future work.

2. Literature Review

2.1 Overview of China's Residential Buildings and Energy Retrofits

2.1.1 Evolution of residential buildings

With the rapid pace of urbanisation and the continuous improvement of living standards, both the total floor area and operational energy consumption of residential buildings in China have grown significantly. By 2019, the total existing building stock had reached approximately 64.4 billion square metres, with residential buildings comprising approximately 79% of the total building area, equivalent to 51 billion square metres (Wu and Ying, 2024). According to Yan et al. (2022), carbon emissions from the operation of residential buildings increased at an average annual rate of 4.53% between 2000 and 2018, with the expansion of residential floor area identified as the primary contributing factor.

To better understand the evolution of energy use in Chinese residential sector, the development of residential buildings can be categorised into three major phases (Peng et al., 2021). The first phase started in 1949 and ended in 1976. Under the welfare-oriented public housing allocation policy, over 0.5 billion square metres of building areas were constructed by governments or state-owned companies (Wang, 2016). Most of these buildings have reached the end of their designed service lifetime and have been gradually abandoned and demolished for new residential building developments. The second phase commenced in the early 1980s and concluded in the late 1990s. Following the implementation of major economic reforms, China progressively transformed its planned economic structure, resulting in rapid urbanisation. During this period, more than 184 million individuals relocated from rural regions to urban centres, increasing the urbanisation rate from 18.96% to 30.40%. Simultaneously, the residential building areas expanded from 1.4 to 4.9 billion square metres, reflecting a total expansion of 3.5 billion square metres to meet the rising urban population (China, 2019). Despite ongoing utilisation, these buildings are

progressively approaching the conclusion of their intended lifespan. The final phase mostly spans from the commencement of the 21st century until the present. Over 400 million individuals relocated to urban regions over this timeframe, and almost 24.1 billion square metres of residential construction were developed (Deng et al., 2018). The majority of the structures are high-rise buildings, ranging from 20 to 30 floors in height, with a standard design lifespan of 50 to 70 years (Baldwin et al., 2018).

Before 1986, China had no energy-saving requirements for residential buildings, and the residential buildings constructed in this period typically lacked both thermal insulation and airtight construction (Zhang et al., 2014). In 1986, the first building energy efficiency standard, i.e. the energy-saving Standard for Civil Buildings (JGJ26-1986), was published by the central government of China, which required that residential buildings should save 30% energy compared with buildings constructed in the early 1980s (Lang, 2004). The energy-saving standards for residential buildings in severe cold and cold regions were subsequently raised to 50% in 1995, 65% in 2010, and 75% in 2018 (Yu et al., 2019). However, due to limited financial support and low environmental protection awareness, most residential buildings were still built without insulation in the 1990s (Peng et al., 2021). Even some newly constructed residential buildings in recent years have struggled to meet these standards (Ma et al., 2019). According to Liu et al. (2023), more than 90% of the existing residential buildings in China still demonstrate substandard energy performance.

2.1.2 Progress and challenges of residential energy retrofits

To support China’s goals of peaking carbon emissions by 2030 and achieving carbon neutrality by 2060, the Chinese government launched large-scale energy retrofit programmes in 2007, specifically targeting ageing residential communities to improve the energy performance of existing buildings and accelerate their low-carbon transformation. By 2019, the Chinese government had retrofitted a total of 514 million square metres of existing residential area, involving 21 million households (Guo et al., 2024). However, compared to the vast stock of existing residential buildings, current

retrofit projects remain limited in scale. According to the statistics of the Ministry of Housing and Urban-Rural Development of China, more than 170,000 ageing residential communities in China are still in need of retrofitting, involving more than 100 million residents (Wang et al., 2019b). This highlights the significant gap between ongoing retrofit practices and the urgent need for a large-scale, low-carbon transformation of Chinese residential building stock.

Currently, the primary energy retrofit measures for residential buildings in China include improving the insulation of building envelopes, upgrading windows and doors, optimising HVAC systems, and guiding occupant energy-use behaviours (Huang and Xu, 2024). Enhancing the thermal performance of building envelopes, such as external walls and roofs, represents a fundamental measure in achieving deep energy retrofits of ageing residential buildings (El-Darwish and Gomaa, 2017). Adding insulation layers to external walls and roofs significantly reduces heat transfer losses in winter and cooling loads in summer. A study conducted in Shanghai, a typical city within China's hot-summer and cold-winter climate zone, demonstrated that applying insulation to the walls of older residential buildings effectively reduced both heating and cooling energy consumption, particularly resulting in significant savings for heating (Liu et al., 2023). Therefore, envelope insulation is generally prioritised in energy retrofit strategies due to its notable impact on reducing HVAC loads. Additionally, windows represent one of the weakest thermal components of building envelopes. Older residential buildings commonly use single-pane glass with ordinary frames, causing substantial air leakage and heat loss. Upgrading windows to high-performance alternatives can significantly reduce heat loss during winter and improve air-conditioning efficiency during summer (Li and Wu, 2025). For instance, replacing single-pane and double-pane windows with Semi-transparent photovoltaic (STPV) windows can save up to 18% and 16% energy consumption in Hong Kong (Zhang et al., 2016). Beyond reducing building envelope loads, retrofitting and optimising HVAC systems themselves can further enhance energy efficiency (Liao et al., 2025). For residential buildings with centralised heating systems in northern China, improving the efficiency of heating systems can be achieved by upgrading to

high-efficiency boilers or heat sources, optimising secondary network balancing, and installing indoor thermostatic valves and household heat-metering systems (Liu et al., 2020). Moreover, studies indicate that promoting energy-saving occupant behaviours, such as properly adjusting air conditioning temperature setpoints and promptly turning off unnecessary lighting and appliances, can achieve energy-savings of approximately 10% to 25% without additional investment in Beijing (Zhang et al., 2020). Numerous studies highlight occupant behaviour as a critical factor affecting building energy consumption and indoor environmental quality, noting that neglecting occupant behaviour can lead to a performance gap between expected and actual retrofit outcomes (Maghsoudi Nia et al., 2022). Therefore, energy retrofit projects increasingly emphasise occupant engagement and behavioural interventions, including soliciting occupant opinions during the design phase, providing user guidance during implementation and operation stages, and offering energy-use feedback to enhance occupants' awareness and motivation. Some studies have also indicated that early involvement of occupants in retrofit planning and enhanced communication can effectively reduce implementation resistance and improve overall retrofit energy performance (Maghsoudi Nia et al., 2022, Lang et al., 2021).

Although retrofit techniques and policy support for residential building energy retrofits in urban areas continue to improve, significant challenges still arise during practical implementation. The first step in making retrofit decisions involves assessing the energy performance and energy-saving potential of existing buildings (Fina et al., 2019). Systematically evaluating and comparing the energy-saving potentials of various retrofit strategies can assist decision-makers in scientifically selecting optimal retrofit solutions (Zheng et al., 2019). However, in practice, decisions regarding building retrofit strategies often heavily rely on the experience and intuition of experts (Xue et al., 2022). Due to limitations in practitioners' knowledge and experience, the trade-offs between multiple retrofit objectives may not be optimally identified, potentially leading to suboptimal retrofit solutions (Ma et al., 2023). Additionally, while numerous building energy modelling tools have been developed, their use typically requires extensive professional knowledge, significant computational resources, and

labour-intensive data collection (Dino et al., 2020). In reality, many ageing residential communities lack detailed design documents, and technical personnel frequently have insufficient understanding of the existing building materials, structures, and equipment performance. Thus, developing efficient assessment methods to quickly formulate customised retrofit plans for ageing residential communities remains a critical technological challenge. Furthermore, economic considerations play a central role in decisions regarding residential energy retrofits. Retrofit projects typically require substantial initial investments, whereas the savings in energy expenses accumulate slowly over several years. This time-value mismatch often causes residents to hesitate due to uncertainties regarding financial returns, significantly undermining their motivation to participate in energy retrofit initiatives.

2.2 Automated Residential Floor Plan Analysis

Automated building floor plan analysis has been studied across multiple research domains. Targeting the specific goal of energy assessment, this study narrows the scope to automated generation of Building Energy Models (BEMs) directly from raw floor plan data. Accordingly, this section takes 2D floor plans as the entry point and reviews methods for automated spatial reconstruction and semantic information extraction. This section concentrates on two fundamentals: (1) space reconstruction techniques that recover accurate architectural geometry and topology, and (2) room function prediction that enables assigning energy usage patterns and schedules to the reconstructed spaces.

2.2.1 Space reconstruction from 2D floor plan

Space reconstruction involves segmenting floor plan images to identify architectural components such as walls, doors, and windows. Traditional approaches rely on rule-based pattern recognition, where predefined rules are used to classify pixels into different building elements. For instance, Gimenez et al. (2016) examined all

segments and arcs in the building floor plan to classify them into walls and openings with pre-defined rules. Or et al. (2005) employed the floor plan image after vectorisation to distinguish wall pixels, and located windows and doors through thin boxes and arcs detection. Ahmed et al. (2011) used a 3 × 3 square mask for floor plan image erosion and dilation to extract external walls. A thick/thin line separation algorithm was then deployed to identify internal walls, windows and doors. Feltes et al. (2014) further improved edge detection by incorporating corner recognition, enabling their method to handle rotated and diagonal walls. These rule-based approaches generally achieve high classification accuracy when applied to floor plans that strictly follow specific annotation styles, such as hatched walls or arc-shaped doors. However, their applicability is limited and lacks robustness. When analysing floor plan images of different styles, scales, or levels of complexity, these methods often require extensive manual pre-processing and fine-tuning parameters. For example, floor plans with different image sizes may require different kernel sizes for geometric feature extraction, which often involves repeated trial-and-error to find the optimal parameter settings. As the complexity of floor plan layouts increases, so does the number of pixels and geometric elements that need to be analysed, leading to increased computational cost. The sensitivity to annotation styles and demanding computation time makes such rule-based approaches less desirable for large-scale automated energy modelling workflows.

With the development of artificial intelligence, machine learning-based approaches become popular in addressing floor plan segmentation problems. By learning the hidden complex relationships between the raw input floor plan image and manually drawn ground truth, the machine learning model can classify pixels into specific architectural categories. De Las Heras et al. (2014) split the floor plan image into overlapped squared patches and developed a K-means clustering model to classify pixel patches into walls, windows and doors. Liu et al. (2017) employed the Convolutional Neural Network (CNN) to detect junction points in the building plan drawings and used the Integer Programming algorithm to aggregate junctions into wall primitives. Cho et al. (2020) adopted style-transfer algorithms based on

Conditional Generative Adversarial Networks (CGAN) to extract the wall, door, and window pixels; their network also performed well on the hand-drawn diagrams. Zeng et al. (2019) designed a room-boundary guided multi-task network. They defined several convolution kernels to recognise different wall shapes, including those with opening pixels from sophisticated building plans with curve boundaries. Chang et al. (2025) introduced a key point-based detection approach (CPN-Floor) for identifying walls, doors, windows, and scale indicators by generating and filtering candidate primitives through axial alignment rules and geometric constraints, achieving 87% precision and 88% recall. Kratochvila et al. (2024) designed two multi-scale pixel segmentation models based on the MDA-Unet and MACU-Net architectures with improved skip connections and attention mechanisms for multi-household floor plan analysis. In addition to recognising standard architectural elements such as walls, windows, and doors, their models also enable the identification of more complex features, including glass walls, sliding doors, railings, and stairs.

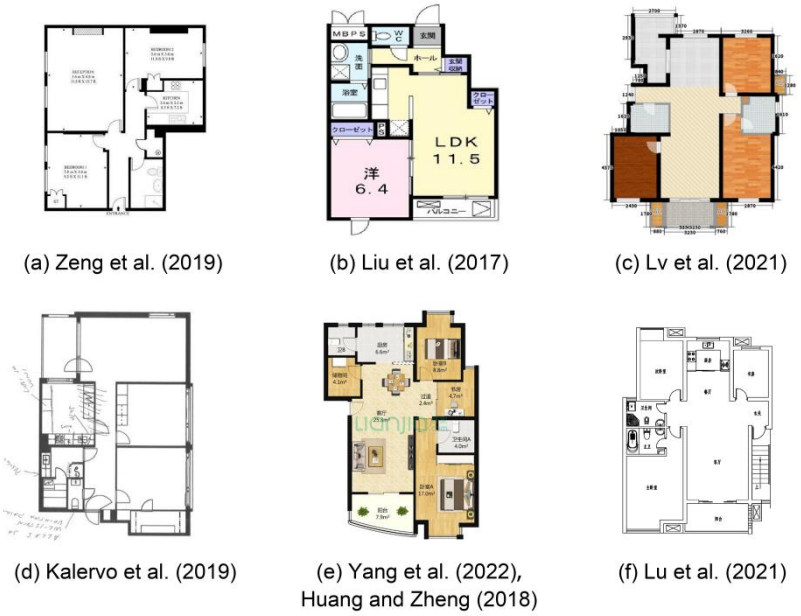
Unlike traditional rule-based approaches, machine learning models extract deep colour and texture features from floor plan images, learning the hidden complex relationships between raw input images and manually annotated ground truth, which enables the classification of each pixel into specific architectural categories. By including floor plan samples with varied annotation styles in the training dataset, machine learning models can generalise across multiple drawing styles, significantly improving the robustness of architectural element recognition. More importantly, machine learning models greatly accelerate the image segmentation process. For example, Kratochvila et al. (2024) evaluated different machine learning models for floor plan segmentation using a dataset of 1,453 images, reporting a total processing time ranging from 95 to 499 seconds, with an average of just 0.06 to 0.34 seconds per image. This rapid processing capability facilitates large-scale spatial reconstruction of architectural floor plans, laying the foundation for automated BEM construction.

2.2.2 Room function classification

Machine-learning-based floor plan segmentation models have also been widely applied to room semantic function classification. Huang and Zheng (2018) annotated floor plans into seven room types and used the Generative Adversarial Networks (GAN) to classify room functions. The network accurately identifies well-defined spaces, such as kitchens and bathrooms, while showing uncertainty in open areas such as living rooms and walkways. Lu et al. (2021) and Ahmed et al. (2012) assigned room functions through textual information detection. This approach is limited when the textual information is not available in the floor plan image. Lv et al. (2021) proposed a multi-model for room-type classification by combining semantic segmentation, symbol detection, and text recognition through a voting mechanism. Their method significantly improves the robustness of functional room labelling through a multi-source fusion strategy. Yang et al. (2022) proposed a three-stage pipeline that significantly improves the representation of indoor building spaces by integrating deep learning and optimisation techniques. Their method first uses a modified CNN for hierarchical semantic segmentation to identify basic geometric elements like corners, edges, and room areas. A mixed-integer programming (MIP) approach then assembles these elements into a planar graph, incorporating both spatial cues and structural constraints. Finally, a polygonal coordinate descent strategy refines the topology, producing accurate and consistent vectorised representations. Liu et al. (2017) present a hybrid method combining deep learning and Integer Programming (IP) to convert raster floor plan images into vectorised representations. Their approach infers room functions by first predicting per-pixel semantic maps for 11 room types using a CNN, then refining these predictions through IP that enforces spatial and topological consistency. Kalervo et al. (2019) proposed the CubiCasa5K dataset and an improved multi-task CNN model to enhance room-type classification in floor plan images by combining semantic segmentation with key point detection and using an uncertainty-weighted loss. Zeng et al. (2019) proposed a room boundary-guided Deep Multi-Task Network (DMTN) to

improve room-type prediction accuracy. They designed four direction-aware convolutional kernels to better capture diverse room boundary shapes, and a cross-and-within-task weighted loss function was introduced to effectively balance the learning of both room boundary detection and room type classification tasks.

Similar to the segmentation of architectural elements, room function prediction also relies on learning deep colour or texture features from floor plan images to predict the room type associated with each pixel (O’ Mahony et al., 2020). The difference is that existing research in room function prediction primarily relies on features such as furniture symbols, floor texture patterns, or colour schemes within the floor plan. Fig. 1 presents floor plan datasets of different styles used for training pixel-based room function classification models. In Fig. 1(b), (c), and (e), room categories are visually distinguished by colour-coded regions, and both (c) and (e) further include distinct floor texture patterns across different room types. Meanwhile, Fig. 1 (a), (d), and (f) contain furniture layouts or textual labels indicating room functions. These examples reflect the reliance of pixel-based models on explicit visual cues in the training dataset. However, room areas that lack such indicators, such as the empty spaces in (a) and (d) where no furniture is present, often exhibit significant local noise in the segmentation results, which leads to misclassification.



(a) Zeng et al. (2019) (b) Liu et al. (2017) (c) Lv et al. (2021)

(d) Kalervo et al. (2019) (e) Yang et al. (2022), Huang and Zheng (2018) (f) Lu et al. (2021)

Fig. 1 Single-household floor plan examples

Object-based room type classification predicts room functions by quantified room features. Mewada et al. (2020) extracted the room's geometrical features such as room area and aspect ratio to identify a closed room polygon. Bloch and Sacks (2018) collected room features, including area, number of doors and windows adjacent to the room, and number of boundary lines, and used the Artificial Neural Network (ANN) to classify room functions into 15 types. Buyuklieva (2020) calculated the adjacency matrix of rooms within a floor plan and employed Visual Graph Analysis (VGA) measurement to distinguish room types. Wang et al. (2022) created a graph representation of the room layout, where rooms were defined as nodes and the connections between rooms were defined as edges. The Graph Neural Networks (GNN) was then used to classify room functions. Similarly, Dong et al. (2021a) propose a GAN-based framework that first segments floor plans into vectorised representations and then predicts room functions using a GNN by analysing room features like area, window count, and inter-room connectivity. Zhao et al. (2022) collected both room geometrical features and connectivity relationships to classify room types.

In brief, pixel-based prediction methods that require specific visual features, such as detailed furniture layouts or colour schemes within floor plans, could hinder the applicability. For instance, during the early design stages, furniture arrangements are often not included in the floor plan. Object-based prediction demonstrates greater potential by leveraging geometric features and topological relationships of rooms for function classification. However, existing studies primarily focus on single-household floor plans, where the number of rooms is limited and spatial connectivity is relatively straightforward. In contrast, the complexity of spatial configurations can grow considerably in multi-household residential layouts, which remain under-investigated (Pizarro et al., 2022). In multi-household layouts, household units are often organised around shared circulation spaces, such as corridors and staircases. The presence of such circulation spaces poses additional challenges in room type classification, particularly at the household boundary detection and shared communal spaces recognition.

In summary, current research on automated geometric reconstruction and room semantic extraction largely focuses on simple single-household floor plans and lacks investigations of complex multi-household layouts. In multi-household plans, both the number of architectural components and the connectivity among rooms are far more complex than in single-household cases, which greatly increases the difficulty of automated analysis. In addition, existing studies rarely include the step that links recognised information with energy modelling; how to use the extracted geometric and semantic information for building energy model construction remains a challenge. This includes, for example, defining rules for aggregating rooms into thermal zones, and establishing mappings from room functions to occupant density, lighting or equipment power density, and operating schedules.

2.3 Current Building Energy Performance Evaluation Methods

In the process of building energy retrofitting, reliable energy performance evaluation models are essential for diagnosing the baseline energy use of existing buildings, evaluating the energy-saving potential of retrofit measures, and verifying post-retrofit energy performance. At present, modelling approaches for building energy prediction have generally been categorised into three main types: (1) physical energy models, also known as white-box models, which simulate energy consumption based on thermodynamic principles; (2) data-driven models, or black-box models, which predict energy use based on building feature data using machine learning (ML) or deep learning (DL) techniques; and (3) hybrid models, or grey-box models, which combine elements of both physical and data-driven approaches.

2.3.1 Physical energy models

Physical energy models, also known as the engineering method or white-box models, calculate building energy consumption through precise physical heat transfer equations (Chen et al., 2022). Commonly used software tools include EnergyPlus,

eQuest, Trnsys, Esp-r, IES, and DeST. These commercial or open-source simulation platforms are equipped with comprehensive physical modelling engines, allowing users to efficiently construct and simulate BEMs. With adequate input building information such as the thermal properties of building envelopes, HVAC system configurations and control strategies, internal loads and occupancy schedules, spatial zoning, geographic location, and weather data, physical models can perform accurate building energy simulations.

Numerous studies have examined building energy consumption using various simulation models or forecasting engines. For example, Macías et al. (2018) used the Transfer Function Method (TFM) to evaluate energy-saving potential of residential buildings in Mexico. The building data consisting of household size and HVAC unit capacity were collected from 300 local houses. The results presented that almost 40% of households used oversized air conditioning equipment, and the replacement of old low-efficiency air conditioning equipment can lead to an annual energy consumption saving of 32%. Shen et al. (2019) performed a parametric energy simulation using Python and EnergyPlus. The model is validated to predict the future hourly energy consumption based on EnergyPlus simulation results. Pasichnyi et al. (2019) employed DesignBuilder based on the EnergyPlus simulation engine to study the energy-saving potential in terms of heat recovery ventilation, energy-efficient windows, and a combination of these in Stockholm. Physical energy models are also usually combined with optimisation methods to find the appropriate retrofit option. Zangheri et al. (2018) used EnergyPlus and multi-stage optimisation techniques to identify the most cost-optimal retrofit package to reach nearly zero-energy building (NZEB) in different climate zones. The study selected four different building typologies of 60s-70s from ten climate areas in the European Union. It was found that the cost-optimal scenarios were highly associated with the initial investment cost, and the energy efficiency potential was substantial between 36% and 88%. Piira et al. (2022) combined the physical energy simulator and the genetic algorithm (GA) to select the most appropriate building retrofitting package. They first collected real-time energy consumption from the actual building operation. The baseline model was then

developed through utilising a GA to find the building parameter set that results in the best match between simulated and measured energy consumption. The gene in the GA represented the building characteristics and set according to different retrofitting packages, which was iteratively calculated to reach the optimal set. He et al. (2020) developed a deterministic decision-making mechanism to find the optimal retrofit options of high-rise residential buildings in China. The Energy Performance Calculator (EPC) was employed to assess the building thermal performance and the net present value (NPV) was employed to evaluate the retrofit cost. The study found that the identified optimal retrofit options can achieve a maximum energy-saving of 60% compared with previous annual consumption.

Physical energy models calculate energy consumption based on thermodynamic principles, making the relationship between inputs and outputs transparent and interpretable. The model input parameters represent actual physical meanings, which makes it easy to adjust and analyse the impact of various retrofit measures. For example, they can be used to evaluate energy consumption differences under different envelope materials, equipment configurations, and operational conditions, as well as explaining the reasons behind those differences. Moreover, with detailed building information as inputs, physical simulations can provide accurate energy performance results. However, the process of collecting detailed building information such as material properties, HVAC system configurations, and operation schedules, along with the modelling process, is a labour-intensive task. Meanwhile, such complete building information is often unavailable, especially for ageing residential communities, due to missing design documentation, degradation of building components, and inconsistencies in documentation management systems (Feng et al., 2022). In addition, physical simulation involves high computational complexity, making the evaluation of large-scale retrofit scenarios extremely time-consuming, which hinders the ability to quickly compare alternative solutions in practical projects. As a result, physical models are more suitable for qualitative assessment and detailed analysis during the design phase, rather than for scenarios that require rapid iteration and real-time energy prediction (Chen et al., 2025).

2.3.2 Data-driven models

Different from the physical energy models, data-driven models (also called the black-box model) predict building energy performance by learning the hidden patterns between input building features and target energy variables (Runge and Zmeureanu, 2019). Compared with physical models, data-driven models usually do not require detailed physical parameter inputs. Even when comprehensive building information is unavailable, data-driven approaches can establish reliable prediction models using existing energy consumption data and partial building characteristics. Based on the algorithms applied, data-driven approaches can be further classified into traditional ML models and advanced DL models. Traditional ML methods include algorithms such as Support Vector Machines (SVM), Random Forests (RF), Decision Trees (DT), and Multiple Linear Regression (MLR). In contrast, DL models utilise emerging neural network architectures including Multi-Layer Perceptron (MLP), CNN, Recurrent Neural Networks (RNN/LSTM), and GNN.

Data-driven models have been widely applied in various research studies for predicting building energy performance under different feature combinations. For example, Dong et al. (2021b) collected hourly electricity consumption data over a full year from an office building and developed an hourly energy consumption prediction framework based on time-related variables and meteorological factors, such as dew point temperature, air humidity, and air pressure. The operational conditions of the case office building were categorised into four distinct energy consumption patterns using a decision tree. For each pattern, an individual stacking ensemble model was trained independently, combining Support Vector Regression (SVR), ANN, and MLR. Experimental results showed that even with only 60% of the available training data (4,991 hourly records), the proposed method achieved a coefficient of variation of the root mean squared error (CVRMSE) of 15.4%. Seo et al. (2022) collected data from 361 residential buildings and selected seven key input features to predict annual building energy use, including building geometry, envelope thermal characteristics. They compared three data-driven models, namely MLR, SVR, and ANN. The results

showed that the ANN model outperformed the others, achieving the lowest root mean squared error (RMSE) of 23.4 kWh/m^2 , and the fastest prediction speed, completing 10,000 case predictions in just 0.215 seconds. Olu-Ajayi et al. (2022) conducted a comprehensive comparative analysis of nine widely used models using a large dataset of 5,000 residential buildings in the UK. They extracted 22 input features to predict the annual energy consumption of each building. Results showed that the Deep Neural Network (DNN) achieved the highest prediction accuracy (RMSE = 1.16 kWh/m^2), outperforming conventional models such as Linear Regression (LR) and DT, particularly under large-scale data conditions. Ibrahim et al. (2022) employed MLR and MLP to study the effect of eight building variables on heating and cooling load, including building size, floor height, glazing area, wall area, window to wall ratio (WWR), U-value of glazing, U-value of roofs and external walls. Comparative results showed that the MLP consistently outperformed MLR, achieving higher predictive accuracy with lower MAE and RMSE values. Kim and Cho (2019) proposed a DL model that combines CNN and LSTM to predict household electricity load and other forms of energy consumption. The model takes as input hourly historical electricity consumption time series for each household, along with related influencing factors such as voltage and current, meter readings from different rooms, and calendar information. The CNN subnetwork automatically extracts feature relationships among various variables affecting energy use, while the LSTM subnetwork captures the temporal dynamics of electricity consumption over time. This hybrid approach integrates spatial feature extraction with time-series modelling, enabling more comprehensive pattern discovery from electricity usage data. Moveh et al. (2025) proposed a Temporal Graph Neural Network (TGNN) framework that combines GNN and LSTM networks, and a weather integration module to forecast energy consumption across 150 commercial buildings. The model captures spatial relationships through a weighted adjacency matrix based on building proximity and similarity, while temporal dynamics are learned via LSTM layers. Weather data are processed through a separate neural branch and fused with spatial-temporal features using a multi-head attention mechanism. The framework achieves high accuracy

(mean absolute percentage error (MAPE) 3.2% for 15-minute forecasts) and maintains robust performance under seasonal variation.

In recent years, DL models have made significant progress in the field of building energy consumption prediction. Compared to traditional ML approaches, DL leverages multi-layer neural networks to automatically extract features, thereby greatly reducing the need for manual feature engineering. These models are capable of capturing complex nonlinear relationships and temporal patterns in energy usage, which enhances their predictive performance and robustness. Numerous studies have demonstrated that DL models generally outperform classical ML models in terms of both accuracy and computational efficiency, particularly in scenarios requiring rapid assessment of large-scale retrofit options (Dong et al., 2021b, Seo et al., 2022, Olu-Ajayi et al., 2022, Ibrahim et al., 2022). Notably, DL models exhibit strong generalisation and real-time prediction capabilities, and they are less dependent on fine-grained building detail inputs, making them well-suited for evaluating a wide range of retrofit configurations. However, DL methods are inherently data-intensive and require large volumes of high-quality energy performance data for effective training. When applied to atypical buildings or conditions that fall outside the training distribution, their predictions may suffer from reduced accuracy (Chen et al., 2022). Furthermore, due to their complex internal structures and black-box nature, DL models lack interpretability, making it difficult to trace how individual input parameters contribute to final energy predictions. This limitation poses challenges for practical decision-making in energy retrofit planning, where transparency and explainability are often critical.

2.3.3 Hybrid models

Hybrid models, also called grey-box models, are based on the coupling of physical energy models and data-driven models, combining the strengths of both approaches (Li et al., 2021). For example, Koponen et al. (2019) used the physical model to forecast control responses and load saturation and used the ML model to predict the

residue term. The final load forecast was comprised of the sum of these component forecasts. (Amasyali and El-Gohary, 2022) developed a hybrid model for the prediction of energy use based on occupant behaviour. This model included two parts: (1) the creation of ML models that predicted the impacts of both climate and occupant behaviour; and (2) a hybrid model that predicted hourly building cooling energy consumption according to the results from the ML models. Their model can achieve a prediction accuracy of hourly cooling energy consumption of 91.93%. Another typical way is to utilise a physical-based simulator to generate the dataset for a data-driven model. Płoszaj-Mazurek et al. (2020) used Grasshopper to create a parametric model for simulating the BEP using Ladybug and Honeybee. A total of 1500 cases were generated parametrically for simulation, and the results were then used to train the Gradient Boosting Regressor (GBR) model for the energy consumption prediction. Li and Yao (2021) simulated building cooling and heating energy consumption through Urban Modelling Interface (UMI) and combined it with building detailed characteristics to create the dataset. They trained ten ML models with the generated dataset to predict the intensity of energy use and compared their performance. The polynomial kernel SVR performed best in predicting heating and cooling energy use with 92.7% and 95.7% accuracy. Imani et al. (2025) presented a hybrid framework combining parametric physical simulation and data-driven analysis to optimise retrofitting strategies across UK climate zones. Using Grasshopper and the TT Toolbox, they generated 3,808 retrofit scenarios varying in window-to-wall ratios, insulation, and heating systems. Each scenario was simulated using EnergyPlus, and the resulting Energy Use Intensity (EUI) and thermal comfort indicators (PMV) were analysed to assess performance. This comprehensive dataset enabled a data-driven comparison of retrofitting options, revealing that air-source heat pumps (ASHPs) were most effective in mixed and cool-humid climates, while water-source heat pumps (WSHPs) performed better in colder regions. To examine the influence of retrofit initiatives on multi-scale energy consumption, Nutkiewicz et al. (2021) simulated the energy consumption of 29 buildings through EnergyPlus with historical hourly weather data and used simulation results to train a LSTM model. The study found that taking the

urban environment into account, the effects of retrofits on the energy consumption of individual buildings and surrounding buildings could increase by 7.4%. Nutkiewicz et al. (2018) proposed a hybrid model called DUE-S to predict multi-scale urban building energy consumption. They created several urban buildings and simulated the energy consumption with a time series of 15-minute intervals. The results showed that the DUE-S model can accurately predict urban-scale energy consumption at hourly, daily, and monthly intervals. Kannari et al. (2021) trained an ANN model with data from 11700 simulations. The model is forced to learn the energy use pattern from different building types to approximate the physical simulator, and finally achieved an average accuracy in predicting building energy demand of 0.95 and performed a good generalisation in office, apartment, and detached house buildings. Truong et al. (2021) proposed a comprehensive framework that employs a fine-tuned deep ANN model to predict the hourly energy consumption of residential buildings based on occupancy rate and seasonality. Due to the lack of high-resolution real-world energy usage data, the authors developed a synthetic load profile generator that simulates household energy consumption behaviour under constraint conditions, incorporating different household types (with 1 to 5 occupants), appliance usage patterns, and daily schedules. This generator produced a training dataset of 100,000 hourly samples, each containing season, number of occupants, hour of the day, and corresponding energy consumption. The energy consumption was computed using a combination of deterministic appliance power models and stochastic noise. Results showed that the proposed deep ANN model achieved an RMSE of 111W and R^2 of 97.5%.

In summary, physics-based energy models, data-driven models, and hybrid modelling models each offer distinct advantages in the task of building energy consumption prediction. Physics-based models, grounded in thermodynamic principles, building construction, and operational boundary conditions, can produce accurate energy simulations. However, their high modelling and simulation costs, along with the complexity of parameter acquisition, make them impractical for the rapid evaluation of large-scale retrofit scenarios. Data-driven models, particularly DL models, demonstrate strong capabilities in automatic feature extraction, rapid prediction, and

the capture of complex energy usage patterns. These characteristics make them well suited for evaluating building energy performance in large-scale retrofit projects. Nevertheless, the effectiveness of DL models relies heavily on the availability of large volumes of high-quality data, and their interpretability remains limited. Hybrid models address some of these shortcomings by combining advantages of both physical and data-driven models, but their complex integration and limited generalisability pose challenges. An emerging and powerful strategy uses simulation platforms such as EnergyPlus or TRNSYS to conduct large scale parametric studies that generate synthetic energy consumption datasets covering a wide range of building types and retrofit strategies. These datasets are then used to train DL models, ensuring high quality training data without extensive field measurements and allowing the models to internalise the thermal principles embedded in the physical simulations. By uniting the precision of physics-based simulation with the efficiency of DL, this strategy is particularly well suited to rapid evaluation and optimisation of many retrofit options, and it has become a major research trend in building energy performance evaluation.

2.4 Current Applications of Deep Learning in Residential Building Energy Retrofit

Deep learning has been widely adopted for building energy performance assessment in the field of building energy retrofits, with numerous model architectures developed to address different building types and forecasting requirements. In light of the diverse geometries, varied occupant behaviours, and complex thermal interactions between adjacent units found in China’s multi-household residential buildings, the following review examines existing DL models from three perspectives: building geometrical feature extraction, occupant behaviour modelling, and inter-unit thermal interaction encoding.

2.4.1 Building geometrical feature extraction

The geometrical characteristics of residential buildings, such as orientation, building shape, envelope configuration, and floor plan layout, have a direct impact on energy performance (Kistelegdi et al., 2022). Existing studies often rely on numerical features to describe building geometry, such as shape factor, floor area, window-to-wall ratio, and insulation thickness (Ibrahim et al., 2022, Seo et al., 2022, Olu-Ajayi et al., 2022). While these parameters can partially reflect the building geometrical properties, they fall short in fully capturing its two-dimensional spatial configuration. For example, two buildings with the same floor area but distinctly different shapes or internal room layout may exhibit significantly different energy consumption behaviours (Uddin et al., 2022).

To more effectively describe building geometry features, some studies have explored the potential of CNNs to learn 2D building spatial representations or 3D features. Singh and Smith (2023) proposed a 3D CNN framework that utilises voxelised massing models ($100 \times 100 \times 24$ grids) as input to capture complex building shape characteristics. The 3D building voxel input is passed through several convolutional and average pooling layers to extract deep geometrical features. In addition, 21 numerical features consist of envelope thermal properties, orientation, internal loads, and HVAC efficiencies are concatenated to predict building EUI. The proposed 3DCNN achieved a RMSE of 0.77 kWh/m^2 • year using 10,000 samples, demonstrating its ability to accurately learn the nonlinear interactions between building geometry and energy performance. Mahan Singh and Geyer (2021) proposed a CNN-based model that uses 50×50 pixel grayscale floor plan images as input to extract spatial features such as self-shading and building compactness. The image is processed through multiple convolution and pooling layers, then flattened and concatenated with other numerical parameters (e.g., floor height, window-to-wall ratio, heating/cooling performance coefficients) to predict the annual EUI. The model was trained on 5,000 medium-sized office building variants, and its performance was evaluated on an additional 5,000 unseen test samples. Results showed that the CNN

model achieved a MAPE of 1.51% and a RMSE of 1.06 kWh/m^2 • year, indicating strong generalisation capability in capturing shape-related features for energy prediction. Streltsov et al. (2020) proposed a novel three-stage framework that leverages high-resolution aerial imagery to estimate residential energy use across large urban areas. Their method involves (1) segmenting building footprints using a U-Net CNN, (2) classifying buildings as residential or commercial via a ResNet-152 classifier, and (3) extracting features such as building area, perimeter, local density, and deep semantic features for energy prediction using a RF regressor. Evaluated on two U.S. cities, Gainesville and San Diego, the model achieved R^2 values of 0.28 and 0.38 at the individual building level, with performance improving significantly under spatial aggregation (R^2 = 0.91 at the 200 m × 200 m neighbourhood level). Notably, their study demonstrated that simple geometric attributes like building footprint area derived from remote sensing imagery can outperform more complex models reliant on socio-demographic data. Sheng et al. (2022) developed a deep multi-modal learning framework for predicting residential building energy use by integrating textual data from EPC and visual data from Google Street View (GSV) images. The study employed a dual-stream neural architecture combining MLP for structured EPC attributes and CNN for street-view images, fusing them at an intermediate layer to enhance prediction accuracy. Applied to over 5,000 properties in the UK, the multi-modal model achieved a low mean absolute error of 0.01 kWh/m^2 • year. This research highlights the potential of facade imagery in capturing building geometrical characteristics.

Overall, for DL based building energy consumption prediction tasks, CNNs are commonly used to extract local geometric features (such as spatial configuration, window or door distributions) from input images. Through multiple convolutional layers, these local features are gradually aggregated into deep geometric representations. Building geometry is typically expressed using two-dimensional inputs, such as floor plans, elevations, and contour maps, or through three-dimensional representations like voxel models. Floor plans primarily capture the spatial layout of a building (e.g., room distribution), while elevation drawings emphasise facade features such as

window-to-wall ratios and envelope materials. Contour maps (e.g., projections from satellite imagery) highlight the external shape of buildings. Three-dimensional voxel models discretise building volumes into a grid structure suitable for input into 3D CNN, enabling the capture of height and volumetric distribution. However, due to the variable floor dimensions and heights of residential buildings, it is challenging to standardise 3D inputs. Moreover, such models significantly increase the computational burden. In contrast, two-dimensional image inputs processed by CNNs provide an efficient way to supplement traditional parametric features with rich geometric shape information. This approach is particularly useful for capturing the relationships between design variables, such as floor plan configuration and spatial layout, and energy performance. Notably, since non-geometric factors like building materials, equipment power density, and occupant behaviour also have a significant influence on energy consumption, image-based geometric features alone are insufficient to capture semantic patterns. Therefore, current research commonly integrates semantic features with image-derived representations through feature concatenation strategies, in order to improve prediction accuracy.

2.4.2 Occupant behaviour modelling

Energy consumption in residential buildings is highly dependent on occupant behaviour patterns, including HVAC settings, electricity usage habits, and daily routines such as occupancy schedules and travel patterns. Behavioural differences across individual households often result in significant temporal fluctuations in energy demand, making occupant behaviour one of the primary sources of uncertainty in evaluating retrofit effectiveness. For example, some studies have shown that the energy consumption of the same building can vary by as much as a factor of seven under different occupant behaviour scenarios (Amasyali and El-Gohary, 2021). Therefore, accurately representing and incorporating occupant behaviour features in DL models is essential for improving both prediction accuracy and generalisability.

Common occupant behaviour representation methods include state vector sequences,

probabilistic occupancy matrices, and activity category encoding. State vector sequences explicitly indicate the occupancy status or number of occupants for each time interval in a time-series format, typically recorded on an hourly or finer scale (Qaisar et al., 2023). These sequences reflect whether a space is occupied and how many people are present at a given time. In DL models, daily occupancy schedules can be directly input as feature vectors, such as a 24-hour binary vector representing daily occupancy states. Alternatively, distinct encoding schemes can be used for weekdays and weekends to capture periodic behavioural patterns. Probabilistic occupancy matrices are used to express the uncertainty of occupant behaviour. For example, a 7×24 matrix can represent the probability of occupancy for each hour across a week. Such matrices are often generated using stochastic models like Markov chains, which assign a probability value to each hour indicating the likelihood that a space is occupied (Banfi et al., 2024). This approach better captures the randomness and variability of human behaviour. Activity category encoding involves classifying occupant behaviour into specific categories (e.g., "sleeping", "cooking", "working") and encoding them either through one-hot vectors or numerical representations (Banfi et al., 2024). These activity types are treated as discrete features, helping the model distinguish energy use patterns associated with different behavioural states. In addition to commonly used behaviour modelling, Bansal et al. (2025) proposed a sinusoidal encoding scheme that transforms periodic features like hour-of-day or day-of-week into continuous representations using sine and cosine functions. This method preserves the circular structure of time and enables the model to more accurately capture seasonality and periodicity in energy consumption. Empirical results demonstrated that integrating sinusoidal encoding into ensemble learning models (e.g., XGBoost, LightGBM) led to substantial improvements in prediction accuracy. Notably, LightGBM with sinusoidal encoding achieved an RMSE of 0.3983 and an R^2 score of 0.8356, outperforming models using traditional encodings by up to 12.6% in error reduction. Sinusoidal encoding effectively addresses the challenge of conventional approaches that often introduce artificial discontinuities at period boundaries, such as the transition from hour 23 to hour 0,

which hinders the model's ability to learn smooth temporal patterns.

Occupant behaviour in residential buildings typically exhibits clear periodic patterns alongside random fluctuations. Therefore, DL models must be capable of capturing both the temporal dependencies and cyclic variations inherent in user behaviour. Gated Recurrent Units (GRUs) and LSTM networks are widely used DL architectures for time series modelling. These models are designed to process sequential data by maintaining temporal dependencies through recurrent connections. Among them, LSTM networks are popular in building energy prediction tasks, owing to their gated memory cells that effectively capture and retain long-term dependencies. For instance, Anan et al. (2024) proposed an occupant-aware energy consumption prediction framework based on a Long Short-Term Memory (LSTM) neural network. The study utilised a real-world dataset containing minute-level electricity consumption data from various subsystems (HVAC, lighting, plug loads, and elevators), along with outdoor temperature and occupancy information. The data were resampled into multiple temporal resolutions — hourly, 3-hourly, daily, and quarterly, to support multi-scale prediction. To capture temporal dependencies, the authors constructed time-lagged input sequences (e.g., 1-hour, 24-hour, 48-hour, and 168-hour lags) as input to the LSTM model. The proposed model achieved its best performance on the 1-hour lag configuration, with a RMSE of 0.2198 and a R^2 of 0.9964. Baek et al. (2022) proposed an enhanced LSTM based framework for community energy consumption prediction that leverages shared electricity usage data among buildings within a cluster. To capture implicit occupant behaviour patterns, the model incorporates temporal electricity consumption profiles from seven additional buildings in the same community as part of the input. The time series data were organised using a four-hour sliding window, and each input sample consisted of seventeen features. Experimental results demonstrated that the data-sharing model achieved an average RMSE reduction of 9.22% across individual buildings and a 5.11% improvement in total community-level prediction accuracy. Mahjoub et al. (2022) proposed a comparative framework for short- to medium-term power consumption forecasting using three recurrent neural network architectures: LSTM, GRU, and Drop-GRU. The study

utilised 30-minute interval smart meter data from a French city and employed a sliding window approach, with input window sizes tailored to different forecasting horizons (1, 3, 7, and 15 days). Among the three models, Drop-GRU consistently achieved the lowest error metrics and the highest correlation coefficients, particularly in short-term predictions. The integration of dropout layers effectively mitigated overfitting, enhancing the model's generalisation capability and predictive stability.

Unlike recurrent models such as LSTM that process sequence data step by step, Transformers leverage a global self-attention mechanism to capture dependencies across all time steps simultaneously. This enables parallel computation and the ability to model long-range and non-local temporal dependencies, making them particularly effective for identifying key time points that significantly influence the prediction outcome in sequences with multiple periodicities and complex temporal patterns. For example, Qaisar et al. (2023) proposed OPTnet, a Transformer-based framework for predicting occupancy presence in multi-zone buildings. The data were collected at a 1-minute resolution during HVAC operating hours (9:00 – 19:00 on weekdays), resulting in a total of 324,000 data records across six zones and nine input features. OPTnet adopts a sequence-to-sequence Transformer architecture, utilising 30-minute historical multi-sensor data as input to forecast occupancy presence across various future horizons (1, 2, 5, 10, 20, and 30 minutes). The Transformer employs a self-attention mechanism to capture both temporal and cross-feature dependencies. Model performance was evaluated using accuracy and MSE. Across nearly all zones and prediction horizons, OPTnet consistently outperformed traditional methods such as DT, LSTM, and MLP, excelling in modelling long-term dependencies, handling non-stationary patterns, and maintaining robustness to noisy data. Deng et al. (2024) proposed a TCN-Attention-BiGRU mode, which integrates Temporal Convolutional Networks (TCN), Bidirectional Gated Recurrent Units (BiGRU), and an attention mechanism. The TCN module uses causal and dilated convolutions to capture long-range temporal dependencies efficiently, while the BiGRU module models both forward and backward temporal information, improving the representation of sequential patterns. The attention mechanism assigns dynamic weights to different

time steps, enabling the model to focus on important periods with periodic or anomalous energy usage.

The complexity and randomness of occupant behaviour make it challenging for DL models to accurately describe this feature. Existing studies have addressed this issue through various data strategies, such as using probabilistic occupancy matrices, distinguishing occupancy status sequences for weekdays, weekends, and holidays. For DL architectures, attention mechanisms allow models to automatically focus on repetitive behaviours within time series, enhancing long-term memory of periodic trends. By assigning adaptive weights to different time steps, the attention layer enables the model to emphasise temporally significant patterns, such as consistently high energy consumption during specific daily or weekly intervals. This selective weighting not only reinforces recognition of recurring trends but also improves the model's responsiveness to sudden, irregular anomalies by dynamically reallocating focus to unexpected variations in the input sequence. As a result, the model becomes more capable of distinguishing between normal cyclic patterns and unpredictable events in building energy use. Nevertheless, for evaluating residential building energy retrofits, how to integrate occupant behaviour features across different levels of granularity remains a critical challenge (Li et al., 2019a). For instance, retrofit planning and energy consumption prediction at the building level require the aggregation of occupant behaviour across multiple households. Even within each household, the use of individual rooms often follows distinct temporal patterns—bedrooms are primarily used at night, while kitchens are typically active around noon and in the evening. Deep learning models therefore need to capture these fine-grained behavioural variations and integrate them effectively. Moreover, another key challenge is how to organically integrate occupant behaviour with building geometry. Different household units occupy different spatial positions within the building and may exhibit substantially different energy usage patterns. For example, a household located on the top floor and facing west may experience more solar heat gains and greater exposure to external conditions, resulting in higher energy demand. However, most existing studies either focus on single-household houses or rely on averaged

occupant behaviour across households, which fails to reflect spatially differentiated energy needs among units on different floors or with different orientations. This highlights the need for DL models capable of simultaneously capturing both spatial and temporal characteristics of buildings, and effectively combining multi-dimensional features to accurately predict energy performance.

2.4.3 Inter-unit thermal interaction encoding

Significant heat transfer can occur between different spaces within residential buildings, such as adjacent rooms or households, through structural elements like walls, doors, and windows. This spatial coupling can alter the temperature distribution and energy consumption patterns of individual zones, thereby affecting the overall accuracy of load prediction. Previous studies have shown that thermal coupling between spatial units has a notable impact on heat load prediction in multi-zone buildings (Jia et al., 2022).

To capture building inter-thermal interaction, recent studies have proposed a component-based modelling approach. This method decomposes the building into multiple sub-components, such as walls, windows, roofs, floor slabs, heaters, and chillers. Surrogate models are then used to link the parameters of these components, and systems engineering paradigms are applied to manage the resulting complexity (Geyer and Singaravel, 2018). For instance, Singh et al. (2021) proposed a Component-Based Machine Learning (CBML) framework structured into three hierarchical layers: element, zone, and building. At the element level, separate neural networks predict heat flows through components such as walls, windows, roofs, floors, and infiltration, using inputs like area, U-values, g-values, and heat capacities. These outputs feed into zone-level models, which estimate heating and cooling loads based on internal gains, operating hours, and aggregated envelope heat exchange. At the building level, a final model integrates zone loads with system parameters including heating/cooling COP and boiler efficiency, to predict total annual energy demand. Geyer and Singaravel (2018) introduced a component-based ML framework that

combines static and dynamic modelling. The building is decomposed into construction-level components, such as walls, windows, and roofs, as well as zone-level thermal modules. Each construction component is modelled using ANNs to estimate annual heat transfer and solar radiation gains. The predicted outputs, including transmission losses and radiation gains, are aggregated and provided as inputs to the zone-level models. These models employ LSTM networks to capture the dynamic relationships between thermal loads and weather conditions. By linking component-level outputs with zone-level energy demand modelling, the framework enables accurate monthly predictions of heating and cooling requirements. Li et al. (2019b) decomposed the building into several standard units, trained models for each unit separately, and then summed the predicted energy consumption of all units to obtain the total building energy use. The greatest advantage of component-based modelling lies in its fine-grained geometric representation and strong transferability. By decomposing a building into generic components and modelling them individually, this approach enables model reuse and modular combination (Chen et al., 2021). For example, a sub-model trained for a specific type of component, such as windows, can be reused across different buildings, significantly improving adaptability to new designs. This modular approach, known as component-based machine learning, has been shown to enhance both interpretability and maintainability. It allows researchers to trace the contribution of each component to the overall energy consumption and to better understand how design modifications influence performance (Dalach et al., 2025). However, since buildings can vary greatly in the types and quantities of their components, especially in complex high-rise residential buildings that may include thousands of elements, the computational cost of this method can become very high. More importantly, component-based models typically make independent predictions for each component, which limits their ability to capture spatial topologies or thermal interactions among components. Furthermore, as predictions are passed between different component levels, small errors may accumulate at each step, potentially resulting in large deviations in the final energy consumption estimate.

With the development of graph-based learning, some recent studies have begun to

use graph related DL algorithms to represent the topological relationships between building spaces. Each thermal zone or building component is often abstracted as a graph node, with edges connecting adjacent or functionally related spaces. Edge weights may be defined according to geometric distance, shared surface area, or other attributes, thereby explicitly capturing the building's internal spatial relationships and thermal interactions (Jia et al., 2022). Vontzos et al. (2024) proposed a spatial-temporal power consumption prediction framework that integrates GCNs with LSTM models to effectively capture both spatial and temporal dependencies in multizone buildings. To represent the spatial relationships among thermal zones in a multistorey educational building, six graph construction methods were explored: three correlation-based approaches, including Pearson Correlation Coefficient (PCC), Absolute PCC (PCCA), and Scaled PCC (PCCS), and three distance-based approaches, including Euclidean Distance Scaled (EDS), Euclidean Distance with Threshold (EDT), and Euclidean Distance with Gaussian Kernel (EDGK). These adjacency matrices were used to define the graph structure input to a single-layer GCN, which extracted spatial features from neighbouring zones. The GCN output was then fed into an LSTM network to model temporal dependencies in zone-level power consumption. Experimental results demonstrated that the proposed GCN-LSTM framework with EDT achieved superior prediction accuracy and robustness, with CV(RMSE) values well below the ASHRAE Guideline 14 threshold of 25%. Cao et al. (2022) proposed a GNN framework for predicting building energy performance based on architectural floor plans. In this approach, each building layout is modelled as a graph, where nodes represent functional spaces (e.g., offices, laboratories, classrooms) and edges represent spatial adjacency. Node features encode semantic information such as room type, while the adjacency matrix captures the topological structure of the layout. The energy prediction task is formulated as a graph-level regression problem, with GNNs trained to estimate aggregated energy consumption from the layout graph. Several GNN architectures, including GCN, GraphSAGE, MoNet, GAT, and GatedGCN were tested, all achieving stable performance with mean absolute error (MAE) values consistently below 0.17. Jia et al. (2024) introduced a

Temporal Graph Attention Network (TGAT) for predicting heating loads in multi-zone buildings. Each thermal zone is represented as a node and the connections between zones are defined based on physical adjacency. Spatial dependencies are captured through a multi-head GAT, which dynamically learns the importance of neighbouring zones using attention weights derived from their node features. Temporal patterns are modelled using GRUs, enabling the network to capture sequential dynamics in heating loads. A set of eleven node features, including solar radiation, air temperature, humidity, and relative sun-zone orientation, were selected for each zone node. Halaçlı et al. (2023) proposed a zone-level urban building energy modelling framework (GUBEM) that leverages an edge-enhanced graph neural network to explicitly capture thermal interdependencies between adjacent zones in multi-unit residential buildings. In this framework, each building is represented as a bidirectional graph where nodes correspond to individual thermal zones, characterised by features such as envelope thermal properties, internal loads, and occupancy-related parameters. Edges encode physical interactions between neighbouring zones, incorporating attributes like shared wall areas and U-values to model inter-zone heat transfer. Yang et al. (2024) proposed a physics-constrained GNN framework for modelling multi-zone building thermal dynamics, where the building layout is represented as a graph with nodes denoting thermal zones and edges capturing heat transfer based on physical adjacency. Each node includes both dynamic features (e.g., mass flow rates, supply and ambient temperatures) and static trainable parameters such as a composite thermal time constant reflecting the zone's thermal inertia. A reduced-order RC model is embedded in each node to govern temperature evolution through a physics-informed exponential decay function. Internal gains from HVAC systems, outdoor conditions, and adjacent zones are processed via a neural network to estimate steady-state temperatures. Edge weights are computed using simplified thermal conduction laws based on temperature differences and inter-zone resistance. To enhance physical consistency and interpretability, the model introduces a loss penalty enforcing a monotonic relationship between zone size and the learned thermal parameters.

Graph-based models offer a powerful and flexible framework for representing the intricate spatial and thermal relationships within residential buildings (Kiavarz et al., 2023). Unlike traditional models that treat buildings as uniform entities or rely solely on aggregated features, graph models can explicitly encode the topological structure of buildings. Each node in the graph can represent an individual unit, room, or thermal zone, while edges capture the physical or functional connections between them, such as shared walls, adjacency, or thermal interaction paths. This structural adaptability makes graph models especially suitable for residential buildings with irregular geometries and varying numbers of household units. In large-scale residential complexes, where each unit may exhibit distinct occupancy patterns, thermal characteristics, and renovation needs, graph representations enable localised modelling while preserving the global structure (Moveh et al., 2025). For instance, a GNN can be used to predict the energy demand of a single household while accounting for the influence of its spatial neighbours. This is particularly valuable when evaluating retrofit interventions: improving insulation in one unit may alter the thermal flow to adjacent units, and graph-based models can quantify such indirect effects, allowing for more holistic and accurate assessment of retrofit benefits. More importantly, graph models can incorporate physical priors to enhance the physical plausibility of deep learning predictions. For example, Yang et al. (2024) introduced physics-informed constraints into GNN, embedding prior knowledge such as incorporating exponential decay functions to mimic thermal dissipation across nodes (thermal zones) and embedding thermal parameters such as the U-value and thermal resistance of wall components into edge features. These approaches not only enhance the interpretability of the model but also constrain it to produce outputs consistent with fundamental physical laws. However, graph models are relatively complex and require detailed spatial relationship data along with multi-unit energy consumption data. Currently, there is a lack of large-scale, publicly available energy datasets with graph-structured representations to support comprehensive training and validation of such models.

2.5 Research Gaps

With the acceleration of urban renewal policies in China, the energy retrofit of ageing residential communities has become a key task in achieving national energy conservation and carbon reduction goals (Congxiang et al., 2024). About 4 billion square metres ageing urban residential communities need to be retrofitted, involving more than 42 million households (Peng et al., 2025). These buildings are mostly constructed decades ago, suffering from poor envelope insulation, deteriorated windows and doors, outdated HVAC systems, and overall high energy consumption. However, despite the massive demand, both research and practical implementation face significant limitations and technical gaps.

First, current energy retrofit practices in old communities lack systematic and scientific energy performance evaluation methods. In many cases, retrofit decisions are made based on the empirical knowledge of technicians or generic templates, with limited ability to tailor strategies to the specific characteristics of each building. This experience-driven approach lacks quantitative support, which can lead to sub-optimal interventions and inefficient use of financial and material resources. Moreover, the retrofitting of old communities is constrained by limited financial funding (Wang et al., 2015). In the absence of quantitative evaluation of cost-effectiveness and return on investment (ROI), it becomes difficult to compare and optimise among alternative retrofit strategies (Baset and Jradi, 2024).

For automated floor plan analysis, current research has made progress in geometric reconstruction and room semantic classification of single-household residential building floor plans. However, current research lacks attention to complex multi-household residential buildings that are more common in reality. More importantly, the current literature generally lacks the engineering connection from identification to modelling. Even if the automatic identification of geometry and semantics is completed, how to transform these results into structured inputs that can be directly used by energy consumption assessment tools still lacks interfaces, which makes it difficult to stably import the automatic results into the subsequent

assessment links, and carry out large-scale construction.

From the perspective of energy performance assessment, although physics-based simulation models such as EnergyPlus or TRNSYS are widely used in building energy prediction and offer high accuracy by simulating detailed thermal and mechanical processes, their practical application in large-scale retrofit assessments is hindered by two key challenges. On one hand, such models require extensive input data, including building geometry, material properties, occupancy schedules, and system settings, which are often unavailable or outdated in older buildings. On the other hand, the computational cost of simulating multiple scenarios across many buildings makes them unsuitable for rapid, large-scale retrofit evaluation (Runge and Zmeureanu, 2021).

In this context, data-driven approaches, particularly DL models, have emerged as promising alternatives. These models can automatically extract complex patterns from large building samples, support rapid inference, and have shown great potential in predicting energy performance. In retrofit planning scenarios involving thousands of buildings and numerous design options, DL models can provide instant energy performance predictions, enabling fast optimisation of retrofit strategies (Ali et al., 2024). However, the application of deep learning in residential retrofit contexts introduces a new set of technical challenges.

Firstly, residential buildings exhibit highly complex spatiotemporal characteristics. Temporally, household behaviours vary significantly between families, leading to highly fluctuating energy demand profiles. This includes differences in thermal comfort preferences, appliance usage patterns, and occupancy schedules. Spatially, residential buildings vary in terms of floor plan, orientation, envelop configuration, and room layout. Deep learning models used for prediction in such contexts must therefore be capable of modelling both temporal sequences to capture occupant behaviour dynamics and spatial structures to reflect the building geometric and thermal topology. Secondly, traditional DL approaches often flatten building information into one-dimensional feature vectors, failing to represent the spatial dependency or thermal interaction among zones or units. In contrast, graph-based

models allow for a more expressive representation of buildings by modelling spaces (e.g., rooms or households) as nodes and physical or functional connections (e.g., shared walls, thermal paths) as edges. These models can explicitly capture topological relationships and inter-zone thermal couplings, improving prediction accuracy. However, constructing such models is complex, and publicly available graph-structured energy datasets are currently lacking, limiting the broader application and validation of this approach. Third, from an architectural perspective, it is necessary to build a multi-modal DL framework that incorporates multiple types of input: graph structures to capture spatial topologies, time series inputs to model occupant behaviour, and geometric features (e.g., floor plans, elevation drawings) extracted via CNNs. Such a model would need to implement efficient feature fusion mechanisms that combine these heterogeneous data sources into a unified representation for accurate energy prediction. However, existing research in this area remains limited. Most multi-modal models rely on simple feature concatenation and lack systematic integration strategies or physical interpretability.

To overcome the challenge of data availability and support DL model development, some researchers have proposed using large-scale parametric physical simulations to generate synthetic energy datasets (Zheng et al., 2022). By systematically varying building parameters (geometry, materials, usage patterns, weather conditions, etc.), these simulations can create a comprehensive dataset for model training without requiring labour-intensive field measurements. This strategy provides an effective alternative for data-scarce contexts like old residential neighbourhoods. However, it also introduces a significant concern—the performance gap between simulated and actual energy consumption. Due to simplifications in modelling assumptions, inaccuracies in user behaviour modelling, or environmental inputs, physics-based simulations may deviate from real-world performance (Wang et al., 2022). If these simulated datasets are directly used to train deep learning models without correction, the resulting models may internalise systematic biases and produce inaccurate predictions. Therefore, it is necessary to implement calibration strategies by incorporating a small amount of real-world measured energy data to fine-tune the DL

models. Such an approach can significantly improve model accuracy and practical applicability while leveraging the coverage of simulation-based datasets.

In summary, despite significant progress in data-driven modelling techniques, several critical research gaps remain in the context of energy retrofits for old Chinese residential buildings:

- Lack of a scientific and quantitative evaluation framework for tailored retrofit strategy selection;
- Absence of economic performance evaluation tools to support investment decisions under limited budgets;
- A shortage of automated analysis tools and BEM creation frameworks for complex multi-household floor plans.
- High data requirements and computational cost limit the applicability of physics-based models in large-scale retrofit assessments;
- Insufficient modelling of spatiotemporal characteristics unique to residential buildings, especially multi-household behaviour and complex building geometries;
- Inadequate integration of multi-modal input sources and limited research on physically informed deep learning frameworks;
- Lack of standardised, large-scale parametric simulation datasets tailored to residential retrofit contexts;
- Absence of mechanisms for integrating measured data to correct simulation-trained deep models and reduce performance gap;

3. Methodology

In response to the current research questions and gaps, this study develops an integrated data – model – decision framework to support retrofit decision-making for ageing residential communities. As shown in Fig. 2, the methodology comprises four stages. First, based on existing data and floor plans of ageing residential communities, an automated building energy modelling workflow is developed to generate

preliminary BEMs directly from multi-household floor plans. Second, the automatically generated BEMs undergo detailed parameter enrichment, followed by large-scale energy simulations to construct the residential energy simulation dataset. In the third stage, a graph-based multi-modal feature fusion model, FusionGNN, is proposed. The model is first trained on the simulation dataset and then calibrated via transfer learning using real-world household information and energy data. Finally, FusionGNN is applied to four typical multi-household residential typologies to predict energy performance under multiple retrofit scenarios. Coupled with retrofit cost estimation, a multi-objective optimisation is conducted to identify the optimal retrofit strategy.

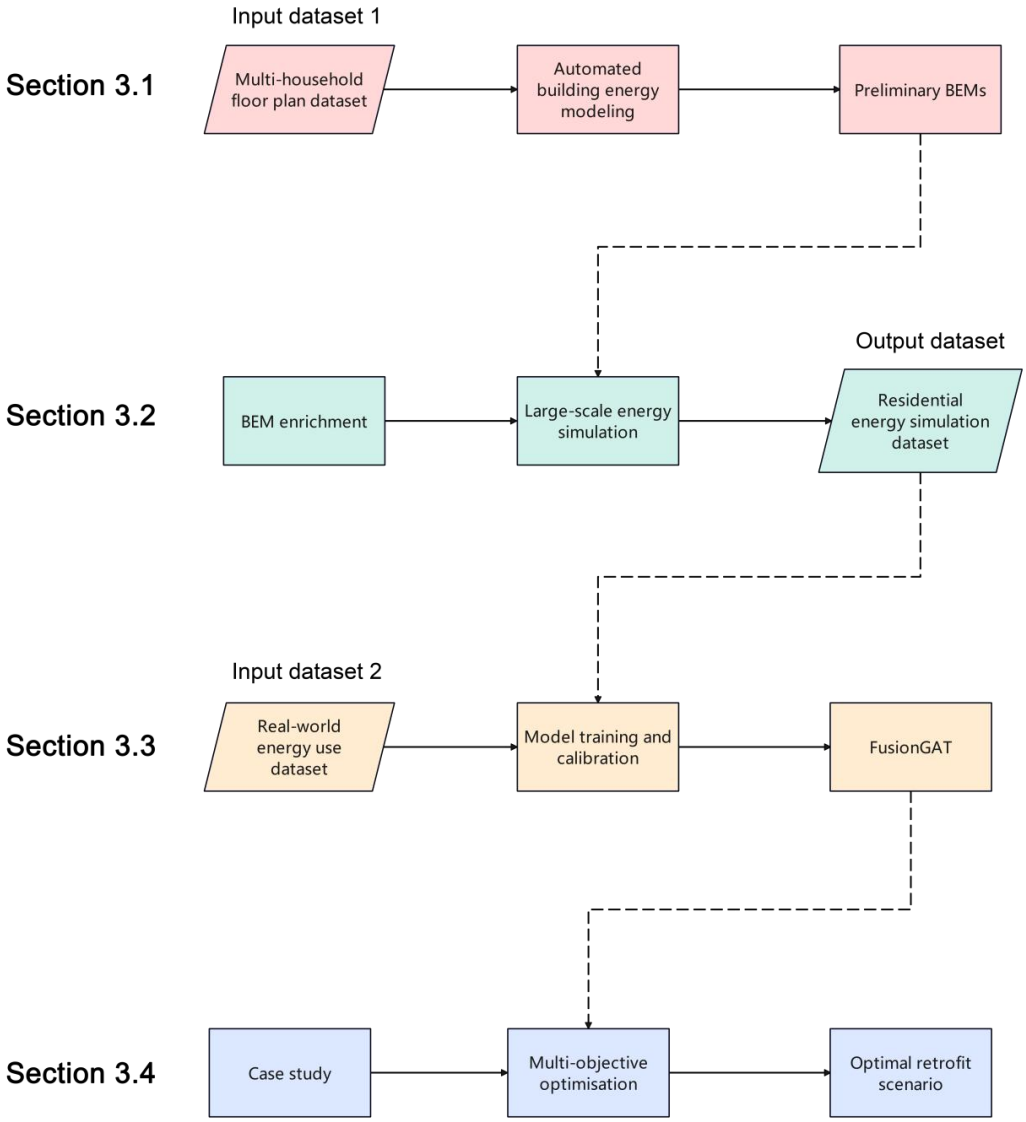


Fig. 2 Methodology framework

3.1 Automated Building Energy Modelling

This section provides a detailed explanation of the automated building energy modelling process based on floor plan images. The study began by selecting the target city and representative ageing residential communities, followed by the collection of raw data from the identified buildings. Through automated architectural element recognition and semantic analysis, the original floor plans were converted into preliminary BEMs suitable for energy simulation. These automatically generated BEMs form the foundation for subsequent large-scale parametric energy simulations. This step significantly reduces the manual effort and time typically required for energy model construction and plays a crucial role in enabling the efficient generation of a large and diverse simulation dataset.

3.1.1 Residential building data collection

City selection and residential building target

China is a vast country with a national territory of 9.6 million square kilometres. As shown in Fig. 3, it can be divided into five major climate zones: severe cold, cold, hot summer and cold winter, mild, and hot summer and warm winter (Zhao et al., 2015). Among these, the cold and severe cold zones are primarily located in northern China, covering more than half of the country's total land area. According to statistics, the total residential building area in northern China amounts to approximately 5.45 billion square metres, of which around 4.125 billion square metres were constructed during the second and early third construction phases and generally exhibit poor energy performance (Zhou, 2009). Due to high heating demand and inadequate thermal insulation, residential buildings in northern China account for nearly 70% of the region’s total energy consumption (Jiao and Rong, 2022). This study selects Jinan, a representative city in the cold climate zone, as the case study. Jinan experiences a typical continental monsoon climate characterised by hot, humid summers and cold, dry winters. The average temperature in winter is around -0.4 ℃, while in summer it

reaches as high as 27.5 ℃, with extreme temperatures exceeding 40 ℃. Under such climatic conditions, both heating and cooling demands contribute significantly to building energy consumption. Meanwhile, as the capital of China's second most populous province, Jinan has a permanent population of 9.33 million and a total residential building area of 367 million square metres (Deng et al., 2021). The city underwent rapid urban development in the 1990s, marked by the construction of universities, research institutes, and commercial complexes, which attracted a large influx of residents. To accommodate the growing housing demand, a substantial number of residential buildings were constructed during this period. Although these buildings are expected to remain in use for another 20 to 40 years, their poor thermal performance and ageing mechanical systems make them urgently in need of energy-efficiency retrofitting (Huang et al., 2020).

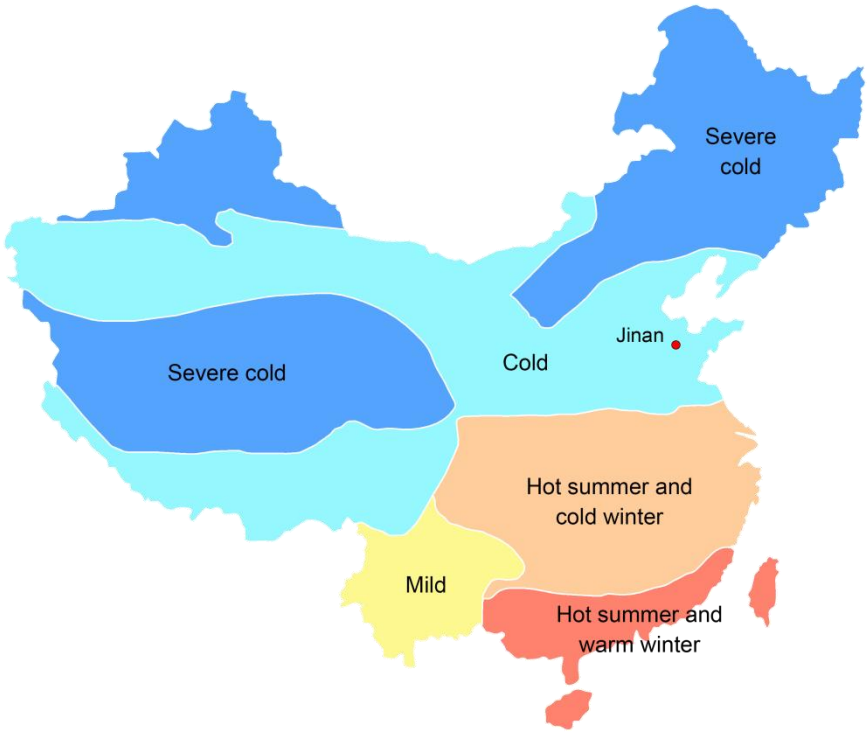


Fig. 3 Climate zones in China

Residential building data and floor plan collection

Based on official policy documents issued by the Jinan municipal government, 64

residential communities included in the city's 2020 energy retrofit plan for ageing neighbourhoods were selected as the research subjects. Housing data for these communities were collected using a web crawler developed for Lianjia, one of China's leading second-hand housing platforms, resulting in a total of 2,383 valid records. Key building attributes, including construction year, household floor area, and number of floors, were statistically analysed, as illustrated in Fig. 4.

In terms of the construction year, most of the buildings in the 64 communities were built in the 1990s (46.3%), followed by the 2000s (26.6%) and the 1980s (22.6%), while the proportion of buildings built before the 1980s and after 2010 was small (2.4% and 2.2% respectively). In terms of household floor area, small and medium-sized apartments are the main type, with 60-90 square metres accounting for 40.2%, 90-120 square metres accounting for 21.1%, and less than 60 square metres accounting for 23.4%. In terms of building floors, 4-6 floors buildings are the most common, accounting for 68.1% of the sample, 7-9 floors account for 16.9%, and those higher than 10 floors account for about 13.8%. In summary, the selected buildings have the characteristics of long service life, small and medium-sized units, and low and medium floors, which are in line with the typical characteristics of old residential communities.

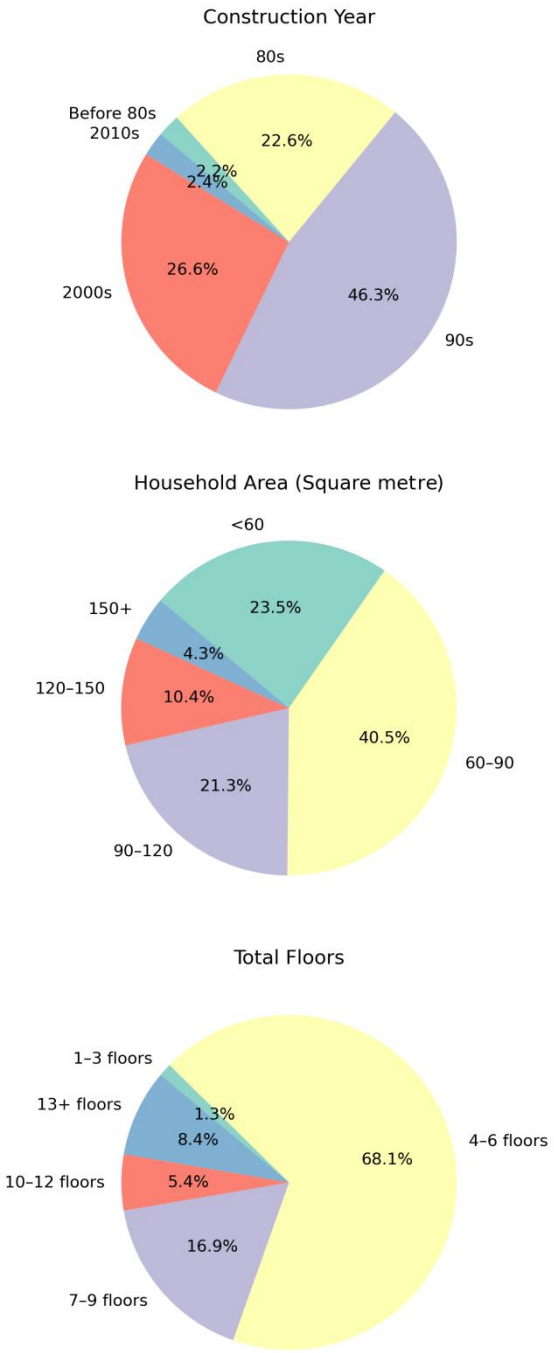


Fig. 4 Statistics of residential communities in Jinan

In addition, floor plans were collected for the 64 ageing residential communities, yielding a total of 300 multi-household residential floor plans. As shown in Fig. 5, the selected floor plans illustrate the full-floor spatial organisation of residential buildings, including both shared communal areas (such as staircases and corridors) and various individual household unit layouts. The collected residential floor plans are presented as black-and-white raster images, where walls are depicted as solid black regions,

windows are represented by double or triple lines, and doors are illustrated as arcs. Some sliding doors are shown as rectangular shapes.

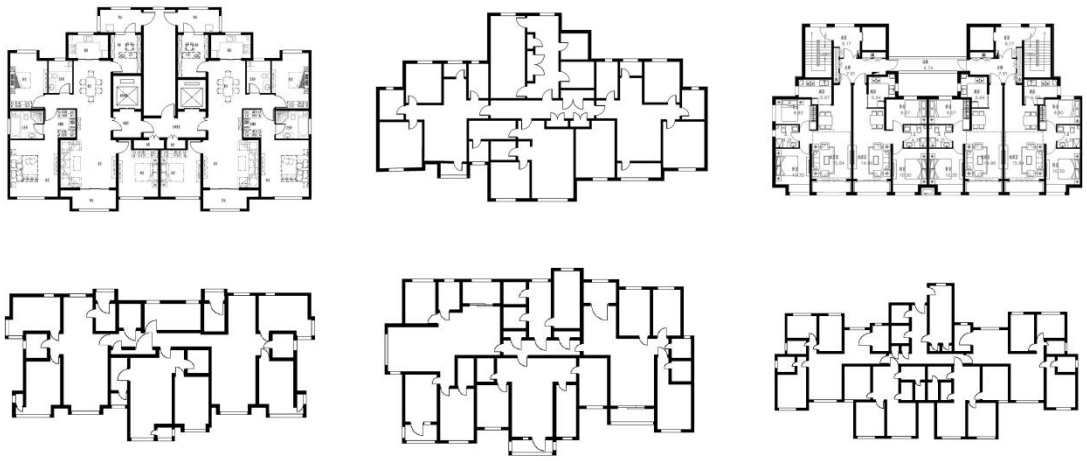


Fig. 5 Example of collected floor plans

A statistical analysis of the stair-to-unit ratio across the 300 plans was conducted to identify four representative floor plan typologies. As shown in Fig. 6, each coloured block represents an individual household unit. The four typical floor plan layouts are described as follows:

(a): One stair with two units layout with symmetrical plan arrangement;

(b): One stair with three units layout clustered around a central circulation core;

(c): One stair with four units layout clustered around a central circulation core;

(d): Two stairs with four units layout with corridor-style access.

The distribution of these floor plan types is also illustrated in the Fig. 6. Types (a) and (b) dominate, accounting for a combined 85.7% of all samples, whereas types (c) and (d) are less common, representing 11.6% and 2.6%, respectively.

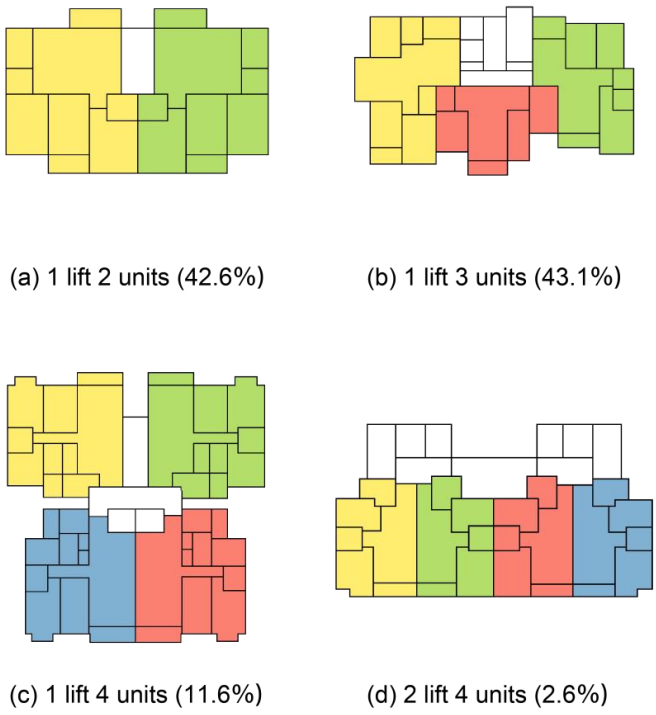


Fig. 6 Four typical floor plan typologies in 64 ageing residential communities

Floor plan process

From the 300 collected multi-household residential floor plans, 120 were manually annotated to construct the multi-household residential floor plan dataset. The dataset includes the original floor plan images (Fig. 7a), pixel-level annotations of walls, windows, doors, and room regions (Fig. 7b), as well as room object clustering and room type labels (Fig. 7c and Fig. 7d). This dataset is used to develop and test the automated building energy modelling workflow, enabling automatic conversion from 2D plans to 3D energy simulation models; for the remaining 180 floor plans, the corresponding energy models are generated by the proposed automated energy modelling pipeline. Overall, the dataset serves as the foundation for constructing preliminary BEMs and for generating large-scale simulation data.

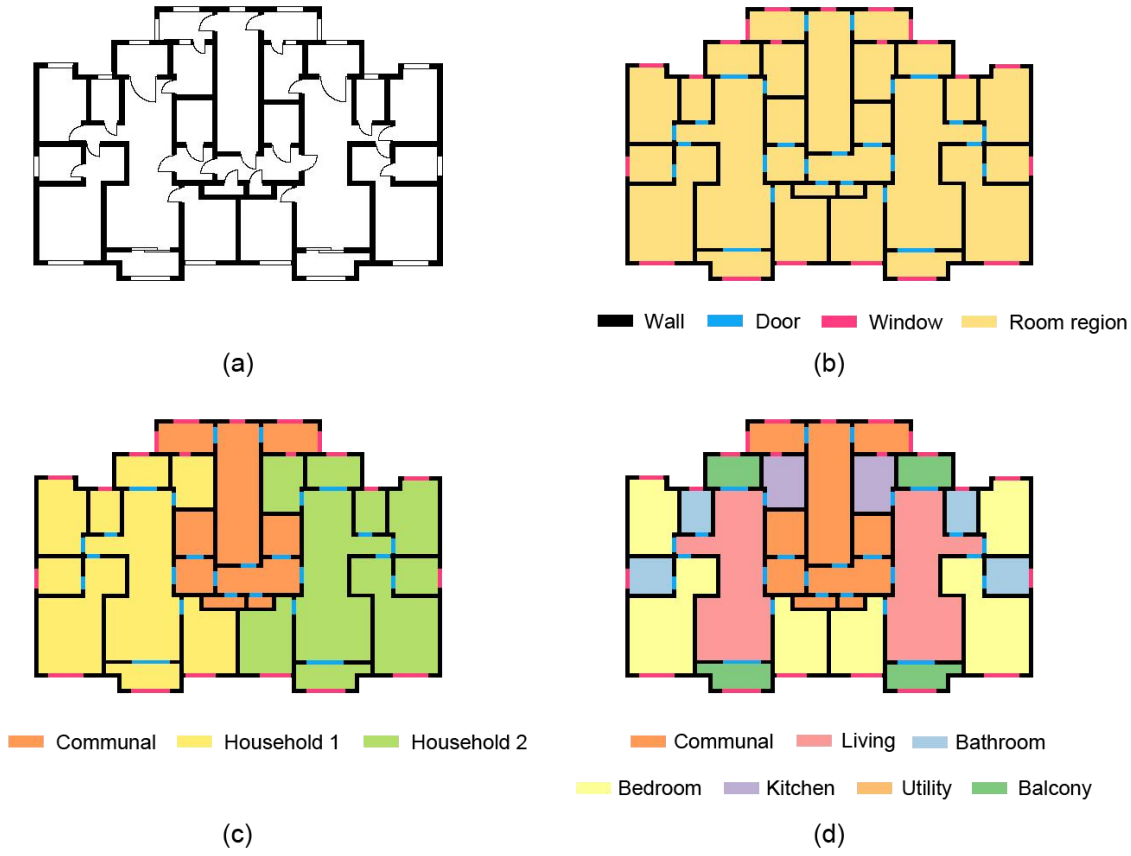


Fig. 7 Annotation examples (a) raw floor plan image (b) floor plan annotation with wall, door, window, and room region (c) visualisation of room cluster annotation (d) visualisation of room type annotation

3.1.2 Automated building energy modelling

The automated building energy modelling process based on floor plans consists of four steps: architectural element recognition, multi-household reconfiguration, room type classification, and energy modelling. Architectural element recognition focuses on extracting walls, windows, doors, and room regions from the floor plan, providing the geometric foundation for energy model construction. Multi-household reconfiguration involves clustering individual rooms to reconstruct household layouts, enabling energy simulations at different spatial resolutions. Room type classification is used to assign semantic labels to each room, which supports the automatic assignment of occupancy schedules. Finally, the energy modelling step transforms the extracted geometric and semantic information into preliminary BEM, which serve

as the basis for subsequent large-scale energy simulations.

Architectural element recognition and vectorisation

The architectural element recognition module builds upon the Deep Multi-Task Network (DMTN) (Zeng et al., 2019) for pixel-level floor plan segmentation. Specifically, this study introduces two modifications: (1) refining the ‘Opening’ class into two distinct ‘Window’ and ‘Door’ classes, and (2) redeveloping the room type classification using generic room features and inherited connectivity relationships. These two modifications aim to enhance the accuracy of architecture element recognition and improve the applicability for versatile 2D floor plan images.

The first modification distinguishes the original ‘Opening’ class into sub-classes of ‘Window’ and ‘Door’. Further classification of opening types supports room feature extraction, spatial connectivity analysis, and detailed 3D building energy modelling. The second modification of the room type classification involves incorporating generic room features, such as distance between rooms, and connectivity relationships. These modifications aim to enhance the applicability and robustness needed for a wide range of floor plan images, with or without furniture arrangements and additional texture features. As a result, the output layers of the architectural element recognition include four fundamental element classes—Wall, Door, Window and Room region. Fig. 8 illustrates the floor plan segmentation, showing the input floor plan on the left and generated segmented results for four architecture elements on the right.

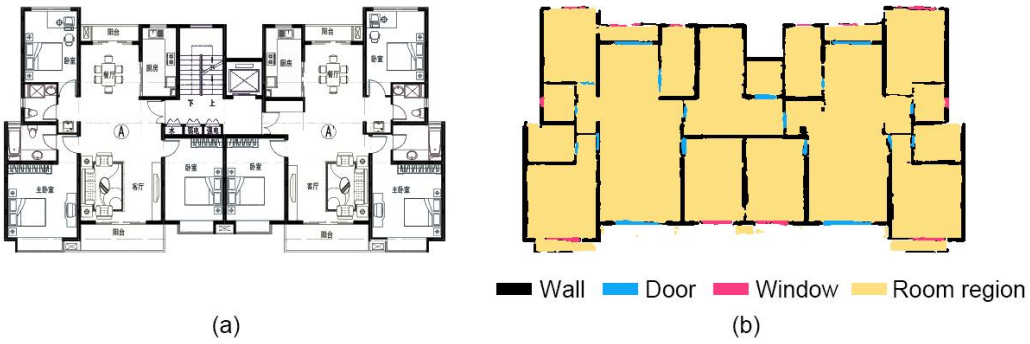


Fig. 8 Floor plan segmentation using the modified DMTN model (a) the input sample floor plan (b) four architectural element segmentation results

Fig. 9 illustrates the required image processing steps to vectorise the pixel-level segmentation results shown in Fig. 8. First, the segmented room region pixels are grouped into individual room objects, and those with an area smaller than 5% of the total room pixel area are removed. The external boundaries of the remaining room objects are then extracted (Fig. 9b). These boundaries undergo an iterative dilation process (Fig. 9d) to achieve alignment with the structural element mask (Fig. 9c). Subsequently, the medial axes of the room boundaries are extracted using the thinning algorithm (Huang et al., 2003). The resulting single-line room boundaries are finally integrated with the extracted window and door pixels to reconstruct a vectorised floor plan (Fig. 9e). This vector representation encompasses geometric information of architectural elements, and serves as the basis for multi-household reconfiguration and energy model geometry construction.

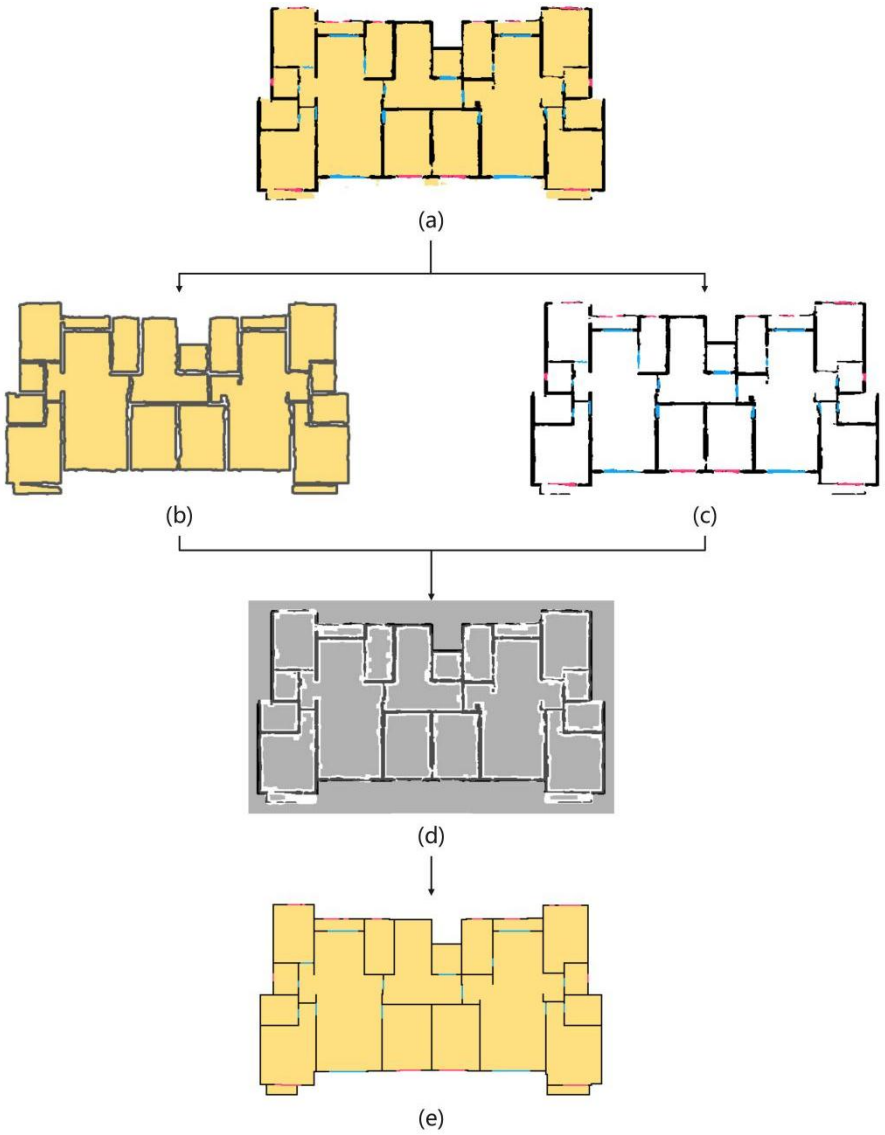


Fig. 9 Image processing for floor plan vectorisation (a) segmented floor plan (b) room pixels (c) wall, door and window pixels (d) iterative room boundary dilation (e) vectorised floor plan

Multi-household reconfiguration

The spatial connectivity complexity in multi-household floor plans increases with the number of rooms and significantly correlates with the shared communal spaces connecting multiple household units. To uncover multi-household reconfigurations, this study constructs an underlying spatial connectivity graph for each floor plan to facilitate household layout analyses. The first step classifies core rooms in the spatial connectivity graph, as shown in Fig. 10 (a+b). Core rooms are room objects that

consist of multi-connectivity relationships in the floor plan and represent communal and circulation spaces shared between or within the households. Three core room types identified in this study are communal, circulation and function rooms. Communal core rooms are hallways or staircases that connect multi-household entrances. Circulation core rooms are living rooms, or halls, within each household. Function rooms are utility or specific rooms that connect to multiple spaces within the household. This study proposes a new Core Topology Room Clustering (CTRC) algorithm with three core rooms to decompose the multi-household graph into distinct household and communal clusters.

Fig. 10 illustrates the multi-household reconfiguration process, starting with the construction of a multi-household graph from a given vectorised floor plan (Fig. 10a). The multi-household graph consists of nodes representing room objects and edges representing the connectivity relationships through doors between rooms. Room nodes with more than one direct connection are identified as core nodes, shaded in yellow (Fig. 10b). Fig. 10c shows the intermediate result of the multi-household connectivity graph with core nodes identified in three categories:

- Communal spaces that connect multiple units
- Circulation spaces (e.g., living rooms) in a household
- Functional spaces (e.g., master bedrooms with private bathrooms and balconies) with multi-connectivity relationships in a household

To classify these core nodes, this study employs a Decision Tree (DT) classifier with five room features from each core node: (1) area ratio (node area/total area), (2) sum area ratio of adjacent core nodes, (3) sum area ratio of adjacent non-core nodes, (4) number of adjacent core nodes, and (5) mean depth. Specifically, the mean depth is a measurement that reflects the centrality or peripherality of a particular node within a spatial network (Turner, 2001), calculated by averaging the shortest path distances (depths) from a given core node to all other nodes in the network. This study applies mean depth to evaluate hierarchical accessibility patterns. For instance, communal spaces exhibit the highest accessibility (lowest mean depth), reflecting their role in organising the entire floor plan layout. Household circulation spaces demonstrate

intermediate accessibility, as they primarily connect rooms within their respective units while maintaining comparable accessibility levels across different households. Function rooms exhibit the lowest accessibility (highest mean depth), since they typically serve only specific, localised spaces within the network. In total, 935 core nodes are extracted, of which 80% are used for training the DT classifier and the remaining 20% were used for testing.

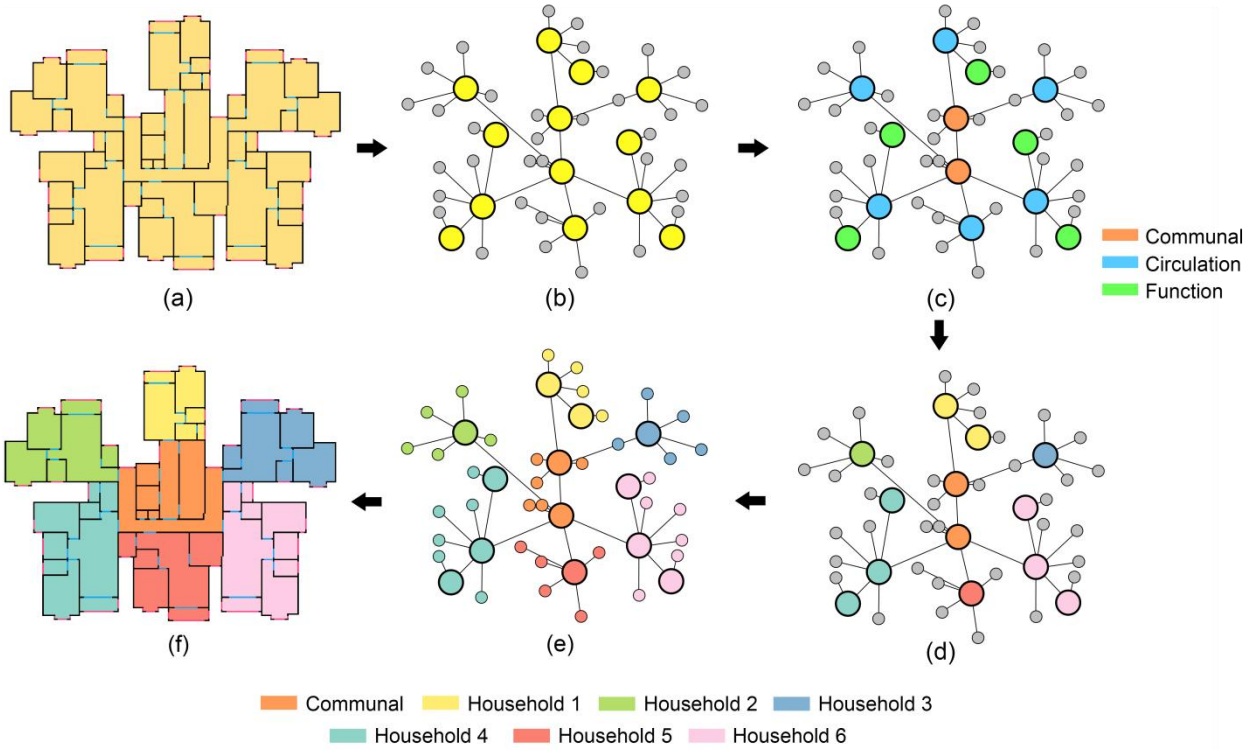


Fig. 10 Multi-household reconfiguration through the CTRC algorithm (a) vectorised floor plan; (b) core nodes highlighted in the spatial connectivity network; (c) core node classification; (d) household cluster identification (e) room node search (f) reconfigured floor plan

With the classified core nodes in the spatial connectivity graph (Fig. 10d), the proposed CTRC algorithm carries out room clustering, including cluster creation, cluster expansion, and cluster merging:

Cluster creation (Fig. 10e):

Create a Circulation Cluster (CL) containing all circulation core nodes.

Create individual clusters for each household/function core node:

Household Clusters: H_1 – H_m (where m = household core node count)

Function Clusters: F_1 – F_n (where n = function core node count)

Cluster Expansion (Fig. 10f):

For each non-core room node RN_k in the Room Graph:

Calculate topological depth to all core nodes

Identify nearest core node CN_t

Assign RN_k to the cluster of CN_t

Cluster Merging (Fig. 10g):

For each function core node FCN_X and corresponding cluster F_x:

Calculate topological depth to all other core nodes

Identify all nearest core nodes [CN_y]

If length [CN_y] = 1:

Merge F_x to the cluster of CN_y

Else:

If [CN_y] contains more than 1 household core node:

Merge F_x into CL

Else:

Merge F_x into the nearest Household Cluster

As a result, each room is clustered into either a communal or one of the household clusters. Fig. 10h shows the result where rooms are reconfigured and remapped onto the input floor plan.

Room type classification

This step focuses on classifying individual room types within each household cluster in Step 2. As shown in Fig. 11, each household cluster is processed individually, with

seven features collected per room: (1) area ratio (room area/total household cluster area), (2) long-to-short axis ratio, (3) window-to-wall ratio, (4) the number of adjacent rooms, (5) sum of adjacent room area ratios, (6) mean depth and (7) entrance distance. The entrance distance is computed by first identifying entrance doors as those doors that connect household clusters to circulation clusters. The room directly adjacent to an entrance door is defined as the entrance room. For each room within the household cluster, the Manhattan distance between the room centre and the entrance location is calculated. When multiple entrance doors exist within a household cluster, the entrance distance for each room is computed with respect to each entrance and averaged. To ensure comparability across households of varying sizes, the entrance distances are normalised by the maximum distance within the same household cluster. This feature encodes the accessibility hierarchy of rooms and provides functional cues that are particularly informative for distinguishing private spaces (e.g., bedrooms) from public or service spaces (e.g., living rooms and kitchens). To predict room functions, we implement a Graph Attention Network (GAT), which extends conventional Graph Neural Networks (GNNs) by incorporating attention mechanisms that dynamically assign various weights to neighbouring nodes. The attention mechanisms allow the model to assess the relative influence of surrounding rooms during training, thereby capturing complex spatial dependencies. The GAT architecture consists of three layers: the first two use multi-head attention to aggregate and transform node features, and the final layer performs classification to output the predicted room type. The model is trained using the Adam optimiser with a learning rate of 0.005 and a categorical cross-entropy loss function, based on 400 household graphs (80% for training and 20% for testing) derived from the 120 multi-household floor plans. The model outputs six residential room types, including living room, bedroom, kitchen, bathroom, balcony, and utility. Among these, living room, bedroom, kitchen, and bathroom have distinct energy usage patterns in residential buildings (Dong et al., 2023), which supports the construction of the energy model in the next step. Communal cluster rooms are directly labelled as communal spaces without classification.

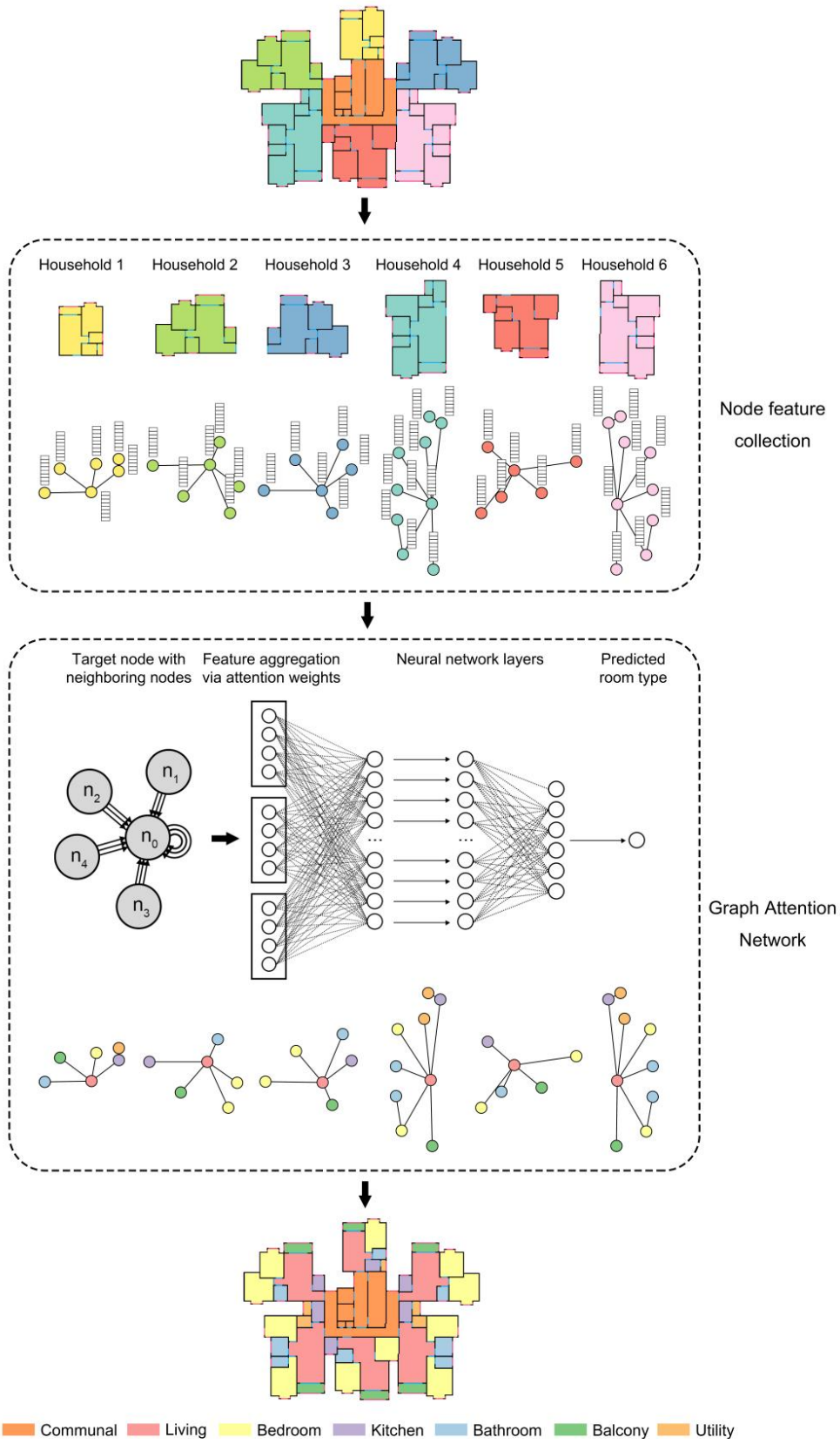


Fig. 11 Individual household room feature collection and room type prediction using Graph Attention Network

Automated energy model construction

This step automates the construction of preliminary BEMs by integrating the vectorised architectural elements, household clusters, and room type classifications. These BEMs represent simplified but complete energy modelling structures, containing the essential geometric, topological, and semantic information required for simulation, and are encoded into EnergyPlus-compatible Input Data Files (IDFs) (Shen and Yarnold, 2021). This automated process reduces manual efforts while ensuring the accuracy and efficiency of BEM construction.

The Python-based EnergyPlus Application Programming Interface (API) library, Geomeppy, is employed to automate the generation of IDF files (Bull J, et al. 2016). Geomeppy constructs thermal zone geometries by extracting room boundary, window, and door coordinates from vectorised floor plans and combining them with user-specified floor and component heights. Each thermal zone is assigned a unique identifier that embeds the storey information, automatically derived household-cluster and room-type. Fig. 12 shows the resulting energy model geometry visualised in Grasshopper, and Table. 1 provides examples of the thermal zone identifier naming scheme.

In this study, the storey height is uniformly set to 3m, the window base height to 0.9m, the window height to 1.2m, and the door height to 2m (SHI et al., 2024). Using this automated workflow, preliminary BEMs (saved as IDF files) were generated for 300 multi-household residential floor plans, forming the basis for large-scale energy simulations in the subsequent stage.

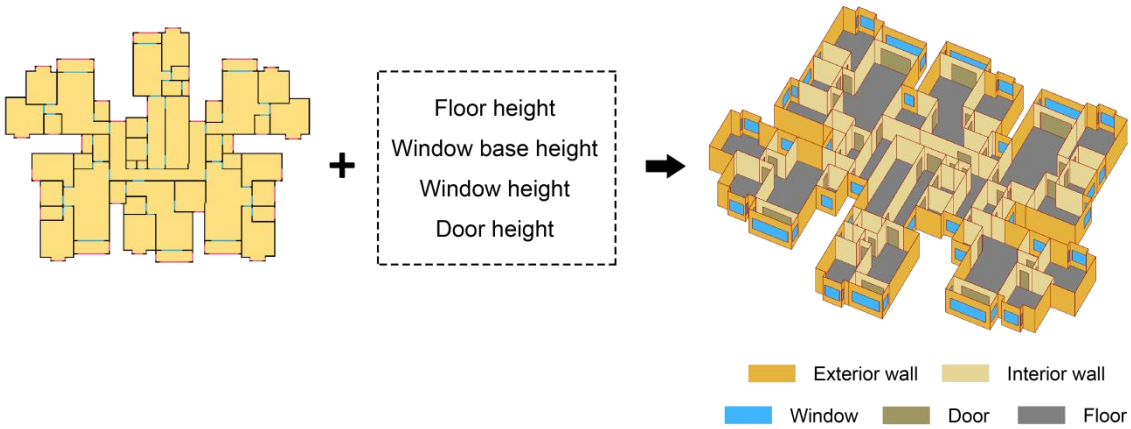


Fig. 12 Visualisation of input floor plan image and automatically generated energy model geometry

LV1	Household1	Bedroom1
		Bedroom2
		Livingroom1
	...	...
	Household6	Kitchen1
		Bathroom1
		Balcony1

Table. 1 Example of thermal zone identifier naming scheme

Validation

To evaluate the reliability of the proposed method, three groups of energy simulations were conducted. The Ground Truth group was constructed manually, serving as the baseline energy model. The Group 1 used manually constructed geometries, while the household clustering and room type prediction were performed automatically, followed by automated schedule assignment. The Group 2 used raw floor plan images to fully automate both the geometry construction and schedule assignment. All three simulation groups used the same climate file to calculate the annual EUI as an energy indicator.

To evaluate the reliability of the proposed method, three comparative simulation groups were conducted. For consistency, standard power density and usage schedule templates were assigned to the six room categories. The Ground Truth group was totally created manually, with geometry and room templates produced and assigned by hand, serving as the baseline energy model. Group 1 used manually constructed geometries, while household clustering and room type prediction were performed automatically; room power densities and schedule templates were then assigned automatically based on the predicted room types. Group 2 started from raw floor plan images with fully automated geometry construction and room template (power density and schedule) assignment. All three groups used the same climate file to simulate annual EUI as the evaluation metric.

3.2 Residential Building Energy Consumption Dataset Development

This section describes the parameter enrichment of the automatically generated preliminary BEMs, including urban-context reconstruction, simulation parameter settings, occupant-behaviour definition, and parameter assignment. Subsequently, large-scale energy simulations were conducted to comprehensively capture the impacts of variations in building morphology, envelope performance, internal loads, and occupant behaviour on building energy use. For each building sample, input parameters and the corresponding energy outputs were recorded to construct a comprehensive residential building energy simulation dataset. Organised at the room, household, and building levels, this dataset reflects residential building characteristics and energy-use patterns, facilitates prioritisation of retrofit measures at different granularities, and provides high-quality training data for deep learning model development.

3.2.1 Preliminary BEM enrichment

Urban context reconstruction

To simulate the impact of the urban environment on building energy consumption, this study reconstructed the urban context of 64 ageing residential communities. Based on publicly available building footprint shapefiles for the Jinan region, a total of 1,246 building footprint records were extracted from the selected communities. Each footprint was manually populated using the preliminary BEM generated in Section 3.1.2. In total, 4,282 BEM samples were assembled to populate all 1,246 footprints across the 64 communities. An example of the reconstructed urban layout is illustrated in Fig. 13.

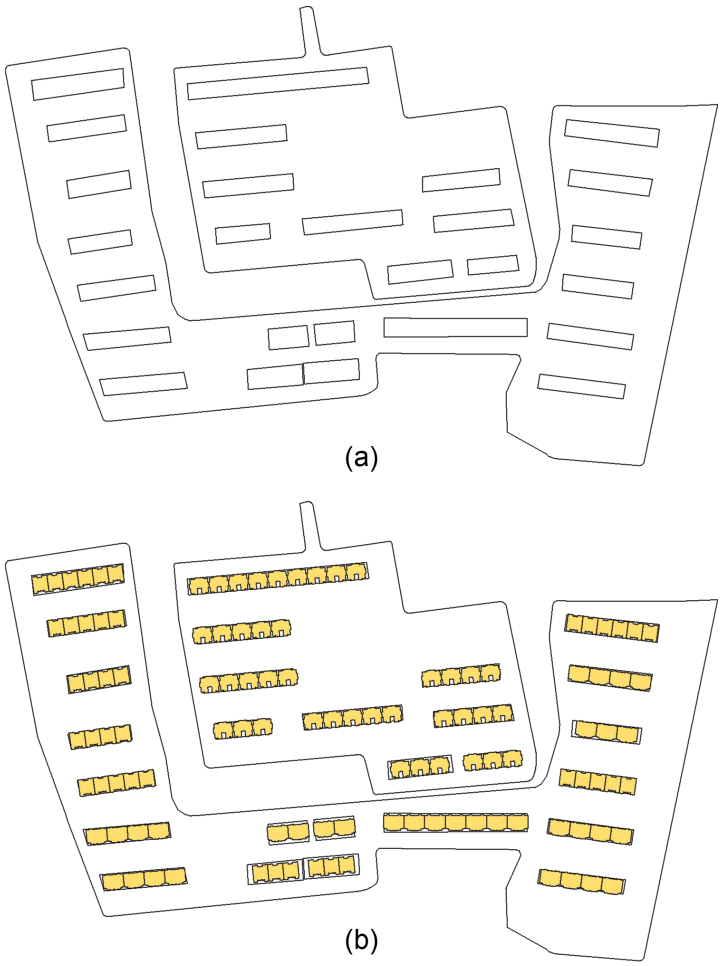


Fig. 13 Urban context reconstruction through building footprint data (a) building footprint (b) reconstructed urban context

Based on the reconstructed urban context, buildings located within a 30-metre radius of each target BEM were included as shading obstacles in the energy simulation process. According to height attributes provided in the shapefile, the number of floors for each BEM was determined by dividing the building footprint height by the floor height (3 metres). The height of shading buildings was directly taken from the footprint height data. Fig. 14 shows an example of a target BEM and the surrounding shading context used for simulation.

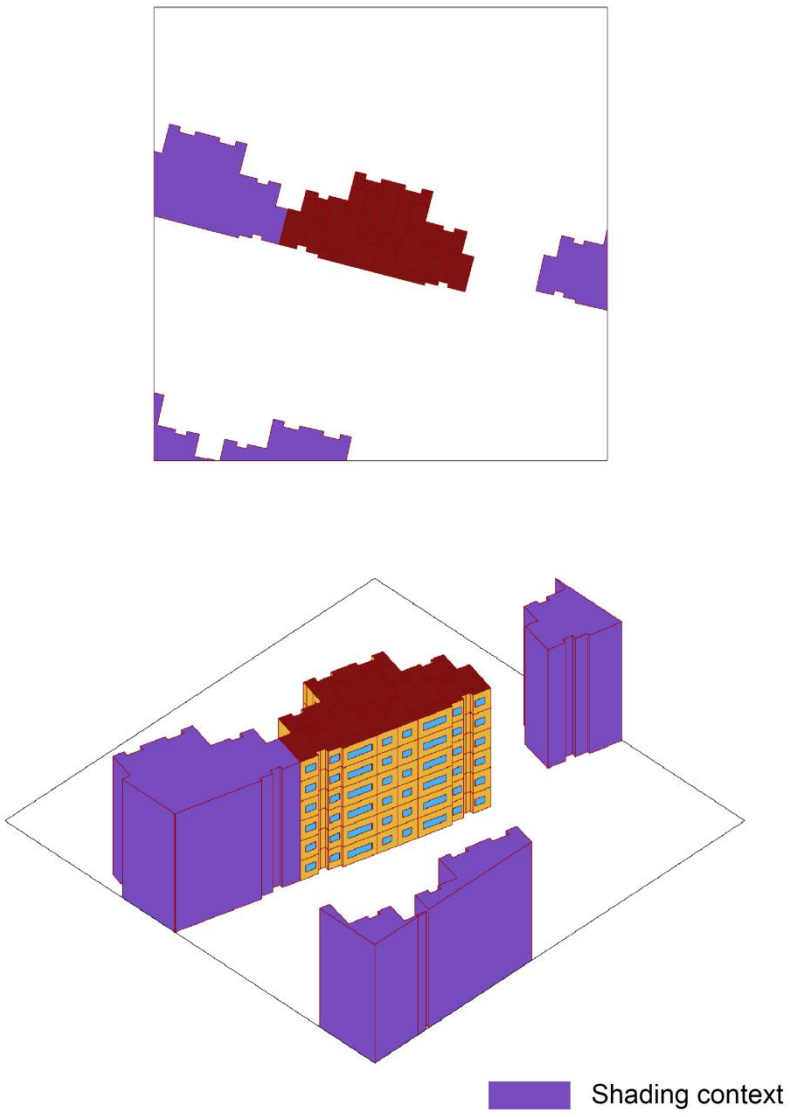


Fig. 14 Contextual building shading visualisation

Simulation parameter definition

To accurately represent the thermal performance of residential building envelopes while maintaining computational efficiency in large-scale energy simulations, this study parameterises the thermal performance of building envelope elements using U-values rather than detailed multi-layer material definitions. This enables the creation of a uniform and continuous feature space, facilitating the model's ability to capture the nonlinear relationship between building envelope thermal properties and

energy consumption. The adopted U-value ranges for different envelope types are summarised in

Table. 2. For exterior walls, un-insulated brick or concrete structures exhibit high U-values, ranging from 1.5 to 2.0 W/m^2 • K or even up to 3.0 in some rural cases (Guo et al., 2021, Zhang et al., 2023). In contrast, U-values of wall with insulation layers compliant with energy-saving standards typically fall below 1.0, often around 0.4 – 0.6 (Wang et al., 2019a). Therefore, this study adopts a U-value range of 0.4 – 3.0 for exterior walls. Interior walls, often constructed from un-insulated 120 mm hollow brick partitions, are assigned U-values in the 1.5 – 3.0 range (Guo et al., 2021). For ceilings, U-values range from 0.3 – 3.5, reflecting both un-insulated exposed roofs and insulated slab roofs (Wang et al., 2019a). Floors, depending on whether they are above ventilated crawlspaces or insulated ground slabs, are modelled with U-values from 0.4 to 2.5 (Guo et al., 2021). For doors, a U-value range of 0.7 – 2.5 is applied to cover both insulated metal or composite products and basic wooden doors (Sadhukhan et al., 2020).

Windows were modelled with U-values between 0.8 and 6.0, reflecting the spectrum from high-performance double-glazed low-emissivity windows to outdated single-pane aluminium-frame windows (Zou et al., 2021, Zhu et al., 2021). The Solar Heat Gain Coefficient (SHGC) was set between 0.5 and 0.86, covering glass coatings from low-e to fully transparent glazing (Zhu et al., 2021).

The air change rate (ACH) was assigned a range of 0.3–3.0 h^{-1}, based on observed leakage conditions from poorly sealed apartments to recently upgraded households (Wang et al., 2019a, Zhang et al., 2023). Older buildings with deteriorated windows typically exhibit higher infiltration rates (>1.0 ACH), while tight envelopes with weatherisation measures achieve lower values (<0.5 ACH).

The occupancy density was set between 0.01–0.10 persons/m², consistent with observed household sizes and floor areas in urban and suburban China (Chen et al., 2019). For lighting power density, a range of 1.0–10.0 W/m² was used, corresponding to modern LED usage and traditional incandescent scenarios (Luo and Du, 2024). The equipment power density was set between 1.0 and 20.0 W/m², encompassing

variation in plug-load intensity and appliance ownership patterns (Luo and Du, 2024). Heating and cooling setpoints were defined as 16–20 °C for heating and 24–28 °C for cooling, reflecting national guidelines and measured occupant behaviour in northern regions. Central heating systems commonly operate around 18–20 °C, while rural households may heat to as low as 15 °C due to limited infrastructure (Zhang et al., 2023). Cooling preferences vary but tend to cluster around 26 °C, consistent with observed AC usage patterns (Luo and Du, 2024). To control whether a room had active space-conditioning demand, a HVAC-conditioned flag was implemented to describe rooms as mechanically conditioned (served by HVAC) or free-running (unconditioned); only conditioned rooms were assigned and enforced with heating and cooling set points in the simulations.

Parameter	Unit	Min	Max	Step	Source
Exterior wall U-value	$W/m^2 \cdot K$	0.4	3.0	0.1	(An et al., 2023)
Interior wall U-value	$W/m^2 \cdot K$	1.5	3.0	0.1	(An et al., 2023)
Floor U-value	$W/m^2 \cdot K$	0.4	2.5	0.1	(An et al., 2023)
Ground floor U-value					
Ceiling U-value	$W/m^2 \cdot K$	0.3	3.5	0.1	(An et al., 2023)
Door U-value	$W/m^2 \cdot K$	0.7	2.5	0.1	(An et al., 2023)
Window U-value	$W/m^2 \cdot K$	0.8	6.0	0.1	(An et al., 2023)
ACH	h^{-1}	0.3	3.0	0.1	(Ahn and Kim, 2022, Zou et al., 2021)
People density	persons/ m^2	0.01	0.10	0.01	(Dong et al., 2023)
Lighting power	W/ m^2	1.0	10.0	0.5	(Luo and Du, 2024)

density					
Equipment power density	W/ m^2	1.0	20.0	0.5	(Luo and Du, 2024)
Heating set point	℃	16	20	0.5	(Dong et al., 2023)
Cooling set point	℃	24	28	0.5	(Dong et al., 2023)
HVAC-conditioned		0	1		

Table. 2 Energy simulation parameter configurations

Occupant behaviour

Based on Dong et al. (2023), this study defines four typical occupant behaviours, including office-worker household, retired household, household with an infant, and student household. As shown in Fig. 15, each profile is associated with a distinct 24-hour schedule for four primary functional zones: bedroom, living room, kitchen, and bathroom. These schedules reflect typical occupancy patterns and room usage intensity throughout the day. Balconies and utility rooms were assigned low-activity profiles (0.1 or 0.2 during active periods), while communal spaces were given a low background presence drawn uniformly from 0.1 to 0.4 across the day. To simulate seasonal absence behaviours (e.g. long-term travel or holidays), each household was randomly assigned a vacancy duration ranging from 0 to 3 months.

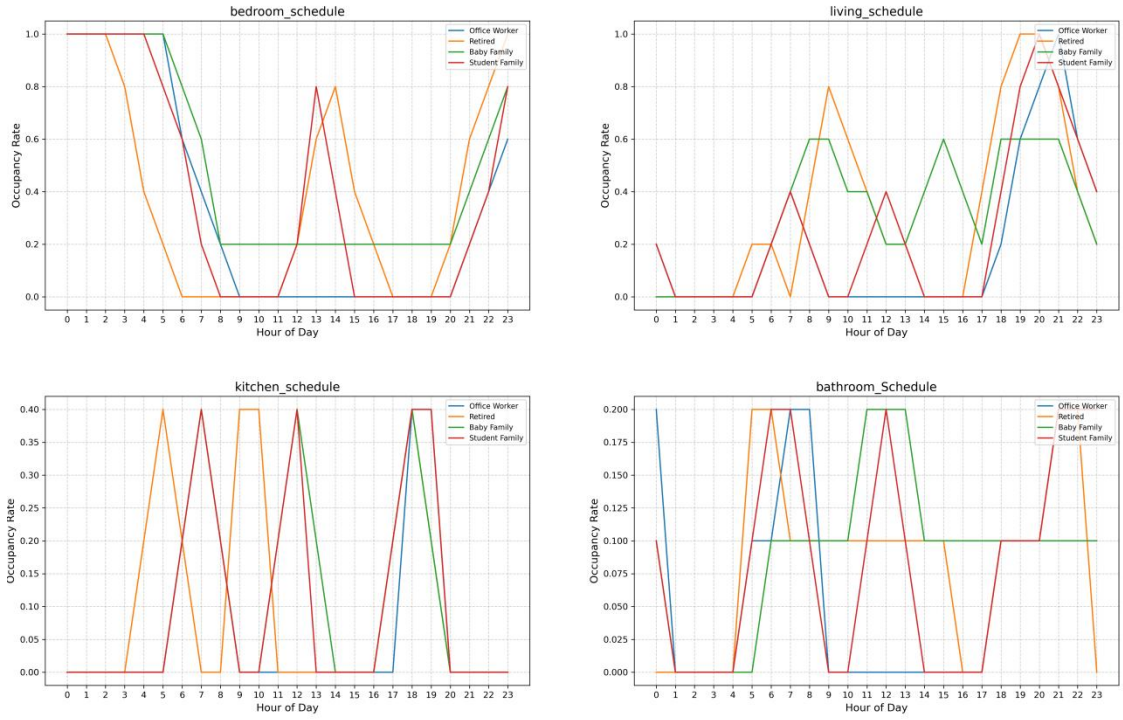


Fig. 15 Typical Chinese residential occupant behaviours

To comprehensively simulate the diversity of occupant behaviour, we introduced two types of stochastic variation: time displacement and intensity adjustment. Time displacement shifts the active occupancy periods by up to ± 3 hours to reflect variability in daily routines, such as differences in wake-up times among households (e.g., between 6:00 and 8:00 AM). Intensity adjustment modifies the duration of occupancy periods by up to $\pm 25\%$, simulating differences in how often and how intensively individuals use specific spaces. This stochastic strategy introduces sample-level variability while preserving the underlying behavioural structure, thereby enhancing the realism of occupancy patterns. These variations were incorporated into the parametric simulation framework to generate energy use profiles representative of diverse residential lifestyles. As illustrated in Fig. 16, the randomly perturbed bedroom schedule exhibits a one-hour delay and irregular intensity variations compared to the typical schedule of an office worker, demonstrating how stochastic adjustments affect the temporal and quantitative aspects of room usage.

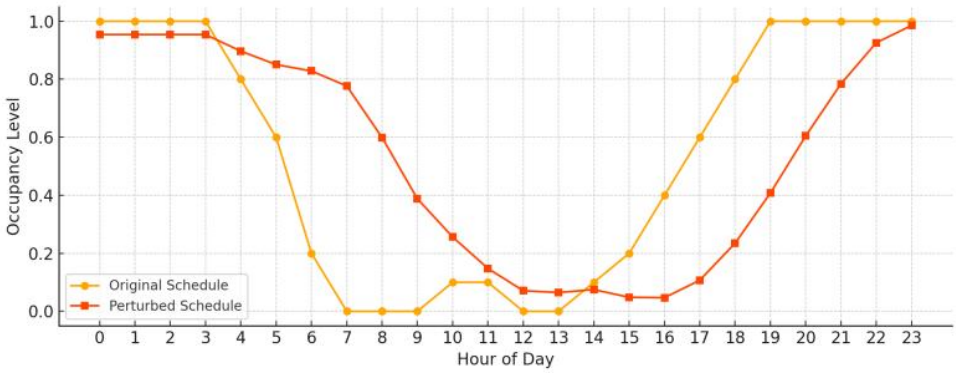


Fig. 16 Example of random schedule perturbation

Parameter assignment

Based on the defined parameter ranges, this study used Grasshopper with the Honeybee plugin to enrich the automatically generated preliminary BEMs. The IDF files were imported into Grasshopper, and room identifiers were parsed and mapped to the corresponding Honeybee inputs to execute randomised parameter assignments. All simulation parameters were categorised into three hierarchical levels: building-level, household-level, and room-level, as summarised in Table. 3.

Building-level parameters include the U-values of exterior walls, interior walls, floors, ground floors, ceilings, exposed roofs, windows and doors. As these envelope components are typically installed uniformly during construction, building-level parameters were randomly generated within the predefined ranges and assigned uniformly to each BEM to ensure consistent thermal properties across the entire building.

Household-level parameters include the ACH, people density, and typical occupant behaviour type. Each household was randomly assigned one of the four typical occupant behaviour profiles, with room-type-specific 24-hour occupancy schedules allocated accordingly. The assignment of room schedules was automated by parsing the room type field included in the room identifier. Household-level parameters were assigned based on the number of households identified in each BEM. By parsing the household field in the room identifiers, consistent parameters were assigned to all rooms belonging to the same household.

Room-level parameters include lighting power density, equipment power density, cooling and heating set point and a 24-hour occupancy schedule subject to stochastic perturbation. Room schedules were perturbed based on the assigned household behaviour profile to enhance behavioural diversity. Room-level parameters were sampled independently for each room. Among all room types, only bedrooms and living rooms were designated as mechanically conditioned (HVAC-conditioned = 1) with cooling and heating set points applied; all other rooms were modelled as unconditioned.

All parameter assignments were automated using the room name attributes embedded in the BEMs generated in Section 3.1.2, enabling comprehensive coverage of diverse household scenarios and occupant lifestyles.

Level	Parameter
Building level	Exterior wall U-value
	Interior wall U-value
	Floor U-value
	Ground floor U-value
	Ceiling U-value
	Exposed ceiling U-value
	Window U-value
	Door U-value
Household level	ACH
	People density
	Typical occupant behaviour
Room level	Occupant rate (24h)
	Lighting power density
	Equipment power density
	Cooling set point
	Heating set point

	HVAC-conditioned

Table. 3 Energy simulation parameter assignment in different levels

3.2.2 Large-scale parametric energy simulation

After the parameter enrichment of preliminary BEMs, large-scale energy simulations were conducted in the Grasshopper environment using the Honeybee plugin, with EnergyPlus as the simulation engine. The run period covered a full calendar year, and climate conditions were set using Jinan's standard EPW weather file. To accelerate computation, the simulation time step was set to 1 per hour. Room level simulation outputs included annual EUI for cooling, heating, lighting, and equipment (kWh/m^2). All input parameters and simulation results were archived to construct the residential building energy simulation dataset.

3.2.3 Dataset construction

Built from large-scale energy simulations, a comprehensive residential building energy simulation dataset is constructed. The dataset comprises six categories of information:

(1) Indexing data

Room-level indices include: building id, floor, household id, room identifier, room global id, and room id within household. These keys serve as the primary indices for multi-scale retrieval and analysis across the building, household, and room levels.

(2) Room-level simulation parameters

For each room, twelve attributes are recorded: room boundary coordinates, area, exposed area, exterior window area, people density, lighting power density, equipment power density, heating set point; cooling set point, ACH, HVAC conditioned, and a 24-hour occupancy rate.

(3) Envelope thermal properties

For each building, coordinates and corresponding U-values are recorded for exterior and interior walls, exterior and interior windows, exterior and interior doors, floors,

ground floors, roofs, exposed roofs.

(4) Context shading geometry

For each building, the boundary coordinates and heights of shading buildings are recorded.

(5) Household adjacency

Using overlapping boundary geometries, adjacency between households within the same building is computed. For each neighbour pair the dataset stores the neighbour household identifier, shared surface area, adjacency type (wall, floor, or ceiling), and the associated U-value.

(6) Energy simulation outputs

For each room, annual end-use EUI is reported for cooling, heating, lighting, and equipment (kWh/m^2), together with the total EUI aggregated from these components.

The entire dataset is stored in CSV format and supports multi-scale energy modelling at the building, household, and room levels through room-based indexing. It provides a high-quality mapping between spatial configuration, thermal properties, operational schedules, and energy outcomes, forming a robust foundation for analysing the relationship between multi-scale building features and energy performance, and for developing data-driven energy prediction models.

3.3 FusionGNN: A Multi-Modal Graph-Based Energy Prediction Model

To model the spatial, physical, and behavioural complexities of energy use in multi-household residential buildings, this study proposed FusionGNN, a multi-modal graph-based deep learning architecture that takes the household as the basic modelling unit. Room level geometry, schedules, and physical attributes are fused into a household level representation via a Feature-wise Linear Modulation (FiLM) mechanism with attention-based aggregation. A building level graph is then constructed with households as nodes and edge weights encoding spatial adjacency and potential heat exchange. The annual household EUI is predicted by a gated

mixture of a Multi-Layer Perceptron (MLP) head and the graph head, allowing the model to adaptively balance local features and neighbourhood context.

FusionGNN is trained on the residential building energy simulation dataset and calibrated with 120 measured household records. Aggregating household predictions at the building level enables multi-scale evaluation and decision-making from households to buildings.

3.3.1 Data preparation and feature construction

Based on the residential building energy simulation dataset constructed in Section 3.2.3, three categories of features with different dimensions are extracted for each household node to predict the household EUI, including one-dimensional physical features, two-dimensional spatial features, and temporal features reflecting occupant behaviour. Room-level features are aggregated within each household to form household node inputs. In addition, the connections between household nodes and the associated edge features are defined according to their spatial adjacency, resulting in a building graph representation.

Physical features

This study constructs two types of one-dimensional physical features, namely room-level and household-level. These parameters typically do not exhibit continuous spatial variation and are therefore represented as scalar values to support energy-use prediction. A total of 10 room-level physical features is included: area, exposed exterior area, exterior window area, people density, lighting power density, equipment power density, heating set point, cooling set point, ACH, and an HVAC-conditioned indicator. Six household-level physical features are included: household area, household exposed exterior area, household exterior window area, floor level, ground-floor indicator, and top-floor indicator. The household area-related features are computed as the sums of the corresponding room areas within each household. All physical features are normalised by the maximum value observed across the

samples; this normalisation ensures consistency and comparability among different samples in the dataset. For each building sample, the physical features of all rooms are represented as a matrix of shape (K_r, 10), where K_r denotes the number of rooms; the households physical features are represented as a matrix of shape (K_h, 6), where K_h denotes the number of households.

2D spatial features

To capture the 2D building spatial features, this study constructs five rasterised 2D feature maps for each household node: envelope U-value distribution (floor), envelope U-value distribution (ceiling), orientation (sin θ), orientation (cos θ), and context shading building height, as shown in Fig. 17. Each map is represented as a 300 × 300 pixel grid. In addition to these feature maps, a room index map is created to support room-specific feature extraction. The room index map is generated by projecting the polygonal room boundaries onto the raster grid, with each pixel assigned a room ID based on its spatial location.

The two envelope U-value distribution maps are constructed by rasterising the locations of external walls, internal walls, doors, windows, and floors or ceilings as defined in the simulation dataset. Each pixel corresponding to these elements is assigned the corresponding U-value, thereby encoding the spatial distribution of thermal properties.

Orientation channels sin θ and cos θ are constructed per household as a unit direction field over its occupied region. We rasterise all room boundaries to obtain the household occupancy mask and compute a 2D Euclidean distance transform from the mask boundary. The gradient of Euclidean distance is mapped to the world axes to form a normalised direction vector field sin θ and cos θ , where cos θ encodes the west-east component from -1 to 1 and sin θ encodes the south-north component from -1 to 1.

The context shading height map encodes the relative height difference between the current floor level and contextual shading buildings. Pixels corresponding to the floor

plan are set to zero, while obstructing structures are filled with their height differences to represent shading and radiative obstruction effects.

All feature maps are pixel-aligned and stacked along the channel dimension to form a multi-channel spatial input tensor. This structured representation enables the model to learn both local spatial patterns within rooms and global spatial interactions influenced by envelope configuration and external obstructions.

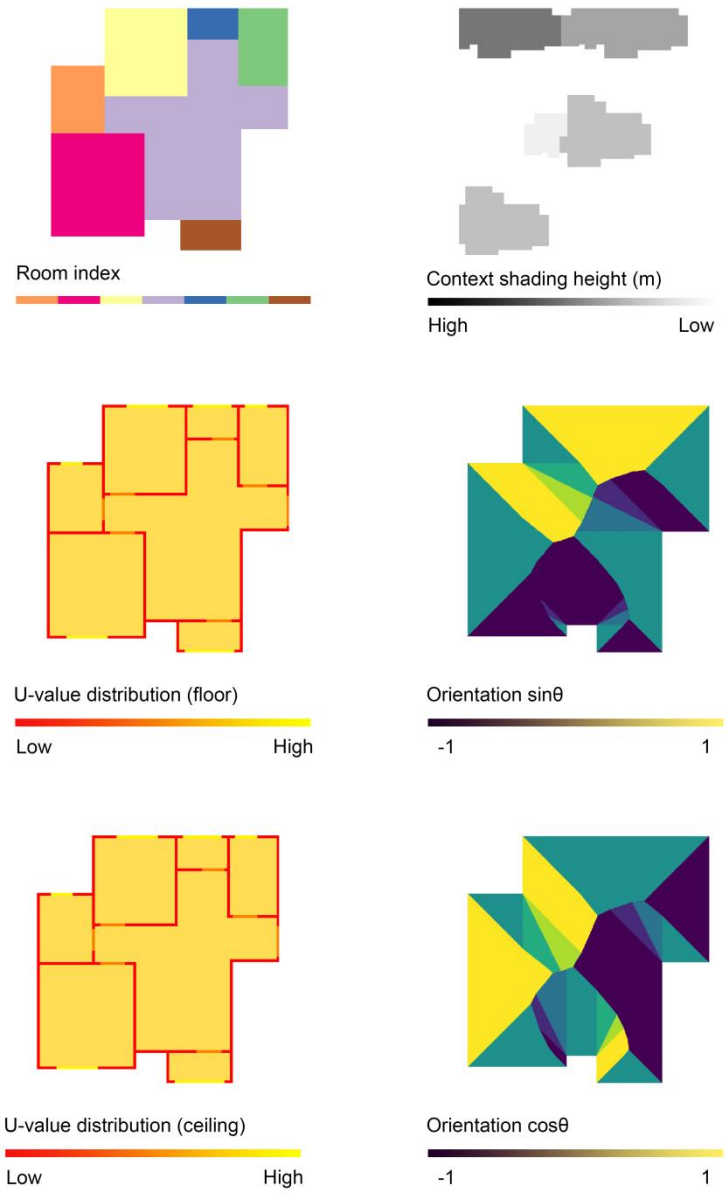


Fig. 17 Five input 2D feature channels and the room index map

Occupant behaviour encoding

To characterise how room usage patterns contribute to household-level energy use, a 36-dimensional temporal behaviour feature vector is constructed for each room, based on the occupant behaviour parameter settings described in Section 3.2.2. This vector consists of a 24-hour occupancy rate vector and a 12-month usage status vector.

The 24-hour occupancy rate vector is a real-valued vector of length 24, where each element represents the average occupancy intensity or probability of presence for a given hour in a typical day, with values ranging from 0 to 1. This feature captures the daily usage pattern of each room.

The 12-month usage status vector is a binary vector of length 12, where each element indicates whether the room is actively used during the corresponding month (1 for active, 0 for inactive), reflecting seasonal usage patterns.

The combined behaviour feature for each room can be represented as:

$$O_r = [o_r^1, o_r^2, \ldots, o_r^{35}, o_r^{36}]$$

These two types of features are generated based on the room's functional type and typical usage patterns, using rule-based and stochastic procedures in the simulation dataset. Together, they capture the diversity and temporal dynamics of occupant behaviour, encoding latent behavioural attributes such as periodicity, peak load hours, and functional variability.

Edge definition

Physical adjacency edges are defined in the household-level graph representation of each building to model direct conductive heat transfer between households. These edges capture structural thermal coupling and include two types of adjacency relationships: (1) lateral adjacency between households on the same floor, connected via shared internal walls, and (2) vertical adjacency between households across

different floors, connected via shared floors or ceilings. Each physical adjacency edge is associated with one key feature: the contact area of the shared building component multiplies its corresponding U-value. Together, edge and the thermal feature quantify the strength of thermal conduction between adjacent households.

3.3.2 Model development

To effectively address the multi-scale and multi-modal characteristics of residential building energy consumption, this study proposes a feature-fusion graph learning framework named FusionGNN. Each building is modelled as a graph where households are treated as nodes and whose edges reflect spatial proximity and potential thermal coupling. For each household, we start from room level information and integrate 2D spatial feature maps, physical features, and occupant behaviour encoding into a single household representation. A graph neural network then performs message passing over the building graph to capture inter-household thermal interactions and contextual dependencies, producing household-level EUI.

The model constructs the household node feature by concatenating two branches: a room-branch fused feature and a household-branch fused feature. In the room branch, masked average pooling over each room mask is performed on the shared CNN feature map to obtain a room geometry vector. This geometry vector is used as the FiLM conditioner to modulate the MLP-encoded 36-dimensional room behaviour feature, which is then fused with an independently MLP-mapped 10-dimensional room physical feature to produce a per-room fused representation. All room features within the same household are subsequently aggregated via attention to form a room-branch household representation. In the household branch, global pooling over the same shared CNN feature map yields a household geometry vector, which is fused with an independently MLP-mapped 6-dimensional household physical feature to produce the household-branch representation. The two branch representations are concatenated into the final household node feature and fed to both an MLP head and a graph head. The graph head uses GraphSAGE, retaining only the top-K most

relevant neighbours per household ranked by edge weights for weighted aggregation, thereby suppressing weak-edge noise and reducing the influence of long distant households. Finally, a gated mixture adaptively weights the predictions from the graph and MLP heads to produce the household's EUI prediction, while the average gate value indicates the relative importance of graph structure versus local household features under the current data.

Multi-Layer Perceptron

For the physical features and occupant behaviour features of each room, a MLP is employed to project the original physical feature vector and behaviour feature into a fixed-dimensional latent representation, facilitating subsequent feature fusion. As described in Section 3.3.1, each room is associated with a 10-dimensional physical feature vector and a 36-dimensional schedule feature, encompassing room-level occupancy rates/states and parameters used in energy simulation. All features are normalised prior to being fed into the MLP to eliminate unit or scale inconsistencies.

Given a vector of occupant behaviour features $r_s^i \in \mathrm{R}^{36}$ for each room i, the MLP projects r_s^i into a latent feature space $f_s^{(room)} \in R^{64}$ through a two-layer MLP:

$$f_{s,\,i}^{(room)} = ReLU(W_s^{(2)} ReLU(W_s^{(1)} r_s^i + b_s^{(2)}) + b_s^{(1)}),\ f_s^{(room)} \in R^{64}$$

Given a vector of physical features $r_p^i \in R^{10}$ for each room i, a single-layer MLP with ReLU projects r_p^i into a latent feature space $f_p^{(room)} \in R^{64}$ through:

$$f_{p,\,i}^{(room)} = ReLU(W_p r_p^i + b_p),\ f_p^{(room)} \in R^{64}$$

These two MLP encoders are complementary: the former focuses on behavioural signals such as temporal patterns and usage intensity, while the latter provides static priors linked to geometry and envelope properties, enabling a joint representation in

the subsequent fusion stage. The MLP mapping embeds the physical features into a unified representation space, enabling them to participate in the subsequent fusion process. This design ensures that the thermal and control characteristics of the building are effectively captured and utilised in the final energy consumption prediction.

Feature Extraction via CNN and Mask Pooling

To capture the spatial structure and envelope thermal performance encoded in the 2D geometrical feature maps, this study designs a shared Convolutional Neural Network (CNN) module that extracts both room-level and household-level spatial features. Each household plan is represented by five spatial channels two U-value maps (U_{floor}, $U_{ceiling}$), two orientation maps ($\sin\theta$, $\cos\theta$), and a context shading height map together with a room index map used only for masking, as described in Section 3.3.1. Based on semantics, the five input channels are processed by two branches that share the same CNN weights: the room branch and the household branch. The Shared CNN first encodes the five channels with two convolutional layers (step=3, stride=2, padding=1) followed by group normalisation and ReLU activation, yielding an output feature map with 32 channels at one-quarter spatial resolution, and the parameters are shared across both branches.

(1) Room branch. We derive a binary mask for each room from the room index map and perform masked average pooling on the Shared-CNN feature map to obtain a room-specific CNN descriptor. In addition, we apply channel-aware mask dilation on the input maps to capture multi-scale context: the two orientation maps use the original room mask (no dilation), the two U-value maps use the mask dilated by 5 pixels (to capture near-neighbour envelope conditions), and the context shading height map uses the mask dilated by 25 pixels (to capture long distance shading distribution). We compute masked means from these three scales and concatenate them with the room's CNN descriptor to form the room-level geometry feature, denoted as $f_{geo}^{(room)} \in \mathrm{R}^{128}$. (2) Household branch. On the household side, we apply

global average pooling over the Shared-CNN feature map to produce the household-level geometry feature, denoted as $f_{geo}^{(household)} \in \mathrm{R}^{128}$. This provides a compact summary of the household's overall spatial organisation and envelope condition, complementary to the fine-grained room representation.

Room-level FiLM feature fusion

To characterise room occupant behaviour over time and the associated physical attributes, this study independently encodes the room-level schedule features and the room-level physical features, and then use the Feature-wise Linear Modulation (FiLM) to integrate geometric context for multimodal fusion.

After obtaining $f_s^{(room)} \in R^{64}$ and $f_p^{(room)} \in R^{64}$ for each room, we apply FiLM to inject geometric context into the temporal schedule representation. Specifically, we take the room-level geometry feature as the conditioning input. A lightweight gating MLP generates channel-wise scale and bias parameters that match the dimensionality of $f_s^{(room)}$, and we apply feature-wise scaling and shifting to obtain the geometry-conditioned temporal representation $\tilde{f}_s^{(room)}$. This preserves dimensionality while allowing identical schedules to be expressed differently under varying spatial thermal and orientation distributions.

Finally, $\tilde{f}_s^{(room)}$ is concatenated with $f_p^{(room)}$ to form the final room-level fused representation $h_r^{(room)} \in R^{128}$, which jointly captures temporal usage patterns from the schedule, simulation parameters from the physical features, and geometry-conditioned differences introduced via FiLM.

Household-level Feature Fusion

In the household-level branch, we use a CNN encoder that shares parameters with the room branch to perform convolutional down sampling and semantic extraction over each household's five-channel planar inputs. We then apply global average

pooling to the resulting feature map to obtain a global channel-wise summary. To align with the model's common representation scale, this summary is passed through a linear layer, ReLU, and LayerNorm and projected to a unified geometric dimensionality of 128, yielding the household-level geometry feature.

Next, the household-level physical feature vector is 6-dimensional and is fed into an independent household physical encoder. This encoder has the same structure as the room physical MLP (a single-layer MLP followed by ReLU) but does not share parameters with it, in order to avoid semantic interference arising from the different scales of room- and household-level signals.

In parallel, the room branch has already produced a per-room geometric representation using masks and multi-scale geometric statistics, and it derives attention weights within each household via a learnable scoring function. These weights are used in two consistent ways: first, to compute a weighted aggregation of room geometry, forming the room geometry fused vector; second, to apply the same weights to the 10-dimensional room physical features, producing a weighted average that is subsequently mapped by the room physical encoder (single-layer MLP + ReLU) to obtain the room physical aggregated embedding (64D). This attention consistency ensures that rooms emphasised by the model in geometric space are simultaneously emphasised in the physical statistics, improving cross-scale coherence and interpretability between room- and household-level information.

Finally, at the household node we concatenate four components: the room geometry fused vector (128D), the household geometry feature (128D), the household physical embedding (64D), and the room physical aggregated embedding. This yields an information-dense, scale-aligned node representation with a default total dimensionality of 384. The representation preserves fine-grained room heterogeneity while capturing the household's global context, providing a stable, interpretable, and generalisable input to the subsequent Top-K graph aggregation and parallel MLP branches.

Graph Neural Network

This study constructs a building-level graph with households as nodes and inter-household thermal couplings as edges. Each node corresponds to one household, whose feature is a 384-dimensional vector that fuses room-level and household-level information. Edge weights come from a pre-computed adjacency matrix (adjacent area multiplied by adjacent component U-value), reflecting the strength of thermal coupling due to spatial adjacency and shared envelope components. Graphs are built independently for each building (no cross-building edges), enabling modelling of inter-household energy interactions within a consistent architectural context.

To reduce computation and overfitting, we use a weighted-mean GraphSAGE as the message-passing core: for each node, we keep only the Top-K neighbours (K=4) with the largest weights, first apply a linear mapping to the node features, then perform a linearly weighted sum using row-normalised edge weights, and add a residual self-projection; a ReLU activation is applied after the aggregation layer. Compared with using the full adjacency, this design markedly lowers training complexity and naturally focuses information flow on closer or more strongly coupled units, avoiding noise from distant or weakly coupled nodes.

The resulting 40-dimensional hidden representation is then fed into a lightweight two-layer MLP to produce a scalar regression of the household's annual EUI.

Training Process

This study adopts an end-to-end training strategy to predict household-level annual EUI on building-level graphs. The dataset is generated from large-scale energy simulations and is randomly split by building into 80% for training and 20% for validation. During training, we use a batch size of 24 buildings per step, with each building containing a variable number of household nodes. The shared CNN and the graph model are optimised jointly.

We optimise a robust regression objective consisting of MAE + 0.1 $\times$ Huber, using

Adam with an initial learning rate of 0.001 and L2 weight decay 0.0004. Training runs for 100 epochs, with early stopping triggered if the validation loss does not improve for 20 consecutive epochs. Experiments are conducted on a CUDA device (NVIDIA RTX 4070 GPU). To improve computational efficiency and memory usage, we enable AMP mixed precision, performing forward/backward passes and normalisation on the GPU. Model evaluation is performed at the household level using three regression metrics, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination R^2. Let y_i denote the ground-truth annual household EUI, $\hat{y}_i$ the prediction, and $\bar{y}$ the mean of y_i over the evaluated set of N households. The metrics are computed as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}} (\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i})^2}$$

$$\mathrm{MAE} = \frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}} |\mathrm{y_i} - \hat{\mathrm{y}}_\mathrm{i}|$$

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

MAE reflects the average absolute deviation in the original units of EUI, while RMSE penalises larger errors more strongly due to the squared residuals. R^2 measures the proportion of variance in the ground-truth EUI explained by the model relative to a mean-only baseline. An R^2 value of 1 indicates perfect agreement, 0 indicates no improvement over predicting $\bar{y}$, and negative values indicate worse performance than the mean baseline.

3.3.3 Model calibration

While simulation-based data provides comprehensive and fine-grained inputs for model training, it may deviate from actual building operation due to assumptions in material properties, occupant behaviour, or environmental conditions. To improve the model's generalisability, a transfer learning-based calibration strategy was developed. This strategy utilises a limited set of real household energy consumption data, collected through on-site questionnaire surveys, to fine-tune the pretrained FusionGNN model for enhanced performance in real-world applications.

On-site questionnaire survey

To collect realistic household usage patterns and energy consumption data for model calibration and validation, this study conducted an on-site questionnaire survey targeting occupants of existing residential buildings. The survey aimed to capture both quantitative and qualitative factors that influence energy demand at the household level. Specifically, the questionnaire (Appendix A) covered the following aspects:

- Annual electricity information collected from utility bills, including annual electricity bills and, where available, self-reported annual electricity use (kWh);
- Floor plan layout and functional zoning of the household;
- Number of occupants and their typical daily occupancy schedules, including working hours, stay-at-home rates, and room-level activity distribution;
- Usage frequency and temperature settings of air conditioning and heating systems during different seasons;
- Ownership, power ratings, and usage frequency of major household appliances such as refrigerators, washing machines, lighting systems, and kitchen devices.

The questionnaire was administered through in-person, structured interviews during field visits to improve completeness and reduce missing entries. Recruitment was conducted through intercept invitations in public squares and communal areas near ageing residential estates. The sampling frame was the 64 ageing residential communities identified in Jinan's 2020 retrofit plan (Section 3.1.1), which also

underpin the simulation dataset. To be included, participants needed to be able to provide annual household electricity bills, have lived in the dwelling for at least one year, and use the home as their primary residence (i.e., non-commercial use). Residents who did not meet these requirements, or could not provide reliable annual bill information, were excluded. No financial or material incentives were offered. The survey compiled annual electricity use from utility bills, floor plan layout and functional zoning, number of occupants and daily occupancy schedules, heating and cooling set points and usage frequency, and the ownership and usage frequencies of major appliances. To align with the simulation-driven training data, all samples were mapped to FusionGNN's household-level inputs, including room geometry and functional assignment, 36-dimensional room-level activity schedules, room ACH, lighting and plug-load intensities, and heating/cooling set points. Household building geometry was reconstructed by asking participants to select the best-matching dwelling layout from a curated library of floor plans collected for their residential community, thereby enabling consistent extraction of room geometry and functional zoning. Urban context was further matched by linking each household to its residential block and generating surrounding context using Jinan building footprint data, in order to derive exposure and shading-related inputs in a manner consistent with the FusionGNN feature definition. Envelope thermal parameters were obtained by querying the publicly released retrofit specifications in Jinan's 2020 ageing residential community energy retrofit programme, which document envelope upgrade measures and the corresponding thermal-performance targets. Finally, the annual household EUI was calculated by using self-reported annual electricity use (kWh) when available; otherwise, it was estimated from the annual electricity bills by applying the local electricity tariff to convert annual expenditure to electricity use, and then normalised by the reported household floor area.

Ethics approval and data management

Ethical approval was granted by the School of Architecture and Landscape, University

of Sheffield. All participants received an information sheet (Appendix B) and signed a consent form (Appendix C) prior to participation. No direct identifiers, such as addresses, phone numbers, or email addresses, were collected in the dataset, and survey responses were recorded in written or digital form. To align with simulation inputs and enable context matching, the survey recorded only high-level location and household descriptors, and all such descriptors were stored as non-identifiable codes rather than original labels. No household-level identifiers that could be used to trace records back to an individual dwelling within a community were retained in the analysis dataset; the dataset is therefore treated as anonymised. Consent forms were stored separately from survey responses, and access to raw data was restricted to the author and supervisors. Raw files were stored on the author's local computer. Data will be retained for 5 years and then securely deleted.

Transfer learning

To enable supervised fine-tuning and fair generalisation assessment of the proposed FusionGNN, 120 households were split into 100 for calibration training and 20 for held-out testing. Because these households come from different buildings, complete building-level graphs could not be formed. Each household was therefore modelled as an isolated node. During calibration and inference, graph message passing was disabled by setting all edge weights to zero, so the model predicts from independent node-level features. The physical, geometric, and behavioural attributes of each room were encoded from the survey questionnaire and household geometry and then mapped to the input format required by FusionGNN.

To avoid over-fitting and preserve pre-trained representations, all backbone parameters are frozen, including the shared CNN, the room-level and household-level physical encoders, the FiLM fusion head, and the Top-K SAGE layer, with only the GNN regression head updated. Calibration minimises the mean squared error (MSE) between predicted and measured household EUI:

$$\mathcal{L}_{\text{calib}} = \frac{1}{\mathrm{N}} \sum_{\mathrm{h}=1}^{\mathrm{N}} \left(\hat{\mathrm{y}}^{(\mathrm{h})} - \mathrm{y}_{\text{real}}^{(\mathrm{h})} \right)^2$$

where N is the total number of household samples used for calibration, $\hat{\mathrm{y}}^{(\mathrm{h})}$ is the predicted household EUI and $\mathrm{y}_{\text{real}}^{(\mathrm{h})}$ is the measured household EUI. To assess the effectiveness of the calibration strategy, the model's performance before and after fine-tuning was quantitatively evaluated on the test set using RMSE and MAE computed at the household level.

3.4 Case Study

To evaluate the practical applicability of the proposed model in real-world residential energy retrofit decision-making, this study selects four representative residential floor plan typologies as case studies. The FusionGNN model is applied to predict and analyse the energy performance under various retrofit scenarios, and the associated costs of each strategy are calculated. A multi-objective optimisation framework is then employed to support decision-making by balancing energy-savings and economic investment.

3.4.1 Energy performance evaluation under different retrofit scenarios

Retrofit scenarios

To systematically evaluate how different floor plan typologies affect the prioritisation of retrofit measures, this study conducts an optimisation analysis using the four typical residential floor plan layouts identified in Section 3.1.1 as case studies. Reflecting the floor level distribution commonly observed in ageing residential estates, all four comparison buildings are set to six stories. The four representative floor plans and their corresponding visualised energy models are shown in Fig. 18. To analyse the prioritisation of retrofit strategies across layouts, the four cases adopt a consistent set

of 1990s envelope thermal properties and operational parameters based on Deng et al. (2025)’s research of different construction period building characteristics in Jinan, as summarised in Table. 4. Occupant behaviour and schedules within each building follow the "office-worker household" configuration defined in Section 3.2.1.

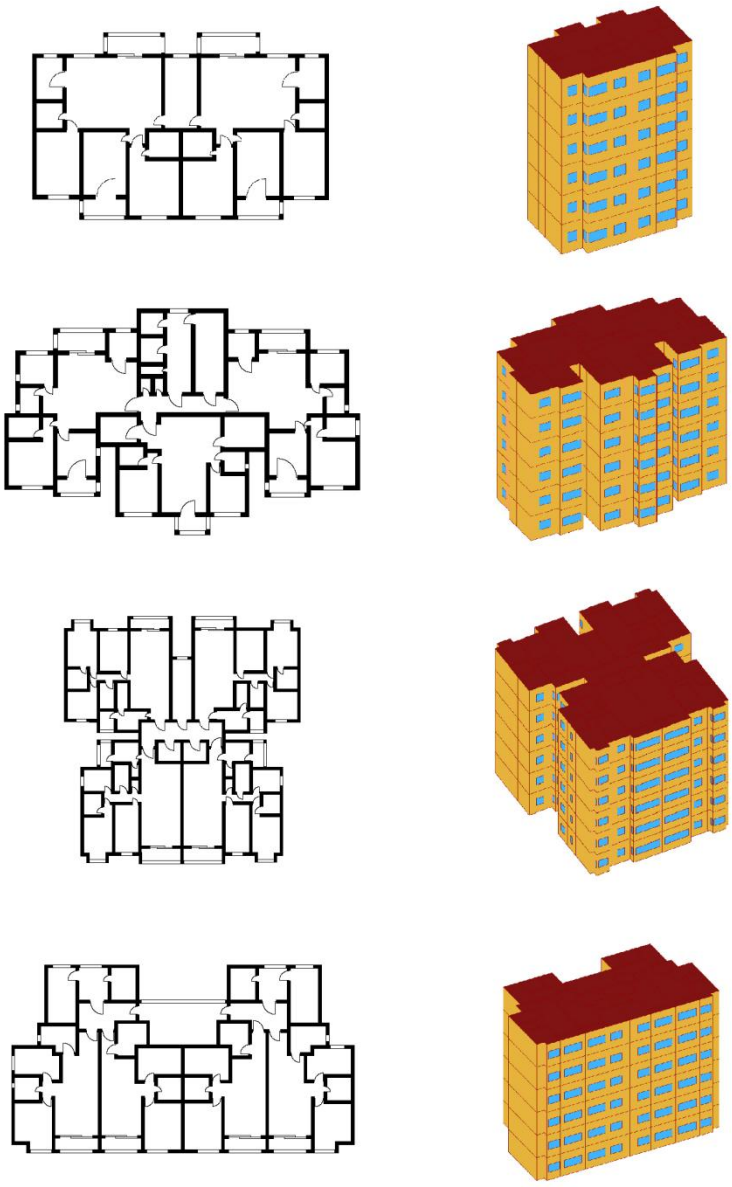


Fig. 18 Four typical residential floor plan typologies and corresponding energy models

Parameter	Baseline setting
Wall	20mm Cement-sand mortar 360mm Clay brick 20mm Internal insulation mortar
Exterior Roof	20mm Cement-sand mortar 120mm Reinforced concrete slab 50mm PVC board 20mm Cement-sand mortar
Window	12 plastic steel windows
People density	0.04 person/㎡
ACH	0.5
Lighting power density	5W/㎡
Equipment power density	10 W/㎡
Cooling set point	26℃
Heating set point	18℃

Table. 4 Building energy-related parameter configurations

Based on the retrofit measures defined in the referenced study, a set of representative energy retrofit scenarios was formulated. These include the addition of external wall insulation, roof insulation, and window replacement. The parameters and cost assumptions of each measure are summarised in Table. 5 and Table. 6, and a unique code is assigned to each scenario for subsequent analysis.

Code	Material	U-value ($W/m^2 \cdot K$)	Density (kg/m^3)	Specific heat capacity ($J/kg \cdot K$)	Cost for wall/roof (CNY/m^2)	Reference
M1	EPS	0.36	20	1380	27.29/35.02	(Guo et al., 2021)
M2	XPS	0.32	29	1500	28.12/37.50	(Guo et al., 2021)
M3	PU	0.27	30	1380	32.50/45.52	(Guo et al., 2021)
M4	PF	0.20	50-80	1400	22.5/33.40	(Zheng et al., 2025)
M5	RW	0.23	140	1220	36.96/46.20	(Guo et al., 2021)

Table. 5 Insulation materials for building energy retrofit

Code	Material	U-value ($W/m^2 \cdot K$)	SHGC	Price (CNY/m^2)	Reference
W1	Low-E film coated hollow (low transmittance type)	2.1	0.426	500	(Liu et al., 2025)
W2	6 mm Clear glass + 12 mm Air + 6 mm Transparent glass	2.4	0.487	350	(Liu et al., 2025)
W3	Ordinary insulating glass (6+12A+6)	2.9	0.722	323	(Liu et al., 2025)
W4	6 mm Medium transparent glass + 12 mm Air + 6 mm	3.1	0.722	450	(Liu et al., 2025)

	Transparent glass plastic				
W5	Plastic steel hollow window	2.3	0.513	320	(Liu et al., 2025)
W6	Broken bridge aluminium Hollow window	2.2	0.624	627	(Guo et al., 2021)
W7	Broken Bridge Aluminium Vacuum Window	2.0	0.455	720	(Guo et al., 2021)
W8	Plastic Steel Vacuum Window	2.6	0.523	460	(Guo et al., 2021)

Table. 6 Windows for building energy retrofit

To comprehensively evaluate the impact of different retrofit strategies on building energy performance, this study generates a wide range of retrofit scenarios based on a set of representative energy-saving measures listed in Table. 5 and Table. 6. A flexible combination strategy is adopted, allowing each scenario to include any subset of 1 to 3 retrofit categories. To represent subset selections within a unified three-variable formulation, each decision variable includes a baseline no-retrofit option. Consequently, a package that upgrades only one category can be expressed by selecting the baseline option for the remaining categories. This retrofit scenario space captures a wide range of strategies, from light retrofits (e.g., window replacement only) to deep retrofits involving the simultaneous upgrade of envelope. It enables a detailed analysis of the nonlinear trade-offs between retrofit investment and energy-savings. Moreover, the diversity of this scenario space ensures that the subsequent multi-objective optimisation process can explore a sufficiently rich solution space to identify optimal retrofit strategies under varying economic and performance constraints. For clarity and reproducibility, the scenario coding scheme

used to label each retrofit option is summarised in Table. 7, where wall insulation, roof insulation, and window performance measures are represented using standardised code notations (e.g., M1 – M5 and W1 – W8) to support systematic scenario generation and comparison.

No.	Parameter	Code Notation
1	Wall insulation (U-value)	M1 / M2 / M3 / M4 / M5
2	Roof insulation (U-value)	M1 / M2 / M3 / M4 / M5
3	Window performance (U-value)	W1 / W2 / W3 / W4 W5 / W6 / W7 / W8

Table. 7 Retrofit scenario codes

FusionGNN prediction

The FusionGNN model is employed to predict the energy performance of buildings under different retrofit scenarios. Each building is represented as a graph constructed from its original floor plan, using the baseline envelope constructions and operational parameters specified in Table. 4 for the 1990s settings, and parameterising retrofit options with the envelope U-values listed in Table. 5 and Table. 6, together with occupant behaviour and power density. For each retrofit scenario, the feature vectors of relevant room nodes are updated based on the selected combination of retrofit measures, including external wall insulation, roof insulation, and window type. For example, replacing windows modifies the glazing U-value, while applying insulation alters the U-value of exterior walls or roofs. The spatial connectivity among rooms (graph edges) remains unchanged across scenarios, allowing rapid prediction by updating only the node attributes. The modified building graph is then input into the FusionGNN model for forward inference, which outputs the annual EUI of each household. The overall building-level EUI is subsequently calculated by taking a floor area weighted average of household-level EUIs.

3.4.2 Retrofit cost calculation

To quantitatively evaluate the economic performance of each energy retrofit strategy, this study estimates the initial investment cost based on envelope material unit prices reported in the literature, thereby providing consistent economic indicators for subsequent multi-objective optimisation analysis. The initial investment cost of a retrofit scenario represents the upfront expenditure required to implement a specific set of retrofit measures. In this thesis, only envelope material costs are considered, including external wall insulation, roof insulation, and window replacement. To remain consistent with the unit-price data available in the literature and to ensure fair comparability across scenarios, other cost items are excluded, such as labour, scaffolding, transportation, design and management fees.

Envelope-related measures are evaluated based on the product of retrofit area and the corresponding unit price of materials (e.g., in CNY/m^2). The overall investment cost for the k th retrofit scenario is thus calculated as:

$$IC_k = \sum_{i=1}^{a} A_i \cdot c_i$$

where A_i is the area of the i th envelope component (e.g., wall, roof, window), and c_i is its unit cost (CNY/m^2); a is the total number of materials considered in the scenario.

3.4.3 Multi-objective optimisation

To achieve a balanced trade-off between energy performance and initial economic investment, this study proposes a multi-objective optimisation framework for residential building energy retrofitting. The framework integrates the previously trained FusionGNN model for energy consumption prediction and a cost model to evaluate retrofit scenarios from both thermal and financial perspectives. The two optimisation objectives are:

- Minimise post-retrofit annual EUI;
- Minimise initial investment cost;

Each retrofit scenario is defined by three decision variables: external wall insulation, roof insulation, and window type. These variables form a combinatorial search space, with each unique combination representing a feasible retrofit package. To evaluate the performance of each candidate solution, the FusionGNN model is employed to predict household-level EUIs, which are then aggregated to yield building-level energy consumption. In parallel, the retrofit cost is computed based on unit pricing and material area for each upgrade measure. To solve this high-dimensional, non-linear Multi-Objective Optimisation (MOO) problem, the Non-dominated Sorting Genetic Algorithm II (NSGA-II) is adopted. NSGA-II is particularly well-suited for large discrete design spaces and ensures both convergence and diversity of solutions across the Pareto front. In the context of this study, Pareto-optimal solutions refer to retrofit packages for which no other option can improve one objective (e.g., energy-savings) without worsening another (e.g., cost). The solutions on the Pareto front are therefore considered the most efficient trade-offs, from which decision-makers can select strategies according to project-specific constraints or policy goals.

4. Results

4.1 Results of Automated Building Energy Modelling

4.1.1 Architectural element recognition

This study modified the DMTN model proposed by Zeng et al. (2019) for floor plan segmentation, with the classification performance results summarised with the Confusion Matrix in Table. 8. Given the focus on accurately identifying architectural elements from floor plan images to support BEM geometry construction, we evaluate segmentation performance at the object level. Specifically, a predicted object is considered correctly classified if it overlaps by at least 80% with its corresponding ground-truth annotation. The multi-household residential floor plan dataset used in

this study includes a total of 3,340 annotated room regions, 3,606 windows, and 3,780 doors across all floor plan images. To assess model performance, the evaluation metrics include precision, recall, and F1 scores, which respectively reflect the accuracy, completeness, and overall effectiveness of architectural element recognition.

As shown in Table. 8, room region detection achieves the highest precision of 0.93, the highest recall of 0.89, and the highest F1 score of 0.91. This indicates that the modified DMTN model can effectively identify room regions, providing a reliable foundation for the construction of thermal zones in BEM. For window detection, the precision reaches 0.87, suggesting that most of the predicted window objects are correct. However, the relatively low recall of 0.62 indicates that a considerable number of windows were not successfully identified. For door detection, the recall reaches 0.79, reflecting the model's ability to detect most actual doors, while the precision is relatively lower at 0.71, indicating the presence of misclassified door objects. Misclassifications of doors and windows can change the connectivity relationships between rooms, thereby affecting subsequent household clustering and room type classification. Moreover, errors in door detection can directly impact room region recognition. Since doors often serve as key separators between adjacent rooms, missing a door may result in two distinct rooms being incorrectly merged into a single region, which can reduce the recall rate for room regions. Overall, the model shows robust performance in room region recognition, with comparatively lower performance in detecting smaller architectural components such as windows and doors. The F1 scores for windows and doors are 0.72 and 0.75, respectively, suggesting comparable overall recognition performance for these elements, despite differences in precision and recall due to varying patterns of missed detections and misclassifications.

	Precision	Recall	F1 score
Room region	0.93	0.89	0.91
Window	0.87	0.62	0.72
Door	0.71	0.79	0.75

Table. 8 Precision, recall and F1 score of architectural element segmentation (modified DMTN)

4.1.2 Multi-household reconfiguration

Core node classification and the CTRC algorithm are employed for multi-household room configuration. The results show an average core nodes classification accuracy of 95.3% and 98.5% for household clustering, highlighting the effectiveness of the proposed method in handling complex multi-household residential layouts. A DT classifier was trained to classify core nodes and evaluate 187 core nodes, with the classification confusion matrix shown in Table. 9. Notably, the DT classifier achieves an optimal accuracy of 1.0 for household nodes, which is particularly significant for the subsequent step of room type classification that relies on accurate household cluster identification. A small number of communal and household function nodes were misclassified, indicating limited confusion between these two categories. In particular, a few communal nodes were incorrectly classified as household nodes.

	Communal	Household	Function
Communal	0.89	0.04	0.07
Household	0	1.0	0
Function	0.03	0	0.97

Table. 9 Confusion matrix of the core node classification

Fig. 19 illustrates some clustering results for multi-household residential floor plans with the number of households ranging from 2 to 6. Based on the high classification accuracy of core nodes, the CTRC algorithm can effectively reconfigure rooms and

separate communal and household clusters. However, some misclassifications in core node resulted in incorrect household cluster assignments. For instance, M1 in Fig. 19 shows rooms that should belong to a communal cluster being incorrectly grouped into a household cluster, while M2 highlights a communal cluster misclassified as a household cluster. These examples indicate that occasional classification errors in core nodes can affect the accuracy of household clustering.

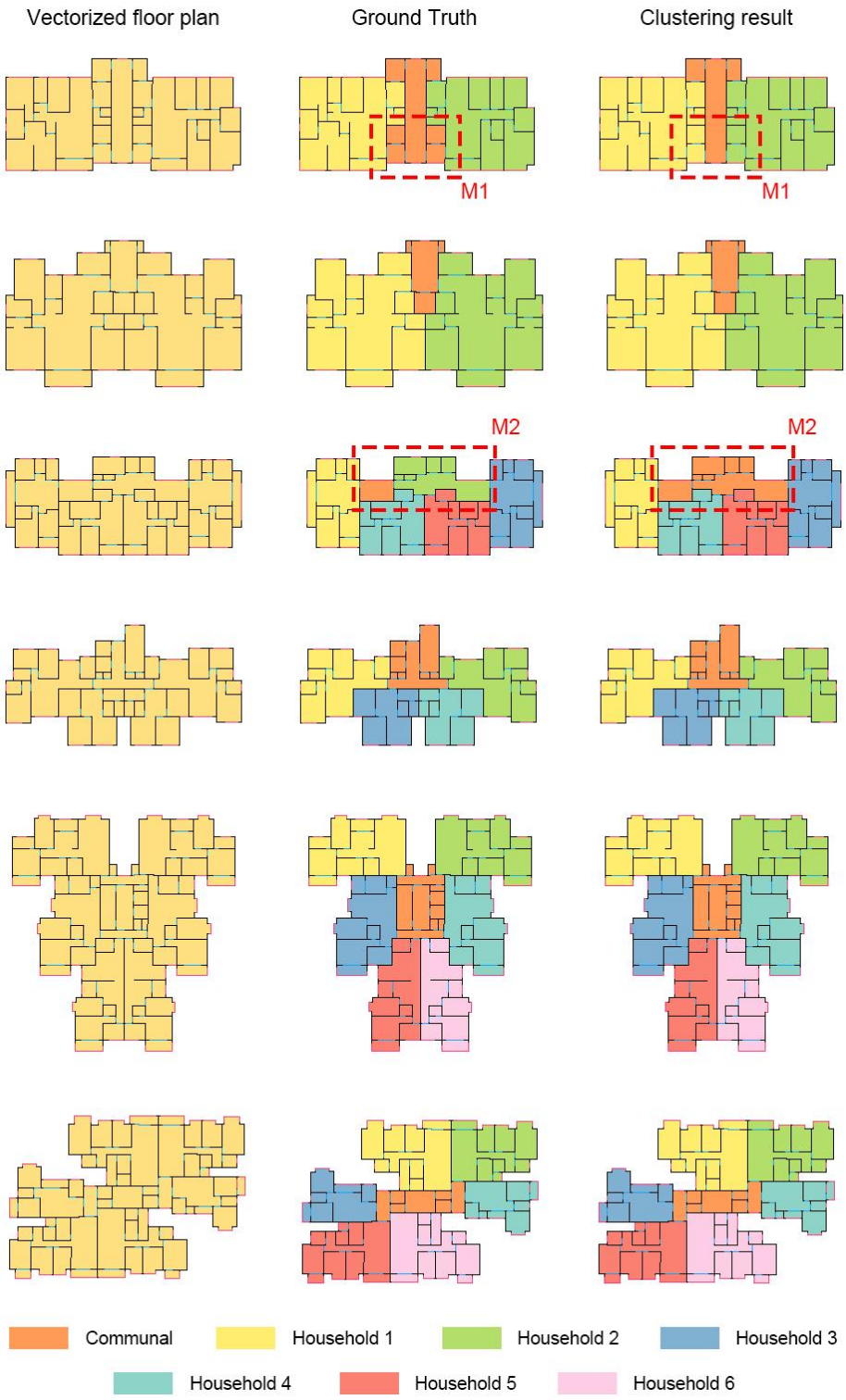


Fig. 19 Multi-household reconfiguration result using the CTRC algorithm (left: vectorised floor plan, middle: room cluster ground truth, right: household clustering result through CTRC)

4.1.3 Room type classification

Table. 10 presents the confusion matrix for classifying room types using the GAT model on 572 room nodes across 80 household graphs, with an overall classification accuracy of 93.7%. Among the six room types, the living room achieves the highest classification accuracy of 1.0, showing no misclassifications, which indicates the model's capability to identify its unique topological position within household layouts. Bedrooms, bathrooms, and balconies also show high classification accuracies of 0.94, 0.95, and 0.97, respectively. Among these misclassifications observed, 3% of bedrooms are misclassified as balconies, and 2% as bathrooms, while 3% of bathrooms are misclassified as balconies. Kitchen and utility rooms exhibit slightly lower classification accuracies, at 0.87 and 0.89, respectively. Specifically, 13% of kitchens are misclassified as bedrooms and 11% of utility rooms are misclassified as bathrooms, suggesting some overlap in their connectivity patterns or physical features in certain household layouts.

	Bedroom	Living	Kitchen	Bathroom	Balcony	Utility
Bedroom	0.94	0.02	0	0.02	0.03	0
Living	0	1.0	0	0	0	0
Kitchen	0.13	0	0.87	0	0	0
Bathroom	0	0	0.03	0.95	0.03	0
Balcony	0.03	0	0	0	0.97	0
Utility	0	0	0	0.11	0	0.89

Table. 10 Confusion matrix of the room type classification

4.1.4 Automated energy modelling

In this study, a total of 120 building floor plans were evaluated through energy simulations conducted across three groups. The Ground Truth group was entirely constructed manually and served as the baseline for comparison. Group 1 retained

manually defined building geometries but automated the processes of household clustering and room type classification. Group 2 employed a fully automated workflow, including geometry creation through floor plan segmentation and vectorisation, and room schedule assignment through semantic analysis. Fig. 20 summarises the relative errors and sample proportions of Group 1 and Group 2 compared to the Ground Truth. Group 1 achieved an average annual EUI error of 0.043, with over 88.7% of samples exhibiting errors within 8% of the Ground Truth. The error distribution of Group 1 was highly concentrated, with only a few outliers. In contrast, Group 2 reported a higher mean EUI error of 0.191, with 62.1% of samples exhibiting errors within 20%, and a more dispersed error distribution that included several extreme deviations.

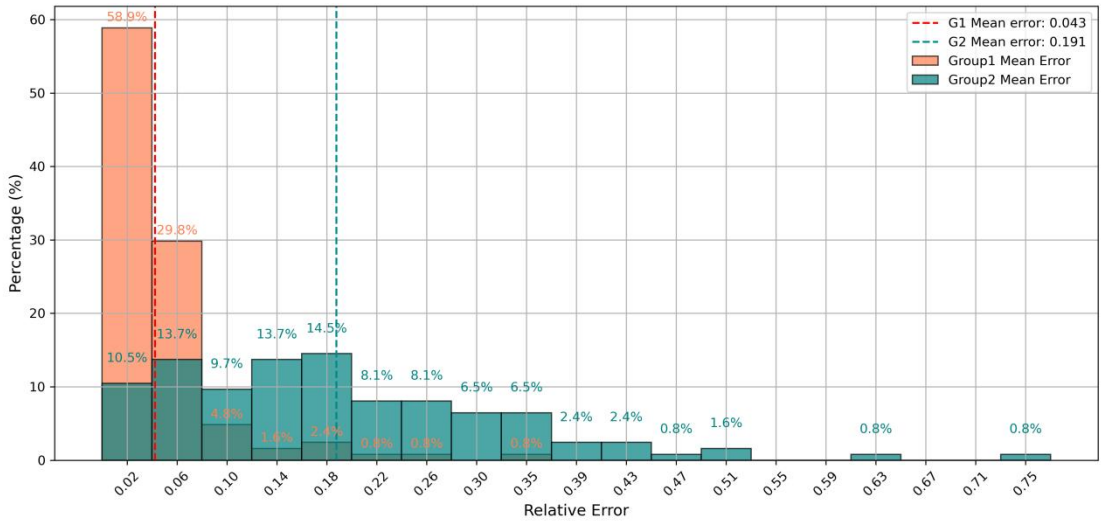


Fig. 20 Relative error statistics of Group1 and Group2 against the ground-truth models

4.2 Results of Large-scale Parametric Energy Simulation

Building on automatically generated BEMs and subsequent model enrichment, this study conducted a large-scale energy simulation to assemble a high-quality residential building energy simulation dataset for energy analysis of multi-household residential buildings. The study reconstructed and simulated 4,282 residential buildings across 64 old residential communities in Jinan, China, applying room-level

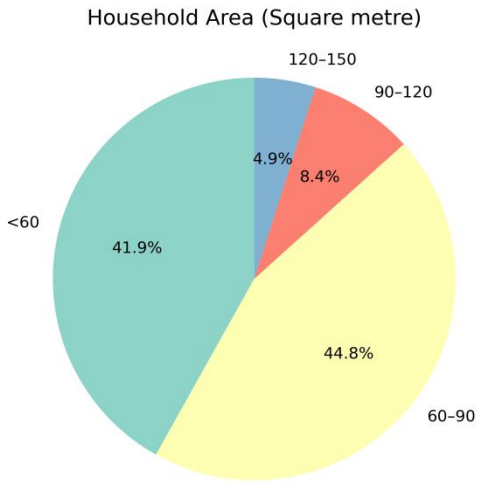
thermal zoning inferred from floor plans and ultimately covering 89,739 households and 650,019 rooms; among these rooms, 74.1% have heating/cooling demand, while 25.9% are communal or functional spaces (e.g., balconies and bathrooms) for which only lighting and equipment energy were simulated. Annual simulations were run in Grasshopper using the Honeybee plugin, recording inputs and outcomes at the room, household, and building levels, including building geometry, surrounding shading context, envelope thermal properties, inter-apartment connectivity, room internal loads, and occupant schedules, together with the corresponding annual EUIs for cooling, heating, lighting, and equipment. To better reflect real-world operational diversity, we determined physically plausible parameter ranges and step sizes from the literature, thereby ensuring the simulations span a wide space of envelope and behavioural parameters.

4.2.1 Overview

The large-scale residential building energy simulation dataset is organised into six parts: indexing data, room-level simulation parameters, envelope thermal properties, urban shading context, household adjacency, and energy simulation results. The urban context records the distribution and heights of surrounding buildings within a radius of 30 m for each simulated building. The simulation parameters and energy simulation results section stores each room's boundary coordinates, twelve physical features (including the room's operating schedule), and four simulated energy outcomes. The household adjacency captures the connections between households within each building, including the adjacency matrix, the shared area, and the connection type (floor slab or wall). The envelope thermal behaviour part provides the start coordinates of doors, windows, and walls, distinguishing interior and exterior, together with the U-values of ten building components. All four files use a unified indexing scheme consisting of building ID, household ID, and room ID, which enables fast retrieval and extraction of information at multiple levels of granularity.

To assess the correspondence between the simulated buildings and the old

residential estates, we compiled statistics on household floor areas, building storey counts, and floor plan layouts of the 4,282 buildings in the dataset. Regarding the apartment floor-area distribution (Fig. 21a), 64.1% of units fall within 60–90 m², followed by 14.7% within 90–120 m², and only 1.7% exceed 150 m², indicating that most households are relatively compact. The building floor-count distribution (Fig. 21b) shows that 81.4% of buildings have 1–9 floors, and only 7.1% exceed 13 floors, reflecting the dominance of low to mid-rise stock in the sample. In terms of household layout configuration shown in Fig. 21c, one core with two units accounts for 39.0%, one core with three units for 42.0%, while one core with four units and two cores with four units account for 11.2% and 5.3%, respectively. Overall, the architectural characteristics of the simulated dataset closely match those documented for older residential communities described in Section 3.1.1, indicating that the dataset is representative of typical legacy estates. All records adopt a consistent indexing scheme across building, apartment, and room levels, supporting spatial energy use analyses at multiple granularities and providing a reliable basis for training and validating predictive models for energy-saving assessment and retrofit strategy optimisation.



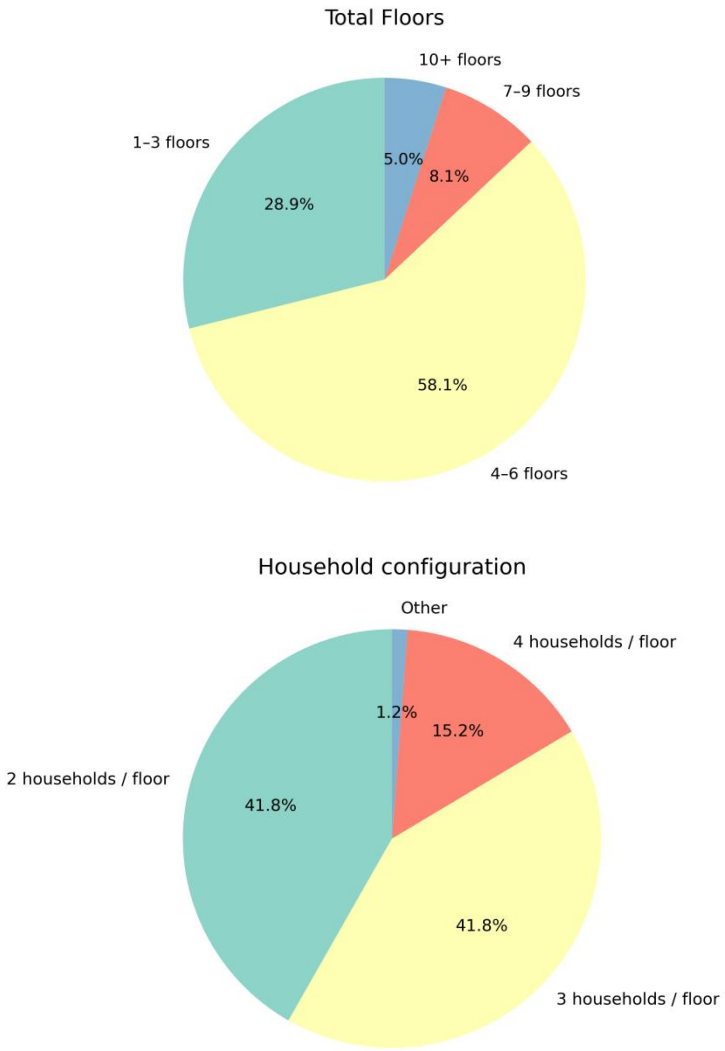


Fig. 21 Statistics of household area, total floors and household configuration in the simulation dataset

4.2.2 Parameter distribution

Shading building distribution

For each simulated building, the distribution and heights of surrounding shading buildings were recorded. For every shading building, the polygon vertices of its plan footprint and its height were recorded. Using these data, an orientation-resolved analysis was conducted. The building centre was taken as the origin, with 0° and 180° defined as west and east, respectively. Over 0–360°, two metrics were computed for equal-angle sectors: the obstruction area ratio (the fraction of the sector covered by

plan projections of shading buildings) and the average obstruction height (the mean height of shading buildings within the sector). The results are shown in Fig. 22. The statistics reveal pronounced directional differences. The obstruction area ratio varies periodically with angle, reaching higher values towards the west and east where coverage exceeds 0.40, while the south and north directions are typically below 0.20. The average obstruction height exhibits a similar pattern: in the east and west sectors it is generally around 16–17 m, whereas in the north and south sectors it is usually below 15 m. These findings indicate that shading buildings are more numerous and typically taller on the east and west sides, suggesting a predominant east–west parallel building block arrangement.

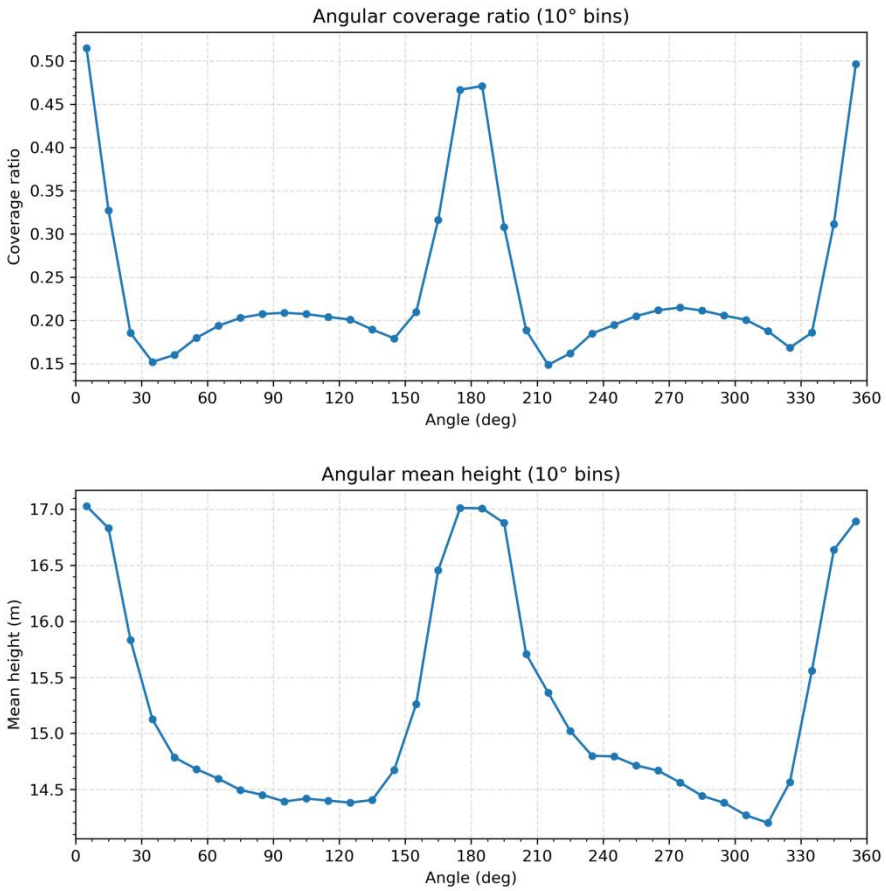


Fig. 22 Statistics of shading building coverage ratio and mean height across different orientations

Simulation parameter distribution

Table. 11 reports the statistical distributions of all physical input parameters. For the envelope thermal properties, exterior walls, interior walls, floors, and doors exhibit relatively modest dispersion, with interquartile ranges (IQRs) of approximately 0.8, 0.8, 1.1, and 1.0 W/m²·K, respectively. By contrast, ceilings and windows show higher variability: standard deviations are 0.95 and 3.40 W/m²·K, and the IQRs span 1.6 and 2.6 W/m²·K, respectively. The U-values of envelops cover a continuous spectrum from uninsulated constructions to insulated assemblies. Operational parameters display comparable diversity. The ACH has an IQR of 0.90–2.30 h^{-1} with a mean of 1.65 h^{-1}, capturing both tight and leaky envelope conditions. The occupant-load distribution accommodates household configurations from single-occupancy to multi-household co-residence. Median lighting and equipment power densities are 5.50 W/m² and 10.50 W/m², with wide IQRs that reflect a spectrum from efficient lighting to traditional fixtures, and from low-plug-load spaces (e.g., bedrooms, balconies) to high-intensity rooms (e.g., kitchens). Heating and cooling setpoints cluster around 18.0 °C and 26.0 °C, aligning with common comfort preferences while still providing sufficient spread to capture variation in occupant behaviour. Overall, the thermal properties of the envelope, operational settings, internal loads, and temperature setpoints together form a diverse and physically plausible input space. The distributions are consistent with those reported in the literature, yet retain enough dispersion to encompass a broad range of insulation levels and operational scenarios, thereby supporting robust model training and performance evaluation.

Parameter	Unit	Mean	Std	25%	50%	75%
Exterior wall U-value	$W/m^2 \cdot K$	1.20	0.49	0.80	1.20	1.60
Interior wall U-value	$W/m^2 \cdot K$	2.25	0.46	1.90	2.30	2.70
Floor U-value	$W/m^2 \cdot K$	1.45	0.64	0.90	1.50	2.00

Ceiling U-value	$W/m^2 \cdot K$	1.90	0.95	1.10	1.90	2.70
Door U-value	$W/m^2 \cdot K$	1.60	0.55	1.10	1.60	2.10
Window U-value	$W/m^2 \cdot K$	3.40	1.53	2.10	3.40	4.70
ACH	h^{-1}	1.65	0.81	0.90	1.60	2.30
Occupant load	persons/ m^2	0.06	0.03	0.03	0.06	0.08
Lighting power density	W/m^2	5.49	2.75	3.00	5.50	8.00
Equipment power density	W/m^2	10.46	5.62	5.50	10.50	15.50
Heating set point	℃	18.01	1.29	17.00	18.00	19.00
Cooling set point	℃	26.00	1.29	25.00	26.00	27.00

Table. 11 Statistics of building parameters in the simulation dataset

Occupant behaviour distribution

A total of 650,019 room-level schedules were analysed. The 24-hour distribution of average hourly occupancy (Fig. 23) shows a clear nocturnal peak: 00:00–03:00 registers the highest values (0.24-0.26). Occupancy then declines through the morning and reaches a trough around 12:00-14:00 (0.13-0.15), after which it gradually increases from 15:00 onward, rising through the evening and approaching 0.22-0.23 by 22:00-23:00. Aggregating by period, the nighttime average (22:00-05:59) is 0.238, substantially higher than the daytime average (09:00-17:00) of 0.155, consistent with residential "night-in, day-out" occupancy patterns. At the household level (n = 89,739), Fig. 23b summarises the 12-month activation status. Always active (12/12 months) accounts for approximately 0.62 of households; 1 month inactive and 2 months inactive account for roughly 0.14 and 0.11, respectively; 3 months inactive and 4 months inactive decline further to about 0.08 and 0.05. The inactive months are distributed fairly evenly across the calendar year, with no pronounced seasonality observed.

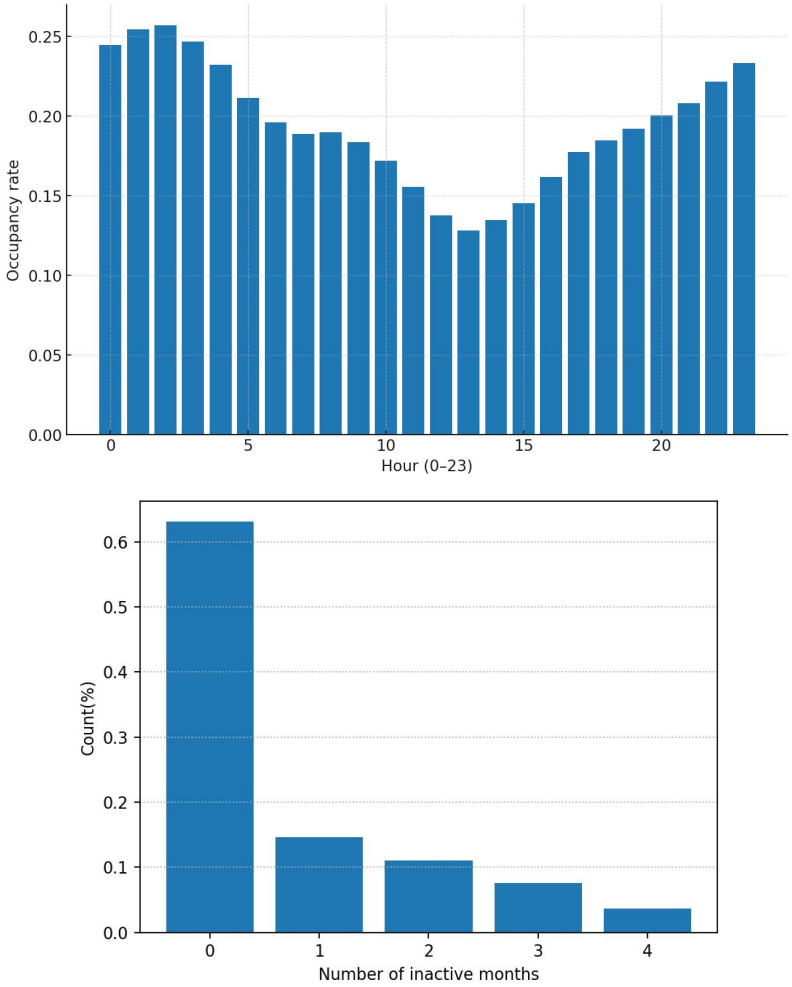
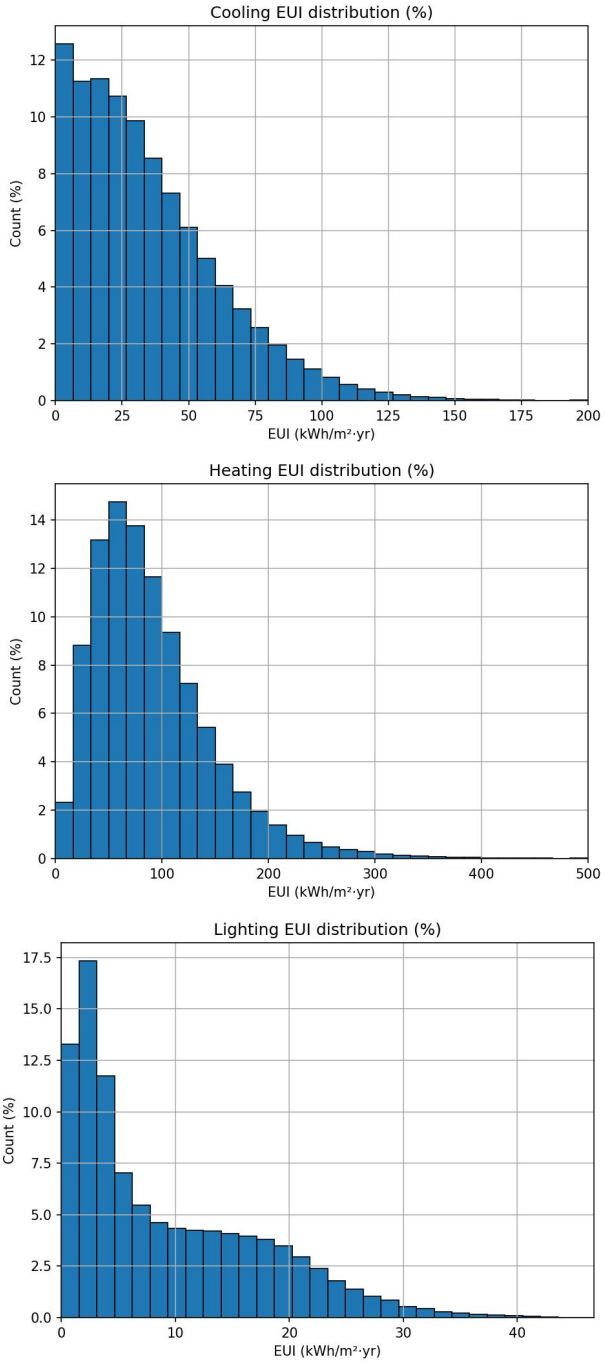


Fig. 23 Statistics of room occupancy rate and household inactive months

4.2.3 Energy use intensity distribution

Fig. 24 summarises the distributions of room-level cooling, heating, lighting, equipment, and total EUI for 650,019 rooms. Cooling EUI spans 0 – 200 kWh/m², whereas heating EUI extends to 0 – 500, indicating that winter heating demand generally exceeds summer cooling demand, which consistent with Jinan's cold-climate context and its higher heating requirements. Both lighting and equipment EUI lie within 0 – 50; however, lighting is heavily concentrated in the 0 – 5 range, while equipment is predominantly 0 – 15. Total EUI ranges from 0 – 600. The 0 – 20 bin accounts for the largest share around 27%, largely corresponding to communal spaces and functional rooms such as balconies and bathrooms that lack space

conditioning. For the remainder of the distribution 20 – 400, the shape closely mirrors that of the heating EUI, suggesting that among the four end uses, heating is the primary driver of total energy use.



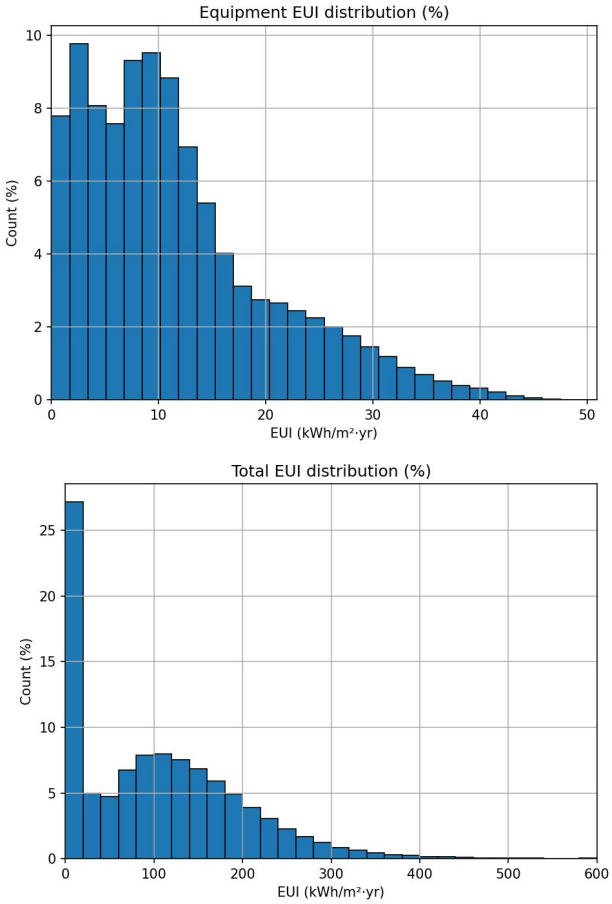


Fig. 24 Statistics of simulated EUIs in the simulation dataset

4.2.4 Cross-scale sensitivity analysis

This study uses a residential building energy dataset generated from large scale parametric energy simulations to compute Spearman correlations between physical and operational parameters and five end use energy use intensities, including cooling, heating, lighting, equipment, and total EUI, at three spatial scales: room, household, and building. By comparing the dominant drivers across scales, we identify how key parameter to EUI relationships evolve with aggregation. The correlation analysis at the room level is conducted only for rooms with space conditioning demand, because these rooms represent a large share of the dataset (74.1%) and contribute the most energy relevant heating and cooling loads. To emphasise drivers that are more meaningful for retrofit decision making, only variables with absolute correlation coefficients $|\rho| > 0.1$ are reported. Tables. 12 to 14 summarise the main correlates for each end use at each scale.

As shown in Table. 12, at the room scale, cooling EUI is strongly negatively correlated with the cooling set point (ρ = -0.53), followed by positive correlations with exterior window area and a negative correlation with being on the ground floor. Heating EUI is positively correlated with exterior wall U-value, heating set point, and exterior window area, with correlation coefficients all around 0.3, indicating that no single factor stands out as a dominant driver at this scale. Lighting EUI shows strong positive correlations with lighting power density (ρ = 0.68), room area (ρ = 0.62), and exposed area (ρ = 0.49), suggesting that higher lighting power density and larger, more exposed rooms tend to have higher lighting EUI. Equipment EUI is positively correlated with room area (ρ = 0.47), equipment power density (ρ = 0.40), and exposed area (ρ = 0.35). Total EUI at the room scale is primarily positively correlated with exposed area (ρ = 0.33), exterior window area (ρ = 0.31), and room area (ρ = 0.28), implying that total energy use at this scale is jointly influenced by room size and the degree of exterior exposure.

Parameter 1	Parameter 2	ρ
Cooling EUI	Cooling set point	-0.53
	Exterior window area	0.34
	Is ground floor	-0.24
Heating EUI	Exterior wall U-value	0.33
	Heating set point	0.33
	Exterior window area	0.28
Lighting EUI	Lighting power density	0.68
	Area	0.62
	Exposed area	0.49
Equipment EUI	Area	0.47
	Equipment power	0.40

	density	
	Exposed area	0.35
Total EUI	Exposed area	0.33
	Exterior window area	0.31
	Area	0.28

Table. 12 Top correlations between EUI (Parameter 1) and parameters (Parameter 2) at the room level

At the household scale (Table. 13), correlations for some room scale geometric descriptors become weaker, while the influence of a household's position within the building becomes more pronounced. Cooling EUI is negatively correlated with being on the ground floor (ρ = -0.43) and positively correlated with floor level (ρ = 0.36). It remains negatively correlated with the cooling set point (ρ = -0.35). Heating EUI shows a relatively strong positive correlation with exterior wall U-value (ρ = 0.61) and is also positively correlated with the top floor indicator and the heating set point. Lighting and equipment EUIs are still driven primarily by their respective power densities, with correlations of ρ = 0.76 for lighting power density and ρ = 0.77 for equipment power density. All other features have correlations below 0.1 for these two end uses. Household total EUI correlates most strongly with exterior wall U-value (ρ = 0.52) and also exhibits clear position effects, with higher total EUI for top floor households (ρ = 0.31) and lower total EUI for ground floor households (ρ = -0.23).

Parameter 1	Parameter 2	ρ
Cooling EUI	Is ground floor	-0.43
	Floor level	0.36
	Cooling set point	-0.35
Heating EUI	Exterior wall U-value	0.61
	Is top floor	0.26
	Heating set point	0.20

Lighting EUI	Lighting power density	0.76
Equipment EUI	Equipment power density	0.77
Total EUI	Exterior wall U-value	0.52
	Is top floor	0.31
	Is ground floor	-0.23

Table. 13 Top correlations between EUI (Parameter 1) and parameters (Parameter 2) at the household level

At the building scale (Table. 14), aggregation further smooths within building variability, making the importance of envelope thermal performance more evident. Cooling EUI is positively correlated with exterior wall U-value (ρ = 0.42), negatively correlated with the cooling set point, and weakly positively correlated with exposed ceiling U-value (ρ = 0.16). Heating EUI is very strongly positively correlated with exterior wall U-value (ρ = 0.73), negatively correlated with building area (ρ = -0.24), and positively correlated with exterior window U-value (ρ = 0.20). At this scale, lighting and equipment EUIs continue to be dominated by lighting power density (ρ = 0.74) and equipment power density (ρ = 0.82), respectively. Building total EUI correlates most strongly with exterior wall U-value (ρ = 0.71), is negatively correlated with building area (ρ = -0.23), and shows a weak positive correlation with exposed ceiling U-value (ρ = 0.17).

Parameter 1	Parameter 2	ρ
Cooling EUI	Exterior wall U-value	0.42
	Cooling set point	-0.20
	Exposed ceiling U-value	0.16

Heating EUI	Exterior wall U-value	0.73
	Area	-0.24
	Exterior window U-value	0.20
Lighting EUI	Lighting power density	0.74
Equipment EUI	Equipment power density	0.82
Total EUI	Exterior wall U-value	0.71
	Area	-0.23
	Exposed ceiling U-value	0.17

Table. 14 Top correlations between EUI (Parameter 1) and parameters (Parameter 2) at the building level

4.3 Results of FusionGNN

Based on the constructed simulation residential building energy simulation dataset, we trained and evaluated the proposed FusionGNN model to assess its effectiveness in predicting household-level annual EUI in multi-household residential buildings. The model integrates four types of inputs: (1) physical features at both the room and household levels, (2) spatial features extracted using a Convolutional Neural Network (CNN), (3) behavioural time-series features, and (4) graph-structured building data that include intra-building connections and thermally relevant edge attributes. The prediction output is the annual EUI (kWh/m^2) for each household unit. The training results show that the model stabilises after 60 epochs, achieving an average validation MAE of 4.29 and a train MAE of 2.13, which indicates high predictive

accuracy for household-level energy use and strong generalisation across diverse building geometries and usage scenarios.

4.3.1 Training performance

As shown by the learning curves (Fig. 25) over 0 to 60 epochs, In the first 0 – 5 epochs, the training MAE drops rapidly from a high level and the validation MAE declines in tandem, indicating that the model quickly learns the dominant patterns and corrects most systematic errors. Between 10 and 30 epochs, both curves continue a smooth downward trend; around 20 – 25 epochs there is a small local fluctuation that quickly subsides, and the MAE is largely compressed to below 5. Between 30 and 60 epochs, the training curve exhibits a gradual fall, while the validation curve stabilises with few fluctuations, indicating that training stays consistent with multi-modal inputs, while the generalisation error has nearly reached its lower limit in the current model setting. At epoch 60, the validation MAE is 4.29 and the training MAE is 2.13. The small and stable gap indicates high reliability in modelling household-level energy use and good generalisation. Because the training MAE continues to fall while the validation MAE no longer improves, the validation performance at epoch 60 is the best in this run; therefore, the model at epoch 60 is selected as the final result to avoid potential over-fitting from further training and to balance accuracy and robustness.

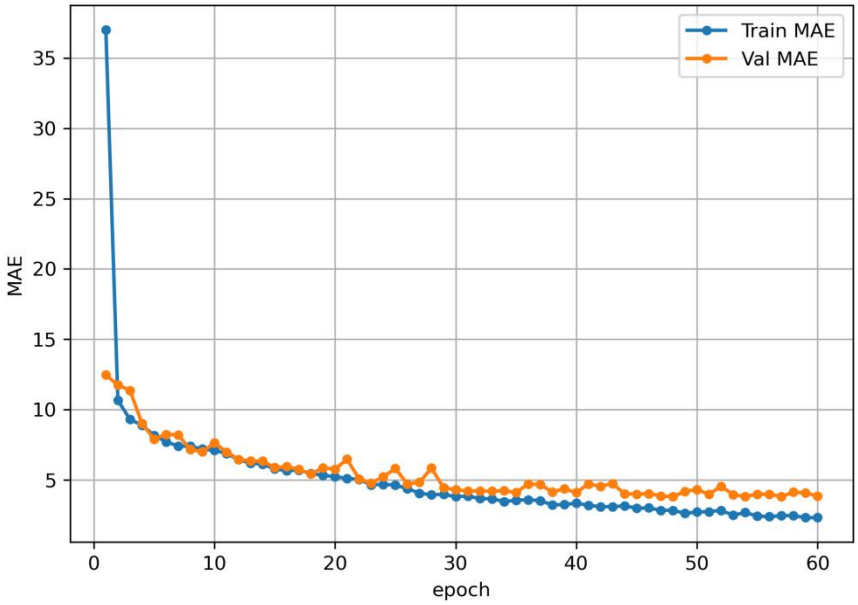


Fig. 25 Learning curve of FusionGNN

Fig. 26 compares the model's predictive performance at epochs 40, 50, and 60. For each stage, the left panel shows the predicted vs actual scatter plot with the coefficient of determination (R^2) to quantify degree of fit; the right panel shows the residual histogram, presenting the share of samples in each residual bin in percentage terms.

At epoch 40, most predictions line up well with the simulated values (R^2 = 0.9961), but there are deviations in the 100 to 200 range. The residuals are centred near zero and 86.88% of samples fall within ±10, while the spread is still wide with a noticeable share of errors lies outside that band. At epoch 50, the deviations in the 100 to 200 range remain (R^2 = 0.9962). The residuals become a little narrower, the share within ±10 rises to 87.89%, and the overall error level is close to that at epoch 40. At epoch 60, the scatter plot reaches the highest R^2 (0.9969), showing the best overall fit, and the deviations in the 100 to 200 range are smaller. The residuals of 90.39% samples are within ±10, which means the errors are more tightly concentrated and the results are more stable.

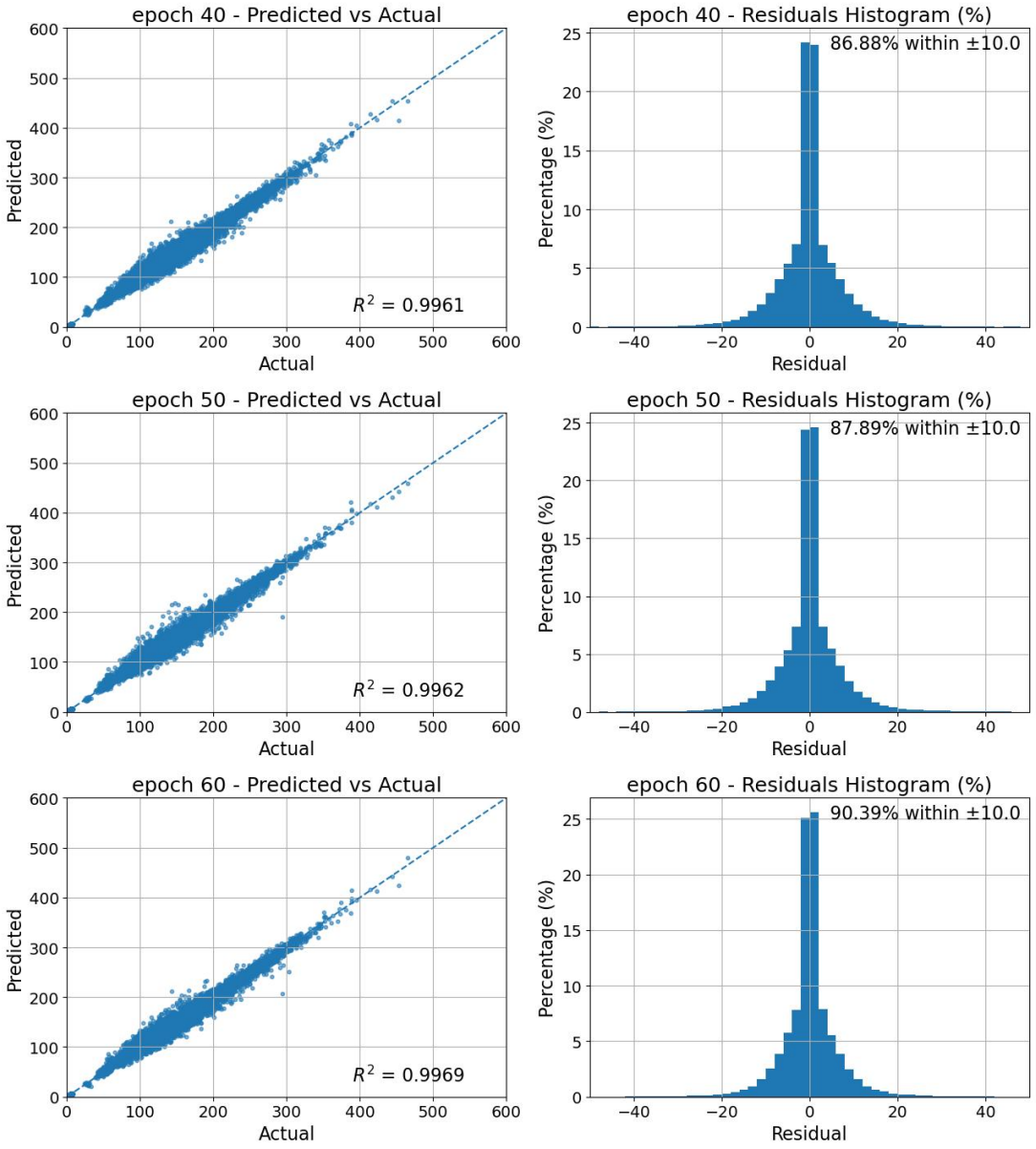


Fig. 26 Predictions vs. actual labels and residuals at epochs 40, 50, and 60

To further assess the prediction accuracy of the final epoch 60 model at a more granular level, Fig. 27 shows the distribution of prediction errors for all nodes in the dataset and the corresponding share of nodes. The errors are expressed as EUI (kWh/m^2) for all residential household units. The histogram in Fig. 27 indicates that the model is accurate for most samples: 50.7% of household units have an absolute error within 0 to 2 kWh/m^2, and another 21.8% fall within 2 to 5. More than 72% of predictions differ from the ground truth by less than 5. A further 17.8% fall within 5 to 10, while larger errors account for only a small share: 5.8% in 10 to 15, 2.0% in 15 to

20, and 1.9% greater than 20. A total of 9.7% of nodes have errors greater than 10. These results show that the model can predict the vast majority of residential household units with small deviations.

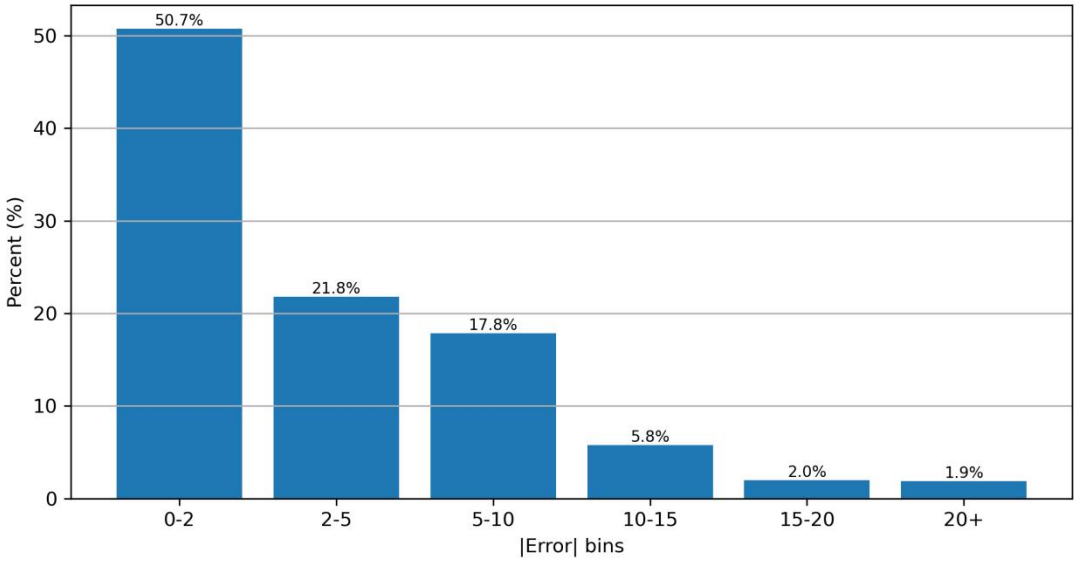
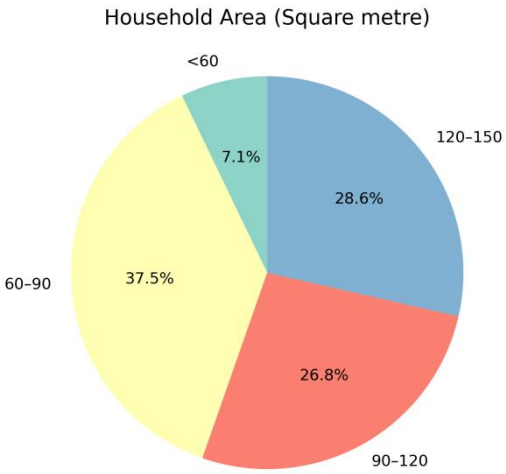


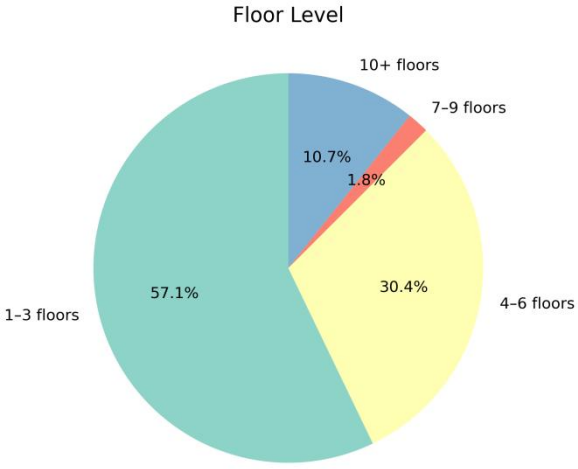
Fig. 27 Household node prediction loss range

4.3.2 Statistics of the real-world energy use dataset

This study conducted an on-site household questionnaire survey in ageing residential communities in Jinan to collect and construct a real-world energy-use dataset for transfer-learning calibration. The dataset contains information from 120 households, including building characteristics, occupant behaviour, usage patterns of common household appliances, and annual electricity bills. Fig. 28 summarises the household floor area, floor level, household configuration, and occupant behaviour archetypes for the 120 samples. In terms of household area, most households fall within 60–90 m^2 (37.5%), followed by 120+ m^2 (28.6%) and 90–120 m^2 (26.8%), while households smaller than 60 m^2 account for only 7.1%. Regarding floor level, the sample is dominated by lower floors: 57.4% of households are located on floors 1–3 and 29.6% on floors 4–6, together accounting for 86% of the total. In terms of stair–household configuration, one staircase serving two households per floor is the prevailing typology (75.0%), followed by one staircase serving three households per

floor (16.1%). With respect to occupant behaviour, the dominant archetypes are retired households (35.7%), student households (32.1%), and office-worker households (26.8%), while households with an infant are relatively rare (5.4%).





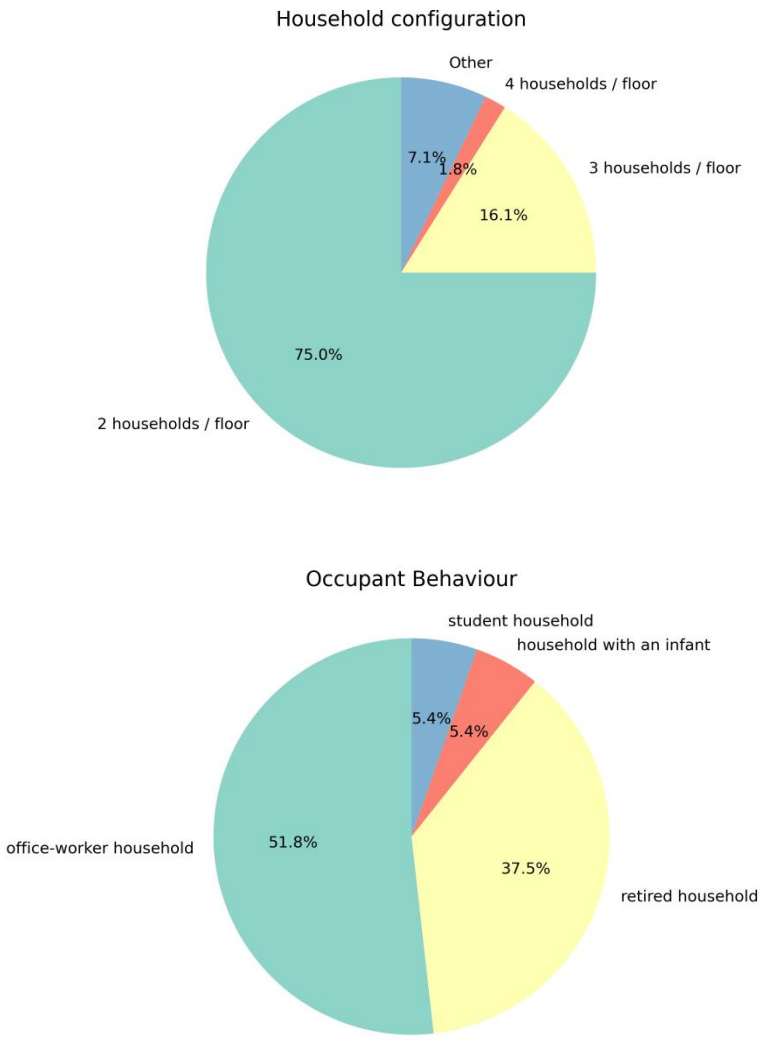


Fig. 28 Statistics of household area, floor levels, household configuration and occupant behaviour archetype from questionnaire

The annual household EUI distribution is shown in Fig. 29. Measured electricity consumption was obtained from collected bills, converted to kWh using the 2024 Jinan electricity tariff, and then divided by household floor area to calculate EUI. Household annual EUI spans 30 – 90 kWh/m^2, with the largest concentration in the 30 – 34 kWh/m^2 bin (23 households). In total, 86 households fall within 30 – 60 kWh/m^2 (72% of the samples), while only 9 households exceed 80 kWh/m^2, indicating that most households cluster in the low-to-medium consumption range and relatively few exhibit high energy use.

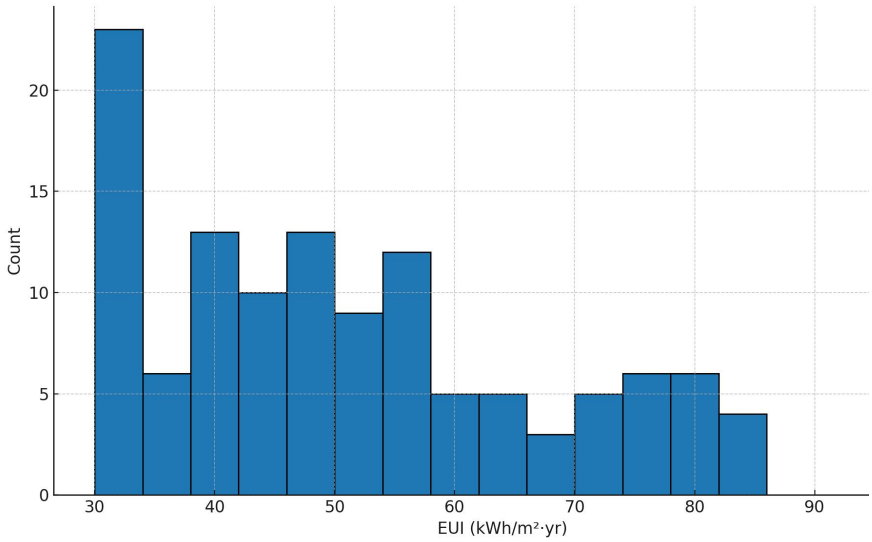


Fig. 29 Measured EUI distribution from questionnaire

4.3.3 Performance gap between simulation and measurements

To quantify the performance gap between measured household energy use and energy simulations, this study conducted household-level energy simulations for all 120 surveyed households using the collected building characteristics and occupant-behaviour inputs. Fig. 30 presents the scatter distribution of simulated annual EUI versus measured annual EUI. Overall, the two series show a moderate level of agreement, with a linear fit of $R^2 = 0.61$, indicating that simulation can reflect the variability in energy use between different households. Nevertheless, the scatter remains visibly dispersed, suggesting a significant simulation–measurement performance gap, with a subset of households exhibiting pronounced deviations.

In terms of error magnitude, the overall mean absolute error (MAE) is 9.77 kWh/m². 26 households (21.7%) fall within an absolute deviation of 0–5 kWh/m², and 101 households (84.2%) are within 0–15 kWh/m². Large deviations are relatively rare: only 6 households (5.0%) exceed 20 kWh/m², implying that for most cases the simulations are broadly consistent with the measured values, while a small number of outliers remain. Regarding error direction, 82 households (68.3%) have simulated EUI higher than measured EUI, with a mean overestimation magnitude of 9.57 ± 4.64 kWh/m². In contrast, 38 households (31.7%) show measured EUI higher than simulated EUI, with

a mean underestimation magnitude of 10.19 ± 8.06 kWh/m², which is more dispersed and exhibits a clearer long-tail pattern. Taken together, these statistics confirm a structured simulation–measurement performance gap: under the current survey-informed input setting, simulations tend to produce systematically higher estimates for a majority of households, while underestimating energy use for a smaller subset of higher-consumption cases, leading to a heavy-tailed error distribution.

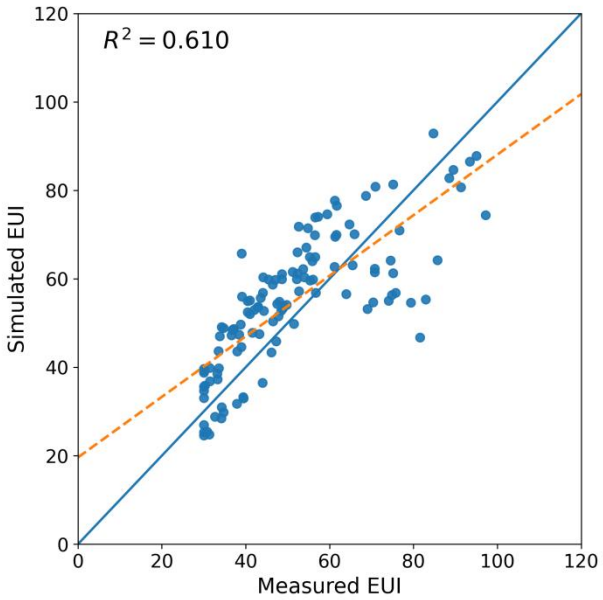


Fig. 30 Simulated EUI and measured EUI distribution

4.3.4 Transfer learning calibration

Against this distributional backdrop, we performed small-sample calibration of the pretrained model. The model achieved its best performance at epoch 6: test-set MAE decreased from 15.53 to 9.24 kWh/m^2 , a relative reduction of 40.5%. This shows that, without unfreezing the backbone, a modest amount of measured data suffices to learn and correct the systematic discrepancies between simulation-based and real-world energy use. Fig. 31 presents scatter plots of the 20 test households before (a) and after (b) transfer learning. It should be noted that Fig. 30 compares survey-informed EnergyPlus simulations with measurements, whereas Fig. 31 compares FusionGNN predictions with measurements; therefore, the bias direction in Fig. 31 does not necessarily mirror that in Fig. 30. In addition, Fig. 31 is based on a small

20-household test subset, so its bias pattern may deviate from the overall trend in the full 120-household sample. In Fig. 31 (a), the correlation between predicted and measured EUI is weak and the fitted slope is close to zero, indicating no meaningful linear correlation. In addition, predicted EUI of most samples are less than measured EUI, indicating that the model usually underestimation on actual samples. After calibration, the correlation improves markedly with R^2 is 0.598. Meanwhile, most predicted-EUI data points fall above the fitted line, indicating that the systematic underestimation is largely mitigated. Together, these results suggest that the regression head effectively learned the residual structure between simulated and measured data, restoring scale consistency while preserving the pretrained ranking information.

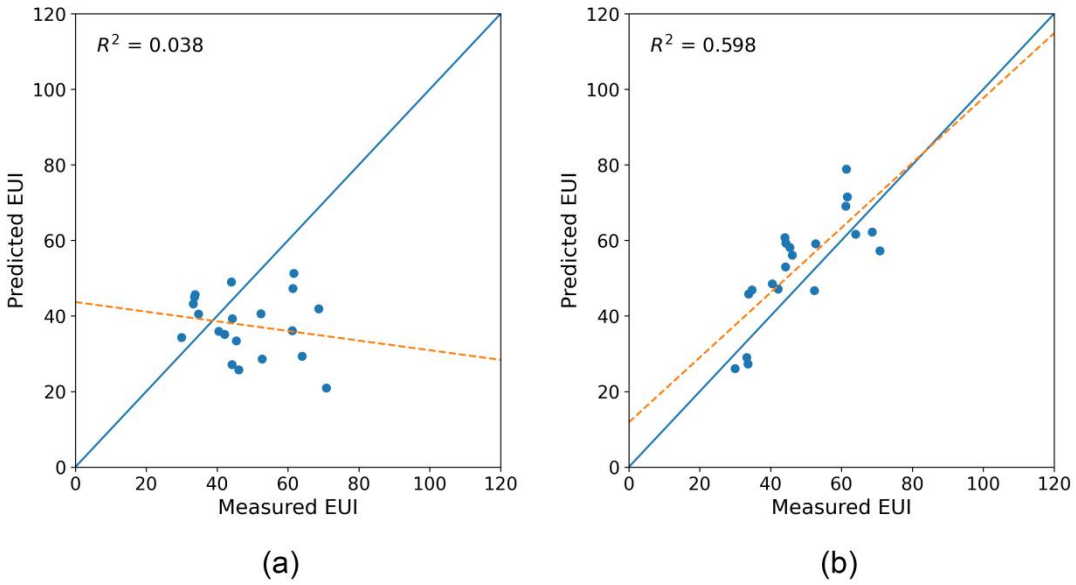


Fig. 31 Predicted and measured EUI before (a) and after (b) transfer learning

4.4 Results of Case Study

The study conducts retrofit decision optimisation for four representative floor-plan types: (a) one stair with two units, (b) one stair with three units, (c) one stair with four units, and (d) two stairs with four units. By combining upgrades to external walls, roof, and windows, 324 retrofit schemes were generated for each prototype, yielding 1,296

schemes in total. Household-level energy use was predicted with FusionGNN and aggregated to the building level; costs were calculated from envelope window areas and unit prices. Pareto-front analysis indicates that the cost distributions for the four archetypes lie roughly within 40,000-600,000 CNY, and post-retrofit EUI is typically reduced by 10–45% relative to the baseline.

4.4.1 Typology A - one stair with two units

Fig. 32 presents the multi-objective scatter plot for Typology A. Feasible solutions fall mainly within a 40,000 to 180,000 CNY cost range, while the Pareto front (orange crosses) is concentrated in 40,000 to 140,000 CNY with energy-savings around 15 to 20%. Savings increase monotonically with cost along the front and show clear diminishing returns toward the right. In the 40,000 to 60,000 CNY band, most optimal packages are window-only upgrades (W5 or W6), achieving about 6 to 11% savings. Within 60,000 to 100,000 CNY, adding roof insulation M3 or M4 on top of high-performance windows raises savings to 17 to 18.5%. Adding M4 wall insulation and W7 windows can achieve a higher of 20% energy-saving but with a noticeably higher cost. The cluster of sub-optimal points on the right typically corresponds to packages that apply higher-grade wall/roof insulation (M3 or M4) yet deliver only modest additional savings relative to their cost. Consistent with the plan characteristics of Typology A which has the smallest envelope area among the four prototypes and the highest WWR, these results confirm that window replacement yields the most pronounced improvement in energy performance for this building type.

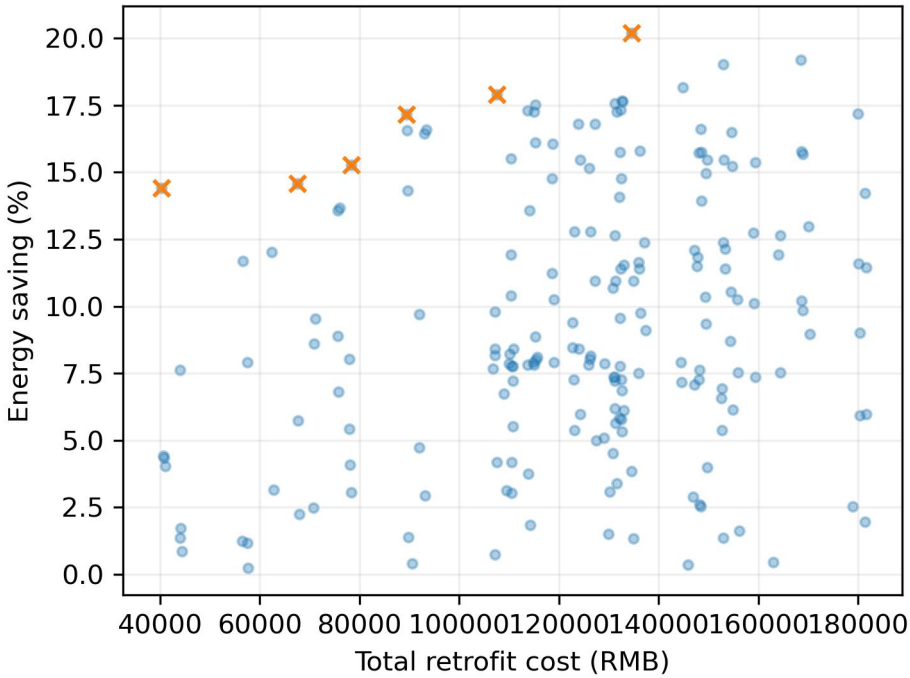


Fig. 32 Pareto front between retrofit cost and energy-saving of Typology A

4.4.2 Typology B - one stair with three units

The multi objective scatter (Fig. 33) of Typology B shows that solutions concentrate in the 60,000 to 320,000 CNY cost range, with a maximum energy-saving of about 23% to 44%. The Pareto front rises monotonically from left to right. On the left around 60,000 to 90,000 CNY there are several low cost and high return orange points. A dense front band appears in the 120,000 to 180,000 CNY range. Only a few high specification front points remain around 200,000 to 260,000 CNY, and to their right many dominated blue points indicate higher cost without clear gains. Within 60,000 to 90,000 CNY, using W5 or W6 high performance windows together with roof insulation M3 and wall insulation M2 achieves about 24% to 41% savings, which offers the best value for money. Upgrading to M4 insulation on both walls and roof and W7 windows can push savings to about 44%, but the cost is higher at around 280,000 CNY. The sub optimal clusters mostly correspond to packages that select high grade wall insulation while windows remain W2 or W3, or to schemes that add wall and roof insulation first while keeping windows unchanged. These patterns are consistent with the plan characteristics of Typology B. The roof to exterior wall proportion is the

highest among the four prototypes, and the building has large north south window areas with additional windows on the east and west. This supports the observation that retrofits on the roof and on windows deliver the most favourable cost effectiveness for this typology.

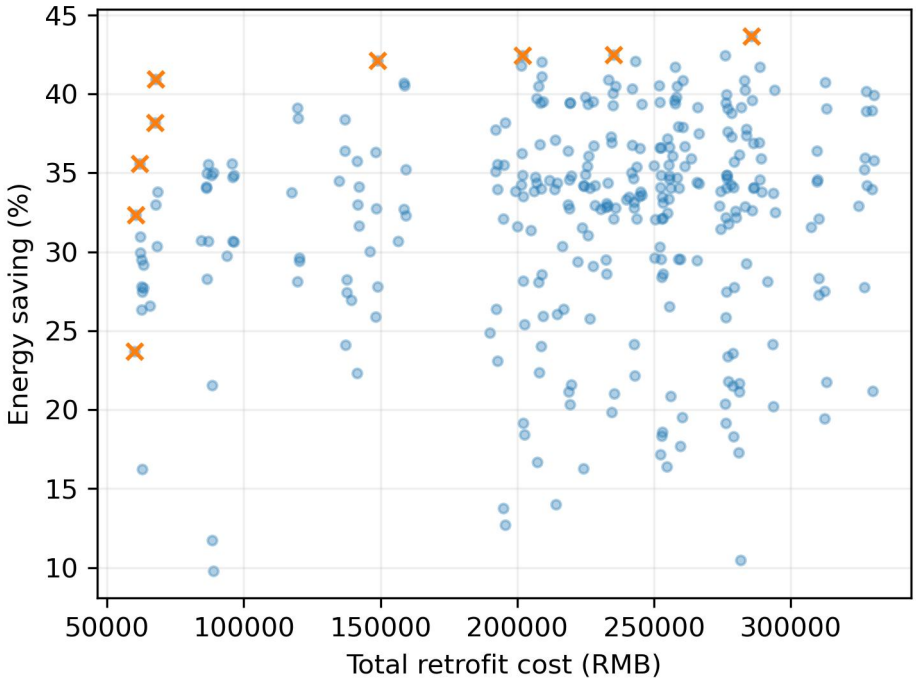


Fig. 33 Pareto front between retrofit cost and energy-saving of Typology B

4.4.3 Typology C - one stair with four units

The scatter plot (Fig. 34) shows that optimal retrofit solutions of Typology C mostly fall between 100,000 and 650,000 CNY, and the Pareto front spans roughly 23 to 34% energy-savings, with clear diminishing returns toward the right end. Typology C has a large exposed exterior wall area and a small WWR, which makes windows contribute less to annual EUI than walls and the roof. Consequently, the best low cost and high energy-saving region on the front is dominated by wall and roof combinations. For the orange Pareto points: within 100,000 to 170,000 CNY, adding M2 or M3 insulation to walls and roof and replacing windows with W5 yields about 23 to 30% savings. With 180,000 to 300,000 CNY, upgrading to M3 wall insulation and M3 or M4 roof insulation

and replacing windows with W5 or W6 achieves around 31% savings. In the 400,000 to 450,000 CNY range, using M4 insulation on both walls and roof plus W7 windows can raise savings to roughly 34%, but with a pronounced increase in cost. The many sub-optimal points scattered across the figure mainly stem from large-scale replacement of low-U windows while keeping walls/roof at higher-U materials (M1 or M5). Therefore, for Typology C, upgrading wall and roof insulation offers a better cost-effectiveness than large-scale window replacement.

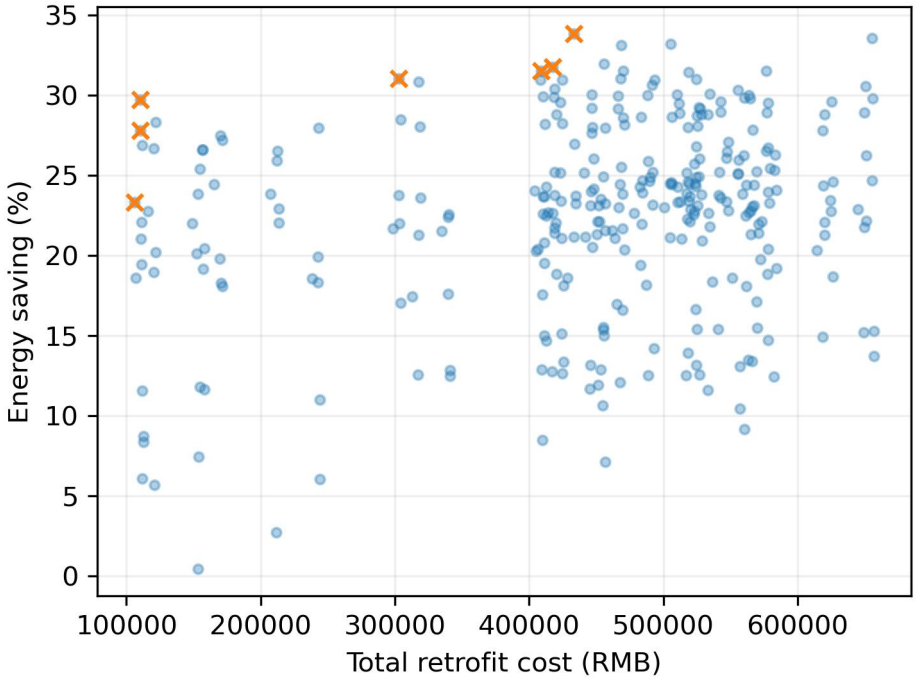


Fig. 34 Pareto front between retrofit cost and energy-saving of Typology C

4.4.4 Typology D - two stairs with four units

The multi-objective scatter plot Fig. 35 shows that optimal solutions for Typology D lie mainly between 50,000 and 300,000 CNY, while the Pareto front concentrates in the 50,000 to 160,000 CNY band and rises from about 9% to 24% energy-savings; on the right, many high-cost but low-gain dominated solutions appear. The front can be read in three bands. At low cost around 50,000 to 70,000 CNY, window replacement with W5 or W6 alone yields 9 to 15% savings. Building on window upgrades, adding roof insulation M3 or M4 increases savings to 15 to 20% with a 70,000 to 110,000 CNY

budget. Within 110,000 to 160,000 CNY, further upgrading external walls to M2 or M3 raises savings to 20 to 24%. The large cluster of sub-optimal points on the right mostly corresponds to packages that add roof or wall insulation while window performance remains weak. This pattern matches Typology D's plan characteristics: among the four layouts it has the smallest exposed exterior wall area, while south-facing windows are the primary pathway for heat exchange. Therefore, for Typology D, prioritising high-performance window replacement achieves the best energy-savings within a limited budget.

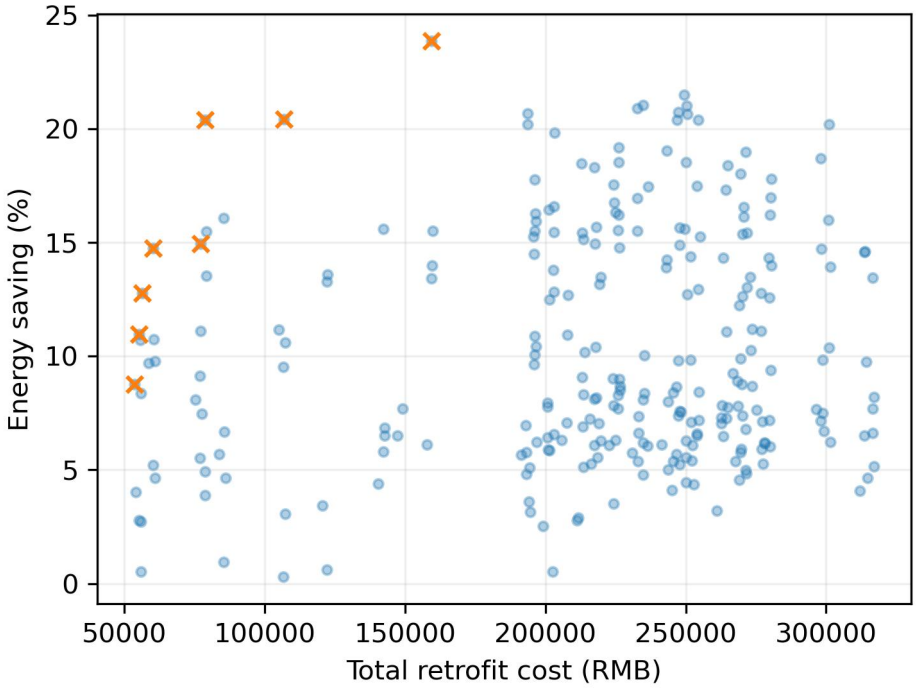


Fig. 35 Pareto front between retrofit cost and energy-saving of Typology D

5. Discussion

5.1 Discussion of Automated Building Energy Modelling

This study presented an automated building energy modelling framework that streamlines energy model generation from residential floor plans. By accurately

identifying architectural elements and classifying room types, the automated framework reduces the efforts needed to reconstruct the energy model creation with occupancy schedules. The streamlined building energy assessments using floor plans can be effectively integrated into energy retrofit planning, affording rapid and iterative evaluations of design alternatives. The proposed CTRC algorithm and GAT model achieved accuracies of 98.5% for room reconfiguration and 93.7% for room type classification, demonstrating their effectiveness in retrieving semantic information from complex multi-household residential layouts.

5.1.1 Floor plan segmentation to vectorisation

In this study, we employed a modified DMTN model to perform image segmentation on raw floor plans, followed by image processing techniques to generate simulation-ready vectorised representations. The evaluation results demonstrate that the modified DMTN model performs well in detecting room regions but shows relatively lower performance in recognising windows and doors. The low recall rate in window detection is attributed to annotation ambiguity, as sliding doors and windows are often represented by similar rectangular symbols in architectural drawings, as illustrated in Fig. 36 (R4). This visual similarity can lead to misclassification between windows and doors.

The limitation of a fixed-size convolutional kernel in the DMTN model necessitates input images to be rescaled to 512×512 pixels for segmentation. This rescaling process introduces distortion, particularly in floor plans with fine-grained architectural elements such as thin walls or closely spaced components. As shown in Fig. 36, R1 demonstrates a case where two originally separate rooms were mistakenly merged into one due to a missed door detection, causing room segmentation errors. In R2 and R3, the proximity of walls after rescaling causes the model to incorrectly predict non-existent doors, which not only reduces door detection accuracy but also erroneously splits originally complete room regions, further lowering room recognition performance. These results highlight the need for improved feature extraction

strategies to meet image resolution requirements effectively. Future work could explore adaptive convolution mechanisms (Wang et al., 2025) or introduce pyramid pooling layers to enable variable size floor plan inputs (He et al., 2015), thereby improving the preservation of spatial details during segmentation and mitigating errors caused by image resizing.

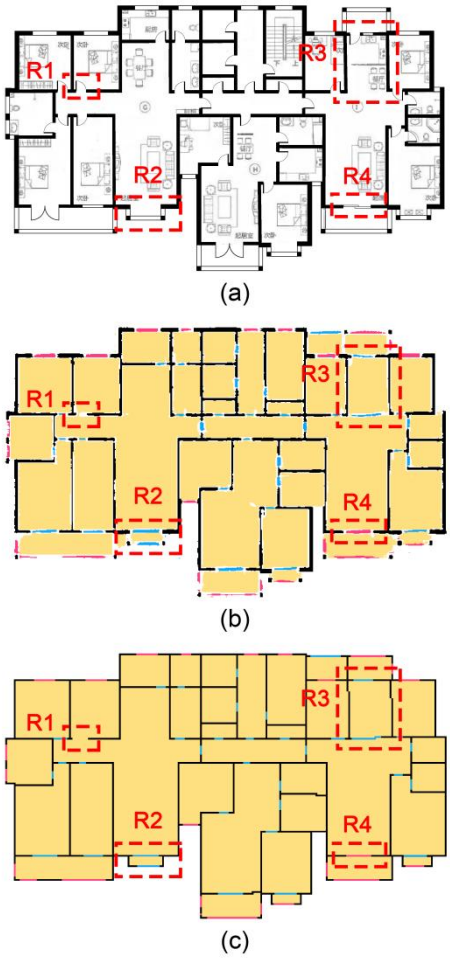


Fig. 36 Wall errors of architectural element extraction (a) input floor plan (b) segmented floor plan (c) vectorised floor plan

5.1.2 Multi-household reconfiguration

Core node classification using the DT classifier and the CTRC algorithm are employed to perform multi-household reconfiguration. The high accuracy of 95.3% for core node classification and 98.5% for multi-household clustering indicates the

robustness of the proposed method. In particular, the classification of household nodes is essential to support successful room-type classifications. Occasional misclassifications were observed between communal and function core nodes, leading to inaccurate household clustering results. Fig. 37 (a) and (b) illustrate two example scenarios: (a) misclassification between communal and function core nodes and (b) misclassification of communal core nodes as household core nodes. As shown in Fig. 37 (a), residential households with multiple pathways connecting to communal spaces often exhibit complex connectivity relationships that influence the household clustering results. Fig. 37 (b) presents a scenario where a communal core node is misclassified as a household core node, leading to an additional household cluster. Such misclassifications caused by unseen spatial connectivity characteristics could result in inaccurate household clustering and reduce the overall reconfiguration accuracy.

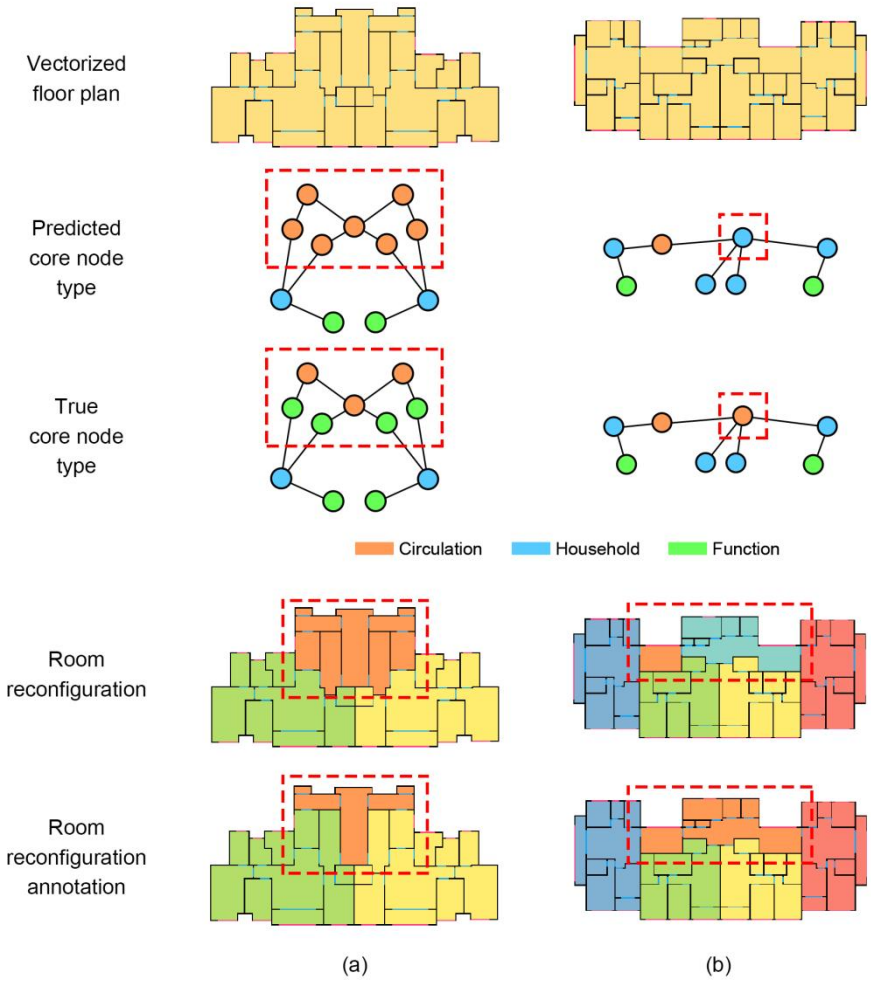


Fig. 37 Examples of core-node misclassification (a) misclassification of communal and function core nodes (b) misclassification of communal and household core nodes

5.1.3 Room type classification

The accuracy of 93.7% for room type classification indicates that the GAT model effectively learns and utilises the geometric and topological relationships among rooms within household graphs for room type identification. The classification accuracy for the living room is the highest, reaching 1.0, indicating the superior capability to capture the unique connectivity patterns of living rooms within the household unit. The classification accuracies for bedroom, bathroom, and balcony are all around 0.95, demonstrating that the GAT effectively captures both node features and connectivity relationships to distinguish these room types. For example, living rooms typically function as central organising spaces that connect to multiple rooms,

characterised by distinctive edge features and a lower mean depth, making them easily identifiable. Balconies, on the other hand, often exhibit a higher window-to-wall area ratio, which serves as a key distinguishing feature from other room types. However, the classification accuracy for kitchens and utility rooms was lower. The confusion between kitchens and bedrooms arises from their similar area ratios and both rooms are often connected only to living rooms. Utility rooms are often similar in size to bathrooms and tend to be in less accessible corners of the layout, resulting in misclassification with bathrooms. Additional informative features are needed to better distinguish between kitchens and utility rooms.

In comparison, this study considers the widely utilised R3D dataset to assess its classification accuracy with those from other models (as shown in Table. 15). The R3D dataset contains 215 residential building floor plans, all of which are single-household layouts, with an average of six rooms per floor plan (Liu et al., 2015). The results show that the proposed GAT model outperforms in five out of six room types, achieving an overall highest accuracy of 0.921. The GAT model demonstrates significant improvements in prediction accuracy across various room types, including bedroom, bathroom, balcony, hallway and closet. Specifically, the prediction accuracy for the balcony, hallway, and closet increased by over 10%. Despite FPNet outperforming in living room classification by 0.02, the GAT model still achieved the second highest accuracy of 0.94 among the four models. These comparative results demonstrate the effectiveness and robustness of the proposed GAT model for classifying room types in residential floor plan layouts.

	Accuracy	bedroom	living	bathroom	balcony	hallway	closet
DMTN	0.697	0.79	0.93	0.81	0.49	0.61	0.61
DeepLabV3+	0.393	0.40	0.90	0.57	0.0027	0.44	0.048
FPNet	0.765	0.83	0.96	0.83	0.61	0.74	0.62
GAT	0.921	0.9	0.94	0.91	0.97	0.85	0.96

Table. 15 Room type classification accuracy among different approaches

5.1.4 Automated energy modelling

This study represents the first integration of automated floor plan analysis with energy modelling, significantly reducing the manual effort traditionally required in the BEM creation process. The automated framework for architectural element construction, household configurations, and room type classifications demonstrates its practicality in supporting building energy performance evaluation. In particular, automating the assignment of predicted room occupancy schedules for building energy assessments is particularly valuable for early-stage building design and building energy retrofit planning, where multiple design iterations required rapid and accurate evaluations with limited information. The comparison results of energy simulations highlight that the accuracy of architectural element recognition and vectorisation has a significant impact on the performance of the proposed automated BEM modelling workflows. Group 1, which retained manually defined building geometries while automating household clustering, room type classification, and schedule assignment, demonstrated a low average error of 0.043 with a highly concentrated error distribution. In contrast, Group 2, which employed a fully automated modelling process, yielded a higher average error of 0.191 and exhibited a more dispersed error distribution. Further analysis revealed that the primary source of error in Group 1 stemmed from incorrect room type predictions. Specifically, 93.5% of the samples exhibited minor room classification errors, resulting in energy simulation errors generally within 12%. Samples with errors exceeding 12% were typically caused by household clustering mistakes, which subsequently led to multiple room type misclassifications. As discussed in Section 4.2, the misidentification between household and communal spaces caused multiple rooms originally belonging to communal clusters to be misclassified, leading to a maximum energy use error of 38%. For Group 2, the major source of error was the misidentification or omission of architectural elements, which often lead to incorrect household configurations. In particular, high-error samples with deviations exceeding 30% are frequently caused by the omission of walls or doors between household and communal core nodes.

These misidentifications result in the unintended merging of rooms that should belong to separate clusters, severely disrupting the automated room reconfiguration process and ultimately leading to significant inaccuracies in energy simulation outcomes. While full automation offers clear advantages in efficiency, manual verification and calibration of geometric modelling and component recognition are still essential to ensure the accuracy of automatically generated BEMs. Future research should focus on improving the precision of boundary segmentation and enhancing the robustness of architectural element recognition to reduce simulation errors caused by component identification failures.

In summary, this study introduces a novel automated building energy modelling framework for automating building information extraction and energy model generation from 2D residential floor plans. Its ability to handle complex multi-household layouts demonstrates strong potential for large-scale residential energy analysis. By accelerating the generation of BEM inputs and reducing dependency on manual processes, the proposed method not only enhances simulation efficiency but also supports broader goals in building energy performance optimisation, conservation, and the transition toward sustainability.

5.1.5 Limitations

Despite the strong overall accuracy achieved across all processing stages, the proposed automated building energy modelling framework introduces interdependent error propagation, where inaccuracies in earlier steps systematically compromise downstream results. For example, segmentation errors, such as undetected room boundaries or the misclassification of windows and doors, can distort the vectorised geometry used for thermal zone construction. These geometric inaccuracies can lead to incorrect surface area estimations and misrepresentation of fenestration properties, significantly compromising the accuracy of the energy model. Similarly, misclassification of core nodes during room reconfiguration, particularly when communal spaces are erroneously labelled as household nodes, can cascade into

subsequent steps, where room type assignments are based on faulty household clustering. This results in semantic mismatches, such as assigning living room schedules to communal zones, or misestimating internal loads due to incorrect zone labelling.

Moreover, the multi-household residential floor plan dataset used in this study primarily consists of Chinese residential floor plans, which may limit the generalisability of the framework across diverse architectural and cultural contexts. For instance, the current room type classification is limited to six categories commonly found in Chinese apartments. However, residential layouts from other countries may include additional functional spaces such as garages, closets, or pantries that are not accounted for in the current model, potentially reducing classification accuracy when applied to these cases.

These findings underscore the need for future enhancements such as iterative cross-stage validation mechanisms, including topological consistency checks during geometry vectorisation and semantic verification prior to cluster reconfigurations to reduce compounded errors in complex floor plan layouts. In addition, extending the dataset to include a broader range of residential typologies and cultural contexts will further improve the robustness and adaptability of the proposed framework.

5.2 Discussion of Large-scale Parametric Energy Simulation

This study leverages automatically generated BEMs, enhanced through model and parameter enrichment, to run large-scale energy simulations and assemble a residential-building energy dataset. Section 4.2 reports summary statistics for the ranges of physical parameters and the distributions of simulated energy use. In this chapter, we examine correlations between these physical parameters and five end-use EUIs, namely cooling, heating, lighting, equipment, and total EUI. Based on the dataset's consistent building and household identifiers, we conduct correlation analyses at three levels of granularity: room, household, and building. By comparing

parameter – EUI relationships across these scales, we reveal how the influence of different drivers changes with aggregation, providing evidence to support energy-efficiency retrofit strategies tailored to each level.

5.2.1 Dataset representativeness and uncertainty

This study conducted large-scale parametric energy simulations for 64 representative legacy residential communities in Jinan, China, and constructed a multi-household residential building energy simulation dataset. The dataset covers 4,282 buildings, 89,739 households, and 650,019 rooms, providing a solid training foundation for subsequent development of the FusionGNN energy prediction model and retrofit strategy optimisation. Compared with simulation datasets that record energy use only at the building or household level, the proposed dataset captures inputs and outputs at room, household, and building scales, including geometry and thermal parameters, surrounding shading context, inter-apartment adjacency, and end-use EUIs. This multi-level, structured representation supports both cross-scale energy pattern analysis and graph-based learning that leverages spatial topology and physically meaningful features. The parameter distributions demonstrate physical diversity, with envelope U-values spanning a continuum from uninsulated to insulated specifications, lighting and equipment power densities representing a wide range of internal load intensities, and set points and behavioural parameters showing sufficient dispersion. The resulting EUI distributions further reflect the heating-dominated nature of the local stock under Jinan's climate, while the low-energy cluster associated with communal and functional spaces highlights the structural role of non-conditioned rooms in shaping the overall distribution. Collectively, these results indicate that the dataset provides a representative and information-rich basis for modelling energy performance in typical residential communities in Jinan.

Nevertheless, to enable large-scale simulation and parametric sampling, several modelling simplifications were adopted, which may introduce uncertainty when transferring beyond the target climate context or aligning with real measured energy

use. First, all simulations were conducted under a single typical meteorological year weather file for Jinan, meaning that the derived parameter – energy relationships primarily reflect responses under this specific climate and do not capture inter-annual variability or other climate zones. Future work could treat climate as an explicit variable and develop multi-climate simulation datasets to strengthen cross-region applicability. Second, envelope performance was simplified using component-level U-values, which does not fully capture thermal bridges, construction quality variation, ageing, moisture effects, or heterogeneous assemblies. This simplification may bias the absolute magnitude of heating demand and influence the inferred importance of envelope parameters. In addition, applying uniform envelope parameters to all households within the same building specification is a pragmatic representation at scale, but it may deviate from real-world conditions where window and door upgrades often vary across households within the same building, contributing to intra-building energy differences that are not explicitly represented. Third, simplified assumptions regarding infiltration and ventilation, such as uniform ACH within a household or across rooms, do not reflect room-specific ventilation behaviour. In the Chinese residential context, occupants may ventilate living rooms and master bedrooms more frequently, while kitchens and bathrooms often exhibit episodic, event-driven ventilation after use. Such spatially heterogeneous behaviour can alter local heat exchange and internal conditions and may lead to measured patterns that deviate from simulated distributions. Future work may incorporate finer-grained behavioural evidence, for example via questionnaires, time-use data, or lightweight sensing where ethically feasible, to parameterise room-level ventilation and occupancy more realistically.

Despite these limitations, the dataset remains valuable for both methodological development and applied analysis. On the one hand, it integrates real floor plans and reconstructed urban context and provides multi-scale, structured records of shading, envelope properties, internal loads, behavioural schedules, and end-use EUIs, enabling reproducible surrogate modelling and deep learning training at scale. On the other hand, the inclusion of household adjacency and shared-boundary information

offers essential inputs for representing inter-unit thermal coupling and spatial dependence in multi-household buildings, thereby supporting graph-based models that more closely reflect the interconnected physical reality of apartment buildings. Therefore, the proposed dataset not only underpins FusionGNN training and validation, but also provides systematic support for cross-scale energy analysis and multi-objective retrofit optimisation.

5.2.2 Occupant behaviour modelling and uncertainty

To represent occupant behaviour in the large scale parametric simulations, this study adopted four schedule archetypes derived from Dong et al. (2023), covering office worker households, student households, retired households, and households with an infant. Each archetype provides a distinct 24 hour occupancy profile for the main functional zones, including bedroom, living room, kitchen, and bathroom, while secondary rooms such as balconies and utility rooms were assigned low activity levels and communal spaces were modelled with a low background presence. In addition, a month scale vacancy variable was introduced to approximate longer absences, with each household assigned an inactive duration between 0 and 3 months. Together, these modelling choices enable the dataset to capture both within day room use patterns and seasonal or holiday related vacancy dynamics, providing behaviour inputs that can be linked directly to room, household, and building energy outcomes.

To better reflect behavioural diversity beyond a single fixed schedule, stochastic variation was incorporated through time displacement and intensity adjustment. Active periods were shifted by up to plus or minus 3 hours to represent variability in daily routines, and occupancy duration or intensity was adjusted by up to plus or minus 25 percent to represent differences in how frequently and how intensively spaces are used. The resulting distributions show that the generated schedules reproduce plausible residential rhythms at scale. Across 650,019 rooms, the average hourly occupancy exhibits a clear nocturnal peak around midnight to early morning and a

midday trough, consistent with typical night dominant residential presence. At the household level, the vacancy indicator produces a realistic spread of activation states, with most households active throughout the year and smaller proportions experiencing one to several inactive months, while the inactive months are distributed broadly across the calendar without strong seasonality. These results suggest that the stochastic sampling strategy increases sample level diversity while preserving the underlying behavioural structure of the archetypes, supporting robust model training and reducing over reliance on a narrow set of deterministic routines.

Nevertheless, behavioural assumptions remain a major source of uncertainty when transferring from simulation to real households. Even with stochastic perturbations, archetype based schedules may not capture non-standard lifestyles or atypical working patterns, such as households with systematically inverted day night activity. Likewise, real households often exhibit episodic and event driven changes that are difficult to represent through static archetypes and a simple month scale vacancy indicator, including temporary co residence of visiting relatives, short term changes in room function, or holiday periods with additional family members. Such events can cause abrupt shifts in occupancy and appliance use intensity and may lead to measured energy patterns that deviate from the simulated distributions. Future work should therefore prioritise collecting empirical behavioural evidence, for example through refined questionnaires, time use diaries, or sensing where ethically feasible, to expand the set of archetypes and parameter ranges. A practical extension is to introduce multiple behavioural modes for each household, such as guest visit weeks, school holiday mode, or night shift mode, and to model behaviour as a probabilistic mixture of modes that switch across days or weeks. Incorporating such multi-mode stochastic schedules into simulation would expose learning models to a wider behavioural space, reduce sensitivity to mismatched behavioural assumptions, and improve generalisability in real world deployment.

5.2.3 Differences in correlations across room, household, and building scales

The correlation results indicate that the key correlates of cooling and heating energy use are concentrated in set points, the degree of spatial exposure and floor position, and the thermal performance of the building envelope. This is consistent with basic building physics: set points determine the indoor to outdoor temperature difference and thus the intensity of heating and cooling demand; exposure and positional conditions determine the boundary area through which a room exchanges heat with the outdoors; and envelope performance directly affects conductive heat losses and solar or ambient heat gains. In contrast, lighting and equipment energy use are primarily driven by their respective power densities, suggesting that non HVAC end uses are more directly constrained by internal load assumptions and usage intensity. Statistically, total energy use follows patterns closer to heating and cooling, implying that under the current simulation sample and parameter settings, space conditioning loads are the primary source of variation in total EUI and therefore show consistent correlation changes with factors that jointly influence thermal demand and heat exchange with the exterior.

From the perspective of end uses, cooling EUI is most sensitive to the cooling set point and exhibits a practically meaningful monotonic relationship: lower set points lead to higher cooling demand. The positive correlation between cooling EUI and exterior window area may reflect increased solar heat gains associated with larger glazing areas, which elevates cooling energy. In addition, ground floor rooms show slightly lower cooling EUI, which may be explained by reduced roof exposure, lower susceptibility to direct solar radiation in summer, and heat exchange with the ground, making these spaces less prone to pronounced increases in summer cooling loads compared with upper floors. At the room scale, heating EUI is mainly driven by exterior wall U-value and the heating set point, which is physically intuitive. Poorer insulation, represented by a higher U-value, increases winter heat losses, while a higher heating set point enlarges the indoor to outdoor temperature difference and

therefore increases heating demand. The positive correlation between exterior window area and heating EUI is also consistent with the principle that larger opening areas lead to greater heat transfer losses. Compared with space conditioning end uses, lighting and equipment exhibit stronger consistency. Across all three scales, lighting power density and equipment power density remain the most significant correlates of their corresponding EUIs, indicating that when operating schedules and behavioural assumptions are broadly consistent, internal load settings largely determine the statistical variation of lighting and equipment energy use.

The correlation analysis also reveals that some parameter to energy relationships attenuate as the analysis scale is aggregated. For example, at the room scale, exposed area and exterior window area are key correlates of total EUI, whereas at the household and building scales these variables no longer show notable correlations. This suggests that floor area weighted averaging at the household level cancels part of the room to room variability, and subsequent averaging at the building level further cancels differences among households. As a result, many geometric descriptors that are salient at the room scale appear as weaker statistical associations at larger scales. In contrast, envelope thermal parameters are more likely to emerge as dominant correlates at the household and building scales, because they are usually consistent or highly similar within the same building and therefore better explain variation between buildings, particularly for heating dominated outcomes. This pattern provides a useful basis for interpreting energy drivers and for formulating multi-level retrofit strategies. At the room or household level, measures can focus on exposure sensitivity and operational settings, such as prioritising shading, window upgrades, or set point management for spaces with larger glazing areas, higher exposure, or specific floor positions. At the building level, more robust strategies typically involve improving overall envelope performance, especially exterior wall insulation and potential upgrades to glazing and roof thermal performance, because these measures systematically reduce conductive loss pathways and yield more stable reductions in heating and total energy use.

The current correlation analysis also has limitations. Some variables may act as

proxies for unobserved factors, for example exposure measures capturing both geometry and boundary condition differences. To further improve interpretability and decision maker trust, future work could complement correlation screening with targeted sensitivity tests for key parameters or develop interpretable surrogate models to quantify the influence of individual variables on building energy use, thereby strengthening the explanatory value of the framework for practical retrofit decision support.

5.3 Discussion of FusionGNN

Results on the simulation dataset show that FusionGNN effectively leverages multimodal inputs and delivers accurate household-level energy predictions. Building on the training outcomes in Section 4.3 and the transfer-learning calibration against measured data, this chapter aims to analyse the characteristics of high-loss nodes, examining the sources and magnitude of uncertainty in the real-data calibration.

5.3.1 Distribution of high-loss nodes

We analysed the top 10% of households by MAE (threshold $\approx$ 10.2) and, as shown in Fig. 38, evaluated ten household-level features using the standardised mean difference (Cohen's d): HVAC conditioned, ACH, equipment per area, cooling set point, exterior window area, lighting per area, people per area, heating set point, exposed area, and total area. The standardised mean difference (d) quantifies how different the feature distributions are between the high-loss group and the rest; larger values indicate stronger distributional divergence. Household-level features were aggregated from room-level inputs by area-weighted averaging (e.g., equipment per area), simple averaging (e.g., cooling set point), or summation (e.g., exposed area). Results show that HVAC conditioned has the largest standardised mean difference, 0.823, implying that households with more rooms requiring cooling/heating and with more frequent or more stable operation are more prone to model bias in energy-use

prediction. This points to insufficient representation of equipment on/off states and control logic in the current feature set, leading to systematic error. ACH has d = 0.801, indicating that uncertainty in ventilation and infiltration is especially sensitive for error: a fixed per-household ACH cannot capture real window-opening behaviour, so room-level mismatches accumulate and amplify at the household level. For internal loads, equipment power density is notably higher at 0.78, suggesting that internal gains from equipment operation are not being effectively expressed. Regarding setpoints, cooling set point is strongly positively related to error, while heating set point is negatively related, meaning higher cooling set points and lower heating setpoints are both associated with larger errors, reflecting the model's sensitivity to gaps between assumed comfort strategies and actual operation. Overall, high-loss households systematically deviate from the rest by exhibiting more conditioned spaces, greater ACH and exterior window area, higher internal loads, and more aggressive cooling or heating set points.

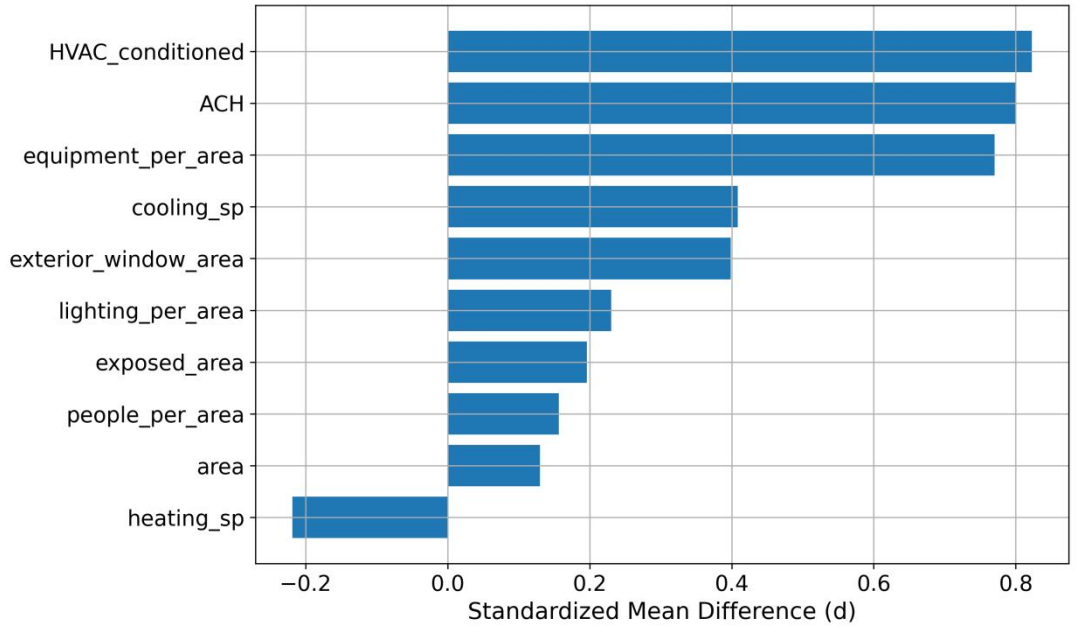


Fig. 38 Top 10 features with the highest standardised mean differences for high-loss nodes

To address the current distribution of high loss nodes, improvements can be made in

both feature engineering and model architecture. On the feature side, encode HVAC conditioning status and ACH as time resolved signals rather than static flags, that is, map binary on or off or single scalars to multidimensional, time dependent embeddings. Concretely, refine HVAC on or off and ACH with schedule data to better express when HVAC is enabled or disabled and when ventilation demand is high. For equipment and lighting, build a household level decomposition of intensity times daily usage hours to reduce bias introduced by area only mappings. For windows, introduce thermal parameters beyond the U-value, such as SHGC, so the model can learn interactions among multiple thermal properties and rely less on geometry only U maps. At the model architectural level, elevate HVAC status and setpoints from constants to time conditioned drivers. Use the same FiLM mechanism to modulate HVAC on or off, cooling and heating setpoints, and ACH with user behaviour time series, producing physically meaningful features weighted by temporal patterns that reflect behavioural uncertainty. Introduce node level gating in the graph or aggregation stage to down weight features with high uncertainty or variance and prevent them from contaminating neighbouring nodes. In view of the observed high loss distribution, the dominant error sources are underestimating uncertainty when scalar simplifications of time sensitive features interact with occupant behaviour, and insufficient thermophysical representation when relying on a single U map channel for the envelope. Targeted preprocessing of these time sensitive features is expected to shrink errors in high loss households and improve overall generalisation stability.

Overall, FusionGNN currently delivers strong accuracy and stable generalisation on multi-household residential energy prediction, effectively learning the mappings between temporal, spatial, and physical features and energy use. The high loss pattern indicates that some physical variables are highly sensitive to temporal dynamics, and behavioural uncertainty can undermine their expressiveness. This suggests that FiLM-based fusion should incorporate temporal and physical modulation, enabling the feature structure to move beyond the current combined time-space and physics formulation toward a genuinely integrated multi-modal representation across time, space, and physics.

5.3.2 Model transferability and scalability

The proposed framework supports large-scale screening of retrofit scenarios by generating training data through extensive energy simulations and then using a trained data-driven model for rapid prediction. In the current implementation, FusionGNN was trained for 60 epochs on an NVIDIA RTX 4070 GPU, with a total training time of 57 hours and 48 minutes. After training, inference is computationally lightweight. Across 4,282 buildings, 32 retrofit scenarios were evaluated, resulting in 137,024 building-level graph inferences completed in 22 minutes, equivalent to approximately 0.01 second per building prediction. By contrast, during the simulation stage, a single energy simulation for one building typically takes around 300 seconds per run. This speed contrast indicates that the surrogate model is well suited to large-scale energy assessment and retrofit decision support.

Although the present case study focuses on Jinan, cross-region deployment is feasible because the model inputs explicitly encode building geometry, envelope thermal parameters, and occupant schedules, while climate effects can be incorporated through simulation-driven pre-training followed by adaptation. A practical pathway is domain adaptation via transfer learning: pre-train the model on a synthetic dataset generated under the target region's weather file, and then fine-tune using a limited number of measured or locally simulated samples from the new region. This approach is justified because many feature-energy relationships, such as envelope U-values, window-to-wall ratios, and occupancy intensity, are structurally shared across regions, whereas climate discrepancies can be corrected through limited local calibration. Future work could explore multi-climate pre-training or climate-informed embeddings to improve robustness across diverse climates.

Despite the computational scalability of inference, data availability remains the dominant practical bottleneck for real-world deployment. Current calibration is constrained by two key gaps. First, complete building metered datasets are scarce, meaning energy measurements for all households within the same building are often unavailable, which limits recalibration of graph-based inter-household interactions.

Second, the observed diversity of occupant behaviour profiles is limited, reducing calibration effectiveness for under-represented lifestyles. These gaps may prevent the model from fully closing the simulation-to-reality performance gap in some contexts. To strengthen deployment readiness, future work should prioritise collecting within-building, multi-household sub-metering and behaviourally diverse samples, such as additional household archetypes and event-driven occupancy patterns. With richer real-world datasets, transfer learning could update a broader set of parameters beyond the regression head, including message passing and edge-related parameters as well as selected encoders, thereby further improving generalisability.

5.3.3 Representativeness of the real-world energy use dataset

In terms of building characteristics, although the field dataset is limited in size, it shows consistent trends with the simulation dataset on several key attributes and aligns with the spatial features of typical ageing residential estates. For example, both datasets are dominated by low rise buildings, and staircase based configurations with two or three households per floor constitute the prevailing layout types. This indicates that the measured sample covers the most common building typologies and plan organisations found in legacy estates. However, the field sample is clearly skewed in floor area: 92.9% of measured households are larger than 60 m², whereas only 58.1% of simulated households exceed 60 m², implying that the fieldwork sample is concentrated in low rise and relatively larger units. In terms of occupant behaviour, retired households, student households, and office worker households appear in comparable proportions, while households with an infant are relatively under represented. The comparison of energy use distributions further clarifies what the dataset can and cannot support. Billing based measured annual EUI mainly falls within 30 to 90 kWh per square metre per year, with most samples concentrated in the 30 to 60 band. This provides useful evidence for relationships between building features, occupant behaviour, and energy use under low to medium intensity conditions. By contrast, the simulation dataset spans a much wider EUI range of 0 to

600 and is largely concentrated in 0 to 200, reflecting broader end use coverage and a wider spread of operating conditions that are not fully represented in the measured sample. As a result, the measured dataset offers limited empirical support for high consumption extremes.

It is important to state explicitly that the 120 household sample is not intended to be statistically representative of a wider population. Recruitment relied on intercept invitations and inclusion criteria centred on the availability of complete annual electricity bills and stable residence, which is appropriate for obtaining reliable measured anchors but not for population level inference. Accordingly, the appropriate generalisations from this dataset are methodological rather than demographic. It can be used to characterise the form and magnitude of simulation measurement errors for households similar to those surveyed, and to test whether a small set of measured samples can calibrate model outputs to the real world EUI scale. In contrast, it is not appropriate to extrapolate the observed distributions of EUI, household area, or behaviour archetype shares to the broader housing stock in Jinan, nor to make strong claims about model performance at the high EUI end where measured coverage is sparse. Consistent with this intended use, the measured dataset is primarily used to quantify the simulation measurement discrepancy and to evaluate calibration feasibility. As shown in Section 4.3.3, household level energy simulations based on the measured building and behaviour inputs reveal a structured deviation between simulated and measured EUIs, and the results further demonstrate that transfer learning can effectively align the prediction domain to the measured energy domain, providing a practical calibration pathway for data scarce community level assessments.

5.3.4 Simulation to real performance gap

The structured bias and outliers observed in Fig. 30 indicate a typical simulation to real performance gap. Specifically, 68.3% of the samples exhibit simulated EUI higher than measured EUI, and these cases are mainly concentrated in the low consumption

range (30 to 60 kWh/m^2). In contrast, 31.7% of the samples show simulated EUI lower than measured EUI, which tends to occur toward the higher consumption end (approximately 60 to 100 kWh/m^2). This pattern suggests a directional and structured mismatch driven by modelling assumptions and uncertainties in occupant behaviour.

For low consumption households, energy simulation often assumes fixed daily operating periods and full space utilisation, whereas in real operation occupants may use only part of the household or operate systems intermittently. Such partial time and partial space use can reduce annual energy use below what is expected under a normal occupancy assumption, leading to systematic overestimation. In addition, reporting errors and omissions in questionnaires can further amplify this bias, particularly when occupants have limited knowledge of appliance power ratings or fail to report intermittent but high impact devices, such as portable electric heaters, kitchen appliances, or post occupancy lighting retrofits. Conversely, for higher consumption households, cases where measured EUI is markedly higher than simulated EUI are often associated with behavioural shifts that increase occupancy intensity or alter operating conditions. For example, retired households may host grandchildren during winter or summer breaks, increasing time spent at home and adopting more extreme cooling or heating setpoints in certain rooms, which can substantially raise energy use. Similarly, unreported high power devices or atypical usage patterns can elevate measured consumption beyond simulation predictions.

Part of the discrepancy may also stem from temporal and boundary inconsistencies. Measured consumption corresponds to specific billing periods and the actual weather conditions of the monitored year, whereas simulations commonly use typical meteorological year files that may not be fully aligned with the measurement period, introducing additional deviations.

These findings highlight practical directions for improving both data collection and modelling to enhance calibratability. Field surveys should explicitly record vacancy and visitor periods, including their durations, during critical seasons such as summer holidays and the Spring Festival, and collect thermostat settings, ventilation habits, and high power appliances with greater granularity. Where feasible, nameplate

collection or short term metering can reduce under reporting and omissions. On the modelling and calibration side, robustness can be improved by assigning higher weights to samples that are more likely to exhibit larger bias, such as high floor or high power density cases. In addition, the measured dataset in this study is concentrated in a moderate EUI band, with relatively few high consumption households, which may limit calibration performance under extreme scenarios. Future surveys should therefore increase the representation of the high EUI end to test robustness and reduce extrapolation risk. Overall, Fig. 30 demonstrates that even when energy simulation is based on survey information, a measurable and structured simulation measurement gap remains. This directly motivates calibration strategies for models trained on large scale simulation data. Once the backbone has learned a transferable mapping from building features to energy use, a small measured sample is still required to align predictions to the real world EUI scale.

5.3.5 Transfer learning

The calibration results indicate that head-only transfer learning is effective because the pretrained backbone already captures transferable nonlinear mappings from household geometry, physical attributes, and behaviour encodings to energy outcomes in the simulation domain. When transferred to measured data, the dominant error manifests as domain-level bias and scaling mismatch rather than a failure of feature extraction. Updating only the regression head therefore provides a stable and data-efficient correction mechanism, particularly when measured supervision is sparse. In practice, this supports a two-stage workflow for neighbourhood retrofit assessment. First, automated energy modelling and large-scale simulation are used to train a transferable model. Second, a small measured sample is used to rapidly align the model to real-world EUI scales for the remaining unmetered households.

The proposed model achieves strong performance when trained and evaluated on the large-scale simulation dataset, with high R^2 and prediction accuracy, indicating that it

can effectively learn complex relationships between building characteristics, occupant behaviour representations, and energy outcomes within the simulation domain. To reduce the simulation-to-reality gap, the model was further calibrated using transfer learning on a field dataset of 120 households, where the calibrated model shows a clear reduction in prediction error on the 20 held-out validation set compared with the pre-calibration model.

However, the real-world measurement dataset remains limited in both scale and structure, which constrains the confidence of generalisation. First, the 120 records are household-level samples collected across multiple buildings, and do not provide complete coverage of all households within any single building. As a result, calibration was performed with message passing disabled, freezing the graph-based information propagation and updating only the final regression head. This design choice improves stability under sparse field supervision but may limit the model's ability to adapt its inter-household interaction mechanisms to real buildings. In particular, the parameters governing edge-related effects, which are intended to represent inter-unit thermal interactions, remain largely inherited from simulation training. Without measured adjacent-household energy data within the same building, these edge effects cannot be directly recalibrated, potentially affecting generalisation when inter-unit heat transfer is a dominant factor.

The field dataset also reflects a skewed coverage of occupant behaviour profiles. The dominant household types in the sample are retired households and working households with students, whereas other usage patterns (e.g., baby families and working households without students) are under-represented. This imbalance limits calibration effectiveness for those less-observed behavioural modes and may lead to larger errors when the model is applied to households with substantially different schedules and appliance-use patterns.

Despite these limitations, the observed reduction in validation error after transfer learning provides evidence that pre-training on large-scale simulation data followed by calibration with small, real-world samples is a viable technical pathway for bridging the simulation-to-reality performance gap. Future work could strengthen validation and

improve generalisability by collecting building-complete datasets that include energy measurements for all households within the same building, enabling calibration of graph edges and message passing, and by assembling a more behaviourally diverse sample spanning additional household types. With such datasets, transfer learning could be extended beyond the regression head to update a broader set of parameters, including GraphSAGE or message-passing weights, edge-related coefficients, and selected feature extractors such as CNN and MLP encoders, thereby further reducing prediction error and narrowing the gap between simulation-domain and real-domain performance.

5.4 Discussion of Case Study

This study optimises envelope retrofit strategies for four representative multi-household floor-plan configurations, selecting solutions that strike the best balance between energy-savings and upfront cost. This chapter discusses the optimisation outcomes, highlighting how the cost performance and the composition of Pareto-optimal packages differ by typology on the household level.

5.4.1 Discussion of typology-dependent optimisation

The optimisation results across the four representative floor plan typologies (Figs. 32 – 35) reveal three consistent and practically relevant findings. First, the most cost effective retrofit driver is strongly typology dependent, primarily due to differences in envelope composition and dominant heat transfer pathways across the prototypes. For typologies with a relatively high window to wall ratio and glazing dominated heat exchange (Typologies A and D), Pareto optimal solutions are mainly achieved by prioritising window upgrades, while roof and wall insulation provide secondary improvements. In Typology B, where roof exposure is pronounced and facade openings span multiple orientations, the Pareto front indicates that the best value for money region is typically characterised by combined window and roof measures;

once a moderate performance level is reached, wall insulation tends to play a supporting role. By contrast, Typology C features a larger exposed exterior wall area and a smaller glazing share; accordingly, the Pareto optimal region is dominated by wall and roof combinations, whereas window led packages are more likely to be cost inefficient. Overall, the typology specific shift in the primary driver implies that a uniform retrofit priority list is often suboptimal. Instead, retrofit sequencing should be selected according to which envelope component dominates heat loss or heat gain for a given layout.

Second, the Pareto fronts consistently exhibit diminishing returns at the high investment end, and the dominated clusters help explain why some packages are inefficient. Across typologies, many dominated solutions arise from unbalanced allocations. For example, high grade wall or roof insulation may be deployed while windows remain comparatively weak, or significant investment may be made in high performance windows while major opaque components are left under improved. Once the dominant heat transfer pathway has been substantially mitigated, further strengthening the same component yields smaller incremental savings and increases the cost per unit saving. This explains why efficient solutions tend to concentrate within a narrow band along the Pareto front, while the number of dominated points grows rapidly beyond it. Importantly, these patterns do not simply indicate that higher investment always delivers better outcomes; rather, they highlight the need to treat the envelope as an interacting system, in which the marginal value of any measure depends on the upgrade status of the other components.

Third, these findings support a staged retrofit strategy that can be generalised beyond the four prototypes. In the first stage, the dominant heat transfer pathway should be identified from typology characteristics and the corresponding primary upgrade should be implemented. For example, window first upgrades are appropriate for high window to wall ratio layouts, whereas wall insulation should be prioritised for layouts with large exposed wall areas. In the second stage, the next most influential component should be upgraded to strengthen the envelope and avoid bottlenecks, such as adding roof insulation after window upgrades, or applying targeted window improvements after

wall and roof measures. Finally, the most expensive measures, such as very high grade insulation or premium glazing, should be reserved for the final stage and adopted only when budgets allow or when higher performance targets are required, while recognising that marginal savings may be limited. In this way, the Pareto front patterns can be translated into an actionable sequencing framework that avoids dominated solutions and maximises cost effectiveness under different budget constraints.

5.4.2 Cost assumption and uncertainty

This thesis adopts a simplified retrofit cost model to compare the relationship between investment and energy performance across different floor plan typologies. Unit material costs were compiled exclusively from published literature, including product prices for external wall and roof insulation (Guo et al., 2021, Zheng et al., 2025); and window system prices (Guo et al., 2021, Liu et al., 2025). As reported in Table. 5 and 6, these sources typically provide prices in CNY per square metre. The initial investment cost for each scenario is therefore calculated as retrofitted area multiplied by the corresponding unit price, summed across envelope components.

Several assumptions were made to keep the optimisation framework lightweight and consistent with the core objective of this thesis, namely to examine how floor plan typology influences the relative prioritisation of envelope retrofit measures under a unified economic indicator. First, the model includes material costs only, while excluding labour, installation, and construction preliminaries such as scaffolding, vertical transport, site management, and safety measures. Second, literature reported unit prices are used directly, without converting them to a common base year or applying explicit regional price indices or adjustment factors. These simplifications inevitably introduce uncertainty in the absolute magnitude of the estimated costs, especially when sources span different publication years or reflect different market conditions and regions.

The excluded cost items have clear implications for bias. Omitting labour and

preliminaries will generally underestimate real implementation costs, and the underestimation is likely larger for external wall insulation than for roof insulation or window replacement, because facade works typically require more access systems and organised external operations. As a result, when measures differ substantially in labour intensity, the model may overstate the cost effectiveness of facade insulation relative to other options. In addition, demolition, waste disposal, and reinstatement are not included, such as removal of existing finishes, facade repair, window opening repairs, and local interior restoration after window replacement. This omission further lowers estimated costs and may affect window replacement and facade related measures more strongly. Finally, the model assumes a constant cost per square metre that does not vary with geometric complexity. In practice, irregular facades, balconies, and thermal bridge detailing can increase material waste and installation time, which means costs may be underestimated for complex facades and overestimated for simpler ones.

Overall, the reported investment costs should be interpreted as a lower bound proxy, so cost effectiveness metrics may appear optimistic. Nevertheless, because the same cost rules are applied consistently across all typologies and scenarios, the optimisation results remain useful for identifying relative trends. Future work could improve transparency by converting prices to a common base year, applying regional adjustments for the Jinan and Shandong context, separating material and non-material components, and testing robustness via a simple sensitivity analysis using plus or minus 10 to 20 percent changes in unit costs.

5.4.3 Interpretability and decision-maker-facing outputs

This study develops FusionGNN as a rapid energy prediction and retrofit decision-support model for multi-household residential buildings in China. By combining model-based energy estimation with retrofit cost calculation, the framework can evaluate a large number of retrofit packages and identify optimal solutions under different objectives. For practical deployment, FusionGNN represents each household

as a node and outputs household-level energy predictions under alternative retrofit scenarios, enabling multi-scale explanations tailored to different stakeholder groups.

For individual households, the model can compare scenarios and translate predicted energy changes into bill-relevant metrics, including annual electricity use and the corresponding change in yearly electricity cost under a local tariff. For example, residents can be shown the expected reduction in annual consumption and bill savings associated with different measures. This presentation helps residents understand the practical value of each option and can increase willingness to participate by linking retrofit benefits to reductions in electricity bills. For public decision-makers, FusionGNN can aggregate household-level predictions to the building level to estimate total saving potential and budget requirements for a single building or an entire community. This enables policy-relevant questions to be addressed directly, such as identifying the retrofit package that maximises savings under a fixed budget, or minimising cost while meeting a targeted reduction. Because the model can evaluate thousands of scenario combinations rapidly, it supports large-scale screening and prioritisation across communities. Beyond point predictions, the framework can generate a synthetic dataset of energy outcomes under systematically varied retrofit measures and then perform global sensitivity analysis, such as one-at-a-time screening, variance-based indices, or permutation importance, to rank measures by their marginal impact. This shifts the model from a black-box predictor to a decision-support tool that highlights which measures matter most for a given typology, climate, and occupancy profile.

To further enhance trust and transparency, several complementary interpretability techniques can be integrated. For example, feature importance can be derived from permutation tests or ablation experiments by removing specific retrofit inputs or behaviour features and observing the resulting performance drop. In addition, locally-fitted surrogate models, such as shallow decision trees or linear approximations, can be used to mimic FusionGNN behaviour for explanation and to reduce computational overhead when needed. Finally, household-level predictions can be visualised to map expected savings and costs across households and floors,

enabling customised retrofit guidance within the same building. Together, these approaches present model outputs as clear comparisons, ranked recommendations, and transparent trade-offs, supporting more informed retrofit decisions by both residents and public stakeholders.

6. Conclusion

Addressing the evaluation of energy retrofit strategies for multi-household residential buildings, this study develops an integrated pipeline that connects data, modelling, validation, and decision making, with Jinan as the demonstration region. The workflow begins with automated building energy modelling from floor plans, where architectural element recognition and vectorisation, household reconfiguration, and room function classification automatically produce simulation-ready preliminary models. These models are then enriched through parametric urban-context reconstruction, parameter refinement, and occupant behaviour definition, enabling large-scale simulations and the creation of a multi-household residential building energy simulation dataset. On this basis, a graph-structured, multi-modal prediction model, FusionGNN, is trained and calibrated using a small set of field measurements. Finally, four typical multi-household layouts are used as case studies, where FusionGNN's predictions are combined with cost evaluation and multi-objective optimisation to rapidly identify retrofit portfolios that balance energy-savings and economic performance.

6.1 Answers to the Research Questions

Automated building energy modelling. An automated workflow converts multi-household floor plans into simulation-ready preliminary models through element recognition and vectorisation, household reconfiguration, and room function

classification. The workflow reduces manual effort and turnaround time while maintaining the fidelity required for energy analysis, thereby improving the feasibility of portfolio-scale assessments.

High-quality dataset construction. Building on the preliminary models, urban context, physical parameters, and occupant schedules are incorporated to conduct large-scale simulations and form a multi-household residential building energy simulation dataset. The dataset integrates multi-modal information and adjacency relations at room, household, and building scales, supporting deep-learning training, ablation studies, and cross-scale evaluation.

Multi-scale feature contributions. Correlation analyses across room, household, and building scales show that the influence of the same type of feature is not uniform. At the room scale, window area and air changes per hour strongly affect energy intensity. At the household scale, spatial position and orientation lead to notable variation in use. At the building scale, the exterior wall U-value exerts a stronger effect on the overall baseline. Accordingly, window and airtightness upgrades should be prioritised at the room scale, top-floor and corner units should receive greater attention at the household scale, and exterior wall insulation should be emphasised at the building scale.

Multi-modal fusion and thermal-coupling modelling. The proposed FusionGNN fuses one-dimensional physical attributes, two-dimensional geometric information, and time-series occupancy, and explicitly encodes inter-unit thermal coupling via edge weights derived from shared envelope areas and U-values. Compared with single-modality or weak-fusion models, this integration improves predictive accuracy and robustness and more faithfully represents thermal interactions among units.

Energy–cost optimisation across layouts. For four typical legacy multi-household layouts, the optimal retrofit mix varies with plan typology. FusionGNN supports rapid scenario screening; combined with a cost model to form Pareto fronts, the analysis yields solution sets that balance energy-saving rates and capital cost, providing actionable guidance for retrofit decisions under budget constraints.

6.2 Contributions

To clearly distinguish different types of contributions, this thesis differentiates between methodological contributions, which comprise the proposed methods and workflows; empirical contributions, which draw on evidence and findings from the Chinese case study and validation; and strategic contributions, which demonstrate how the approach can inform retrofit decision-support practice and policy.

6.2.1 Methodological contributions

First, the thesis contributes an end-to-end workflow that connects automated floor-plan understanding, physics-based simulation and large-scale dataset generation, graph-based learning and calibration, and multi-objective optimisation into a complete and transferable pipeline. This workflow can support retrofit decision making for residential buildings in other cities and climate zones. Starting from existing floor plans, the approach can be adapted to local contexts by defining region-specific simulation parameter ranges, collecting a small set of measured samples for calibration, and estimating retrofit costs using local pricing assumptions. In this way, an energy prediction model aligned with the building characteristics of a target city can be developed to inform local retrofit planning.

Second, the automated floor plan interpretation and energy modelling workflow addresses a practical bottleneck in large-scale retrofit assessment for multi-household buildings. By converting complex multi-household floor plans into structured representations (architectural elements, rooms, households, and adjacency relationships), the thesis substantially reduces reliance on manual modelling. This is particularly relevant for ageing residential communities where digital building information and models are often unavailable and the building stock is large, making large-scale energy performance evaluation feasible.

Third, FusionGNN provides a calibratable multi-modal graph-learning architecture for multi-household energy prediction. Methodologically, the model integrates heterogeneous inputs (physical attributes, geometric representations, and occupancy

time series) while encoding thermal coupling through edge features derived from shared envelopes. This design enables the model to reflect inter-building heat transfer mechanisms that are often simplified or neglected in household-level predictions. The calibration strategy further strengthens the approach by correcting simulation-to-reality discrepancies using a limited amount of measured data.

Finally, the thesis integrates prediction with decision making through a multi-objective optimisation framework. By combining energy performance prediction with cost evaluation and Pareto-front screening, the pipeline enables rapid exploration of large numbers of retrofit packages. This methodological integration turns the predictive model into a practical decision-support tool, allowing practitioners to compare trade-offs and identify robust retrofit portfolios under budget constraints.

6.2.2 Empirical contributions

Empirically, the thesis contributes a large, multi-scale simulation dataset for Chinese multi-household residential buildings and demonstrates how it can be used to learn, test, and explain retrofit performance at different scales. The dataset captures room-, household-, and building-level descriptors alongside adjacency relationships, enabling scale-specific drivers to be quantified rather than inferred from qualitative assumptions. As a high-quality training resource, the dataset can also support the further development and benchmarking of building energy prediction models for residential housing.

The thesis further provides quantitative evidence on the simulation-to-measurement performance gap. By analysing measured household electricity data and matched household-level simulations, the work quantifies both the magnitude and direction of the discrepancy and demonstrates that transfer-learning calibration can reduce systematic error. These results indicate that simulation-trained models can be adapted more effectively to real operational contexts when only limited field data are available.

Through four representative layout case studies, the thesis also offers empirical

insights into how typology influences the most cost-effective retrofit strategy. The Pareto-front patterns show that the composition of cost-effective packages differs by typology, reflecting how envelope configuration and exposure conditions vary across plan typologies. This evidence is directly useful for selecting retrofit options that deliver high energy savings at relatively low cost, and for prioritising measures in a typology-oriented manner.

6.2.3 Conceptual or strategic contributions

At the strategic level, the FusionGNN-based energy prediction model, combined with cost-informed multi-objective optimisation, can support government decision-makers in designing retrofit programmes and can also be used to communicate expected energy savings to residents to increase participation. Specifically, this study addresses a key limitation of conventional retrofit planning, which often relies primarily on experience rather than quantified evidence of energy savings. The proposed approach enables rapid evaluation of energy performance across a large number of retrofit scenarios for different buildings, providing a scientific basis for developing and prioritising retrofit strategies. More importantly, the framework evaluates energy use at the household level, which helps move beyond retrofit interventions applied uniformly across a community or a building. This household-centred perspective supports multi-scale and customised retrofit decision making by accounting for systematic heterogeneity within the same community (e.g., differences in exposure, position, and inter-unit interactions). The cost-integrated multi-objective optimisation framework further helps public authorities identify the most cost-effective retrofit options under limited budgets. In addition, energy prediction under different retrofit scenarios enables annual energy savings to be estimated by comparing post-retrofit outcomes with baseline consumption. These savings can be translated into expected reductions in electricity bills, providing a clear and locally relevant message for public engagement and helping to strengthen residents' willingness to participate in retrofit initiatives.

6.3 Limitations

The limitations have been discussed at multiple points throughout the thesis; this section provides a consolidated summary of the main limitations across the workflow and outlines corresponding directions for future improvement.

(1) Robustness of automated floor-plan recognition and energy modelling. The automated generation of energy models from floor plans depends on the accuracy of floor plan recognition, which is sensitive to annotation styles, resolution, and drawing scale. For large and complex plans, omission or misclassification of doors, windows, and walls may occur, which can reduce the quality of the resulting energy models. As a result, manual checking and correction are still required at the current stage. Future work could improve robustness by expanding the floor plan dataset to include a wider range of resolutions and annotation conventions, and by introducing adaptive convolutional kernels (or other domain-robust architectures) to strengthen recognition performance, thereby improving the reliability of automatically generated energy models.

(2) Uncertainty in occupant behaviour modelling. In the large-scale simulations and subsequent feature construction, this study represents occupant behaviour using 24-hour occupancy profiles and 12-month occupancy states, and increases behavioural diversity by introducing stochastic variations in occupancy intensity and time shifts. However, real-world behaviour is inherently uncertain. For example, atypical events such as visits from family or friends can temporarily alter energy-use patterns beyond routine schedules. In future work, behavioural representation could be extended by incorporating explicit non-typical or event-driven usage states, enabling the model to better capture rare but influential deviations from standard patterns.

(3) Constraints on model interpretability. FusionGNN improves predictive performance by learning non-linear relationships across heterogeneous features and across neighbouring households, but this expressiveness introduces interpretability challenges for decision makers. Although the thesis provides cross-scale correlation

analyses and case-study explanations, graph-based deep models may still be perceived as “black boxes” in practical deployment, potentially reducing trust and adoption. Interpretability could be strengthened by reporting prediction uncertainty intervals and by incorporating physically consistent constraints or monotonicity checks for key parameters (e.g., U-values and ACH), so that model behaviour aligns more closely with engineering intuition and stakeholder communication needs. In addition, the current envelope representation relies primarily on U-values, which makes it difficult to capture material diversity and reduces transparency in explaining how specific construction choices influence predicted outcomes. Future work can address this by introducing richer thermal descriptors (e.g., thermal resistance, layer thickness, and other relevant parameters) to better represent envelope thermal properties, thereby improving the model’s expressiveness and interpretation.

(4) Transferability and scalability of the prediction model. FusionGNN predicts household-level EUI by integrating building geometry, occupant behaviour, physical attributes, and adjacency relationships; however, climate conditions are not explicitly encoded, which may constrain transferability and scalability across regions and climates. Future work should test the full workflow in different climate zones, build cross-climate energy datasets, and explicitly introduce climate features into the prediction model to enhance transferability. In addition, the current study uses Jinan as the case study, and the resulting dataset is constrained by the local building characteristics and energy-use levels in Jinan. Future work can extend the workflow to other cities and use lightweight transfer learning strategies, supported by a small number of additional simulations or measured samples, to rapidly capture city-specific building and energy characteristics. This would allow the model to align its predictions with the target city domain more efficiently, thereby improving scalability for broader deployment.

(5) Limited scale and diversity of measured energy data. Although the field survey and electricity bill data provide an essential real-world reference, the measured dataset remains limited in scale and diversity. Moreover, the current samples often do not form complete building graphs (i.e., metered data for all households within the same

building), which constrains calibration of inter-household thermal coupling effects. Future work should expand measurement coverage to improve representativeness, prioritise within-building complete metering where possible, and collect finer-grained behavioural and system-operation signals (e.g., window-opening ventilation, HVAC operation logs) to reduce noise and improve calibration stability.

(6) Simplifications in the cost model. The optimisation analysis adopts a simplified cost model that mainly includes material costs and directly uses unit prices reported in the literature, without full normalisation to a common base year or regional price adjustment. Therefore, absolute cost estimates should be interpreted as a lower-bound approximation, and relative cost-effectiveness may be biased when measures differ substantially in labour intensity and delivery requirements (e.g., scaffolding, transport, and management). Future work could incorporate labour, enabling works, demolition and reinstatement, and other delivery-related costs; introduce regional price indices and base-year conversion; and extend the assessment from upfront investment to life-cycle cost or cost-effectiveness under uncertainty.

6.4 Future work

Future work can build on the above limitations and extend the proposed framework along three complementary directions.

First, expand and diversify floor plan data and improve the robustness of automated energy modelling. Future studies should broaden the floor plan dataset to cover a wider range of residential typologies, drawing styles, and levels of plan complexity across regions and cultures. This includes improving tolerance to low-quality scans, mixed annotations, and heterogeneous drafting standards. Methodologically, robustness can be further strengthened through domain-adaptive vision models and recognition strategies that generalise across styles (e.g., style augmentation, domain adaptation, and more robust parsing modules), reducing the need for manual

checking and correction when generating energy models at scale.

Second, enrich field measurements and strengthen model transfer to real-world operation. While the current thesis demonstrates calibration with limited measured data, future efforts should prioritise richer and more representative measurement campaigns. In particular, collecting metered data for all households within the same building, together with sub-metered end-use information and behavioural signals (e.g., HVAC operation, window opening, setpoints), would enable stronger calibration of inter-household thermal coupling and reduce noise in the training process. Building on this, transfer learning and domain adaptation can be extended beyond regression-head fine-tuning to more systematic approaches, improving predictive accuracy and stability across diverse real settings.

Third, enhance transferability, interpretability, and scalability through introducing climate features and multi-physics representations. Future work should explicitly incorporate meteorological and microclimate variables as model inputs, enabling climate-conditioned prediction and improving generalisation across climate zones. At the building-physics level, envelope and system representations can be enriched beyond nominal U-values by introducing additional descriptors such as thermal resistance, layer thickness, thermal mass indicators, SHGC, and system control parameters where available. These additions would improve the model's expressiveness and the physical interpretability of learned relationships, while also reducing the need for extensive re-simulation when extending the framework to new retrofit measures. Finally, scalability can be improved by leveraging lightweight transfer learning when extending the workflow from the Jinan case study to other cities.

In summary, this thesis presents a practical path from floor plan to decision. By unifying automated modelling, multi-modal learning with explicit thermal coupling, and multi-objective optimisation, it advances methods for energy assessment of multi-household housing and provides directly applicable tools and evidence for retrofit package selection and investment planning. The proposed future directions further outline how the framework can evolve towards more robust automation,

stronger real-world validity, and wider cross-region deployment.

Appendix A

Household Appliances and Occupant Behaviour Data Collection Sheet

Completion notes

- Time format: Use the 24-hour clock (e.g., 06:30-08:00).
- Recommended occupancy intensity values: 0 = unoccupied; 0.2 = low (short stay / few people); 0.6 = medium (normal activities); 1.0 = high (multiple people / intensive activities). If unsure, you may use only 0 / 0.6 / 1.0.
- Active time periods format: start-end@intensity; separate multiple periods with semicolons. Example: 06:00-08:00@0.6; 18:00-22:00@0.8; 22:00-06:00@1.0
- Months of use: Write months as a list or range from 1-12, e.g., 6-9 or 1,2,12.
- If a room does not exist: mark 'No' in the room list; related tables may be left blank.

A. Household basic information

Item	Value
Residential community name	
Building / block code	
Floor level	
Total number of floors	
Main orientation (South / South-North / East-West)	
Floor area (m²)	
Construction period / year	
Number of occupants	
Layout (No. bedrooms / living rooms / bathrooms)	
Household type (working adults /	

retired / student / with infants / other)	
Vacant months per year	
Stair-to-household ratio	

B. Envelope and openings

Item	Value
Window frame material (timber / uPVC / aluminium / steel / other)	
Number of glazing layers (single / double / triple / other)	
Entrance door (steel / aluminium alloy / timber / glass / other)	
Internal doors (steel / aluminium alloy / timber / glass / other)	

C. Room list

ID	Room type	Area (m², optional)	Air-conditioning (Y/N)	Heating (Y/N)	Main window orientation (N/E/S/W/multi)	Notes
R1	Living room					
R2	Bedroom 1					
R3	Bedroom 2					
R4	Bedroom 3					
R5	Bedroom 4					
R6	Kitchen					

R7	Bathroom 1					
R8	Bathroom 2					
R9	Balcony					
R10	Utility					
R11	Other (please specify)					

D. Room usage frequency

ID	Room type	Inactive intensity (default 0 or 0.1)	Weekday active periods (start-end; use ';' for multiple)	Weekend active periods	Months of use (1-12 or ranges)
R1	Living				
R2	Bedroom 1				
R3	Bedroom 2				
R4	Bedroom 3				
R5	Bedroom 4				
R6	Kitchen				
R7	Bathroom 1				
R8	Bathroom 2				
R9	Balcony				
R10	Utility				
R11	Other (please specify)				

E. Cooling (air-conditioning)

AC unit	Served room IDs (e.g., R1,R2)	Type (central / floor-standing / wall-mounted / other)	Rated cooling capacity (kW)	Typical set-point (°C)	Months of use	Operating periods

AC 1						
AC 2						
AC 3						
AC 4						
AC 5						

F. Heating

Item	Value
Heating type (electric heater / AC heating / gas boiler / underfloor / other)	
Heating season months (e.g., 11-3)	
Typical set-point (°C) (e.g., 18-22)	
Operation mode (all-day / scheduled / intermittent; note start-end periods)	
Controllability (dwelling/room control via valves/thermostat / not controllable)	
Rooms consistently turned off (room IDs)	

ID	Heated (Y/N)	Heating frequently turned off (Y/N)	Notes
R1			
R2			
R3			

R4			
R5			
R6			
R7			
R8			
R9			
R10			
R11			

G. Ventilation and window opening

ID	Winter window-opening frequency	Summer window-opening frequency	Typical window-opening periods	Extract fan use (kitchen/bathroom; frequency)	Notes
R1					
R2					
R3					
R4					
R5					
R6					
R7					
R8					
R9					
R10					
R11					

H. Lighting

ID	Main lamp type (incandescent / fluorescent / LED / other)	Total lighting power in this room (W, estimated)	Weekday use periods	Weekend use periods	Months of use	Notes
R1						
R2						
R3						
R4						
R5						

R6						
R7						
R8						
R9						
R10						
R11						

I. Appliances and plug loads

Add rows as needed.

Category	Quantity	Rated power (W)	Room ID	Always on (Y/N)	Use frequency (times/week or h/day)	Duration (min or h)	Months of use	Time periods of use

J. Domestic hot water

Item	Value
Water-heating system type (electric storage heater / gas water heater / solar / other)	
Capacity / power (e.g., 60 L or 2000 W)	
Mode (instant / rapid / storage)	
Average shower frequency (per person per day)	
Typical hot water temperature (°C, estimated)	
Typical use periods (e.g., 19:00-22:00)	

K. Annual energy information

Item	Value
Annual electricity cost (CNY)	
Annual electricity use (kWh, if known)	
Billing period (e.g., 2023.4-2024.4)	

Privacy statement: This questionnaire is used only for academic research and model calibration. Data will be anonymised and stored and deleted in accordance with research ethics requirements.

Appendix B

Participant Information Sheet

Critical building parameters and occupant behaviour impacting building energy performance

You are invited to take part in a research project being carried out by researchers at the University of Sheffield. Before deciding whether to take part it is important you understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you like more information – contact details are below. Take time to decide whether or not your wish to take part. Thanks for reading this.

1. What is the project's purpose?

The aim of the project is to identify critical building parameters and occupant behaviour that affects building energy performance. A dataset will be developed from local households in Jinan, and the critical parameters will be identified through sensitivity analysis. The identified parameters can help the decision-making for building energy-efficient retrofit considerations, thus reducing the building energy consumption and carbon emissions.

2. What will happen to me if I take part?

You will first be asked to sign a consent form, and you can freely choose whether to sign it or not. If you sign the consent form, your house information and annual household electricity consumption will be collected through either face-to-face interviews or online questionnaires. The house information includes your residential community name, floor, occupancy density, and power density and operation hours of your household appliances. The interview or questionnaire should take about 20

minutes to complete, during the data collection process, you can ask to have a break or to terminate the collection at any time.

3. What are the possible disadvantages and risks of taking part?

Taking part in this research is entirely voluntary. If you decide to take part you will be asked to sign a consent form. You can still withdraw at any time without giving a reason. The occupant's behaviour and schedule may involve your privacy, to ensure personal privacy security, all the data will be anonymous and will not be shared with a third party.

4. What are the possible benefits of taking part?

This research would help to conduct the energy-efficiency retrofit for your residential community. You will also be contributing to an important piece of scientific research that will influence how the optimal energy-efficiency retrofit scenarios be selected.

5. What happens if the research study stops earlier than expected?

It is possible we may complete a session earlier than the 20 minutes anticipated. In this situation, we will finish earlier.

6. What if something goes wrong?

If something goes wrong in the experiment and you are unhappy with how you have been treated you can raise a complaint with the research project lead or the University. For this research, Dr. Tsung-Hsien Wang is the overall lead for the whole project (tsung-hsien.wang@sheffield.ac.uk). If you feel your complaint was not handled satisfactorily you can also contact the University’s Registrar and Secretary at registrar@sheffield.ac.uk.

7. Will taking part in this project be kept confidential?

All personal data collected from you will be stored confidentially and will only be used in an aggregated, anonymised format. You will not be identifiable in any research reports or papers, or other publications. The data collected from you during the research sessions will be collated along with data collected from other participants and analysed in order to draw conclusions about the research questions being addressed during this project. The findings from this research may be published in an academic journal or on the University's website but you will not be identifiable and individual, personal data will not be published. If you would like to be updated about the findings from the research project we will send you any resultant reports or papers once complete. Your anonymised data may be used in future research projects, but again you will never be identified.

8. What will happen to the results of the research project?

The results will be published in academic journals and may also be presented at academic conferences. The results will also be used as evidence during the review and revisions of standards, guidelines, and design criteria related to the building energy-efficiency retrofits.

9. Who is organising and funding the research?

The project is being undertaken by Xingjian Zhao, a PhD student in the School of Architecture, and is led by Dr. Tsung-Hsien Wang.

10. Who has ethically reviewed the project?

The project has been approved by the School of Architecture's Ethics Review Committee at the University of Sheffield.

11. Contact for further information

Please contact Dr. Tsung-Hsien Wang, project lead, at tsung-hsien.wang@sheffield.ac.uk.

Thank you for reading this. We hope you would like to take part in this research project. To do so please contact Xingjian Zhao at xzhao66@sheffield.ac.uk.

Appendix C

The University Of Sheffield.

[A graph-based multi-modal deep learning framework for automated energy performance assessment and retrofit decision support in Chinese residential buildings]

Consent Form

Please tick the appropriate boxes	Yes	No
Taking Part in the Project		
1. I have read and understood the project information sheet or the project has been fully explained to me. (If you answer No to this question please do not proceed with this consent form until you are fully aware of what your participation in the project involves)	☐	☐
2. I have been given the opportunity to ask questions about the project.	☐	☐
3. I agree to take part in the project. I understand that taking part in the project will include being interviewed and recorded.	☐	☐
4. I understand that I can withdraw my data from the study at any time; I do not have to give any reasons for why I no longer want to take part and there will be no adverse consequences if I choose to withdraw.	☐	☐
How my information will be used during and after the project		
1. I understand my personal details such as my name, phone number, address and email address will not be revealed to people outside the project.	☐	☐
2. I understand and agree that my words may be quoted in publications, reports, web pages, and other research outputs. I understand that I will not be named in these outputs unless I specifically request this.	☐	☐
3. I give permission for an anonymised transcript of my interview to be deposited in University of Sheffield's Data Repository and Registry (ORDA), so it can be used for future research and learning.	☐	☐
4. I understand and agree that other authorised researchers will have access to this data only if they agree to preserve the confidentiality of the information as requested in this form.	☐	☐
5. I understand and agree that other authorised researchers may use my data in publications, reports, web pages, and other research outputs, only if they agree to preserve the confidentiality of the information as requested in this form.	☐	☐
So that the information you provide can be used legally by the researchers		
I agree to assign the copyright I hold in any materials generated as part of this project to The University of Sheffield.	☐	☐

Name of participant Signature Date

Name of Researcher: Xingjian Zhao Signature Date

Project contact details for further information:
Researcher:
Xingjian Zhao, xzhao66@sheffield.ac.uk, Arts Tower, University of Sheffield, +447379637887/ +8619853706002.

The template of this consent form has been approved by the University of Sheffield Research Ethics Committee and is available to view here: https://www.sheffield.ac.uk/rs/ethicsandintegrity/ethicspolicy/further-guidance/homepage

References

AHMED, S., LIWICKI, M., WEBER, M. & DENGEL, A. Improved automatic analysis of architectural floor plans. 2011 International conference on document analysis and recognition, 2011. IEEE, 864-869.

AHMED, S., LIWICKI, M., WEBER, M. & DENGEL, A. Automatic room detection and room labeling from architectural floor plans. 2012 10th IAPR international workshop on document analysis systems, 2012. IEEE, 339-343.

AHN, Y. & KIM, B. S. 2022. Prediction of building power consumption using transfer learning-based reference building and simulation dataset. *Energy and Buildings,* 258**,** 111717.

ALI, U., BANO, S., SHAMSI, M. H., SOOD, D., HOARE, C., ZUO, W., HEWITT, N. & O'DONNELL, J. 2024. Urban building energy performance prediction and retrofit analysis using data-driven machine learning approach. *Energy and Buildings,* 303**,** 113768.

AMASYALI, K. & EL-GOHARY, N. 2021. Machine learning for occupant-behavior-sensitive cooling energy consumption prediction in office buildings. *Renewable and Sustainable Energy Reviews,* 142**,** 110714.

AMASYALI, K. & EL-GOHARY, N. 2022. Hybrid approach for energy consumption prediction: Coupling data-driven and physical approaches. *Energy and Buildings,* 259**,** 111758.

AN, J., WU, Y., GUI, C. & YAN, D. Chinese prototype building models for simulating the energy performance of the nationwide building stock. Building Simulation, 2023. Springer, 1559-1582.

ANAN, M., KANAAN, K., BENHADDOU, D., NASSER, N., QOLOMANY, B., TALEI, H. & SAWALMEH, A. 2024. Occupant-Aware Energy Consumption Prediction in Smart Buildings Using a LSTM Model and Time Series Data.

BAEK, J., PARK, H. & CHANG, S. 2022. Enhanced LSTM-based community energy consumption prediction model leveraging shared building cluster datasets. *Journal of building performance simulation,* 15**,** 717-734.

BALDWIN, A. N., LOVEDAY, D. L., LI, B., MURRAY, M. & YU, W. 2018. A research agenda for the retrofitting of residential buildings in China–A case study. *Energy Policy,* 113**,** 41-51.

BANFI, A., FERRANDO, M., LI, P., SHI, X. & CAUSONE, F. 2024. Integrating Occupant Behaviour into Urban-Building Energy Modelling: A Review of Current Practices and Challenges. *Energies,* 17**,** 4400.

BANSAL, A., BALAJI, K. & LALANI, Z. 2025. Temporal encoding strategies for energy time series prediction. *arXiv preprint arXiv:2503.15456*.

BASET, A. & JRADI, M. 2024. Data-Driven Decision Support for Smart and Efficient Building Energy Retrofits: A Review. *Applied System Innovation,* 8**,** 5.

BLOCH, T. & SACKS, R. 2018. Comparing machine learning and rule-based inferencing for semantic enrichment of BIM models. *Automation in Construction,* 91**,** 256-272.

BUYUKLIEVA, B. 2020. Machine Learning for Exploring Spatial Affordance Patterns. *arXiv preprint arXiv:2005.08106*.

CAO, J., ZHANG, H., SAVOV, A., HALL, D. & DILLENBURGER, B. Energy-aware design: predicting building performance from layout graphs. Proceedings of the 2022 European Conference on Computing in Construction, 2022. University of Turin, 130-137.

CHANG, J., LV, Y., WANG, J., PANG, H. & LIU, Y. 2025. Raster Image-Based House-Type Recognition and Three-Dimensional Reconstruction Technology. *Buildings,* 15**,** 1178.

CHEN, G., LU, S., ZHOU, S., TIAN, Z., KIM, M. K., LIU, J. & LIU, X. 2025. A Systematic Review of Building Energy Consumption Prediction: From Perspectives of Load Classification, Data-Driven Frameworks, and Future Directions. *Applied Sciences,* 15**,** 3086.

CHEN, X., CIMILLO, M., CHOW, D. & CHEN, B. 2024. Assessing the Energy Performance and Retrofit Potential of the 1980–1990s' Residential Building Stock in China's Jiangsu Province: A Simulation-Based Study. *Energies,* 17**,** 1260.

CHEN, X., SINGH, M. M. & GEYER, P. Component-based machine learning for predicting representative time-series of energy performance in building design. 28th International Workshop on Intelligent Computing in Engineering, Berlin, 2021.

CHEN, Y., GUO, M., CHEN, Z., CHEN, Z. & JI, Y. 2022. Physical energy and data-driven models in building energy prediction: A review. *Energy Reports,* 8**,** 2656-2671.

CHEN, Y., HONG, T., LUO, X. & HOOPER, B. 2019. Development of city buildings dataset for urban building energy modeling. *Energy and Buildings,* 183**,** 252-265.

CHINA, N. B. O. S. O. 2019. China statistical yearbook 2019. *National Bureau of Statistics of China*.

CHO, D., KIM, J., SHIN, E., CHOI, J. & LEE, J. K. Recognizing architectural objects in floor-plan drawings using deep-learning style-transfer algorithms. 25th International Conference on Computer-Aided Architectural Design Research in Asia, CAADRIA 2020, 2020. The Association for Computer-Aided Architectural Design Research in Asia …, 719-727.

CONGXIANG, T., AHMAD, N. A., ABD RASED, A. N. N. W., SUQI, W. & HAINING, T. 2024. Establishing Energy-efficient Retrofitting Strategies in Rural Housing in China: A Systematic Review. *Results in Engineering***,** 103653.

CRISTINO, T. M., NETO, A. F., WURTZ, F. & DELINCHANT, B. 2022. The evolution of knowledge and trends within the building energy efficiency field of knowledge. *Energies,* 15**,** 691.

DALACH, A., WANG, Z., NOUSIAS, S. & BORRMANN, A. 2025. Exploring building energy performance prediction using graph neural networks.

DE LAS HERAS, L.-P., AHMED, S., LIWICKI, M., VALVENY, E. & SáNCHEZ, G. 2014. Statistical segmentation and structural recognition for floor plan interpretation: Notation invariant structural element recognition. *International Journal on Document Analysis and Recognition (IJDAR),* 17**,** 221-237.

DEB, C. & SCHLUETER, A. 2021. Review of data-driven energy modelling techniques for building retrofit. *Renewable and Sustainable Energy Reviews,* 144**,** 110990.

DENG, K., CUI, Y., DENG, Q., LIU, R., CHEN, Z. & WANG, S. 2025. Multi-Objective Optimization of Urban Residential Envelope Structures in Cold Regions of China Based on Performance and Economic Efficiency. *Buildings,* 15**,** 2365.

DENG, Q., WANG, G., WANG, Y., ZHOU, H. & MA, L. 2021. A quantitative analysis of the impact of residential cluster layout on building heating energy consumption in cold IIB regions of China. *Energy and Buildings,* 253**,** 111515.

DENG, W., XIE, J. & PENG, Z. 2018. Material transitions and associated embodied energy input of rural buildings: case study of Qinyong Village in Ningbo China. *Sustainability,* 10**,** 2016.

DENG, Y., YUE, Z., WU, Z., LI, Y. & WANG, Y. 2024. TCN-Attention-BIGRU: Building energy modelling based on attention mechanisms and temporal convolutional networks. *Electronic Research Archive,* 31.

DINO, I. G., SARI, A. E., ISERI, O. K., AKIN, S., KALFAOGLU, E., ERDOGAN, B., KALKAN, S. & ALATAN, A. A. 2020. Image-based construction of building energy models using computer vision. *Automation in Construction,* 116**,** 103231.

DONG, Q., MA, Z. & SUN, C. Occupancy of rooms in urban residential buildings by users in cold areas of China. Building Simulation, 2023. Springer, 483-497.

DONG, S., WANG, W., LI, W. & ZOU, K. 2021a. Vectorization of floor plans based on EdgeGAN. *Information,* 12**,** 206.

DONG, Z., LIU, J., LIU, B., LI, K. & LI, X. 2021b. Hourly energy consumption prediction of an office building based on ensemble learning and energy consumption pattern classification. *Energy and Buildings,* 241**,** 110929.

EL-DARWISH, I. & GOMAA, M. 2017. Retrofitting strategy for building envelopes to achieve energy efficiency. *Alexandria engineering journal,* 56**,** 579-589.

FELTES, M., AHMED, S., DENGEL, A. & LIWICKI, M. Improved contour-based corner detection for architectural floor plans. Graphics Recognition. Current Trends and Challenges: 10th International Workshop, GREC 2013, Bethlehem, PA, USA, August 20-21, 2013, Revised Selected Papers 10, 2014. Springer, 191-203.

FENG, K., LU, W., WANG, Y. & MAN, Q. 2022. Energy-efficient retrofitting under incomplete information: a data-driven approach and empirical study of Sweden. *Buildings,* 12**,** 1244.

FINA, B., AUER, H. & FRIEDL, W. 2019. Profitability of active retrofitting of multi-apartment buildings: Building-attached/integrated photovoltaics with special consideration of different heating systems. *Energy and Buildings,* 190**,** 86-102.

GEYER, P. & SINGARAVEL, S. 2018. Component-based machine learning for performance prediction in building design. *Applied energy,* 228**,** 1439-1453.

GIMENEZ, L., ROBERT, S., SUARD, F. & ZREIK, K. 2016. Automatic reconstruction of 3D building models from scanned 2D floor plans. *Automation in Construction,* 63**,** 48-56.

GUO, H., HU, Z. & LI, W. 2024. Effectiveness evaluation of owners' driving force of residential buildings energy-saving retrofit using ANP-FCE in China. *Environment, Development and Sustainability***,** 1-21.

GUO, H., SONG, W., SHAO, Y. & CALAUTIT, J. K. 2021. Influence of Wall Insulation Structure on Summer Overheating Risk of Residential Buildings in the Severe Cold Region of China: A Case Study Inharbin. *Available at SSRN 4344184*.

GUO, H., ZHOU, S., QIN, T., HUANG, L., SONG, W. & YIN, X. 2020. Energy sustainability of bio-based building materials in the cold and severe cold regions of China—A case study of residential buildings. *Applied Sciences,* 10**,** 1582.

HALAçLı, E. G., CANLı, İ., İŞERI, O. K., YAVUZ, F., AKGüL, Ç. M., KALKAN, S. & DINO, I. G. A Novel Graph Neural Network for Zone-Level Urban-Scale Building Energy Use Estimation. Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, 2023. 169-176.

HE, K., ZHANG, X., REN, S. & SUN, J. 2015. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE transactions on pattern analysis and machine intelligence,* 37**,** 1904-1916.

HE, Q., NG, S. T., HOSSAIN, M. U. & AUGENBROE, G. L. 2020. A data-driven approach for sustainable building retrofit—A case study of different climate zones in China. *Sustainability,* 12**,** 4726.

HUANG, H., NAZI, W. I. B. W. M., YU, Y. & WANG, Y. 2020. Energy performance of a high-rise residential building retrofitted to passive building standard–A case study. *Applied Thermal Engineering,* 181**,** 115902.

HUANG, L., WAN, G. & LIU, C. An improved parallel thinning algorithm. ICDAR, 2003. Citeseer, 780.

HUANG, W. & XU, Q. 2024. Sustainable-Driven Renovation of Existing Residential Buildings in China: A Systematic Exploration Based on Review and Solution Approaches. *Sustainability,* 16**,** 3895.

HUANG, W. & ZHENG, H. Architectural drawings recognition and generation through machine learning. Proceedings of the 38th annual conference of the association for computer aided design in architecture, Mexico City, Mexico, 2018. 18-20.

IBRAHIM, D. M., ALMHAFDY, A., AL-SHARGABI, A. A., ALGHIETH, M., ELRAGI, A. & CHICLANA, F. 2022. The use of statistical and machine learning tools to accurately quantify the energy performance of residential buildings. *PeerJ Computer Science,* 8**,** e856.

IMANI, E., DAWOOD, H., WILLIAMS, S. & DAWOOD, N. 2025. Physics-based and data-driven retrofitting solutions for energy efficiency and thermal comfort in the UK: IoT-validated analysis. *Buildings,* 15**,** 1050.

JIA, Y., WANG, J., HOSSEINI, M. R. & SHOU, W. Graph neural networks in building life cycle: A review. EC3 Conference 2022, 2022. European Council on Computing in Construction, 0-0.

JIA, Y., WANG, J., HOSSEINI, M. R., SHOU, W., WU, P. & MAO, C. 2024. Temporal graph attention network for building thermal load prediction. *Energy and Buildings,* 321**,** 113507.

JIAO, L. & RONG, X. 2022. Analysis of Principal Factors on Energy Consumption of Expressway Service Buildings. *Energies,* 15**,** 4392.

KALERVO, A., YLIOINAS, J., HäIKIö, M., KARHU, A. & KANNALA, J. Cubicasa5k: A dataset and an improved multi-task model for floorplan image analysis. Image Analysis: 21st Scandinavian Conference, SCIA 2019, Norrköping, Sweden, June 11–13, 2019, Proceedings 21, 2019. Springer, 28-40.

KANNARI, L., KILJANDER, J., PIIRA, K., PIIPPO, J. & KOPONEN, P. 2021. Building heat demand forecasting by training a common machine learning model with physics-based simulator. *Forecasting,* 3**,** 290-302.

KIAVARZ, H., JADIDI, M. & ESMAILI, P. 2023. A graph-based explanatory model for room-based energy efficiency analysis based on BIM data. *Frontiers in Built Environment,* 9**,** 1256921.

KIM, T.-Y. & CHO, S.-B. 2019. Predicting residential energy consumption using CNN-LSTM neural networks. *Energy,* 182**,** 72-81.

KISTELEGDI, I., HORVáTH, K. R., STORCZ, T. & ERCSEY, Z. 2022. Building geometry as a variable in energy, comfort, and environmental design optimization—a review from the perspective of architects. *Buildings,* 12**,** 69.

KOPONEN, P., NISKA, H. & MUTANEN, A. Mitigating the weaknesses of machine learning in short–term forecasting of aggregated power system active loads. 2019 IEEE 17th International Conference on Industrial Informatics (INDIN), 2019. IEEE, 303-310.

KRATOCHVILA, L., DE JONG, G., ARKESTEIJN, M., BILIK, S., ZEMCIK, T., HORAK, K. & RELLERMEYER, J. S. 2024. Multi-Unit Floor Plan Recognition and Reconstruction Using Improved Semantic Segmentation of Raster-Wise Floor Plans. *arXiv preprint arXiv:2408.01526*.

LANG, M., LANE, R., ZHAO, K., THAM, S., WOOLFE, K. & RAVEN, R. 2021. Systematic review: Landlords' willingness to retrofit energy efficiency improvements. *Journal of cleaner production,* 303**,** 127041.

LANG, S. 2004. Progress in energy-efficiency standards for residential buildings in China. *Energy and Buildings,* 36**,** 1191-1196.

LI, J., YU, Z. J., HAGHIGHAT, F. & ZHANG, G. 2019a. Development and improvement of occupant behavior models towards realistic building performance simulation: A review. *Sustainable Cities and Society,* 50**,** 101685.

LI, X. & WU, Y. 2025. A review of complex window-glazing systems for building energy-saving and daylight comfort: Glazing technologies and their building performance prediction. *Journal of Building Physics,* 48**,** 496-540.

LI, X. & YAO, R. 2021. Modelling heating and cooling energy demand for building stock using a hybrid approach. *Energy and Buildings,* 235**,** 110740.

LI, Y., O'NEILL, Z., ZHANG, L., CHEN, J., IM, P. & DEGRAW, J. 2021. Grey-box modeling and application for building energy simulations-A critical review. *Renewable and Sustainable Energy Reviews,* 146**,** 111174.

LI, Z., DAI, J., CHEN, H. & LIN, B. An ANN-based fast building energy consumption prediction method for complex architectural form at the early design stage. Building Simulation, 2019b. Springer, 665-681.

LIANG, X., YU, T. & GUO, L. 2017. Understanding stakeholders' influence on project success with a new SNA method: A case study of the green retrofit in China. *Sustainability,* 9**,** 1927.

LIAO, X., WANG, C., LI, B., LI, B. & DU, C. 2025. Professional Barriers in Energy Efficiency Retrofits—A Solution Based on Information Flow Modeling. *Buildings (2075-5309),* 15.

LIU, C., SCHWING, A. G., KUNDU, K., URTASUN, R. & FIDLER, S. Rent3d: Floor-plan priors for monocular layout estimation. Proceedings of the IEEE conference on computer vision and pattern recognition, 2015. 3413-3421.

LIU, C., SHARPLES, S. & MOHAMMADPOURKARBASI, H. 2023. A review of building energy retrofit measures, passive design strategies and building regulation for the low carbon development of existing dwellings in the hot summer–cold winter region of China. *Energies,* 16**,** 4115.

LIU, C., WU, J., KOHLI, P. & FURUKAWA, Y. Raster-to-vector: Revisiting floorplan transformation. Proceedings of the IEEE International Conference on Computer Vision, 2017. 2195-2203.

LIU, G., LI, X., TAN, Y. & ZHANG, G. 2020. Building green retrofit in China: Policies, barriers and recommendations. *Energy Policy,* 139**,** 111356.

LIU, J., QIAN, Y., YANG, Y. & YANG, Z. 2022. Can artificial intelligence improve the energy efficiency of manufacturing companies? Evidence from China. *International Journal of Environmental Research and Public Health,* 19**,** 2091.

LIU, P., LI, M., ZHU, J. & YUAN, W. 2025. Study on energy retrofits for rural residential envelopes in Northwest China. *Scientific Reports,* 15**,** 1-22.

LU, Z., WANG, T., GUO, J., MENG, W., XIAO, J., ZHANG, W. & ZHANG, X. 2021. Data-driven floor plan understanding in rural residential buildings via deep recognition. *Information Sciences,* 567**,** 58-74.

LUO, X. & DU, L. Energy consumption simulations of rural residential buildings considering differences in energy use behavior among family members. Building Simulation, 2024. Springer, 1335-1358.

LV, X., ZHAO, S., YU, X. & ZHAO, B. Residential floor plan recognition and reconstruction. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2021. 16717-16726.

MA, D., LI, X., LIN, B. & ZHU, Y. 2023. An intelligent retrofit decision-making model for building program planning considering tacit knowledge and multiple objectives. *Energy,* 263**,** 125704.

MA, M., CAI, W. & WU, Y. 2019. China act on the energy efficiency of civil buildings (2008): A decade review. *Science of The Total Environment,* 651**,** 42-60.

MACíAS, J. S., TELLO, C. P., RAMíREZ, A. A., ARISTA, A. L., ALMAGUER, H. M., ESCOBEDO, P. R. & PUENTE, A. R. 2018. Assessment of electrical saving from energy efficiency programs in the residential sector in Mexicali, Mexico. *Sustainable Cities and Society,* 38**,** 795-805.

MAGHSOUDI NIA, E., QIAN, Q. K. & VISSCHER, H. J. 2022. Analysis of occupant behaviours in energy efficiency retrofitting projects. *Land,* 11**,** 1944.

MAHAN SINGH, M. & GEYER, P. CNN-based quick energy prediction model using image analysis for shape information. Building Simulation 2021, 2021. IBPSA, 1311-1316.

MAHJOUB, S., CHRIFI-ALAOUI, L., MARHIC, B. & DELAHOCHE, L. 2022. Predicting energy consumption using LSTM, multi-layer GRU and drop-GRU neural networks. *Sensors,* 22**,** 4062.

MEWADA, H. K., PATEL, A. V., CHAUDHARI, J., MAHANT, K. & VALA, A. 2020. Automatic room information retrieval and classification from floor plan using linear regression model. *International Journal on Document Analysis and Recognition (IJDAR),* 23**,** 253-266.

MOVEH, S., MERCHáN-CRUZ, E. A., ABUHUSSAIN, M., ALHUMAID, S., ALMAZAM, K. & DODO, Y. A. 2025. Multi-Building Energy Forecasting Through Weather-Integrated Temporal Graph Neural Networks. *Buildings,* 15**,** 808.

NUTKIEWICZ, A., CHOI, B. & JAIN, R. K. 2021. Exploring the influence of urban context on building energy retrofit performance: A hybrid simulation and data-driven approach. *Advances in Applied Energy,* 3**,** 100038.

NUTKIEWICZ, A., YANG, Z. & JAIN, R. K. 2018. Data-driven Urban Energy Simulation (DUE-S): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow. *Applied energy,* 225**,** 1176-1189.

O'MAHONY, N., CAMPBELL, S., CARVALHO, A., HARAPANAHALLI, S., HERNANDEZ, G. V., KRPALKOVA, L., RIORDAN, D. & WALSH, J. Deep learning vs. traditional computer vision. Advances in Computer Vision: Proceedings of the 2019 Computer Vision Conference (CVC), Volume 1 1, 2020. Springer, 128-144.

OLU-AJAYI, R., ALAKA, H., SULAIMON, I., SUNMOLA, F. & AJAYI, S. 2022. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. *Journal of Building Engineering,* 45**,** 103406.

OR, S.-H., WONG, K.-H., YU, Y.-K., CHANG, M. M.-Y. & KONG, H. 2005. Highly automatic approach to architectural floorplan image understanding & model generation. *Pattern Recognition***,** 25-32.

PASICHNYI, O., LEVIHN, F., SHAHROKNI, H., WALLIN, J. & KORDAS, O. 2019. Data-driven strategic planning of building energy retrofitting: The case of Stockholm. *Journal of cleaner production,* 233**,** 546-560.

PENG, W., HUANG, Y., LI, C. & WANG, Y. 2025. Exploration of Resident Satisfaction and Willingness in the Renovation of a Typical Old Neighborhood. *Buildings,* 15**,** 293.

PENG, Z., ZHAO, S., SHEN, L., MA, Y., ZHANG, Q. & DENG, W. 2021. Retrofit or rebuild? The future of old residential buildings in urban areas of China based on the analysis of environmental benefits. *International Journal of Low-Carbon Technologies,* 16**,** 1422-1434.

PIIRA, K., KANTOROVITCH, J., KANNARI, L., PIIPPO, J. & VU HOANG, N. 2022. Decision support tool to enable real-time data-driven building energy retrofitting design. *Energies,* 15**,** 5408.

PIZARRO, P. N., HITSCHFELD, N., SIPIRAN, I. & SAAVEDRA, J. M. 2022. Automatic floor plan analysis and recognition. *Automation in Construction,* 140**,** 104348.

PŁOSZAJ-MAZUREK, M., RYŃSKA, E. & GROCHULSKA-SALAK, M. 2020. Methods to optimize carbon footprint of buildings in regenerative architectural design with the use of machine learning, convolutional neural network, and parametric design. *Energies,* 13**,** 5289.

QAISAR, I., SUN, K., ZHAO, Q., XING, T. & YAN, H. 2023. Multi-sensor-based occupancy prediction in a multi-zone office building with transformer. *Buildings,* 13**,** 2002.

RUNGE, J. & ZMEUREANU, R. 2019. Forecasting energy use in buildings using artificial neural networks: A review. *Energies,* 12**,** 3254.

RUNGE, J. & ZMEUREANU, R. 2021. A review of deep learning techniques for forecasting energy use in buildings. *Energies,* 14**,** 608.

SADHUKHAN, D., PERI, S., SUGUNARAJ, N., BISWAS, A., SELVARAJ, D. F., KOINER, K., ROSENER, A., DUNLEVY, M., GOVEAS, N. & FLYNN, D. 2020. Estimating surface temperature from thermal imagery of buildings for accurate thermal transmittance (U-value): A machine learning perspective. *Journal of Building Engineering,* 32**,** 101637.

SEO, J., KIM, S., LEE, S., JEONG, H., KIM, T. & KIM, J. 2022. Data-driven approach to predicting the energy performance of residential buildings using minimal input data. *Building and Environment,* 214**,** 108911.

SHEN, P., BRAHAM, W. & YI, Y. 2019. The feasibility and importance of considering climate change impacts in building retrofit analysis. *Applied energy,* 233**,** 254-270.

SHEN, Y. & YARNOLD, M. 2021. A novel sensitivity analysis of commercial building hybrid energy-structure performance. *Journal of Building Engineering,* 43**,** 102808.

SHENG, Y., WARD, W. O., ARBABI, H., ÁLVAREZ, M. & MAYFIELD, M. Deep multimodal learning for residential building energy prediction. IOP Conference Series: Earth and Environmental Science, 2022. IOP Publishing, 012038.

SHI, Z., WANG, X., WANG, J., WANG, Y., JIAN, Y., HUANG, C. & YAO, J. A Method for Real-Time Prediction of Indoor Natural Ventilation in Residential Buildings. Proceedings of CAADRIA, 2024. 9-18.

SHU, L. & ZHAO, D. 2023. Decision-making approach to urban energy retrofit—a comprehensive review. *Buildings,* 13**,** 1425.

SINGH, M. M., SINGARAVEL, S. & GEYER, P. 2021. Machine learning for early stage building energy prediction: Increment and enrichment. *Applied Energy,* 304**,** 117787.

SINGH, M. M. & SMITH, I. F. 2023. Convolutional neural network to learn building-shape representations for early-stage energy design. *Energy and AI,* 14**,** 100293.

STRELTSOV, A., MALOF, J. M., HUANG, B. & BRADBURY, K. 2020. Estimating residential building energy consumption using overhead imagery. *Applied Energy,* 280**,** 116018.

SUN, W., SUN, Y., XU, L., CHEN, X. & ZAI, D. 2022. Research on energy consumption constitution and energy efficiency strategies of residential buildings in China based on carbon neutral demand. *Sustainability,* 14**,** 2741.

TRUONG, L. H. M., CHOW, K. H. K., LUEVISADPAIBUL, R., THIRUNAVUKKARASU, G. S., SEYEDMAHMOUDIAN, M., HORAN, B., MEKHILEF, S. & STOJCEVSKI, A. 2021. Accurate prediction of hourly energy consumption in a residential building based on the occupancy rate using machine learning approaches. *Applied Sciences,* 11**,** 2229.

TURNER, A. Depthmap: a program to perform visibility graph analysis. Proceedings of the 3rd international symposium on space syntax, 2001. Citeseer, 31-12.

UDDIN, M. N., CHI, H.-L., WEI, H.-H., LEE, M. & NI, M. 2022. Influence of interior layouts on occupant energy-saving behaviour in buildings: An integrated approach using Agent-Based Modelling, System Dynamics and Building Information Modelling. *Renewable and Sustainable Energy Reviews,* 161**,** 112382.

VAN DEN BROM, P., MEIJER, A. & VISSCHER, H. 2018. Performance gaps in energy consumption: household groups and building characteristics. *Building Research & Information,* 46**,** 54-70.

VONTZOS, G., LAITSOS, V., CHARAKOPOULOS, A., BARGIOTAS, D. & KARAKASIDIS, T. E. 2024. Estimating spatio-temporal building power consumption based on graph convolution network method. *Dynamics,* 4**,** 337-356.

WANG, D., RONG, X., SUN, S., HU, Y., ZHU, C. & LU, J. 2025. Adaptive Convolution for CNN-based Speech Enhancement Models. *arXiv preprint arXiv:2502.14224*.

WANG, J., LI, Y., GAO, R. X. & ZHANG, F. 2022. Hybrid physics-based and data-driven models for smart manufacturing: Modelling, simulation, and explainability. *Journal of Manufacturing Systems,* 63**,** 381-391.

WANG, W., GE, X. & XIONG, H. Thermal design optimization and analysis on heating load of rural buildings in northern China. E3S Web of conferences, 2019a. EDP Sciences, 03018.

WANG, X., FENG, W., CAI, W., REN, H., DING, C. & ZHOU, N. 2019b. Do residential building energy efficiency standards reduce energy consumption in China?–A data-driven method to validate the actual performance of building energy efficiency standards. *Energy Policy,* 131**,** 82-98.

WANG, X., LU, M., MAO, W., OUYANG, J., ZHOU, B. & YANG, Y. 2015. Improving benefit-cost analysis to overcome financing difficulties in promoting energy-efficient renovation of existing residential buildings in China. *Applied Energy,* 141**,** 119-130.

WANG, Y. 2016. A Review of Beijing's Urban Development in the Twentieth Century. *A Century of Change: Beijing's Urban Structure in the 20th Century***,** 11-52.

WU, J. & YING, X. 2024. Development trend of green residential buildings in China under the guidance of the low-carbon concept: A policy review and analysis. *Journal of Urban Management*.

WU, S., SUN, K., BI, Y., SHI, M. & YU, L. 2024. Carbon reduction potential of housing retrofits: Evidence from China. *Renewable and Sustainable Energy Reviews,* 200**,** 114580.

XU, Y., LI, G. & ZHOU, J. 2021. Energy-efficiency retrofitting strategies for existing residential building envelope system—a case study in China. *Journal of Building Construction and Planning Research,* 9**,** 12-25.

XUE, Y., TEMELJOTOV-SALAJ, A. & LINDKVIST, C. M. 2022. Renovating the retrofit process: People-centered business models and co-created partnerships for low-energy buildings in Norway. *Energy Research & Social Science,* 85**,** 102406.

YAN, R., XIANG, X., CAI, W. & MA, M. 2022. Decarbonizing residential buildings in the developing world: Historical cases from China. *Science of the total environment,* 847**,** 157679.

YANG, B., JIANG, T., WU, W., ZHOU, Y. & DAI, L. 2022. Automated semantics and topology representation of residential-building space using floor-plan raster maps. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing,* 15**,** 7809-7825.

YANG, Z., GAIDHANE, A. D., DRGOŇA, J., CHANDAN, V., HALAPPANAVAR, M. M., LIU, F. & CAO, Y. 2024. Physics-constrained graph modeling for building thermal dynamics. *Energy and AI,* 16**,** 100346.

YU, Z., XU, W., CHEN, X., SUN, D., ZHANG, J., LU, F. & LIU, C. Progress in energy efficiency standards of residential buildings in China's severe cold and cold zones. E3S Web of Conferences, 2019. EDP Sciences, 03032.

ZANGHERI, P., ARMANI, R., PIETROBON, M. & PAGLIANO, L. 2018. Identification of cost-optimal and NZEB refurbishment levels for representative climates and building typologies across Europe. *Energy Efficiency,* 11**,** 337-369.

ZENG, Z., LI, X., YU, Y. K. & FU, C.-W. Deep floor plan recognition using a multi-task network with room-boundary-guided attention. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019. 9096-9104.

ZHANG, J., CHEN, P., CHEN, M. & CHEN, Y. Analysis of Energy-Saving Effects of Different Building Exterior Wall Insulation Materials. International Conference on Environmental Pollution and Governance, 2023. Springer, 1433-1439.

ZHANG, W., LU, L., PENG, J. & SONG, A. 2016. Comparison of the overall energy performance of semi-transparent photovoltaic windows and common energy-efficient windows in Hong Kong. *Energy and Buildings,* 128**,** 511-518.

ZHANG, X., SKITMORE, M. & PENG, Y. 2014. Exploring the challenges to industrialized residential building in China. *Habitat International,* 41**,** 176-184.

ZHANG, Y., BAI, X. & MILLS, F. P. 2020. Characterizing energy-related occupant behavior in residential buildings: Evidence from a survey in Beijing, China. *Energy and Buildings,* 214**,** 109823.

ZHANG, Y., ZHANG, M., GAO, F. & LI, J. 2024. Experimental study on internal heat transfer among adjacent rooms for building energy efficiency with housing vacancy consideration. *Case Studies in Thermal Engineering,* 62**,** 105188.

ZHAO, M., KüNZEL, H. M. & ANTRETTER, F. 2015. Parameters influencing the energy performance of residential buildings in different Chinese climate zones. *Energy and Buildings,* 96**,** 64-75.

ZHAO, X., WANG, T.-H. & PENG, C. Automatic room type classification using machine learning for two-dimensional residential building plans. Proceedings of the 40th Conference on Education and Research in Computer Aided Architectural Design in Europe (eCAADe 2022), 2022. 593-600.

ZHENG, D., YU, L. & WANG, L. 2019. A techno-economic-risk decision-making methodology for large-scale building energy efficiency retrofit using Monte Carlo simulation. *Energy,* 189**,** 116169.

ZHENG, J., LEI, H., LU, Y. & XIONG, S. 2025. Comprehensive Performance Evaluation of Insulation Materials for Low-Carbon Renovation of Enclosure Structures in Old Communities Based on Cloud Model and Matter–Element Extension Method. *Sustainability,* 17**,** 1907.

ZHENG, X., XU, N., TRINH, L., WU, D., HUANG, T., SIVARANJANI, S., LIU, Y. & XIE, L. 2022. A multi-scale time-series dataset with benchmark for machine learning in decarbonized energy grids. *Scientific Data,* 9**,** 359.

ZHOU, N. 2009. Energy for 500 million homes: drivers and outlook for residential energy consumption in China.

ZHU, M., DIAO, R., FU, J., JIANG, X. & ZHANG, Y. Study on the influence of key parameters of exterior window structure on building energy-saving effect. IOP Conference Series: Earth and Environmental Science, 2021. IOP Publishing, 012051.

ZOU, Y., XIANG, K., ZHAN, Q. & LI, Z. 2021. A simulation-based method to predict the life cycle energy performance of residential buildings in different climate zones of China. *Building and Environment,* 193**,** 107663.