EG-ICE 2025
GLASGOW

Exploring Building Energy Performance Prediction Using Graph Neural Networks

AGATA DALACH[12], ZIJIAN WANG[12], STAVROS NOUSIAS[12], ANDRÉ BORRMANN[12]

[1] TUM Georg Nemetschek Institute, Munich, Germany
[2] Technical University of Munich, Chair of Computing in Civil and Building Engineering, Munich, Germany

ABSTRACT
Building energy performance prediction is critical for optimizing energy efficiency, and recent studies have increasingly focused on replacing computationally intensive physics-based simulations with faster, data-driven methods. However, the wide variability in building typologies poses challenges when converting building energy models into effective machine learning training samples. The graph data format offers a promising way to simplify the representation of building energy models and introduce relationships among their components. This paper explores a data-driven approach of representing energy models as graphs and adopting graph neural networks (GNN) to predict their energy performance. To achieve that, we first select the total end-use building energy intensity (EUI) as the key building performance criterion to be predicted. We construct a synthetic component-level building energy graph dataset where nodes are building elements, and edges are their spatial and hierarchical relationships. We formulate the energy prediction as a graph regression task and evaluate the performance of several GNN architectures. The results demonstrate that all evaluated GNN models show promising predictive accuracy, highlighting their potential for delivering rapid building performance predictions, especially suitable for the early design stages.

KEYWORDS:
Energy simulation, surrogate models, performance-driven design, artificial intelligence, building design.

1. INTRODUCTION

In the context of climate change and growing global energy demands, accurate building performance simulations are essential for optimizing energy efficiency in building design. A common metric for assessing building energy efficiency is, among others, the total end-use energy intensity (EUI), typically expressed in kWh/m²·year, which provides a standardized measure to compare building performance across different designs and configurations (ASHRAE, 2021; Guy *et al.*, 2023). Traditional physics-based simulation methods, while precise, often require extensive, high-quality input data and substantial computational resources. This limitation reduces their practicality, especially in the early design stages, and has led to the exploration of data-driven, Artificial Intelligence (AI) -based methods that promise faster and more flexible energy predictions even with incomplete building information.

Early data-driven energy prediction approaches, including regression models (Tsanas and Xifara, 2012), have evolved into more advanced machine learning (ML) techniques, including Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), and Gradient Boosting Machines (GBMs)

(Wei et al., 2018). Hybrid techniques, combining ANNs with Genetic Algorithms, have enabled design optimization (Maderspacher et al., 2016), while deep learning frameworks employing Long Short-Term Memory (LSTM) networks have further improved energy forecasting by capturing temporal dependencies (Zhang et al., 2020). Meta-reviews (Elwy and Hagishima, 2024) confirm AI's growing role in building energy performance prediction, with 62% of studies focusing on early-phase predictions.

Despite recent advancements, AI-driven methods still face several critical limitations. Key issues include poor generalization, discrepancies between predicted and actual performance, and inconsistencies in optimization approaches (Elwy and Hagishima, 2024). Data-driven models often struggle with interpretability and generalizability (Geyer, Singh and Chen, 2024), and their predictive accuracy can degrade significantly when models are simplified, leading to significant gaps between estimated and actual energy consumption (Singh et al., 2020). Many of those challenges are also bound to limitations of AI models, such as the need for specifically formatted data and uniformly sized input samples. Data availability and quality also remain major concerns, often requiring substantial expert input to ensure robust predictions (Chen, Singh and Geyer, 2024). Furthermore, many traditional ML approaches to building energy performance prediction rely on monolithic models, which lack the capacity to generalize effectively across diverse building topologies (Chen, Singh and Geyer, 2024).

To address these gaps, this paper proposes a novel Graph Neural Network (GNN)-based method for predicting building energy performance. GNNs excel at modeling complex relationships in unstructured data and have been already used to accelerate engineering simulations (Jiang and Chen, 2023). They have been applied in engineering prediction tasks, such as electricity demand forecasting (Campagne et al., 2024), HVAC piping layout optimization (Wang et al., 2023), and smart city energy modeling (Sunder et al., 2024). Despite their potential, GNN applications in building energy performance remain underexplored, particularly regarding capturing building geometry and properties, as well as the environmental factors.

Graph-based representations model buildings through nodes (e.g., spaces, components) and edges (e.g., spatial or functional relationships), offering insights into energy flows and topological dependencies (Wang *et al.*, 2021; Wang, Sacks and Yeung, 2022). Unlike previous component-based ML strategies that model building elements independently (Chen, Singh and Geyer, 2024; Geyer, Singh and Chen, 2024), this research leverages GNNs to model the holistic interdependencies between building components, properties, and environmental conditions through graph representations.

This study, in an experimental form, has two main objectives. First, we prepare and evaluate synthetic dataset of building energy models to develop a GNN capable of capturing the complex interdependencies of building components. Second, to predict a single performance metric, the total end-use energy intensity (EUI), aggregated at the graph level, serving as a proof-of-concept for early design energy predictions.

2. CURRENT RESEARCH

2.1 GNNs in building energy performance prediction

Recent literature highlights the growing application of GNNs in building energy prediction tasks. For instance, Halaçlı et al. (2023) explore the use of GNNs to predict inter-zone thermal interactions. Their dataset comprises 64462 thermal zones across 5866 buildings in a residential district of Ankara. After evaluating various architectures, they introduce their own GNN model, GUBEM, which outperforms all baseline models in heating load estimation, achieving the highest R^2 score of 0.9202 and the lowest RMSE value of 13.8548 kWh/m^2 on the test set.

Similarly, Hu et al., (2022) utilize a spatio-temporal graph convolutional network (ST-GCN) to model urban-scale building energy usage, explicitly accounting for inter-building shading effects. Their approach explicitly incorporates solar-based interdependencies between buildings, where graph nodes represent buildings, and edges capture the shading effects of neighboring structures. Their model achieves a high prediction accuracy (lower RMSE than other models), and their findings confirm that buildings with a greater number of interdependencies (i.e., more connections in the graph) benefit the most from their approach.

Kiavarz, Jadidi and Esmaili (2023) apply the GraphSAGE algorithm for a classification task aimed at identifying critical rooms with the highest energy consumption. Using openly available IFC (Industry Foundation Classes) building models, they construct a dataset where nodes represent individual rooms, and edges are defined by weighted adjacency based on spatial relationships. Node features include geometric properties, energy qualities, and occupancy intensity. Their model achieves a classification accuracy of 91.2% when tested on a single building, highlighting its effectiveness in identifying energy-intensive zones.

Vontzos et al. (2024) combine GCNs and LSTMs for energy forecasting in buildings, leveraging

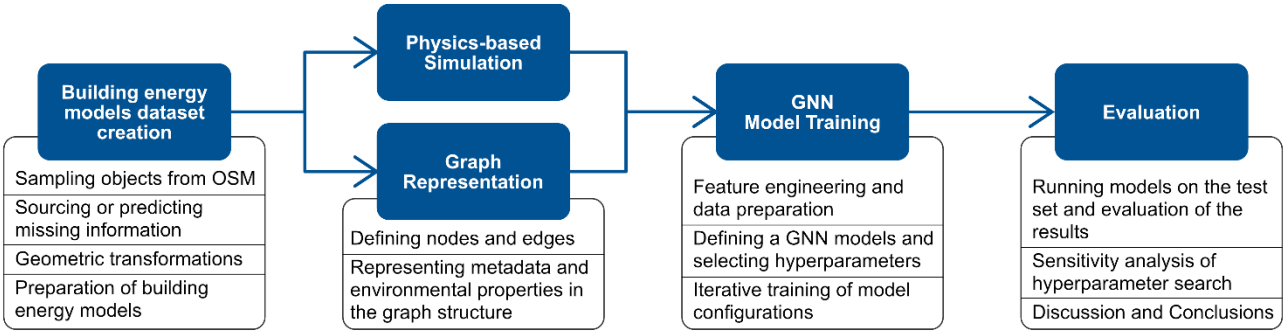

Figure 1 Overview of the experiment

spatial and temporal dependencies in a detailed single-building dataset. Like Halaçlı et al. (2023), they focuse on the interrelations between thermal zones, particularly emphasizing distance-based adjacency. Their proposed model is validated using a highly detailed dataset from a single building, comprising 33 time-series data with nearly 800,000 power consumption observations. Beyond achieving high predictive performance, their study demonstrate the significant advantage of incorporating spatial context into energy prediction.

Finally, Moveh et al. (2025) similarly use a hybrid GCN-LSTM approach to integrate static spatial relationships and the dynamic influence of environmental factors. Their dataset includes 150 commercial buildings with energy data spanning three years. In their graph representation, buildings are modeled as nodes, while edges capture geographical proximity and operational similarity. Their hybrid model utilizes detailed weather data to search for temporal dynamics, and in the end, it achieves the best performance with 3.2% MAPE.

2.2 Research gap

While existing GNN applications predominantly model buildings or thermal zones as nodes, the application of GNNs at more granular, detailed component levels has yet to be explored. Such an approach was introduced as Component-Based Machine Learning (CBML) research, which has demonstrated the benefits of modeling individual building elements, such as walls, windows, and zones, each with distinct properties (Chen, Singh and Geyer, 2024; Geyer, Singh and Chen, 2024). CBML research underscores how this form of model decomposition enhances interpretability, reusability, and generalization. However, these methods often involve significant domain knowledge and complex model assembly processes. A GNN-based method has the potential to integrate the component-based fine-grained modeling with offering a more scalable and unified framework for capturing both local interactions and global energy dynamics within a single holistic model.

2.3 Research Aim

This study aims to explore the effectiveness of GNNs in predicting building energy performance by leveraging detailed graph-based component-level building representations. The proposed approach offers a new perspective on integrating structured data with advanced neural networks and aims to achieve robust and generalizable predictions of building performance indicators. To demonstrate the feasibility of this approach, the current study focuses on predicting a single, aggregated performance metric, EUI, based on detailed, component-level graph representations of building energy models.

3. METHOD

This research introduces a more granular approach to graph-based energy simulation models for predicting total end-use building energy intensity, a building energy performance criterion.

Unlike previous GNN methods that analyze graphs representing entire buildings or zones, this study investigates a detailed graph structure in which graph nodes correspond to individual building elements (such as walls, windows, and shades) and edges represent their spatial and hierarchical relationships.

Methods utilized in this experiment involve a top-down approach to create a diverse dataset aiming towards high generalization of the proposed models, as well as adaptability across different building types, functions and geographic locations, as well as robustness to missing or faulty data.

The experiment has three core objectives:

1. Construct a diverse, globally distributed dataset of detailed building energy models, represented at the component level.
2. Create a Building Energy Graphs dataset for graph neural network computations.
3. Evaluate and compare GNN models in predicting critical building performance metrics. In this experimental setup we aim for a graph-level prediction of one value: the building total EUI.

Figure 1presents the overview of the main steps further elaborated in the Method section.

3.1 Building Energy Graphs

The data schema of building energy graphs is presented in **Error! Reference source not found.**. In the graph, nodes represent different model components. Edges capture the relationships between nodes.

Building Nodes: Each data sample (i.e., each building) is represented by the first node. We use these nodes to store building-level properties and external environmental conditions retrieved from EPW (EnergyPlus Weather) and DDY (Design Day) files.

Room Nodes: These nodes capture, among others, time-based occupancy and operational schedules, as well as system setpoints, inferred from the building's program type. To manage data complexity, full-resolution daily schedules are summarized into key statistical measures (average, peak values, peak occurrence times, and variation ranges).

Geometry Nodes: These nodes represent the components of the building, including their geometric and physical properties. The "geometry" category is subdivided into: external and internal walls, roofs, floors, foundations, and apertures (windows).

Shades Nodes: These nodes represent the surrounding context of the building, specifically the faces of neighboring structures. They are assigned the same geometric parameters as the geometry nodes (area, normal vector, centroid, aspect ratio) but do not include material properties due to their lower relevance in energy simulations.

Inheritance Edges: These directional edges represent the hierarchical links, for example, a wall belonging to a room or a window being part of a wall Those relationships are directly extracted from the energy model's hierarchical structure.

Adjacency Edges: These connect nodes that are physically adjacent, which is determined by comparing coordinate points. Those edges are uni-directional, as a wall adjacent to the floor is the same as a floor adjacent to the wall.

Shading Influence Edges: These capture interactions due to shading effects from nearby structures. They contain only geometric information.

After processing, the final graph includes only the relevant nodes and edges (see **Error! Reference source not found.**). All values are normalized, and the nodes are interconnected with edges that realistically reflect their structural and environmental relationships in a human-understandable format.

3.3 Graph Neural Network Models

All the GNNs follow the general message-passing algorithm, including neighbor aggregation, transformation of the aggregated features, updating with an activation function and propagation through stacking. The key difference is in how they are performing the mentioned tasks.

The first tested architecture is a **Graph Convolutional Network (GCN).** GCNs aggregate features from a node's neighbors, including the node itself, and perform a spectral convolution (Kipf and Welling, 2017). As the aggregation is uniform, all the node neighbors have the same importance in the final prediction.

GraphSAGE model, instead of aggregating all the neighbours, samples neighbor features using permutation-invariant functions, for example mean-pooling. Therefore, it can be applied on large graphs and can better generalize to unseen data (Hamilton, Ying and Leskovec, 2018).

Graph Attention Network (GAT) models, instead of simple aggregation of neighbor node features, they learn the attention weights allowing the model to work on the most important relationships (Veličković *et al.*, 2018). It uses an attention mechanism on each edge, calculating attention coefficients between nodes and edges.

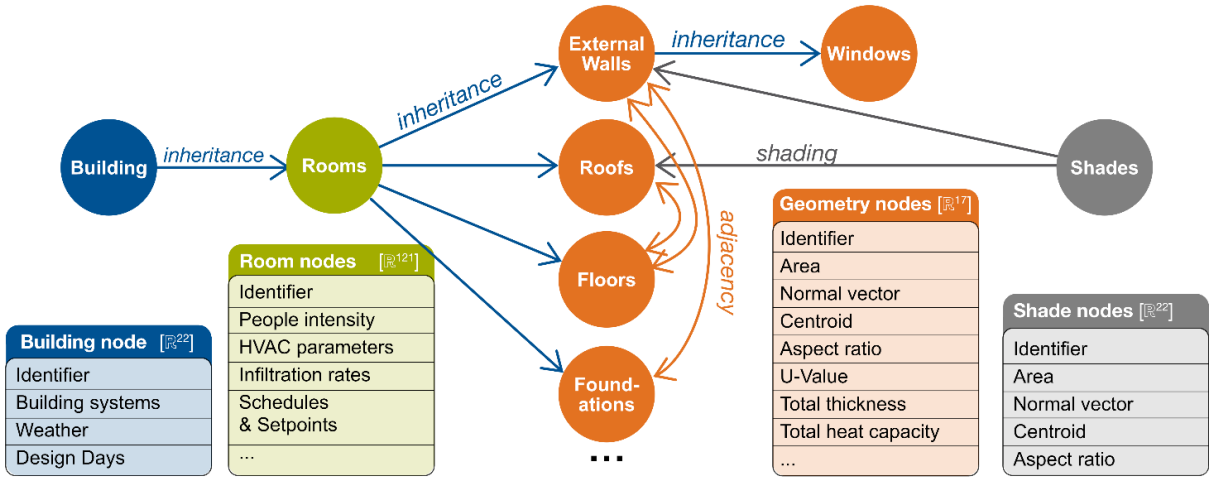

Figure 2 Building Energy Graph Structure

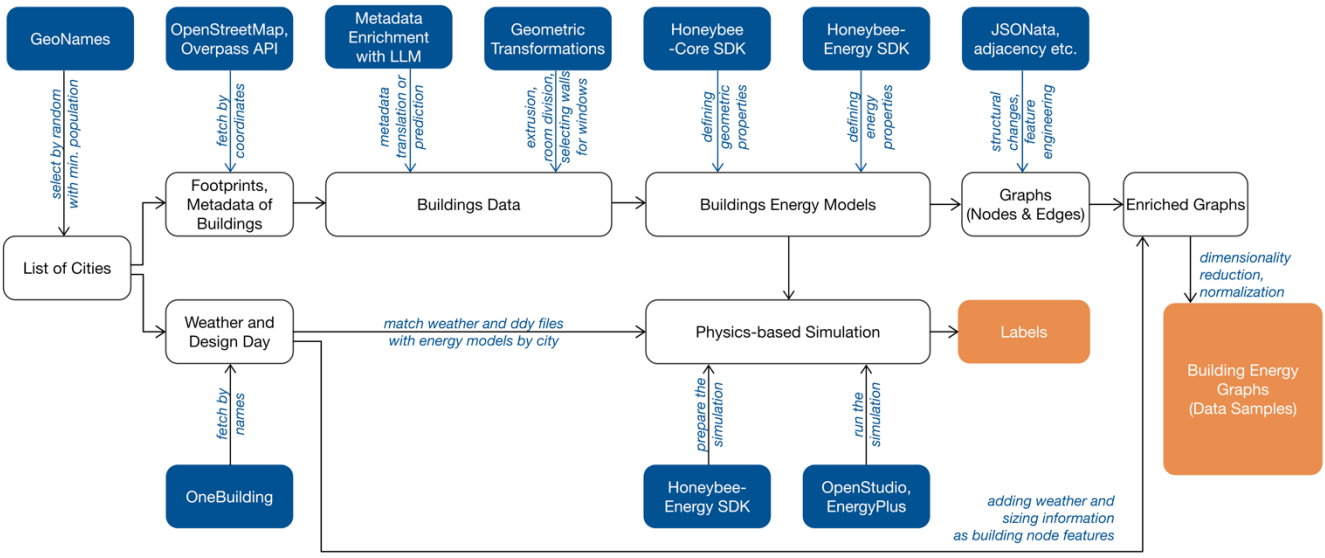

Figure 3 Dataset creation for the GNN application

4. EXPERIMENT

Figure 3 provides a concise summary of subsections 4.1 through 4.4.

4.1 Dataset Collection and Preparation

We prepare a dataset capturing a broad range of climate zones and urban typologies by sampling cities with at least 300,000 inhabitants using OpenStreetMap (OSM). In particular, we randomly select 107 city centres from Europe and the Middle East (see Figure 4). From each city, we sample 10 buildings along with their closest neighbours, accepting only samples with key metadata (e.g., building height, number of levels). This way each sample includes a "main building" and its context, thereby capturing effects of adjacency, shading, and urban density.

Figure 4 Map of the locations included in the dataset

The acquired data is structured in a hierarchical JSON format, listing buildings with their geometric attributes and available metadata. We then assign each building a climate zone. Due to the lack of publicly available ASHRAE climate zone maps for Europe, we classify the climate using Köppen-Geiger categories (Kottek et al., 2006) and map these categories to the corresponding ASHRAE Climate Zones for simulation purposes.

Approximately 65% of the samples include the building function, but only 3% provide material information and 4% indicate a construction date. To address this missing metadata, we employ a large language model (LLM), OpenAI's GPT-4o, to infer absent values based on available details such as city, country, function, climate, height, and gross area. We manually validate a selection of these predicted values using publicly available data such as Google Street Maps and photos to ensure data accuracy.

Building geometries, originally defined in geographic coordinates (latitude/longitude), are transformed into a local Euclidean coordinate system. Each building is extruded according to its provided height or floor number value. Neighboring structures are simplified to improve computational efficiency and irrelevant faces (far, occluded) are removed using ray-tracing. The main building is further subdivided into floors and convex zones. Using simple geometric rules based on distances and adjacency, we automatically select external walls that are likely to host windows.

4.2 Building Energy Models and Simulation

Each geometry model is converted into an energy model using the Honeybee framework (Sadeghipour Roudsari, Pak and Viola, 2013), based on both its geometric information and metadata. We apply physical and energetic properties using templates from the "OpenStudio Standards Gem", based on the predefined building function, construction type, building age, and ASHRAE climate zone.

Neighboring structures are included as shading elements to capture their impact on solar access and thermal interactions. The models are stored in HBJSON format, allowing us to either run EnergyPlus simulations (see Figure 5) as well as convert the models into graphs (see Figure *6*).

To ensure climate data consistency, we match each model with its closest EnergyPlus Weather (EPW) to provide hourly meteorological data for typical meteorological years (TMY) and Design Day (DDY) files to define extreme climatic conditions for HVAC system sizing.

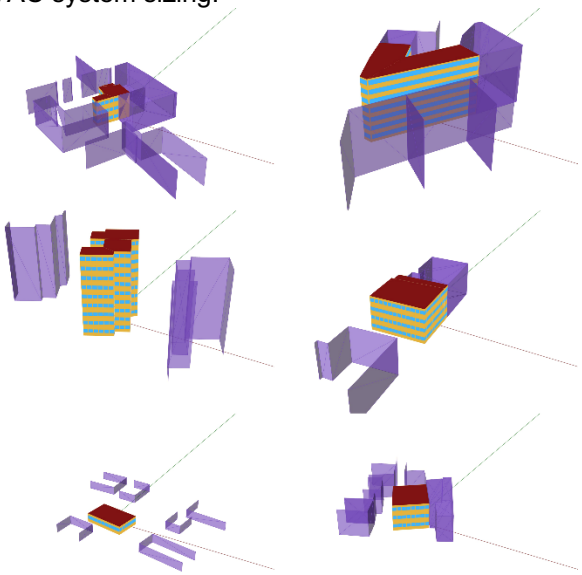

Figure 5 Example Building Energy Models

We iteratively simulate their energy performance prediction using the physics-based EnergyPlus 23.2.0 engine. On average, a simulation takes approximately 1 minute and 50 seconds, although more complex models can require up to 19 minutes and 20 seconds to process. To manage simulation times under ca. 20 minutes, we exclude samples with energy model files larger than 5 MB from the simulation queue. The raw simulation results are stored in databases ranging from 14 MB to 1.8 GB and include detailed tabular data covering simulation setups, sequential partial results, and summaries. We extract key performance indicators (such as annual energy demand and peak heating and cooling loads) to serve as ground truth for further ML applications.

4.3. Data Transformation into Graph Format

We transform each HBJSON file into a graph representation, where nodes represent building components and edges capture their relationships, as described in Section 0.

The building node contains information about the building environment. Those are obtained from the EPW and DDY files. EPW files contain hourly data (8760 hours per year) for a given location, while DDY files include structured climate parameters such as extreme dry-bulb temperatures, and humidity levels, which are critical for determining heating and cooling loads in buildings. In both cases, we use multiple strategies to reduce dimensionality and the final building node properties are obtained by concatenating the components from both datasets.

For the room nodes, time-based occupancy and operational schedules are condensed into key statistics, including average and peak values, peak occurrence times, and variation ranges. Initially, this process produced 548 different features; however, after filtering out parameters that appeared in less than 20% of nodes and those with no variation (min=max or std=0), followed by conversion and normalization, the final room schema comprises 129 parameters.

For geometry nodes, we extract key physical properties of each building element rather than storing the detailed attributes of individual material layers. Using predefined materials from template construction sets, we compute additional physical properties like U-value, total thickness, and heat capacity. The original geometry, stored as lists of vertices, is reduced to parameters such as surface area, normal vectors, centroids, and aspect ratios.

We create edges in three steps. First, we extract inheritance edges directly from the structure of the input JSON file. Second, we create adjacency edges based on vertex matching. Third, we generate shading edges using a ray-tracing approach. A selection of Building Energy Graphs is presented in *Figure 6*.

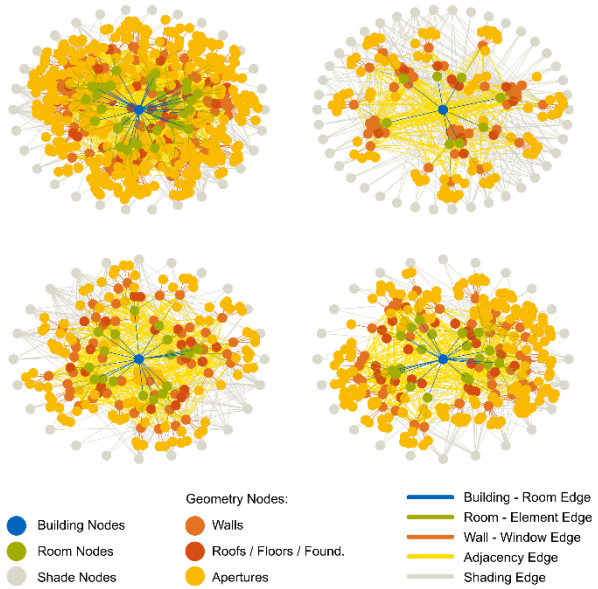

Figure 6 Example Building Energy Graphs

4.4 Labels

Labels are EUI values obtained from simulations, as described in Section 4.2. Our task is a graph-level regression where the model predicts EUI values derived from physics-based simulation results. We select this metric because it is decoupled from the building size (divided by m²), presenting a more challenging task where the model cannot merely rely on building dimensions to predict higher energy consumption. A critical initial step involves inspecting the raw distribution of these labels to identify and address the right-tail skewness.

Our initial dataset of 1,176 models shows a mean EUI of roughly 380 kWh/m²·year (median ~329 kWh/m²·year), with some outliers reaching over 16,000 kWh/m²·year. After removing 29 outliers, the maximum drops to around 867 kWh/m²·year, and the overall statistics (mean ~350 kWh/m²*year, median ~326kWh/m²*year) decrease accordingly (see Figure 7). This may be slightly above the typical European commercial building median of just under 200 kWh/m²·year, with some values reaching around 800 kWh/m²·year (Guy *et al.*, 2023).

Available real-life data sources for Europe often focus on heating demands, thus, we explore this component additionally. After outlier cleaning, the yearly heating intensity (mean ~211 kWh/m²·year, median ~190 kWh/m²·year) is above the TABULA/EPISCOPE data for residential heating, which typically ranges from 90 to 170 kWh/m²·year (Loga, Stein and Diefenbach, 2016). Overall, our simulation results indicate that the synthetically created dataset matches the range of observed real-life values, with a distinct right-skewness, even after outlier removal.

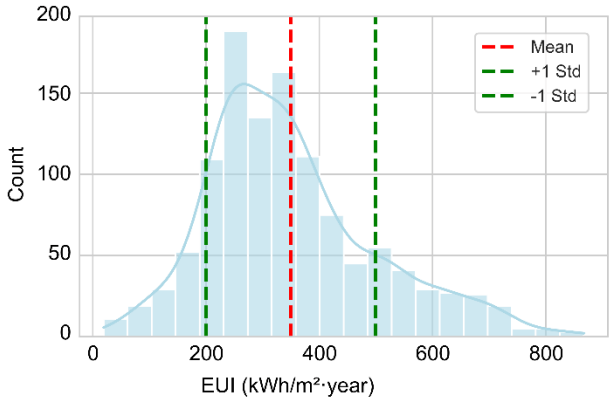

Figure 7 Label Distribution of EUI (kWh/m²·year) after outlier removal

For training efficiency, the EUI values (labels) are normalized to the range (0,1) using Min-Max scaling, resulting in normalized values with a mean of 0.42 and a standard deviation of 0.19. The normalized labels are then mapped to their corresponding model identifiers, forming the ground truth for this GNN regression task.

4.5 GNN Models Architecture and Training Setup

We evaluate three GNN architectures - GCN, GraphSAGE, and GAT - to investigate how different message-passing and aggregation strategies influence performance, as well as to explore the impact of various hyperparameters. Each model employs a two-stage aggregation process:

Message Passing Aggregation: During message passing, each model updates the node feature matrix by aggregating features from neighboring nodes. We test GraphSAGE with various aggregator functions: mean, pooling, or LSTM, while GAT utilizes attention-based weighted aggregation.

Graph-Level Aggregation: After message passing, we aggregate the node-level outputs into a single graph-level feature vector. For this final aggregation, we apply mean pooling over all nodes and then pass the resulting vector through a final linear layer to generate a single prediction.

4.6 Hyperparameters searching

To identify the best-performing models for this task, we conduct a grid search over a selection of the following hyperparameters:

- Hidden layer dimensions: 32, 64, 128
- Number of layers: 2, 3, 4
- Learning rates: 1×10^{-2}, 1×10^{-3}, 1×10^{-4}
- Dropout rates: 0.0, 0.2, 0.5
- Activation functions: ReLU, LeakyReLU
- GAT - Number of attention heads: 2, 4, 8, 16
- GraphSAGE - Aggregation functions: Mean, Pooling, LSTM

4.7 Evaluation Criteria

All models are trained using the same data splits (70% training, 10% validation, and 20% test), with batches of the size of 32 for 50 epochs, and loss function, which we have chosen as the Mean Squared Error (MSE).

Once training is complete, we evaluate the model on unseen test data by computing various regression metrics. In addition to MSE used for loss minimization during training, we calculate the root mean squared error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). The coefficient of determination (R^2) further illustrates the differences in model performance. An R^2 of 1.0 implies perfect prediction, while 0.0 indicates a model predicting the mean of the target variable.

4.8 Results

The test results on the best combinations of hyperparameters are presented in Table 1.

Table 1 Test results of GNN models

Model Type	GCN	Graph SAGE	GAT
MSE (Norm. 0-1)	0.009	0.008	0.007
RMSE [kWh/m²·year]	82.12	77.00	75.68
R^2	0.71	0.75	0.76
MAPE [%]	19.00	16.64	15.97
MAE [kWh/m²·year]	53.34	49.61	46.86

When evaluated on the original scale, the GAT model performed best, with an RMSE of 75.68 kWh/m²·year, compared to 77 kWh/m²·year GraphSAGE and 82.12 kWh/m²·year for GCN. Similarly, with an MAE of 46.86 kWh/m²·year, it outperformed GraphSAGE (49.61 kWh/m²·year) and GCN (53.34 kWh/m²·year).

This indicates that the best model overall deviates from the ground truth on average by around 47 kWh/m²·year - a substantial error in an engineering context (see Figure 8). The best GAT model achieved an R^2 of 0.76, reflecting reasonably strong performance for a regression task, though additional improvements would be necessary for deployment in engineering applications.

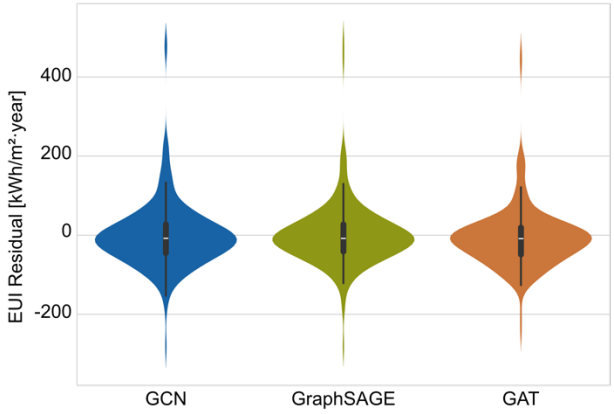

Figure 8 Distribution of residuals (EUI ground truth – prediction) across the best variants of each GNN Model

5. DISCUSSION

This study created a diverse dataset of building energy models and investigated GNNs' ability to predict building energy performance.

To achieve that, we constructed a synthetic dataset with 1,176 buildings in different countries.

We selected EUI as our prediction task. Because EUI is defined per unit area, it partially decouples total consumption from building geometry, making prediction task more challenging.

5.1 Contributions

In contrast to previous GNN-based energy prediction studies that treat each building or zone as a single node (e.g., Hu *et al.*, 2022; Halaçlı *et al.*, 2023), this work introduces a more granular approach, where each wall, window, or shading object becomes a node in the graph.

This element-level approach aligns with component-based energy modeling methods in the literature (Chen, Singh, and Geyer 2024; Geyer, Singh, and Chen 2024), but our method involves GNN architectures rather than conventional, non-geometric ML. Our top-down data creation process ensures broad coverage of climates, functions, and urban contexts, aiming at generalization across varied building configurations.

After feature engineering and dimensionality reduction, the resulting graph representations are well-suited for GNN applications. Our best model performance, when comparing the RMSE with this metric's value in other studies, is slightly worse, what is expected considering a much more diverse dataset.

5.2 Limitations

The study's experimental design meant that the dataset was built on many assumptions and simplifications. First, a substantial amount of building metadata (e.g., year built, material properties) is unavailable in OpenStreetMap (OSM), while the available data is not spread evenly. This gap leads to data imputation or broad categorization, here additionally aided by large language models, thus introducing possible biases and inaccuracies. Utilizing available templates significantly reduced the range of material definitions and building programs, while existing building stock shows higher variation in construction technologies and material-specific properties that affect performance (Kaltenegger et al., 2025).

Another constraint involves the graph representation. Throughout the iterative graph-defining process, numerous node and edge definitions were tested, but no formal, systematic evaluation method was employed to compare configurations. Especially the edge properties can potentially further improve the graph representation.

Likewise, the chosen feature-engineering and dimensionality-reduction techniques, while effective at managing high-dimensional data, might lose certain information relevant to more precise energy performance predictions. Finally, signs of overfitting emerged in some GNN models, indicating that further regularization methods or data augmentation should be explored to enhance robustness and real-world applicability.

5.3 Future Work

Immediate next steps will focus on further refining the dataset, graph representations, and GNN architectures. First, improvements in data quality will be prioritized by integrating additional data sources, such as building databases, sensor data, or normative reference buildings, and by systematically. Second, graph-structure choices (i.e., node and edge definitions) will be further investigated, and systematic methods will be applied to evaluate various feature sets and dimensionality-reduction strategies. Moreover, this should also lead to the development of more advanced models, which could possibly analyze the various types of model and node properties more intelligently (geometries, energy, time and weather properties).

Finally, the future goal of the study is to acquire component-level predictions, seeking to quantify each element's impact on total energy demand, in contrast to other GNN-based approaches. Eventually, advanced techniques such as reinforcement learning or optimization algorithms may be integrated to provide real-time design recommendations, enabling architects to rapidly assess how changes in building geometry, room configurations, or materials can affect predicted energy performance. By doing so, this research can contribute not just to predictive accuracy but also to meaningful input for sustainable building design.

6. CONCLUSION

In this study, we focused on evaluating the applicability of GNNs in graph-level building energy performance prediction using a component-level graph structure. After creating a diverse synthetic dataset, we trained many hyperparameter configurations for GCN, GraphSAGE, and GAT and obtained satisfactory results, where all GNNs performed well in terms of predictive accuracy, thereby confirming both the feasibility of this graph-based data structure and the viability of the approach itself.

Quick building energy performance predictions are a topic of active research, and our work contributes by exploring a more granular method for GNN-based modeling. Results show that such models can potentially replace computationally and time-intensive physics-based simulations, particularly during the early design stages, although an end-user interface for real-world deployment has yet to be developed. Future work will explore the topic of energy performance optimization, possibly integrating advanced techniques for adaptive or automated design decision support.

It is crucial to acknowledge that even physics-based simulations can manifest a performance gap when compared to measured data, and introducing higher levels of abstraction via machine learning methods could amplify discrepancies between the predicted and the real-life energy consumption. Consequently, incorporating real-life data is a vital step to reduce divergence and ensure that surrogate models remain both credible and robust.

Nevertheless, considering the importance of the energy demands of the building sector, searching for enhancements in predictive and optimization workflows remains an essential research objective. Although it is difficult to predict the exact final performance of any building design, we should aim to compare variants and choose the most energy-efficient one. This study contributes to this goal by introducing a new approach to this topic.

ACKNOWLEDGEMENTS

This work was supported by the funds of TUM Georg Nemetschek Institute.

REFERENCES

American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc. (ASHRAE) (2021) 'ASHRAE Handbook, Chapter 19, Energy Estimating and Modeling Methods'. American Society of Heating, Refrigerating and Air-Conditioning Engineers, Inc. (ASHRAE). Available at: https://app.knovel.com/hotlink/khtml/id:kt012MH6J1/ashrae-handbook-fundamentals/typical-applications.

Campagne, E. *et al.* (2024) 'Leveraging Graph Neural Networks to Forecast Electricity Consumption'. arXiv. Available at: https://doi.org/10.48550/arXiv.2408.17366.

Chen, X., Singh, M.M. and Geyer, P. (2024) 'Utilizing domain knowledge: Robust machine learning for building energy performance prediction with small, inconsistent datasets', *Knowledge-Based Systems*, 294, p. 111774. Available at: https://doi.org/10.1016/j.knosys.2024.111774.

Elwy, I. and Hagishima, A. (2024) 'The artificial intelligence reformation of sustainable building design approach: A systematic review on building design optimization methods using surrogate models', *Energy and Buildings*, 323, p. 114769. Available at: https://doi.org/10.1016/j.enbuild.2024.114769.

Geyer, P., Singh, M.M. and Chen, X. (2024) 'Explainable AI for engineering design: A unified approach of systems engineering and component-based deep learning demonstrated by energy-efficient building design', *Advanced Engineering Informatics*, 62, p. 102843. Available at: https://doi.org/10.1016/j.aei.2024.102843.

Guy, H. *et al.* (2023) 'Benchmarking the energy performance of European commercial buildings with a bayesian modeling framework', *Energy and Buildings*, 299, p. 113595. Available at: https://doi.org/10.1016/j.enbuild.2023.113595.

Halaçlı, E.G. *et al.* (2023) 'A Novel Graph Neural Network for Zone-Level Urban-Scale Building Energy Use Estimation', in *Proceedings of the 10th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation*. *BuildSys '23: The 10th ACM*

International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, Istanbul Turkey: ACM, pp. 169–176. Available at: https://doi.org/10.1145/3600100.3623747.
Hamilton, W.L., Ying, R. and Leskovec, J. (2018) 'Inductive Representation Learning on Large Graphs'. arXiv. Available at: https://doi.org/10.48550/arXiv.1706.02216.
Hu, Y. *et al.* (2022) 'Times series forecasting for urban building energy consumption based on graph convolutional network', *Applied Energy*, 307, p. 118231. Available at: https://doi.org/10.1016/j.apenergy.2021.118231.
Jiang, C. and Chen, N.-Z. (2023) 'Graph Neural Networks (GNNs) based accelerated numerical simulation', *Engineering Applications of Artificial Intelligence*, 123, p. 106370. Available at: https://doi.org/10.1016/j.engappai.2023.106370.
Kaltenegger, J., Frandsen, K.M. and Petrova, E. (2025) 'An ontology-driven framework for digital transformation and performance assessment of building materials', *Building and Environment*, 271, p. 112565. Available at: https://doi.org/10.1016/j.buildenv.2025.112565.
Kiavarz, H., Jadidi, M. and Esmaili, P. (2023) 'A graph-based explanatory model for room-based energy efficiency analysis based on BIM data', *Frontiers in Built Environment*, 9, p. 1256921. Available at: https://doi.org/10.3389/fbuil.2023.1256921.
Kipf, T.N. and Welling, M. (2017) 'Semi-Supervised Classification with Graph Convolutional Networks'. arXiv. Available at: https://doi.org/10.48550/arXiv.1609.02907.
Kottek, M. *et al.* (2006) 'World Map of the Köppen-Geiger climate classification updated', *Meteorologische Zeitschrift*, 15(3), pp. 259–263. Available at: https://doi.org/10.1127/0941-2948/2006/0130.
Loga, T., Stein, B. and Diefenbach, N. (2016) 'TABULA building typologies in 20 European countries—Making energy-related features of residential building stocks comparable', *Energy and Buildings*, 132, pp. 4–12. Available at: https://doi.org/10.1016/j.enbuild.2016.06.094.
Maderspacher, J. *et al.* (2016) 'Energy optimization of an existing building based on a neural network and a genetic algorithm'.
Moveh, S. *et al.* (2025) 'Multi-Building Energy Forecasting Through Weather-Integrated Temporal Graph Neural Networks', *Buildings*, 15(5), p. 808. Available at: https://doi.org/10.3390/buildings15050808.
Sadeghipour Roudsari, M., Pak, M. and Viola, A. (2013) 'Ladybug: A Parametric Environmental Plugin For Grasshopper To Help Designers Create An Environmentally-conscious Design', in. *2017 Building Simulation Conference*. Available at: https://doi.org/10.26868/25222708.2013.2499.
Singh, M.M. *et al.* (2020) 'Quick energy prediction and comparison of options at the early design stage', *Advanced Engineering Informatics*, 46, p. 101185. Available at: https://doi.org/10.1016/j.aei.2020.101185.
Sunder, R. *et al.* (2024) 'An advanced hybrid deep learning model for accurate energy load prediction in smart building', *Energy Exploration & Exploitation*, 42(6), pp. 2241–2269. Available at: https://doi.org/10.1177/01445987241267822.
Tsanas, A. and Xifara, A. (2012) 'Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools', *Energy and Buildings*, 49, pp. 560–567. Available at: https://doi.org/10.1016/j.enbuild.2012.03.003.
Veličković, P. *et al.* (2018) 'Graph Attention Networks'. arXiv. Available at: https://doi.org/10.48550/arXiv.1710.10903.
Vontzos, G. *et al.* (2024) 'Estimating Spatio-Temporal Building Power Consumption Based on Graph Convolution Network Method', *Dynamics*, 4(2), pp. 337–356. Available at: https://doi.org/10.3390/dynamics4020020.
Wang, H. *et al.* (2023) 'GNNs based method for evaluating HVAC terminal piping layout solutions in office buildings', in. *2023 Building Simulation Conference*. Available at: https://doi.org/10.26868/25222708.2023.1275.
Wang, Z. *et al.* (2021) 'Room Type Classification for Semantic Enrichment of Building Information Modeling Using Graph Neural Networks'.
Wang, Z., Sacks, R. and Yeung, T. (2022) 'Exploring graph neural networks for semantic enrichment: Room type classification', *Automation in Construction*, 134, p. 104039. Available at: https://doi.org/10.1016/j.autcon.2021.104039.
Wei, Y. *et al.* (2018) 'A review of data-driven approaches for prediction and classification of building energy consumption', *Renewable and Sustainable Energy Reviews*, 82, pp. 1027–1047. Available at: https://doi.org/10.1016/j.rser.2017.09.108.
Zhang, C. *et al.* (2020) 'A hybrid deep learning-based method for short-term building energy load prediction combined with an interpretation process', *Energy and Buildings*, 225, p. 110301. Available at: https://doi.org/10.1016/j.enbuild.2020.110301.