Deep Learning-Based Commercial Building Energy Consumption Prediction Model Considering Occupant Density Using Mobile Network Big Data: A Case Study of Seoul

Havva Elmalı, Seung-Nam Kim
Chung-Ang University Graduate School, Seoul, South Korea
havva.elmali@outlook.com, snkim@cau.ac.kr

Abstract
This study proposes a multidimensional modelling framework to estimate annual energy consumption in commercial buildings across Seoul by integrating physical, environmental, economic, and behavioural variables. While previous research has predominantly focused on structural and environmental factors, this study addresses a critical gap by incorporating high-resolution spatial data and temporally disaggregated floating population variables derived from mobile network data, segmented into three time periods: late night, daytime, and evening. This enables a more refined representation of occupant dynamics in dense urban settings. Separate regression models for annual electricity and gas/heating consumption per unit floor area provide a contextual and scalable analytical approach. In the next phase of the research, a non-linear modelling technique based on a Multi-Layer Perceptron (MLP) will be implemented, along with SHAP (SHapley Additive exPlanations) analysis to enhance model interpretability. These advanced methods aim to capture complex interdependencies among variables and improve predictive performance. Ultimately, the proposed framework offers a robust foundation for data-driven energy policies and smart urban planning strategies in high-density metropolitan areas.

Keywords: Energy Consumption; Occupant Density; Commercial Buildings; Deep Learning; Prediction Model

1. Introduction
The building sector accounts for approximately one-third of global final energy consumption, making it a vital domain for achieving energy efficiency and climate resilience (IEA, 2022). Among various building types, commercial buildings located in dense urban areas represent key targets for sustainability interventions due to their high energy intensity and operational complexity. In highly urbanised contexts such as Seoul, energy consumption at the building level is shaped not only by structural and environmental characteristics but also by spatiotemporal variations in occupant density, especially in areas with high human mobility.

Traditional models of building energy consumption have typically emphasised physical variables such as gross floor area (GFA), building age, and structural type. While such models provide a basic analytical framework, they often fail to account for the behavioural dimension of energy use, particularly in cities where occupancy patterns fluctuate throughout the day. A comprehensive understanding of urban energy demand therefore requires a multidimensional modelling approach that incorporates physical, environmental, economic, and behavioural factors.

Recent advancements in spatial data infrastructure and mobile sensing technologies have made it feasible to integrate time-segmented floating population data into energy consumption models. Unlike static population indicators, these data capture dynamic occupant presence across three key time zones—late night (00:00–06:00), daytime (06:00–18:00), and night-time (18:00–24:00)—enabling more nuanced insights into the behavioural drivers of energy demand in commercial buildings.

In response to these developments, this study proposes a multidimensional analytical framework to estimate the annual electricity and gas/heating consumption of commercial buildings in Seoul. The model integrates four categories of explanatory variables: physical (e.g., GFA, building age, structural type), environmental (e.g., green space ratio, average temperature), economical (e.g., land price), and behavioural (e.g., floating population by time zone). The analysis begins with a set of multiple linear regression models to assess the explanatory power of each variable group, using energy consumption per unit of gross floor area as the dependent variable. In subsequent phases of the research, deep learning methods—Multi-Layer Perceptron (MLP) models—along with SHAP (SHapley Additive exPlanations) values will be applied to enhance prediction accuracy and interpretability.

Through the integration of high-resolution spatial datasets and advanced modelling techniques, this study aims to contribute both theoretically and practically to the understanding of urban energy demand.

The findings are expected to inform data-driven strategies in building energy management, policy development, and sustainable urban planning.

2. Literature Review

Building energy consumption has been widely studied across four major dimensions: physical attributes, socio-economic context, behavioural dynamics, and environmental factors.

Physical characteristics such as building age, structural system, and gross floor area (GFA) are frequently cited as key determinants of energy use. Empirical research indicates that newer buildings typically consume less energy due to improved insulation, efficient systems, and stricter building codes (An et al., 2014). Structural systems also significantly influence energy performance; for instance, reinforced concrete buildings tend to be more energy-efficient than masonry structures due to higher thermal mass and better envelope integrity (Kim, 2013). While GFA is often used as a proxy for operational scale, especially in commercial facilities, its explanatory power improves when paired with use type and occupancy characteristics that influence internal loads and activity schedules (Zhang et al., 2020).

Socioeconomic indicators help capture spatial and behavioural heterogeneity in urban energy consumption. Factors such as income level, land value, and population density have been used as proxies for energy service demand and urban activity intensity. In the Korean context, where income data is often limited at the building level, official land price has emerged as a practical proxy for economic intensity and urban development pressure (Kim et al., 2017; Shi et al., 2021). These variables provide important context for interpreting energy patterns within the built environment.

A growing body of research emphasises the importance of human activity and occupant dynamics. Traditional models often rely on static or assumed schedules, failing to capture the actual spatial and temporal variability in occupancy. To address this gap, recent studies have introduced floating population data derived from mobile communication networks as a spatially disaggregated and temporally sensitive measure of real-time human presence (Kim and Park, 2021; Chen et al., 2022). Such data is especially useful in commercial zones, where occupancy patterns fluctuate markedly across different times of day. Segmenting this data into discrete time periods—such as late night, daytime, and night-time—offers a more accurate behavioural input for urban energy models.

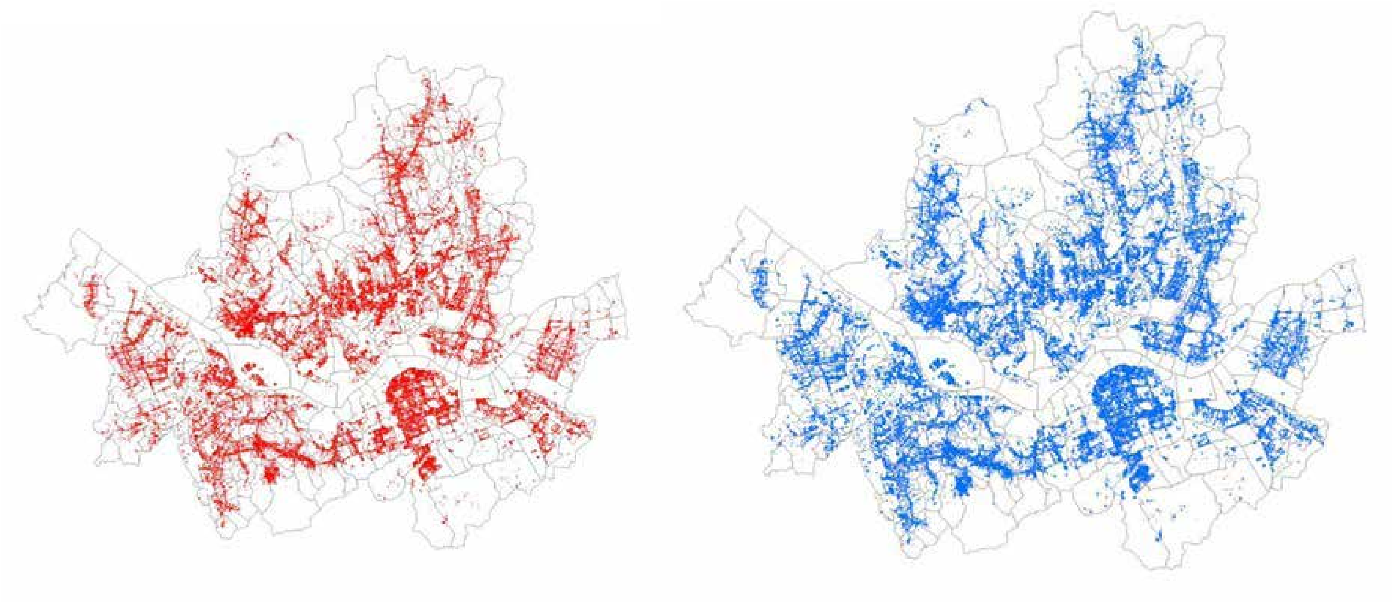
Environmental conditions also remain central to understanding energy demand. Temperature extremes significantly affect electricity and gas/heating consumption by increasing cooling and heating loads (Li et al., 2018; Jo et al., 2022). Furthermore, environmental features such as vegetation, green spaces, and water bodies can mitigate urban heat through shading and evapotranspiration, contributing to lower cooling energy needs, especially during summer months (Ewing and Rong, 2008; Ko, 2013).

Despite these advances, many previous studies rely on aggregated spatial units—such as neighbourhoods or districts—which may mask variability at the building level. Moreover, the lack of high-resolution behavioural and environmental data has limited the explanatory power of many models. This study addresses these limitations by utilising a building-level dataset that integrates physical, environmental, economic, and behavioural indicators for commercial buildings in Seoul. Specifically, it incorporates mobile-based floating population data segmented by time of day to better reflect occupancy-driven variation in energy consumption. By doing so, the study aims to bridge the gap between temporal human activity and granular urban energy modelling, providing a more realistic and actionable framework for analysis.

3. Research Subject and Data

This study focuses on Seoul, the capital of South Korea, as the spatial and analytical unit for investigating commercial building energy consumption. As a highly urbanised and technologically advanced metropolis, Seoul is characterised by dense population clusters, a compact and heterogeneous building stock, and the availability of rich, high-resolution spatial datasets. These features make the city a particularly suitable case for building-level energy consumption analysis. The scope of the study is limited to commercial buildings, specifically those categorised as first- and second-class neighbourhood living facilities and office buildings. These building types were selected due to their wide spatial distribution, extended operating hours, and relatively high user intensity throughout the urban landscape. The spatial distribution of valid energy data for electricity (a) and gas/heating (b) consumption is presented in Figure 1a and Figure 1b, respectively.

The study utilises two dependent variables to represent energy intensity: electricity consumption intensity and gas/heating consumption intensity, both based on data from the year 2023. These are defined as the natural logarithm of the total annual energy consumption (electricity or gas/heating) divided by the natural logarithm of the gross floor area (GFA), allowing for size-normalised comparisons across buildings. After data cleaning and preprocessing, the final sample consists of 88,293 buildings for electricity consumption and 67,918 for gas/heating consumption.



1a. 1b.

Figure 1a and Figure 1b: Spatial distribution of electricity (a) and gas/heating (b) consumption

To comprehensively reflect the determinants of urban energy demand, the independent variables are categorised into four analytical dimensions: physical, environmental, economic, and behavioural. Physical variables include the gross floor area, building age, structural type (classified as reinforced concrete, brick, or other), and usage function (office, first-class neighbourhood facility, or second-class neighbourhood facility). Environmental characteristics are measured through average annual temperature and the green area ratio within a 50-metre buffer around each building, derived from NDVI satellite imagery. Economic factors are represented by officially registered land price values, which act as proxies for development intensity and land-based economic value. Behavioural characteristics are captured through mobile network-based floating population data, segmented by three time periods—late night (00:00–06:00), daytime (06:00–18:00), and night-time (18:00–24:00)—as well as an overall daily average.

All variables used in this study were obtained from official government sources or authorised public data platforms. Annual energy consumption data were provided by the Ministry of Land, Infrastructure and Transport (MOLIT), while building characteristics were derived from the national spatial data platform V-World. Environmental indicators were sourced from the Seoul Metropolitan Government's S-DoT sensor network and Sentinel-2 satellite imagery. Official land price data were retrieved from the national land valuation registry, and floating population data were obtained from mobile network big data repositories. A summary of all variables, including definitions, measurement units, types, and data sources, is presented in Table 1. Because the study models electricity and gas/heating consumption separately, all statistical outputs and visualisations are reported accordingly to reflect the unique patterns of each energy type.

Table 1: Summary of variables and descriptions

Variable Name	Variable Type	Description	Unit	Data Source
elect_gfa	Dependent	Log-transformed electricity consumption per gross floor area (2023)	log(kWh/m²)	MOLIT (2023)/ V-World
gasheat_gfa	Dependent	Log-transformed gas and heat consumption per gross floor area (2023)	log(kWh/m²)	MOLIT (2023)/ V-World
GFA_log	Independent	Gross floor area	log(m²)	V-World
bldg_age	Independent	Age of the building	0	V-World
str_reinfo	Independent (Dummy)	Structural type: Reinforced concrete	0/1	V-World
str_brick	Independent (Dummy)	Structural type: Brick	0/1	V-World
str_others	Independent (Dummy)	Structural type: Others	0/1	V-World
bldg_use_1neigh	Independent (Dummy)	Use type: Type 1 Neighborhood Facility	0/1	V-World
bldg_use_2neigh	Independent (Dummy)	Use type: Type 2 Neighborhood Facility	0/1	V-World
bldg_use_office	Independent (Dummy)	Use type: Office Building	0/1	V-World
green_ratio	Independent	Proportion of green space within a 50-meter buffer around the building (NDVI)	%	Copernicus Data Space
temp_avg	Independent	Average temperature within a 50-meter buffer around the building (IDW)	°C	S-DoT Sensor Data
gongsi_jiga	Independent	Official land price of the parcel where the building is located	KRW/m²	MOLIT
avg_lifepop_tot	Independent	Annual average floating population (00:00–24:00)	Persons	Mobile Network Data
avg_lifepop_late	Independent	Annual average floating population (00:00–06:00)	Persons	Mobile Network Data
avg_lifepop_day	Independent	Annual average floating population (06:00–18:00)	Persons	Mobile Network Data
avg_lifepop_night	Independent	Annual average floating population (18:00–24:00)	Persons	Mobile Network Data

4.Methodology

The methodological approach of this study consists of three primary phases: data preprocessing and integration, statistical modelling using multiple linear regression, and the application of a deep learning-based prediction model in the form of a Multi-Layer Perceptron (MLP).

In the initial phase, datasets covering energy consumption, building characteristics, environmental variables, land value, and behavioural data were integrated spatially using unique building identifiers or geocoded addresses. This integration ensured both spatial alignment and content consistency across all records. Missing values were addressed using listwise deletion, and extreme outliers were removed by applying the interquartile range (IQR) method, with thresholds set at the 1st and 99th percentiles. These decisions were visually validated using boxplots.

Categorical variables, including structural type and building use, were transformed into dummy variables to enable their inclusion in the regression model. Additionally, natural logarithmic transformations were applied to electricity consumption, gas/heating consumption, and gross floor area to reduce skewness and improve interpretability. Other continuous variables, such as land price, temperature, and green area ratio, were retained in their original form. In all regression models, dummy variables were constructed for categorical features. Specifically, office buildings were set as the reference group for building use type, and others for structural type, enabling interpretation of coefficients relative to these baseline categories.

Environmental indicators were assigned at the building level using ArcGIS Pro. Average annual temperature was interpolated from point-based sensor data using the Inverse Distance Weighted (IDW) method. The green area ratio was calculated based on NDVI values within a 50-metre buffer surrounding each building. Floating population data, originally provided in a 50-by-50-metre grid format, were overlaid with building footprints. For each building, average values were computed from the intersecting grid cells corresponding to late night, daytime, night-time, and daily periods.

All preprocessing, spatial operations, and feature engineering tasks were conducted using Python, with relevant libraries including Pandas, GeoPandas, and Scikit-learn. Following data preparation, multiple linear regression models were estimated separately for electricity and gas/heating consumption intensities.

All explanatory variables from the four thematic categories were entered simultaneously to evaluate their joint and relative contributions to energy use.

In the subsequent phase of the research, a deep learning approach is planned to be implemented using a Multi-Layer Perceptron (MLP) model. The dataset will be randomly divided into training (70%), validation (15%), and test (15%) subsets. All continuous variables will be standardised, and the MLP architecture will include multiple hidden layers with ReLU activation functions. The model will be optimised using the Adam algorithm, with mean squared error (MSE) as the loss function. To mitigate the risk of overfitting, regularisation techniques such as dropout and early stopping will be employed. Model performance will be assessed using standard evaluation metrics, including R^2, root mean square error (RMSE), and mean absolute error (MAE), based on the test set.

To enhance transparency and interpretability of the deep learning models, SHAP (SHapley Additive exPlanations) analysis will be conducted to quantify the marginal impact of each explanatory variable on the prediction outcomes. All model results, performance evaluations, and SHAP visualisations will be presented separately for electricity and gas/heating, offering differentiated insights into the determinants of each energy type. This methodology enables a high-resolution, multidimensional analysis of urban energy consumption that incorporates spatial, temporal, and behavioural complexity.

5.Analysis Results

This study investigated the annual electricity and gas/heating consumption per unit of gross floor area in commercial buildings in Seoul through a series of multiple linear regression models, constructed separately for each energy type. The analysis utilised large-scale 2023 data, comprising 88,293 building-level observations for electricity and 67,918 for gas/heating. A five-stage modelling framework was adopted to evaluate the contribution of physical, environmental, economic, and behavioural variables, with a particular focus on the impact of time-segmented floating population data. The results reveal distinct determinants for each type of energy consumption and underscore the explanatory value of incorporating spatiotemporal population data.

In the case of electricity consumption, the baseline model includes fixed variables representing physical attributes (log-transformed gross floor area, building age, building use, and structural type), environmental factors (green area ratio and average temperature), and economic conditions (official land price). This initial model demonstrates relatively high explanatory power, with an adjusted R^2 of 0.584. Interestingly, the log of gross floor area shows a statistically significant negative coefficient, suggesting that larger buildings tend to consume less electricity per square meter, potentially reflecting economies of scale. Building age has a positive and significant effect, indicating that older structures are associated with higher energy use, possibly due to outdated equipment or insulation.

In terms of building use, neighbourhood living facilities consume less electricity compared to office buildings, while reinforced concrete structures display lower consumption than other structural types. Environmental variables exhibit notable effects: the green area ratio is positively associated with electricity use, possibly reflecting increased human activity in greener areas. Average annual temperature also contributes positively, likely driven by cooling demands during warmer periods. The official land price emerges as a statistically significant and positive factor, implying that buildings located in higher-value zones tend to consume more electricity, which may be indicative of more intensive building operations. (see Table 2 for regression results of Electricity Consumption per GFA).

Table 2. Regression Result of Model 1 of Electricity Consumption per GFA

Category	Variable Name	Coef.	Std.Err.	t-value	p-value	VIF
Physical	Building Age	0.000644	2.68E-05	24.07526	1.2E-127***	1.157005
	Gross Floor Area (log)	-0.16122	0.000679	-237.501	0***	1.907705
	Structural Type: Reinforced Concrete	-0.1244	0.002012	-61.8237	0***	2.842248
	Structural Type: Brick Masonry	-0.08187	0.002222	-36.8445	6.3E-295***	2.491508
	Use Type: Class 1 Neighborhood Facility	-0.04137	0.002282	-18.1293	2.54E-73***	5.116019
	Use Type: Class 2 Neighborhood Facility	-0.04245	0.0022	-19.2938	8.97E-83***	4.848605
Environmental	Green Area Ratio	0.000436	3.45E-05	12.61746	1.82E-36***	1.021823
	Avg. Temperature	-0.00083	0.000623	-1.32818	0.184122	1.013653
Economic	Official Land Price	4.5E-09	7.25E-11	62.02604	0***	1.148543

N=88293, Durbin-Watson: 1.6837, Adj. R-square: 0.584
* p < 0.05, ** p < 0.01, *** < 0.001

Dependent Variable: Annual electricity consumption per unit of gross floor area
Independent Variables: Physical, Environmental, and Economic Characteristics of Buildings

As floating population variables are introduced progressively in subsequent models, the explanatory power improves modestly. The final model, which includes late-night population density (00:00–06:00), achieves the highest adjusted R^2 value of 0.588 and the lowest Akaike Information Criterion (AIC), suggesting an improved model fit. This outcome highlights the role of 24-hour commercial activities or night-time operations—such as lighting, security, and maintenance services—in shaping electricity demand. All models yield Durbin–Watson statistics close to 1.68, indicating no significant autocorrelation. F-statistics and p-values confirm the overall statistical robustness of the models on Table 3.

Table3: Regression Results for Models with Floating Population on Electricity Consumption per GFA

Model	Variable	Coef.	t-value	p-value	Adj. R2	DW
Model 2	Average Total Population	0.000110404	19.80503643	4.17E-87***	0.5859	1.6835
Model 3	Late Night Population (00:00-06:00)	-0.000296387	-20.12045846	7.78E-90***	0.5878	1.6847
Model 4	Daytime Population (06:00-18:00)	0.000107928	4.846908526	1.26E-06***	0.586	1.6841
Model 5	Evening Population (18:00-24:00)	0.000501439	21.58245971	4.86E-103***	0.5881	1.6797

Durbin-Watson: 1.6837, * p < 0.05, ** p < 0.01, *** < 0.001

For gas and heating consumption, the baseline model including fixed variables explains 53.1% of the variance. Similar to the electricity model, the log of gross floor area is negatively and significantly associated with consumption intensity, reinforcing the view that larger buildings operate more efficiently. Building age once again exerts a positive influence, consistent with ageing infrastructure leading to increased energy usage.

In contrast to electricity, the effects of building use and structural type appear less pronounced in gas/heating models, suggesting that heating demand is more strongly governed by physical and climatic factors. Average annual temperature shows a strong and statistically significant negative effect, as expected in colder environments. However, the green area ratio is found to be statistically insignificant, possibly due to its limited relevance to internal heating systems. As with electricity consumption, the land price variable remains a significant positive predictor. (see Table 4 for regression results of Gas/Heat Consumption per GFA).

Table 4. Regression Result of Model 1 of Gas/Heat Consumption per GFA

Category	Variable Name	Coef.	Std.Err.	t-value	p-value	VIF
Physical	Building Age	0.000218	4.08E-05	5.341784	9.23E-08***	1.186283
	Gross Floor Area (log)	-0.19464	0.000965	-201.675	0***	1.924024
	Structural Type: Reinforced Concrete	-0.1125	0.003159	-35.6173	2.7E-275***	3.470785
	Structural Type: Brick Masonry	-0.07384	0.003456	-21.3633	6.3E-101***	3.23347
	Use Type: Class 1 Neighborhood Facility	-0.10211	0.003137	-32.5453	1.5E-230***	5.255343
	Use Type: Class 2 Neighborhood Facility	-0.09747	0.00301	-32.3832	2.6E-228***	4.945725
Environmental	Green Area Ratio	0.000565	4.83E-05	11.7026	1.33E-31***	1.020019
	Avg. Temperature	0.001474	0.000844	1.746352	0.080754	1.010858
Economic	Official Land Price	2.16E-09	1.19E-10	18.24758	3.26E-74***	1.209314

N=67918, Durbin-Watson: 1.8720, Adj. R-square: 0.531
* p < 0.05, ** p < 0.01, *** < 0.001

Dependent Variable: Annual gas/heat consumption per unit of gross floor area
Independent Variables: Physical, Environmental, and Economic Characteristics of Buildings
Incorporating floating population variables leads to a modest increase in explanatory power, reaching an adjusted R^2 of 0.533. Notably, the model including late-night population data performs best in terms of model fit, exhibiting the lowest AIC value. This supports the interpretation that night-time human presence is linked to greater heating requirements, particularly in buildings with 24-hour occupancy or operational activities. Durbin–Watson values hover around 1.87, indicating no evidence of autocorrelation in the residuals. (Table 5 and Figure 3)

Table 5: Regression Results for Models with Floating Population on Gas/Heat Consumption per GFA

Model	Variable	Coef.	t-value	p-value	Adj. R2	DW
Model 2	Average Total Population	0.000129	16.9001	6.07E-64***	0.533	1.8736
Model 3	Late Night Population (00:00~06:00)	-0.00016	-7.5759	3.61E-14***	0.5334	1.8737
Model 4	Daytime Population (06:00~18:00)	0.000121	3.7243	0.000196***	0.5331	1.8739
Model 5	Evening Population (18:00~24:00)	0.000155	4.619	3.86E-06***	0.5332	1.8729

Durbin-Watson: 1.6837, * p < 0.05, ** p < 0.01, *** < 0.001

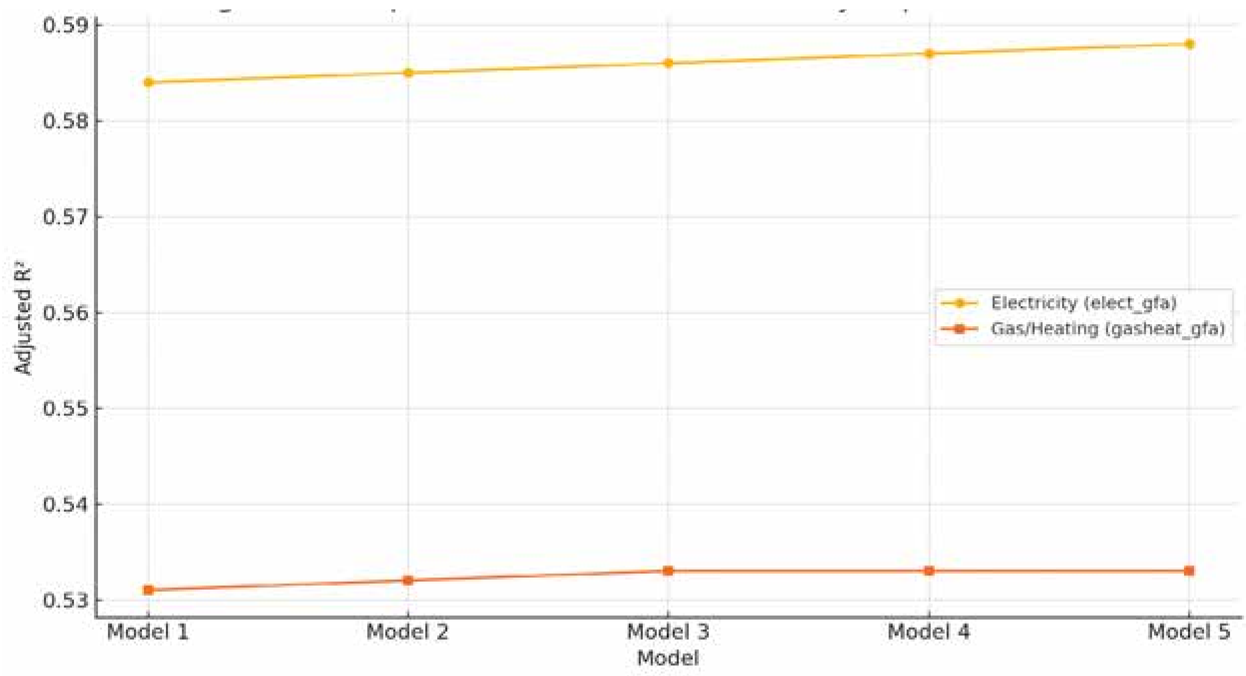


Figure 3: Comparison of Models Performance by Dependent Variables

6.Conclusion

This study proposes a multidimensional modelling framework to examine energy consumption in commercial buildings in Seoul by integrating physical, environmental, economic, and behavioural factors within a unified analytical structure. Separate regression models for electricity and gas/heating consumption reveal both shared and distinct drivers across energy types. Among the explanatory variables, temperature emerges as a particularly strong determinant of heating energy use, while official land price reflects patterns of energy demand in economically dense urban areas. Although floating population variables contributed modestly to the overall explanatory power of the models, they nonetheless provided valuable insight into the behavioural dimension of urban energy use.

The proposed framework enables a more comprehensive interpretation of energy consumption patterns by incorporating temporally segmented mobility data alongside building attributes and environmental indicators. Such integration allows for more context-sensitive predictions, especially in highly dynamic urban environments. In the subsequent phase of this research, a non-linear modelling approach using a Multi-Layer Perceptron (MLP) will be implemented to capture complex interdependencies that are beyond the scope of traditional linear models. The application of SHAP (SHapley Additive exPlanations) will further enhance model interpretability by quantifying the contribution of each input variable. Together, these methodological advancements aim to expand both the predictive power and the policy relevance of building-level energy models.

Overall, the findings underscore the importance of adopting a holistic perspective in urban energy modelling—one that accounts not only for physical infrastructure but also for the social and environmental dynamics that influence energy demand. The study contributes a scalable, data-driven approach that can inform future energy efficiency strategies, smart city initiatives, and evidence-based urban sustainability policies.

These findings carry several policy implications for energy-efficient urban development. First, retrofit programmes should prioritise ageing commercial buildings, particularly in areas with high official land values, where energy consumption is significantly elevated. Second, urban energy models should incorporate behavioural data, such as time-segmented floating population, to better reflect actual occupancy-driven energy demand. Third, planning authorities may consider expanding green infrastructure and optimising thermal environments to indirectly influence energy use patterns. Lastly, the integration of spatially granular data and advanced predictive models can support smarter zoning decisions, resource allocation, and the design of targeted energy policies in high-density urban contexts such as Seoul.

7.Discussion and Limitations

This study provides important insights by analysing the energy consumption of commercial buildings in Seoul within a multivariate analytical framework. However, several methodological and contextual limitations should be taken into account when interpreting the findings.

First, the analysis relies exclusively on cross-sectional data from the year 2023. This temporal limitation prevents the examination of seasonal fluctuations or longer-term trends in energy use. In practice, energy consumption is influenced by dynamic factors such as economic cycles, climate variability, and policy shifts. Future research could address this limitation by incorporating panel datasets or time-series analyses, allowing for a more robust understanding of temporal patterns.

Second, the study focuses solely on Seoul, a city with unique urban characteristics including high population density, advanced infrastructure, and extensive availability of digital data. These context-specific conditions enhance the resolution of the analysis but also constrain the generalisability of the findings. Comparative studies across different urban settings would be beneficial to assess the external validity of the proposed modelling framework.

Third, the behavioural variables in this study—namely, floating population indicators—are derived from mobile network data, which may contain inherent biases due to variations in data collection methodologies, coverage across providers, or limitations in capturing nuanced human mobility. Future work should consider integrating multiple data sources, such as public sensor networks or multimodal mobility datasets, to enhance measurement accuracy and behavioural representation.

Finally, the regression-based methodology employed in this study assumes linear relationships between variables. While useful for interpretability, this assumption may oversimplify the complex interactions that characterise real-world energy consumption. To overcome this limitation, the study outlines a plan to implement a deep learning-based approach using a Multi-Layer Perceptron (MLP), which can capture non-linear dynamics and improve prediction accuracy.

Despite these limitations, the study contributes meaningfully to the urban energy modelling literature by demonstrating how high-resolution, multi-source data can be operationalised within a consistent analytical framework. The proposed methodology supports both academic inquiry and practical decision-making in the pursuit of energy-efficient and data-informed urban environments.

References

An, J., Kim, H. & Lee, H. (2014) 'A study on residential energy consumption according to dwelling type, area and building age', *Journal of the Architectural Institute of Korea*, 30(4), pp. 233-240.

Chen, Y., Wang, S., Li, H., Zhang, Y. & Wang, Y. (2022) 'Exploring the impact of floating population with different household registrations on urban crime: A case study of Guangzhou, China', *ISPRS International Journal of Geo-Information*, 11(8), pp. 443. Available at: https://doi.org/10.3390/ijgi11080443

Chou, J.-S. & Bui, D.-K. (2014) 'Modeling heating and cooling loads by artificial intelligence for energy-efficient building design', *Energy and Buildings*, 82, pp. 437-446. Available at: https://doi.org/10.1016/j.enbuild.2014.07.036

Ewing, R. & Rong, F. (2008) 'The impact of urban forms on US residential energy use', *Housing Policy Debate*, 19(1), pp. 1-30. Available at: https://doi.org/10.108/10511482.2008.9521624

Google Earth Engine (2023) 'Sentinel-2 surface reflectance imagery dataset'. Available at: https://code.earthengine.google.com/

International Energy Agency (IEA) (2022) 'World Energy Outlook 2022'. Paris: IEA. Available at: https://www.iea.org/reports/world-energy-outlook-2022

Jo, J., Kim, H. & Park, C. (2022) 'Seasonal variation of electricity consumption in residential buildings according to temperature', *Korean Energy Economics Review*, 21(1), pp. 35-52.

Jung, H., Kim, M. & Yoon, Y. (2015) 'Analysis of building energy consumption according to urban form factors', *Journal of the Korean Institute of Architectural Sustainable Environment and Building Systems*, 9(3), pp. 219-226.

Kim, H. (2013) 'Comparative analysis of energy consumption by residential building structure type', *Journal of the Korean Solar Energy Society*, 33(2), pp. 47-56.

Kim, H.J. & Park, Y. (2021) 'Spatial equity of urban park distribution: Examining the floating population in Seoul, South Korea', *Land*, 10(3), pp. 243. Available at: https://doi.org/10.3390/land10030243

Kim, J., Choi, M. & Lee, S. (2017) 'Effect of land price on energy consumption in urban commercial buildings', *Journal of Korean Geographical Society*, 52(3), pp. 323-337.

Ko, Y. (2013) 'Urban green infrastructure as a climate change adaptation strategy', *Sustainable Cities and Society*, 9, pp. 89-100. Available at: https://doi.org/10.1016/j.scs.2013.03.004

Li, X., Zhou, Y., Asrar, G.R., Imhoff, M. & Li, X. (2019) 'The surface urban heat island response to urban expansion: A panel analysis for the conterminous United States', *Science of The Total Environment*, 658, pp. 709-717. Available at: https://doi.org/10.1016/j.scitotenv.2018.12.161

Li, Y., Li, X. & Liu, X. (2018) 'Influence of temperature fluctuations on electricity demand in urban buildings', *Energy Reports*, 4, pp. 397-403. Available at: https://doi.org/10.1016/j.egyr.2018.06.007

Santamouris, M. (2016) 'Innovating to zero the building sector in Europe: Minimizing energy consumption, eradication of energy poverty and mitigating the local climate change', *Solar Energy*, 128, pp. 61-94. Available at: https://doi.org/10.1016/j.solener.2016.01.021

Shi, Y., Zhao, X., Wang, J., Li, X. & Zhao, Y. (2021) 'Can the marketization of urban land transfer improve energy efficiency? Evidence from China', *Journal of Cleaner Production*, 278, pp. 123956. Available at: https://doi.org/10.1016/j.jclepro.2020.123956

Yalcintas, M. & Ozturk, U.A. (2007) 'An energy benchmarking model based on artificial neural network method utilizing US Commercial Buildings Energy Consumption Survey (CBECS) database', *International Journal of Energy Research*, 31(4), pp. 412-421. Available at: https://doi.org/10.1002/er.1232

Zhang, Y., Yu, Z. & Li, H. (2020) 'Exploring the relationship between building size and energy consumption in commercial facilities', *Applied Energy*, 262, p.114567.