# A novel sensitivity analysis of residential building retrofitting strategies: Artificial neural network vs. linear regression

*Huan* Tang[1], *Wanyu* Lai[1], *Zuliang* Lu[1], *Yating* Zhang[2], and *Rudai* Shan[1*]

[1]Jangho Architecture College, Northeastern University, Shenyang, China
[2]China northeast architecture design & research institute Co., LTD., Shenyang, China

**Abstract.** This study compares the implementation of traditional methods (linear regression) and artificial neural network (ANN) algorithms in sensitivity analysis for residential building retrofitting strategies. The impact of building retrofitting strategies, such as wall insulation, roof insulation, glazing types, and shading systems, on the life-cycle carbon emissions ($LCCO_2$) and life-cycle costs (LCC) of a residential building in the severe cold climate of China is investigated. The results demonstrated that the ANN modelling was more accurate and stable than the linear regression method. Through impact factor analysis, it was found that window area and insulation types strongly impact the LCC, while shading, insulation type and insulation thickness strongly influence the $LCCO_2$ in the building. This study validated the feasibility and efficiency of the ANN methodology in sensitivity analysis for residential building retrofitting strategies.

## 1 Introduction

The building sector represents a significant contributor to global greenhouse gas emissions and primary energy consumption. This underscores the substantial potential of building retrofits to enhance energy efficiency and reduce carbon emissions [1]. A number of predictive methods exist for assessing a building's energy consumption. Among these, artificial neural networks (ANN) and linear regression models have proven effective, particularly in forecasting weekday electricity use [2]. ANNs are employed in a variety of fields, including energy prediction modelling, pattern recognition, intelligent control, and information processing [3], [4]. Their efficacy has been demonstrated through a high degree of conformity between predicted outcomes and actual experimental data. Moreover, ANNs have been shown to be effective in optimizing the life-cycle carbon emissions ($LCCO_2$) and life-cycle costs (LCC) of buildings, as well as in predicting the impact of various retrofit variables [5].

This study assesses the efficacy of linear regression and ANN algorithms in predicting the $LCCO_2$ and LCC of buildings. The study examines the impact of retrofit strategies, including insulation thickness, glazing type, and shading systems, on these metrics in residential buildings located in China's severe cold climates. In order to assess the influence of each input data element, impact value factors were applied to these datasets, thus enabling a nuanced analysis of the impact of retrofit strategies on building performance.

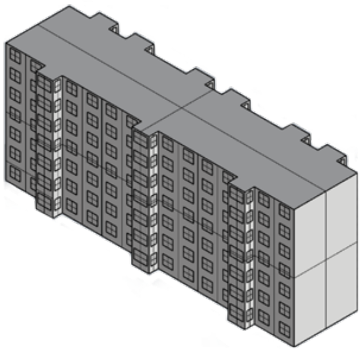
### 1.1 Case study

The case study is an existing residential building located in the city of Shenyang (41°48′N, 123°25′E), which is the extreme cold climate of China. The average temperature of the coldest month is -11.2°C, while the average temperature of the hottest month is 25°C. District-wide central heating systems are a common feature of the region, with usage extending from late October to early April. **Fig. 1** illustrates the 7-storey residential building analyzed in this study, which has a total floor area of 3519 m².

**Table 1** provides the details of the materials and construction for the existing building facade.

* Corresponding author: shanrudai@mail.neu.edu.cn


(a) Photograph of the case study building



(b) Perspective view of the 3D model

**Fig. 1.** Photograph and 3D model for the case study building.

**Table 2** provides detailed information of the input variables employed in this analysis. In order to enhance the realism and representativeness of the simulations, the Latin Hypercube Sampling (LHS) method was employed for the generation of input combinations. This method ensures a uniform distribution across variables, which is crucial for the subsequent sensitivity analysis. The energy simulation was performed using the EnergyPlus simulation engine.

### 1.2 Sensitivity analysis based on Machine-Learning model

Recent advances in building performance optimization have employed a variety of models, including Artificial Neural Networks (ANNs), Multi-Fidelity models and Bayesian networks, with ANNs being noted for their reliability and accuracy [6]. ANNs are particularly valued for solving complex non-linear problems due to their fast prediction capabilities and accurate results. This study aims to refine analysis methods by contrasting the effectiveness of ANNs with linear regression models in performing sensitivity analysis.

To evaluate the robustness of the machine learning model, samples were generated using the Latin Hypercube Sampling (LHS) method in varying sizes: 20, 40, 60, 80, 100, 120, 160, 180, 200, 300, 400, and 500. The predictive accuracy of sensitivity analysis models was assessed via out-of-sample root mean square error (RMSE) and the regression (R) coefficient, both derived from cross-validation. RMSE measures the average prediction error, while R quantifies the variance explained by the model; thus, lower RMSE and higher R values signify enhanced predictive performance. It can be seen in that the RMSE and R coefficient stop performance better when the sample size is higher than 120. Therefore, the sample size of 120 is appropriate for this study, with an acceptable RMSE of 1.290 and 1.543 for ANN and liner regression, and R coefficient of 0.953 and 0.938 for ANN and liner regression, respectively.

**Table 1.** Structural and material properties of existing building envelope.

| Construction component | Layers (from exterior to interior) |
|---|---|
| Roof | waterproof layer + 20 mm cement mortar + 100 mm cement per-lite + 20 mm cement mortar+ 120 mm reinforced concrete |
| Exterior wall | 20 mm cement mortar + 370 mm clay brick + 20 mm cement mortar |
| Exterior window | 6 mm transparent glass |

**Table 2.** Variations of input variables for case study.

| Categories | Variables | Range | Cost | Carbon emission factor |
|---|---|---|---|---|
| External insulation | EPS | 0.041 W/(m$^2$·K) | 360 CNY/m$^3$ | 5020 kgCO$_2$e/t |
| | XPS | 0.033 W/(m$^2$·K) | 680 CNY/m$^3$ | 5020 kgCO$_2$e/t |
| | Mineral wool | 0.045 W/(m$^2$·K) | 540 CNY/m$^3$ | 1980 kgCO$_2$e/t |
| | South wall | 100~300 mm (step 20 mm) | | |
| | North wall | 100~300 mm (step 20 mm) | | |
| | East/West wall | 100~300 mm (step 20 mm) | | |

| | | | | |
|---|---|---|---|---|
| | Roof | 200~300 mm (step 20 mm) | | |
| Glass | South window | Glass #1: 2.59 W/(m$^2$·K), SHGC: 0.75, t-Value: 0.81 | 775 CNY/m$^2$ | 121 kgCO$_2$e/m$^2$ |
| | | Glass #2: 2.58 W/(m$^2$·K), SHGC: 0.61, t-Value: 0.61 | 890 CNY/m$^2$ | 121 kgCO$_2$e/m$^2$ |
| | | Glass #3: 2.44 W/(m$^2$·K), SHGC: 0.35, t-Value: 0.29 | 890 CNY/m$^2$ | 121 kgCO$_2$e/m$^2$ |
| | | Glass #4: 1.79 W/(m$^2$·K), SHGC: 0.43, t-Value: 0.57 | 920 CNY/m$^2$ | 121 kgCO$_2$e/m$^2$ |
| | | Glass #5: 1.44 W/(m$^2$·K), SHGC: 0.45, t-Value: 0.62 | 1200 CNY/m$^2$ | 121 kgCO$_2$e/m$^2$ |
| | North window | Glass #6: 1.35 W/(m$^2$·K), SHGC: 0.27, t-Value: 0.32 | 1900 CNY/m$^2$ | 194 kgCO$_2$e/m$^2$ |
| | | Glass #7: 1.23 W/(m$^2$·K), SHGC: 0.42, t-Value: 0.62 | 2300 CNY/m$^2$ | 194 kgCO$_2$e/m$^2$ |
| | | Glass #8: 1.01 W/(m$^2$·K), SHGC: 0.27, t-Value: 0.32 | 2600 CNY/m$^2$ | 194 kgCO$_2$e/m$^2$ |
| Shading | Overhang depth | 0~1500 mm (step 10 mm) | 40 CNY/m$^3$ | 2530 kgCO$_2$e/t |

## 2 Results

This section first compares the predictive performance of the two analytical models for the optimization objective, which is assessed mainly by comparing the RMSE and R values. The last subsection presents the results of the sensitivity analyses using partial correlation coefficient (PCC) and standardized rank regression coefficient (SRRC) for all variables.

**Table 3.** The RMSE and R for ANN and linear regression based on different sample size.

| Sample size | 20 | 40 | 60 | 80 | 100 | 120 | 140 |
|---|---|---|---|---|---|---|---|
| RMSE for ANN | 1.755 | 1.735 | 1.616 | 1.431 | 1.33 | 1.29 | 1.318 |
| RMSE for linear regression | 1.75 | 1.821 | 1.689 | 1.54 | 1.54 | 1.543 | 1.569 |
| R for ANN | 0.81 | 0.912 | 0.929 | 0.924 | 0.951 | 0.953 | 0.95 |
| R for linear regression | 0.787 | 0.896 | 0.928 | 0.911 | 0.935 | 0.938 | 0.935 |

**Table 3** continued.

| Sample size | 160 | 180 | 200 | 300 | 400 | 500 | 160 |
|---|---|---|---|---|---|---|---|
| RMSE for ANN | 1.367 | 1.407 | 1.393 | 1.309 | 1.361 | 1.343 | 1.367 |
| RMSE for linear regression | 1.598 | 1.626 | 1.611 | 1.556 | 1.587 | 1.578 | 1.598 |
| R for ANN | 0.947 | 0.944 | 0.942 | 0.95 | 0.945 | 0.947 | 0.947 |
| R for linear regression | 0.933 | 0.929 | 0.922 | 0.926 | 0.921 | 0.924 | 0.933 |

### 2.1 Accuracy and stability of ANN and linear regression

**Table 3** depicts the correlation between sample size and the predictive efficacy of the ANN and linear regression models in estimating the energy consumption of residential buildings. As the sample size increases, the root mean square error (RMSE) decreases and the R values increases, indicating an improvement in model accuracy (**Fig. 2**). Notably, the ANN model consistently outperforms the linear regression model, particularly with smaller sample sizes. This trend is evident in the significantly higher R values and lower RMSE for the ANN, which demonstrates its superior predictive capabilities. The data further indicate that as the sample size increases, the accuracy of the ANN model continues to improve, thereby underscoring its efficacy in applications involving diverse data sets.

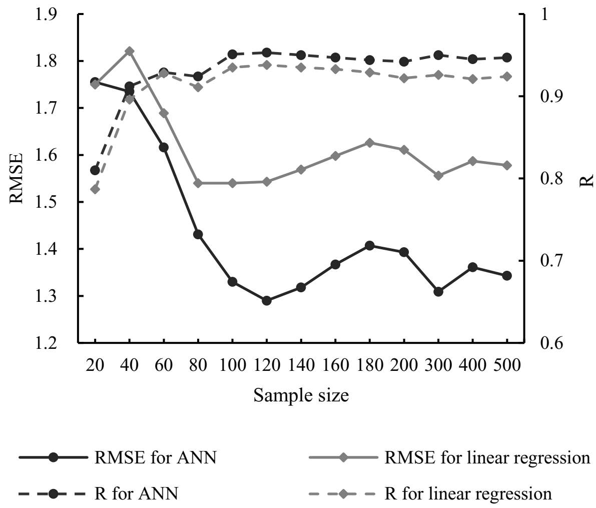



**Fig. 2.** Comparison of predictive performance of ANN and linear regression models.

### 2.2 The effect of design variables in two sensitivity analysis methods

In order to quantitatively evaluate the influence of various building design variables on specific objectives, partial correlation coefficient (PCC) and standardized rank regression coefficient (SRRC) were computed and are presented in **Fig. 3**. The aforementioned coefficients elucidate the nature of the relationships, whereby positive values signify a positive correlation, while negative values indicate an inverse relationship. The analysis indicates that roof insulation thickness exerts a minimal influence on both evaluated objectives. Conversely, the configuration of the shading system demonstrates a limited impact on the LCC and a similar minimal influence of the north window type on $LCCO_2$ emissions. In contrast, variables such as the south window type and wall insulation display varying degrees of impact on both objectives. Notably, the coefficients for the north window type are significantly positive for LCC, suggesting a beneficial effect. In contrast, they show negative implications for $LCCO_2$ emissions. The type of window, particularly in the south-facing orientation, exerts the most pronounced effect in favor of minimizing the LCC. In order to reduce $LCCO_2$ emissions, it is recommended that a cap be implemented on specific design elements. Furthermore, the shading system setting displays contrasting effects on both LCC and $LCCO_2$ emissions, emphasizing the necessity for a balanced approach in design optimization.

## 3 Conclusion

This paper examines the efficacy of two sensitivity analysis methods, namely the ANN model and the linear regression model. The sensitivity analyses of the two models yielded reliable results, enabling the identification of the most influential variables in terms of $LCCO_2$ and LCC of the building.

The Latin Hypercube Sampling (LHS) method is used to evaluate the robustness of the machine learning model to generate varying sizes. The sample size of 120 is appropriate for this study, with an acceptable RMSE of 1.290 and 1.543 for ANN and liner regression, and R coefficient of 0.953 and 0.938 for ANN and liner regression, respectively. The results demonstrated that the ANN model exhibited superior accuracy and stability in predicting electricity consumption compared to the linear regression model.

An investigation of the influencing factors revealed that glazing type has a significant impact on the $LCCO_2$ and LCC of an existing building, while the thickness of the insulation layer is also a key factor that cannot be ignored. The comparison of the performance of the ANN model and the linear regression model demonstrated that the ANN model is faster in prediction and has more accurate predictive ability in sensitivity prediction. It is recommended that the ANN model be employed more extensively in future building renovation optimization in order to provide reliable results for the sensitivity analysis.

This research was supported by Liaoning Provincial Natural Science Foundation Joint Fund, grant number 2023-MSBA-102, and Northeastern University National Project Incubation Fund (Science and Technology), grant number N2311003.

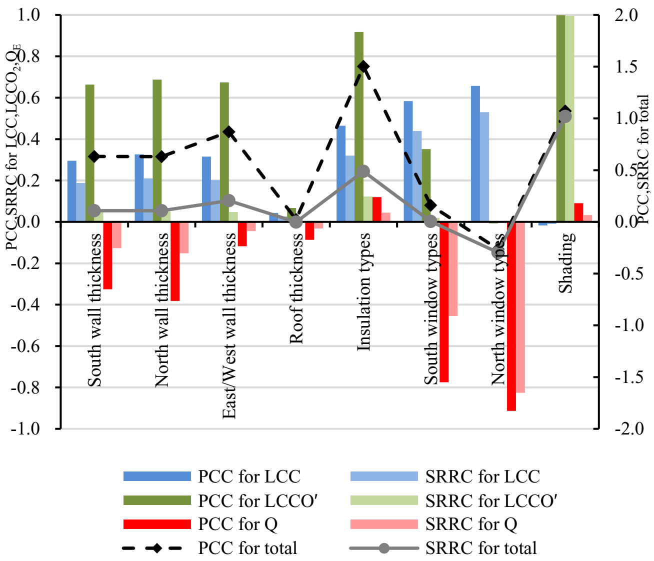



**Fig. 3.** PCC and SRRC of design variables for LCC and $LCCO_2$.

## References

1. A. S. Ahmad *et al.*, "A review on applications of ANN and SVM for building electrical energy consumption forecasting," *Renew. Sustain. Energy Rev.*, vol. 33, pp. 102–109, May 2014, doi: 10.1016/J.RSER.2014.01.069.
2. M. K. Kim, Y.-S. Kim, and J. Srebric, "Predictions of electricity consumption in a campus building using occupant rates and weather elements with sensitivity analysis: Artificial neural network vs. linear regression," *Sustain. Cities Soc.*, vol. 62, p. 102385, 2020, doi: https://doi.org/10.1016/j.scs.2020.102385.
3. F. Mostafazadeh, S. J. Eirdmousa, and M. Tavakolan, "Energy, economic and comfort optimization of building retrofits considering climate change: A simulation-based NSGA-III approach," *Energy Build.*, vol. 280, p. 112721, 2023, doi: https://doi.org/10.1016/j.enbuild.2022.112721.
4. F. Bre, N. Roman, and V. D. Fachinotti, "An efficient metamodel-based method to carry out multi-objective building performance optimizations," *ENERGY Build.*, vol. 206, 2020, doi: 10.1016/j.enbuild.2019.109576.
5. Q. Xue, Z. Wang, and Q. Chen, "Multi-objective optimization of building design for life cycle cost and CO2 emissions: A case study of a low-energy residential building in a severe cold climate," *Build. Simul.*, vol. 15, no. 1, pp. 83–98, Jan. 2022, doi: 10.1007/s12273-021-0796-5.
6. C. Lu, S. Li, and Z. Lu, "Building energy prediction using artificial neural networks: A literature survey," *Energy Build.*, vol. 262, p. 111718, 2022, doi: https://doi.org/10.1016/j.enbuild.2021.111718.