



Performance evaluation of machine learning models for multi-output building energy prediction

Tidarat Luangrungruang ^a and Gawalee Phatai ^{a*}

^a *Department of Computer, Faculty of Science and Technology, Sakon Nakhon Rajabhat University, Sakon Nakhon, 47000, Thailand*

Abstract

Article history:

Received: 27-03-2026

Revised : 04-07-2026

Accepted: 07-07-2026

Published: 19-07-2026

Keywords:

Multi-output regression;
Building energy prediction;
Heating and cooling load;
LightGBM;
Gradient boosting;
Multi-task learning

Accurate prediction of building heating and cooling loads is central to energy-efficient design, and because these loads are strongly correlated, they represent a suitable multi-output regression problem. This study compares six models from three families—linear (Linear Regression, Polynomial Regression), kernel-based (Support Vector Regression, Kernel Ridge Regression), and gradient-boosting ensembles (XGBoost, LightGBM)—for predicting the heating (y_1) and cooling (y_2) loads, and additionally evaluates a joint neural network (Joint NN) that predicts both outputs together. The models were assessed on the publicly available Energy Performance of Buildings dataset, comprising 768 samples with eight input variables and two highly correlated outputs ($r = 0.98$). Performance was measured by mean absolute error (MAE), mean relative error (MRE), and mean squared error (MSE) using 10-fold cross-validation repeated 100 times, reported as mean $\pm$ standard deviation. LightGBM achieved the lowest error on both outputs (MAE = 0.0077 ± 0.0009 for heating and 0.0173 ± 0.0022 for cooling), reducing MAE by 15.4% and 37.5% relative to the next-best model, XGBoost. Paired statistical tests confirmed that this advantage was significant across models. Although the two outputs are strongly correlated, the joint neural network did not outperform the independent ensembles on this moderate-sized dataset, suggesting that strong per-output models are better suited to data of this size. Residual analysis supported these findings, with LightGBM producing residuals randomly dispersed around zero. Overall, the study combines multiple error metrics, statistical testing, and residual diagnostics to guide model selection in correlated multi-output energy prediction.

* Corresponding author : gawalee@snru.ac.th

DOI: <https://doi.org/10.55674/cs.v18i3.266914>

© [2026] The Author(s).

1. Introduction

Multi-output prediction has become increasingly important in data science, as many real-world problems involve interdependent target variables [1]. Applications in medicine, economics, engineering, and environmental science often show strong correlations among outputs, and treating them independently through single-output models can discard structural information and reduce accuracy. Building energy prediction illustrates this well: interdependent loads such as heating and cooling must often be estimated together, which places the problem naturally within multi-output regression.

Growing concern over energy inefficiency has intensified research on the energy performance of buildings (EPB). HVAC systems account for a large share of building energy use because of their role in maintaining thermal

comfort [2]. Accurate estimation of heating and cooling loads is therefore essential for sizing these systems appropriately, and depends on factors such as building geometry, occupancy behavior, and climate.

Building energy simulation tools are widely used for this purpose, but their complexity and computational cost have motivated data-driven and machine learning alternatives [3, 4]. These models handle large datasets and capture nonlinear relationships efficiently, which makes them well suited to load prediction.

Despite this progress, prior work on multi-output regression has often been limited in scope. Many studies adopt single-output frameworks or focus only on linear models, overlooking methods better able to capture dependencies among targets [5, 6]. Deep and multi-task learning approaches have also been applied to correlated

energy loads. For example, hybrid CNN–GRU architectures with hard-parameter sharing have been used to jointly forecast cooling, heating, and electricity loads [7], and transformer-based models have been proposed for floor-level multi-load forecasting in real buildings [8]. Such models capture complex interdependencies effectively but typically rely on large training sets, which is not always practical for moderate-sized design datasets such as EPB.

Comparative analyses spanning linear, kernel-based, and ensemble families on a common multi-output energy task therefore remain scarce. To address this gap, this study systematically compares six representative models—Linear Regression (LR), Polynomial Regression (PR), Support Vector Regression (SVR), Kernel Ridge Regression (KRR), XGBoost, and LightGBM—for the simultaneous prediction of heating and cooling loads on the EPB dataset. To test whether explicitly modeling the two outputs together is beneficial, a joint multi-output neural network is also evaluated against these independent learners. Few studies have jointly compared models from all three families, together with a joint multi-output baseline, under a repetition-based cross-validation protocol on the EPB dataset.

This study addresses three questions that go beyond simply ranking the models: (i) how model complexity relates to predictive accuracy across correlated outputs; (ii) whether jointly modeling the two outputs improves accuracy over predicting them independently; and (iii) whether aggregate error metrics alone adequately reflect model quality, or whether residual diagnostics expose structural limitations that these metrics miss.

The insights gained here are not confined to building energy analysis; they can inform other multi-output problems with interrelated targets, such as multivariate economic analysis, multi-dimensional environmental forecasting, and complex engineering system modeling [9 - 11].

The main contributions of this study are as follows:

- We compare six models from three methodological families—linear, kernel-based, and gradient-boosting ensembles—on a common multi-output building energy task.
- We evaluate a joint multi-output neural network against independent per-output models to test whether exploiting the correlation between outputs improves accuracy.
- We adopt a repetition-based protocol (10-fold cross-validation repeated 100 times) with paired statistical testing, reporting both accuracy and the stability of each model.
- We combine residual diagnostics with error metrics, showing that models with similar error magnitudes can differ in structural adequacy, and provide insight into the asymmetry between heating- and cooling-load prediction.

2. Materials and Methods

This section describes the regression task, the models compared, the dataset and its preprocessing, the model

configurations, and the evaluation procedure. Fig. 1 summarizes the workflow, which follows a standard machine learning pipeline from data preparation to evaluation.

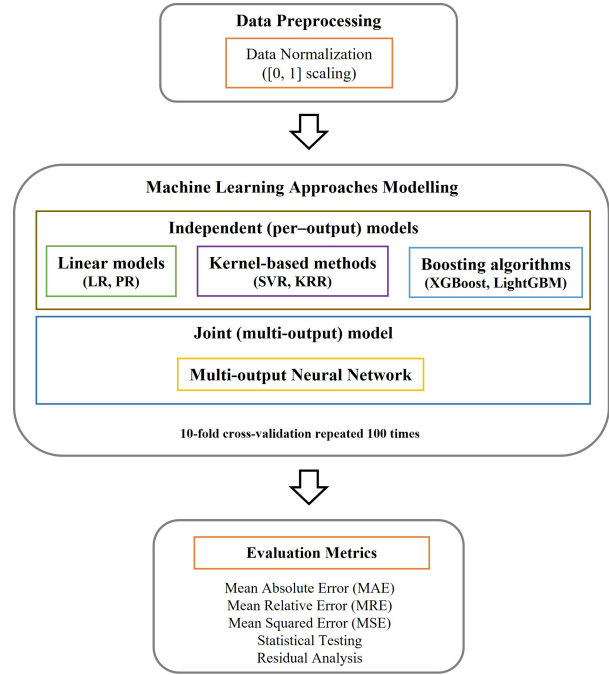


Fig. 1 Experimental workflow for the multi-output EPB benchmarking study.

Multi-Output Regression

Multi-output regression, also called multi-target or multivariate regression, predicts several continuous outputs at once [12]. It is useful when the targets are related or share underlying factors, since modeling them together can improve accuracy and reduce prediction bias. In building energy analysis, this setting arises naturally because loads such as heating and cooling depend on the same design variables and are often required together [13, 14].

Predictive Models

Six independent models from three families of increasing complexity were compared, together with a joint neural network.

Linear Regression (LR) models the relationship between the inputs and each output as a linear equation [15]. It is simple, fast, and easy to interpret, and serves here as a baseline against which the more flexible models are judged [16, 17]. Its main limitation is that it cannot represent nonlinear behavior.

Polynomial Regression (PR) extends LR by adding higher-order and interaction terms, allowing it to fit curved relationships while keeping the linear-model framework [18 - 20]. It provides an intermediate step between linear and nonlinear methods, though high-degree terms can lead to overfitting on small datasets.

Support Vector Regression (SVR) maps the inputs into a high-dimensional feature space through a kernel function and fits a regression function within a tolerance margin [21 - 23]. With an RBF kernel it captures nonlinear patterns and remains stable on small or noisy data, which suits energy and environmental problems.

Kernel Ridge Regression (KRR) combines ridge regression with the kernel trick, solving a regularized least-squares problem with an L2 penalty [24 - 26]. Its closed-form solution makes training fast, and the RBF kernel lets it model nonlinear input-output relationships.

Extreme Gradient Boosting (XGBoost) is a gradient-boosted tree ensemble in which each new tree corrects the errors of the previous ones [27]. It adds L1 and L2 regularization and efficient tree construction, giving strong accuracy on structured, high-dimensional data [28, 29].

Light Gradient-Boosting Machine (LightGBM) is also a gradient-boosted tree ensemble, but it grows trees leaf-wise rather than level-wise and uses techniques such as gradient-based one-sided sampling and exclusive feature bundling to speed up training [30 - 32]. These features often yield lower loss and faster computation.

In addition to these six independent models, a joint multi-output neural network (Joint NN) based on shared hidden layers was evaluated, following the multi-task

learning principle of exploiting correlations among related outputs [7, 8]. The independent models train a separate regressor for each output, whereas the network predicts both loads at once through shared hidden layers, allowing it to use the correlation between the two outputs. This gives a direct comparison between independent and joint approaches to multi-output prediction.

Dataset and Preprocessing

The experiments used the Energy Performance of Buildings (EPB) dataset [33], which is publicly available and contains 768 building samples. Each sample has eight input variables (x1–x8) describing the building geometry and two output variables, the heating load (y1) and the cooling load (y2). Table 1 lists the variables. All values are numeric, and no entries are missing. The two outputs are strongly correlated (Pearson $r = 0.98$), which motivates comparing independent and joint modeling.

Both inputs and outputs were scaled to the range [0, 1] using min-max normalization. To prevent information leakage, the scaler was fitted only on the training fold of each split and then applied to the corresponding test fold. Because the outputs were also scaled, the error values reported later (MAE, MRE, MSE) are expressed on the normalized [0, 1] scale rather than in physical units.

Table 1 Variables in the EPB dataset.

Role	Symbol	Variable	Type	Range
Input	x1	Relative compactness	numeric	0.62 – 0.98
Input	x2	Surface area	numeric	514.50 – 808.50
Input	x3	Wall area	numeric	245.00 – 416.50
Input	x4	Roof area	numeric	110.25 - 220.50
Input	x5	Overall height	numeric	3.50 – 7.00
Input	x6	Orientation	numeric	2 – 5
Input	x7	Glazing area	numeric	0.00 – 0.40
Input	x8	Glazing area distribution	numeric	0 – 5
Output	y1	Heating load	numeric	6.01 – 43.10
Output	y2	Cooling load	numeric	10.90 – 48.03

Table 2 Model configurations.

Model	Implementation	Key settings
LR	scikit-learn, LinearRegression	Ordinary least squares
PR	scikit-learn, PolynomialFeatures + LinearRegression	Degree 2 with interaction terms
SVR	scikit-learn, SVR	RBF kernel; C = 1.0; epsilon = 0.01; gamma = scale
KRR	scikit-learn, KernelRidge	RBF kernel; alpha = 0.01; gamma = 1.0
XGBoost	xgboost 3.3.0	100 estimators; max depth 3; learning rate 0.1; L2 $\lambda = 1.0$
LightGBM	lightgbm 4.6.0	200 estimators; 31 leaves; learning rate 0.1; L2 $\lambda = 1.0$
Joint NN	scikit-learn, MLPRegressor	Hidden layers (32, 16); ReLU; learning rate 0.01; Adam; early stopping

Model Configuration

All models were implemented in Python 3.12.10. The six independent models and the neural network used scikit-learn 1.9.0, except for the two boosting models, which used the official XGBoost (v3.3.0) and LightGBM (v4.6.0)

libraries. Each independent model was trained separately for y1 and y2, whereas the joint neural network predicted both outputs together, explicitly using their correlation. Hyperparameters were fixed rather than tuned by an automated search, except for the neural network

architecture, which was selected by a preliminary comparison of several configurations. Table 2 lists the settings for each model.

Evaluation Strategy

Each model was evaluated with 10-fold cross-validation repeated 100 times, giving 1,000 train–test splits per model and output. In every repetition the data were shuffled and divided into ten folds; the model was trained on nine folds and tested on the remaining one. Results are reported as the mean and standard deviation of MAE, MRE, and MSE over all 1,000 splits, where the standard deviation reflects how stable each model is across repetitions.

To assess whether the differences between models were significant, the best model was compared with each of the others using a paired Wilcoxon signed-rank test over the 1,000 evaluations, with the effect size given by Cohen's d and the 95% confidence interval of the error differences.

MRE can become very large when a target value is close to zero, so it was computed only for test samples whose normalized target exceeded 0.1. All reported figures come from the test folds, so they measure how well each model generalizes to unseen data rather than how well it fits the training data.

The three metrics are the mean absolute error (MAE), the mean relative error (MRE), and the mean squared error (MSE), defined in Eq. (1) – (3).

The MAE quantifies the average magnitude of the prediction error regardless of its direction, as defined in Eq. (1).

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\text{Actual}_i - \text{Predict}_i| \quad (1)$$

The MRE expresses the average deviation relative to the actual value, which allows comparison across variables with different scales; a smaller MRE indicates higher accuracy. As noted above, it was computed only over samples whose normalized target exceeded 0.1. It is defined in Eq. (2).

$$\text{MRE} = \frac{1}{n} \sum_{i=1}^n \frac{|\text{Actual}_i - \text{Predict}_i|}{\text{Actual}_i} \quad (2)$$

The MSE is the average of the squared errors, which gives more weight to large deviations. It is therefore sensitive to models that produce occasional large errors, as defined in Eq. (3).

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (\text{Actual}_i - \text{Predict}_i)^2 \quad (3)$$

In these equations, Actual_i is the true value of the output variable, Predict_i is the predicted value, and n is the number of samples, with i ranging from 1 to n .

Beyond these aggregate metrics, model adequacy was examined through residual analysis. For each model, the residuals (actual minus predicted) of a representative fold were plotted against the predicted values to check whether the errors scatter randomly around zero; such a pattern indicates that no systematic structure remains unmodeled. These plots are presented in Fig. 2–8.

3. Results and Discussion

Seven models were evaluated: six that predict each output independently—Linear Regression (LR), Polynomial Regression (PR), Support Vector Regression (SVR), Kernel Ridge Regression (KRR), XGBoost, and LightGBM—and a joint neural network (Joint NN) that predicts both outputs together. The independent models span three families of increasing complexity, from simple linear baselines to kernel-based methods and gradient-boosting ensembles, while the Joint NN tests whether modeling the two outputs together improves accuracy.

The data were drawn from the Energy Performance of Buildings (EPB) dataset [33], whose eight input variables (x_1 – x_8) are used to predict the heating load (y_1) and the cooling load (y_2). The two outputs are strongly correlated (Pearson $r = 0.976$), which is what makes a joint approach plausible. Each model was evaluated with 10-fold cross-validation repeated 100 times, and MAE, MRE, and MSE were recorded on the test folds in every run, so the results reflect generalization to unseen data. Tables 3–5 summarize the three metrics for the six independent models, and the Joint NN is discussed separately below.

Table 3 MAE performance of each model on unseen test data.

	LR	PR	SVR	KRR	XGBoost	LightGBM
y1	0.0564±0.0059	0.0152±0.0014	0.0342±0.0055	0.0253±0.0022	0.0092±0.0010	0.0077±0.0009
y2	0.0611±0.0066	0.0301±0.0036	0.0446±0.0058	0.0377±0.0035	0.0277±0.0034	0.0173±0.0022

Table 4 MRE performance of each model on unseen test data.

	LR	PR	SVR	KRR	XGBoost	LightGBM
y1	0.1533±0.0152	0.0462±0.0054	0.0902±0.0125	0.0774±0.0079	0.0242±0.0033	0.0210±0.0033
y2	0.1900±0.0187	0.0775±0.0069	0.1321±0.0155	0.1245±0.0129	0.0718±0.0068	0.0487±0.0053

Table 5 MSE performance of each model on unseen test data.

	LR	PR	SVR	KRR	XGBoost	LightGBM
y1	0.0063±0.0012	0.0004±0.0001	0.0040±0.0012	0.0010±0.0002	0.0002±0.0000	0.0001±0.0000
y2	0.0075±0.0016	0.0020±0.0004	0.0050±0.0011	0.0024±0.0005	0.0017±0.0004	0.0006±0.0002

Table 3 reports the MAE. The two boosting ensembles give the lowest errors on both outputs, with LightGBM best (0.0077 ± 0.0009 for heating and 0.0173 ± 0.0022 for cooling), followed by XGBoost. PR ranks next, ahead of the kernel-based models, as its interaction terms capture much of the nonlinearity; LR shows the highest errors, reflecting its inability to represent nonlinear behavior.

Table 4 reports the MRE, which expresses the deviation relative to the true value. The ranking matches the MAE results, with LightGBM again attaining the smallest values (0.0210 ± 0.0033 for heating and 0.0487 ± 0.0053 for cooling). The wide margin between the boosting ensembles and the remaining models reflects the advantage of gradient boosting in learning complex input–output relationships.

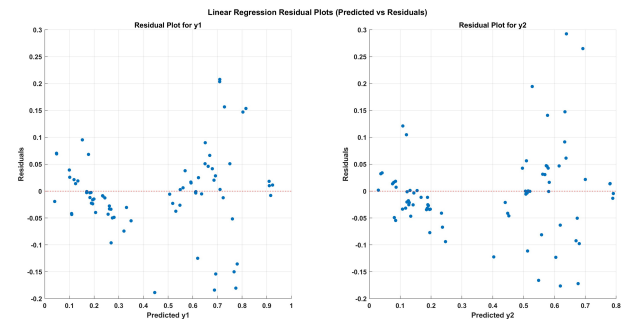
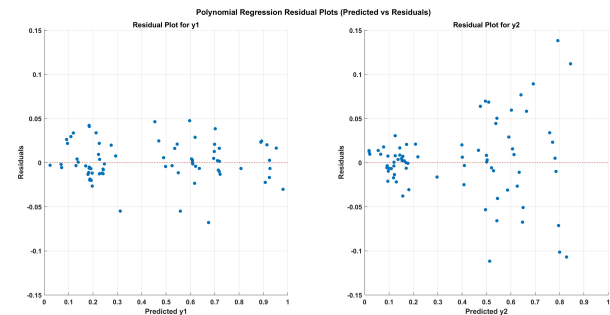
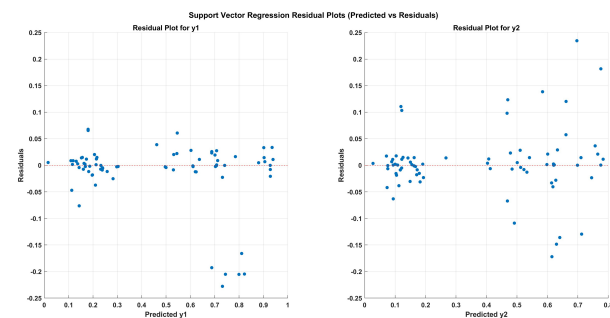
Table 5 reports the MSE, which penalizes large errors more heavily. LightGBM again gives the lowest values (0.0001 ± 0.0000 for heating and 0.0006 ± 0.0002 for cooling), with XGBoost close behind. PR remains competitive, ahead of the kernel-based models (KRR and SVR), while LR trails, consistent with its weaker fit to the nonlinear structure of the data.

Across all three metrics, the ranking is consistent: LightGBM is best, XGBoost second, PR third, followed by the kernel-based models and then LR. LightGBM also shows the smallest variance across repetitions, so it is not only the most accurate but also the most stable. The difference between the two outputs is consistent for every model: the cooling load (y2) is harder to predict than the heating load (y1), with higher errors on all metrics. For LightGBM, for example, the cooling MAE (0.0173) is more than double the heating MAE (0.0077), suggesting that the cooling load depends on the inputs in a more complex way.

To test whether these differences are statistically significant, LightGBM was compared with each other model using a paired Wilcoxon signed-rank test over the 1,000 evaluations, with Cohen's d and 95% confidence intervals of the error differences. LightGBM's advantage is significant against every model ($p < 0.001$), with large effect sizes in almost all cases. The one exception is the heating load against XGBoost, where the two are statistically close (95% CI of the MAE difference $[-0.0003, 0.0033]$, spanning zero); for the cooling load, however, LightGBM is clearly ahead (Cohen's d = 4.08). This matches the tables, where the two ensembles are near-identical on heating but diverge on cooling.

The Joint NN did not improve on the independent models. Even after its architecture was selected by a preliminary search, its errors (MAE 0.0267 for heating and 0.0429 for cooling) place it among the mid-ranked models—

better than LR but well behind LightGBM and XGBoost—and its variance is much larger (for heating, ± 0.0132 versus ± 0.0009 for LightGBM). Two factors explain this. First, the dataset is moderate in size (768 samples), which limits a neural network. Second, because the two outputs are strongly correlated and already well predicted from the shared inputs, training them jointly adds little over training them separately. On this dataset, strong per-output learners are therefore the more effective and more stable choice, consistent with the earlier expectation that deep and joint models tend to need larger datasets.

**Fig. 2** LR residual plots.**Fig. 3** PR residual plots.**Fig. 4** SVR residual plots.

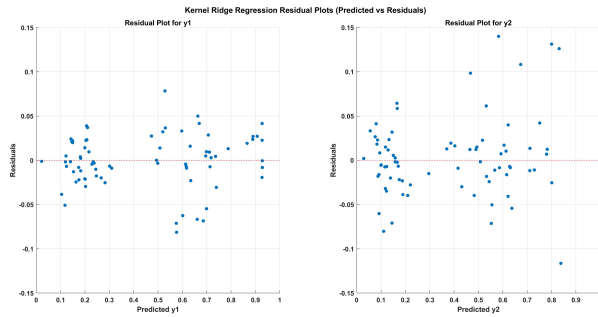


Fig. 5 KRR residual plots

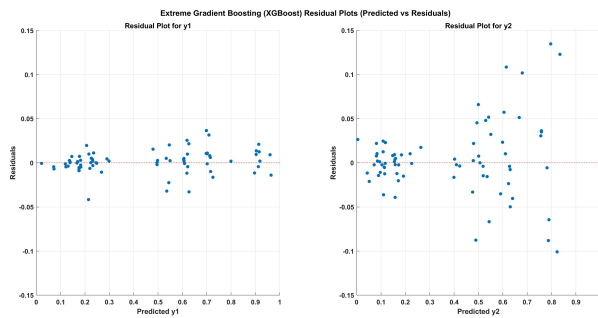


Fig. 6 XGBoost residual plots.

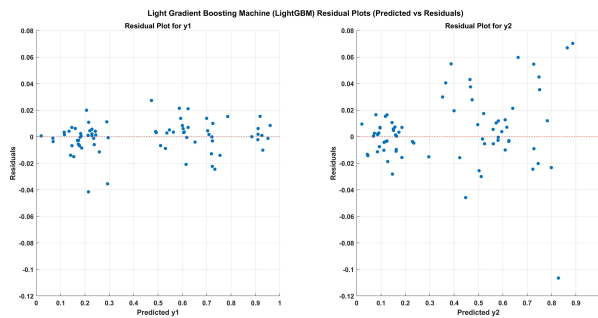


Fig. 7 LightGBM residual plots.

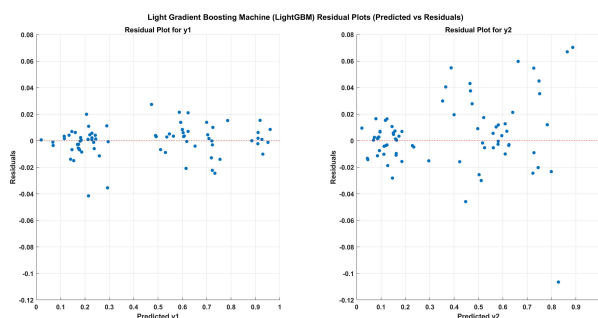


Fig. 8 Joint NN residual plots.

Feature importance from LightGBM offers some insight into the predictions. For both outputs, the relative compactness (x_1), glazing area distribution (x_8), glazing area (x_7), and orientation (x_6) are the most influential

variables, while the surface area (x_2) and overall height (x_5) contribute almost nothing. The negligible role of x_2 and x_5 is likely because they are geometrically tied to the relative compactness, so their information is already captured by x_1 .

Aggregate error metrics do not reveal whether a model systematically over- or under-predicts in particular regions, so residual analysis was used to complement them. For a well-specified model, the residuals should scatter randomly around zero with no visible pattern. The residual plots for all seven models are shown in Fig. 2–8.

The residual plots agree with the metric-based findings. LR shows clear patterns and heteroscedasticity, confirming that it misses much of the nonlinear structure. SVR and KRR reduce these effects but still leave residuals clustered in some regions, and PR, despite its low errors, retains mild structure at the extremes. XGBoost produces a more balanced spread, and LightGBM shows the most random and even dispersion around zero, indicating the least systematic bias across both outputs. The Joint NN's residuals are more scattered, consistent with its higher and less stable errors.

4. Conclusion

This study compared six independent models and a joint multi-output neural network for predicting the heating and cooling loads of buildings, using the EPB dataset under a repetition-based cross-validation protocol. Among all models, LightGBM achieved the lowest error on both outputs, and paired statistical tests confirmed that its advantage over the other models was significant, except on the heating load, where it was statistically comparable to XGBoost. LightGBM also showed the smallest variance across repetitions, indicating stable as well as accurate predictions.

Two findings stand out beyond the overall ranking. First, the cooling load was consistently harder to predict than the heating load across all models, pointing to a more complex dependence on the design variables. Second, although the two outputs are strongly correlated, the joint neural network did not outperform the independent models; on a dataset of this size, strong per-output learners were both more accurate and more stable. Feature-importance analysis further indicated that relative compactness, glazing area distribution, glazing area, and orientation were the main drivers of both loads. These findings are based on a single benchmark dataset with fixed hyperparameters and should be interpreted within those limits.

In practical terms, these results suggest that gradient-boosting models such as LightGBM are well suited to building load estimation during early design, where fast and reliable predictions from a limited set of design parameters can support energy-efficient decisions and HVAC sizing without the cost of detailed simulation. The feature-importance results also highlight which design variables most influence the loads, which can help prioritize design choices. To support reproducibility, the full implementation,

including all model configurations and the evaluation procedure, is provided as supplementary material.

5. Suggestions

This study points to several directions for future work. First, the joint model tested here was a single feedforward network; other multi-task architectures—such as shared-representation networks combining gradient boosting with neural feature extraction—may exploit output correlations more effectively, particularly on larger and more dynamic datasets where neural models are better supported. Second, hyperparameters were fixed in this study, so systematic tuning through grid, random, or metaheuristic search could further improve model generalization. Third, the feature-importance analysis could be extended with dedicated explainable-AI methods, such as SHAP, to examine variable interactions in more depth and support the adoption of these models in building management systems. Finally, applying the framework to other datasets and domains—such as renewable energy forecasting, environmental monitoring, or urban energy networks—would help establish how well the findings generalize beyond a single benchmark.

6. Acknowledgement

This research was partially supported by the Department of Computer, Faculty of Science and Technology, Sakon Nakhon Rajabhat University, Sakon Nakhon, Thailand. The authors would like to express their sincere gratitude for the support and resources provided.

7. Declaration of Generative AI in Scientific Writing

The authors acknowledge the use of generative AI tools to enhance language clarity, grammar, and readability.

8. CRediT Author Statement

Tidarat Luangrungruang: Conceptualization, Methodology, Data analysis, Writing Original Draft. **Gawalee Phatai:** Conceptualization, Methodology, Data analysis, Writing Original Draft, Review, EditingText.

9. Research Involving Human and Animals Rights

This study did not involve any research or experiments on human participants or animals. Therefore, it does not pertain to human or animal rights, and ethical approval and informed consent were not required.

10. Ethics Approval and Consent to Participate

Not applicable, as this study did not involve human participants or animals; therefore, ethical approval and informed consent were not required.

11. Declaration of Competing Interest

The authors declare no conflicts of interest.

12. References

- [1] Xu, D., Shi, Y., Tsang, I. W., Ong, Y.-S., Gong, C., & Shen, X. (2020). Survey on Multi-Output Learning. *IEEE Transactions on Neural Networks and Learning Systems*, 31(7), 2409–2429. <https://doi.org/10.1109/TNNLS.2019.2945133>
- [2] Hashempour, N., Taherkhani, R., & Mahdikhani, M. (2020). Energy performance optimization of existing buildings: A literature review. *Sustainable Cities and Society*, 54, 101967. <https://doi.org/10.1016/j.scs.2019.101967>
- [3] Jeong, J.-Y., Kang, J.-S., & Jun, C.-H. (2020). Regularization-based model tree for multi-output regression. *Information Sciences*, 507, 240–255. <https://doi.org/10.1016/j.ins.2019.08.034>
- [4] Ali, U., Bano, S., Shamsi, M. H., Sood, D., Hoare, C., Zuo, W., Hewitt, N., & O'Donnell, J. (2024). Urban building energy performance prediction and retrofit analysis using data-driven machine learning approach. *Energy and Buildings*, 303, 113768. <https://doi.org/10.1016/j.enbuild.2023.113768>
- [5] Wang, Z., Feng, Z., Ma, Z., & Peng, J. (2024). A multi-output regression model for energy consumption prediction based on optimized multi-kernel learning: A case study of tin smelting process. *Processes*, 12(1), 32. <https://doi.org/10.3390/pr12010032>
- [6] Rajaperumal, T. A., & Christopher Columbus, C. (2025). Enhanced wind power forecasting using machine learning, deep learning models and ensemble integration. *Scientific Reports*, 15(1), 20572. <https://doi.org/10.1038/s41598-025-05250-3>
- [7] Zhao, H., Liu, X., & Peng, D. (2025). Multi-energy load forecasting for integrated energy system based on multi-task learning. In Z. Dong, G. Li & Z. Li (Eds.), *2025 5th Power System and Green Energy Conference (PSGEC)* (pp. 252–256). IEEE. <https://doi.org/10.1109/PSGEC66102.2025.11150989>
- [8] Cen, S., & Lim, C. G. (2024). Multi-Task Learning of the PatchTCN-TST Model for Short-Term Multi-Load Energy Forecasting Considering Indoor Environments in a Smart Building. *IEEE Access*, 12, 19553–19568. <https://doi.org/10.1109/ACCESS.2024.3355448>
- [9] Sahoo, K., & Thakur, M. (2025). A Multi-output ensemble learning approach for multi-day ahead index price forecasting. *AppliedMath*, 5(1), 6. <https://doi.org/10.3390/appliedmath5010006>
- [10] Setiawan, K. E., Shiddiqi, H. A., & Wicaksono, P. (2024). Multi-output machine learning regression for climate prediction: a comparative study of precipitation and temperature forecasting in Jakarta and East Kalimantan, Indonesia. *Communications in Mathematical Biology and Neuroscience*, 2024, 97. <https://doi.org/10.28919/cmbn/8820>
- [11] Tran, N. K., Kühle, L. C., & Klau, G. W. (2024). A critical review of multi-output support vector regression. *Pattern Recognition Letters*, 178, 69–75. <https://doi.org/10.1016/j.patrec.2023.12.007>
- [12] Borchani, H., Varando, G., Bielza, C., & Larrañaga, P. (2015). A survey on multi-output regression. *WIREs Data Mining and Knowledge Discovery*, 5(5), 216–233. <https://doi.org/10.1002/widm.1157>
- [13] Ahmed, M. I., Nurjahan, Mahbub-Or-Rashid, M., & Islam, F. (2023). Prediction of Death Counts Based on Short-term

- Mortality Fluctuations Data Series using Multi-output Regression Models. *International Journal of Advanced Computer Science and Applications*, 14(5), 1165–1171. <https://doi.org/10.14569/IJACSA.2023.01405120>
- [14] Brindha, G. R., Rishiikeshwer, B. S., Santhi, B., Nakendraprasath, K., Manikandan, R., & Gandomi, A. H. (2022). Precise prediction of multiple anticancer drug efficacy using multi target regression and support vector regression analysis. *Computer Methods and Programs in Biomedicine*, 224, 107027. <https://doi.org/10.1016/j.cmpb.2022.107027>
- [15] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis* (6th ed.). John Wiley & Sons.
- [16] Rath, S., Tripathy, A., & Tripathy, A. R. (2020). Prediction of new active cases of coronavirus disease (COVID-19) pandemic using multiple linear regression model. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(5), 1467–1474. <https://doi.org/10.1016/j.dsx.2020.07.045>
- [17] Yang, S. J. H., Lu, O. H. T., Huang, A. Y. Q., Huang, J. C. H., Ogata, H., & Lin, A. J. Q. (2018). Predicting Students' Academic Performance Using Multiple Linear Regression and Principal Component Analysis. *Journal of Information Processing*, 26, 170–176. <https://doi.org/10.2197/ipsjip.26.170>
- [18] Cheng, C., & Schneeweiss, H. (1998). Polynomial Regression With Errors in the Variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 60(1), 189–199. <https://doi.org/10.1111/1467-9868.00118>
- [19] Shi, M., Hu, W., Li, M., Zhang, J., Song, X., & Sun, W. (2023). Ensemble regression based on polynomial regression-based decision tree and its application in the in-situ data of tunnel boring machine. *Mechanical Systems and Signal Processing*, 188, 110022. <https://doi.org/10.1016/j.ymssp.2022.110022>
- [20] Amrutphale, J., Rathore, P., & Malviya, V. (2020). A Novel Approach for Stock Market Price Prediction Based on Polynomial Linear Regression. In R. Shukla, J. Agrawal, S. Sharma, N. Chaudhari, & K. Shukla (Eds), *Social Networking and Computational Intelligence: Proceedings of SCI-2018* (pp. 161–171). Springer. https://doi.org/10.1007/978-981-15-2071-6_13
- [21] Zhang, F., & O'Donnell, L. J. (2020). Support vector regression. In A. Mechelli & S. Vieira (Eds), *Machine Learning* (pp. 123–140). Elsevier. <https://doi.org/10.1016/B978-0-12-815739-8.00007-9>
- [22] Ma, Z., Ye, C., & Ma, W. (2019). Support vector regression for predicting building energy consumption in Gradient Boosting Machine Model. In N. Priyadarshi, S. Padmanaban, R. K. Ghadai, A. R. Panda, & R. Patel (Eds), *Advances in Power Systems and Energy Management. ETAEERE ETAEERE 2020 2020. Lecture Notes in Electrical Engineering*, vol 690. (pp. 413–422). Springer. https://doi.org/10.1007/978-981-15-7504-4_39
- [32] Ramalingam, K., Yadalam, P. K., Ramani, P., Krishna, M., Hafedh, S., Badnjević, A., Cervino, G., & Minervini, G. (2024). Light gradient boosting-based prediction of quality southern China. *Energy Procedia*, 158, 3433–3438. <https://doi.org/10.1016/j.egypro.2019.01.931>
- [23] Hasan, M. S., Tarequzzaman, M., Moznuzzaman, M., & Ahad Juel, M. A. (2025). Prediction of energy consumption in four sectors using support vector regression optimized with genetic algorithm. *Heliyon*, 11(2), e41765. <https://doi.org/10.1016/j.heliyon.2025.e41765>
- [24] Vovk, V. (2013). Kernel ridge regression. In B. Schölkopf, Z. Luo, & V. Vovk, (Eds), *Empirical inference: Festschrift in honor of vladimir n. vovk* (pp. 105–116). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-41136-6_11
- [25] Panigrahy, S. K., & Chiang, K.-N. (2021). Study on an Artificial Intelligence Based Kernel Ridge Regression Algorithm for Wafer Level Package Reliability Prediction. In L. O'Conner (Ed), *2021 IEEE 71st Electronic Components and Technology Conference (ECTC)* (pp.1435–1441). <https://doi.org/10.1109/ECTC32696.2021.00229>
- [26] Zhang, R., Li, Y., & Gui, Y. (2023). Prediction of rock blasting induced air overpressure using a self-adaptive weighted kernel ridge regression. *Applied Soft Computing*, 148, 110851. <https://doi.org/10.1016/j.asoc.2023.110851>
- [27] Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp.785–794). Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>
- [28] Islam, S. F. N., Sholahuddin, A., & Abdullah, A. S. (2021). Extreme gradient boosting (XGBoost) method in making forecasting application and analysis of USD exchange rates against rupiah. *Journal of Physics: Conference Series*, 1722(1), 012016. <https://doi.org/10.1088/1742-6596/1722/1/012016>
- [29] Noorunnahar, M., Chowdhury, A. H., & Mila, F. A. (2023). A tree based eXtreme Gradient Boosting (XGBoost) machine learning model to forecast the annual rice production in Bangladesh. *PLOS ONE*, 18(3), e0283452. <https://doi.org/10.1371/journal.pone.0283452>
- [30] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. In *Proceedings of the 31st International Conference on Neural Information Processing Systems* (pp.3149 - 3157). Curran Associates Inc.. <https://dl.acm.org/doi/10.5555/3294996.3295074>
- [31] Chowdhury, S. R., Mishra, S., Miranda, A. O., & Mallick, P. K. (2021). Energy Consumption Prediction Using Light of life among oral cancer-treated patients. *BMC Oral Health*, 24(1), 349. <https://doi.org/10.1186/s12903-024-04050-x>
- [33] Tsanas, A., & Xifara, A. (2012). Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. *Energy and Buildings*, 49, 560–567. <https://doi.org/10.1016/j.enbuild.2012.03.003>