A COMPARATIVE ANALYSIS OF MULTI-TARGET FEATURE SELECTION METHODS IN DATA-DRIVEN MODELS FOR BUILDING ENERGY AND THERMAL PERFORMANCE PREDICTION

Esther Borkowski, Xiaojun Fan and Arno Schlueter
Architecture and Building Systems, Institute of Technology in Architecture, ETH Zurich, Zurich, Switzerland

Abstract

Building energy management increasingly utilises Machine Learning (ML) to use data from sensor-rich environments. A significant challenge in this context is managing high-dimensional data, which can affect model performance. This study addresses this by applying multi-target feature selection, an underexplored method that reduces dimensionality by analysing inter-feature relationships. From 182 features, two were key for developing three ML models predicting the energy and thermal performance of the HiLo living lab. These models achieved a robust fit with an average Root Mean Squared Error (RMSE) of 0.18 kW and 1.03 °C, demonstrating multi-target feature selection's effectiveness in enhancing building performance predictions.

Introduction

Buildings, accounting for 36% of global energy consumption, play a key role in the transition towards a more sustainable energy grid (United Nations Environment Programme, 2021). With the projected increase in the electrification of many end-use sectors and the broader integration of renewable energy sources, the focus shifts from supply-side to demand-side flexibility in grid operations. This shift is crucial for buildings to contribute significantly to decarbonisation targets by managing their energy consumption and generation dynamically, a concept known as energy flexibility (Li et al., 2022).

In this evolving landscape, buildings are expected to provide nearly half of the flexible demand capacity by 2030. This transformation is accelerated by the 'Internet of Things', which has increased the use of sensors in Heating, Ventilation and Air-Conditioning (HVAC) systems in buildings, resulting in the generation of vast amounts of data (Kathirgamanathan et al., 2019). However, these data are often high-dimensional, which creates challenges for Machine Learning (ML) algorithms due to the presence of irrelevant, redundant or noisy features that can slow down the learning process and degrade performance. Feature selection is recognised as a crucial method for reducing dimensionality. It involves choosing a subset of relevant features based on certain criteria, thereby simplifying model interpretation and decreasing memory usage, computational costs and learning time (Hashemi et al., 2021). Furthermore, should these models be applied in real buildings, the careful selection of sensors through feature selection would necessitate only essential sensors. Such efficient sensor usage would enhance economic viability and conserve critical raw materials, thereby supporting resource efficiency and circularity in buildings by reducing sensor network requirements.

In feature selection, a distinction is made between single-target and multi-target data. While single-target data involves predicting one output per sample, multi-target data consists of samples with multiple outputs, where the correlation between features and these outputs determines feature relevancy. This distinction directly influences the choice and application of feature selection methods. According to Hashemi et al. (2021) and as shown in Figure 1, feature selection methods can be broadly categorised into search methods and supervision methods, with search methods being often designed with single-target problems in mind.

On the other hand, supervision methods can address both single-target and multi-target feature selection. Within supervised ML algorithms, multi-target feature selection methods can be further divided into binary transformation and algorithm adaptation. Binary transformation initially converts multi-target data into several independent single-target data and then evaluates features for each target separately. However, this method may overlook the correlations between different targets (Masmoudi et al., 2020). In contrast, algorithm adaptation is specifically designed for multi-target data and aims to preserve and utilise these inter-target relationships.

Recent years have seen the development of various single-target feature selection methods, with numerous studies evaluating their application in ML models related to building controls (e.g. Olu-Ajayi et al., 2023; Kathirgamanathan et al., 2019; Zhang and Wen, 2019). Yet, real-world complexities often require addressing multi-target problems, especially in buildings where simultaneous prediction of multiple targets is common. Despite their significance, multi-target feature selection methods have received limited attention in research, both generally and within the built environment. While some studies have explored binary transformation for multi-target issues, to the best of the authors' knowledge, research on algorithm adaptation in this area is notably lacking. This study set out to fill this gap by comparing multi-target feature selection methods, with a particular focus on algorithm adaptation in data-

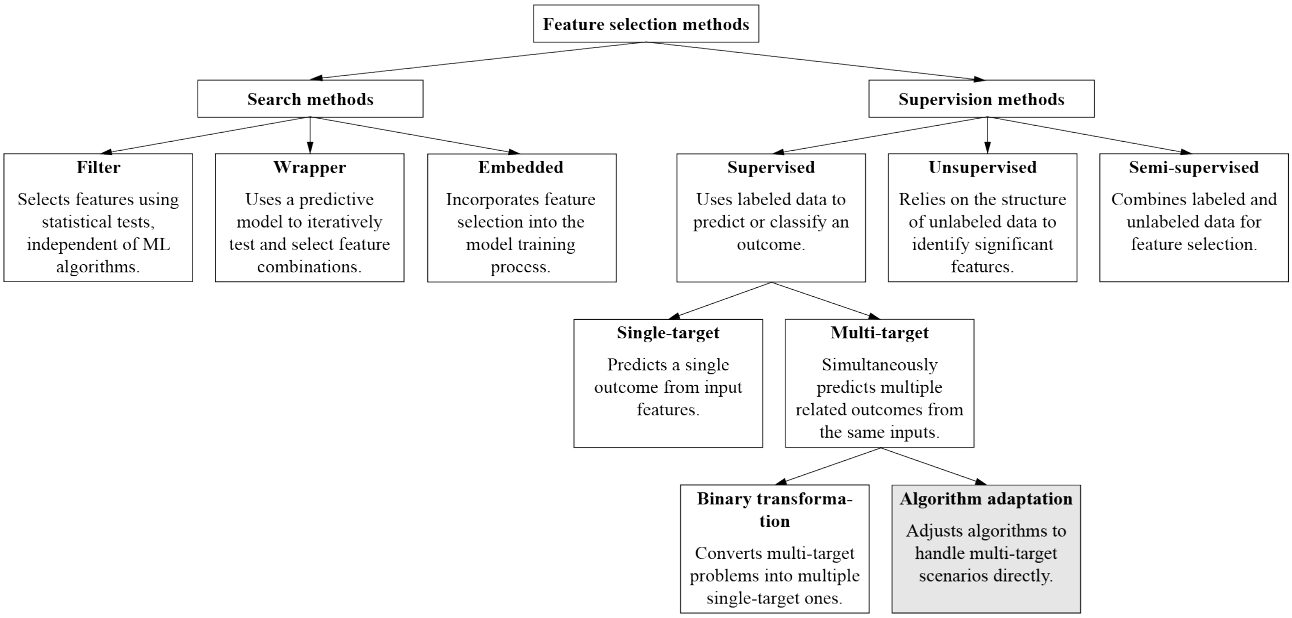

Figure 1: Categorisation of feature selection methods, highlighting the method used in this study. Based on categorisation by Hashemi et al. (2021).

driven models for accurately predicting building energy and thermal performance.

To achieve this aim, investigations have been conducted on data from the HiLo living lab in Dübendorf, Switzerland, as can be seen from Figure 2 (Block et al., 2017). Designed with the guiding principle of 'high performance, low emissions', HiLo features innovative building systems, such as Thermally Activated Building Systems (TABS) embedded into lightweight concrete structures. These TABS use a hydronic pipe network integrated into HiLo's structural components to transform the internal ceilings of two of its offices into radiant surfaces for heating and cooling (Lydon et al., 2019). Utilising concrete's thermal properties for heat storage and thermal inertia, TABS facilitate a time shift of peak heat load and a time lag of sensible heat transfer from the concrete to the interior air. These characteristics make TABS particularly effective for enhancing building energy flexibility (Arteconi et al., 2014).

Figure 2: Exterior view of HiLo from the southwest. Photo: Roman Keller.

However, controlling TABS in heavyweight concrete structures, which typically exhibit extended time constants, presents significant challenges as the regulation of heat output's timing and amount often becomes complex (van der Heijde et al., 2017). In response, this research project investigates the use of TABS in lightweight concrete structures, as exemplified in two of HiLo's offices. To model one of these similarly designed offices, i.e. Office 1, a data-driven (black-box) approach was selected due to its computational efficiency and accuracy against a more resource-intensive physics-based (white-box) model.

This study's central focus is applying multi-target feature selection during data pre-processing. This step was instrumental in accurately predicting two key aspects: (1) the thermal power of the TABS to ensure optimal control over the timing and amount of heat output and (2) the indoor air temperature in the office to maintain it within comfortable ranges. By comparing various multi-target feature selection methods, the study aims to identify the most influential sensors within HiLo's extensive sensor network. The aim is to develop a model that predicts TABS thermal power and indoor air temperature and aligns closely with the measured data. Achieving this aim is expected to provide valuable insights into improving building energy efficiency and occupant comfort and enhancing the methodological tools available for built environment research.

Methods

This section details the study conducted at the HiLo living lab, which was used to assess the effectiveness of multi-target feature selection methods in predicting the thermal power of the TABS and the indoor air temperature in

HiLo's Office 1.

Dataset

The study utilised a comprehensive dataset from HiLo, which is equipped with over 1,500 data points. For this study, a subset of 184 sensors was selected based on their relevance to predicting the cooling performance of the TABS in HiLo's Office 1. These sensors recorded data at ten-minute intervals for two summers (i.e. from 1 June to 31 August 2022 and from 1 June to 21 August 2023), covering a range of parameters from environmental conditions to HVAC system performance metrics. This dataset was then used to study the effect of 182 input variables on two key output variables: the TABS's thermal power and the office's indoor air temperature. Data extraction and analysis were conducted using Python, with libraries such as scikit-learn, pandas and matplotlib.

Data pre-processing

Data pre-processing is crucial for ML algorithm performance, as it addresses raw data imperfections and irregularities, such as high dimensionality and missing data (Olu-Ajayi et al., 2023). Therefore, data were pre-processed in this study to ensure dataset quality and avoid difficulties during model development. The process involved seven main steps:

1. **Data splitting:** Initially, the dataset was divided into features and targets, followed by splitting these into training and testing subsets for model evaluation.
2. **Biased reduction:** To prevent model overfitting and ensure accuracy, features strongly correlated with targets were removed, reducing the feature count to 162.
3. **Constant feature removal:** Non-variable features were eliminated for lacking predictive value, narrowing down the dataset to 106 features.
4. **Redundancy elimination:** Highly correlated features with a Pearson correlation above 0.9 were identified, and duplicates were removed to prevent multicollinearity, leaving 73 features.
5. **Missing data imputation:** Missing values, often a result of communication errors, were imputed using the K-Nearest Neighbours (KNN) algorithm to maintain dataset integrity.
6. **Cyclical feature encoding:** Cyclical aspects like time and solar position were encoded using sine and cosine transformations, leading to 97 features in the dataset.
7. **Supervised learning transformation:** To effectively model the TABS response, 12 lagged observations spanning a two-hour window were included to capture immediate past effects, expanding the feature set to 1,261.

Feature selection methods

The feature selection process leveraged the algorithm adaptation method, employing three ML models to determine the most predictive features:

- **Random Forest (RF):** This model combines multiple decision trees to form a more robust and accurate prediction. It is particularly adept at handling large datasets with complex relationships between features.
- **Gradient Boosting Machine (GBM):** Employing a stage-wise additive model, this technique progressively builds a model to optimise a loss function, which makes it effective for both bias reduction and variance control.
- **Support Vector Regression (SVR):** SVR applies the principles of support vector machines for regression purposes, hence providing a flexible approach to capture both linear and non-linear relationships within the data.

These models were selected due to their ability to handle diverse data characteristics and provide comprehensive insights into feature relevance.

Given that each of these models has its inherent method of assigning feature weights, which can differ significantly in scale and interpretation, feature weights were normalised to ensure comparability. This normalisation process adjusted the feature weights from each model to a common scale, thereby allowing for a direct and meaningful comparison of feature importance across different methods.

The effectiveness of each model in feature selection was then assessed based on these normalised weights, along with the impact on model performance and the computation time. The comparison aimed at identifying the most effective feature selection method for the HiLo dataset.

Model development

This study used the scikit-learn library to implement the three ML models listed above: RF, GBM and SVR. These models were chosen because of their performance in various prediction tasks, especially in scenarios involving complex and high-dimensional data like the real-world data in this study. The development process focused on optimising the predictive performance through strategic feature selection and iterative model tuning. Standard hyperparameters from scikit-learn were used due to the study's exploratory nature.

The process to determine the optimal feature set for each model involved systematic experimentation. Starting with the most predictive feature for the thermal power of the TABS and the most predictive feature for the indoor air temperature of Office 1, features were added incrementally – first in pairs, then in larger groups of ten and eventually by hundreds. This approach allowed testing all features up to the entire dataset of 1,261 features. Duplicate features were carefully removed to ensure a balanced emphasis on feature importance.

Model evaluation

The model's performance was assessed primarily using the Root Mean Squared Error (RMSE) metric, where a lower RMSE value indicates better model performance with fewer errors in prediction (Naser and Alavi, 2023).

For predicting the thermal power of the TABS (0-1 kW range) and the indoor air temperature of Office 1 (22-28 °C range), RMSE values up to 0.3 kW and 1 °C, respectively, were considered indicative of satisfactory model accuracy. This evaluation ensured that the model was not only accurate but also reliable and applicable in practical settings.

Results analysis and discussion

The study conducted with data from the HiLo living lab revealed that only a selected group of features significantly influenced the prediction of the thermal power of the TABS and the indoor air temperature in Office 1. The feature selection methods used – RF, GBM and SVR – differed in identifying these influential features. Notably, RF and GBM models assigned high importance to a few features, whereas the SVR model distributed importance more evenly across a broader range of features, as shown in Figures 3 and 4. Descriptions of the symbols for the top predictive features are provided in Table 1.

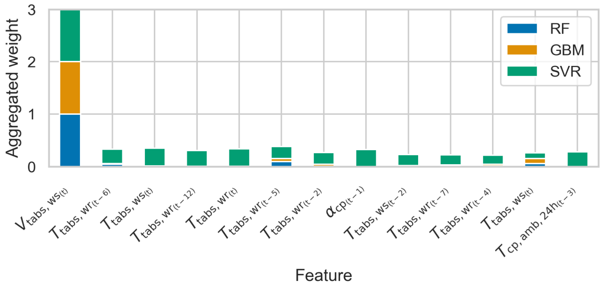

Figure 3: Aggregated weights of top six features for predicting TABS thermal power across RF, GBM and SVR models.

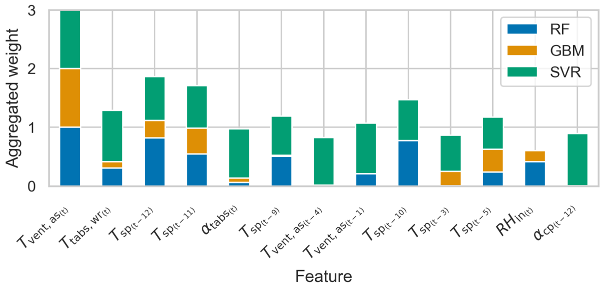

Figure 4: Aggregated weights of top six features for predicting indoor air temperature across RF, GBM and SVR models.

Despite methodological differences, all models consistently recognised two features as most important for their predictions: the supply water flow rate of the TABS at the current time step, $V_{\text{tabs,s}_{(t)}}$, for predicting the TABS's thermal power, and the air supply temperature of the ventilation system at the current time step, $T_{\text{vent,s}_{(t)}}$, for predicting the indoor air temperature. Beyond these top two features, the divergence in feature selection became more apparent. For instance, when predicting the indoor air temperature, RF and GBM models identified the zone regulation's setpoint temperature at various preceding time steps as the next two most important features. RF focused on the measurements twelve and ten steps prior, $T_{\text{sp}_{(t-12)}}$ and $T_{\text{sp}_{(t-10)}}$, whereas GBM focused on the measurements eleven and

Table 1: Top feature symbols and descriptions without time-lag notation.

Symbol	Description
$V_{\text{tabs,s}}$	Supply water flow rate of TABS
$T_{\text{tabs,r}}$	Water return temperature of TABS
$T_{\text{tabs,s}}$	Water supply temperature of TABS
α_{cp}	Valve opening of circulation pump
$T_{\text{amb,24h}}$	24h average of ambient temperature
$T_{\text{vent,s}}$	Air supply temperature of ventilation
T_{sp}	Setpoint temperature of zone regulation
RH_{in}	Indoor relative humidity

five steps prior, $T_{\text{sp}_{(t-11)}}$ and $T_{\text{sp}_{(t-5)}}$. In contrast, the SVR model yielded a different pair of features as most influential: the circulation pump's valve opening percentage twelve steps before the current time, $\alpha_{\text{cp}_{(t-12)}}$, and the water return temperature of the TABS at the current time, $T_{\text{tabs,r}_{(t)}}$.

This variance in feature selection could be attributed to the inherent characteristics of each model. RF and GBM, both being ensemble tree-based models, tended to focus on similar aspects of feature interaction and non-linear relationships. On the other hand, SVR's unique approach, particularly its sensitivity to data characteristics, led to a distinct interpretation of feature importance.

These differences also slightly influenced model performance. As Figures 5 and 6 illustrate, RMSE values remained relatively stable across varying feature counts and models, which may suggest that the models' reliability was not heavily dependent on the feature count or selection methodology. Overall, all models demonstrated a strong fit to the data, with average RMSE values of 0.18 kW for thermal power and 1.03 °C for indoor air temperature across various feature counts and selection methods.

A closer examination of the RMSE values for TABS's thermal power prediction revealed that they remained relatively constant, regardless of the number of features added, as detailed in Table 2. Including just the top two features yielded RMSE values of 0.15 kW for RF, 0.13 kW for GBM and 0.14 kW for SVR, closely aligning with the overall average. This suggests that more features did not necessarily improve RMSE. Interestingly, expanding the feature set to four or six initially degraded prediction ac-

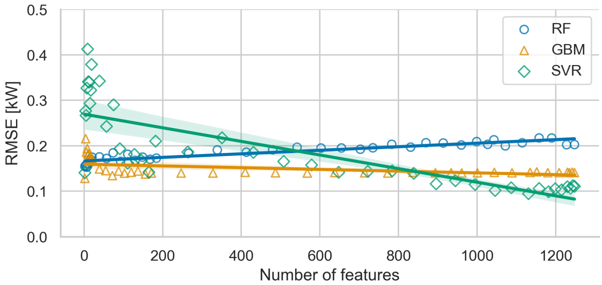

Figure 5: RMSE comparison for TABS's thermal power prediction across RF, GBM and SVR models and varying feature counts.

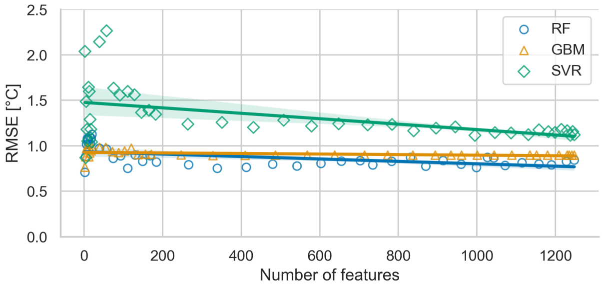

Figure 6: RMSE comparison for indoor air temperature prediction across RF, GBM and SVR models and varying feature counts.

curacy, followed by a slight improvement. However, even with the full feature set, RMSE values were 0.20 kW for RF, 0.14 kW for GBM and 0.11 kW for SVR, which indicated minimal gains from adding more features. This pattern suggests that optimising predictions for TABS's thermal power and indoor air temperature in HiLo's Office 1 might not require more than two important features, which demonstrates effectiveness in feature utilisation from an RMSE standpoint.

Table 2: RMSE comparison for TABS's thermal power prediction [kW] across RF, GBM and SVR models with top two, four and six features and overall feature average μ.

Model	Feature count 2	4	6	μ
RF	0.15	0.16	0.15	0.19
GBM	0.13	0.22	0.19	0.15
SVR	0.14	0.28	0.27	0.19

RF and GBM models showed a similar trend for indoor air temperature predictions to those observed for the TABS's thermal power predictions. Starting with the top two features provided better-than-average results, with an RMSE of 0.87 °C for RF and 0.91 °C for GBM. This performance diminished with the inclusion of four or six top features, then slightly improved again as more features were considered. This pattern, however, diverged for the SVR model. With only the top two features, the SVR's RMSE was 2.04 °C, which improved to 0.87 °C with the four top features but worsened again with six. As detailed in Table 3, the fluctuations observed with the SVR model underscore the distinct feature selection strategies inherent to each model. With its unique handling of features, the SVR model underscores the critical influence of feature selection methodology on model performance. This variability further accentuates the need for model-specific feature selection to enhance predictive accuracy and reliability.

Table 3: RMSE comparison for indoor air temperature prediction [°C] across RF, GBM and SVR models with top two, four and six features and overall feature average μ.

Model	Feature count 2	4	6	μ
RF	0.71	0.88	1.01	0.87
GBM	0.76	0.88	0.97	0.91
SVR	2.04	0.87	1.49	1.32

The runtime analysis complements these findings. As depicted in Figure 7, the runtime increased significantly with the addition of more features, underscoring the need for a careful balance between computational efficiency and model accuracy. The comparative analysis of RMSE and runtime suggests that a model with two features provided an optimal balance between accuracy and efficiency. Using two features, the runtime was 0.14 seconds for RF, 0.56 seconds for GBM and 0.63 seconds for SVR. This indicates that SVR was 77.77% slower than RF and 11.23% slower than GBM, a significant efficiency loss that could be disadvantageous in more complex modelling scenarios. This balance was consistent across different feature selection methods, offering a practical and robust solution for predicting TABS thermal power and air temperature in the given context.

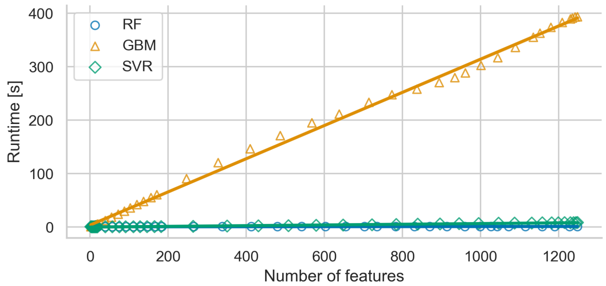

Figure 7: Runtime comparison across RF, GBM and SVR models and varying feature counts.

The overall performance metrics of the models, characterised by average RMSE values of 0.18 kW for thermal power and 1.03 °C for indoor air temperature across various feature counts and selection methods, indicate a robust fit to the dataset. This alignment is further evidenced by Figures 8 and 9, which compare the model's predicted outcomes against the measured data for TABS thermal power and indoor air temperature in HiLo's Office 1. These comparisons span a one-week period in mid-August 2023 within the test set, with each model iteration using the top two features identified by the respective feature selection method.

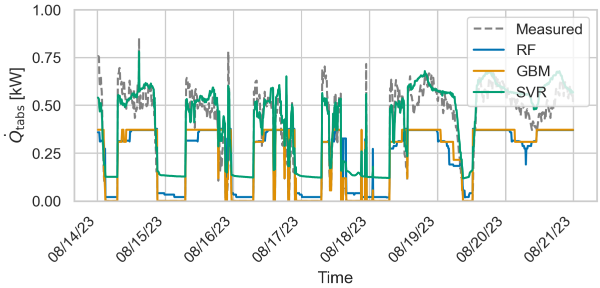

Figure 8: Comparison of predicted and measured TABS thermal power across RF, GBM and SVR models for one week in the test set.

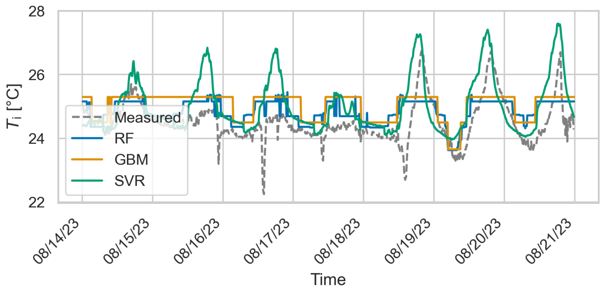

Figure 9: Comparison of predicted and measured indoor air temperature in Office 1 across RF, GBM and SVR models for one week in the test set.

The figures highlight the differences in the algorithms of RF, GBM and SVR models, despite utilising the same features. RF and GBM models yielded more constant outcomes, while the SVR model, characterised by greater variability, performed better at capturing peak values. This was particularly noticeable on the weekend, as indicated in the last three days of the graphs. A plausible explanation for this divergence could be SVR's sensitivity to extreme values (peaks) that arose from its objective to minimise the error within a certain threshold. This inherent sensitivity allowed SVR to better capture and predict extreme values or peaks in the data, leading to its observed variability but effective handling of peaks.

The analysis underlines the importance of model selection based on specific use cases, which factor in considerations such as prediction accuracy, computational efficiency and the ability to manage peak values. While RF and GBM showed higher accuracy and faster execution with the top two features, they did not match SVR's capability in considering peak values. This capability to accurately predict peak values might be crucial for certain applications and emphasises the need to choose models that align with particular performance criteria and application-specific requirements.

Conclusions

This study set out to evaluate how effectively multi-target feature selection methods, through algorithm adaptation, identify influential sensors to enhance building performance predictions. Employing a data-driven approach, the research focused on modelling and predicting the thermal power of the TABS and the indoor air temperature, using a subset of features from the extensive sensor network in the HiLo living lab.

The key findings highlighted that while RF and GBM models prioritised a similar set of features, the SVR model demonstrated a preference for a different subset. Despite these variations, the comparative analysis of the RMSE metric and the computational runtime suggested that models incorporating just two features achieved an optimal balance between accuracy and effectiveness. This balance is particularly crucial in real-time building control applications where both precision and computational speed are essential.

However, the study faced several limitations. First, while the focus was on balancing model complexity and effectiveness, further refinement through hyperparameter tuning or exploring other ML algorithms could enhance performance. Additionally, the inherent differences in how each feature selection method evaluates and ranks feature importance might have impacted the model's ability to generalise across different scenarios.

Despite these limitations, this research underscores the advantages of algorithm adaptation in multi-target feature selection, particularly its ability to leverage inter-feature relationships. This is crucial for complex building systems like TABS, where the interactions between different features significantly impact the system's performance.

Looking ahead, this study opens avenues for future research. A direct extension could utilise the findings to explore the energy flexibility potential of the TABS within HiLo's lightweight concrete structures. This would contribute to a more nuanced understanding of energy flexibility in building systems and its implications for sustainable building operations.

On a broader scale, future research should delve into other multi-target feature selection methods, exploring their strengths and weaknesses in different contexts. Such studies would enrich the toolbox of ML techniques available for building energy management, paving the way for more intelligent control strategies. This would be a significant step toward advancing smart building technologies, which align with global efforts to reduce energy consumption and carbon emissions in the built environment.

Acknowledgments

The data used in this study were collected and made available by the Swiss Federal Laboratories for Materials Science and Technology (Empa) with the support of the Swiss Federal Office of Energy (SFOE) and the Swiss National Science Foundation. It was partly financed by Mitsubishi Electric R&D Centre Europe B.V. [contract numbers EMPA NEST2019 / TT 2019-124, ETH ID No 16234, ETH ID No 20500, ETH ID-No 23035] and the SFOE [contract number SI/502623-01].

References

Arteconi, A., Costola, D., Hoes, P., and Hensen, J. L. M. (2014). Analysis of control strategies for thermally activated building systems under demand side management mechanisms. Energy and Buildings, 80:384–393.

Block, P., Schlueter, A., Veenendaal, D., Bakker, J., Begle, M., Hischier, I., Hofer, J., Jayathissa, P., Maxwell, I., Echenagucia, T. M., Nagy, Z., Pigram, D., Svetozarevic, B., Torsing, R., Verbeek, J., Willmann, A., and Lydon, G. (2017). NEST HiLo: Investigating lightweight construction and adaptive energy systems. Journal of Building Engineering, 12:332–341.

Hashemi, A., Dowlatshahi, M. B., and Nezamabadi-pour, H. (2021). VMFS: A VIKOR-based multi-target feature selection. Expert Systems with Applications, 182:115224.

Kathirgamanathan, A., De Rosa, M., Mangina, E., and Finn, D. P. (2019). Feature Assessment in Data-driven Models for Unlocking Building Energy Flexibility. pages 366–373, Rome, Italy.

Li, R., Satchwell, A. J., Finn, D., Christensen, T. H., Kummert, M., Le Dréau, J., Lopes, R. A., Madsen, H., Salom, J., Henze, G., and Wittchen, K. (2022). Ten questions concerning energy flexibility in buildings. Building and Environment, 223:109461.

Lydon, G. P., Caranovic, S., Hischier, I., and Schlueter, A. (2019). Coupled simulation of thermally active building systems to support a digital twin. Energy and Buildings, 202:109298.

Masmoudi, S., Elghazel, H., Taieb, D., Yazar, O., and Kallel, A. (2020). A machine-learning framework for predicting multiple air pollutants' concentrations via multi-target regression and feature selection. Science of The Total Environment, 715:136991.

Naser, M. Z. and Alavi, A. H. (2023). Error Metrics and Performance Fitness Indicators for Artificial Intelligence and Machine Learning in Engineering and Sciences. Architecture, Structures and Construction, 3(4):499–517.

Olu-Ajayi, R., Alaka, H., Sulaimon, I., Balogun, H., Wusu, G., Yusuf, W., and Adegoke, M. (2023). Building energy performance prediction: A reliability analysis and evaluation of feature selection methods. Expert Systems with Applications, 225:120109.

United Nations Environment Programme (2021). 2021 Global Status Report for Buildings and Construction: Towards a Zero-emission, Efficient and Resilient Buildings and Construction Sector. Technical report, Nairobi, Kenya.

van der Heijde, B., Sourbron, M., Vega Arance, F. J., Salenbien, R., and Helsen, L. (2017). Unlocking flexibility by exploiting the thermal capacity of concrete core activation. Energy Procedia, 135:92–104.

Zhang, L. and Wen, J. (2019). A systematic feature selection procedure for short-term data-driven building energy forecasting model development. Energy and Buildings, 183:428–442.