Component-Based Machine Learning for HVAC Systems Component Modeling

Seyed Azad Nabavi, Ueli Saluz, Sahar Mohammadi, Philipp Geyer
Sustainable Building Systems Group, Leibniz University Hannover, Germany
Azad.nabavi@iek.uni-hannover.de

Abstract. Heating Ventilation and Air Conditioning (HVAC) systems are responsible for a significant portion of building energy consumption, accounting for up to 38% and 12% of global energy consumption. Predicting energy consumption for HVAC systems in the early design phases is important due to their significant impact on energy use and user comfort. However, it is a challenging task due to the complex and dynamic nature of these systems traditionally requiring the effort of building simulation. The main aim of this research is to use machine learning (ML) techniques to model the components of HVAC systems in buildings and to predict the system's performance. We analyze the HVAC components individually to assess the proposed component-based machine learning method's ability to predict their performance and explore their interdependence. The components are structured in two alternative hierarchies to examine alternative modeling approaches: the first hierarchy's order follows the direction of energy flows with the order Z-S-P (zone, secondary HVAC, and primary HVAC components), while the second one follows the logic of design and engineering with the order Z-P-S. A random forest regression algorithm serves as a component ML model. The R^2 value for the CBML model is, respectively, 0.98, 0.99, and 0.99 in forecasting the zone, primary HVAC, and secondary HVAC components in the Z-P-S hierarchy. Hence, the component-based ML method is highly effective in forecasting HVAC system components especially, in the Z-P-S hierarchy. Moreover, in forecasting the secondary HVAC components, the hierarchy following the design and engineering logic shows a significantly higher accuracy for the heat transfer coefficient. The comparison of the prediction accuracy of the CBML method in both hierarchies highlights the critical role of design dependencies in defining such data-driven prediction hierarchies. The primary HVAC component configuration playing a crucial role in modeling secondary HVAC components is a representative example of such a situation.

1. Introduction

Globally, buildings are responsible for almost one-third of final energy consumption (Nejat et al., 2015). Heating, Ventilation, and Air Conditioning (HVAC) systems, which use up a significant amount of energy within buildings, account for 40-60% of the total energy usage (Solano et al., 2021). Moreover, HVAC systems have a great impact on building energy demand and thermal comfort (Afram & Janabi-Sharifi, 2015). The common load forecasting method for building energy modeling, not specifically HVAC systems, is using simulation approaches. However, on the one hand, the building HVAC systems are influenced by numerous factors in the design/operation stage. On the other hand, different degrees of assumptions and simplification in the process of building the model is necessary. Hence, numerous simplifications and assumptions are required for building energy simulations. These design assumptions and simplifications result in significant differences between the forecasting results and the actual situation (Qian et al., 2020).

Machine learning techniques have gained significant importance in energy modeling for HVAC systems. This is due to their ability to handle non-linear and complex problems. Moreover, these methods are independent of assumptions and simplification. These models have proven to be efficient and robust, delivering accurate prediction results in a shorter time frame. As a result, they have gained the attention of designers and stakeholders who can benefit from their quick and reliable prediction results for making informed decisions (Yu et al., 2022). In this regard, Liu et al. (2019) introduced a combination of Autoencoder and Deep Deterministic Policy Gradient (DDPG) algorithm for short-term prediction of HVAC system energy consumption. Moreover, an application of Bayesian Network techniques in selecting the most energy-efficient Primary HVAC (PHVAC) systems based on critical factors was proposed based on a survey (Tian et al., 2019). Furthermore, Woods & Bonnema, (2019) proposed a regression model plug-in in EnergyPlus to estimate the impact of newly designed HVAC systems on building energy performance.

The aforementioned studies highlight the potential of machine learning and data-driven methods in improving the forecasting accuracy and control of HVAC energy consumption in buildings. However, these methods have some limitations such as limited generalizability, interpretability, and explainability due to their black-box nature. Generalizability is critical for machine learning models since they are often applied in the early design phase of non-existing buildings. To improve the generalizability of ML models, a novel approach that combines simulation and transfer learning to enhance the accuracy of heating and cooling load forecasting was suggested by Qian et al. (2020).

Additionally, black box models provide limited guidance for the design process, and therefore, explainability and interpretability are important to understand how the models work and to improve prediction results. Geyer et al. (2021) and Chen et al. (2022) have discussed the importance of explainability, interpretability, and generalizability in machine learning models applied to the prediction of building energy performance. Moreover, the building HVAC systems mostly have been modeled using monolithic black-box models that receive building information and forecast the energy demand of HVAC systems. In other words, the sub-components of HVAC systems have been neglected in modeling HVAC systems. This adds a high level of uncertainty and deteriorates the explainability of developed ML models.

To address the limitations of traditional ML models in HVAC system modeling, we propose a component-based machine learning (CBML) model that creates separate models for individual HVAC components rather than a single model for the entire system. This approach enhances the interpretability and explainability of ML models, allowing designers and engineers to reason with the models more effectively. Additionally, unlike monolithic models, CBML models enable error traceability as each component has its own ML model with performance metrics. While CBML models may experience slight errors due to error propagation, they outperform traditional methods when generalization or sparse data is involved (Chen et al., 2023). Moreover, breaking down the HVAC system into components reduces complexity, improving the transferability of ML models to new cases. Finally, in this study, we developed CBML models for two different hierarchies of information flow among the HVAC components to evaluate the impact of each component on other components.

2. Methods

In this study, a component-based machine learning approach is used to model the HVAC systems components. Accordingly, the HVAC system components are divided into three main components; zone components, primary components, and secondary components. The primary component often referred to as the "plant" converts fuel and electricity to provide heating and cooling to a building through secondary systems. The secondary component typically referred to as a "system" includes distribution systems between the primary system and the building zones as the last component.

The CBML model in this study includes three levels of data-driven modeling for zone component heating demand, secondary HVAC components in each zone, and primary HVAC components for each building. These component models are connected in two hierarchical structures. The zone, secondary HVAC, and primary HVAC components are connected in the first hierarchy (Z-S-P), respectively. In the second hierarchy (Z-P-S), the zone, secondary HVAC, and primary HVAC components are respectively structured.

2.1 Case Studies and System Boundary

For the generation of synthetic training data, an automated pipeline has been implemented that takes parameter ranges as input and accordingly produces simulation files. After the simulation process has finished, a second automated pipeline extracts simulation results and outputs the training data. The parameter ranges cover geometric, energy system, and simulation-related parameters.

For the building geometry a generic H-shape office building with eleven rectangular zones on the normal floors and three service zones in the basement. For the building HVAC systems, three different PHVAC systems were modeled: boiler, central air-source heat pump (ASHP), and district hot water (DHW). All systems are assumed to have convective hot water baseboards for their secondary HVAC system. The hot water loop temperature was 50°C for ASHP and 80°C for the boiler and DHW. The piping system was modeled as adiabatic. The heating setpoint scales linearly with a typical office hour schedule to a new target heating setpoint. Non-working hours (starting from 6 pm), only 75% of the setpoint is set, and starting at 6 am setpoints are increasing hourly to 85%, 95%, and at 9 am to 100%. No cooling system or mechanical ventilation was modeled. The zone ventilation was defined by the air change rate per hour.

2.2 Hierarchy of the proposed method

The hierarchical structures presented in this paper are demonstrated in Figure 1. The Z-S-P structure begins with the zone component, as noticeable in Figure 1-a. In the zone component, a machine learning model receives the zone information and forecasts the Annual Heating Demand (AHD) and Peak Heating Demand (PHD) for each zone. At the secondary HVAC system component, the predicted annual heat demand, the predicted peak heating demand, and the zone information are input to the ML model. This level is composed of two ML models for forecasting the Maximum Design Flow Rate (MDFR) and coefficient of heat transfer (CHT) of the secondary systems, respectively. The U-factor times the area of the building elements is considered the coefficient of heat transfer from that element in this paper. The first ML model forecasts the maximum design flow rate which later is inserted as an input to the second ML model to forecast the coefficient of heat transfer for each secondary HVAC system. Lastly, at the PHVAC system component, the sum of the coefficient of the heat transfer value of all SHVAC systems, the heating setpoint for each building, and PHVAC system types in each building are imported as inputs to the ML model to forecast the design capacity and design outlet temperature of the PHVAC system.

In the Z-P-S hierarchy, as presented in Figure 1-b, the first component modeling is the zone component, similar to the Z-S-P hierarchy. The next component in this hierarchy is the primary HVAC component. The input variables for these components are the predicted results from the zone component models, building orientation, and PHVAC type. In these components, the ML model forecasts design capacity and flow rate for the PHVAC components. The predicted results of PHVAC components are transferred to SHVAC components. In the SHVAC components, the ML model forecasts the design coefficient of heat transfer and design flow rate of the secondary HVAC components.

a) Z-S-P

b) Z-P-S

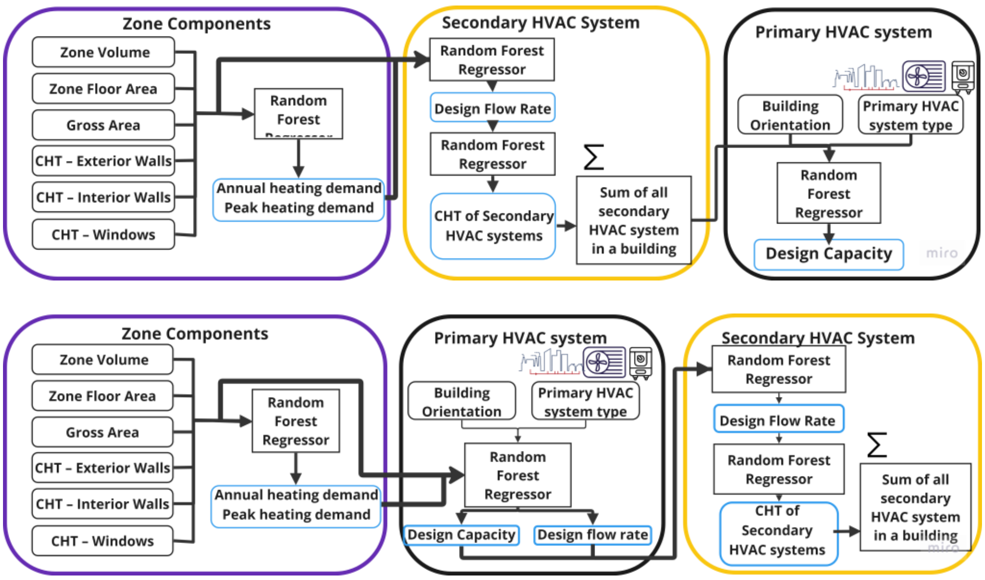


Figure 1 - Hierarchical structure of components and information transfer in modeling.

3. Results

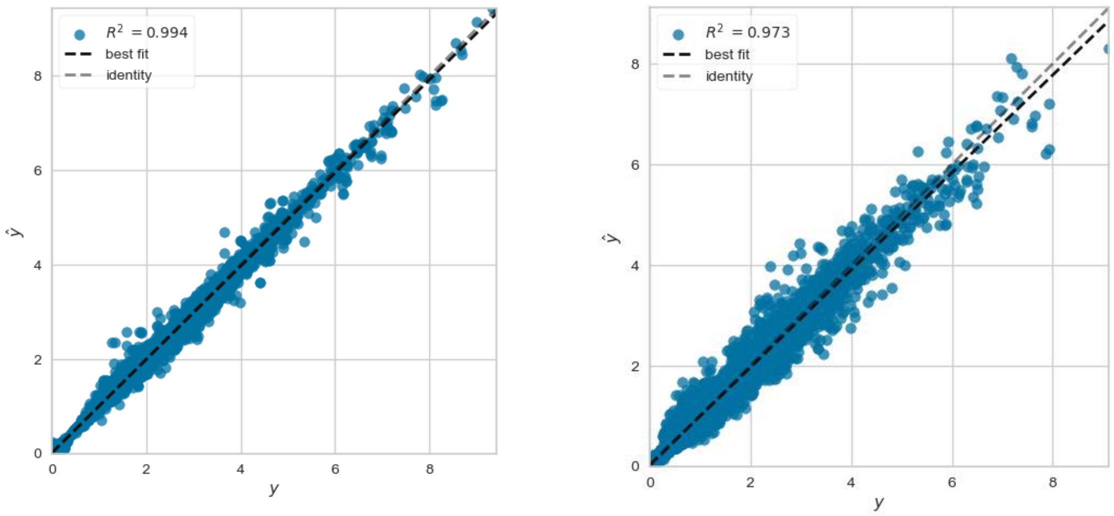
According to our proposed method, we composed three ML models by coupling them with their input and output parameters for forecasting three components of HVAC systems including zones, secondary HVAC systems, and primary HVAC systems. We have structured these three components in two different hierarchies. We used a random forest regression method to model these components. RFR is used to forecast annual heat demand and peak heat demand of each zone in a building. Also, this ML method is used to predict the maximum design flow rate and the coefficient of heat transfer of the SHVAC systems and the capacity and design flow rate of the PHVAC systems. In this paper, we used several performance measures including Root Mean Square Error (RMSE), Mean Squared Error (MSE), Mean Average Percentile Error (MAPE), and R-Squared, to assess how well the proposed methods perform. These performance measures are calculated for each component in both hierarchies and presented in Table 1.

Table 1: Performance metrics components in Z-S-P and Z-P-S hierarchies.

Hierarchy			**Z-P-S**	**Z-S-P**	**Z-P-S**	**Z-S-P**	**Z-P-S**	**Z-S-P**	**Z-P-S**	**Z-S-P**
Component	**Variable**	**UNIT**	**MAE**	**MAE**	**RMSE**	**RMSE**	**MAPE**	**MAPE**	R^2	R^2
Zone Component	PHD	W	666	*666*	1368	*1368*	5.8%	*5.8%*	0.99	*0.99*
	AHD	GJ	6.36	*6.36*	11.8	*11.8*	21%	*21%*	0.97	*0.97*
PHVAC Component	Capacity	W	25180	*18828*	35646	*24450*	1.6%	*1.7%*	0.99	*0.99*
SHVAC Component	CHT	W/K	58.7	*219*	119.7	*562*	3.4%	*15%*	0.99	*0.84*
	MDFR	m^3/s	1.15 e^{-5}	*12.25*	1.97e^{-5}	*37.62*	0.2%	*0.3%*	0.99	*0.99*

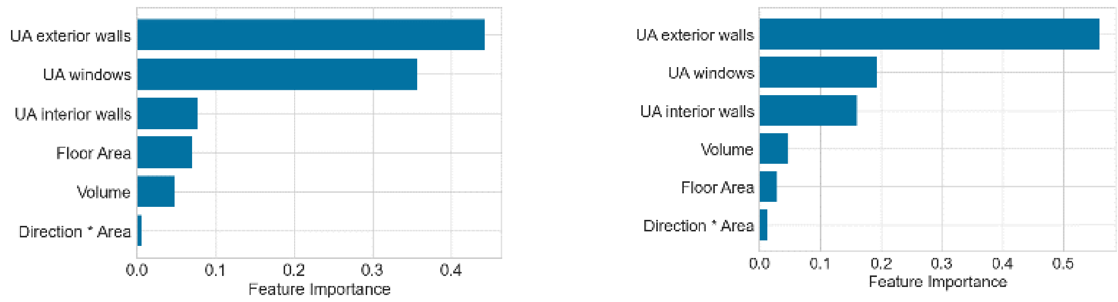
3.1 Zone components

In this study, firstly, we used RFR models to forecast the annual heat demand and peak heat demand of zones. The zone component is similar in both hierarchies as it is the first component of both hierarchies. The prediction error plots of the developed models for forecasting annual and peak heat demand are demonstrated in Figure 6 and Figure 7. As noticeable in these figures, the RFR approach has a high performance in forecasting both annual and peak heat demand. Moreover, in Figure 2, the importance of input variables on the forecasting result using the RFR model is depicted. As noticeable in Figure 2-a and Figure 2-b, the most important feature in forecasting both annual and peak heating demand is the coefficient of heat transfer of exterior walls. The important features in forecasting peak and annual heating demand are respectively, the coefficient of heat transfer of windows, and interior walls of each zone.



d) Peak heating demand prediction *e)* Annual heat demand prediction

Figure 6 – Prediction error plot of peak heat demand prediction



a) Peak heating demand feature importance b) Annual heat demand

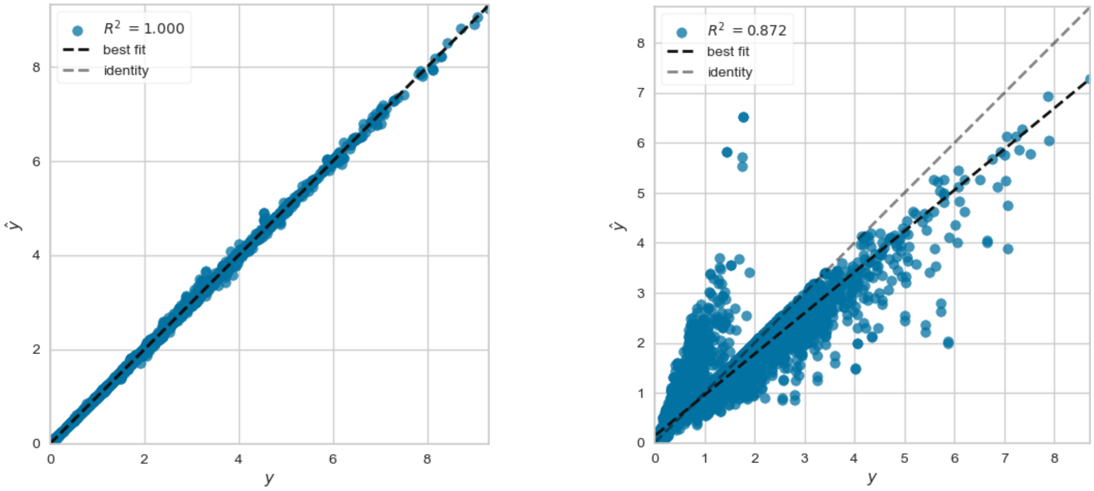
Figure 2 - Feature importance analysis of the input variable in peak heat demand (a) and peak annual demand (b).

3.2 Zone-SHVAC-PHVAC Hierarchy

In the Z-S-P hierarchy, we first use an RFR approach to forecast the maximum design flow rate of the secondary HVAC components using the prediction results of the zone components and zone information. As noticeable in Figure 3-a, the prediction accuracy of RFR in the test set is significantly high. This shows the high capability of ML models in correlating peak heat demand and maximum design flow rate of the secondary HVAC components. In the next step, we predict the coefficient of heat transfer value for each secondary component in the zones. Figure 3-b demonstrates the prediction error of the RFR model in forecasting the coefficient of heat transfer value for the secondary HVAC component.

In Figure 4, the feature importance analysis of the RFR models in forecasting maximum design flow rate and coefficient of heat transfer-SHVAC is demonstrated. As presented in Figure 4-a, the prediction of the maximum design flow rate is highly correlated to the peak heat demand value of the zones. Moreover, the most important feature in forecasting the coefficient of heat transfer values of the secondary HVAC systems, as depicted in Figure 4-b, are maximum design flow rate, coefficient of the heat transfer value of windows, peak heat demand, and annual heat demand features, respectively.

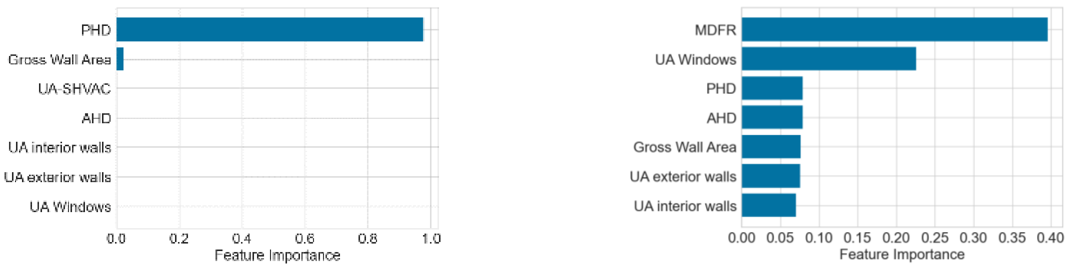
Moreover, the correlation between the coefficient of the heat transfer value of windows and the coefficient of the heat transfer value of SHVAC systems shows the great role of the windows in heat loss and respectively the need for bigger SHVAC systems with higher U-factors.



a) Prediction error plot of maximum design flow rate prediction

b) Prediction error for the coefficient of heat transfer prediction of SHVAC systems

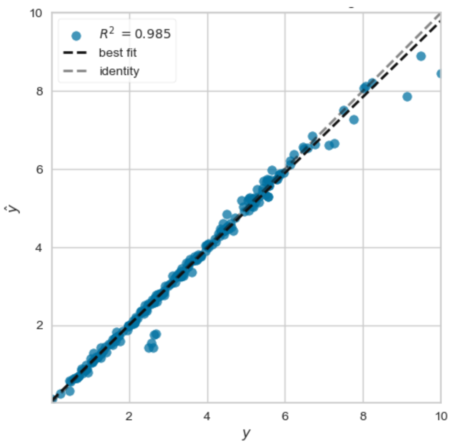
Figure 3 - Prediction error plot of maximum design flow rate (a) and U-factor time area (b) prediction.

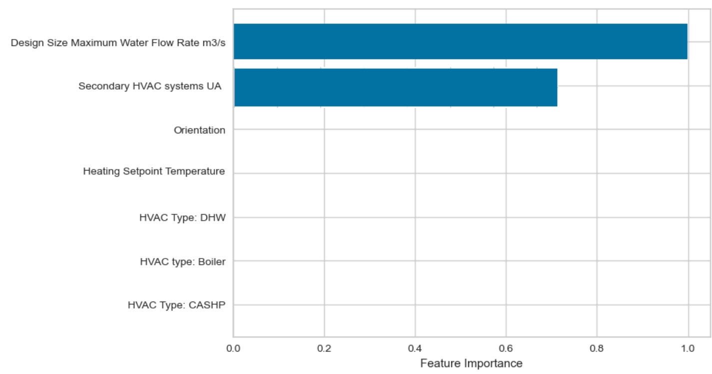


a b

Figure 4 Feature importance analysis of the input variable in maximum design flow rate (a) prediction and coefficient of heat transfer-SHVAC (b) prediction.

In this hierarchy, the design capacity of the PHVAC component as the last component is predicted. The input variables are the sum of the coefficient of heat transfer values for all SHVAC systems, PHVAC system type, heating temperature setpoint, and the orientation of the buildings. In Figure 5, the prediction error and the feature importance of the RFR model in forecasting the PHVAC component's design capacity are illustrated. The RFR model shows high performance in forecasting the capacity of the PHVAC systems as demonstrated in Figure 5-a. As noticeable in this Figure 5-b, the design size maximum flow rate and coefficient of heat transfer of SHVAC systems have the highest importance in forecasting Primary HVAC capacity, respectively. The building orientation and the heating setpoint temperature are the third and fourth important features in forecasting the primary HVAC system's capacity.





a) Prediction error plot of Capacity. b) Feature importance analysis of capacity prediction.

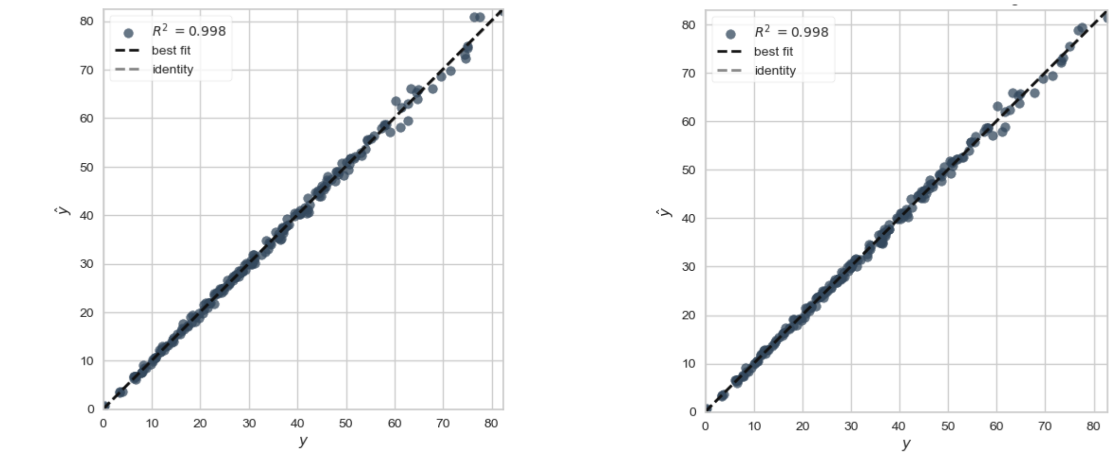
Figure 5 - Prediction error plot of PHVAC prediction results (a) and Feature importance analysis of the input variables in the PHVAC capacity prediction.

3.3 Zone-PHVAC-SHVAC Hierarchy

In this hierarchy, the PHVAC component receives information from the zone component and accordingly forecasts the design capacity and design size maximum flow rate of the PHVAC components. Then, the information from the PHVAC component is used as input for SHVAC component forecasting.

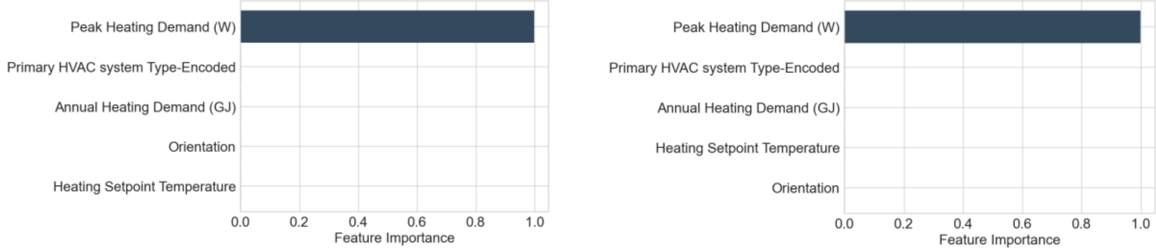
The RFR model forecasts design capacity and design size maximum flow rate using peak heating demand, annual heating demand, primary HVAC system type, building orientation, and heating setpoint temperature as inputs. Figure 6 depicts the prediction error for both output variables of the PHVAC component. In Figure 6-a and Figure 6-b, the high performance of the RFR model in forecasting the design capacity and design size maximum flow rate of the PHVAC component is demonstrated, respectively. In addition, the feature importance analysis of the RFR model for the design capacity and design size flow rate are presented in Figure 7-a and Figure 7-b, respectively. As depicted in these figures, the most important variable in forecasting both design capacity and design size maximum flow rate for PHVAC components.

In the Z-P-S hierarchy, the SHVAC components receive predicted results from PHVAC and zone components information. Accordingly, the coefficient of heat transfer and design size maximum flow rate for SHVAC components are predicted considering peak heating demand, annual heating demand, primary HVAC system type, heating setpoint temperature, building orientation, PHVAC component design capacity, and PHVAC design size maximum flow rate as input variables. The forecasting results show the high performance of the proposed method in predicting the coefficient of heat transfer and design size maximum flow rate. In Figure 8-a and Figure 8-b, the prediction error plot of the developed ML models for the coefficient of heat transfer and design size maximum flow rate of the SHVAC components are presented, respectively. The prediction results of both output variables of the SHVAC components show significantly high accuracy of the proposed method with an R-Squared value of 0.997 and 0.994. Moreover, the feature importance analysis of the proposed ML model, presented in Figure 9 shows the importance of each input variable on the output variables. As presented in Figure 9-a, the peak heating demand, PHVAC system type, and annual heating demand are respectively the most important features in forecasting the SHVAC components' coefficient of heat transfer. In addition, Figure 9-b shows the significant influence of the peak heating demand in forecasting the design size maximum flow rate of the SHVAC components.



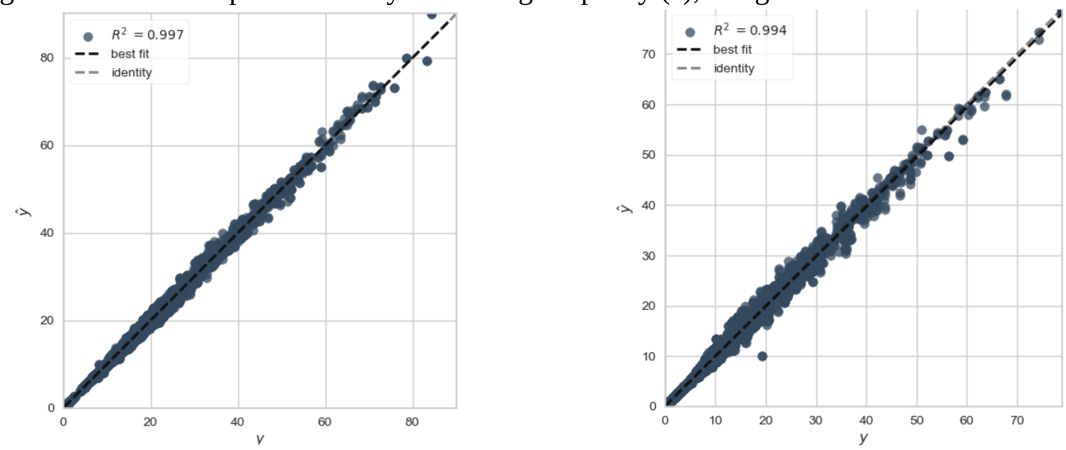
a) Design capacity prediction. b) Design size maximum flow rate prediction.

Figure 6 - Prediction error diagram of design (a) capacity and design size maximum flow rate (b).



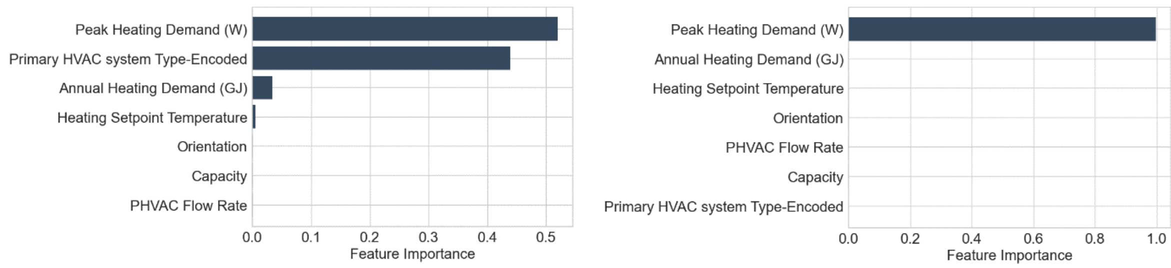
a) Design capacity prediction. b) Design size maximum flow rate prediction.

Figure 7 – Feature importance analysis of design capacity (a), design size maximum flow rate (b).



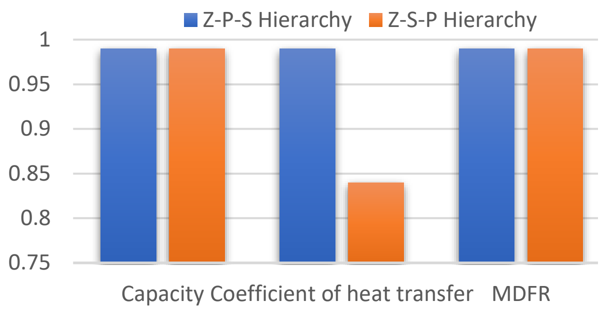
a) Coefficient of heat transfer b) Design size maximum flow rate

Figure 8 - Prediction error in forecasting coefficient of heat transfer (a), design size maximum flow rate (b) of the SHVAC components.

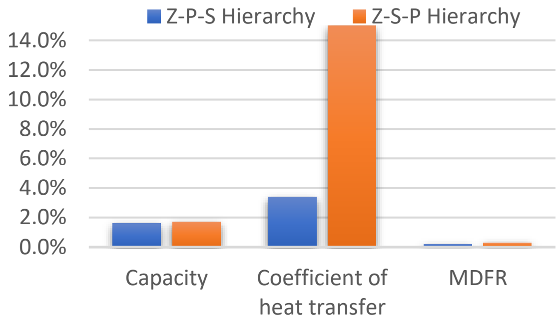


a) Coefficient of heat transfer b) Design size maximum flow rate

Figure 9 - Feature importance analysis of the coefficient of heat transfer (a), design size maximum flow rate (b) of the SHVAC components.



a) R-squared



b) Mean absolute percentile error

Figure 10 – Comparison of the proposed hierarchies in forecasting HVAC components; R^2 (a) MAPE (b)

Moreover, the comparison of the MAPE and R-squared metrics for both hierarchies are presented in Figure 10-a and Figure 10-b, respectively. As depicted in Figure 10, the CBML method has significantly higher accuracy in forecasting the coefficient of heat transfer of secondary HVAC components in the Z-P-S hierarchy compared with the Z-S-P hierarchy.

4. Discussion

In this subsection, we compare the performance of the proposed CBML model in forecasting the HVAC components in both proposed hierarchies. According to the results in Table 1, the Z-P-S hierarchy has significantly higher accuracy than the Z-S-P hierarchy, especially in predicting secondary HVAC components. The estimated R^2 value for the coefficient of heat transfer in secondary HVAC components for Z-P-S and Z-S-P hierarchies is 0.99, and 0.84, respectively. The main reason for this difference in prediction performance is the influence of the primary HVAC system type on the secondary HVAC components. In the Z-S-P hierarchy (Figure 3-b), the prediction results of the coefficient of heat transfer have a group of samples with a high prediction error. This high prediction error is related to a specific group of SHVAC components that have an ASHP as the primary HVAC component. This shows the important role of the primary HVAC system type in forecasting secondary HVAC components. As the information about the primary HVAC components is not used as input to the secondary HVAC components modeling in the Z-S-P hierarchy, the prediction accuracy of the CBML model decreases. However, in the Z-P-S hierarchy, this problem has been solved as the primary HVAC component's features are considered in modeling secondary HVAC components. The reason is a dynamic dependency of the secondary HVAC system on the primary system. The type of the primary system is defined before the specification details of the secondary system are defined. Therefore, this calls for not just following the direction of energy flow in defining such hierarchies but aligning them with design and engineering logic. In the demonstrated case, the prediction error does not have an effect on the capacity prediction and thus the dimensioning of the main system. However, it has a significant impact on the prediction of the heat transfer coefficient in the secondary system and would cause faulty dimensions if the hierarchy following the energy follows would be used. This provides evidence for the need for a careful selection of prediction pipelines in such hierarchies. We expect that not only in this case but in general the consideration of the design and engineering logic according to domain knowledge excels pure physical logic.

5. Conclusion

The main objective of this study is to model HVAC system components in buildings using ML methods. We apply component-based ML to represent the main components of the HVAC system connected by their parameters in a hierarchical way and examined the influence of two alternative component structures. According to the prediction results, the component-based machine learning method shows high performance in forecasting HVAC system components in both hierarchies with only unimportant differences: The R-squared value of the primary HVAC components' capacity prediction is in both cases 0.99, and the MAPE 1.6% respectively 1.7%. The comparison of prediction accuracy of interim hierarchy levels in both hierarchies shows the important impact of primary HVAC components on the secondary HVAC components. The predicted coefficient of heat transfer of the secondary HVAC components shows a significant difference: The R-squared and MAPE metrics are 15%, and 0.84 in the Z-S-P hierarchy and 3.4% and 0.99 in the Z-P-S hierarchy. The case of the importance that primary HVAC component configuration plays for the secondary HVAC components demonstrates the need to follow design and engineering logic instead of pure physical logic.

6. Acknowledgements

We gratefully acknowledge the German Research Foundation's (DFG) support for funding the project under grant GE 1652/3-2 in the Researcher Unit FOR 2363 and under grant GE 1652/4-1 as a Heisenberg professorship.

7. References

Afram, A., & Janabi-Sharifi, F. (2015). Gray-box modeling and validation of residential HVAC system for control system design. *Applied Energy*, *137*, 134–150. https://doi.org/10.1016/j.apenergy.2014.10.026

Chen, X., Guo, T., Kriegel, M., & Geyer, P. (2022). A hybrid-model forecasting framework for reducing the building energy performance gap. *Advanced Engineering Informatics*, *52*. https://doi.org/10.1016/j.aei.2022.101627

Chen, X., Sing, M. M., & Geyer, P. (2023). Utilizing Domain Knowledge: Robust Machine Learning for Building Energy Prediction with Small, Inconsistent Datasets. *ArXiv Preprint ArXiv:2302.10784*.

Geyer, P., Singh, M. M., & Chen, X. (2021). Explainable AI for engineering design: A unified approach of systems engineering and component-based deep learning. *ArXiv Preprint ArXiv:2108.13836*.

Nejat, P., Jomehzadeh, F., Taheri, M. M., Gohari, M., & Majid, M. Z. A. (2015). A global review of energy consumption, CO2 emissions and policy in the residential sector (with an overview of the top ten CO2 emitting countries). *Renewable and Sustainable Energy Reviews*, *43*, 843–862.

Qian, F., Gao, W., Yang, Y., & Yu, D. (2020). Potential analysis of the transfer learning model in short and medium-term forecasting of building HVAC energy consumption. *Energy*, *193*. https://doi.org/10.1016/j.energy.2019.116724

Solano, J. C., Caamaño-Martín, E., Olivieri, L., & Almeida-Galárraga, D. (2021). HVAC systems and thermal comfort in buildings climate control: An experimental case study. *Energy Reports*, *7*, 269–277. https://doi.org/10.1016/j.egyr.2021.06.045

Tian, Z., Si, B., Shi, X., & Fang, Z. (2019). An application of Bayesian Network approach for selecting energy efficient HVAC systems. *Journal of Building Engineering*, *25*. https://doi.org/10.1016/j.jobe.2019.100796

Woods, J., & Bonnema, E. (2019). *Regression-based approach to modeling emerging HVAC technologies in EnergyPlus: A Case study using a Vuilleumier-cycle heat pump*.

Yu, J., Chang, W. S., & Dong, Y. (2022). Building Energy Prediction Models and Related Uncertainties: A Review. In *Buildings* (Vol. 12, Issue 8). MDPI. https://doi.org/10.3390/buildings12081284