2025 European Conference on Computing in Construction

CIB W78 Conference on IT in Construction

Porto, Portugal
July 14-17, 2025

SYMBOLIC REGRESSION FOR COOLING LOAD FORECASTING: ADDRESSING DATA EFFICIENCY AND COLD START CHALLENGES

Mohamed S. Kandil and J.J. McArthur
Dept. Architectural Science, Toronto Metropolitan University, Canada

Abstract

Conventional machine learning models for cooling load prediction require large datasets and lack interpretability, limiting their use in data-scarce and real-time scenarios. This study proposes a symbolic regression-based method to develop compact, transparent models using limited data. Trained on five days of hospital data, the model achieved high accuracy and outperformed linear and gradient boosting benchmarks under out-of-distribution conditions. The approach addresses cold start challenges while enabling deployment on embedded systems. Results demonstrate that symbolic regression offers a robust, interpretable alternative to black-box models, supporting scalable and energy-efficient control in building HVAC systems.

Introduction

Heating, Ventilation, and Air Conditioning (HVAC) systems play a vital role in maintaining thermal comfort and energy efficiency in modern buildings. Among the core components of HVAC systems, cooling load forecasting is critical for ensuring optimal operation, reducing energy consumption, and minimizing operational costs. Accurate cooling load predictions enable effective scheduling, resource allocation, and control strategies, contributing to enhanced performance and sustainability in building management systems.

Traditional approaches to cooling load forecasting rely on physics-based models or machine learning techniques. While physics-based models provide interpretability, they often require extensive domain knowledge and calibration, making them inflexible and computationally intensive. Machine learning and deep learning methods, on the other hand, have shown superior predictive capabilities but are heavily dependent on large datasets. This reliance poses challenges in scenarios with limited data availability, a condition commonly referred to as the cold start problem. Additionally, these data-driven methods often act as black-box models, making their predictions difficult to interpret and limiting their adoption in practical applications where transparency and trust are essential.

To address these limitations, this paper proposes a symbolic regression-based approach for forecasting cooling loads. Symbolic regression generates explicit mathematical expressions that describe the relationships between input features and outputs. This approach not only enhances model interpretability but also reduces dependency on large datasets, making it a suitable solution for data-scarce environments. The ability to generate compact and transparent models allows for seamless integration into energy management systems and supports adaptability to new operating conditions without extensive retraining.

In addition to its interpretability, the proposed framework offers the flexibility to control model complexity, enabling the creation of simplified models that can be deployed on embedded systems. This capability supports decentralized and edge-based deployments, broadening the range of practical applications and facilitating localized decision-making in distributed energy systems.

The model leverages key weather features, such as outdoor air temperature and wet bulb temperature, along with temporal features, including hour of the day and day of the week. Data collected from a hospital in British Columbia, Canada, is used for training and validation, demonstrating the model's effectiveness in delivering accurate predictions, addressing the cold start problem, and ensuring robust performance under varying conditions.

Objectives of the Study

The main objectives of the study are as follows:

1. Develop a data-efficient and interpretable forecasting model for cooling loads in HVAC systems, addressing the cold start problem and minimizing dependency on large datasets.
2. Prioritize model simplicity, supporting applications in embedded systems and decentralized deployments.
3. Validate the proposed framework using real-world data to demonstrate its robustness, practicality, and readiness for integration into real-time control and energy management systems.

This study highlights symbolic regression as a versatile and scalable alternative to traditional machine learning approaches, offering enhanced interpretability, computational efficiency, and suitability for decentralized HVAC optimization.

Literature Review

The accurate forecasting of cooling loads in HVAC systems has garnered significant attention in recent years, given its critical role in optimizing energy efficiency, reducing operational costs, and ensuring system reliability. A range of methodologies has been explored, from traditional regression techniques to advanced machine learning (ML) and deep learning (DL) models. However, these approaches often present trade-offs between accuracy, interpretability, and data efficiency, making them less suitable for certain applications such as projects with limited budgets and resources. This section critically evaluates existing work on load forecasting, emphasizing the limitations of black-box models and the need for data-efficient, interpretable alternatives like symbolic regression.

Approaches to Load Forecasting: Trends and Challenges

Yildiz et al. (2017) reviewed regression and ML models for commercial building electricity load forecasting. They found that while ML methods generally outperformed traditional regression models in accuracy, they required substantial computational resources and large training datasets. Furthermore, the risk of overfitting in ML models, especially in small or imbalanced datasets, was highlighted as a critical limitation. These findings underscore the importance of developing data-efficient methods to ensure robustness and scalability.

Wang et al. (2020) compared shallow ML models, such as XGBoost, with DL approaches like Long Short-Term Memory (LSTM) networks for building thermal load prediction. While both methods achieved high accuracy, the authors noted that DL models are particularly data-hungry, requiring large, high-quality datasets for effective training. This data dependency makes them unsuitable for projects with limited resources, where collecting extensive datasets is challenging. Wang et al. (2020) suggest using the less accurate but simpler heuristic models, which trade accuracy for simplicity and data efficiency, particularly in resource-constrained environments. This aligns with the objective of developing interpretable and computationally efficient models.

Zhang et al. (2021) provided a comprehensive review of ML techniques in building load prediction, highlighting their impressive performance across various forecasting horizons. However, they also identified key limitations, including the complexity of data preprocessing, feature selection, and the opaque nature of many ML models. The lack of interpretability poses a barrier to their adoption in critical applications, where understanding model behavior is essential for decision-making.

Chen, et al. (2023) emphasized the importance of interpretable ML in building energy management. While black-box models like neural networks offer high predictive accuracy, their complexity often makes them unsuitable for applications requiring transparency and trust. The authors argue for approaches that balance accuracy with interpretability, advocating for models that provide explicit relationships between input features and outputs. Although interpretable ML improves trust in the application of ML for building energy management, the research is still in its infancy and faces significant challenges. Current ad hoc techniques like SHAP and LIME have limitations. Therefore, a symbolic regression framework is proposed as an effective and robust solution to fill this gap.

Gao et al. (2022) proposed a hybrid method that combines extreme learning machines (ELM) with random forest (RF) for feature selection and the improved parallel whale optimization algorithm for parameter tuning. The authors claim that this approach outperforms other methods in terms of accuracy and computational efficiency. However, the model's complexity and reliance on specific configurations make it less applicable for practical applications, particularly in scenarios requiring simplicity and adaptability.

Lu et al. (2022) examined the use of artificial neural networks (ANNs) in building energy prediction, noting their ability to model complex nonlinear relationships. However, the authors acknowledged that ANNs often act as black boxes, with limited interpretability and significant data requirements. This lack of transparency is a major drawback in HVAC applications, where clear insights into system behavior are essential for effective control and optimization.

Myat et al. (2024) introduced a hybrid model combining multivariate fast iterative filtering (FIF) with LSTM networks for ultra-short-term cooling load prediction. While this approach improved predictive accuracy by capturing both linear and nonlinear patterns, the computational complexity of LSTM models and their reliance on extensive datasets remain significant challenges. The trade-offs between accuracy and simplicity are evident, particularly in resource-constrained environments.

Symbolic Regression as an Alternative

Symbolic regression has emerged as a promising alternative to traditional ML and DL models for load forecasting, offering a balance between accuracy, interpretability, and data efficiency. Unlike black-box models, symbolic regression generates explicit mathematical expressions that describe the relationships between inputs and outputs, making it inherently transparent and suitable for decentralized applications.

Quade et al. (2016) explored symbolic regression for predicting dynamical systems, demonstrating its ability to handle nonlinear relationships and produce compact, interpretable models. The authors emphasized its computational efficiency and scalability, supporting its application in real-time energy management systems.

Ozawa et al. (2023) demonstrated the use of symbolic regression for HVAC control, showing that it can produce interpretable models without the need for extensive datasets. The study highlighted its adaptability to varying conditions, making it ideal for projects with limited data availability or computational resources.

Yousaf et al. (2024) presented a gray-box modeling approach for unitary air conditioners using symbolic regression. By combining physical insights with data-driven techniques, the authors achieved a balance between accuracy and simplicity. Importantly, their models were compact enough for deployment on embedded systems, aligning with the objective of creating computationally efficient solutions for decentralized HVAC control.

Summary and Gaps in the Literature

Existing approaches to load forecasting demonstrate significant advancements in accuracy and capability. However, they are often limited by their reliance on large datasets, lack of interpretability, and high computational demands. Black-box models, such as DL and hybrid techniques, excel in predictive performance but are resource-intensive and opaque. In contrast, symbolic regression offers a compelling alternative, addressing the cold start problem and enabling the development of simple, interpretable models suitable for embedded systems and decentralized applications.

This study builds on the strengths of symbolic regression, leveraging its data efficiency and transparency to address the limitations of existing approaches. By validating the proposed model using real-world data, this work contributes to advancing load forecasting methodologies, emphasizing computational efficiency, model simplicity, and practical applicability in HVAC systems.

Methodology

Context and Data Acquisition

This study focuses on a hospital in the Greater Vancouver area, located in ASHRAE climate zone 4. The hospital's central cooling plant, equipped with chillers and cooling towers, supports its critical medical services, including trauma, cardiac care, and neonatal intensive care. The plant includes two chiller units, each with a capacity of 4396 kW (1250 tons). However, these chillers do not operate simultaneously; one serves as a backup while the other operates, with their roles rotating to ensure reliability and maintenance efficiency. Reliable cooling is essential for maintaining stringent environmental conditions, making this facility an ideal case for assessing cooling load forecasting.

Data were gathered from the hospital's cooling system over three months, spanning July to September 2024. The dataset included key variables such as chilled water supply and return temperatures, water flow rates, and weather conditions, including dry-bulb outdoor air temperature and relative humidity. Sampling was conducted at 15-minute intervals to capture the system's dynamic behavior effectively.

Data Preparation and Feature Engineering

To prepare the dataset for modeling, preprocessing and feature engineering steps were applied. Data anomalies and missing values were systematically removed, yielding a final dataset of high-quality records.

Temporal features, such as hour of the day, day of the week, and a binary weekend indicator, were added to account for time-dependent variations in cooling demand. The wet-bulb temperature was calculated using an empirical expression developed by (Stull, 2011), which integrates relative humidity and dry-bulb temperature. This formula provides a reliable approximation of the wet-bulb temperature, capturing critical thermodynamic conditions influencing cooling system performance. Additionally, the cooling load (Q, in kW) was calculated using Eq. 1:

$$Q = 4.186 * \dot{V} * (T_{chwr} - T_{chws}) \quad (1)$$

where $\dot{V}$ is the water flow rate (L/s), T_{chwr} and T_{chws} represents chilled water return and supply temperatures respectively in degrees Celsius. Figure 1 shows cooling load data recorded from 8th July to 15th July 2024.

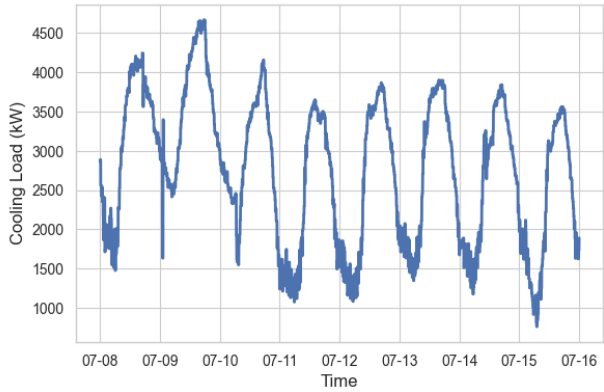

Figure 1 Cooling load data recorded from 8th July to 15th July 2024

Symbolic Regression Modeling

Symbolic regression was conducted using *PySR* (Cranmer, 2023), a Python library designed to leverage genetic algorithms for identifying mathematical patterns in data. Unlike traditional regression techniques, symbolic regression eliminates the need for predefined model architectures, enabling the discovery of both interpretable and accurate equations directly from the data.

PySR integrates Julia as its computational backend, facilitating efficient and comprehensive exploration of the search space. The study utilized a set of core mathematical operators—addition (+), subtraction (-), multiplication

(*), and division (/). To guide the optimization process, the mean squared error (L2DistLoss) was employed as the loss function, ensuring a balance between model precision and interpretability.

Training and Model Validation

For model training, five days of data from July 8th to July 12th was selected to represent a broad range of operational scenarios. The data were split into training (80%) and testing (20%) sets, preserving the chronological order to respect the time-series nature of the data. The cooling load was designated as the target variable, while the remaining variables served as input features.

The symbolic regression model was trained to identify the mathematical relationship between the input variables: dry-bulb outdoor air temperature (OAT_{dry}), wet-bulb temperature (OAT_{wet}), hour of the day (H), day of the week (D), and the target cooling load (Q).
Model performance was evaluated using multiple metrics: R-squared (R^2) to quantify the proportion of variance explained by the model, Root Mean Squared Error (RMSE) for absolute prediction error, Coefficient of Variation of RMSE (CVRMSE) for normalized error, and Mean Absolute Percentage Error (MAPE) for relative accuracy.

Validation involved testing the model on a separate dataset consisting of 96 hours of cooling load data recorded from July 17th to July 20th, 2024. This independent evaluation ensured the model's accuracy and generalizability under new conditions, which is critical for real-world deployment.

Benchmarking and Robustness Assessment

To evaluate the robustness and relative performance of the proposed symbolic regression model, two widely used machine learning algorithms were selected as benchmarking baselines: Linear Regression (LR) and Extreme Gradient Boosting (XGBoost). These models were chosen to represent two ends of the interpretability-performance spectrum — with LR offering a simple, interpretable gray-box model, and XGBoost serving as a high-performing black-box model.

Both LR and XGBoost models were implemented using the Python libraries *scikit-learn* and *xgboost*, respectively. To ensure a fair comparison, these models were trained using the same pre-processed dataset and feature set employed for the symbolic regression models. This included input variables such as outdoor air temperature, wet-bulb temperature, hour of the day, and day of the week. The training period was limited to five consecutive days of data (from July 8th to July 12th, 2024), consistent with the symbolic regression training protocol.

To assess model robustness under out-of-distribution (OOD) conditions, a distinct validation dataset was selected spanning September 1st to 15th, 2024. This period was chosen to introduce a temporal and seasonal gap from the training window (July 8th to 12th, 2024), representing a shift in environmental conditions, operational patterns, and building dynamics. By validating all models — symbolic regression, linear regression, and XGBoost — on this separate dataset, we simulate real-world deployment scenarios where models are expected to generalize beyond the specific conditions seen during training.

Model evaluation was conducted using four key performance metrics to ensure a comprehensive assessment of accuracy and generalization. The Coefficient of Determination (R^2) was used to quantify the proportion of variance in the target variable explained by the model. Root Mean Squared Error (RMSE) measured the average magnitude of prediction error in absolute terms. To account for differences in load magnitude, the Coefficient of Variation of RMSE (CVRMSE) was calculated by normalizing RMSE against the mean observed load. Finally, the Mean Absolute Percentage Error (MAPE) provided a scale-independent measure of relative prediction accuracy, capturing average percentage deviations between predicted and actual values.

This benchmarking setup enables a comprehensive comparison of model accuracy, interpretability, and robustness under real-world, data-limited scenarios, and highlights the trade-offs between model complexity and generalization. The results of this comparative analysis are discussed in the following section.

Results

Performance of Symbolic Regression Models

The symbolic regression modeling process using *PySR* produced multiple candidate expressions for predicting the cooling load. Among these, the two key models (A and B) were selected for their performance and interpretability.

Model A was the best-performing model, as per the selected performance metrics, and is based on Eq. (2):

$$Q = OAT_{dry} + OAT_{wet}\left(252.20 + \frac{1.68}{6.95 - H}\right) + D(1.81H + 0.75) + D - \left(\frac{38649.35}{OAT_{dry} - 0.83}\right) \quad (2)$$

This model achieved the highest accuracy with a value of 0.9239, an RMSE of 217.55, a CVRMSE of 7.24%, and a MAPE of 6.59% on the validation dataset. It offers a detailed representation of how the cooling load depends on weather variables (outdoor air temperature and wet-bulb temperature) and temporal features (hour and day of the week). The inclusion of these parameters ensures the model can capture the complex interactions between environmental and operational conditions.

The package, however, recommends the simpler expression shown in Equation (3), referred to herein as Model B:

$$Q = 255.52\, T_{wet} + H - \frac{40525.42}{OAT_{dry}} \quad (3)$$

This recommendation is based on the “model_selection” parameter set to best, which prioritizes simplicity and interpretability by choosing models that achieve high scores while maintaining a loss within 1.5 times that of the most accurate model. Despite using an approximation of Equation (2), Model B’s Equation (3) maintains high performance with a value of 0.9239, an RMSE of 232.91, a CVRMSE of 7.75%, and a MAPE of 7.11%.

Model A slightly outperforms Model B across all metrics, making it the preferred choice when accuracy is paramount. However, the computational simplicity and interpretability of Equation (3) make it suitable for real-time applications or scenarios with limited computational resources. The performance of the recommended model (B) on the validation dataset is illustrated through two key visualizations:

Figure 2 depicts the predicted cooling load alongside the actual load over the validation period. The predicted load closely follows the actual load, demonstrating the model’s ability to capture temporal variations in cooling demand effectively. The periodic patterns of cooling load fluctuations are accurately replicated, highlighting the model’s robustness in accounting for both weather-dependent and temporal features.

Figure 3 presents a scatter plot comparing the predicted and actual cooling loads. The data points align closely with the 45-degree line, indicating a strong correlation between the predicted and observed values. This high degree of alignment reinforces the model’s accuracy and reliability in estimating the cooling load under diverse operating conditions.

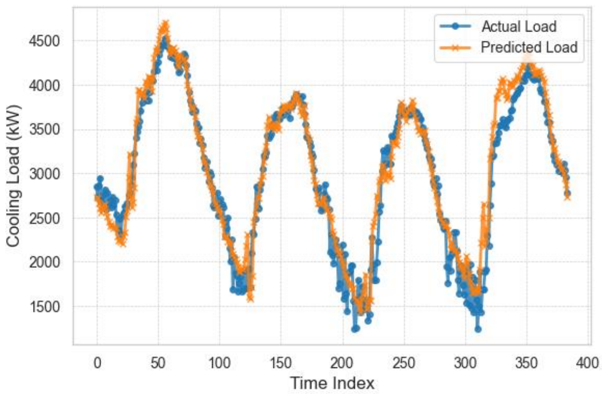

Figure 2 Timeseries representation of actual load alongside predicted values on the validation dataset for model B

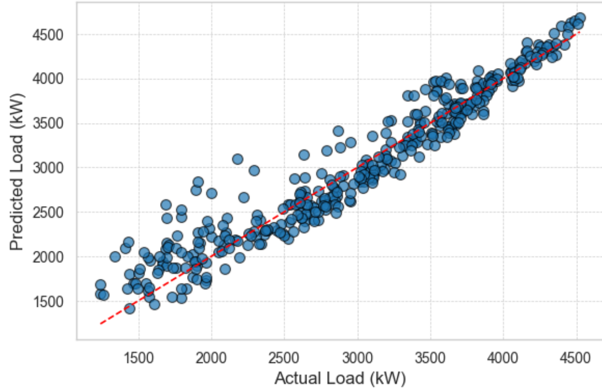

Figure 3 Scatter plot actual cooling load vs predicted values on the validation dataset for model B

Impact of Training Dataset Size

A critical aspect of this study is the development of data-efficient models for cooling load prediction. To evaluate the impact of training dataset size, a model trained on 3 days of data (model C) was compared to the recommended model (model B) trained on 5 days, with identical hyperparameters ensuring consistency. The symbolic regression model derived from the 3-day dataset is given as:

$$Q = 400\, T_{wb} - 4380 \quad (4)$$

Visual analysis provides further insight into the models' performance differences. Figure 4 illustrates the predicted vs. actual cooling loads for both models over time. Both models capture the overall trend of cooling load variations, but the 5-day model exhibits closer alignment with the actual data, especially during peak load periods. Figure 5 presents a scatter plot comparing predicted and actual cooling loads, where the 5-day model shows a denser clustering along the 45-degree line, reflecting improved accuracy and reliability. Figure 6 visualizes the performance metrics for both models, clearly demonstrating the performance gains achieved with additional training data.

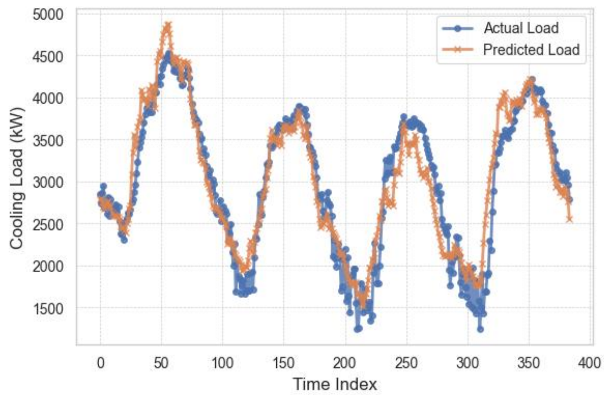

Figure 4 Timeseries representation of actual load alongside predicted values on the validation dataset for model C

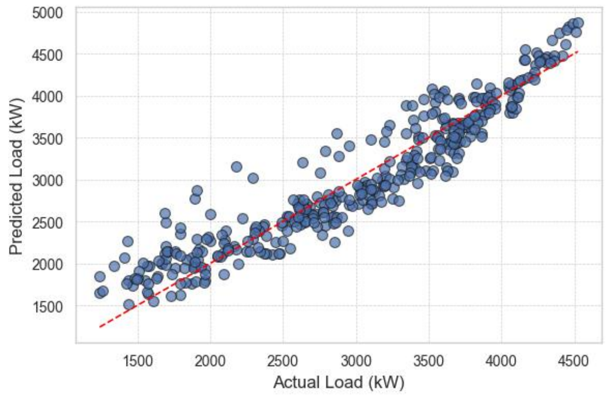

Figure 5 Scatter plot actual cooling load vs predicted values on the validation dataset for model C

The model trained on 3 days of data achieved reasonable performance, with an R^2 value of 0.89, indicating a strong correlation between predicted and actual cooling loads. However, error metrics highlight the limitations of using a smaller dataset, with an RMSE of 280.74 kW, a CVRMSE of 9.34%, and a MAPE of 8.62%. These results suggest that while the 3-day model is sufficient for general predictions, it is less precise compared to the 5-day model.

These visualizations provide compelling evidence of the recommended model's capability to generalize well on unseen data, confirming its suitability for practical deployment in real-world HVAC systems.

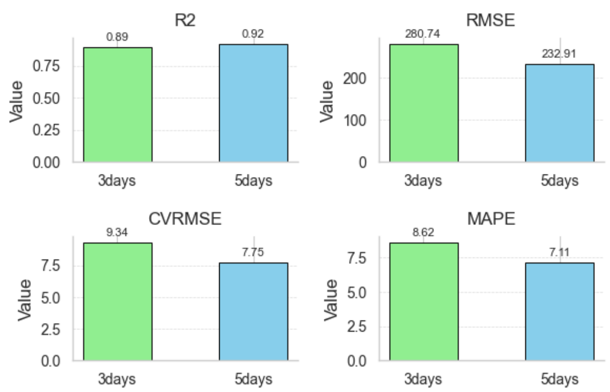

Figure 6 Comparison of evaluation metrics between Model B (5 days) and Model C (3 days)

Benchmarking and Robustness Results

To evaluate the performance of the symbolic regression models against common machine learning baselines, two benchmarking models were introduced: Linear Regression (LR) and Extreme Gradient Boosting (XGB). LR represents a simple, interpretable model (gray-box), while XGB is a high-capacity black-box model known for strong predictive performance. All models were trained on the same five-day dataset (July 8–12, 2024) using identical input features and preprocessing steps to ensure a fair comparison.

The first evaluation was conducted using a four-day validation set from July 17–20, 2024, which closely matches the distribution of the training data. Table 2 summarizes the results. As expected, XGB achieved the best overall performance with an R^2 of 0.94, slightly outperforming both symbolic regression models. Model A (SR_A) yielded strong performance ($R^2 = 0.92$, MAPE = 6.56%) and was closely followed by the simpler Model B (SR_B), which prioritized interpretability. LR showed lower performance across all metrics, though still within an acceptable range.

Table 1 Model performance on validation set (July 17–20, 2024)

Model	R^2	RMSE	CVRMSE	MAPE
SR_A	0.92	216.33	7.20	6.56
SR_B	0.92	232.91	7.75	7.11
LR	0.89	282.13	9.39	8.65
XGB	0.94	203.57	6.77	6.17

To assess generalization under shifting conditions, all models were evaluated on a separate out-of-distribution (OOD) test set spanning September 1–15, 2024. This period differs from the training window both temporally and seasonally, representing a distinct operational regime. Table 3 presents the results.

While symbolic regression models experienced a drop in performance, they retained significantly stronger accuracy than both benchmarking models. SR_B achieved an R^2 of 0.83 with a MAPE of 44.80%, while SR_A produced similar results. In contrast, XGB and LR showed severe degradation, with MAPE values exceeding 79% and R^2 dropping below 0.67. Notably, XGB, despite being the best model on the standard validation set, was the worst performer under OOD conditions.

Table 2 Model performance on out-of-distribution (OOD) test set (September 1–15, 2024)

Model	R^2	RMSE	CVRMSE	MAPE
SR_A	0.83	476.43	26.26	44.12
SR_B	0.83	473.55	26.10	44.80
LR	0.66	680.59	37.51	79.09
XGB	0.63	707.79	39.01	81.29

These results highlight the strong robustness and generalization capability of symbolic regression models, particularly under data distribution shifts, reinforcing their suitability for real-world deployment.

Discussion

This study set out to develop a data-efficient, interpretable, and deployable framework for forecasting cooling loads in HVAC systems using symbolic regression (SR). The results from real-world experiments confirm that the proposed SR models not only achieve competitive predictive performance but also satisfy the core objectives of interpretability, robustness, and practical integration into real-time energy management systems. The following discussion outlines the key findings and implications of the study.

Model Performance and Interpretability

The symbolic regression models produced using PySR successfully balance accuracy and simplicity. Model A, which includes more intricate relationships between variables, delivers the highest accuracy across all evaluation metrics on the standard validation set. In contrast, Model B provides a simplified version with marginally lower performance but improved interpretability and computational efficiency. This trade-off aligns with real-world requirements where resource constraints and system transparency are critical, such as in embedded systems or on-device forecasting modules.

Dependency on Weather Conditions

Both SR models underscore the critical role of weather conditions, particularly outdoor air temperature and wet-bulb temperature, in determining the cooling load. These findings emphasize the need for high-fidelity weather forecasting data to achieve reliable predictions, especially in climates with variable conditions.

Training Data Size and Data Efficiency

Symbolic regression demonstrated strong performance with minimal data. The comparison between Model B (trained on five days of data) and Model C (trained on three days) highlights the benefits of using more diverse data. Although Model C provided a reasonable baseline, the 5-day model achieved substantial improvements—a 17% decrease in RMSE and an 18% improvement in MAPE—illustrating the method's sensitivity to modest increases in training data coverage. Nonetheless, the fact that accurate models can be developed from as little as 3–5 days of data directly addresses the cold start problem, achieving the first objective of minimizing dependency on large datasets.

Benchmarking Against Machine Learning Models

To further validate the SR models, they were benchmarked against two common machine learning approaches: Linear Regression (LR) and Extreme Gradient Boosting (XGB). Under standard validation conditions, XGB slightly outperformed SR models in accuracy; however, this came at the cost of model transparency and complexity. SR_B, the recommended symbolic model, achieved performance within 1% of XGB in MAPE, demonstrating that interpretability does not require sacrificing predictive power.

LR, while simple and interpretable, consistently underperformed relative to both SR and XGB models. These results show that symbolic regression can outperform traditional gray-box methods and approach black-box performance, offering a compelling balance for HVAC forecasting.

Robustness Under Distributional Shift

The most significant finding emerged during testing on the out-of-distribution (OOD) dataset, covering a two-week period that differed seasonally and operationally from the training window. Under these conditions, both SR models maintained strong generalization ($R^2 = 0.83$), while XGB and LR experienced substantial degradation, with MAPE values exceeding 79% and R^2 falling below 0.67. This indicates that symbolic regression is not only interpretable and data-efficient but also more robust when exposed to unseen operational conditions. These attributes make it especially suitable for deployment in dynamic environments where models must adapt without frequent retraining.

Practical Implications and Broader Impact

Symbolic regression offers a unique position within the modeling landscape. Unlike black-box models that often require post-hoc explainability tools, symbolic models generate transparent, human-readable equations by design. These can be directly embedded in control systems, audited for compliance, and tuned by domain experts—without the need for deep learning infrastructure or inference engines. Moreover, the robustness under limited training and OOD conditions suggests strong potential for symbolic regression in cold-start scenarios, decentralized control, and low-computation environments. The proposed framework is well aligned with industry needs for trustworthy, efficient, and deployable energy forecasting solutions.

Conclusions

This study demonstrated the effectiveness of symbolic regression, using PySR, in developing interpretable, data-efficient models for cooling load prediction in HVAC systems. Trained on only five days of data, the models achieved strong performance and generalization, addressing cold start challenges and minimizing dependency on large datasets.

Two symbolic models were developed: one prioritizing accuracy (SR_A) and another emphasizing simplicity and computational efficiency (SR_B). Both models outperformed traditional benchmarks under out-of-distribution conditions, confirming their robustness and practical utility. Compared to black-box models, symbolic regression offered comparable accuracy with the added benefits of transparency and ease of deployment. These results confirm that symbolic regression is a viable and scalable alternative for HVAC forecasting, particularly in applications requiring real-time control, interpretability, and adaptability. Future work should explore broader deployment across building types and integration into adaptive control frameworks.

Acknowledgments

This research was funded by the Natural Science and Engineering Research Council (Canada) [ALLRP-

580958-22] and H. H. Angus and Associates Consulting Engineers Ltd.

Note on the use of Generative AI

GAI was used by the lead author to proofread the final document. No original text was GAI-generated.

References

Chen, Z., Xiao, F., Guo, F. & Yan, J., 2023. Interpretable machine learning for building energy management: A state-of-the-art review. *Advances in Applied Energy,* Volume 9, p. 100123.

Cranmer, M., 2023. Interpretable Machine Learning for Science with PySR and SymbolicRegression.jl. *arXiv preprint arXiv:2305.01582.*

Gao, Z. et al., 2022. A hybrid method of cooling load forecasting for large commercial building based on extreme learning machine. *Energy,* Volume 238, Part C, p. 122073.

Lu, C., Li, S. & Lu, Z., 2022. Building energy prediction using artificial neural networks: A literature survey. *Energy and Buildings,* Volume 262, p. 111718.

Myat, A., Kondath, N., Soh, Y. L. & Hui, A., 2024. A hybrid model based on multivariate fast iterative filtering and long short-term memory for ultra-short-term cooling load prediction. *Energy and Buildings,* Volume 307, p. 113977.

Ozawa, Y. et al., 2023. *Data-driven HVAC Control Using Symbolic Regression: Design and Implementation.* Orlando, FL, USA, s.n.

Quade, M. et al., 2016. Prediction of dynamical systems by symbolic regression. *Physical Review E,* Volume 94, p. 012214 .

Stull, R., 2011. Wet-Bulb Temperature from Relative Humidity and Air Temperature. *Journal of Applied Meteorology and Climatology,* 50(11), p. 2267–2269.

Wang, Z., Hong, T. & Piette, M. A., 2020. Building thermal load prediction through shallow machine learning and deep learning. *Applied Energy,* Volume 263, p. 114683.

Yildiz, B., Bilbao, J. I. & Sproul, A. B., 2017. A review and analysis of regression and machine learning models on commercial building electricity load forecasting. *Renewable and Sustainable Energy Reviews,* Volume 73, pp. 1104-1122.

Yousaf, S., Bradshaw, C. R., Kamalapurkar, R. & San, O., 2024. A gray-box model for unitary air conditioners developed with symbolic regression. *International Journal of Refrigeration,* Volume 168, pp. 696-707.

Zhang, L. et al., 2021. A review of machine learning in building load prediction. *Applied Energy,* Volume 285, p. 116452.