Scalable Building Temperature Prediction for Smart HVAC Control: A Multi-Stage Learning Framework

Rohan Saha [*1] **Tushar Shinde** [*1]

Abstract

Commercial buildings contribute 17% to U.S. carbon emissions, with Heating, Ventilation, and Air Conditioning (HVAC) systems accounting for most energy consumption. We propose a novel multi-stage learning framework for accurate and scalable temperature prediction in smart buildings. Our approach systematically scales from single-zone, single-day forecasts to 123-zone, multi-week predictions, achieving a mean absolute error (MAE) of 0.195°F for single-zone tasks using XGBoost. Our framework advances energy-efficient HVAC control, reducing carbon footprints in commercial buildings.

1. Introduction

Buildings are responsible for 37% of U.S. carbon emissions, with commercial buildings contributing 17% in 2023, as reported by the U.S. Energy Information Administration. Heating, Ventilation, and Air Conditioning (HVAC) systems account for 40–60% of building energy use, making them a critical target for sustainability efforts (Pérez-Lombard et al., 2008). The Smart Buildings Control Suite (Goldfeder et al., 2024) provides a comprehensive open-source benchmark, including real-world data from 11 buildings over six years, lightweight data-driven simulators, and Physically Informed Neural Network (PINN) models. This benchmark addresses the scalability challenges of prior HVAC optimization efforts, which often rely on proprietary data or high-fidelity simulators that are difficult to configure. By offering diverse, real-world telemetric data and scalable modeling tools, the suite enables robust evaluation of control policies across varied climates, building sizes, and management systems.

Traditional HVAC control methods, such as fixed setpoint configurations, are often suboptimal due to their inability to adapt dynamically to varying conditions. Advanced approaches like Model Predictive Control (MPC) and Reinforcement Learning (RL) show promise but are limited by computational complexity and lack of generalizability across buildings (Taheri et al., 2022). Moreover, existing benchmarks often focus on single-zone or short-term predictions, failing to address the multi-zone, long-term forecasting needed for real-world deployment (Luo et al., 2022).

We propose a multi-stage learning framework that systematically scales temperature prediction from single-zone, single-day tasks to 123-zone, multi-week forecasts. Leveraging the Smart Buildings dataset, our approach uses XGBoost to achieve high accuracy (MAE=0.195°F for single-zone predictions) while maintaining computational efficiency.

Contributions:

- A novel four-stage learning framework that scales prediction tasks from single-zone to 123-zone scenarios over extended temporal horizons, addressing scalability challenges in HVAC control.
- A comprehensive evaluation of XGBoost with tailored feature engineering, achieving superior accuracy and efficiency for temperature forecasting.
- Scalability and robustness analysis using 2.5 years of real-world data.

2. Related Work

HVAC control optimization has been extensively studied, with approaches including rule-based systems, MPC, and RL (Fong et al., 2006; Taheri et al., 2022). RL methods, such as Soft Actor-Critic (SAC), have shown promise in optimizing control policies, achieving an 8% improvement over baseline policies in the Smart Buildings Control Suite (Goldfeder et al., 2024). However, RL often requires significant computational resources and struggles to generalize across diverse building configurations due to complex state-action spaces (Yu et al., 2021).

Neural network models, such as Long Short-Term Memory (LSTM) networks, are commonly employed to capture temporal dependencies in building dynamics. However, in multi-zone scenarios, these models face challenges due

[*]Equal contribution [1]IIT Madras Zanzibar, Tanzania. Correspondence to: Tushar Shinde <shinde@iitmz.ac.in>.

Proceedings of the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

to high computational costs and a propensity for overfitting (Sendra-Arranz & Gutiérrez, 2020).

Physically Informed Neural Networks (PINNs) have emerged as a promising approach that incorporates physical priors to enhance model generalization. In the context of smart buildings, PINNs have shown improvements in forecasting building energy consumption and system performance (Jiang & Dong, 2024; Chen et al., 2023). However, designing PINNs that strike an optimal balance between model complexity and accuracy remains a significant challenge, particularly for real-time applications on edge devices.

Gradient boosting methods like XGBoost offer a computationally efficient alternative, excelling in modeling non-linear thermal dynamics with tabular data (Chen & Guestrin, 2016). Unlike LSTMs, XGBoost scales well to multi-zone scenarios and requires less training time, making it suitable for real-time applications. Prior studies, however, often focus on single-zone or short-term predictions, limiting their applicability to large-scale commercial buildings (Zou et al., 2020).

The Smart Buildings Control Suite (Goldfeder et al., 2024) addresses the lack of public, diverse datasets by providing real-world data from 11 buildings, lightweight simulators, and PINN models. Unlike previous datasets focused on residential buildings or non-HVAC metrics (Murray et al., 2017; Miller et al., 2020), this suite enables scalable, interactive RL environments. Our framework builds on this benchmark, introducing a multi-stage approach to systematically address spatial and temporal scalability.

3. Method

3.1. Problem Formulation

We frame building temperature prediction as a component of a Markov Decision Process (MDP), defined by the tuple (S, A, p, R). The state S_t at time t is a vector of sensor measurements (e.g., zone air temperatures, occupancy), the action A_t comprises HVAC setpoints (e.g., supply air temperature), p is the transition probability, and R is the reward function. For prediction, we focus on estimating future states $\hat{T}_{t,z}$ for zone z:

$$\hat{T}_{t,z} = f(X_{t-\tau:t}, U_t, W_t, \theta), \quad \forall z \in \{1, \dots, Z\} \tag{1}$$

where $X_{t-\tau:t}$ includes historical data (lagged temperatures, occupancy), U_t are control inputs (HVAC setpoints), W_t are exogenous variables (weather conditions), and θ are model parameters. This formulation supports both offline prediction and integration with RL-based control policies.

Table 1. Multi-stage learning framework for scalable temperature prediction, detailing zones and prediction horizons.

Stage	#Zones	Horizon	Description
Stage 1	1	24 hours	Single zone, 1 day to establish baseline accuracy with minimal complexity
Stage 2	1	7 days	Single zone, 1 week to evaluate temporal scalability under extended horizons
Stage 3	123	7 days	123 zones, 1 week to assess spatial scalability across multiple zones
Stage 4	123	$\geq$ 28 days	123 zones, 4+ weeks to simulate real-world, long-term deployment scenarios

3.2. Multi-Stage Learning Framework

We present a four-stage learning framework that progressively increases prediction complexity:

$$S_k = (Z_k, H_k), \quad k \in \{1, 2, 3, 4\} \tag{2}$$

where Z_k is the number of zones and H_k is the prediction horizon. The stages are illustrated in Table 1.

3.3. Model and Feature Engineering

We employ XGBoost (Chen & Guestrin, 2016) due to its efficiency in handling tabular data and non-linear relationships. Hyperparameters (learning rate=0.1, max depth=6, $n_{estimators}$=100) are optimized via grid search to balance accuracy and computational cost, ensuring suitability for real-time HVAC applications.

Feature engineering is pivotal for modeling building thermal dynamics using the Smart Buildings dataset. By incorporating lagged temperatures, we capture thermal inertia, reflecting the delayed response of building materials to HVAC actions. Moving averages mitigate short-term sensor noise, enhancing prediction stability. Weather forecasts, including external temperature, humidity, and solar radiation, account for environmental influences critical for accurate modeling.

4. Experimental Setup

4.1. Dataset

The Smart Buildings dataset (SB1) spans January 2022 to June 2024, covering 911 days (261,852 timesteps at 5-minute intervals) across 123 zones in a 93,858 ft² building with 173 HVAC devices (Goldfeder et al., 2024). Table 2 details the dataset partitions, and Table 3 summarizes the sensor network, which includes temperature, setpoint, and flow measurements critical for comprehensive HVAC modeling.

4.2. Evaluation Metrics

We adopt the Temporal Spatial Mean Absolute Error (TS-MAE) from (Goldfeder et al., 2024):

Table 2. Smart Buildings Dataset (SB1) partitions, detailing temporal coverage and data volume for scalable prediction.

Partition	Period	Days	Timesteps	Size (MB)
2022_a	Jan-Jun	180	51,852	1,768.3
2022_b	Jul-Dec	183	53,292	1,844.3
2023_a	Jan-Jun	180	51,852	1,768.3
2023_b	Jul-Dec	183	52,716	1,797.8
2024_a	Jan-Jun	181	52,140	1,778.2
Total	**2.5 Years**	**911**	**261,852**	**8,956.9**

Table 3. Sensor distribution in SB1, enabling robust HVAC modeling with diverse measurements.

Sensor Type	Count	Description
Temperature	123	Zone air temperature sensors
Setpoint	528	HVAC setpoint controls
Command	241	Device command signals
Status	4	Equipment operational status
Other	302	Flow, pressure, frequency sensors
Total	**1,198**	**Complete sensor network**

$$\epsilon = \frac{1}{N}\sum_{t=1}^{N}\left[\frac{1}{Z}\sum_{z=1}^{Z}|T_{real,t,z} - T_{pred,t,z}|\right] \tag{3}$$

where $T_{real,t,z}$ and $T_{pred,t,z}$ are the actual and predicted temperatures for zone z at time t. The coefficient of determination, R^2, measures explained variance to assess model fit. Training time is used to assess computational efficiency, which is critical for real-time HVAC applications.

5. Results and Discussion

Table 4 summarizes XGBoost performance across the four stages. In Stages 1 and 2 (single-zone, 1 day and 1 week), the model achieves excellent accuracy (MAE=0.195°F, R^2=0.957, training time less than 0.4s), demonstrating robust short-term forecasting. Stage 3 (123 zones, 1 week) maintains strong performance (MAE=0.523°F), indicating effective spatial scaling. However, Stage 4 (123 zones, 4+ weeks) shows degraded accuracy (MAE=1.795°F), reflecting challenges in long-term forecasting due to seasonal variability and complex inter-zone dynamics. The MAE distribution across 123 zones in Stage 3 highlights that the zones with stable thermal dynamics (e.g., interior offices) achieve near-perfect predictions (MAE less than 0.3°F), while perimeter zones exposed to external weather show higher variability (MAE up to 0.8°F).

Table 5 compares XGBoost with baseline methods. XGBoost achieves the lowest MAE (0.195°F), scales effectively to 123 zones, and trains in 0.37–93.6s, outperforming linear regression (MAE=1.8°F), and LSTMs (MAE=1.2°F, training time=120s). LSTMs struggle with multi-zone scalability due to high computational complexity, while linear regression fails to capture non-linear thermal dynamics.

Table 4. XGBoost performance across stages, showing trade-offs between accuracy and prediction horizon. Z=Zone(s), D=Day, W=Week(s).

Stage	MAE (°F)	R^2	Time (s)
1Z 1D	0.195	0.957	0.37
1Z 1W	0.195	0.957	0.24
123Z 1W	0.523	0.415	34.5
123Z 4W	1.795	0.400	93.6

Table 5. Method comparison, highlighting XGBoost's superior accuracy, scalability, and efficiency for multi-zone prediction.

Method	MAE (°F)	Scalability	Training Time (s)	Multi-Zone
Linear Regression	1.8	Good	0.5	Yes
LSTM	1.2	Fair	120	Limited
XGBoost (Ours)	**0.195**	**Excellent**	**0.37–93.6**	**Yes**

Scalability Analysis. The multi-stage framework demonstrates robust scalability up to Stage 3, with Stage 3's MAE of 0.523°F and training time of 34.5s indicating feasibility for real-time HVAC control. Stage 4's performance degradation (MAE=1.795°F) suggests diminishing returns beyond 1.5 years of training data, likely due to seasonal variability and unmodeled factors like radiative heat transfer, as noted in the Smart Buildings Suite's limitations (Goldfeder et al., 2024). Increasing feature dimensionality improves robustness but introduces computational overhead, highlighting a trade-off between accuracy and efficiency.

Discussion. Our multi-stage framework, systematically addresses the scalability challenges of HVAC control, achieving high accuracy in single-zone (MAE=0.195°F) and multi-zone (MAE=0.523°F) scenarios. The use of XGBoost with emphasis on lightweight, data-driven models, offering significant advantages over computationally intensive methods. The framework's fast training times (0.37–93.6s) make it suitable for real-time applications, potentially reducing HVAC energy consumption by optimizing setpoint adjustments based on accurate temperature predictions. Zone-level variability reveals that interior zones benefit from stable thermal dynamics, while perimeter zones are more sensitive to external conditions, consistent with the suite's findings on heat exchange rates (Goldfeder et al., 2024). This informs targeted improvements, such as incorporating radiative heat models to enhance perimeter zone predictions.

Limitations and Future Work. The proposed framework, while effective for temperature prediction, is limited by its focus on forecasting rather than direct HVAC control optimization. The absence of a radiative heat model, contributes to performance degradation in Stage 4. Additionally, generalization across diverse buildings remains untested, as evaluations are confined to SB1. Future work includes integrating predictions with reinforcement learning or model predictive control for closed-loop optimization, evaluating performance across multiple buildings to ensure generalizability, exploring hybrid XGBoost-LSTM or PINN models to enhance long-term forecasting, and incorporating radiative heat transfer to improve perimeter zone predictions. Furthermore, the computational and storage demands of deep learning models for edge deployment can be alleviated by applying model compression techniques such as pruning (Han et al., 2015b; Shinde, 2025b) and quantization (Han et al., 2015a; Shinde, 2024; 2025a).

6. Conclusion

We present a multi-stage learning framework for scalable building temperature prediction, achieving high accuracy (MAE=0.195°F for single-zone tasks) using XGBoost and the Smart Buildings Control Suite dataset. By scaling from single-zone to 123-zone predictions over extended horizons. This work advances energy-efficient HVAC systems, reducing the carbon footprint of commercial buildings. By fostering cross-disciplinary collaboration, the framework paves the way for sustainable building management, supporting global climate goals.

References

Chen, T. and Guestrin, C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.

Chen, Y., Yang, Q., Chen, Z., Yan, C., Zeng, S., and Dai, M. Physics-informed neural networks for building thermal modeling and demand response control. *Building and Environment*, 234:110149, 2023.

Fong, K. F., Hanby, V. I., and Chow, T.-T. Hvac system optimization for energy management by evolutionary programming. *Energy and buildings*, 38(3):220–231, 2006.

Goldfeder, J., Dean, V., Jiang, Z., Wang, X., Lipson, H., Sipple, J., et al. The smart buildings control suite: A diverse open source benchmark to evaluate and scale hvac control policies for sustainability. *arXiv preprint arXiv:2410.03756*, 2024.

Han, S., Mao, H., and Dally, W. J. Deep compression: Compressing deep neural networks with pruning, trained quantization and huffman coding. *arXiv preprint arXiv:1510.00149*, 2015a.

Han, S., Pool, J., Tran, J., and Dally, W. Learning both weights and connections for efficient neural network. *Advances in neural information processing systems*, 28, 2015b.

Jiang, Z. and Dong, B. Modularized neural network incorporating physical priors for future building energy modeling. *Patterns*, 5(8), 2024.

Luo, N., Wang, Z., Blum, D., Weyandt, C., Bourassa, N., Piette, M. A., and Hong, T. A three-year dataset supporting research on building energy management and occupancy analytics. *Scientific data*, 9(1):156, 2022.

Miller, C., Kathirgamanathan, A., Picchetti, B., Arjunan, P., Park, J. Y., Nagy, Z., Raftery, P., Hobson, B. W., Shi, Z., and Meggers, F. The building data genome project 2, energy meter data from the ashrae great energy predictor iii competition. *Scientific data*, 7(1):368, 2020.

Murray, D., Stankovic, L., and Stankovic, V. An electrical load measurements dataset of united kingdom households from a two-year longitudinal study. *Scientific data*, 4(1): 1–12, 2017.

Pérez-Lombard, L., Ortiz, J., and Pout, C. A review on buildings energy consumption information. *Energy and buildings*, 40(3):394–398, 2008.

Sendra-Arranz, R. and Gutiérrez, A. A long short-term memory artificial neural network to predict daily hvac consumption in buildings. *Energy and Buildings*, 216: 109952, 2020.

Shinde, T. Adaptive quantization and pruning of deep neural networks via layer importance estimation. In *Workshop on Machine Learning and Compression, NeurIPS 2024*, 2024.

Shinde, T. Model compression meets resolution scaling for efficient remote sensing classification. In *Proceedings of the Winter Conference on Applications of Computer Vision*, pp. 1200–1209, 2025a.

Shinde, T. Towards optimal layer ordering for efficient model compression via pruning and quantization. In *2025 25th International Conference on Digital Signal Processing (DSP)*, pp. 1–5. IEEE, 2025b.

Taheri, S., Hosseini, P., and Razban, A. Model predictive control of heating, ventilation, and air conditioning (hvac) systems: A state-of-the-art review. *Journal of Building Engineering*, 60:105067, 2022.

Yu, L., Qin, S., Zhang, M., Shen, C., Jiang, T., and Guan, X. A review of deep reinforcement learning for smart building energy management. *IEEE Internet of Things Journal*, 8(15):12046–12063, 2021.

Zou, Z., Yu, X., and Ergan, S. Towards optimal control of air handling units using deep reinforcement learning and recurrent neural network. *Building and Environment*, 168: 106535, 2020.