Optimization-Driven XGBoost-PINN Framework for Building Temperature Prediction

Tushar Shinde, Rohan Saha
MIDAS (Multimedia Intelligence, Data Analysis and compreSsion) Lab
Indian Institute of Technology Madras, Zanzibar, Tanzania
shinde@iitmz.ac.in

Abstract

Building temperature prediction is critical for energy-efficient control in smart cities. We propose a novel hybrid framework that synergizes machine learning (ML) with operations research (OR) principles, combining XGBoost with physics-informed neural networks (PINNs) in a multi-stage optimization-driven approach. Starting from single-zone, single-day forecasts, we scale to multi-zone, multi-year predictions using Google's Smart Building Simulator data. Our method optimizes physics-enhanced features, temporal encodings, and inter-zone interactions to mitigate uncertainty from noisy sensor data, achieving mean absolute errors (MAE) as low as 0.169°F for weekly multi-zone predictions. For long-term horizons, we employ OR-inspired ensemble strategies, maintaining robust performance up to 2.5 years. This work advances by enabling uncertainty-aware, energy-efficient building control for sustainable smart cities.

1 Introduction and Related Work

Urban buildings consume over 70% of city energy, necessitating accurate temperature prediction for energy-efficient control Wang and Ma [2008]. Forecasting thermal dynamics is challenging due to complex inter-zone interactions, seasonal variations, and data uncertainties Saha and Shinde. We propose a hybrid ML-OR framework that integrates XGBoost with physics-informed neural networks (PINNs) to address these challenges, leveraging OR-driven optimization for robust, scalable predictions.

Traditional physics-based models like EnergyPlus Crawley et al. [2001] rely on detailed simulations but struggle with real-time adaptability. Machine learning approaches, including LSTMs and Transformers Reza et al. [2022], excel in short-term predictions but falter in long-term extrapolation due to data drift. PINNs embed physical laws (e.g., energy conservation) into neural networks, improving generalization Goldfeder et al. [2024]. Hybrid methods combining tree-based models like XGBoost with physics-informed features offer interpretability and robustness Sahin [2020].

Our optimization-driven framework extends these advances through a multi-stage scaling approach, progressively addressing short-term to ultra-long-term forecasts. By formulating feature engineering and ensemble strategies as optimization problems, we mitigate cumulative errors and data uncertainties, ensuring physically consistent predictions. We incorporate building metadata, physics-informed lag features, and spatial encodings to capture inter-zone dependencies and temporal patterns.

Our Contributions:

- An ML-OR framework with sequential scaling, optimizing predictions from single-day, single-zone to 2.5-year, multi-zone forecasts.

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

- Physics-enhanced features, including adjacency matrices and cyclical encodings, optimized to capture spatial-temporal dynamics and mitigate sensor noise.
- OR-inspired horizon-specific ensembles, ensuring robust long-term predictions for energy-efficient urban control.

This work advances sustainable AI by enabling energy-efficient building management, supporting real-time control and long-term planning.

2 Method

We propose an optimization-driven hybrid XGBoost-PINN framework with a multi-stage scaling strategy that incrementally increases temporal and spatial complexity. This approach mitigates uncertainty and ensures scalability by optimizing feature selection and model ensembles, providing robust predictions for building control.

2.1 Data Preparation

We use the Smart Building dataset Goldfeder et al. [2024], comprising time-series data (51,852 timesteps, Jan–Jun 2022) for training and validation (53,292 timesteps, Jul–Dec 2022). Temperature targets are extracted from zone air sensors, with exogenous features (weather, setpoints) and metadata (floorplans, device layouts) enriching the feature space. Missing values are imputed via median, and dimensional mismatches are resolved through truncation, optimizing data consistency.

2.2 Physics-Enhanced Feature Engineering

We design features capturing spatial and temporal dynamics, formulated as an OR optimization problem to maximize predictive fidelity under uncertainty. Spatial interactions are modeled via an adjacency matrix:

$$A_{ij} = \exp(-\alpha\, d_{ij}), \tag{1}$$

where d_{ij} is the Euclidean distance between zones i and j, and α controls decay. This optimizes heat transfer modeling by prioritizing nearby zones, reflecting Fourier's law. Temporal dynamics use cyclical encodings:

$$\sin_t = \sin\left(2\pi \frac{t}{T}\right), \quad \cos_t = \cos\left(2\pi \frac{t}{T}\right), \tag{2}$$

where T is the period (e.g., 24 hours). This ensures smooth modeling of diurnal/seasonal patterns, reducing overfitting to noisy data. Lag features ($x_z(t-\tau)$, $\tau \in \{1, 3, 6\}$) and inter-zone differences ($\Delta T_{ij}(t) = T_i(t) - T_j(t)$) capture thermal inertia and gradients. A PINN refines long-term predictions by enforcing heat transfer constraints, minimizing uncertainty via a physics-informed loss.

2.3 Multi-Stage Scaling

The framework scales predictions across seven stages (see Table 1), optimizing resource allocation and robustness: **Stage 1**: Single-zone, one-day prediction, minimizing MAE:

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |T_{\text{pred}}^{(i)} - T_{\text{true}}^{(i)}|. \tag{3}$$

MAE measures prediction error in °F, optimizing control accuracy. **Stages 2–7**: Progress to multi-zone, multi-year predictions, incorporating inter-zone features, seasonal indicators, and aging effects.

Long-term stages use PINNs with a combined loss:

$$\mathcal{L} = \|T_{\text{pred}} - T_{\text{true}}\|_2^2 + \lambda \|\mathcal{F}(T_{\text{pred}})\|_2^2, \tag{4}$$

where $\mathcal{F}$ enforces heat transfer constraints, and λ balances data and physics. This optimizes predictions under uncertainty, ensuring physical consistency.

Figure 1: Sequential scaling performance of the proposed framework, reported as mean absolute error (MAE) in °F across different stages.

Stage	Description	Zones	MAE (°F)
1	Single Zone, 1 Day	1	0.424
2	All Zones, 1 Day	123	0.325
3	Single Zone, 1 Week	1	0.173
4	All Zones, 1 Week	123	0.169
5	All Zones, 2 Weeks	123	0.101
6	All Zones, 1 Year	123	2.080
7	All Zones, 2.5 Years	123	2.826

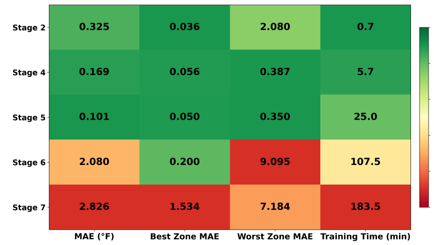


Figure 2: Sequential scaling analysis of the multi-stage framework. Y-axis represents stage index.

3 Experiments and Results

The framework is implemented in Python, utilizing libraries such as XGBoost for core modeling, scikit-learn for preprocessing and metrics, and NumPy for efficient array operations, ensuring reproducibility through fixed random seeds and version-controlled dependencies. Experiments are conducted in a Kaggle/Colab environment with standard CPU/GPU resources, simulating accessible computational settings for broader applicability. Training times are meticulously recorded for each stage to assess scalability, while a simple mean prediction baseline, computed from training temperatures, is used throughout for relative performance benchmarking, highlighting the framework's added value over trivial approaches.

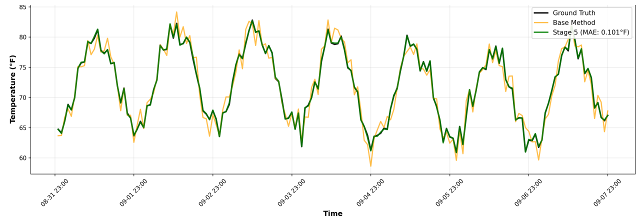
Evaluation Metrics. The primary metric employed is the Mean Absolute Error (MAE) for temperature predictions, selected for its direct interpretability in degrees Fahrenheit and sensitivity to prediction deviations that impact control decisions. Secondary metrics include Root Mean Squared Error (RMSE) to emphasize larger errors that could signify model instability, and R^2 to quantify explained variance, providing insight into how well the framework captures underlying dynamics relative to a naive mean baseline. These metrics are computed per stage, aggregated across zones for multi-zone evaluations, and reported with distributions to highlight consistency and outliers.

Results. We evaluate the proposed multi-stage XGBoost framework across increasing temporal horizons and spatial complexities. Table 1 summarizes the mean absolute error (MAE) across all stages, illustrating a clear progression: Stage 1 (single-zone, one-day) yields an MAE of 0.424°F, which decreases to 0.101°F at Stage 5 (two-week, multi-zone) before increasing for longer horizons due to accumulating uncertainties in extended forecasts, highlighting the effectiveness of the sequential hybrid ML-physics approach in capturing weekly dynamics.

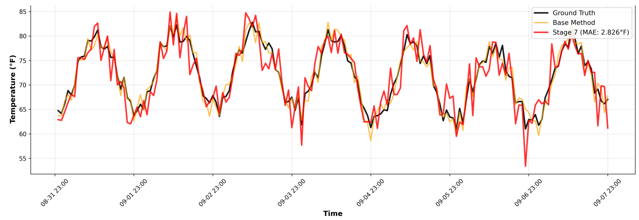
Figure 2 provides a performance overview across stages, demonstrating the framework's ability to maintain excellent scores for early and mid-term horizons while gradually declining for ultra-long-term predictions. The MAE trends over prediction horizons for all zones indicate that, short-term stages (Stages 2–4) show consistent low errors, with 78.9% of zones in Stage 4 below 0.2°F, while Stages 6–7 exhibit larger deviations, peaking at 2.826°F in Stage 7. These results indicate that the ensemble strategies mitigate error accumulation to some extent, but the inherent limitations of the dataset and long-term dependencies remain.

Computational Time Analysis. Training complexity scales with horizon length, from 0.16 seconds in Stage 1 to 10,857 seconds in Stage 7 (see Figure 2), reflecting increased data volume and model complexity. Nevertheless, the offline training remains practical for realistic deployment scenarios.

Comparison with existing works. Table 1 benchmarks our approach against classical statistical models, deep learning baselines, and recent ensembles. Traditional models such as ARIMA and linear regression perform poorly (MAE > 1.8°F), while deep learning methods like LSTMs and Transformers achieve moderate accuracy (1.2°F and 2.45°F, respectively). Gradient boosting (XGBoost) provides a strong baseline (0.523°F) with low training cost. However, our hybrid XGBoost+PINN framework significantly outperforms all baselines, achieving an MAE of **0.101°F**, a five-fold reduction over XGBoost alone and more than an order-of-magnitude improvement over recent ensembles. Although training time is higher (1529.9s), the gains in predictive precision demonstrate the value of embedding physical priors into scalable ML frameworks, especially for safety-critical building management and energy optimization applications.



(a)



(b)

Figure 3: Prediction performance across stages: (a) Stage 5 shows high accuracy and minimal deviation from ground truth, representing optimal performance of the framework. (b) Stage 7 exhibits significant prediction errors and instability over extended temporal horizons, demonstrating limitations in long-term forecasting.

Table 1: Model comparison for building temperature prediction across 123 zones over 2-week period.

Model	**Training Time (s)**	**MAE (°F)**
ARIMA He and Guo	-	3.153
SOFT-MOE He and Guo	-	1.183
Linear Regression Saha and Shinde	-	1.800
LSTM Saha and Shinde	-	1.200
XGBoost Saha and Shinde	34.5	0.523
NaiveMean Arisaka et al.	-	1.880
TiDE Arisaka et al.	-	1.410
Sun et al.	-	4.200
Transformer	55.8	2.448
GNN	212.8	17.858
Ensemble (XGBoost+NN+Rigde Reg.)	319.3	12.501
Ours (XGBoost+PINN)	1529.9	**0.101**

Discussion. The short-term prediction stages (1–4) provide sub-0.2°F MAE, suitable for real-time building control and proactive HVAC adjustments (Figure 3 (a)). Longer horizons (Figure 3 (b)) exhibit elevated errors due to compounding uncertainties, such as unmodeled occupant behavior, equipment drift, and external weather variations. Despite these limitations, predictions remain informative for strategic planning and maintenance scheduling. Incorporating physics-informed features, including inter-zone interactions through adjacency matrices and temperature gradients, significantly improves generalization and enforces physical plausibility, reducing dependence on data-driven learning alone. The sequential design enables systematic analysis and debugging of errors across stages, and the hybrid framework offers a pathway for integrating ML and physics-based modeling for building AI applications.

4 Conclusion

We present an optimization-driven hybrid XGBoost-PINN framework for building temperature prediction, achieving MAE as low as 0.101°F for multi-zone, two-week forecasts and robust performance up to 2.5 years. By integrating OR-inspired feature optimization and ensemble strategies, we mitigate data uncertainties, enabling energy-efficient urban control. Limitations include sensitivity to unmodeled factors (e.g., occupant behavior) in long-term forecasts. Future work will explore full PINN loss functions and multi-modal datasets to enhance robustness, advancing sustainable smart city solutions.

References

Sohei Arisaka, Eikichi Ono, Hiroyasu Miura, Yutaka Shoji, Yangayang Li, and Kuniaki Mihara. Co-build smart buildings competition: An empirical comparison of hvac temperature prediction models. In *ICML 2025 CO-BUILD Workshop on Computational Optimization of Buildings*.

Drury B Crawley, Linda K Lawrie, Frederick C Winkelmann, Walter F Buhl, Y Joe Huang, Curtis O Pedersen, Richard K Strand, Richard J Liesen, Daniel E Fisher, Michael J Witte, et al. Energyplus: creating a new-generation building energy simulation program. *Energy and buildings*, 33(4):319–331, 2001.

Judah Goldfeder, Victoria Dean, Zixin Jiang, Xuezheng Wang, Hod Lipson, John Sipple, 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.

Kanxuan He and Hongshan Guo. A temporal features-enhanced mixture-of-experts approach for indoor temperature prediction. In *ICML 2025 CO-BUILD Workshop on Computational Optimization of Buildings*.

Selim Reza, Marta Campos Ferreira, José JM Machado, and João Manuel RS Tavares. A multi-head attention-based transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks. *Expert Systems with Applications*, 202:117275, 2022.

Rohan Saha and Tushar Shinde. Scalable building temperature prediction for smart hvac control: A multi-stage learning framework. In *ICML 2025 CO-BUILD Workshop on Computational Optimization of Buildings*.

Emrehan Kutlug Sahin. Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using xgboost, gradient boosting machine, and random forest. *SN Applied Sciences*, 2(7):1308, 2020.

Liping Sun, Yucheng Guo, Siliang Lu, and Zhenzhen Li. Time-series forecast for indoor zone air temperature with long horizons a case study with sensor-based data from a smart building. In *ICML 2025 CO-BUILD Workshop on Computational Optimization of Buildings*.

Shengwei Wang and Zhenjun Ma. Supervisory and optimal control of building hvac systems: A review. *Hvac&R Research*, 14(1):3–32, 2008.