Multi-Horizon Forecasting of Indoor $PM_{2.5}$ Using Machine Learning in a Digital Twin Environment

Zihao Zheng[1], Karunakar Reddy Mannem[2], Borja Garcia de Soto[1]

[1]S.M.A.R.T. Construction Research Group, Division of Engineering, New York University Abu Dhabi (NYUAD), Experimental Research Building, Saadiyat Island, P.O. Box 129188, Abu Dhabi, United Arab Emirates
[2]Center for Research Computing, New York University Abu Dhabi (NYUAD), Saadiyat Island, P.O. Box 129188, Abu Dhabi, United Arab Emirates
zihao.zheng@nyu.edu, krm7@nyu.edu, garcia.de.soto@nyu.edu

Abstract –
Indoor air quality management remains largely reactive due to limited multi-horizon forecasting capabilities and deployment challenges under real-world constraints. This paper presents a Digital Twin (DT)-enabled framework for multi-horizon indoor $PM_{2.5}$ forecasting from 1 to 6 hours ahead using machine learning and multi-source environmental data from a university campus in Abu Dhabi, UAE. Indoor sensor measurements are combined with outdoor monitoring data and ERA5 wind variables used with a realistic five-day availability delay. CatBoost and Transformer models are evaluated using seasonal cross-validation, with CatBoost demonstrating advantages at short-term horizons while Transformer marginally performs better at the long-term horizon. SHAP analysis reveals that current indoor $PM_{2.5}$ is the dominant predictor for short-term forecasts. An ablation study demonstrates that incorporating lagged wind data yields statistically significant improvements in non-winter seasons.

Keywords –
Indoor air quality; CatBoost; Transformer; Reanalysis data

1 Introduction

Indoor air quality (IAQ) has gained increasing attention globally, particularly as people spend approximately 90% of their time in indoor environments and ambient $PM_{2.5}$ concentrations remain a persistent public health concern in many urban environments [1]. Among various indoor pollutants, particulate matter with a diameter of less than 2.5 micrometers ($PM_{2.5}$) poses significant health risks, including diabetes and prenatal disorders [2]. Indoor $PM_{2.5}$ primarily enters indoor environments through ventilation and infiltration processes, or is generated internally by human activities such as printing and photocopying, cleaning and maintenance activities (e.g., use of disinfectant sprays), frequent door opening that resuspends settled dust, and movement of occupants [2]. Many studies have proposed models to predict indoor PM2.5, but far fewer have dealt with what happens when those models must run continuously with delayed external data, mixed sensor streams, and decision-support needs in a live DT.

Digital Twin (DT) technology offers a transformative paradigm for data-driven IAQ management by creating virtual replicas of physical indoor spaces that enable real-time monitoring, prediction, and optimization. Current $PM_{2.5}$ monitoring systems in DT applications are primarily reactive, alerting occupants after pollutant concentrations exceed threshold levels. With advanced predictive analytics, DTs enable proactive measures for indoor $PM_{2.5}$ risk prevention and support preventive health management [3]. However, the application of predictive models to indoor environments faces unique challenges, particularly the need for real-time predictions with integration of available near real-time data sources such as indoor sensors, outdoor atmospheric monitoring stations, and satellite reanalysis data. What is still limited is the deployment of multi-horizon forecasting that can give operators useful lead time while relying only on data that are available at prediction time. Besides, wind-related variables from ERA5, the global meteorological reanalysis produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) [4], which are widely used in indoor and outdoor $PM_{2.5}$ studies for reliable wind data [5], [6], are subject to an inherent latency of approximately five days before becoming available. To the best of the authors' knowledge, the influence of this latency on indoor PM2.5 prediction has not been fully investigated, representing an operational constraint in real-time prediction deployment in DT applications.

This paper presents a framework that integrates near real-time multi-source data to forecast indoor $PM_{2.5}$ within a DT managed environment. Environmental data from indoor sensors ($PM_{2.5}$, humidity, and pressure), outdoor atmospheric monitoring stations (PM_1, $PM_{2.5}$, PM_{10}, temperature, humidity, pressure, and dew point), and lagged ERA5 meteorological reanalysis data (U and

V wind components, wind gust) are integrated to train CatBoost, Transformer, and persistence models for predictions up to six hours ahead. The contribution is not a new forecasting algorithm; it is a deployment-oriented forecasting workflow that accounts for data latency, tests whether delayed wind information is still useful, and shows how multi-horizon predictions can be pushed into a live DT interface.

2 Previous Studies

2.1 Hourly Indoor $PM_{2.5}$ Prediction

Recent studies have employed diverse machine learning architectures for indoor $PM_{2.5}$ forecasting. Recurrent architectures including LSTM, BiLSTM, and GRU have been widely adopted for capturing temporal dependencies in time-series air quality data [7]. Gradient boosting methods such as XGBoost and LightGBM demonstrate strong performance through ensemble learning approaches [6], [8]. Sparse attention–based Transformer models efficiently capture long-term dependencies in $PM_{2.5}$ time series and demonstrate improved performance for both short-term and long-term air quality forecasting [9]. Feature selection for indoor $PM_{2.5}$ prediction typically incorporates indoor environmental parameters ($PM_{2.5}$, temperature, humidity), outdoor atmospheric conditions (meteorological variables, ambient pollutants), and contextual factors (occupancy, season, time), resulting in a total of 10 to 21 features used for prediction [10], [11]. Even so, most published studies emphasize model accuracy, whereas deployment-oriented questions such as data latency, inference timing, and how much lead time is practically useful receive much less attention.

Current studies on indoor $PM_{2.5}$ forecasting primarily focus on model development and performance validation, with limited attention to real-world deployment challenges and operational constraints.

2.2 Predictive Digital Twin Application

DT technology has emerged as a promising platform for IAQ management, integrating IoT sensors, Building Information Modeling (BIM), and machine learning for real-time monitoring and visualization. Hu et al. [12] proposed a CNN-BiLSTM-MHA model for IAQ failure prediction integrated with a DT platform, demonstrating real-time alert capabilities but emphasizing anomaly detection over continuous predictive forecasting. Khosh-Amadi et al. [13] proposed an AIR-DT framework integrating IoT sensors, BIM, spatial analytics, and short-term IAQ prediction to enable real-time, asthma-oriented indoor air quality monitoring and intervention. Qian et al. [14] presented an intelligent IAQ management system based on DT platforms that integrates multi-source data from IoT and BIM, achieving rapid prediction and control responses within 30 minutes of threshold exceedance.

Despite these developments, critical gaps remain in integrating heterogeneous near real-time data sources from multiple domains (indoor, outdoor, meteorological) and implementing multi-horizon forecasting models that provide actionable lead times for building management interventions.

3 Framework of Digital Twin Enabled Predictive Analysis

The proposed framework comprises three layers encompassing seven processes, shown in Figure 1.

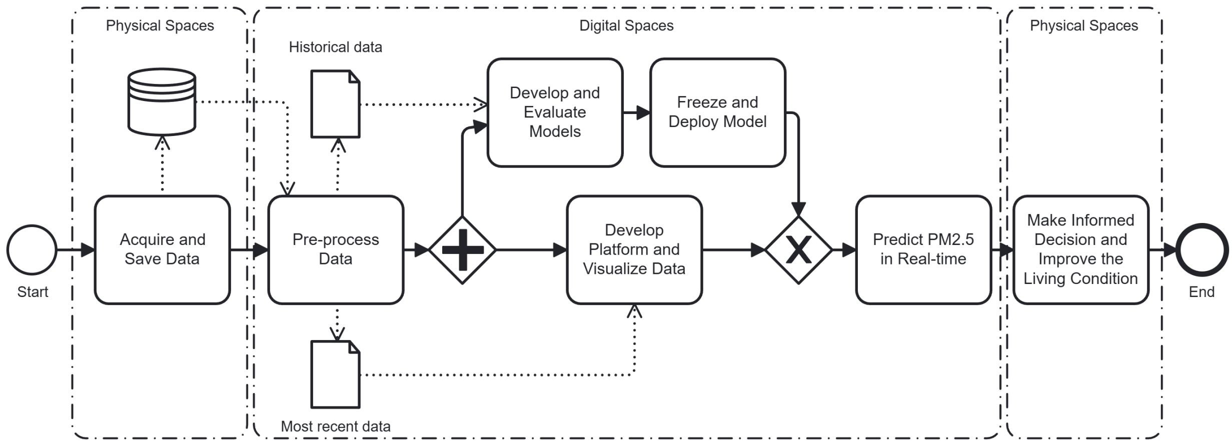

Figure 1: BPMN diagram for Digital Twin-enabled IAQ improvement with indoor $PM_{2.5}$ prediction

Data Acquisition Layer: acquire and save data from indoor sensors, outdoor stations, and ERA5 reanalysis; pre-process data through outlier removal following Tukey 1.5 $\times$ IQR rule, data imputation using KNN method [15], and hourly aggregation and temporal alignment.

<u>Model Development Layer</u>: develop and evaluate models using CatBoost and Transformer through seasonal cross-validation and Diebold-Mariano (DM) tests [16], then freeze and deploy models in the backend DT system.

<u>Broadcasting of Real-time Prediction Layer</u>: develop platform and visualize data, predict $PM_{2.5}$ in real-time, make informed decisions and improve living conditions, enabling bilateral information flow in DT.

4 Implementation

4.1 Data Acquisition

Three categories of environmental data were collected to capture the multi-scale influences on indoor $PM_{2.5}$ concentrations. Indoor environmental parameters were measured using sensors deployed at NYU Abu Dhabi, recording $PM_{2.5}$, humidity, and pressure at 1-minute intervals [17]. Outdoor atmospheric conditions were monitored using two complementary systems. Outdoor air quality stations provided measurements of particulate matter (PM_1, $PM_{2.5}$, PM_{10}), temperature, relative humidity, and atmospheric pressure at 1-minute intervals, and meteorological stations supplied dew point observations at 15-minute intervals.

The realistic 5-day lag constraint of the ERA5 data source was considered in this study. Hourly measurements of 10-m wind gust, as well as zonal (U) and meridional (V) wind components, were extracted from ERA5 for the New York University Abu Dhabi campus location. The complete dataset spans from April 6, 2024, to November 16, 2025, covering approximately 19 months of observations. All variables used in this study are summarized in Table 1.

As part of data preprocessing, 3.2% of indoor sensor readings, 4.0% of outdoor measurements, and 0.7% of meteorological station data were classified as outliers and excluded from further analysis. All variables were subsequently resampled and temporally aligned to an hourly resolution using mean aggregation. Missing values were handled using k-nearest neighbors (KNN) imputation, which filled 3.6% of indoor sensor data, 2.9% of outdoor data, and 22.9% of meteorological station data. Although the meteorological station dataset (dew point) exhibited a higher proportion of missing values, this feature was retained due to its moderately positive Pearson correlation of 0.28 with indoor $PM_{2.5}$. The final dataset comprises 14,129 hourly observations with 13 input features.

An overview of the indoor $PM_{2.5}$ after the data pre-processing is shown in Figure 2. The monthly statistics reveal substantial temporal variability, with mean $PM_{2.5}$ concentrations ranging from 4.1 μg/m³ in May 2024 to 10.2 μg/m³ in October 2025. Lower concentrations were generally observed during winter and spring months (December-May), while elevated levels occurred during summer and fall months (June-November), a pattern consistent with the influence of intense heat, enhanced secondary aerosol formation, frequent dust events, and reduced atmospheric dispersion during warmer periods in Abu Dhabi. Accordingly, a seasonally stratified cross-validation strategy was implemented to evaluate model performance under varying climatic conditions.

Table 1: Parameters collected from different sources

Abbreviation	Parameter	Data source
Ind. PM2.5	Indoor $PM_{2.5}$	Indoor*
Ind. Humi.	Indoor Humidity	Indoor
Ind. Press.	Indoor Pressure	Indoor
Outd. PM1	Outdoor PM_1	Outdoor**
Outd. PM2.5	Outdoor $PM_{2.5}$	Outdoor
Outd. PM10	Outdoor PM_{10}	Outdoor
Outd. Temp.	Outdoor Temperature	Outdoor
Outd. Humi.	Outdoor Humidity	Outdoor
Outd. Press.	Outdoor Pressure	Outdoor
Dew Point	Dew Point	Meteo. station***
Wind Gust	Instantaneous 10-meter wind gust	ERA5
U Wind	10-meter U wind component	ERA5
V Wind	10-meter V wind component	ERA5

** Sensors provided and maintained by the S.M.A.R.T. Construction Research Group at NYUAD*
*** Sensors provided and maintained by CITIES at NYUAD*
**** Sensors provided and maintained by ACCESS at NYUAD*

4.2 Model Development

4.2.1 Model Selection and Configuration

CatBoost gradient boosting and Transformer neural network architectures were selected for systematic comparison based on their respective strengths in ensemble learning with feature interaction capabilities [8] and attention-based temporal modeling [9]. CatBoost is a strong choice for heterogeneous tabular inputs with nonlinear interactions and relatively limited feature engineering, while the Transformer is well suited to learning longer temporal dependencies from lagged sequences. Preliminary experiments with other models, such as Random Forest and BiLSTM, showed almost 50% higher RMSE and were therefore excluded from this study. Additionally, a persistence baseline model, which assumes future $PM_{2.5}$ concentrations remain equal to the current observed value, was included for statistical comparison as a benchmark baseline [10].

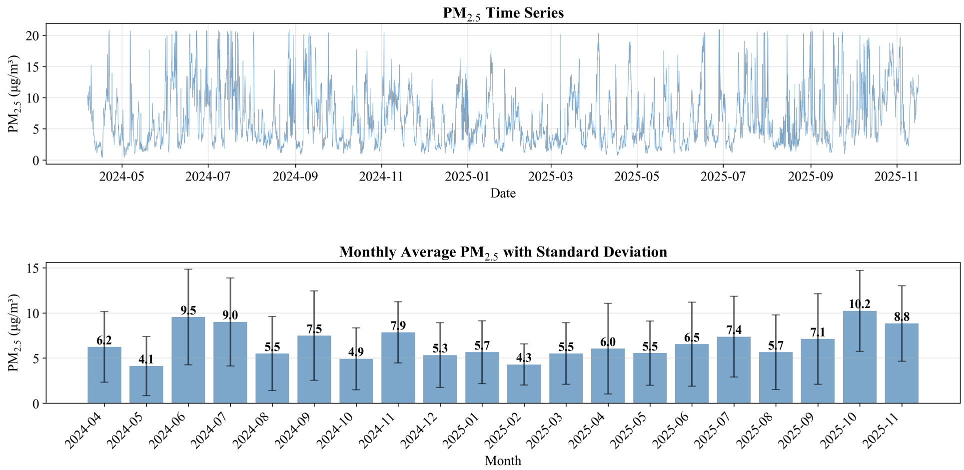

Figure 2: The overview and monthly average of the indoor $PM_{2.5}$

As previously indicated, the ERA5 API exhibits a five-day latency in data acquisition, which poses a limitation for real-time prediction in DT applications. Correlation analysis reveals that Wind Gust, U Wind, and V Wind exhibit moderate Pearson correlations of -0.27, -0.12, and 0.06, respectively, with indoor $PM_{2.5}$ when using contemporaneous data, but this operational lag constraint reduces these correlations to -0.04, -0.03, and -0.01. Nevertheless, the lagged wind features were retained as machine learning models may capture long-term and non-linear relationships that traditional linear regression cannot easily detect, potentially revealing residual atmospheric circulation patterns relevant to $PM_{2.5}$ dynamics. NaN wind data at the beginning of the dataset due to the five-day latency were filled using the earliest available wind observations.

Lag features are applied to all components to capture their temporal dependencies and periodicities [8]. Specifically, for each prediction horizon h, the input feature set includes observations from the current hour and the preceding six hours (at times t through $t-6$), as well as the corresponding hourly observation from 24 hours prior to the target prediction time (at $t+h-24$). All features were normalized using min-max scaling to ensure comparability across features and to support stable model training.

To evaluate model performance under different climatic conditions, a rolling-window cross-validation strategy was employed. Four folds (i.e., temporal splits) were constructed, as summarized in Table 2. For each fold, data from the eight months preceding the target season were used for model training. The data within the target season were then chronologically divided into two equal parts, with the first half used for validation and the second half reserved for testing. Across all folds, the approximate train/validation/test ratio is 72%/14%/14%.

The Transformer model was implemented using PyTorch, while the CatBoost model was implemented using the CatBoost Python library. Configurations of both models are listed in Table 3.

Table 2: Dataset splitting for cross validation

Split	Targeting Season	Time Frame
1	Spring	March, 2025 - May, 2025
2	Summer	June, 2025 - Aug., 2025
3	Fall	Sep., 2025 - Nov., 2025
4	Winter	Dec., 2024 - Feb., 2025

Table 3: Model configurations

CatBoost	Transformer
Max iterations: 2000	Max epochs: 200
Early stopping: yes	Early stopping: yes
Loss function: MAE	Loss function: MAE + Huber (delta=2)
Eval metric: MAPE	Eval metric: MAPE
Learning rate: 0.05	Batch size: 32
Depth: 3	Learning rate: 0.0005
L2 leaf reg: 5	Weight decay: 0.0001
Bagging temperature: 0.8	Embedding dimension: 64
Random strength: 0	Number of attention heads: 8
	Transformer layers: 3
	Dropout rate: 0.1

4.2.2 Results of Model Training

Performance metrics were computed for each cross-validation fold and for each prediction horizon (H1-H6). Model performance was evaluated across RMSE, MAE, MAPE, and R^2 at six prediction horizons, as shown in Figure 3-Figure 6.

1. Performance comparison of horizons

The results indicate a consistent decline in prediction accuracy as the forecast horizon extends from H1 to H6, a trend observed across all models and validation splits.

For CatBoost, mean performance metrics averaged over all splits indicate RMSE values of 0.9412, 1.3176, 1.6461, 1.9479, 2.2253, and 2.4647 µg/m³ (Figure 3), and MAE values of 0.4854, 0.8147, 1.0878, 1.3006, 1.5342, and 1.6996 µg/m³ across the six horizons (Figure 4). Corresponding mean MAPE increased from 8.3% to 28.6% (Figure 5), while mean R^2 decreased from 0.9420 to 0.6332 (Figure 6).

Despite this degradation, CatBoost and Transformer models at the H6 demonstrate predictive ability with average R^2 values of 0.6332 and 0.6278, surpassing the persistence model of 0.5797 by 9.2% and 8.3% respectively.

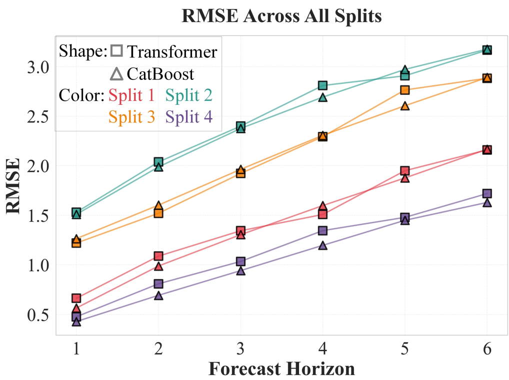

Figure 3: RMSE across all splits and 6 horizons

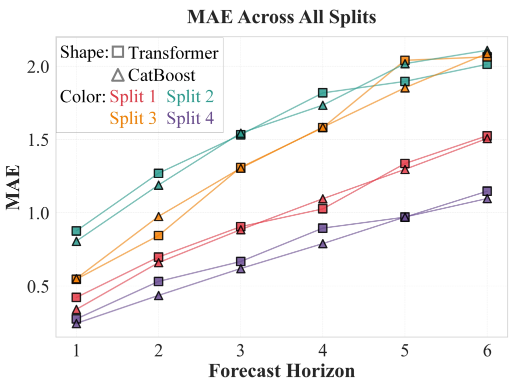

Figure 4: MAE across all splits and 6 horizons

2. Performance comparison of models

CatBoost demonstrated lower predictive error (RMSE, MAE, and MAPE) at short-term horizons compared to both Transformer and persistence models. At H1, CatBoost achieved an average MAE of 0.4854 µg/m³ across the four splits, compared to 0.5297 µg/m³ for Transformer and 0.5062 µg/m³ for persistence, representing improvements of 9.1% and 4.3% respectively. At H6, this reverses, with Transformer achieving 1.6870 µg/m³ while CatBoost achieving 1.6996 µg/m³. Both machine learning models outperformed the 1.7516 µg/m³ of persistence baseline by 3.1% and 3.8% respectively. CatBoost's gradient boosting architecture effectively captures the temporal dynamics and feature interactions governing indoor $PM_{2.5}$ evolution, particularly at shorter prediction horizons while Transformer marginally surpasses CatBoost at long-term prediction at H6.

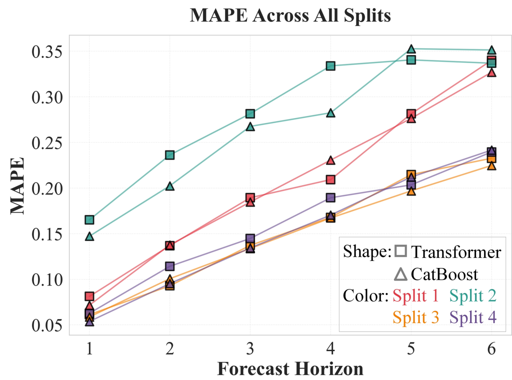

Figure 5: MAPE across all splits and 6 horizons

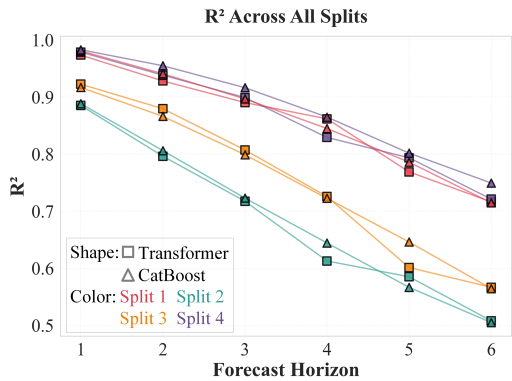

Figure 6: R^2 across all splits and 6 horizons

3. Performance comparison of seasonal split

Seasonal variations significantly influenced model performance across the four cross-validation splits, with Splits 1 (spring, in red) and 4 (winter, in purple) demonstrating notably better performance than Splits 2 (summer, in green) and 3 (fall, in orange). For Split 4, CatBoost achieved the lowest errors with MAE of 0.2448 µg/m³ at H1 and 1.0967 µg/m³ at H6, while for Split 1 CatBoost showed similarly favorable results with MAE of 0.3405 µg/m³ at H1 and 1.5067 µg/m³ at H6. In contrast, Splits 2 and 3 exhibited higher errors, with CatBoost's MAE at H6 reaching 2.1083 µg/m³ and 2.0868 µg/m³, representing increases of 92.2% and 90.3%, respectively, compared with Split 4.

This seasonal performance pattern is consistent with the $PM_{2.5}$ concentration trends shown in Figure 2. Spring and winter are characterized by lower and more stable pollutant levels, resulting in smoother temporal dynamics that are easier for data-driven models to learn. In contrast, summer and fall exhibit higher mean concentrations and greater short-term variability, driven by dust events, heat-induced atmospheric stagnation, and changing wind patterns in the UAE. These factors increase the volatility of indoor $PM_{2.5}$, thereby amplifying forecast uncertainty and contributing to the observed degradation in model performance for Splits 2 and 3.

4.2.3 Statistical Validation

One-sided DM test was employed to evaluate the statistical significance of performance differences between models using absolute error. Table 4 presents the DM statistics and p-values for all pairwise comparisons across prediction horizons. For example, at the two-hour forecast horizon (H2), the DM statistic for CatBoost vs. Transformer is -1.91 with a p-value of 0.0281. Because the statistic is negative and the p-value is below 0.05, this indicates that CatBoost performs significantly better than the Transformer at this horizon across all splits, and the observed improvement is unlikely to be due to random variation.

Table 4: DM test results on different horizons

Horizon	CatBoost vs Persistence	Transformer vs Persistence	CatBoost vs Transformer
1	-3.47 (0.0003)	-1.53 (0.0635)	-0.81 (0.2096)
2	-3.86 (0.0001)	-1.28 (0.0997)	-1.91 (0.0281)
3	-4.07 (0.0000)	-3.08 (0.0010)	-0.61 (0.2714)
4	-4.02 (0.0000)	-2.75 (0.0029)	-1.00 (0.1577)
5	-3.61 (0.0002)	-2.48 (0.0065)	-1.11 (0.1333)
6	-3.20 (0.0007)	-3.02 (0.0013)	-0.19 (0.4242)

Note: Data pair format - DM statistic (p value). Negative DM statistic indicates superior performance of the first model. $p < 0.01$ indicates strong significance; $p < 0.05$ indicates significance; $p < 0.10$ indicates marginal significance.

The CatBoost model significantly outperformed the persistence baseline across all horizons ($p < 0.01$) while the Transformer model significantly outperformed the persistence baseline from H3-H6 ($p < 0.01$) and marginally outperformed persistence baseline at H1 and H2 ($p < 0.1$). Both models showed their ability to capture temporal dynamics beyond simple autocorrelation. CatBoost exhibited the stronger superiority over persistence, with DM statistics ranging from -3.20 to -4.07 ($p < 0.01$). Comparing the two machine learning models, CatBoost demonstrated marginal significance over Transformer at horizon H2 ($p < 0.05$). On the other horizons, the performance differences were not statistically significant ($p > 0.10$), indicating comparable predictive capabilities at these horizons.

4.2.4 Importance Analysis

The model interpretability and the importance of the features are supported by the SHAP (Shapley Additive exPlans). Figure 7 shows an example of feature importance and directional impact for indoor $PM_{2.5}$ at H1 of split 4 of CatBoost model. It shows that indoor $PM_{2.5}$ at the current emerges as the dominant predictor, with high feature values strongly driving higher predictions, while outdoor PM_1 ranks second, reflecting the close coupling between indoor and outdoor particulate matter. Contribution of wind gust feature emerges after indoor $PM_{2.5}$ and other outdoor features.

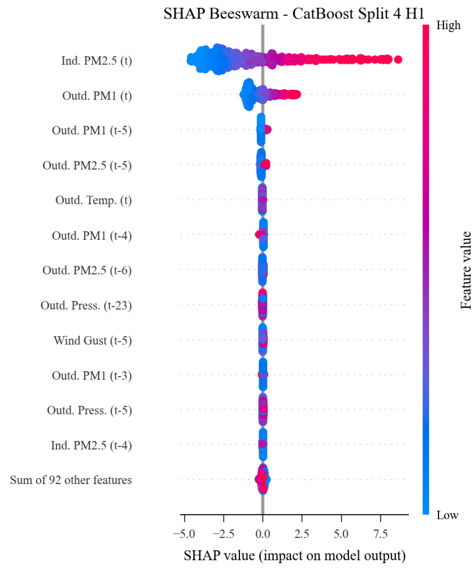

Figure 7: SHAP Beeswarm of CatBoost (H1)

4.3 Broadcasting of Real-time Prediction

A implementation of the real-time forecasting module is shown in Figure 8 and is accessible on the prototype online platform [18]. The data ingestion layer continuously acquires the most recent environmental measurements. Wind inputs are retrieved from the ERA5 reanalysis through the ECMWF API and incorporated into the forecasting workflow under a five-day availability delay, thereby ensuring consistency with operational data-access constraints. The forecasting service executes multi-horizon inference using the trained CatBoost models and publishes prediction outputs to the DT interface for visualization.

Although the models are developed using the data collected at NYUAD campus, the workflow and three-layer framework can be transferred to other buildings that have comparable indoor sensing, outdoor measurements, and meteorological inputs.

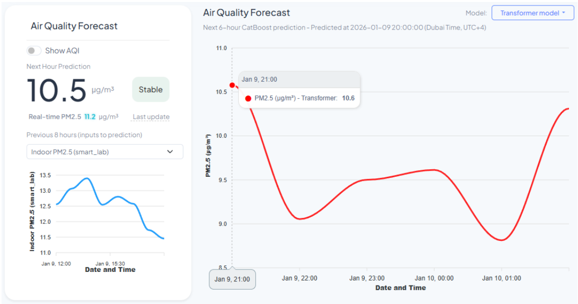

Figure 8: Screenshot of real-time forecasting platform

5 Ablation Study: Lagged Wind Feature

To validate whether using the lagged wind features provides meaningful predictive value, an ablation study was conducted using the CatBoost model aggregating all horizons. Table 5 shows the DM test results comparing the “Lagged ERA5” and “No ERA5” configurations over the four splits. It indicates that predictions with “Lagged ERA5” wind features are strongly significant in Split 1 (spring, $p < 0.01$) and 2 (summer, $p < 0.01$), marginally significant in Split 3 (fall, $p < 0.1$), while not significant in Split 4 (winter).

Table 5: DM test results of “Lagged ERA5” vs. “No ERA5” configurations

Split 1	Split 2	Split 3	Split 4
-3.42 (0.0003)	-3.31 (0.0005)	-1.51 (0.0658)	3.66 (0.9999)

Table 6 compares MAE performance across all prediction horizons for the “Lagged ERA5” and “No ERA5” configurations in Split 1. Overall, incorporating lagged ERA5 wind variables yields modest performance gains, with MAE variations remaining within ±2% across forecast horizons H2–H5.

Table 6: MAE (µg/m³) comparison of “Lagged ERA5” and “No ERA5” configurations in Split 1

Horizon	Lagged ERA5	No ERA5	Improvement
1	0.3405	0.3387	-0.5%
2	0.6594	0.6605	0.2%
3	0.8844	0.8863	0.2%
4	1.0954	1.1091	1.3%
5	1.2960	1.3070	0.9%
6	1.5067	1.5149	0.5%

6 Limitations

The five-day latency in ERA5 wind data yields only modest improvements (under 2%), suggesting a need for more recent wind data sources to enhance predictive performance. Reliance on a single reanalysis dataset (ERA5) and a single imputation method (KNN) may introduce dataset- or method-specific biases. In addition, all predictors used here are observational rather than directly manipulatable. Occupancy-related activities and HVAC operational variables were not available, even though such inputs would be valuable in a DT setting because they are closer to actionable control decisions. Finally, the framework was evaluated on a single campus in Abu Dhabi with a specific building configuration, and its generalizability to other climates and building configurations remains untested.

7 Conclusion

This paper presents a DT-enabled framework for multi-horizon indoor $PM_{2.5}$ forecasting using multi-source environmental data. CatBoost and Transformer models were evaluated using seasonal cross-validation, with CatBoost achieving average MAE of 0.4854 µg/m³ ($R^2 = 0.9420$) at one hour and 1.6996 µg/m³ ($R^2 = 0.6332$) at the six-hour horizon. Both models significantly outperformed persistence baselines, with CatBoost demonstrating statistically significant advantages at several intermediate horizons. SHAP analysis shows that current indoor $PM_{2.5}$ dominates short-term predictions, with its importance declining from 0.408 to 0.300 in CatBoost. An ablation analysis comparing “Lagged ERA5” and “No ERA5” configurations confirmed that incorporating lagged wind features yields statistically significant improvements, strengthening the case for deployment-ready forecasting.

Overall, the proposed framework provides a scalable foundation for data-driven indoor air quality management in DT–enabled environments. Future work should examine alternative real-time meteorological data sources, extend validation across diverse indoor settings and climates, and integrate building operational and occupancy information to further enhance predictive performance.

Acknowledgement

This research was partially supported by the Center for Sand Hazards and Opportunities for Resilience, Energy, and Sustainability (SHORES), funded by Tamkeen under the NYUAD Research Institute Award CG013. The authors also gratefully acknowledge CITIES and ACCESS for providing access to sensor data used in this paper.

References

[1] J. Cai, J. Chen, Y. Hu, S. Li, and Q. He, "Digital twin for healthy indoor environment: A vision for the post-pandemic era," *Front. Eng. Manag.*, vol. 10, no. 2, pp. 300–318, Jun. 2023, doi: 10.1007/s42524-022-0244-y.

[2] N. R. Martins and G. Carrilho da Graça, "Impact of PM2.5 in indoor urban environments: A review," *Sustainable Cities and Society*, vol. 42, pp. 259–275, Oct. 2018, doi: 10.1016/j.scs.2018.07.011.

[3] J. D. Castaño Molina, Z. Zheng, and B. García de Soto, "Predicting Indoor PM2.5 Levels Using Deep Learning for Enhanced Digital Twin Applications," in *Proceedings of the 2025 European Conference on Computing in Construction*, Porto, Portugal, Jul. 2025, pp. 1–8. doi: 10.35490/EC3.2025.277.

[4] H. Hersbach *et al.*, "ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS)." Accessed: Jan. 09, 2026. [Online]. Available: doi: 10.24381/cds.adbb2d47

[5] Z. Wang, P. Chen, R. Wang, Z. An, and L. Qiu, "Estimation of PM2.5 concentrations with high spatiotemporal resolution in Beijing using the ERA5 dataset and machine learning models," *Advances in Space Research*, vol. 71, no. 8, pp. 3150–3165, Apr. 2023, doi: 10.1016/j.asr.2022.12.016.

[6] W. Yu, B. Nakisa, S. W. Loke, S. Stevanovic, Y. Guo, and M. N. Rastgoo, "Indoor PM2.5 forecasting and the association with outdoor air pollution: a modelling study based on sensor data in Australia," May 13, 2024, *arXiv*: arXiv:2405.07404. doi: 10.48550/arXiv.2405.07404.

[7] B. Eren, İ. Aksangür, and C. Erden, "Predicting next hour fine particulate matter (PM2.5) in the Istanbul Metropolitan City using deep learning algorithms with time windowing strategy," *Urban Climate*, vol. 48, p. 101418, Mar. 2023, doi: 10.1016/j.uclim.2023.101418.

[8] A. Singh, M. Islam, and N. Dinh, "Forecasting Indoor Air Quality Using Machine Learning Models," in *2024 IEEE International Conference on Consumer Electronics (ICCE)*, Jan. 2024, pp. 1–6. doi: 10.1109/ICCE59016.2024.10444387.

[9] Z. Zhang and S. Zhang, "Modeling air quality PM2.5 forecasting using deep sparse attention-based transformer networks," *Int. J. Environ. Sci. Technol.*, vol. 20, no. 12, pp. 13535–13550, Dec. 2023, doi: 10.1007/s13762-023-04900-1.

[10] S. A. Asklany, D. Mohammed, I. K. Youssef, M. Nawaz, and W. Al Malwi, "Forecasting urban air quality in Paris using ensemble machine learning: A scalable framework for environmental management," *PLoS One*, vol. 20, no. 11, p. e0336897, Nov. 2025, doi: 10.1371/journal.pone.0336897.

[11] R. Das, A. Acharyya, and S. Majumdar, "Influences of Human Presence on the Indoor Air Quality of Educational Institutions: Concurrent Multipollutant Sensing Approach," *IEEE Open Journal of Instrumentation and Measurement*, vol. 4, pp. 1–8, 2025, doi: 10.1109/OJIM.2025.3583294.

[12] W. Hu, C. Chang, H. Wu, K. Lai, and Y. Cai, "Digital Twin-enabled failure prediction for indoor air quality: A CNN-BiLSTM model with multi-head attention mechanism," *Journal of Building Engineering*, vol. 117, p. 114755, Jan. 2026, doi: 10.1016/j.jobe.2025.114755.

[13] N. Khosh-Amadi, S. Talebi, F. Cheung, and M. Al-Adhami, "Development of a spatially contextualised and AI-enabled digital twin for asthma-specific indoor air quality management," *Journal of Building Engineering*, vol. 118, p. 115027, Jan. 2026, doi: 10.1016/j.jobe.2025.115027.

[14] Y. Qian, J. Leng, K. Zhou, and Y. Liu, "How to measure and control indoor air quality based on intelligent digital twin platforms: A case study in China," *Building and Environment*, vol. 253, p. 111349, Apr. 2024, doi: 10.1016/j.buildenv.2024.111349.

[15] Y. Kim *et al.*, "Positive matrix factorization outperforms machine learning in imputing missing PM2.5 and further identifying spatial patterns in multi-sites without external data," *Urban Climate*, vol. 62, p. 102552, Aug. 2025, doi: 10.1016/j.uclim.2025.102552.

[16] F. X. Diebold and R. S. Mariano, "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, vol. 20, no. 1, pp. 134–144, Jan. 2002, doi: 10.1198/073500102753410444.

[17] Z. Zheng, J. D. Castaño Molina, E. Mengiste, S. Prieto, and B. García de Soto, "Digital Twins for Healthier Spaces: A Scalable Framework for Monitoring Indoor Environmental Quality," in *ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci.*, 2025, pp. 1053–1060. doi: 10.5194/isprs-annals-X-G-2025-1053-2025.

[18] Z. Zheng, J. D. Castaño Molina, and B. García de Soto, "S.M.A.R.T. Forecast." Accessed: Jan. 09, 2026. [Online]. Available: https://digitaltwin.smart-crg.com/spaces/smart_lab/forecast/