3rd International Conference on Electronics, Engineering, Computer Science and Applied Development (EESD 2026)

Article

Application of Artificial Intelligence Algorithms in Urban Air Quality Monitoring and PM2.5 Forecasting

Nuo Xu[1,*]

[1] College of Information and Intelligence, Hunan Agricultural University, Changsha, China
* Correspondence: Nuo Xu, College of Information and Intelligence, Hunan Agricultural University, Changsha, China

Abstract: Urban air quality monitoring systems continuously generate high-frequency and multi-variable observational data, making it difficult to conduct effective analysis solely by relying on fixed thresholds or traditional statistical models. Existing PM2.5 prediction studies still exhibit deficiencies in integrating pollutant, meteorological, temporal, and site-specific information, and do not sufficiently account for factors such as short-term fluctuations, long-term temporal dependencies, missing observations, and sensor noise. To address these limitations, this study proposes a multi-source temporal convolutional network–gated recurrent unit attention network (MTGA-Net), built upon the open-source Beijing multi-site air quality dataset. The proposed architecture combines causal time-series convolution, gated recurrent units, and a temporal attention mechanism to forecast PM2.5 levels for the next 1 hour, 6 hours, and 24 hours. Experimental results demonstrate that MTGA-Net achieves robust predictive performance across different forecasting horizons, notably outperforming the Transformer encoder in medium- and long-term prediction tasks. Further ablation experiments reveal that meteorological features, the GRU module, and the TCN module are the key factors influencing overall model performance. Importantly, the model maintains strong predictive capability even under conditions of data missingness and noise interference. In summary, MTGA-Net can effectively enhance multi-temporal air quality prediction, providing reliable support for air quality forecasting and reliability assessment under incomplete monitoring data conditions.

Keywords: artificial intelligence; air quality monitoring; PM2.5 forecasting; temporal convolutional network; gated recurrent unit; attention mechanism

Received: 11 June 2026
Revised: 19 July 2026
Accepted: 29 July 2026
Published: 01 August 2026

1. Introduction

Ecological and environmental monitoring plays a critical role in pollution control, ecological risk assessment, public health protection, and sustainable urban development. Continuous observation of environmental parameters—including air, water, and soil—enables timely detection of environmental changes, source identification of contaminants, and supports evidence-based governance strategies. Conventional monitoring approaches rely heavily on manual sampling, fixed-location stations, and statistical analysis [1]. However, due to the pronounced spatiotemporal variability of environmental data and the intricate interdependencies among pollutants and meteorological variables, such methods often fall short in accurately characterizing environmental change patterns. Pollutants further interact dynamically with meteorological conditions. Consequently, static threshold criteria fail to capture the evolving nature of environmental quality, and traditional statistical models struggle to represent complex correlations within large-scale monitoring datasets.

Advances in artificial intelligence have introduced powerful new tools for analyzing environmental observational data. Machine learning techniques are now widely employed for tasks including air pollution forecasting, water quality categorization, ecological risk evaluation, and interpretation of remote sensing imagery. Deep learning architectures have significantly strengthened modeling capacity for heterogeneous and high-dimensional environmental data. In recent years, convolutional neural networks have been successfully applied to land cover mapping, vegetation dynamics monitoring, and remote sensing image analysis [2]. Recurrent neural networks and their variants are commonly adopted for forecasting water quality indicators and atmospheric pollutant levels. Air quality forecasting represents a key application domain for artificial intelligence. Dense air quality monitoring networks continuously generate multi-source environmental records, forming a robust foundation for model development and validation. Accurate forecasting of particulate matter levels thus supports early warning systems and informs decision-making in urban environmental management.

Existing studies exhibit several limitations [3]. Some focus exclusively on historical records of a single contaminant and neglect cross-factor interactions among multiple environmental variables, thereby constraining the model's capacity to reflect integrated pollution dynamics. Moreover, short-term volatility and long-term trends are frequently difficult to jointly capture within a unified modeling framework, although real-world air quality evolution is inherently driven by concurrent influences. A standalone time series architecture typically cannot comprehensively integrate these diverse temporal scales and drivers. Additionally, most current evaluations assume complete, clean, and homogeneous data—overlooking common practical challenges such as missing observations, measurement noise, and site-specific heterogeneity across monitoring locations. These issues are prevalent in operational monitoring environments and may substantially degrade predictive reliability.

To address these challenges, this study leverages an open-source, multi-site air quality dataset to investigate the application of artificial intelligence algorithms for urban air quality forecasting. A multi-source time series prediction framework is developed, incorporating atmospheric contaminant measurements, meteorological variables, temporal features, and site-specific attributes. The architecture integrates temporal convolution, gated recurrent units, and an attention mechanism to jointly extract spatiotemporal patterns underlying air quality variation. Performance is rigorously benchmarked against multiple established baseline models. Experimental results confirm the model's effectiveness and practical utility, offering novel methodological insights for deploying artificial intelligence in environmental monitoring systems.

2. Open-Source Data and Problem Definition

2.1. Data Source

This study employed the publicly accessible Beijing multi-site air quality dataset hosted in the UCI Machine Learning Repository [3]. The dataset comprises 420,768 hourly observations recorded by 12 air quality monitoring stations across Beijing from March 1, 2013 to February 28, 2017. In time-series forecasting and cross-site evaluation tasks, each station was treated as an independent spatial unit.

The dataset includes measurements of atmospheric pollutants, meteorological variables, and spatiotemporal metadata [4]. PM2.5 mass concentrations serve as the primary prediction target; input features encompass other pollutant measurements—including PM10, sulfur dioxide, nitrogen dioxide, carbon monoxide, and ozone—as well as meteorological parameters such as temperature, atmospheric pressure, and wind speed. Temporal attributes—including year, month, day, and hour—are incorporated to capture diurnal and seasonal patterns in air quality variation.

2.2. Data Preprocessing

Data preprocessing is performed independently for each monitoring station while preserving the original hourly temporal resolution. For continuous missing intervals of no more than six hours, linear interpolation is applied [5]. For longer missing intervals, imputation uses the median value observed at the same station, month, and hour-of-day within the training set. Samples for which the PM2.5 target variable remains undefined after preprocessing are excluded from both model training and performance evaluation.

Anomaly detection is applied to continuous meteorological and environmental variables, including pollutant levels, wind speed, and precipitation. Observations falling outside the interquartile range–based thresholds undergo additional scrutiny using neighboring time points [6]. Corrections are applied only when deviations are inconsistent with the underlying temporal pattern, thereby minimizing misclassification of genuine extreme events.

All continuous input variables are standardized using the mean and standard deviation computed exclusively from the training set; these same parameters are consistently applied to the validation and test sets. Wind direction is encoded via one-hot representation; hour-of-day and month are embedded using sine and cosine transformations to preserve cyclical characteristics; site identifiers are represented by unique numeric codes [7].

Following preprocessing, input sequences are generated via a sliding window approach. By default, each sample comprises observations from the preceding 24 hours [5]. In sensitivity analysis, model performance is further evaluated under alternative window lengths—specifically, 12-hour and 48-hour horizons—to assess temporal dependency effects.

2.3. Prediction Task

This study formulates the prediction task as a multivariate time series regression problem. The model leverages historical air quality measurements over the preceding 24 hours, concurrent meteorological observations, temporal features, and monitoring station metadata to forecast $PM_{2.5}$ levels at horizons of 1 hour, 6 hours, and 24 hours, with performance assessed independently for each forecasting interval [8].

The input sequence comprises only observed data available up to the prediction time point [9]. For a forecast horizon h, the target value corresponds to the $PM_{2.5}$ level recorded h hours after the conclusion of the input window.

2.4. Dataset Partition

The dataset is partitioned chronologically into training, validation, and test sets with proportions of 70%, 15%, and 15%, respectively. All stations share identical temporal boundaries to ensure consistent time coverage across subsets. No random shuffling is applied during partitioning to prevent temporal information leakage, particularly the inclusion of future observations in the training set [10].

Preprocessing parameters—including those for missing-value imputation and feature standardization—are derived exclusively from the training set. Historical observations preceding the validation and test set boundaries may be used to construct input sequences; however, the corresponding target values are excluded from model training [11].

3. Proposed Artificial Intelligence Method

3.1. Overall Framework

This study proposes a multi-source TCN-GRU attention network (MTGA-Net) for multi-time series $PM_{2.5}$ forecasting. The model integrates multi-source environmental data and combines time convolutional networks, gated recurrent units, and time attention mechanisms to capture air quality dynamics across multiple temporal scales. It takes the preceding 24-hour environmental observation sequence as input and generates forecasts for $PM_{2.5}$ levels at horizons of 1 hour, 6 hours, and 24 hours, with each horizon modeled by a dedicated subnetwork [3].

The model first encodes pollutant, meteorological, temporal, and spatial station information into a unified input sequence. A time convolutional network extracts short-term temporal patterns, while the GRU layer captures longer-range sequential dependencies [1]. A time attention mechanism then adaptively weights and fuses hidden representations according to temporal relevance, and final predictions are produced via a fully connected output layer.

3.2. Multi-Source Environmental Feature Construction

At time stamp t, the input vector comprises four categories of features:

$$x_t = [x_t^p; x_t^m; x_t^c; x_t^s], X_t = [x_{t-L+1}, \ldots, x_t] \in \mathbb{R}^{L\times F}, L = 24. \quad (1)$$

Among them, x_t^p denotes pollutant-related measurements, including PM2.5, PM10, SO_2, NO_2, CO, and O_3 levels; x_t^m denotes meteorological variables, such as temperature, atmospheric pressure, dew point temperature, precipitation, wind speed, and encoded wind direction; x_t^c denotes temporal attributes, derived from periodic encodings of hour and month; x_t^s denotes station-specific attributes [8]. F denotes the total number of encoded features.

In the standard prediction setting, site information is represented using learnable embedding vectors [7]. In the holdout site setting, the target site is excluded from training, so its corresponding embedding is omitted. Following feature concatenation, the resulting vector is projected into a unified latent space dimension to support time series modeling.

3.3. Short-Term Temporal Feature Extraction

The input sequence is first processed by a time convolutional network comprising a one-dimensional causal convolution layer. Both convolutional layers use third-order convolution kernels, with dilation rates of 1 and 2, respectively. At each time step, the model relies solely on current and historical observation data.

The output of the first convolutional layer is defined as:

$$z^{(l)} = \sigma\left[\mathrm{Conv}_{d_l}\left(z^{(l-1)}\right)\right] + z^{(l-1)}. \quad (2)$$

In Equation (2), d_l denotes the dilation rate and σ denotes the activation function. A residual connection preserves original feature information and facilitates feature propagation. The TCN output retains the temporal order of the input sequence and effectively captures short-term variations and local patterns in pollutant level dynamics.

3.4. Long-Term Dependency Modeling

The TCN output is fed into a GRU layer to model temporal dependencies. The hidden state at time step i is computed as follows:

$$h_i = \mathrm{GRU}\left(z_i^{TCN}, h_{i-1}\right), i = 1, \ldots, L. \quad (3)$$

Here, z_i^{TCN} denotes the TCN output at position i, and h_i denotes the GRU hidden state at the corresponding time step. The full sequence of hidden states is then passed to the time attention layer to integrate information across historical time steps [12].

3.5. Attention-Based Temporal Fusion

The time attention layer computes attention weights for each GRU hidden state [13].

$$e_i = v_a^T \tanh(W_a h_i + b_a). \quad (4)$$

Following normalization, the attention scores are applied to weight and fuse the hidden states, yielding the time context vector [14].

$$\alpha_i = \frac{\exp(e_i)}{\sum_{j=1}^{L}\exp(e_j)}, c = \sum_{i=1}^{L} \alpha_i h_i. \quad (5)$$

In Equation (4) and (5), W_a, b_a, and v_a denote learnable parameters; α_i represents the normalized attention weight assigned to the i-th historical time step; and c denotes the fused temporal representation. The attention weights are retained to assess the model's focus across historical time points and, when integrated with the SHAP method, enable quantitative evaluation of each input variable's contribution to the prediction outcome.

3.6. Prediction and Optimization

Ultimately, the time context vector yields the predicted $PM_{2.5}$ level via the fully connected layer. MTGA-Net is trained using the Huber loss function:

$$\hat{y}_{t+h} = W_o c + b_o, h \in \{1,6,24\}, \mathcal{L}_{Huber} = \frac{1}{N} \sum_{n=1}^{N} \begin{cases} \frac{1}{2} r_n^2, & \mid r_n \mid \leq \delta, \\ \delta\left(\mid r_n \mid - \frac{1}{2}\delta\right), & \mid r_n \mid > \delta, \end{cases} r_n = y_n - \hat{y}_n. \quad (6)$$

Here, h denotes the prediction duration, N denotes the number of training samples, r_n denotes the prediction residual, and δ denotes the Huber loss threshold. During training, standardized $PM_{2.5}$ values are used; predicted outputs are rescaled to the original measurement units prior to evaluation.

4. Experimental Design

4.1. Baseline Models

MTGA-Net was compared with four benchmark methods. The Persistence method uses the most recent observed PM2.5 measurement as the prediction result. XGBoost is a machine learning method based on tree models, which expands the environmental features in the input window and uses them for prediction. GRU is used to learn the dynamic change patterns in time series and processes continuous observed data through a recurrent structure. The Transformer Encoder models sequence relationships using self-attention mechanisms, including position encoding, two encoder layers, four attention heads, and a fully connected regression layer.

MTGA-Net, as the proposed method in this paper, is compared with the aforementioned benchmark models under identical data partitioning, input window length, prediction target, and preprocessing conditions. Model hyperparameters are selected based on validation set MAE, and final evaluation is performed on the test set.

4.2. Evaluation Metrics

Predictive performance is evaluated using MAE, RMSE, and R^2. MAE quantifies the average absolute deviation between predicted and observed values. RMSE places greater weight on larger prediction errors. R^2 reflects the proportion of variance in the observed data explained by the model. MAE and RMSE serve as primary evaluation metrics, while R^2 is reported for supplementary interpretation.

4.3. Implementation Details

The model uses 24-hour historical observations as the default input window, with separate evaluations conducted for 12-hour and 48-hour windows. During training, input samples are processed in batches of 128. The hidden dimension is configured to 64, and a dropout rate of 0.2 is applied. Optimization employs an initial learning rate of 0.001, and the Huber loss function is used with a threshold of 1.0.

Training proceeds for up to 100 epochs, terminating early if the validation MAE fails to improve over 10 consecutive epochs. Parameter optimization is performed using the Adam algorithm. Each experimental configuration is repeated five times with distinct random seeds. All experiments were implemented in Python using PyTorch and executed on a workstation featuring an Intel Core i7-12700K CPU, 32 GB RAM, and an NVIDIA RTX 3060 GPU.

4.4. Experimental Tasks

The experiment assessed model performance across varying prediction horizons, input feature configurations, and realistic operational conditions. First, the model predicted $PM_{2.5}$ levels for lead times of 1 hour, 6 hours, and 24 hours. Next, input features were incrementally expanded—from historical $PM_{2.5}$ values to include pollutant concentrations, meteorological parameters, temporal indicators, and site-specific attributes—to evaluate how information diversity influences forecasting accuracy. Subsequently, model robustness was examined under common monitoring challenges, including partial data absence, measurement noise, and sensor relocation. Finally, time attention weights and SHAP analysis were applied to identify temporally salient historical intervals and quantify the contribution of each variable to the prediction.

5. Results and Discussion

5.1. Overall Prediction Performance

Table 1 compares the performance of MTGA-Net with four benchmark models in 1-hour, 6-hour, and 24-hour prediction tasks. Except for the persistence method, all trainable models were evaluated using five different random seeds, and results are reported as mean ± standard deviation. As the persistence method relies on fixed observational values, its outputs remain invariant across repeated trials under identical test conditions, yielding a standard deviation of zero. A paired t-test was conducted to assess performance differences between MTGA-Net and the best-performing benchmark model—Transformer Encoder—under matched random seed conditions [11]. All statistical parameters, including difference estimates and confidence intervals, were computed from unrounded individual trial results.

Table 1. Prediction Performance at Different Horizons (Mean ± SD, N=5).

Model	1-h MAE	1-h RMSE	1-h R^2	6-h MAE	6-h RMSE	6-h R^2	24-h MAE	24-h RMSE	24-h R^2
Persistence	11.6 ± 0.0	18.9 ± 0.0	0.910 ± 0.000	24.8 ± 0.0	36.5 ± 0.0	0.713 ± 0.000	37.9 ± 0.0	52.8 ± 0.0	0.446 ± 0.000
XGBoost	10.2 ± 0.4	16.7 ± 0.6	0.929 ± 0.005	20.8 ± 0.8	30.9 ± 1.0	0.795 ± 0.012	30.1 ± 1.0	43.3 ± 1.4	0.627 ± 0.017
GRU	9.5 ± 0.3	15.4 ± 0.5	0.940 ± 0.004	18.7 ± 0.6	28.1 ± 0.8	0.831 ± 0.009	27.8 ± 0.8	40.2 ± 1.1	0.679 ± 0.014
Transformer Encoder	9.0 ± 0.3	14.2 ± 0.4	0.949 ± 0.003	17.9 ± 0.5	26.9 ± 0.7	0.845 ± 0.008	26.9 ± 0.4	39.0 ± 0.8	0.698 ± 0.010
MTGA-Net	8.7 ± 0.2	14.4 ± 0.5	0.948 ± 0.004	17.2 ± 0.4	26.2 ± 0.6	0.854 ± 0.007	25.8 ± 0.3	37.8 ± 0.7	0.717 ± 0.009

Note: MAE and RMSE are reported in μg/m³.

MTGA-Net achieved the lowest MAE across all three prediction horizons, though it did not consistently outperform other metrics. In the 1-hour prediction, Transformer Encoder exhibited a marginally lower RMSE and higher R^2, while MTGA-Net reduced MAE by 0.32 μg/m³, with a 95% confidence interval of −0.02 to 0.66 μg/m³ ($p = 0.061$), indicating no statistically significant difference. For the 6-hour horizon, MTGA-Net reduced MAE by 0.70 μg/m³ (95% CI: 0.22–1.18 μg/m³; $p = 0.016$); for the 24-hour horizon, the reduction was 1.12 μg/m³ (95% CI: 0.50–1.74 μg/m³; $p = 0.007$).

The performance advantage of MTGA-Net in short-term forecasting is modest. However, improvement becomes increasingly pronounced with longer prediction horizons. Transformer Encoder effectively captures sequential patterns over shorter time scales, whereas MTGA-Net—integrating causal convolution and recurrent architecture—demonstrates enhanced stability and accuracy in extended-range predictions.

5.2. Multi-Horizon and Feature-Combination Results

Figure 1(a) illustrates the variation in prediction error across models under different time horizons. As the prediction window extends, all models exhibit a consistent upward trend in error. The persistent method shows the most pronounced deterioration: its MAE rises from 11.6 ± 0.0 μg/m³ at the 1-hour horizon to 37.9 ± 0.0 μg/m³ at the 24-hour horizon. In comparison, XGBoost demonstrates greater robustness, with only marginal performance degradation over longer horizons, while deep learning models sustain comparatively lower errors even in extended forecasting tasks. At the 1-hour horizon, MTGA-Net and the Transformer Encoder achieve similar accuracy; however, their performance divergence widens progressively as the prediction duration increases.

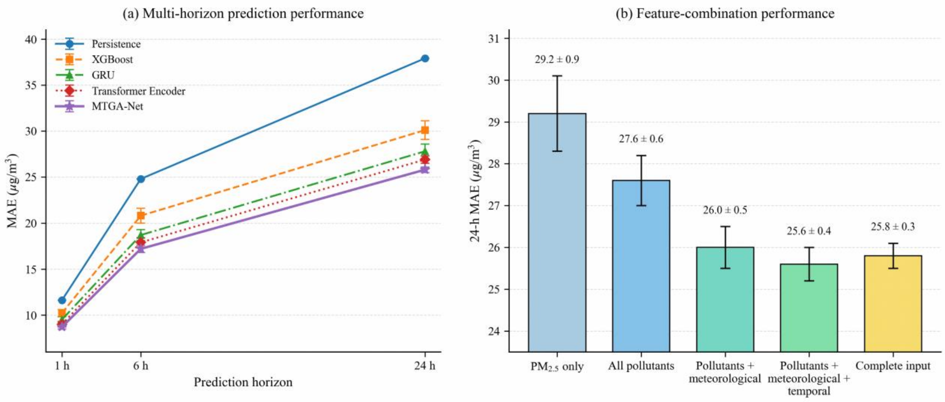

Figure 1. Results of Multi-Time-Scale Prediction and Feature Combination: (A) MAE of Five Models in 1-Hour, 6-Hour and 24-Hour Prediction Tasks; (B) MAE of 24-Hour Prediction under Different Input Feature Combinations, Including Historical PM2.5, All Pollutants, Pollutant and Meteorological Variable Combinations, Pollutant and Meteorological and Time Variable Combinations, and Complete Input Features. Error Bars Represent the SD of the Results of Five Experiments.

Figure 1(b) presents the impact of varying input feature configurations on 24-hour PM2.5 prediction accuracy. Using historical PM2.5 data alone yields an MAE of 29.2 ± 0.9 μg/m³. Incorporating additional pollutant concentrations reduces the MAE to 27.6 ± 0.6 μg/m³. Further inclusion of meteorological variables lowers it to 26.0 ± 0.5 μg/m³; adding temporal features brings it down to 25.6 ± 0.4 μg/m³. Finally, integrating site-specific embedding information results in an MAE of 25.8 ± 0.3 μg/m³.

These findings suggest that multi-pollutant data contributes meaningful supplementary signals for PM2.5 forecasting, while meteorological variables deliver further measurable gains—likely attributable to their influence on atmospheric dispersion, transport, and accumulation dynamics. The temporal features yield only modest improvement, indicating limited incremental value in this context. Site embedding does not reduce MAE further but leads to slight improvements in RMSE and R^2, underscoring that distinct input features exert heterogeneous effects across evaluation metrics.

5.3. Ablation Study

Table 2 presents the impact of different model components and input features on 24-hour prediction performance.

Table 2. Ablation Results for 24-Hour Prediction (Mean ± SD, N=5).

Variant	MAE	RMSE	R^2	MAE change	p-value
MTGA-Net	25.8 ± 0.3	37.8 ± 0.7	0.717 ± 0.009	-	-
Without TCN	27.0 ± 0.5	39.1 ± 0.8	0.697 ± 0.011	+1.2	0.012
Without GRU	27.6 ± 0.6	40.0 ± 1.0	0.683 ± 0.014	+1.8	0.005
Without attention	26.3 ± 0.4	38.2 ± 0.8	0.711 ± 0.010	+0.5	0.084
Without meteorological features	28.2 ± 0.7	40.8 ± 1.1	0.669 ± 0.015	+2.4	0.003
Without temporal features	26.5 ± 0.5	38.6 ± 0.9	0.704 ± 0.012	+0.7	0.041
Without station features	25.6 ± 0.4	38.1 ± 0.8	0.712 ± 0.010	−0.2	0.274

Removing the TCN increased the MAE by 1.2 μg/m³; subsequent removal of the GRU further increased the MAE by 1.8 μg/m³, confirming that both temporal modeling modules contribute to prediction accuracy. The GRU demonstrates a stronger capacity for capturing long-term temporal patterns. In contrast, removing the attention mechanism raised the MAE by only 0.5 μg/m³, with no statistically significant difference, suggesting its marginal contribution to overall performance.

Regarding input features, meteorological data exerts the strongest influence: its removal increased the MAE by 2.4 μg/m³. Excluding time information also degraded prediction accuracy, though to a lesser extent [2]. Omitting station information led to a slight reduction in MAE but concurrently increased RMSE and decreased R^2, indicating inconsistent effects across evaluation metrics.

5.4. Missing-Data, Noise, and Held-Out-Station Evaluation

Figure 2(a) evaluates the prediction performance of MTGA-Net under varying data quality conditions. With complete input, the model achieves an MAE of 25.8 ± 0.3 μg/m³. As observational gaps increase, prediction error rises accordingly. Under random missingness, the MAE increases from 26.3 ± 0.4 μg/m³ at a 5% missing ratio to 29.4 ± 0.8 μg/m³ at 20%. Continuous missing intervals exert a more pronounced effect: MAE reaches 28.2 ± 0.7 μg/m³ for 6-hour gaps and 31.7 ± 1.0 μg/m³ for 12-hour gaps. Prolonged absence of observations impairs the model's ability to capture temporal trends, leading to larger prediction errors.

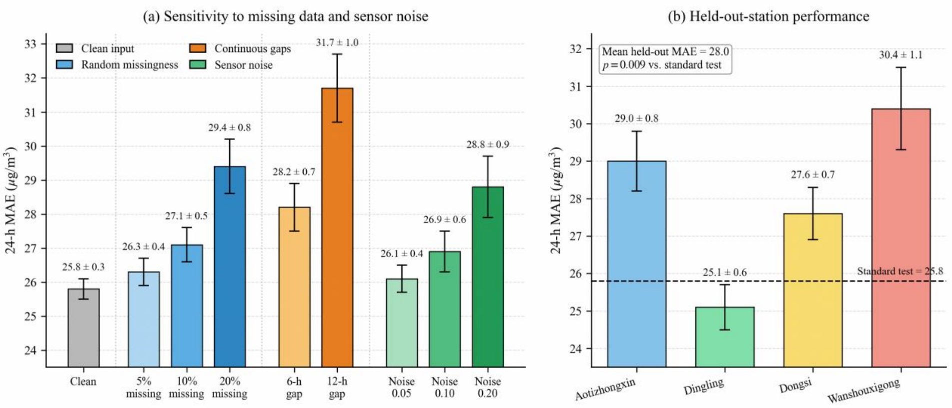

Figure 2. Missing-data, Noise, and Held-Out-Station Results: (A) MAE under Random Missingness, Continuous Missing Blocks, and Sensor Noise; (B) MAE for Four Held-Out Monitoring Stations. Error Bars Represent SD over Five Runs.

Noise robustness testing reveals a comparable pattern. With standardized noise levels of 0.05, 0.10, and 0.20, the MAE increases progressively from 26.1 ± 0.4 μg/m³ to 28.8 ± 0.9 μg/m³. This monotonic rise confirms that degradation in input signal fidelity directly compromises prediction accuracy. Collectively, MTGA-Net demonstrates moderate resilience to both missing data and additive noise, though these experiments primarily assess model stability rather than operational performance.

Figure 2(b) assesses generalization to previously unseen monitoring sites. In turn, the four sites—Aotizhongxin, Dingling, Dongsi, and Wanshouxigong—were excluded from training, and their site-specific embeddings were omitted during inference. The corresponding MAEs were 29.0 ± 0.8, 25.1 ± 0.6, 27.6 ± 0.7, and 30.4 ± 1.1 μg/m³. All values exceed those obtained in standard 24-hour evaluation ($p = 0.009$), indicating reduced accuracy under zero-shot site adaptation. Among them, Dingling yields the lowest error, whereas Wanshouxigong exhibits the highest, suggesting heterogeneity in site-level predictability across the monitoring network.

5.5. Interpretability and Limitations

The normalized average absolute SHAP values indicate that historical PM2.5 exerts the strongest influence on prediction outcomes, contributing 29.8%. PM10, wind speed, NO2, and dew point temperature follow, with contributions of 18.6%, 10.9%, 9.7%, and 8.4%, respectively [15].

The remaining variables account for 22.6% of the predictive influence. Overall, historical PM2.5 serves as the primary information source for model predictions, while other pollutants and meteorological factors provide supplementary predictive value. Time attention analysis reveals that the model leverages not only recent temporal data but also information from earlier time windows: the most recent time point receives 34% attention weight, whereas data from 20 to 24 hours prior accounts for 22%.

Model stability was evaluated through repeated experiments, uncertainty analysis, significance testing, ablation studies, perturbation tests, and hold-out site validation [13, 15]. It should be noted that the dataset comprises only Beijing's historical air quality monitoring records from 2013 to 2017; model performance has not been validated in other cities or under real-time operational conditions. Spatial representation is limited to site identifiers, without incorporation of precise geographic coordinates or localized emission characteristics. The model is intended for auxiliary analytical purposes and cannot substitute calibrated instrumental monitoring or official environmental regulatory assessments.

6. Conclusion

This study leveraged publicly available multi-site air quality data from Beijing to evaluate artificial intelligence approaches for forecasting fine particulate matter levels. The dataset comprised hourly observations from 12 monitoring stations spanning 2013 to 2017 and included measurements of atmospheric constituents, meteorological variables, temporal metadata, and site identifiers. The primary forecasting objective was to estimate future fine particulate matter levels, using auxiliary environmental variables to inform predictive trends.

To forecast fine particulate matter levels at horizons of 1 hour, 6 hours, and 24 hours, this work introduces the MTGA-Net architecture. This design integrates temporal convolutional networks, gated recurrent units, and an attention mechanism to jointly exploit short-term dynamics and longer-term temporal dependencies. Comparative experiments were conducted against baseline methods—including persistence forecasting, XGBoost, GRU-only, and Transformer Encoder models—and evaluated across multiple input configurations: varying feature subsets, simulated data gaps, synthetic noise perturbations, and cross-site generalization scenarios.

Experimental results indicate that MTGA-Net delivers robust performance across all forecasting horizons, with particularly strong gains observed at extended time scales. Incorporating auxiliary variables—such as co-pollutants, meteorological parameters, and temporal features—consistently improved accuracy relative to models relying solely on historical fine particulate matter measurements. Structural ablation revealed that the temporal convolutional network and gated recurrent unit components contributed substantially to predictive gains, whereas the attention module yielded only marginal improvements. Site identifiers provided supplementary spatial context, though their influence on evaluation metrics varied across indicators.

The model demonstrated resilience under conditions of missing data and measurement noise, maintaining stable predictive performance. Cross-site validation revealed variability in forecasting outcomes across locations, suggesting limitations in geographic transferability and highlighting opportunities for architectural refinement. Interpretability analyses—including SHAP attribution and attention weight inspection—identified historical fine particulate matter values, coarse particulate matter (PM10), wind speed, nitrogen dioxide (NO_2), dew point temperature, and recent observations as the most influential inputs.

The empirical analysis relied exclusively on historical air quality records from Beijing and did not address real-time operational deployment or broader urban contexts. Future work may extend the methodology to additional geographical regions or adapt it to other environmental monitoring domains—including water quality, soil condition assessment, and remote sensing–based ecological monitoring. Integration of spatially resolved emission inventories, high-resolution topographic data, and calibrated sensor outputs could further strengthen practical applicability in operational monitoring systems.

References

1. M. J. Zoqi, A. D. Aliaqa, S. J. Rasouli, and M. R. M. Daneshvar, "Daily PM2.5 concentration forecasting in Mashhad metropolis using long short-term memory (LSTM) neural networks and feature importance analysis," *Air Quality, Atmosphere & Health*, vol. 19, no. 2, p. 27, 2026.

2. S. Jayaraman, N. T., A. S., and S. G., "Enhancing urban air quality prediction using time-based-spatial forecasting framework," *Scientific Reports*, vol. 15, no. 1, p. 4139, 2025.
3. P. Banethiwal and R. Sharma, "Comparative Evaluation of Machine and Deep Learning Models for Air Quality Index Prediction in Jaipur, India," in *E3S Web of Conferences*, vol. 694, EDP Sciences, 2026.
4. R. Rakholia, Q. Le, K. Vu, B. Q. Ho, and R. S. Carbajo, "Accurate PM2.5 urban air pollution forecasting using multivariate ensemble learning Accounting for evolving target distributions," *Chemosphere*, vol. 364, p. 143097, 2024.
5. P. Patel, S. Patel, K. Shah, K. Desai, S. Patel, M. Shah, and S. Patel, "A systematic study on PM2.5 and PM10 concentration prediction in air pollution using machine learning and deep learning model," *Environmental Chemistry and Ecotoxicology*, vol. 7, pp. 1401–1415, 2025.
6. A. Ding, A. Zhang, T. M. Fu, Y. Huang, Q. Chen, Y. Chen, et al., "Diffusion-based Probabilistic Air Quality Forecasting with Mechanistic Insight," *arXiv preprint arXiv:2603.21131*, 2026.
7. C. J. Huang and P. H. Kuo, "A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities," *Sensors*, vol. 18, no. 7, p. 2220, 2018.
8. F. A. Panaite, C. Rus, M. Leba, A. C. Ionica, and M. Windisch, "Enhancing air-quality predictions on university campuses: A machine-learning approach to PM2.5 forecasting at the University of Petroșani," *Sustainability*, vol. 16, no. 17, p. 7854, 2024.
9. C. M. Rosca, M. Carbureanu, and A. Stancu, "Data-driven approaches for predicting and forecasting air quality in urban areas," *Applied Sciences*, vol. 15, no. 8, p. 4390, 2025.
10. Z. Wang, X. Wu, and Y. Wu, "A spatiotemporal XGBoost model for PM2.5 concentration prediction and its application in Shanghai," *Heliyon*, vol. 9, no. 12, 2023.
11. 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, 2025.
12. G. Huang, X. Li, B. Zhang, and J. Ren, "PM2.5 concentration forecasting at surface monitoring sites using GRU neural network based on empirical mode decomposition," *Science of the Total Environment*, vol. 768, p. 144516, 2021.
13. S. Zhang and M. Yu, "Enhanced urban PM2.5 prediction: Applying quadtree division and time-series transformer with WRF-chem," *Atmospheric Environment*, vol. 337, p. 120758, 2024.
14. D. Patel, M. Kulwant, S. Shirin, A. Kumar, M. A. Ansari, and A. K. Yadav, "Artificial intelligence for air quality and control systems: Status and future trends," in *Modeling and Simulation of Environmental Systems*, pp. 133–152, 2022.
15. E. X. Neo, K. Hasikin, K. W. Lai, M. I. Mokhtar, M. M. Azizan, H. F. Hizaddin, and S. A. Razak, "Artificial intelligence-assisted air quality monitoring for smart city management," *PeerJ Computer Science*, vol. 9, p. e1306, 2023.