

Bidirectional Temporal Attention Networks for Missing Data Recovery in Urban Air Quality Monitoring

Aklesh Shadaf¹, Ashish Singh², Sandeep³

¹Computer Science & SRMS College of Engineering & Technology

²Computer Science & SRMS College of Engineering & Technology

³Computer Science & SRMS College of Engineering & Technology

Abstract – Urban air quality sensor networks produce continuous multi-pollutant time series that are routinely interrupted by hardware failures, power outages, and scheduled maintenance, creating missing data gaps that compromise downstream analysis, regulatory reporting, and health exposure assessment. Existing imputation approaches — from classical statistical interpolation to recurrent neural networks and variational autoencoders — predominantly exploit only past observations and fail to leverage future context that is available at imputation time. In this paper we propose BiTAN, a Bidirectional Temporal Attention Network that applies stacked self-attention encoders independently in the forward and backward temporal directions and fuses their representations through a learned cross-directional layer. BiTAN incorporates variable-type-specific learnable missing embeddings, motivated by recent variational approaches [1], enabling the model to represent the absence of each pollutant or meteorological variable distinctively. Evaluation on EPA Air Quality System data from 12 Chicago-area monitoring stations under multiple missing-data regimes demonstrates 17–21% RMSE reduction over recurrent baselines, 17.9% reduction over the variational baseline [1], and competitive computational efficiency. Ablation experiments confirm the contribution of bidirectionality, the fusion layer, learnable embeddings, and positional encoding individually.

Key Words: *bidirectional attention, air quality monitoring, missing data imputation, self-attention network, learnable embeddings, urban sensor arrays, temporal deep learning.*

1. INTRODUCTION

Urban air quality monitoring is central to public health protection, regulatory compliance, and environmental policy. Agencies such as the U.S. Environmental Protection Agency (EPA) operate dense networks of

ground-level sensors measuring PM_{2.5}, PM₁₀, NO₂, O₃, CO, and associated meteorological variables including temperature, relative humidity, wind speed, and atmospheric pressure. These networks are indispensable for tracking pollution hotspots, evaluating emission-control programmes, issuing public health advisories, and calibrating atmospheric dispersion models. They also provide ground-truth data for validating satellite-derived aerosol retrievals and informing climate research. Despite their central role, real-world sensor networks are far from perfect data sources: hardware malfunctions, communication failures, power disruptions, calibration drift, and routine maintenance activities create persistent gaps in temporal measurements, with per-station missing rates typically ranging from 3% to 20% annually [2, 5].

The consequences of missing data in air quality records are significant across multiple application domains. Health exposure models that rely on continuous hourly pollutant readings produce biased estimates when gaps are filled naively, potentially mischaracterising exposure during high-pollution episodes when sensors are most likely to malfunction from thermal stress or fouling. Regulatory compliance checks require complete annual averages; gaps inflate or deflate computed averages depending on their temporal distribution. Forecasting models trained on gapped data exhibit systematic underperformance on the hours surrounding observed gaps.

Recovering missing values accurately demands a method that (a) captures the complex nonlinear dynamics of pollutant time series including diurnal cycles, weekly patterns, and episodic events; (b) handles variable-length, irregular gaps without degrading across gap durations; (c) exploits contextual information from all available observations, including those that occur after the missing interval; and (d) scales efficiently to deployment on hundreds of monitoring stations. Classical approaches — mean substitution, linear

interpolation, seasonal-trend decomposition — satisfy none of these requirements adequately for modern monitoring networks [6, 7]. Recurrent neural networks address requirements (a) and (b) but process data in a single temporal direction and cannot exploit post-gap observations [3]. Self-attention models such as SAITS [4] apply bidirectional masking during training but still produce causal encodings at inference. Variational autoencoders with learnable missing embeddings [1] achieve state-of-the-art accuracy but similarly remain single-direction models.

In this paper we propose BiTAN, a Bidirectional Temporal Attention Network that directly addresses the directionality limitation. BiTAN processes each monitoring station's multivariate time window through two independent self-attention encoder stacks — one in the forward direction and one in the backward direction — and fuses their representations through a learned cross-directional layer before projecting to imputed values. BiTAN retains the learnable variable-type-specific missing embeddings introduced in [1], which allow the model to distinguish the absence of PM_{2.5} from the absence of wind speed at the input level, providing a more informative starting point for the attention mechanism. Our main contributions are: (1) a bidirectional self-attention architecture tailored to multivariate environmental time series with structured missing data; (2) a cross-directional fusion layer that combines forward and backward representations through learned gating; (3) comprehensive evaluation across multiple missingness regimes showing consistent improvements over both recurrent and variational baselines; and (4) an ablation study isolating the contribution of each architectural component.

The remainder of the paper is organised as follows. Section 2 reviews related work on statistical, recurrent, attention-based, and variational imputation methods. Section 3 details the BiTAN architecture and training procedure. Section 4 presents experimental results and ablation analyses. Section 5 discusses implications and limitations. Section 6 concludes.

2. RELATED WORK

2.1 Statistical and Classical Methods

Classical approaches to missing data imputation in environmental monitoring include deterministic methods — mean substitution, last-observation-carried-forward, and linear interpolation — and model-based methods

such as seasonal decomposition and Kalman smoothing. Hadeed et al. [2] conducted a systematic evaluation of six imputation techniques for short-term pollutant monitoring gaps, finding that temporal methods outperform spatial interpolation when missing intervals are under 48 hours. Junninen et al. [7] compared deterministic and stochastic approaches across European monitoring networks, establishing benchmarks that remain widely cited. Moshenberg et al. [8] proposed spectral methods exploiting the periodic structure of atmospheric time series for gap-filling. Wang et al. [11] demonstrated imputation improvements for high-altitude monitoring data on the Qinghai-Tibetan plateau using ensemble regression. While computationally inexpensive, classical methods fail to capture nonlinear dynamics, cross-variable dependencies, and the information contained in post-gap observations.

2.2 Recurrent Neural Networks

Recurrent architectures were among the first deep learning methods applied to irregular time series imputation. GRU-D [3] augments gated recurrent units with trainable exponential decay rates that allow hidden states to decay toward variable-specific means in proportion to the elapsed time since the last observation. This formulation handles irregular sampling gracefully but processes observations strictly left-to-right. Cao et al. proposed BRITS, which extends bidirectional RNNs with consistency loss between forward and backward passes. Kim et al. [5] applied deep feedforward and LSTM networks to time-series air quality imputation, demonstrating substantial gains over regression baselines on Korean monitoring data. Hua et al. [6] quantified how imputation quality propagates to downstream air quality prediction accuracy, motivating investment in accurate gap-filling methods. A shared limitation of recurrent approaches is their sequential computation, which limits parallelism and makes training slow on long sequences.

2.3 Attention-Based and Transformer Models

Self-attention mechanisms, popularised by the Transformer architecture, have been adapted for time series imputation. SAITS [4] formulates imputation as a joint reconstruction task using masked self-attention on observed entries, achieving state-of-the-art performance across multiple benchmark datasets. PatchTST and related patching approaches improve handling of long sequences through segment-level attention. However, these models typically apply attention in a single direction or use non-causal masking strategies during training that do not fully exploit post-gap context at

inference time. BiTAN differs from SAITS by explicitly maintaining separate forward and backward encoder stacks with independent parameters, allowing each directional encoder to specialise in extracting contextual signals appropriate to its direction, then combining them through a dedicated fusion mechanism.

2.4 Variational and Generative Approaches

Variational autoencoders have been adapted for missing data imputation through several mechanisms. Nazabal et al. [9] proposed HI-VAE, which extends the VAE likelihood to handle missing values across heterogeneous data types with structured missing patterns. Chen et al. [10] introduced conditional iterative sampling within a VAE to improve imputation consistency across modalities. Kataria et al. [1] introduced variable-type-specific learnable missing embeddings within a temporal VAE specifically designed for environmental time series with multiple temporal scales, achieving 15–20% RMSE improvements over standard VAE baselines on EPA AQS datasets and establishing a strong benchmark. BiTAN builds upon the learnable embedding insight of [1] while replacing the variational encoder-decoder structure with bidirectional self-attention, which is more parameter-efficient and computationally faster while providing comparable or superior accuracy on the benchmark datasets used in this study.

3. METHODOLOGY

3.1 Problem Formulation

Let $X \in \mathbb{R}^{T \times V}$ denote a multivariate time series of V variables over T discrete hourly time steps at a single monitoring station. The V variables comprise pollutant concentrations $\{PM_{2.5}, PM_{10}, NO_2, O_3, CO\}$ and meteorological measurements $\{temperature, relative\ humidity, wind\ speed, wind\ direction, atmospheric\ pressure\}$, partitioned into two type groups $g \in \{pollutant, meteorological\}$. Let $M \in \{0,1\}^{T \times V}$ be a binary observation mask where $M_{t,v}=0$ indicates a missing entry. The observed data are $X_{obs} = X \odot M$. The imputation goal is to estimate $\hat{X} \approx X$ at all positions where $M_{t,v}=0$, with $\hat{X}_{t,v} = X_{t,v}$ when $M_{t,v}=1$ (i.e., observed values are not altered). The window length $T = 48$ hours is chosen to capture both diurnal cycles and their daily interactions.

3.2 Variable-Type Learnable Missing Embeddings

Following [1], for each variable type $g \in \{pollutant, meteorological\}$ we maintain a learnable embedding vector $e_g \in \mathbb{R}^d$ that is used to represent missing entries of that type. For variable v of type $g(v)$ at time step t , the input representation is computed as:

$$\tilde{x}_{t,v} = M_{t,v} \cdot W_{obs} X_{t,v} + (1 - M_{t,v}) \cdot e_{g(v)}$$

where $W_{obs} \in \mathbb{R}^{d \times 1}$ is a learned linear projection applied to observed scalar values. The concatenated input tensor $\tilde{X} \in \mathbb{R}^{T \times Vd}$ is projected to a common token dimension d_{model} via a variable-specific linear layer, yielding the input sequence $Z \in \mathbb{R}^{T \times d_{model}}$ passed to the attention encoders.

3.3 Positional Encoding

Since air quality dynamics exhibit strong diurnal and weekly periodicities, we augment the token sequence with a learned positional encoding $PE \in \mathbb{R}^{T \times d_{model}}$ rather than the fixed sinusoidal encoding of the original Transformer. Learned encodings allow the model to capture the asymmetric importance of hour-of-day and day-of-week effects that are not well-represented by sinusoidal bases. The encoded sequence is $Z_{enc} = Z + PE$, which is fed identically to both the forward and backward encoder stacks.

3.4 Forward and Backward Self-Attention Encoders

The forward encoder F processes Z_{enc} left-to-right using a standard causal (masked) multi-head self-attention mechanism with $H = 4$ heads and $d_{model} = 64$. The backward encoder B processes the reversed sequence $Z_{enc}[T:0:-1]$ with an independent set of parameters, effectively attending right-to-left. Both encoders consist of $L = 4$ identical transformer blocks, each containing multi-head self-attention, position-wise feed-forward sublayers ($d_{ff} = 256$), layer normalisation, and dropout ($p = 0.1$). The forward encoder produces $F = \{f_1, \dots, f_T\} \in \mathbb{R}^{T \times d_{model}}$ and the backward encoder (after re-reversing its outputs) produces $B = \{b_1, \dots, b_T\} \in \mathbb{R}^{T \times d_{model}}$, both aligned to the original time axis.

3.5 Cross-Directional Fusion

Naïve concatenation or summation of forward and backward representations treats both directions symmetrically, ignoring that their relative usefulness varies with position within a gap. We introduce a gated cross-directional fusion layer:

$$g_t = \sigma(W_g[f_t; b_t] + c_g)$$

$$\hat{Z}_t = \text{LayerNorm}(g_t \odot f_t + (1 - g_t) \odot b_t + f_t \odot b_t)$$

where $[\cdot; \cdot]$ denotes concatenation, $W_g \in \mathbb{R}^{d_{\text{model}} \times 2d_{\text{model}}}$, c_g is a bias vector, σ is the sigmoid function, and $\odot$ denotes element-wise product. The gate g_t dynamically weights forward versus backward context at each time step, allowing the model to rely more heavily on post-gap forward context near the end of a gap and on pre-gap backward context near the beginning. The element-wise product term $f_t \odot b_t$ captures interactions between the two directional representations that are missed by simple linear combination.

3.6 Output Head and Training Objective

The fused sequence $\hat{Z} \in \mathbb{R}^{T \times d_{\text{model}}}$ is passed through a two-layer feed-forward head with ReLU activation to produce the imputed output $\hat{X} \in \mathbb{R}^{T \times V}$. The primary training loss is the mean squared error over observed entries:

$$\mathcal{L}_{\text{obs}} = (1/M) \cdot \sum_{(t,v): M_{(t,v)}=1} (\hat{X}_{(t,v)} - X_{(t,v)})^2$$

We additionally apply a self-supervised auxiliary loss by synthetically masking a random fraction (15%) of observed entries during training and computing MSE on those synthetic gaps, following [4]. The total loss is $\mathcal{L} = \mathcal{L}_{\text{obs}} + 0.5 \cdot \mathcal{L}_{\text{aux}}$. The model is trained with Adam ($\text{lr} = 5 \times 10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.98$), batch size 64, for 100 epochs with cosine annealing and early stopping on validation RMSE (patience 15 epochs). Training takes approximately 40 minutes on a single NVIDIA A100 GPU for the 12-station dataset.

4. EXPERIMENTS

4.1 Dataset

We use EPA AQS hourly measurements from 12 urban monitoring stations in the Chicago metropolitan area spanning January 2019 to December 2022 (four full years, 35,064 time steps per station), covering $\text{PM}_{2.5}$, PM_{10} , NO_2 , O_3 , CO, temperature, relative humidity, wind speed, and wind direction. Natural missing rates per station range from 4.1% to 16.3% per year. We evaluate under three synthetic missingness regimes applied on top of natural gaps: (a) MCAR — random missing completely at random; (b) MNAR-block — missing-not-at-random block gaps of 6–24 consecutive hours simulating extended outages; (c) MNAR-seasonal — elevated missing rate ($\times 2.5$) during December–February

to simulate winter maintenance. Each regime targets a 15% synthetic missing rate. The dataset is split temporally: 2019–2021 for training, Q1 2022 for validation, Q2–Q4 2022 for testing, ensuring no temporal data leakage.

4.2 Implementation Details and Baselines

BiTAN uses $d_{\text{model}} = 64$, $H = 4$ attention heads, $L = 4$ encoder layers, $d_{\text{ff}} = 256$, and dropout $p = 0.1$. We compare against: Mean Imputation (MI), Linear Interpolation (LI), Multiple Imputation by Chained Equations (MICE), GRU-D [3], SAITS [4], and Temporal-VAE [1]. All baselines use published hyperparameters where available, tuned on the validation set otherwise. All experiments are repeated with five random seeds; we report mean RMSE and MAE. Statistical significance is assessed by paired Wilcoxon signed-rank test at $p < 0.05$. Evaluation metrics are computed exclusively on synthetically masked positions to avoid confounding with the natural missingness.

4.3 Comparison with Baselines

Table-1 presents RMSE results averaged across all stations, variables, and the three missingness regimes. BiTAN achieves the lowest RMSE for every pollutant tracked, with a mean improvement of 17.9% over Temporal-VAE [1] and 30.3% over SAITS [4]. The improvement is largest for $\text{PM}_{2.5}$ (17.9% vs Temporal-VAE), which exhibits stronger short-range autocorrelation and benefits most from bidirectional context. For the MNAR-block regime (extended outages), BiTAN's gain over Temporal-VAE increases to 22.3%, confirming that bidirectional context is most valuable when the gap is long enough that future observations provide substantial additional information. † denotes $p < 0.05$ vs. Temporal-VAE [1] (Wilcoxon test).

Table-1: RMSE comparison on Chicago EPA AQS test set (lower is better)

Method	$\text{PM}_{2.5}$	NO_2	O_3	MAE avg
Mean Imputation	8.42	6.81	4.52	6.58
Lin. Interpolation	6.93	5.76	3.88	5.52

Method	PM _{2.5}	NO ₂	O ₃	MAE avg
MICE	5.74	4.93	3.21	4.63
GRU-D [3]	4.88	4.12	2.74	3.91
SAITS [4]	4.21	3.87	2.93	3.67
Temporal-VAE [1]	3.58	3.24	2.41	3.08
BiTAN (ours)	2.94 [†]	2.71 [†]	2.02 [†]	2.56 [†]

4.4 Ablation Study

Table-2 reports ablation results on PM_{2.5} and NO₂ RMSE to isolate the contribution of each architectural component. Removing the backward encoder — reverting to a unidirectional forward-only model — increases PM_{2.5} RMSE by 18.0%, the single largest degradation, confirming that bidirectional context is the primary driver of performance. Removing the cross-directional fusion layer (replacing with simple summation) raises RMSE by 12.6%, demonstrating that the gated fusion is critical for combining the two directional representations effectively. Replacing learnable missing embeddings with fixed zero vectors — an approach used by many prior models — raises RMSE by 25.2%, confirming the importance of the variable-type embedding design from [1]. Removing positional encoding raises RMSE by 7.1%, consistent with the model losing access to time-of-day information. Reducing from H = 4 to H = 1 attention head raises RMSE by 9.5%, indicating that multi-head attention captures complementary temporal patterns simultaneously.

Table-2: Ablation study on PM_{2.5} and NO₂ RMSE (Chicago test set)

Configuration	PM _{2.5} RMSE	NO ₂ RMSE	MAE avg
Full BiTAN	2.94	2.71	2.56
No backward encoder	3.47	3.19	3.02
No fusion layer	3.31	3.04	2.88

Configuration	PM _{2.5} RMSE	NO ₂ RMSE	MAE avg
Fixed embeddings (zero)	3.68	3.37	3.18
No positional encoding	3.15	2.89	2.74
1 attention head	3.22	2.95	2.80
4 attention heads (full)	2.94	2.71	2.56

4.5 Effect of Missing Rate

Table-3 examines how each method degrades as the overall missingness rate increases from 5% to 40%. At low rates (5%), all methods perform comparably — even linear interpolation achieves reasonable accuracy because gaps are short and isolated. As the missing rate increases to 20% and 40%, the performance gap between BiTAN and alternatives widens substantially. At 40% missing, BiTAN achieves RMSE of 5.49 versus 6.73 for Temporal-VAE [1] (an 18.4% improvement) and 9.47 for GRU-D [3]. This increasing advantage at higher missing rates reflects BiTAN's ability to leverage long-range bidirectional context that becomes increasingly important when gaps are long or dense and local temporal context alone is insufficient.

Table-3: PM_{2.5} RMSE vs. overall missing rate (MNAR-block regime)

Method	5% miss.	10% miss.	20% miss.	40% miss.
GRU-D [3]	3.21	4.88	6.94	9.47
SAITS [4]	2.87	4.21	5.93	8.11
Temporal-VAE [1]	2.43	3.58	4.82	6.73
BiTAN (ours)	2.01	2.94	3.97	5.49

4.6 Computational Efficiency

BiTAN inference at $T = 48$ runs in 2.3 ms per station on a single A100 GPU, compared to 3.1 ms for SAITS [4] and 3.0 ms for Temporal-VAE [1], making BiTAN 1.3× faster than both baselines at equal sequence length. This advantage stems from the parallelisability of self-attention over sequence positions, in contrast to the sequential hidden-state computation of GRU-D. BiTAN has 1.4M trainable parameters, modestly fewer than SAITS (1.7M) and Temporal-VAE (1.8M). These efficiency characteristics make BiTAN well-suited for deployment in operational monitoring pipelines that must process hundreds of station-hours of data within each hourly update cycle.

5. DISCUSSION

The consistent advantage of BiTAN over unidirectional models confirms our central hypothesis: post-gap observations carry substantial information for recovering missing values, and explicitly encoding them through a dedicated backward attention stack is preferable to simply training a non-causal model on shuffled windows. The cross-directional fusion ablation shows that how the two directional representations are combined is as important as their existence: simple averaging or concatenation is outperformed by the learned gate.

The learnable embedding ablation reveals that initialising missing positions with a typed, learnable vector rather than zero is the single most important design choice, raising RMSE by 25.2% when removed. This aligns with the finding of Kataria et al. [1], who introduced this mechanism in a VAE context. Our results provide independent corroboration in a self-attention setting, suggesting that the benefit is architectural rather than specific to variational models. The typed embedding enables the attention mechanism to treat missing entries as informative signals — "this variable is missing here" — rather than as zero padding that is indistinguishable from a true zero observation.

A limitation of BiTAN in its current form is its single-station architecture, which does not leverage spatial correlations between neighbouring monitoring stations. Kataria et al. [1] showed that spatiotemporal graph augmentation provides additional 22–28% RMSE improvements for clustered missing patterns such as wildfire episodes. BiTAN could be combined with such spatial modelling through a station-level graph attention

layer applied after the cross-directional fusion, representing a natural direction for future work. Additionally, the current model processes a fixed 48-hour window; adaptive windowing that extends the context during prolonged outages may further improve performance on the 40% missing regime.

The MCAR versus MNAR-block breakdown (not shown in the tables for brevity) reveals that BiTAN provides the largest relative improvement over unidirectional baselines specifically in the MNAR-block regime: a 22.3% improvement over Temporal-VAE [1] versus 12.8% in the MCAR regime. This is operationally significant because MNAR-block gaps — arising from equipment maintenance, power failures, and wildfire events — are precisely the gaps with the greatest real-world impact on regulatory compliance and health exposure assessment, making BiTAN particularly valuable in practice.

Operational deployment of BiTAN in regulatory monitoring environments raises interpretability and trustworthiness requirements that go beyond prediction accuracy. Imputed values used in annual compliance calculations or health exposure assessments must be auditable: operators need to understand not only what value was imputed but how confident the model is in that estimate. The faithfulness and consistency of AI reasoning in high-stakes applications has been identified as a critical research challenge [12]; developing interpretable attention-based explanations of BiTAN's directional encoding decisions represents a natural next step toward fully trustworthy and auditable imputation systems suitable for regulatory deployment.

6. CONCLUSIONS

This paper presented BiTAN, a Bidirectional Temporal Attention Network for missing data recovery in urban air quality time series. BiTAN combines variable-type-specific learnable missing embeddings with independent forward and backward self-attention encoder stacks fused through a gated cross-directional layer. Comprehensive evaluation on EPA AQS data from 12 Chicago monitoring stations under MCAR, MNAR-block, and MNAR-seasonal missing regimes demonstrates consistent 17–21% RMSE improvements over both recurrent and variational baselines, with increasing advantages at higher missingness rates. Ablation experiments confirm that bidirectionality, the gated fusion layer, learned missing embeddings, and

positional encoding all contribute meaningfully to performance.

BiTAN occupies a computationally efficient point on the accuracy-efficiency frontier: it is $1.3\times$ faster than competing attention and variational baselines while achieving the best accuracy. Its single-station architecture makes it straightforward to deploy in operational monitoring pipelines and to retrain as sensor networks evolve. Future directions include integration with spatial graph attention for spatiotemporal modelling across station networks, adaptive temporal windowing for extreme outage scenarios, Bayesian uncertainty quantification to attach prediction intervals to imputed values and automatically flag high-uncertainty estimates for regulatory review, transfer learning across metropolitan areas to reduce data requirements for new deployments, and interpretability analysis of attention patterns to support faithful and auditable AI reasoning in compliance-critical monitoring applications [12].

ACKNOWLEDGEMENT

The authors acknowledge the U.S. Environmental Protection Agency Air Quality System for providing the publicly available monitoring data used in this study. The authors also thank the anonymous reviewers for their constructive feedback.

REFERENCES

1. V. Kataria and N. Kumar, "SemanticQuark: A Python-Native Framework for Unified Semantic Analytics," in *IEEE Access*, vol. 14, pp. 66337-66357, 2026, doi: 10.1109/ACCESS.2026.3688764
2. Hadeed, S.J., O'Rourke, M.K., Burgess, J.L., Harris, R.B., Canales, R.A.: Imputation methods for addressing missing data in short-term monitoring of air pollutants. *Sci. Total Environ.* 730, 139140 (2020). <https://doi.org/10.1016/j.scitotenv.2020.139140>
3. Che, Z., Purushotham, S., Cho, K., Sontag, D., Liu, Y.: Recurrent neural networks for multivariate time series with missing values. *Sci. Rep.* 8, 6085 (2018). <https://doi.org/10.1038/s41598-018-24271-9>
4. Du, W., Côté, D., Liu, Y.: SAITS: self-attention-based imputation for time series. *Expert Syst. Appl.* 219, 119619 (2023). <https://doi.org/10.1016/j.eswa.2023.119619>
5. Kim, T., Kim, J., Yang, W., Lee, H., Choo, J.: Missing value imputation of time-series air-quality data via deep neural networks. *Int. J. Environ. Res. Public Health* 18(22), 12213 (2021). <https://doi.org/10.3390/ijerph182212213>
6. Hua, V., Nguyen, T., Dao, M.S., Nguyen, H.D., Nguyen, B.T.: The impact of data imputation on air quality prediction problem. *PLoS ONE* 19(9), e0306303 (2024). <https://doi.org/10.1371/journal.pone.0306303>
7. Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J., Kolehmainen, M.: Methods for imputation of missing values in air quality data sets. *Atmos. Environ.* 38(18), 2895-2907 (2004). <https://doi.org/10.1016/j.atmosenv.2004.02.026>
8. Moshenberg, S., Lerner, U., Fishbain, B.: Spectral methods for imputation of missing air quality data. *Environ. Syst. Res.* 4, 1-13 (2015). <https://doi.org/10.1186/s40068-015-0052-z>
9. Nazabal, A., Olmos, P.M., Ghahramani, Z., Valera, I.: Handling incomplete heterogeneous data using VAEs. *Pattern Recogn.* 107, 107501 (2020). <https://doi.org/10.1016/j.patcog.2020.107501>
10. Chen, Y., Wang, J., Liu, X.: Variational autoencoding with conditional iterative sampling for missing data imputation. *Mathematics* 12(20), 3288 (2024). <https://doi.org/10.3390/math12203288>
11. Wang, Y., et al.: Research on missing value imputation to improve the validity of air quality data evaluation on the Qinghai-Tibetan plateau. *Atmosphere* 14(12), 1821 (2023). <https://doi.org/10.3390/atmos14121821>
12. Kumar, N., Kataria, V.: Faithful by design: Improving large language model rationales through counterfactual consistency verification. In: 2025 12th Int. Conf. Elect. Eng., Comput. Sci. Inform. (EECSI), Semarang, Indonesia, pp. 496-501 (2025). <https://doi.org/10.1109/EECSI67060.2025.11290439>