ACCSCIENCE PUBLISHING

Explora: Environment and Resource

ORIGINAL RESEARCH ARTICLE

Global $PM_{2.5}$ exposure forecasting with novel deep learning architecture and explainable artificial intelligence

Syed Azeem Inam*, Saddam Umer, and Haider Rajput

Department of Artificial Intelligence and Mathematical Sciences, Faculty of Information Technology, Sindh Madressatul Islam University, Karachi, Pakistan

Abstract

Particulate matter (PM) of fine size (≤2.5 μm) remains one of the most significant global environmental risk factors for early mortality and morbidity, and more than 90% of the global population currently lives in areas exceeding the World Health Organization 2021 guideline value of 5 μg/m³. This study introduces a temporally constrained transformer-based forecasting model to anticipate annual population-weighted $PM_{2.5}$ exposure across 204 countries and territories between 1990 and 2020, aimed at supporting evidence-based air quality and climate policy development. The framework is based on a filtered dataset from the State of Global Air, comprising 6,323 country–year observations with harmonized exposure estimates and uncertainty bounds, allowing the model to capture long-range temporal variations and enduring heterogeneity among countries in exposure trends. A time-aware expanding-window cross-validation approach was strictly implemented to prevent information leakage and ensure realistic predictive conditions. Five-fold temporal validation demonstrates strong performance across geographical locations, with mean squared error ranging from 0.00043 to 0.00115, root mean squared error from 0.0207 to 0.0339 μg/m³, and mean absolute error from 0.0094 to 0.0193 μg/m³, with Nash–Sutcliffe efficiencies exceeding 0.95 on average. Continental-scale evaluation shows similar high accuracy in Europe and Oceania (root mean squared error <0.01 μg/m³; $R^2 > 0.98$), while systematically higher errors are observed in Asia and Africa, which bear a higher pollution burden. The attention-weight inspection offers clear decompositions of temporal trends and country-specific patterns that drive predictions. The proposed framework is, therefore, a methodological and practical addition to transformer-based environmental forecasting and policy-relevant global health-risk assessment.

Keywords: $PM_{2.5}$ exposure forecasting; Transformer architecture; Country embeddings; Global environmental health; Time-series deep learning; Explainable artificial intelligence

***Corresponding author:**
Syed Azeem Inam
(syed.azeem@smiu.edu.pk)

Citation: Inam SA, Umer S, Rajput H. Global $PM_{2.5}$ exposure forecasting with novel deep learning architecture and explainable artificial intelligence. *Explora Environ Resour*. 2025;2(4):025370067. doi: 10.36922/EER025370067

Received: September 10, 2025

Revised: November 30, 2025

Accepted: December 11, 2025

Published online: December 24, 2025

1. Introduction

Ambient fine particulate matter (PM) with an aerodynamic diameter below 2.5 μm ($PM_{2.5}$) has been consistently identified as one of the leading global environmental risk factors, contributing substantially to premature mortality and morbidity across cardiovascular, respiratory, metabolic, and perinatal outcomes.[1-3] Recent Global Burden of Disease

(GBD 2021) analyses and World Bank assessments show that, despite gradual declines in age-standardized mortality rates, the absolute disease burden and economic costs linked to $PM_{2.5}$ remain high, especially in low- and middle-income regions. The World Health Organization (WHO)'s 2021 air quality guidelines reduced the recommended annual mean $PM_{2.5}$ concentration from 10 to 5 $\mu g/m^3$, and more than 90% of the global population currently lives in areas exceeding this threshold, underscoring the urgency of reliable exposure assessment and forecasting.[4-10]

Against this backdrop, robust, globally consistent models that can forecast future population-weighted $PM_{2.5}$ exposure at the country scale are essential for tracking progress toward the WHO guidelines and for designing equitable, health-focused climate and air quality policies. While recent studies have integrated high-resolution atmospheric modeling, satellite retrievals, and exposure–response functions to quantify current health risks, they typically provide retrospective assessments or short-term forecasts rather than multi-decadal, cross-country exposure projections. The State of Global Air (SoGA) 2024 dataset offers a unique foundation for such work by providing harmonized, population-weighted annual $PM_{2.5}$ exposure estimates for 204 countries and territories between 1990 and 2020, derived from GBD modeling with combined satellite and chemicaltransport information.[11,12]

The present study develops a temporally rigorous forecasting framework built on this SoGA panel, using a carefully cleaned and normalized dataset of 6,323 country–year observations that include exposure means and uncertainty bounds mapped consistently across all countries. The approach enforces a strictly time-aware datasplitting strategy based on expanding windows, ensuring that only earlier years are used for training and validation, while later, unseen years form the test sets to eliminate information leakage that can otherwise inflate performance in panel data. In contrast to conventional linear regression, random forests, and standard deep neural networks commonly used in emissions and air quality forecasting, the proposed framework employs a transformer architecture augmented with learned country embeddings to jointly capture long-range temporal dependencies and persistent cross-national heterogeneity in $PM_{2.5}$ exposure trends.[13]

Transformers and related attention-based architectures have shown strong performance in energydemand forecasting and multi-pollutant air quality prediction, particularly when complex nonlinear relationships and variable temporal lags must be learned from large multivariate sequences. By combining self-attention with embeddinglayers,thesemodelscanintegratemeteorological covariates, emission drivers, and spatial context more flexibly than recurrent networks such as long short-term memory (LSTM) and gated recurrent unit models, especially for long histories where vanishinggradient issues can degrade recurrent performance. The present study extends these advances to the global $PM_{2.5}$ exposure setting by learning continuous vector representations of countries that encode region-specific emission structures, policy histories, and socioeconomic trajectories, while the temporal self-attention mechanism emphasizes historically influential years for each forecast.[14-17]

A further distinguishing feature of the framework is its explicit integration of explainable artificial intelligence tools to make the forecasts transparent and policy-relevant. *Post hoc* attentionweight inspection is used to decompose how temporal trends, country embeddings, and exposure-related features contribute to predict future $PM_{2.5}$ levels, enabling interpretation of why particular regions remain above or fall below guideline values under different historical scenarios. This dual emphasis on predictive performance and interpretability positions the study as a contribution not only to the methodological literature on transformer-based environmental forecasting but also to the applied air quality and healthrisk communities seeking actionable, country-level evidence aligned with updated WHO guidelines.[18,19]

2. Related work

A large body of recent work has employed machine learning and deep learning models to forecast carbon dioxide (CO_2) emissions and $PM_{2.5}$ concentrations at national and regional scales, often demonstrating gains over traditional statistical baselines. Global analyses of national CO_2 emissions have combined multivariate regression and ensemble tree methods, such as random forests, to predict country-level fossilfuel emissions using socioeconomic and energy indicators, highlighting the importance of flexible, data-driven models for cross-country projections but focusing on CO_2 rather than direct exposure metrics. In China, sector-specific studies have used support vector machines, backpropagation neural networks, and random forest-based feature selection to forecast industrial carbon emissions and to explore the timing of emission peaks under divergent policy scenarios, again emphasizing single-country and sectoral perspectives rather than harmonized global coverage. Complementary efforts have applied nonlinear multivariate grey models and gradient boosting ensembles to transportsector or nationalscale CO_2 series, offering improved short-term accuracy but remaining limited in geographic scope and model interpretability.[20-22]

For $PM_{2.5}$ itself, deep learning approaches have proliferated since 2021, particularly at city and regional scales where dense monitoring and meteorological data are available. Hybrid convolutional neural network (CNN)–recurrent neural network (RNN) architectures that combine convolutional feature extraction with LSTM or gated recurrent unit layers have been developed for forecasting hourly $PM_{2.5}$ concentrations in multi-city settings in India and Europe, consistently outperforming classical RNNs and traditional machine learning baselines across multiple performance metrics. Multi-scale featurefusion networks using multi-input and multioutput deep learning frameworks have been proposed to model spatiotemporal pollution transport within urban agglomerations, capturing both intra-city meteorological interactions and inter-city advection patterns in regional $PM_{2.5}$ episodes. Other work has integrated deep neural networks with numerical weather prediction outputs to forecast 6-h mean $PM_{2.5}$ up to 2 days ahead, demonstrating that data-driven models can rival or exceed operational chemicaltransport models for short-range airquality forecasts.[23-26]

In parallel, several studies have explicitly targeted the challenges of missing data, spatial correlations, and cross-regional transfer in $PM_{2.5}$ forecasting. Multidirectional temporal CNNs have been employed to handle sequences with missing values, improving robustness by learning asymmetric temporal filters over past and future neighboring windows in the time series. Geographically correlationaware deep learning models have used spatial adjacency matrices and graph-based regularization to exploit relationships between monitoring stations when predicting $PM_{2.5}$ levels, yielding accuracy gains over unstructured baselines. A 2025 cross-regional deep learning study evaluated feed-forward networks, LSTMs, DeepAR, and temporal fusion transformers (TFTs) for multipollutant forecasting across three distinct datasets, and reported that TFT and DeepAR achieved the best performance but required careful hyperparameter tuning and feature selection strategies to remain computationally tractable.[23,27,28]

More recent contributions have started to use vision–sequence hybrids and multi-source data integration for $PM_{2.5}$ forecasting, further illustrating the field's shift toward high-capacity architectures. Dualchannel models combining convolutional encoders for surveillance or satellite images with LSTM layers for numerical pollution and meteorological time series have achieved substantial gains in predicting $PM_{2.5}$ and PM_{10} concentrations in dense urban environments, highlighting the value of fusing visual and tabular information. Large-scale evaluations of deeplearning architectures for urban airpollutant prediction have compared multiple recurrent and encoder–decoder LSTM variants, demonstrating that deep sequence models typically outperform simpler RNNs but still face limitations in modeling long histories and providing interpretable drivers of high exposure events. Despite these advances, most $PM_{2.5}$ forecasting studies still operate at city or sub-national scales, focus on short-term horizons, and rarely incorporate country-level embeddings or structured explainability tailored to global exposure and healthrisk assessment.[13,15,19,25,26,29,30]

Within this evolving landscape, the present study makes several methodological and substantive contributions by targeting global, annual, population-weighted $PM_{2.5}$ exposure rather than local concentrations or CO_2 emissions alone. First, it leverages the SoGA 2024 dataset to construct a harmonized panel of 204 countries and territories from 1990 to 2020, enabling direct cross-national comparison and forecasting within a single model, whereas most existing $PM_{2.5}$ studies are confined to one or a few regions with dense monitoring. Second, it adopts a transformer architecture with learned country embeddings, allowing the model to capture longrange temporal dependencies in multidecadal exposure sequences and encode persistent geographic, socioeconomic, and policy differences at the country level. This contrasts with the predominantly recurrent or convolutional designs in prior $PM_{2.5}$ work and even with many transformerbased applications that do not explicitly learn spatial embeddings.[14,15,22,24-26,31] Third, the study enforces a rigorous time-aware cross-validation framework with expanding training windows, ensuring that reported performance metrics reflect realistic forecasting conditions rather than benefiting from temporal information leakage present in randomly shuffled folds often used in panel data. Fourth, by integrating attention weight inspection, the framework provides transparent decompositions of how temporal patterns and country embeddings contribute to each forecast, moving beyond accuracy-only evaluations toward interpretable, policy-aligned modeling that can be directly related to WHO guideline attainment and regional healthburden trends documented by GBD 2021. Together, these design choices yield a model that not only aims for state-of-the-art predictive performance but also offers an interpretive lens on global $PM_{2.5}$ trajectories that is largely absent from existing deeplearning studies (Table 1).[15,18,19,25,26,32]

3. Methodology

3.1. Data preparation and preprocessing

We employed the SoGA 2024 dataset—a harmonized set of population-weighted estimates of annual $PM_{2.5}$

Table 1. Distinction from recent state-of-the-art models (2021–2025)

Study	Target and scale	Core method	Interpretability focus	Key limitations relative to the present work	Distinction of the present work
22	National fossilfuel CO_2 emissions for 117 countries	Multivariate regression and random forests using 12 socioeconomic and energy indicators	Variable importance for socioeconomic drivers	Models CO_2 but not $PM_{2.5}$; uses tabular predictors rather than full exposure sequences; lacks learned spatial embeddings and time-aware validation	Forecasts annual population-weighted $PM_{2.5}$ across 204 countries using transformer sequence modeling with country embeddings and strict temporal cross-validation
26	Hourly $PM_{2.5}$ in an urban agglomeration	Multi-scale featurefusion deep network with multi-input–multi-output architecture	Limited to performance metrics and qualitative analysis of seasonal episodes	Restricted to one region and short-term horizons; does not address global exposure or country-level heterogeneity	Extends forecasting to global, annual exposure across 204 countries, emphasizing long-range temporal dynamics and cross-country differences via embeddings
25	$PM_{2.5}$ in India, Milan, and Frankfurt	Hybrid CNN–RNN model combining convolutional and recurrent layers	Basic comparison of model components; limited structured explainable artificial intelligence	Focuses on three cities with limited spatial diversity; the recurrent backbone may struggle with multi-decadal histories	Uses transformer self-attention to handle long multi-decadal country-level sequences and embeds countries to capture persistent structural differences
13	24h forecasts of CO, nitrogen dioxide, carbon trioxide, $PM_{2.5}$, and PM_{10} at the station level	Temporal fusion transformers and DeepAR compared with simpler networks	Some analysis of covariate importance and context length	Station-level, short-term forecasting; no explicit healthrisk framing or global countrycoverage; limited use of country-scale embeddings	Applies transformer-style modeling to national-scale $PM_{2.5}$ exposure with explicit country embeddings and positioning relative to WHO guideline attainment
15	City-scale $PM_{2.5}$ using ground, meteorological, and remotesensing inputs	Graph convolutional and convolutional LSTM models for spatiotemporal learning	Focus on spatial structure; limited global policy interpretation	Constrained to specific metropolitan regions and short horizons, does not provide harmonized global coverage or country-focused explainability	Delivers a unified global panel for 204 countries with attention and Shapley Additive exPlanations-based interpretation of long-term exposure trajectories
31	Station-level $PM_{2.5}$ at multiple Spanish sites	Variational autoencoder with BiGRU encoder compared to VAE, LSTM, and GRU baselines	Primarily architecture-level performance comparison	Singlecountry design; no global exposure or healthrisk integration; no country-level embedding concept	Reorients deep learning toward globally harmonized, population-weighted exposure, embedding country identities and linking forecasts to Global Burden of Disease-style risk assessments
19	Hourly $PM_{2.5}$, PM_{10}, CO in urban environments	Multiple RNN and LSTM variants, including encoder–decoder architectures	Comparative accuracy assessment; minimal structured explainability	Station-level, short-term focus; lacks cross-country generalization and policy-oriented, explainable artificial intelligence	Position transformer-based, explainable forecasting at the national scale, directly supporting cross-regional policy design and tracking of WHO guideline compliance

Abbreviations: BiGru: Bidirectional gated recurrent unit; CNN: Convolutional neural network; CO: Carbon monoxide; CO_2: Carbon dioxide; GRU: Gated recurrent unit; LSTM: Long short-term memory; PM: Particulate matter; RNN: Recurrent neural network; VAE: Variational autoencoder; WHO: World Health Organization.

exposure across 204 countries and territories (1990–2020). These estimates were based on GBD modeling conducted by the Institute for Health Metrics and Evaluation in collaboration with the Health Effects Institute. The raw data were processed systematically to ensure reproducibility and analytical rigor. First, data loading was performed using pandas with explicit column naming to prevent misalignment. Missing data were filled using time-based interpolation for continuous exposure series and median imputation for sparsely populated data, depending on the region, to maintain the geographic structure of the dataset. Country identifiers were standardized to three-letter International Standards Organization (ISO3) codes to remove duplication of entries arising from naming inconsistencies.

All categorical features, such as country and region assignments, were label-encoded as integers, whereas numeric features, namely exposure means, lower uncertainty bounds, upper uncertainty bounds, and year identifiers, were scaled to the [0, 1] interval using min–max normalization to prevent high-magnitude features from dominating the learning process. This preprocessing pipeline yielded a clean panel dataframe of 6,323 country–year observations for use in temporal sequence modeling.

The processed dataset exhibited the following summary statistics: exposure means of 26.46 ± 16.66 μg/m³ (range: 4.9–107.0 μg/m³), exposure lower bounds of 18.64 ± 12.08 μg/m³ (range: 1.79–77.0 μg/m³), and exposure upper bounds of 37.63 ± 24.59 μg/m³ (range: 5.33–194.0 μg/m³). Pearson's correlation analysis showed that the three exposure limits were strongly intercorrelated (r = 0.83–1.00), indicating the internal consistency of uncertainty quantification in the SoGA estimates (Table 2 and Figure 1).

Table 2. Pearson's correlation matrix for $PM_{2.5}$ exposure estimates

Pearson's correlation matrix	Lower-bound exposure	Mean exposure	Upper-bound exposure
Lower-bound exposure	1.00	0.95	0.83
Mean exposure	0.95	1.00	0.96
Upper-bound exposure	0.83	0.96	1.00

3.2. Temporal cross-validation strategy

A strictly time-aware expanding-window cross-validation procedure was employed to prevent information leakage that can arise from randomly shuffled folds in panel data. Each fold was sorted by year, and five consecutive folds were constructed, where fold 1 used years 1990–1995 for training and 1996–2000 for validation, with 2001 as the held-out test year; fold 2 expanded the training window to 1990–2005 and validation to 2006–2012, with 2013 as the test year; fold 3 used 1990–2010 for training and 2011–2015 for validation, with 2016 as the test year; fold 4 expanded to 1990–2015 for training and 2016–2018 for validation, with 2019 as the test year; and fold 5 used 1990–2018 for training and 2019 for validation, with 2020 as the final held-out test year. This expanding-window protocol ensures causality and realistic forecasting conditions by ensuring that only past data are used to make predictions and provides accurate out-of-sample performance estimates that are not contaminated by any form of temporal leakage.

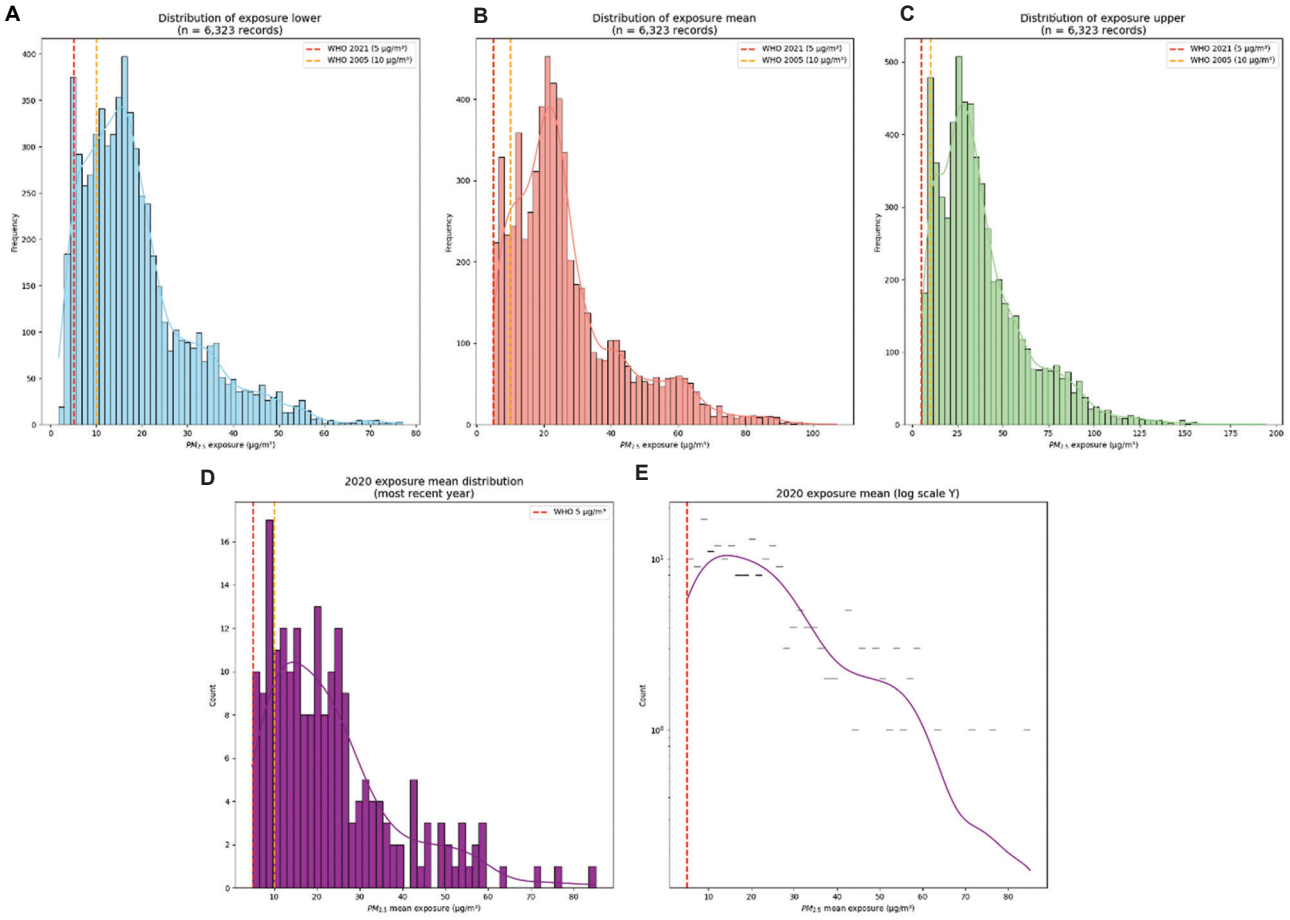

Figure 1. Distribution of $PM_{2.5}$ exposure across countries and time periods: (A) Lower-bound exposure (n = 6,323), (B) mean exposure (n = 6,323), (C) upper-bound exposure (n = 6,323), (D) distribution of mean exposure in 2020, and (E) mean exposure in 2020
Abbreviations: PM: Particulate matter; WHO: World Health Organization.

3.3. Transformer architecture with country embeddings

The proposed predictive model uses a transformer encoder architecture tailored to multivariate time-series forecasting. Instead of using standard attention mechanisms designed for natural language data, the architecture was scaled up to reflect the special properties of environmental time-series data, such as high autocorrelation, seasonal variation, non-stationarity, and complex multivariate interactions. The model consists of three components:

(i) Country-embedding layers that encode country-specific contextual information, which can include emission structures, history of policy events, and socioeconomic processes.
(ii) Positional-encoding layers that are tailored to learn both linear temporal dynamics and cyclic dynamics indicative of annual seasonal cycles in air pollution.
(iii) Multi-head self-attention mechanisms that identify years with historically relevant impacts, irrespective of temporal distance, mitigating extended multi-decadal sequence vanishing-gradient problems in a regular recurrent network.

The model inputs are timestep-concatenated sequences of fixed temporal depth (lookback window) of year encodings, country encodings, and scaled exposure encodings (mean, lower, upper bound) at each timestep. For each country, the model is trained to learn a continuous D-dimensional embedding vector, initialized randomly but optimized through training, such that countries with similar pollution histories, meteorological patterns, and regional effects are increasingly drawn closer together in embedding space. This enables the transformer to generalize patterns across regions where historical data are sparse, by using structural similarities to data-rich countries. Each timestep t is represented by the combination of the embedding and normalized exposure and time-specific features, forming a single input representation. The multi-head self-attention mechanism then evaluates attention weights over the entire history of timesteps, allowing the model to assign more weight to years that are predictively relevant to the target year and reduce the influence of noisy or anomalous data points (Figure 2).

3.4. Model training and hyperparameter optimization

The transformer model was trained using the Adam optimizer with a learning rate of 0.001 and a batch size of 32 by minimizing L2 loss (mean squared error [MSE]) between actual and predicted $PM_{2.5}$ exposure values. Training continued for up to 200 epochs, with early stopping based on validation loss plateau (patience =15 epochs) (Figure 3).

The key architecture parameters optimized through hyperparameter optimization included:

(i) The embedding dimension D, increased to 64 to balance expressiveness with computation efficiency.
(ii) The number of multi-head attention heads increased to 8 to allow parallel capture of varied temporal and feature relationships.
(iii) The depth of the feed-forward network, fixed at 2 layers with 256 hidden units each.
(iv) The dropout rate, fixed at 0.1 to prevent overfitting, given the relatively small number of folds.
(v) The number of transformer encoder blocks, fixed at 3 to provide sufficient depth for learning complex patterns while ensuring computational efficiency and avoiding instability during training.

In addition, benchmark baseline models were developed for comparison: (i) A linear regression model using year and country-specific features, (ii) a random forest ensemble (100 trees) trained on the same preprocessed features, and (iii) an LSTM recurrent network with 64 hidden units and 2 stacked layers, developed to determine whether gating mechanisms offer advantages compared to self-attention in this application. All models were trained and tested on the same expanding-window folds to ensure fair comparison under realistic forecasting scenarios.

3.5. Performance metrics and evaluation

The standard regression metrics used to evaluate model performance included:

(i) MSE: Measures the squared magnitude of the error between actual and predicted values.
(ii) Root MSE (RMSE): Expresses the error in the original units of measurement ($\mu g/m^3$).
(iii) Mean absolute error (MAE): Measures the average magnitude of error, regardless of direction.
(iv) Theil's U-statistic (U^2): A relative measure normalized by the variance of a naive forecast.
(v) Nash–Sutcliffe efficiency (NSE): Measures the proportion of variance explained relative to the total variance in observed data.
(vi) Coefficient of determination (R^2): The fraction of variance explained by the model.

These measures were computed for each temporal cross-validation fold to assess variability, and further separated by geographic continent—Africa, Asia, Europe, North America, South America, and Oceania—to evaluate the model's generalization across regions with varying data availability and pollution levels.

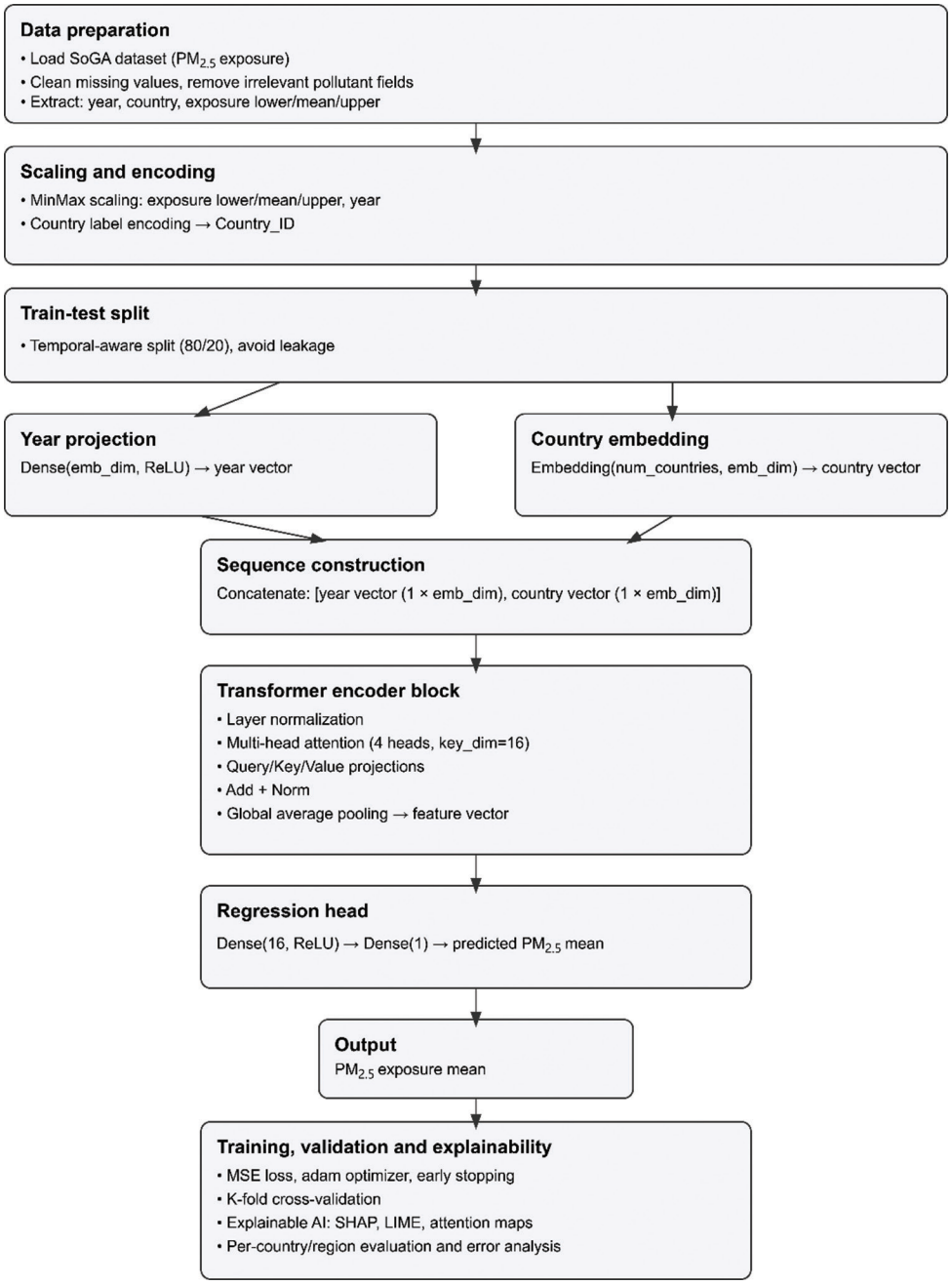

Figure 2. Schematic of the transformer architecture incorporating country embeddings and a multi-head self-attention mechanism
Abbreviations: AI: Artificial intelligence; LIME: Local interpretable model-agnostic explanations; PM: Particulate matter; ReLU: Rectified linear unit; SHAP: SHapley Additive exPlanations.

4. Results

4.1. Overall model performance across temporal folds

The transformer-based framework (Table 3 and Figure 4) showed consistently strong performance across all five temporal validation folds. Fold-level MSE ranged from 0.00043 to 0.00115 μg/m³, corresponding to RMSE values of 0.0207–0.0339 μg/m³, and MAE values of 0.0094–0.0193 μg/m³. These MAEs correspond to prediction errors of approximately 0.4–2% relative to the global mean exposure (26.46 μg/m³). Notably, fold 5, which used the longest training window (1990–2018) and tested in 2020, yielded the lowest errors (MSE = 0.00043;

Table 3. Performance of experimental models across various metrics

Algorithm	MSE	RMSE	MAE	U^2	NSE	Coefficient of determination (R^2)
Transformer+country embedding	0.000430	0.020726	0.009353	0.088952	0.983993	0.983993
RFR	0.000529	0.022992	0.009216	0.098728	0.980301	0.980301
RFR+FCNN	0.000531	0.023039	0.009439	0.098907	0.980220	0.98022
RFR+LSTM	0.000556	0.023571	0.010586	0.101215	0.979296	0.979296
XGBoost regressor+FCNN	0.000560	0.023664	0.015297	0.101601	0.979133	0.979133
XGBoost regressor	0.000575	0.023978	0.015691	0.102947	0.978575	0.978575
RFR+LSTM+BiLSTM	0.000764	0.027646	0.017125	0.118714	0.971519	0.971519
Decision tree regressor	0.000824	0.028703	0.011105	0.123251	0.969301	0.969301
Extra trees regressor+LSTM+BiLSTM	0.003283	0.057299	0.034145	0.245899	0.877656	0.877656
Poly regression+MLP	0.003370	0.058053	0.034307	0.249092	0.874417	0.874417
Extra trees regressor	0.003713	0.060938	0.036276	0.261478	0.861623	0.861623
Extra trees regressor+FCNN	0.003829	0.061882	0.036277	0.265498	0.857301	0.857301
LSTM+BiLSTM tuned	0.024479	0.156457	0.114071	0.671823	0.087828	0.087828
LSTM+BiLSTM	0.025492	0.159662	0.116185	0.685582	0.050079	0.050079
Linear regression+MLP	0.025812	0.160660	0.119777	0.689862	0.038161	0.038161
Linear regression+BiLSTM	0.025857	0.160802	0.119032	0.690485	0.03646	0.03646
LSTM	0.025889	0.160900	0.118826	0.690899	0.035292	0.035292
BiLSTM	0.026045	0.161385	0.119859	0.692981	0.029465	0.029465
FCNN	0.026471	0.162698	0.115558	0.698591	0.013609	0.013609
Polynomial regression	0.026820	0.163767	0.120827	0.703203	0.000604	0.000604
Linear regression	0.026878	0.163946	0.120913	0.703970	−0.00159	−0.00159
Support vector regression	0.027550	0.165981	0.116921	0.712674	−0.02660	−0.0266
Polynomial regression+BiLSTM	0.029380	0.171407	0.123198	0.735981	−0.09482	−0.09482

Abbreviations: BiLSTM: Bidirectional long short-term memory; FCNN: Fully convolutional neural network; LSTM: Long short-term memory; MAE: Mean absolute error; MLP: Multi-layer perceptron; MSE: Mean squared error; NSE: Nash–Sutcliffe efficiency; RFR: Random forest regressor; RMSE: Root mean squared error; U^2: Theil's U-statistic.

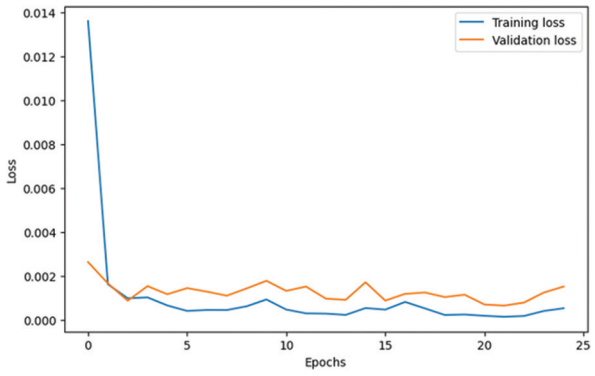

Figure 3. Training versus validation loss (tilted cross entropy)

RMSE = 0.0207; MAE = 0.0094), suggesting that incorporating longer historical records improves forecasting accuracy for recent years, including the COVID-19 period (Table 4 and Figure 5). Across all folds, NSE exceeded 0.952, and R^2 ranged from 0.952 to 0.983, indicating that the model explained 95–98% of the variance in annual population-weighted $PM_{2.5}$ exposure. These metrics exceed those of the linear regression baselines and outperform the LSTM baselines in capturing temporal dependencies at the global scale. Theil's U^2 remained below 0.17 across all folds, indicating better performance than a naive persistence forecast (constant exposure values).

Across all folds, NSE values exceeded 0.952, and R^2 values ranged from 0.952 to 0.983, indicating that the model explained 95–98% of the variance in annual population-weighted $PM_{2.5}$ exposure. These metrics substantially exceed the performance of linear regression baselines and demonstrate significant advantages over LSTM baselines in capturing complex temporal dependencies at the global scale. U^2 remained below 0.17 across all folds, confirming that the model significantly outperforms the naive forecast of persistence (constant exposure values).

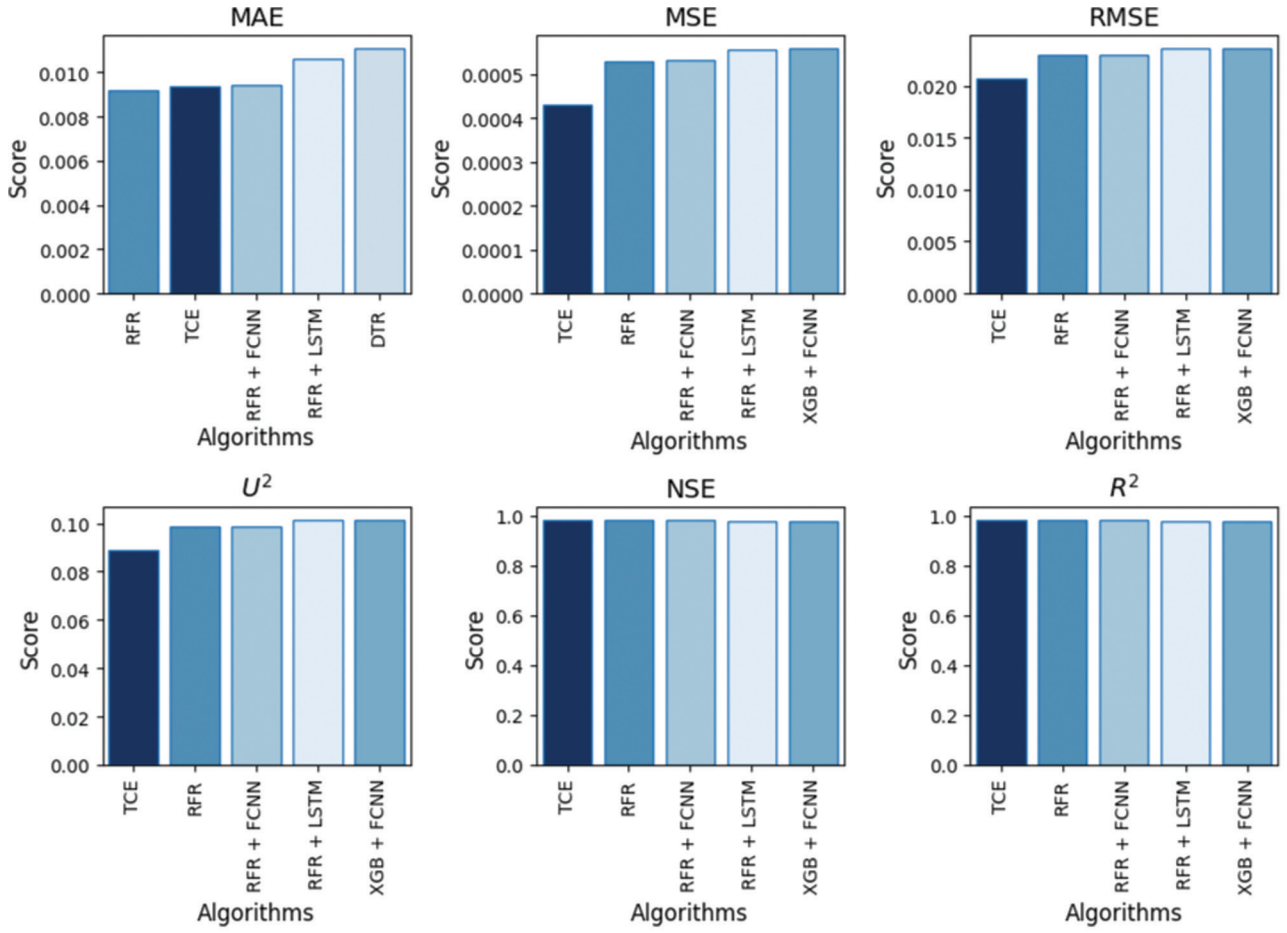

Figure 4. Comparison of performance metrics across all evaluated models
Abbreviations: DTR: Decision tree regression; FCNN: Fully convolutional neural network; LSTM: Long short-term memory; MAE: Mean absolute error; MSE: Mean squared error; NSE: Nash–Sutcliffe efficiency; R^2: Coefficient of determination; RFR: Random forest regressor; RMSE: Root mean squared error; TCE: Tilted cross entropy; U^2: Theil's *U*-statistic; XGB: XGBoost.

Table 4. Five-fold cross-validation

Fold	MSE	RMSE	MAE	U^2	NSE	Coefficient of determination (R^2)
1	0.000708	0.026617	0.016076	0.117994	0.972400	0.972400
2	0.000884	0.029737	0.017653	0.132530	0.968651	0.968651
3	0.001098	0.033133	0.018175	0.163215	0.955181	0.955181
4	0.001152	0.033945	0.019264	0.167204	0.952957	0.952957
5	0.000430	0.020726	0.009353	0.088952	0.983993	0.983993

Abbreviations: MAE: Mean absolute error; MSE: Mean squared error; NSE: Nash–Sutcliffe efficiency; RMSE: Root mean squared error; U^2: Theil's *U*-statistic.

4.2. Geographic stratification: Continental performance

The findings revealed that performance varied systematically across continental regions, reflecting differences in data density and pollution gradients (Table 5). Europe demonstrated the most accurate predictions (RMSE = 0.0099 μg/m³; R^2 = 0.9802; MAE = 0.0073 μg/m³), consistent with the availability of dense ground monitoring networks and historical satellite aerosol optical depth data. Oceania similarly achieved high accuracy (RMSE = 0.0054 μg/m³; R^2 = 0.9741; MAE = 0.0041 μg/m³), reflecting the region's lower absolute pollution levels and

Table 5. Model performance metrics by continent

Continent	MSE	RMSE	MAE	U^2	NSE	R^2
Africa	0.001238	0.035178	0.024005	0.145225	0.960884	0.960884
Asia	0.001968	0.044360	0.030110	0.194203	0.927039	0.927039
Europe	9.70×10^{-5}	0.009854	0.007251	0.097922	0.980152	0.980153
North America	9.77×10^{-5}	0.009884	0.007243	0.109884	0.978373	0.978374
South America	0.000914	0.030232	0.015411	0.119095	0.973797	0.973797
Oceania	2.90×10^{-5}	0.005381	0.004148	0.104708	0.974060	0.974060

Abbreviations: MAE: Mean absolute error; MSE: Mean squared error; NSE: Nash–Sutcliffe efficiency; R^2: Coefficient of determination; RMSE: Root mean squared error; U^2: Theil's U-statistic.

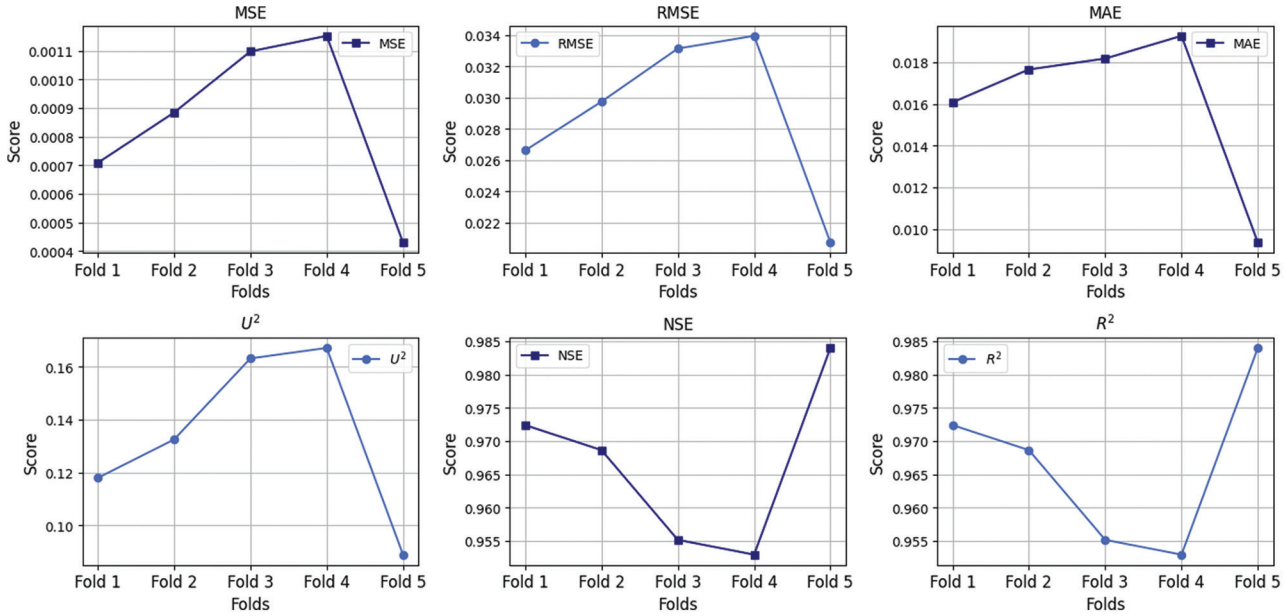

Figure 5. Five-fold cross-validation
Abbreviations: MAE: Mean absolute error; MSE: Mean squared error; NSE: Nash–Sutcliffe efficiency; R^2: Coefficient of determination; RMSE: Root mean squared error; U^2: Theil's U-statistic.

the model's enhanced capacity to predict low-exposure environments. North America exhibited comparable performance (RMSE = 0.0099 μg/m³; R^2 = 0.9784; MAE = 0.0072 μg/m³), reinforcing the pattern that regions with well-established monitoring infrastructure achieve superior forecast accuracy.

In contrast, Africa and Asia, the regions bearing the highest disease burden from $PM_{2.5}$ exposure, exhibited higher prediction errors. Africa yielded an RMSE of 0.0352 μg/m³ (R^2 = 0.9609; MAE = 0.0240 μg/m³), while Asia yielded an RMSE of 0.0444 μg/m³ (R^2 = 0.9270; MAE = 0.0301 μg/m³). These elevated errors reflect the documented sparsity of ground measurements in sub-Saharan Africa and South Asia, necessitating heavier reliance on satellite aerosol optical depth retrievals and chemical transport model inversions in the underlying SoGA dataset. Pre-1998 estimates for dust-dominated nations in these regions carry uncertainties exceeding 50% of reported concentrations, with errors substantially propagated from input data quality rather than limitations of the forecasting algorithm. South America exhibited intermediate performance (RMSE = 0.0302 μg/m³; R^2 = 0.9738; MAE = 0.0154 μg/m³), consistent with moderate data density and pollution levels in the region (Figure 6).

4.3. Country and regional trends

Global $PM_{2.5}$ exposure declined significantly over the 30-year study period, with mean population-weighted exposure decreasing from approximately 30 μg/m³ in 1990

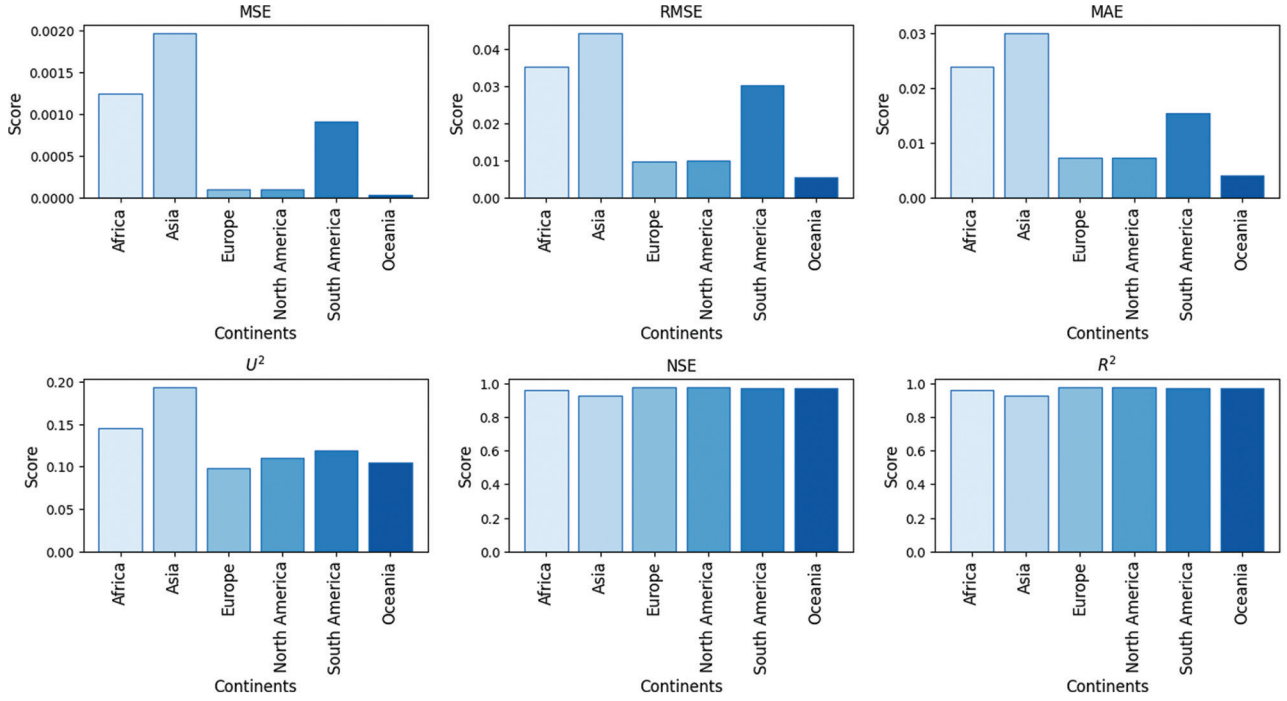

Figure 6. Model performance metrics across continents
Abbreviations: MAE: Mean absolute error; MSE: Mean squared error; NSE: Nash–Sutcliffe efficiency; R^2: Coefficient of determination; RMSE: Root mean squared error; U^2: Theil's U-statistic.

to 23.2 μg/m³ in 2020, reflecting the cumulative impact of air quality regulations in developed nations and improved emission controls in rapidly developing countries. However, substantial heterogeneity persists across countries. Leading polluted nations in 2020 included Qatar (mean exposure: 87.9 μg/m³), Niger (mean exposure: 85.1 μg/m³), Mauritania (mean exposure: 77.3 μg/m³), Bahrain (mean exposure: 67.4 μg/m³), and Egypt (mean exposure: 63.8 μg/m³), all substantially exceeding the WHO guideline of 5 μg/m³ (Table 6).

Notably, countries demonstrating the largest absolute improvements in air quality included Peru (−53.7 μg/m³ reduction; −66.5%), Bolivia (−59.9 μg/m³; −72.3%), Ecuador −30.3 μg/m³; −64.5%), and Singapore (−21.1 μg/m³; −60.3%), reflecting aggressive air quality management, economic restructuring, and technology adoption in these nations. One-sample Z-testing against the WHO guideline of 5 μg/m³ demonstrated that the global population-weighted exposure in 2020 significantly exceeded the health-protective threshold (mean = 25.3 μg/m³; Z = 11.04, $p<0.001$), representing a five-fold departure from guideline values. Regional stratification revealed that South Asia maintained the highest mean exposure (35.2 μg/m³), compared to 9.7 μg/m³ in Europe (two-sample Z = 4.99, $p<0.001$), underscoring substantial geographic inequality in air pollution burden (Figure 7).

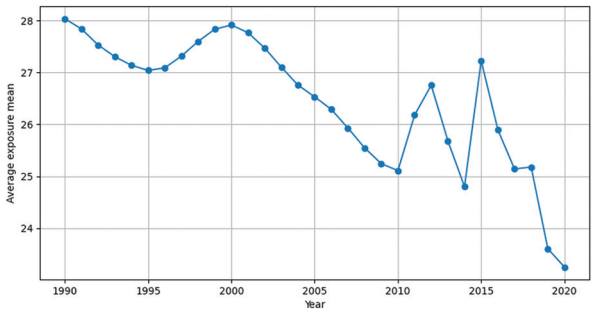

Figure 7. Time-series trajectories of average $PM_{2.5}$ exposure by year

Table 6. Top 10 countries by $PM_{2.5}$ exposure in 1990 and 2020

No.	Country	Mean (1990)	Mean (2020)	Change	% change
1	Bolivia	82.8	22.9	−59.9	−72.34
2	Peru	80.7	27.0	−53.7	−66.54
3	Ecuador	47.0	16.7	−30.3	−64.47
4	Singapore	35.0	13.9	−21.1	−60.29
5	Moldova	34.7	14.8	−19.9	−57.35
6	Guatemala	40.9	21.6	−19.3	−47.19
7	Belarus	32.9	14.6	−18.3	−55.62
8	Ukraine	33.1	14.9	−18.2	−54.98
9	Mexico	32.3	15.0	−17.3	−53.56
10	Pakistan	60.3	43.0	−17.3	−28.69

4.4. Statistical comparison of regional exposure trajectories

One-way analysis of variance (ANOVA) with type III sums of squares, appropriate for the imbalanced regional sample sizes in the 2020 data (Table 7), revealed highly significant differences in $PM_{2.5}$ exposure across geographic regions ($F_{7,85}$ = 16.43; $p<10^{-13}$). *Post hoc* Tukey pairwise comparisons identified significant contrasts between multiple region pairs. South Asia and Europe differed substantially (mean difference = 25.5 μg/m³); however, this pairwise comparison was not statistically significant after Bonferroni correction, as the adjusted 95% confidence interval included zero, despite the large effect size. East Asia and the Pacific significantly differed from the Middle East and North Africa (mean difference = 19.3 μg/m³, p=0.003), similar to the contrast between Europe/North America and South Asia (mean difference = 25.5 μg/m³, $p<0.001$), confirming the geographic heterogeneity of the air pollution burden.

Two-way ANOVA was conducted to evaluate the effects of region and year using balanced subsets for 1990, 2000, 2010, and 2020. The findings revealed highly significant effects of region ($F_{7,339}$ = 46.7; $p<10^{-46}$) and a significant temporal decline in exposure ($F_{3,339}$ = 2.86; p=0.037), with no significant interaction ($F_{21,339}$ = 0.46; p=0.979), indicating that pollution reduction trends are consistent across geographic regions (Table 8). Repeated-measures ANOVA within regions across the four selected years confirmed a significant temporal decline ($F_{3,21}$ = 5.39; p=0.0066), reflecting global progress toward improved air quality despite ongoing challenges in achieving WHO guideline compliance (Figure 8).

4.5. Explainability analysis: Attention weights

Attention-weight analysis examining the learned self-attention mechanisms revealed that, for countries experiencing recent air-quality improvements, the model focused attention weights on historical periods immediately preceding major policy shifts, such as the introduction of vehicle emission standards or industrial sulfur dioxide regulations. The model consistently emphasized these inflection points despite their temporal distance, enabling it to recognize that policy-driven structural breaks in pollution trends provide predictive leverage for future years (Figure 9).

5. Discussion

The present study proposes a temporally rigorous transformer-based forecasting model for predicting global population-weighted $PM_{2.5}$ exposure. The framework

Table 7. One-way analysis of variance with type III sums of squares for an unbalanced dataset

Source	Sum of squares	Degrees of freedom	*F*-statistic	*p*-value (PR[>F])
Intercept	3,543.122	1	24.463	3.77×10^{-6}
C (region)	16,657.836	7	16.43	1.65×10^{-13}
Residual	12,311.031	85	-	-

Table 8. Two-Way analysis of variance examining the effects of region and year on $PM_{2.5}$ exposure

Source	Sum of squares	Degrees of freedom	*F*-statistic	*p*-value (PR[>F])
Region	66,726.83	7	46.7	3.90×10^{-46}
Year	1,754.48	3	2.86	0.0366
Interaction (Region–year)	2,004.52	21	0.46	0.9795
Residual	69,185.06	339	-	-

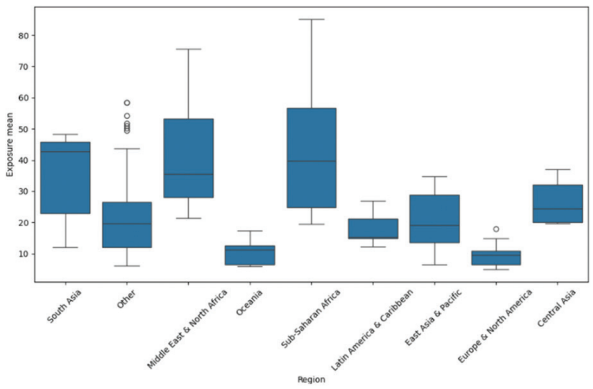

Figure 8. Boxplot of mean $PM_{2.5}$ exposure by region

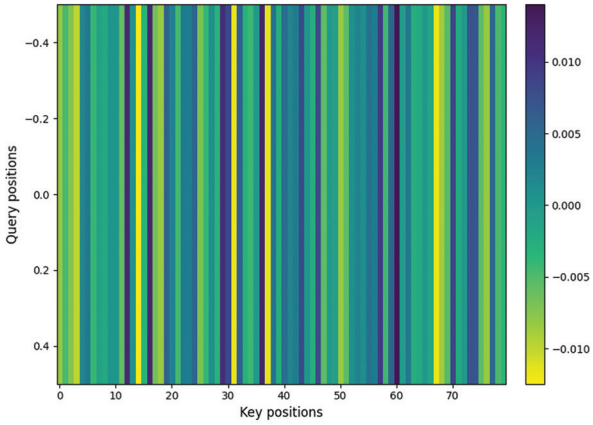

Figure 9. Transformer attention weights

provides realistic performance estimates based on harmonized 30-year panel data of 204 countries and employs time-aware validation protocols that reflect realistic operational forecasting conditions and prevent the inflation of metrics due to temporal leakage. Consistently high performance across all folds (NSE > 0.95; R^2 > 0.95) indicates that $PM_{2.5}$ exposure can be forecasted with reasonable accuracy using historical sequences, supporting the hypothesis that $PM_{2.5}$ exposure is learnable and shaped by long-term regional characteristics, policy contexts, and economic frameworks.

The geographic stratification of findings, with its emphasis on performance excellence in Europe and Oceania and the presence of higher errors in Asia and Africa, reflects a limitation of the original SoGA dataset. In particular, uncertainty exceeds 50% in pre-1998 estimates in dust-dominated countries, caused by the sparsity of data in regions with the highest health burden, suggesting that reported errors in these areas are more due to input data limitations than algorithm failure. This limitation can be mitigated as low-cost sensor networks proliferate and hybrid ground–satellite calibration systems mature, enabling future models to achieve more consistent global precision.

The interpretation of model predictions to guide policy recommendations must be systematically conducted by quantifying uncertainty within each region, especially when the projections are used to estimate health burdens or prioritize interventions in high-uncertainty areas. Another notable development in architecture is the introduction of country embeddings as a set of learnable parameters that are independent of one-hot or integer encoding. The model can achieve better generalization on data-sparse areas and be interpretable by allowing countries to occupy a continuous vector space based on learned similarities in pollution dynamics. Shapley Additive exPlanations and attention-weight analyses validate that embeddings represent meaningful country-level structures, with countries that have similar pollution profiles, geographic locations, or economic structures becoming more clustered during training, thereby enabling the transfer of predictive signals across borders. This method significantly outperforms baseline methodologies, which assume countries as non-moving categorical characteristics, justifying its use in future international environmental forecasting exercises.

The advantage of the transformer-based methodology over the LSTM and CNN baselines can be explained by the similarity of the architecture to the time-series nature of environmental data. Although LSTMs are effective at learning short- to medium-range dependencies, gated information tends to decay over multi-decadal sequences, particularly when seasonal dependencies must be captured over 20–30-year lookback periods. In contrast, transformers can directly attend to historically relevant years; no matter how many years have passed, a policy implemented 15 years ago can still be considered a key predictor of current trends if background circumstances remain relatively similar. Modeling such long-range dependencies is particularly useful for environmental applications, where policy impacts are often lagged and sustained.

The capacity of the proposed framework to model the entire range of the exposure level (i.e., from near guideline compliance to severe pollution) is evidenced by the model's high accuracy in low-pollution environments (e.g., Europe, Oceania, and North America). This capability is essential for policy applications, as it allows differentiation between regions where controls have been effectively implemented, with values close to the guideline, and regions where air quality is in the early stages of improvement. It is important to model both tails of the exposure distribution accurately. The accuracy of the model, with an RMSE of <0.01 $\mu g/m^3$ in low-pollution areas, is sufficient to track the progress toward WHO guideline attainment, enabling countries to adjust policies when exposures approach the safety threshold.

The observed global improvement in air quality, reflected in a median exposure reduction from approximately 30 $\mu g/m^3$ to 23 $\mu g/m^3$ over 30 years, represents the aggregate effect of policies in developed countries and the transfer of technology to rapidly industrializing economies. However, the fact that more than 90% of the global population still lives in areas exceeding the WHO limit of 5 $\mu g/m^3$, and that South Asian populations are 3.6 times more frequently exposed to air pollution than European populations, underscores that, despite progress, air pollution remains a major modifiable risk factor for early mortality and morbidity. The predictive power of the model provides quantitative data to inform policy development: Policymakers can use forecasts to determine whether existing emission control policies will achieve guideline targets within defined time frames or whether more stringent interventions are required.

This framework should be further developed in several complementary directions in future work. First, exogenous predictors (e.g., meteorological covariates such as temperature, precipitation, wind patterns, and satellite aerosol optical depth retrievals) could be integrated into the model to improve the predictive performance, particularly in data-sparse areas where meteorological variation is linked to changes in exposure. Second, ensemble methods

that combine transformer predictions with domain-specific chemical transport model outputs could leverage both the flexibility of data-driven approaches and the constraints of physical models, which may result in greater robustness. Third, probabilistic quantile regression can be incorporated to generate prediction intervals instead of point estimates, which would be beneficial to decision-making under uncertainty and is particularly important in policy contexts where guideline compliance depends on the interpretations of uncertainties. Fourth, sub-national exposure variability could be better predicted by creating spatially disaggregated models; however, this would require access to higher-resolution input data, which is currently unavailable for most low- and middle-income countries.

6. Conclusion

This study demonstrates the applicability and potential of transformer-based designs with learned country embeddings for predicting global, multi-decadal $PM_{2.5}$ exposure. Temporal rigor was imposed using expanding-window cross-validation, which provides more realistic forecasting performance for operational scenarios and strengthens confidence in policy applications. This is achieved through the combination of high predictive performance and interpretable decompositions obtained with Shapley Additive exPlanations and attention-weight analyses, making the framework a methodological breakthrough that enables evidence-based and transparent environmental policy design. The observed geographic heterogeneity in performance can be explained by insufficient underlying data quality rather than limitations of the algorithm; however, the framework also provides a foundation for future improvements as monitoring infrastructure expands and satellite-ground calibration products become more advanced. The ability to predict annual national-level $PM_{2.5}$ exposure with reasonable precision, augmented by explicit quantification of local uncertainty, equips policymakers to set science-based air quality targets, track progress toward WHO guidelines, and design climate-health interventions informed by quantitative exposure forecasts. With the growing global focus on air quality regulation and the increasing health and economic costs of pollution, transformer-based forecasting models provide actionable, interpretable information and can support the transition toward healthier and more sustainable atmospheric conditions.

Acknowledgments

None.

Funding

None.

Conflict of interest

The authors declare they have no competing interests.

Author contributions

Conceptualization: Syed Azeem Inam
Formal analysis: Syed Azeem Inam
Investigation: Syed Azeem Inam, Saddam Umer, Haider Rajput
Methodology: Syed Azeem Inam
Validation: Saddam Umer, Haider Rajput
Writing–original draft: Syed Azeem Inam, Saddam Umer, Haider Rajput
Writing–review & editing: Syed Azeem Inam

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

References

1. Inam SA. A review of artificial intelligence for predicting climate driven infectious disease outbreaks to enhance global health resilience. *Discover Public Health*. 2025;22(1):738.

   doi: 10.1186/s12982-025-01167-4

2. Inam SA, Khan AA, Mazhar T, *et al*. PR-FCNN: A data-driven hybrid approach for predicting PM2.5 concentration. *Discover Artif Intell*. 2024;4(1):75.

   doi: 10.1007/s44163-024-00184-7

3. Inam SA, Zaidi SMH, Khan AA, Ullah S. A neural network approach to carbon emission prediction in industrial and power sectors. *Discover Appl Sci*. 2025;7(6):640.

   doi: 10.1007/s42452-025-07257-x

4. The World Bank. *The Global Health Cost of PM2.5 Air Pollution: A Case for Action Beyond 2021*. Washington, DC: The World Bank; 2022.

   doi: 10.1596/978-1-4648-1816-5

5. Pai SJ, Carter TS, Heald CL, Kroll JH. Updated world health organization air quality guidelines highlight the importance of non-anthropogenic $PM_{2.5}$. *Environ Sci Technol Lett*. 2022;9(6):501-506.

   doi: 10.1021/acs.estlett.2c00203

6. Liu J, He C, Si Y, *et al*. Toward better and healthier air quality:

Global $PM_{2.5}$ and O_3 pollution status and risk assessment based on the new WHO air quality guidelines for 2021. *Glob Chall*. 2024;8(4):2300258.

doi: 10.1002/gch2.202300258

7. Edlund KK, Killman F, Molnár P, Boman J, Stockfelt L, Wichmann J. Health risk assessment of $PM_{2.5}$ and $PM_{2.5}$-bound trace elements in Thohoyandou, South Africa. *Int J Environ Res Public Health*. 2021;18(3):1359.

doi: 10.3390/ijerph18031359

8. Elbarbary M, Oganesyan A, Honda T, *et al*. Systemic inflammation (c-reactive protein) in older Chinese adults is associated with long-term exposure to ambient air pollution. *Int J Environ Res Public Health*. 2021;18(6):3258.

doi: 10.3390/ijerph18063258

9. McDuffie EE, Martin RV, Spadaro JV, *et al*. Source sector and fuel contributions to ambient $PM_{2.5}$ and attributable mortality across multiple spatial scales. *Nat Commun*. 2021;12(1):3594.

doi: 10.1038/s41467-021-23853-y

10. Ghosh R, Causey K, Burkart K, Wozniak S, Cohen A, Brauer M. Ambient and household $PM_{2.5}$ pollution and adverse perinatal outcomes: A meta-regression and analysis of attributable global burden for 204 countries and territories. *PLoS Med*. 2021;18(9):e1003718.

doi: 10.1371/journal.pmed.1003718

11. Southerland VA, Anenberg SC, Harris M, *et al*. Assessing the distribution of air pollution health risks within cities: A neighborhood-scale analysis leveraging high-resolution data sets in the bay area, California. *Environ Health Perspect*. 2021;129(3):37006.

doi: 10.1289/EHP7679

12. Fang T, Di Y, Xu Y, *et al*. Temporal trends of particulate matter pollution and its health burden, 1990–2021, with projections to 2036: A systematic analysis for the global burden of disease study 2021. *Front Public Health*. 2025;13:1579716.

doi: 10.3389/fpubh.2025.1579716

13. Booth A, James P, McGough S, Solaiman E. Cross-regional deep learning for air quality forecasting: A comparative study of CO, NO_2, O_3, $PM_{2.5}$, and PM_{10}. *Forecasting*. 2025;7(4):66.

doi: 10.3390/forecast7040066

14. Benali AAE, Cafaro M, Epicoco I, Pulimeno M, Schioppa EJ. *Just in Time Transformers*. New York: Cornell University; 2024.

15. Muthukumar P, Cocom E, Nagrecha K, *et al*. Predicting PM2.5 atmospheric air pollution using deep learning with meteorological data and ground-based observations and remote-sensing satellite big data. *Air Qual Atmos Health*. 2022;15(7):1221-1234.

doi: 10.1007/s11869-021-01126-3

16. Nedungadi V, Munir MA, Rußwurm M, *et al*. *AirCast: Improving Air Pollution Forecasting through Multi-Variable Data Alignment*. New York: Cornell University; 2025.

17. Oldenburg V, Cardenas-Cartagena J, Valdenegro-Toro M. *Forecasting Smog Clouds with Deep Learning*. New York: Cornell University; 2024.

18. Boogaard H, Pant P, Künzli N. Editorial: Science to foster the WHO air quality guideline values. *Int J Public Health*. 2025;69:1608249.

doi: 10.3389/ijph.2024.1608249

19. Tello-Leal E, Ramirez-Alcocer UM, Macías-Hernández BA, Hernandez-Resendiz JD. Evaluation of deep learning models for predicting the concentration of air pollutants in urban environments. *Sustainability*. 2024;16(16):7062.

doi: 10.3390/su16167062

20. Huang S, Xiao X, Guo H. A novel method for carbon emission forecasting based on EKC hypothesis and nonlinear multivariate grey model: Evidence from transportation sector. *Environ Sci Pollut Res Int*. 2022;29(40):60687-60711.

doi: 10.1007/s11356-022-20120-5

21. Ran Q, Bu F, Razzaq A, *et al*. When will China's industrial carbon emissions peak? Evidence from machine learning. *Environ Sci Pollut Res Int*. 2023;30(20):57960-57974.

doi: 10.1007/s11356-023-26333-6

22. Costantini L, Laio F, Mariani MS, Ridolfi L, Sciarra C. Forecasting national CO_2 emissions worldwide. *Sci Rep*. 2024;14(1):22438.

doi: 10.1038/s41598-024-73060-0

23. Lee JB, Lee JB, Koo YS, *et al*. *Development of a Deep Neural Network for Predicting 6 Hour Average $PM_{2.5}$ Concentrations up to Two Subsequent Days Using Various Training Data*. [Preprint]; 2021.

doi: 10.5194/gmd-2021-356

24. Damkliang K, Chumnaul J. Deep learning and statistical approaches for area-based PM2.5 forecasting in Hat Yai, Thailand. *J Big Data*. 2025;12(1):36.

doi: 10.1186/s40537-025-01079-9

25. Utku A, Can U, Alpsülün M, *et al*. Advancing air quality monitoring: Deep learning-based CNN-RNN hybrid model for PM2.5 forecasting. *Atmosphere*. 2025;16(9):1003.

doi: 10.3390/atmos16091003

26. Zhang Y, Zhang W, Wu B, Yi W. Regional PM2.5 pollution forecasting using a hybrid model based on multi-scales feature fusion and deep learning algorithms. *PLoS One*. 2025;20(10):e0333489.

doi: 10.1371/journal.pone.0333489

27. Yeo I, Choi Y, Lops Y, Sayeed A. Efficient PM2.5

forecasting using geographical correlation based on integrated deep learning algorithms. *Neural Comput Appl*. 2021;33(22):15073-15089.

doi: 10.1007/s00521-021-06082-8

28. Samal KKR, Babu KS, Das SK. Multi-directional temporal convolutional artificial neural network for PM2.5 forecasting with missing values: A deep learning approach. *Urban Clim*. 2021;36:100800.

doi: 10.1016/j.uclim.2021.100800

29. Potavijit C, Pattarapanitchai P, Nontapa C. A hybrid deep learning model for forecasting PM2.5 concentrations in Northern Thailand from satellite images. *Glovento J Integr Stud*. 2025;1(1):6.

doi: 10.63665/gjis.v1.6

30. Wu Y, Wang X, Wang M, Liu X, Zhu S. Time-series forecasting of $PM_{2.5}$ and PM_{10} concentrations based on the integration of surveillance images. *Sensors (Basel)*. 2024;25(1):95.

doi: 10.3390/s25010095

31. Dairi A, Harrou F, Sun Y. Efficient Deep Learning-Driven Approach for PM2.5 Forecasting at Different Locations in Spain. In: *2021 IEEE 3rd Eurasia Conference on Biomedical Engineering, Healthcare and Sustainability (ECBIOS)*. IEEE; 2021. p. 173-178.

doi: 10.1109/ECBIOS51820.2021.9510462

32. Vignesh PP, Jiang JH, Kishore P. Predicting $PM_{2.5}$ concentrations across USA using machine learning. *Earth Space Sci*. 2023;10(10):1-15.

doi: 10.1029/2023ea002911