Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Policy-Oriented Multiresolution CNN–Transformer Fusion for Synoptic-Scale $PM_{2.5}$ Prognostication and Low-Emission Zone Stratification in Hypermetropolitan Aerosol Regimes

Reza Rabie
University of Tehran
Saeed Karimi
karimis@ut.ac.ir

University of Tehran

Article

Keywords: Multiscale Spatiotemporal Forecasting, $PM_{2.5}$ Air Quality Prediction, LEZ Stratification, Hybrid CNN–Transformer, Urban Air Quality Management

Posted Date: November 20th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11486706/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Policy-Oriented Multiresolution CNN–Transformer Fusion for Synoptic-Scale $PM_{2.5}$ Prognostication and Low-Emission Zone Stratification in Hypermetropolitan Aerosol Regimes

Reza Rabie [a], Saeed Karimi [a*]

[a] College of Engineering, Graduate Faculty of Environment, University of Tehran, Tehran, Iran
*Corresponding author

Abstract

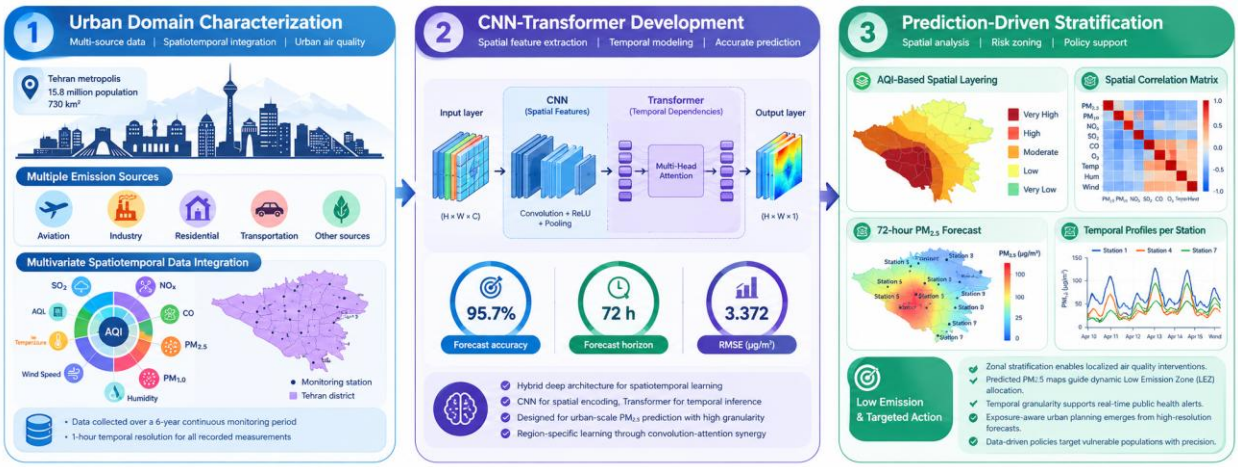
Megacities represent a paradigmatic nexus of anthropogenic density and resource accretion, wherein the proliferation of $PM_{2.5}$ aerosols, characterized by their polymorphic chemical speciation, exerts multifactorial pathophysiological burdens on urban cohorts. Prevailing regulatory frameworks, which homogenize monitoring and management protocols across urban hierarchies, fail to accommodate the spatiotemporal heterogeneities inherent to hypermetropolises, where the confluence of stationary point sources and high-flux mobile emitters generates non-Gaussian dispersion fields. To redress this lacuna, we introduce a hybrid CNN-Transformer ensemble engineered to concurrently resolve high-frequency, localized emission pulses via hierarchical convolutional filters and protracted autocorrelative regimes through multi-head self-attention layers. Leveraging a sextennial dataset of meteorological covariates and station-based $PM_{2.5}$ readings across thirteen delineated districts of the Tehran megacity, our model extends predictive horizons to 72 hours, quadrupling the temporal ambit of antecedent sub-6-hour studies, while preserving sub-3 μg/m³ error magnitudes. The network attained an R^2 exceeding 0.95, an RMSE of 3.372 μg/m³, and an MSE of 11.37 μg/m³, marking the most stringent performance metrics reported to date in urban air-quality prognostication. Spatioclustering analyses revealed discordance between administrative zoning and aerosol concentration gradients, underscoring the necessity for delineating LEZs within megacities to enable targeted mitigation. Moreover, the model's granular outputs facilitate the construction of dynamic public-health protection maps, enabling preemptive exposure abatement for vulnerable demographics, particularly pediatric cohorts, and the temporal-spatial identification of low-pollution refugia. these findings advance the methodological frontier of precision air quality management and furnish a robust decision-support scaffold for the formulation of bespoke regulatory standards tailored to the idiosyncratic pollutant topographies of megacities.

Keywords: Multiscale Spatiotemporal Forecasting, $PM_{2.5}$ Air Quality Prediction, LEZ Stratification, Hybrid CNN–Transformer, Urban Air Quality Management

Reza Rabie
rabie.reza@ut.ac.ir

*Saeed Karimi
karimis@ut.ac.ir
Postal address: Faculty of Environment, University of Tehran, Azin Avenue, Ghods Street, Enghelab Square, Tehran, Iran
Postal code:141785311, Fax:+982166409348

Graphical Abstract



1. Introduction

The advent of pervasive urbanization over the past several decades has precipitated an unprecedented escalation in chronic inhalational exposure to a heterogeneous assemblage of xenobiotic aerosols and gaseous byproducts, namely respirable particulate matter ($PM_{2.5}$ and PM_{10} fractions), volatile organic compounds (VOCs), tropospheric ozone (O_3), and sulfur dioxide (SO_2), each of which exerts multifaceted pathophysiological perturbations within the cardiopulmonary and systemic milieu [1], [2]. Contemporary epidemiological estimates indicate that in excess of 80 % of the global urban populace in low and middle-income countries endures deleterious sequelae attributable to ambient air pollution [3]; within this cohort, particulate matter emerges as the preeminent vector of morbidity and mortality, with over 7.3 billion individuals are exposed to unsafe levels of $PM_{2.5}$ [4]. The intricate bioreactivity of these particulates facilitates translocation across alveolar–capillary barriers, engendering oxidative stress, endothelial dysfunction, dysregulated autonomic modulation, and proinflammatory cytokine cascades, thereby potentiating a spectrum of respiratory, cardiovascular, and neurocognitive pathologies [5], [6], [7]. This pervasive environmental burden underscores the exigency for integrative mitigation strategies predicated on advanced spatiotemporal forecasting, mechanistic toxicology, and precision public health interventions [8], [9].

A mounting corpus of evidence delineates a cogent, quasi-linear correlation between hyperurbanization and exacerbated aerosolized particulate burdens within the troposphere [10], [11]. Since the turn of the millennium, the cadre of megacities, urban agglomerations surpassing 10 million inhabitants, has burgeoned from 27 loci in 2000 to 33 as of 2025, and projections portend an escalation to approximately 47 hyperconurbations harboring in excess of 600 million denizens by mid-century [12]. Presently, over 82 % of these urban behemoths contend with

ambient particulate concentrations that routinely transgress WHO guidelines [13]; for instance, Lahore exhibited a mean $PM_{2.5}$ loading of 102.1 µg/m³ in 2024, while Delhi's mass-weighted $PM_{2.5}$ index lingered between 90 and 100 µg/m³ from 2019–2024, and Dhaka averaged 78 µg/m³ over the same quinquennium [14], [15], [16].

$PM_{2.5}$ functions as the paradigmatic indicator pollutant within global megacities, exceeding the WHO 24-hour guideline threshold on roughly 65% of days over the past decennium [17]. In their seminal receptor-modeling work, G. Zhu et al. [18] demonstrated that on-road vehicular emissions contribute approximately 48% of the cumulative $PM_{2.5}$ burden in Beijing's urban airshed, underscoring the preeminence of mobile-source plumes in the total aerosol mass balance. Likewise, Sandeep et al. [19] apportioned some 32% of ambient fine particulates to industrial combustion and ancillary economic activities in Mumbai, thereby delineating the substantial role of stationary anthropogenic conglomerates. Complementing these source-attribution studies, Ali et al. and Hossain et al. [20], [21] elucidated how orographic confinement, mesoscale circulations, and anomalous synoptic patterns engender marked spatiotemporal heterogeneity in $PM_{2.5}$ concentrations across South Asian megacities. Consequently, urban air pollution emerges as a highly nonlinear, multifactorial phenomenon, wherein the interplay of anthropogenic emissions, topographical constraints, and climatic variabilities defies reduction to any single explanatory variable or monolithic framework [22].

The prevailing paradigm for ambient air quality surveillance and governance, though codified through heterogeneous municipal and sovereign statutes, universally derives from the U.S Environmental Protection Agency's Air Quality Monitoring and Management Program architecture [23]. Under this schema, monitoring nodes are sited via a multi-criteria optimization that accounts for demographic density, proximate emission inventories, prevailing aerodynamic

flux vectors, and physiographic relief [24]. Standard operating procedures typically dictate a heuristic density of one Federal Reference Method sampler per 250 000 inhabitants or per 1000 km² in metropolitan regions; nonetheless, station count and spatial dispersion are ultimately scaled to the demographic and areal expanse of the jurisdiction [25]. Data assimilation protocols then aggregate multi-node observations, yet the locus of maximal concentration, often a hotspot monitor sited adjacent to high-throughput corridors or industrial complexes, is exalted as the referent datum for computing the composite Air Quality Index [11]. While this single-site exceedance algorithm furnishes actionable real-time diagnostics and tactical foresight in small conurbations (< 1 million inhabitants), its transposition to hyperdense megalopolises engenders representativeness error and obfuscates intra-urban heterogeneity, thereby attenuating its utility for strategic urban-scale mitigation planning [26].

The exigency of high-fidelity spatiotemporal characterization of urban aerosol dynamics has precipitated the deployment of advanced machine learning and deep learning paradigms, which now underpin state-of-the-art prognostic frameworks for ambient pollutant dispersion [27], [28]. Contemporary artificial intelligence architectures, spanning long short-term memory (LSTM) RNNs, multi-scale convolutional neural networks (CNNs), graph-convolutional hybrids, and attention-augmented Transformer modules, have demonstrably attained sub-hourly predictive precision at subkilometer resolution [29], [30]. For instance, in their landmark analysis of $PM_{2.5}$ heterogeneity across Beijing, Kumar et al. [31] engineered a GRU–LSTM–attention ensemble that resolved intra-urban concentration gradients with a mean absolute error below 7 µg/m³, simultaneously revealing a pronounced pollution–income decoupling between affluent districts and peri-urban townships. Similarly, Ahmad et al. [32] leveraged a spatiotemporal BiGRU–1DCNN network to quantify pediatric inhalation exposures in Delhi, documenting a 35 %

elevation in cumulative $PM_{2.5}$ dose among children residing in socioeconomically marginalized wards. Notwithstanding these advances, extant models suffer from cascading error accumulation over extended horizons, forecast skill precipitously declines beyond 6 hours, thereby constraining the formulation of robust, horizon-agnostic policy interventions [33].

The principal objective of the present investigation is to proffer a novel prognostic schema predicated upon a CNN–Transformer hybrid architecture, engineered to concurrently distill high-frequency, short-latency perturbations and protracted autocorrelative motifs within urban $PM_{2.5}$ time series. Whereas extant literature predominantly constrains predictive horizons to sub-6-hour windows, our framework seeks to substantially elongate forecast intervals to 24, 48, and 72-hour cadences, thereby ameliorating cumulative error proliferation and furnishing enhanced temporal generalizability. Beyond mere point estimates, the model integrates an interpretative analytics module, leveraging attention-weight disaggregation and saliency-map visualization, to elucidate the relative import of emission surges, meteorological forcings, and land-use heterogeneities upon forecast trajectories. By synthesizing localized convolutional kernels with global self-attention mechanisms, the proposed hybrid engenders synergistic feature hierarchies that underpin both the granularity of short-term fluctuations and the coherence of long-term trends. This dual-pronged approach not only elevates the precision and reliability of $PM_{2.5}$ forecasts across heterogeneous spatial strata but also furnishes an actionable decision-support scaffold for the design and optimization of targeted air pollution mitigation strategies in hyperdense urban milieus.

Over a sextennial observation window, we amalgamated synoptic-scale meteorological covariates and station-based $PM_{2.5}$ measurements from thirteen discrete micro–urban agglomerations within the Tehran megametropolis to construct a high-resolution spatiotemporal dataset. Thereupon, we

instantiated a hybrid convolutional neural network–Transformer ensemble designed to hierarchically encode short-latency emission pulses via localized convolutional kernels while concurrently seizing protracted autocorrelative structures through multi-head self-attention. Following rigorous hyperparameter optimization and backpropagation training across the aggregated epochs, the model projected fine-grained 72-hour $PM_{2.5}$ concentration trajectories for each region. Predictive fidelity was quantitatively appraised using a battery of error metrics, mean absolute error, root mean square error, and Pearson correlation coefficient, against benchmark RNN and pure CNN baselines. The superior performance of our hybrid architecture not only elucidates nuanced spatiotemporal heterogeneities in particulate dispersion but also furnishes a robust decision-support mechanism for informed policy interventions, thereby advancing the paradigm of precision air quality management in hyperdense urban environments.

2. Material and method

Fig.1 schematically delineates the procedural framework of this investigation, encompassing the stratified methodological architecture and operative paradigms employed throughout. This research operationalizes a data-centric paradigm predicated on spatiotemporally disaggregated atmospheric pollutant metrics and synoptic-scale meteorological covariates within the urban morphology of Tehran. The analytical scaffolding is anchored in a hybridized deep learning schema, wherein CNNs are synergistically integrated with Transformer-based attention mechanisms to architect an advanced spatiotemporal inference apparatus. This integrative model is designed to autonomously discern quasi-stationary pollution regimes and transient perturbational anomalies, thereby establishing a neuro-symbolic paradigm for high-resolution environmental monitoring and anomaly-aware forecasting at the intra-urban mesoscale.

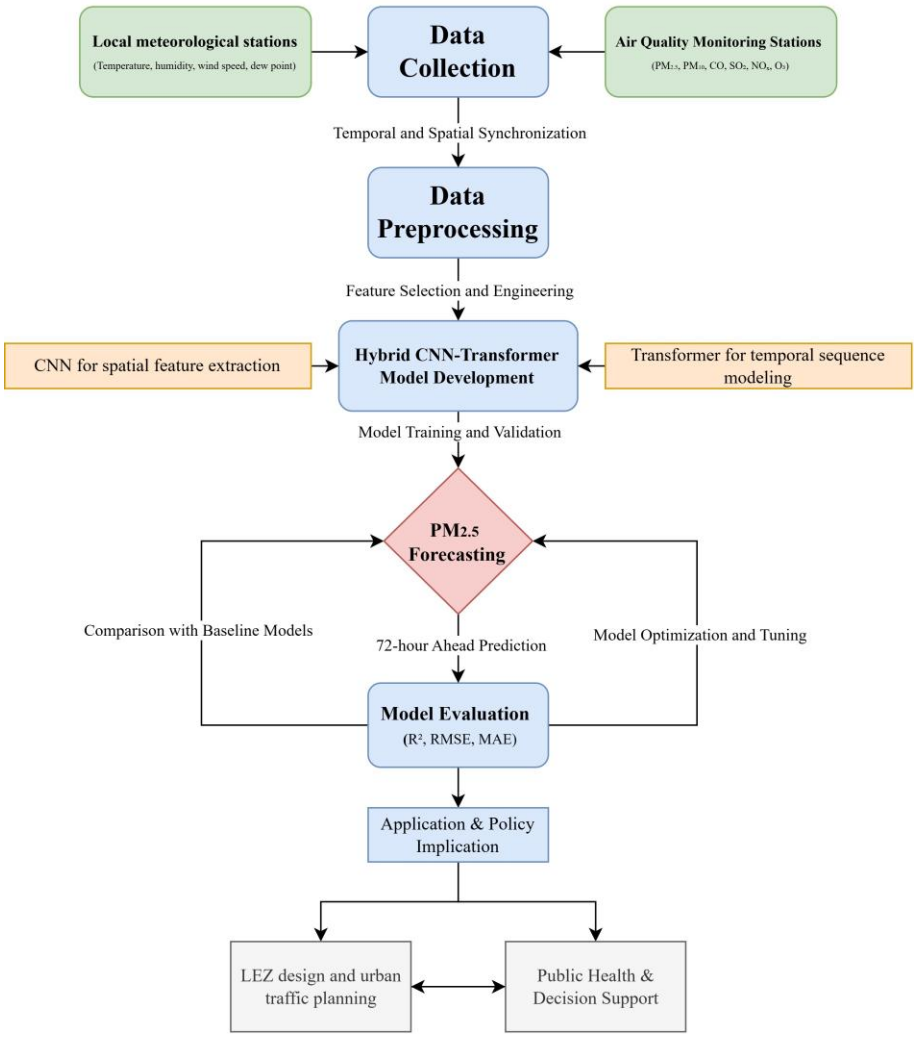


Figure 1- Overview of the methodological workflow for $PM_{2.5}$ prediction, spatial analysis, and policy-oriented application.

2.1. Study area

The empirical domain of this investigation is the metropolitan agglomeration of Tehran, the geopolitical and administrative nucleus of Iran. As the most demographically saturated and atmospherically compromised capital in the West Asian macroregion, Tehran sustains a population exceeding 15.8 million inhabitants, encompassing an approximate areal extent of 730 km² [34]. The city's atmospheric composition is predominantly modulated by anthropogenic emission vectors, with mobile sources (e.g., vehicular exhaust) contributing an estimated 63.8% to ambient

PM burden, followed by stationary point sources (e.g., industrial and residential combustion processes) at 30.5%, and aeolian or biogenic sources accounting for the residual 5.7% [35].

Morphologically, Tehran occupies a heterogeneous topographical gradient delineated by orogenic formations to the north and sedimentary plains to the south, resulting in elevational differentials ranging from ~1,000 to >2,000 meters above sea level across its spatial extent. This physiographic dichotomy, when juxtaposed with the region's thermally stratified atmospheric boundary layer and synoptic stagnation episodes, engenders a highly nonlinear and spatially variegated $PM_{2.5}$ dispersion regime. Consequently, the prognostication of $PM_{2.5}$ concentrations necessitates the deployment of computational models with high-dimensional representational capacity and spatial discriminability [36]. Given the city's intricate urban sprawl and heterogeneous pollution landscape, intra-urban stratification based on administrative zoning (22 legally demarcated municipal districts) is indispensable for region-specific exposure assessment. The spatial geolocation of Tehran and the spatial distribution of its real-time air quality surveillance nodes are illustrated in **Fig.2**.

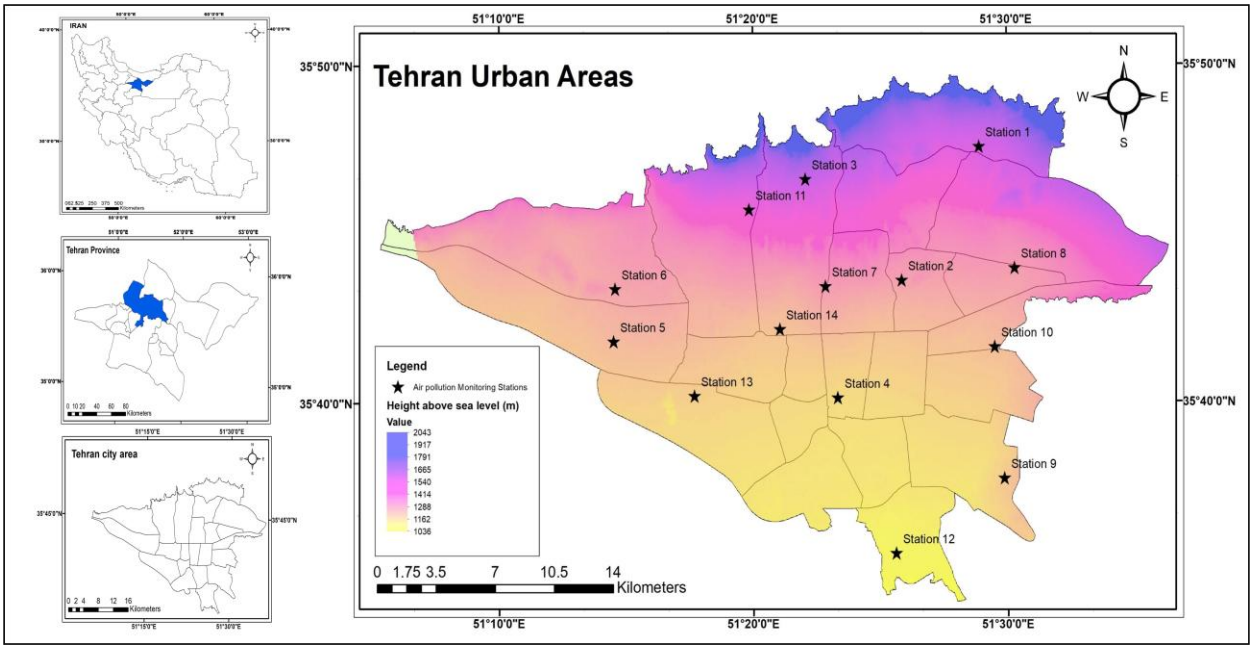


Figure 2- Geographical location of the Tehran metropolitan area within Iran, indicating the spatial extent of the study and the distribution of air quality monitoring stations.

2.2. Data collection and preprocessing

The empirical substrate utilized in this study comprises high-resolution, multivariate spatiotemporal datasets encompassing both atmospheric pollutant concentrations and concomitant meteorological parameters. Specifically, the air quality variables include CO, O_3, SO_2, NO, NO_2, NO_x, coarse particulate matter (PM_{10}), and fine particulate matter ($PM_{2.5}$), while the meteorological covariates consist of ambient temperature (T), wind velocity (WS), relative humidity (RH), and dew point temperature (DP). These measurements were procured from a network of 14 continuous ambient air quality monitoring stations strategically distributed across the Tehran metropolitan macrozone (as spatially indexed in **Fig.2**). The data acquisition protocol spanned a continuous temporal horizon of six years, from March 2018 to February 2025, with an hourly sampling frequency, yielding a canonical data matrix with 52,560 temporally indexed observations and 12 feature dimensions per station, amounting to a total of 630,720 records per spatial node. **Fig.3** shows an example of a monitoring station that measured and recorded the data used in this study.



Figure 3- On-site view of Tehran's air quality monitoring equipment.

Given the protracted observation interval, sporadic data discontinuities emerged due to instrumentational perturbations, maintenance downtime, and sensor degradation. These lacunae were addressed via temporal imputation methodologies, wherein the most recently observed values were propagated forward, constrained by spatiotemporal autocorrelation and domain-informed pollutant behavior assumptions. The imputation strategy presumes the quasi-stationarity of pollutant dynamics in the absence of episodic anomalies. Notably, the incidence of missingness constituted less than 1% of the total data volume, rendering its influence statistically negligible on the convergence behavior and generalization capacity of the predictive model.

To harmonize the disparate feature magnitudes arising from heterogeneous physical units and to mitigate the undue influence of outlier-induced variance inflation, a robust normalization protocol was enacted. Specifically, Robust Scaler transformation, parameterized by the interquartile range (IQR), was

operationalized (as expressed in **Eq.1**), due to its resilience against extremal values and its efficacy in preserving intrinsic feature relationships.

$$scaled\ value = \frac{x - median}{IQR} \quad (1)$$

Outlier diagnostics were concurrently performed as part of anomaly detection preprocessing, ensuring that non-Gaussian or impulsive deviations were either rectified or retained based on contextual relevance to pollution dynamics. Subsequent to data sanitization, the preprocessed dataset was partitioned into mutually exclusive subsets following a stratified hold-out schema: 80% of the samples were allocated to the training set to facilitate model parameter optimization, 15% to the validation set for hyperparameter tuning and overfitting mitigation, and 5% to the test set for post hoc performance appraisal. This tripartite data segregation paradigm ensures both robust learning fidelity and unbiased performance estimation under cross-sectional and temporal variability.

2.3. CNN-Transformer model

Following rigorous data preprocessing and normalization procedures, the architectural implementation phase commenced with the construction of a composite deep learning framework, integrating CNNs with Transformer-based encoders to achieve multiscale spatiotemporal inference. The overarching structure of the proposed model, schematically delineated in **Fig.4** was meticulously designed to harness the complementary proficiencies of CNNs in local feature extraction and temporal localization, and the Transformer architecture in modeling high-order, non-local temporal dependencies through multi-head self-attention and dynamic context aggregation. The rationale behind this dual-path architecture stems from the intrinsic heterogeneity and temporal volatility of $PM_{2.5}$ pollution patterns, which are influenced by non-stationary emission sources, mesoscale meteorological perturbations, and anthropogenic activity cycles such as workdays, holidays, and seasonal climatic transitions [37].

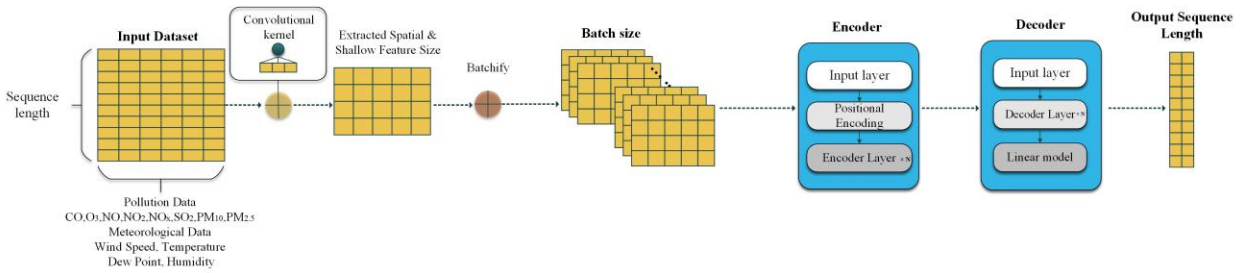


Figure 4- Composite architecture of the CNN–Transformer model for spatiotemporal $PM_{2.5}$ forecasting.

2.3.1. CNN Layer

The structural configuration of the CNN component, delineated in **Fig. 5**, constitutes a critical architectural module for hierarchically extracting salient temporal and spatial representations from multivariate input sequences. In this study, a dual-layer convolutional architecture has been deployed, wherein the first and second convolutional layers employ 32 and 64 filters, respectively, each instantiated with learnable kernels of fixed dimensionality $P \times n$, where n denotes the feature vector dimensionality corresponding to each temporal observation in the input sequence T_{ij} (with $i = 1, \dots, m$ and $j = 1, \dots, n$). The convolutional operation serves as a localized translational filter, enabling the model to autonomously discern high-frequency temporal motifs and transient pollution events embedded within the input tensor. Mathematically, the convolutional transformation is articulated as **Eq. 2**:

$$f^{(k)} = \phi(W^{(k)} \otimes X + b^{(k)} \tag{2}$$

where $f^{(k)}$ is the resulting feature map from the $k-th$ kernel, ϕ denotes the rectified linear unit (ReLU) activation function, $W^{(k)}$ is the convolution kernel, $b^{(k)}$ the bias term, and $\otimes$ indicates the convolution operation over the temporal axis.

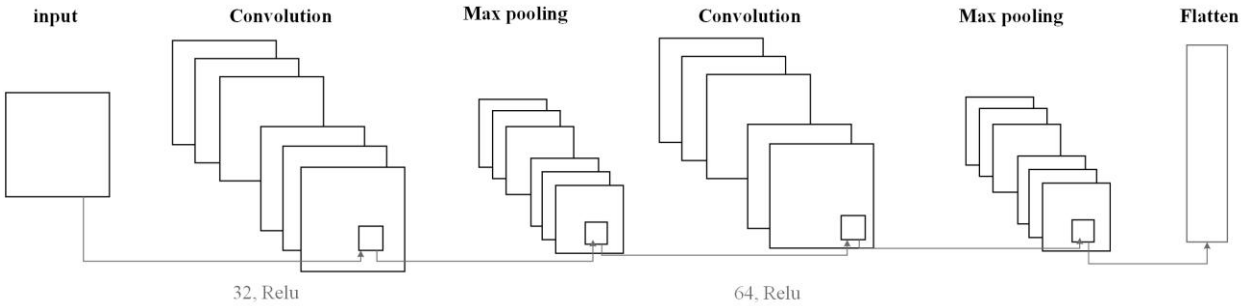


Figure 5- Structural schematic of the CNN layer for spatial feature extraction.

The ReLU activation introduces non-linearity, ensuring that the network can capture nontrivial interactions and heteroscedasticity in pollutant dynamics. Following each convolutional operation, a max-pooling layer is employed to perform non-linear downsampling, whereby only the most prominent activation in each local receptive field is retained. This not only reduces the spatial resolution and computational burden, but also reinforces translational invariance in feature extraction. Subsequently, a flattening operation is performed to vectorize the multi-dimensional feature maps, thereby rendering the CNN outputs compatible for downstream transmission to fully connected or sequence-based architectures such as Transformers.

To ameliorate overfitting and enhance the model's generalization across heterogeneous temporal profiles and pollution regimes, a dropout regularization mechanism with a dropout rate of 0.3 is interposed post-convolution. This stochastic regularization strategy deactivates a randomly sampled subset of neurons during training, compelling the network to learn redundant and robust representations [38]. The CNN module, therefore, serves as a localized pattern encoder, transforming raw, high-dimensional time series inputs into a compact and information-rich latent space that encapsulates the spatiotemporal volatility of $PM_{2.5}$ concentrations. These representations subsequently constitute the input embeddings for the Transformer encoder, thereby facilitating the integration of both local (short-term) and global (long-term) dependencies in the forecasting pipeline.

2.3.2. Transformer layer

The core architecture of the transformer model, depicted in **Fig.6**, is composed of two fundamental components: the encoder and the decoder. In this framework, the features extracted by the preceding CNN model, deemed valuable for predicting $PM_{2.5}$ pollution values, are transformed into data of dimension D_{model} through a fully connected network.

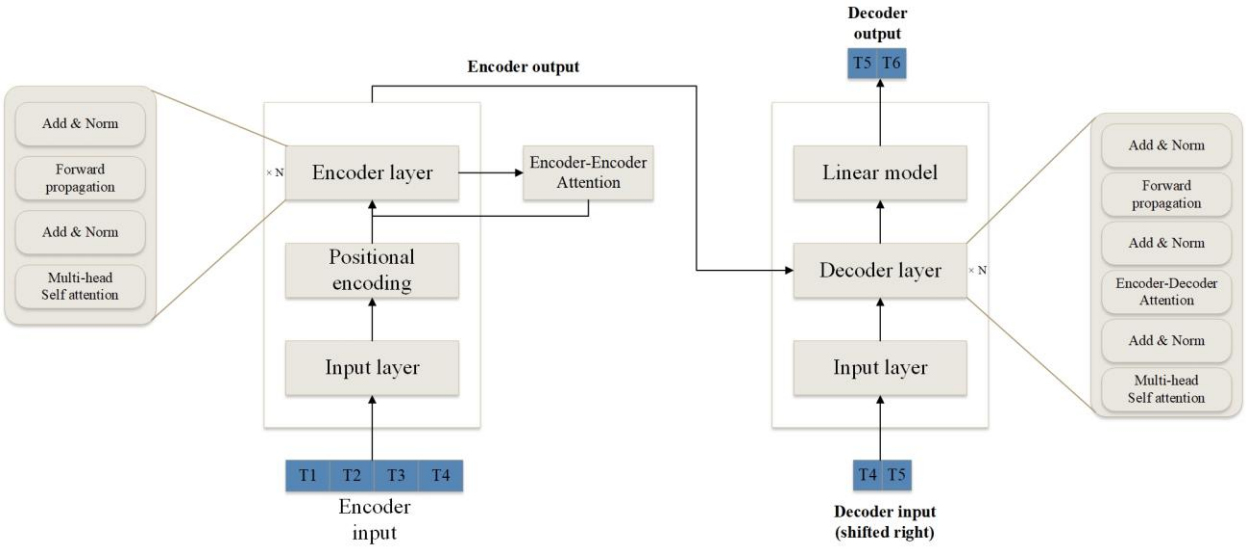


Figure 6- Architectural schema of the Transformer layer, detailing encoder–decoder attention and multi-head self-attention mechanisms.

Recognizing the criticality of temporal sequencing in our data, positional encoding is subsequently applied. For our time series data, assuming a length of n, each sample's specific position in the sequence is denoted by t, while i represents the i-th dimension at position t. The vector R signifies the position of the i-th vector at position t. To derive the R vector, we utilize the function F, as detailed in **Eq.3**. This positional encoding ensures that the model can effectively incorporate the sequential nature of the data into its learning process, enabling it to understand the temporal relationships between different data points.

$$\overrightarrow{P_t^{(i)}} = F(t)^i = \begin{cases} \sin(\omega_k . t), if\ i\ is\ even \\ \cos(\omega_k . t), if\ i\ is\ odd \end{cases} \quad (3)$$

In this equation, the variables i and t represent the dimensions and positions within the sequence, respectively. The frequency w is computed using the **Eq.4**, which K ε {1, … , $[\frac{d_{model}}{2}]$.

After this step, the position vector, which contains sequential information from the input time series data, undergoes an element-wise summation. The resulting output from this layer is then fed into the encoder layer as input. Within the encoder layer, there exists a multi-head self-attention layer and a fully connected forward layer as sub-layers, each followed by an additional element-wise summation and a normalization layer, as illustrated in **Fig.6**.

In the attention function, the input comprises queries, dimension keys (d_k), and dimension values (d_v). To facilitate concurrent computations, sets of queries and keys are packed into Q and K matrices, while sets of values are packed into the V matrix. Subsequently, the attention function A is computed according to **Eq.4.** Given that the multi-head attention mechanism divides attention into h heads, the final output is determined by connecting the outputs of all heads together. This combined output is obtained using **Eq.5**. This approach enables the model to effectively capture intricate patterns and dependencies within the data by leveraging multiple attention heads.

$$A(Q,K,V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right).V \tag{4}$$

$$M(Q.K,V) = Concat(head_1, \dots, head_h).W^O \tag{5}$$

Where $head_i = A(QW_i^Q, KW_i^K, VW_i^V)$. We note that $W_i^Q \varepsilon R^{d_k \times d_{model}}$, $W_i^K \varepsilon R^{d_k \times d_{model}}$ and $W_i^V \varepsilon R^{d_V \times d_{model}}$ are parameters projected linearly from Q , K , and W. $W^O \varepsilon R^{hd_v \times d_{model}}$ is also a projected parameter matrix used to remap the result of concatenation to d_{model} dimensions.

To extract shallow information from the coding sequence, a multi-head attention mechanism known as Encoder-Encoder is employed between different coding layers [39], [40]. This mechanism facilitates the modification of information learned by various coding layers. Subsequently, a sequence of d_{model} is forwarded as input to the decoder layer, after executing all preceding steps. The structure of the decoder layer closely resembles that of the encoder. In addition to the sub-layers utilized in the encoder section, an additional sub-layer is incorporated into the decoder section, referred to as encoder-decoder. The purpose of this sub-layer is to apply the attention mechanism to the encoding output, enhancing the model's ability to capture relevant information from both the input and output sequences. Upon completion of the transformer model process, a fully connected linear layer transforms the output of the final decoder into the target sequence, with dimensions of d_{model}. This transformation prepares the model to generate predictions based on the learned patterns and dependencies within the input data.

The output generated from the decoder is then passed through a dense layer to produce the anticipated outputs, which represent the predictions of $PM_{2.5}$ pollution concentration. In the hybrid CNN-Transformer model, a variety of hyperparameters were employed, each playing a crucial role in shaping the model's performance. The specific values utilized for these hyperparameters are presented in the accompanying **table 1**. These values were carefully selected and fine-tuned to optimize the model's predictive accuracy and overall performance.

Table 1- Optimal hyperparameters tuned for the model

Hyperparameter	Value
Number of epochs	300
Batch size	1024
Look_back	72
Learning rate	0.001
Number of encoder layer	2
Number of decoder layer	2
Number of kernel of 1-D convolution	32
Number of dimensions of inputs for attention	32
Number of dimensions of outputs for attention	64
Number of neurons for deep neural network layer 1	256
Dropout rate for deep neural network layer 1	0.4
Number of neurons for deep neural network layer 2	128
Dropout rate for deep neural network layer2	0.2
optimizer	Adam
Loss function	MSE

After fully developing the hybrid model and adjusting the relevant hyperparameters, predictions were made for the concentration of $PM_{2.5}$ pollutant over the next 72 hours in each of the studied areas. The results obtained from the model were then analyzed and evaluated using statistical indicators. A comprehensive examination of the model's structure and its impact on the quality of results was conducted, along with discussions on applicable control approaches utilizing the predicted values from the model.

2.4. applications and policy making

To rigorously evaluate the spatial autocorrelation and inter-regional dependency of $PM_{2.5}$ concentrations across the Tehran megacity, Moran's I method was used, a robust geostatistical index widely utilized for quantifying the degree of spatial clustering among areal units. This measure provides critical insight into whether high or low pollution values exhibit spatial contiguity or are randomly distributed across the urban topology. The global Moran's I statistic is mathematically expressed as **Eq.6**.

$$I = \frac{N}{W} \cdot \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{N}(x_i - \bar{x})^2} \quad (6)$$

where N denotes the total number of spatial units (monitoring stations), x_i and x_j represent the $PM_{2.5}$ concentration values at locations i and j, respectively, $\bar{x}$ is the mean $PM_{2.5}$ concentration, W_{ij} is the spatial weight matrix element defining the spatial relationship (e.g., contiguity or inverse distance) between locations i and j, and $\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij}$ is the sum of all spatial weights. A positive and statistically significant Moran's I indicates spatial clustering, where similar $PM_{2.5}$ values aggregate in specific regions, while a negative value implies spatial dispersion. In the context of Tehran, this analysis elucidates the heterogeneity of pollution propagation mechanisms and facilitates the identification of localized atmospheric stagnation zones or pollution sinks. By integrating Moran's I within our methodological framework, we enhance the model's spatial explanatory power and offer critical input for spatial policy interventions such as zonal traffic control and localized air quality management.

To quantify the public health implications of predicted $PM_{2.5}$ concentrations, we translated raw pollutant levels into the standardized Air Quality Index (AQI) using the U.S. EPA's breakpoint concentration formula, which allows for the classification of air pollution severity and associated health risks. The AQI for $PM_{2.5}$ was computed using the **Eq.7**.

$$AQI = \frac{I_{High} - I_{Low}}{C_{High} - C_{Low}}(C - C_{Low}) + I_{Low} \quad (7)$$

where C is the predicted $PM_{2.5}$ concentration, C_{High} and C_{Low} are the breakpoint concentrations enclosing C, and I_{High}, I_{Low} are the corresponding AQI values. This transformation enabled spatially and temporally resolved evaluation of health hazard severity across the Tehran metropolitan area. Based on AQI categorizations, we identified critical exposure windows and geographic zones with elevated health risks. This facilitates evidence-based policy making, including the design of emergency response protocols, risk communication strategies for vulnerable populations, and the implementation of dynamic low-emission zones to mitigate acute public health impacts.

3. Results

During the spatiotemporal evaluation window, the hybridized CNN-Transformer framework yielded prognostic outputs delineating the dynamic evolution of $PM_{2.5}$ concentration fields across the Tehran metropolitan agglomeration with exceptional granularity. The inferential model captured pronounced intra-urban heterogeneity, manifesting in temporally phase-shifted pollutant plumes with anisotropic dispersion gradients, particularly accentuated in thermally stratified boundary layers proximal to industrial peripheries and high-density traffic corridors. Predictive fidelity was substantiated by multi-metric validation, wherein statistical indicators such as RMSE, MAE, and R^2 evidenced strong concordance between simulated and empirical values. These outputs served as the substrate for subsequent geostatistical interpolation and exposure stratification across distinct urban morphologies.

3.1. Spatial Attribution of Feature Influence on $PM_{2.5}$

Fig.7 elucidates a multidimensional attributional topology wherein the differential salience of precursor pollutants, O_3, CO, NO, NO_2, NO_x, SO_2, and PM_{10}, is delineated across monitoring loci, revealing spatiotemporal anisotropies in $PM_{2.5}$ genesis pathways. In high-traffic southern nodes, elevated feature weightings for NO_x and CO underscore a combustion-dominated pollution regime, while peripheral stations exhibit amplified O_3 relevance, signaling secondary aerosol formation via photochemical oxidation.

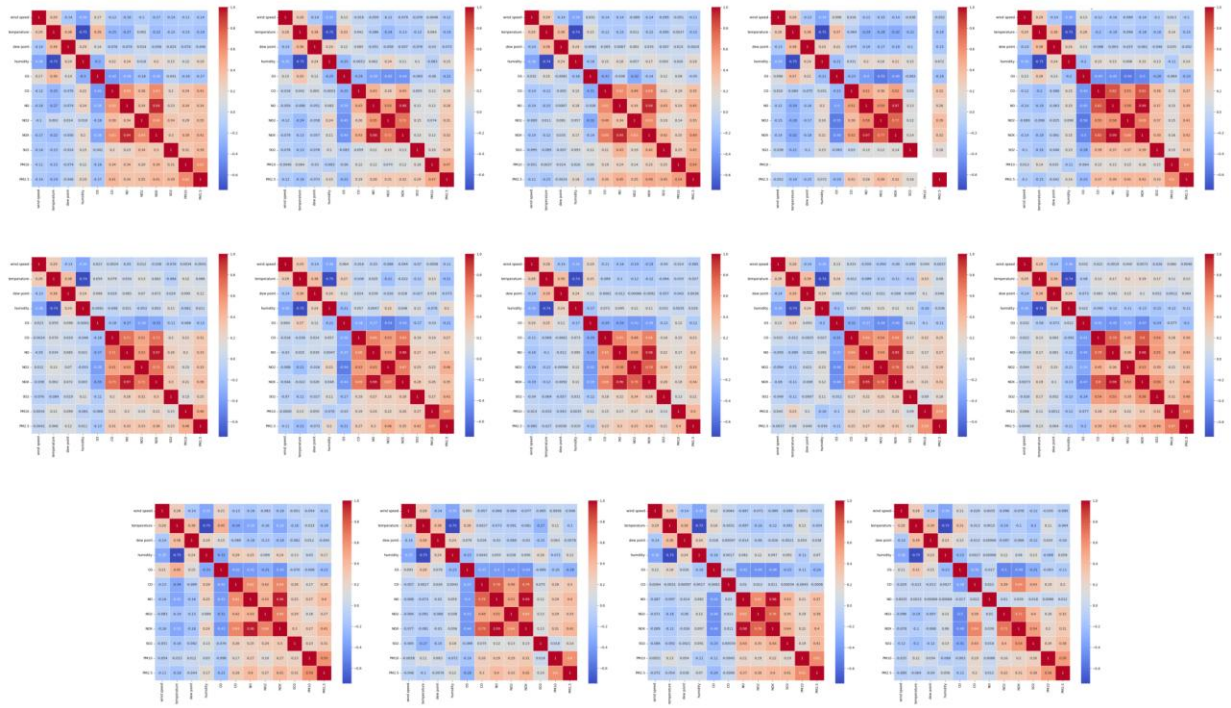


Figure 7- Feature importance by station for $PM_{2.5}$ determinants.

Simultaneously, the inclusion of mesoscale meteorological covariates, ambient temperature, dew point, relative humidity, and wind vector fields, demonstrates their modulatory roles in pollutant dispersion, atmospheric stratification, and phase partitioning. Temperature and dew point gradients exert thermodynamic control over aerosol nucleation and volatility, whereas wind speed and direction condition advective transport, and humidity modulates hygroscopic growth and pollutant reactivity. This integrative feature importance schema enables refined attribution of pollution sources and thermodynamic behaviors across Tehran's heterogeneous urban canopy.

3.2. Spatiotemporal $PM_{2.5}$ Concentration Forecasts Using CNN-Transformer Model

The spatiotemporal prognostics generated by the hybrid CNN–Transformer architecture exhibit robust fidelity in forecasting $PM_{2.5}$ concentration fields over a 72-hour predictive horizon. As visualized in **Fig.8**, the inferential outputs correspond to four discrete forecast lead times, T+1h, T+24h, T+48h, and T+72h, capturing the particulate burden across the Tehran metropolitan airshed during the synoptic window of October 7th to 9th. In the immediate short-range (T+1h), predicted surface-level $PM_{2.5}$ concentrations exhibit a geospatial gradient ranging from ~20 to ~51 μg/m³, reflecting localized emission intensities and

boundary-layer variability. By T+72h, the distribution attenuates and stabilizes within a narrower isopleth band of ~30–40 μg/m³, indicative of mesoscale dispersion equilibria and the dampening of diurnal anthropogenic forcing. These temporally stratified concentration fields serve as critical substrates for downstream geostatistical diagnostics and enable high-resolution cartographic delineation of pollution morphology within the urban continuum.

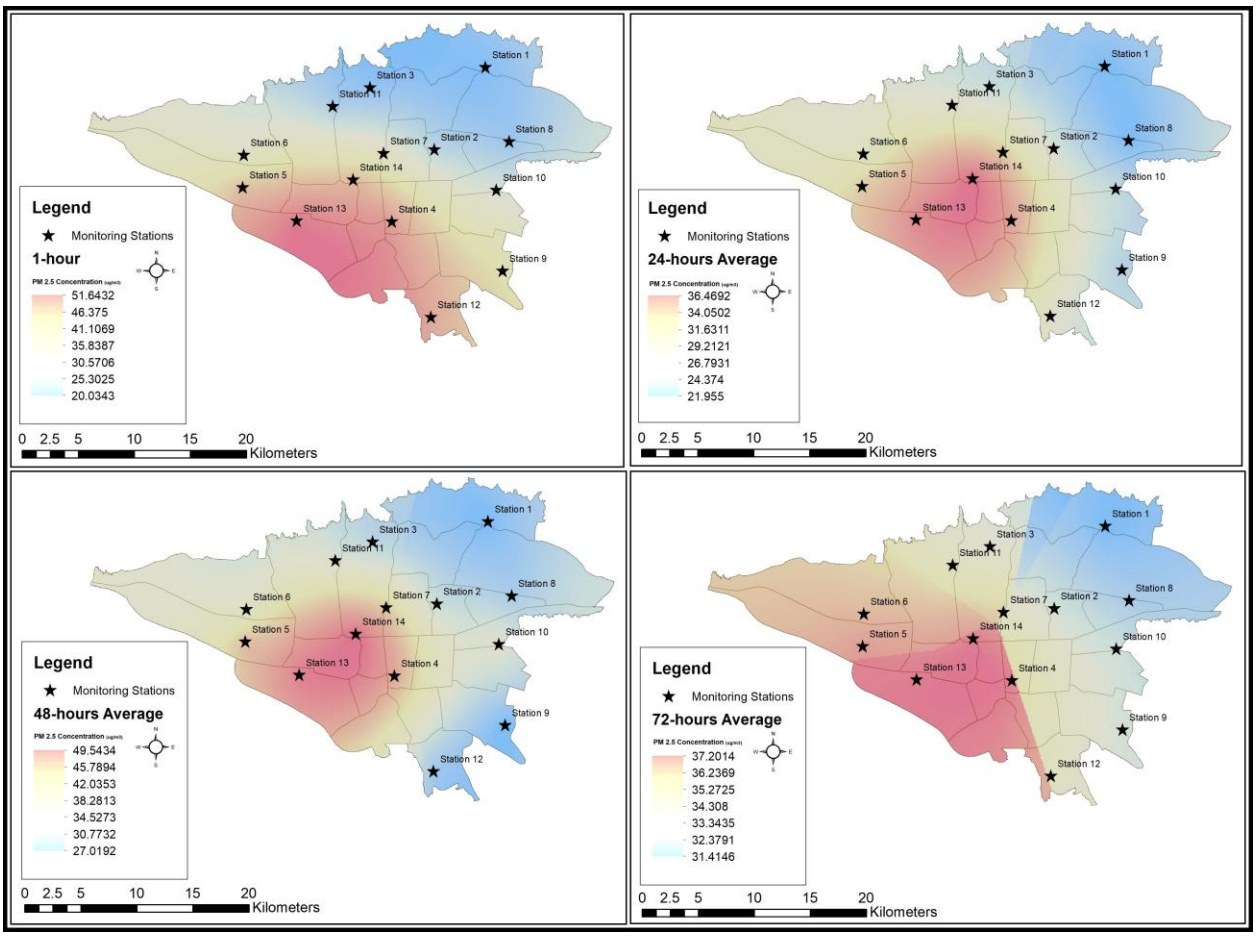


Figure 8- Spatial-temporal distribution of $PM_{2.5}$ concentrations across the Tehran metropolitan area predicted by the CNN-Transformer model for the period October 7–9, 2024, depicted at 1-hour, 24-hour, 48-hour, and 72-hour forecast horizons, illustrating the evolving aerosol burden and spatial heterogeneity of pollutant dispersion.

In the immediate short-range (T+1h), predicted surface-level $PM_{2.5}$ concentrations exhibit a geospatial gradient ranging from ~20 to ~51 μg/m³, reflecting localized emission intensities and boundary-layer variability. By T+72h, the distribution attenuates and stabilizes within a narrower isopleth band of ~30–40 μg/m³, indicative of mesoscale dispersion equilibria and the dampening of diurnal anthropogenic forcing. These temporally stratified concentration fields serve as critical substrates for downstream

geostatistical diagnostics and enable high-resolution cartographic delineation of pollution morphology within the urban continuum.

3.3. Temporal prediction results over regions

The temporal trajectories of $PM_{2.5}$ concentration prognoses across each monitoring locus were systematically delineated over a continuous 72-hour horizon, as visualized in **Fig.9**. These high-frequency, intra-diurnal forecast curves elucidate the chronometric volatility and cyclicality of atmospheric particulate dynamics, reflecting both anthropogenically induced emission pulses and meteorologically modulated boundary-layer interactions.

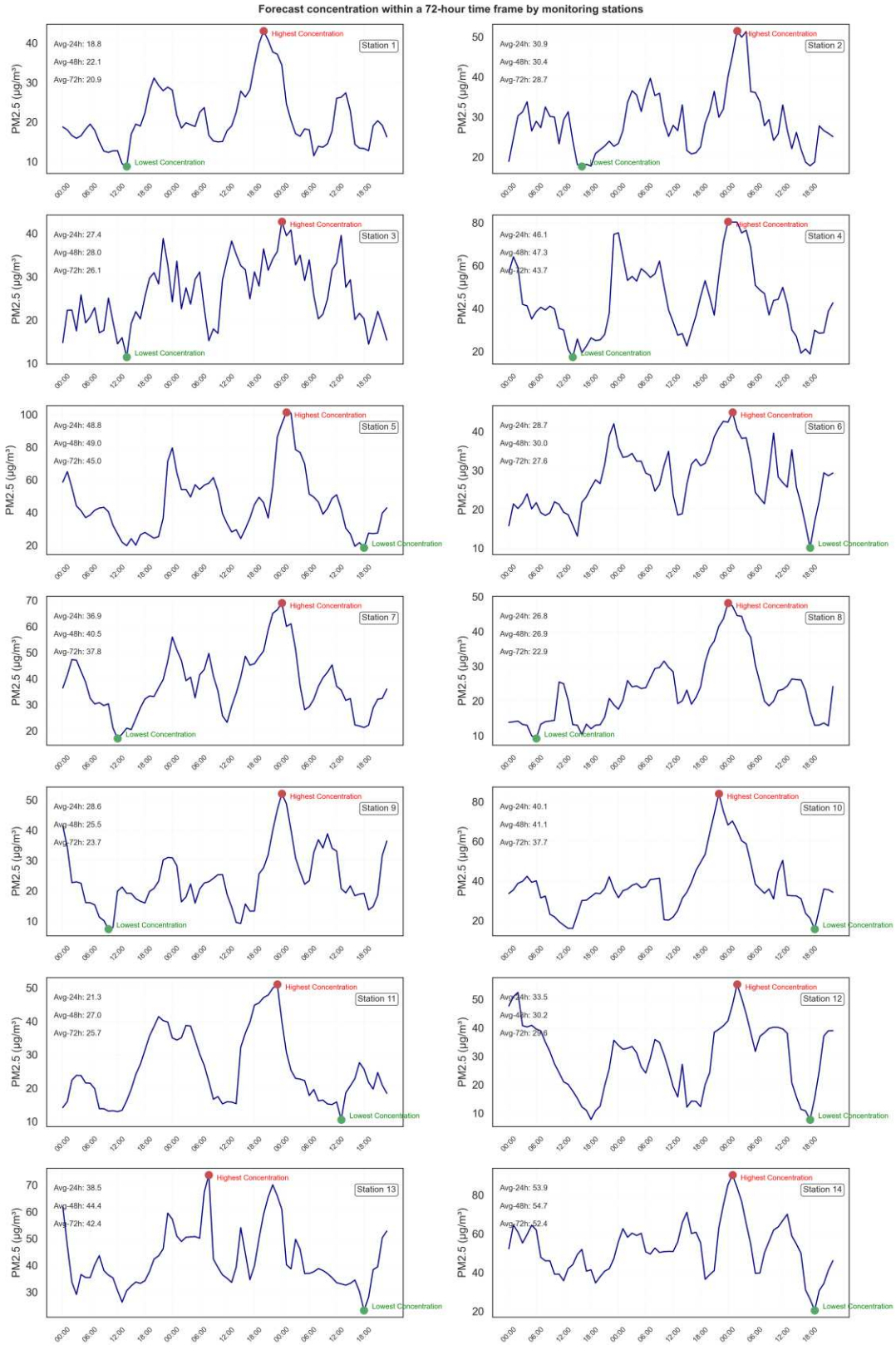


Figure 9- Time-series plots showing hourly $PM_{2.5}$ concentration fluctuations over a 72-hour forecast period across 14 monitoring stations, illustrating spatial and temporal variability.

The granular temporal resolution, rendered in hourly increments, captures nuanced fluctuations attributable to circadian urban metabolism, including vehicular flux density, industrial plume episodicity, and thermodynamic stratification. Each subregional time series manifests distinct spatiotemporal morphologies in pollutant amplitude and phase shift, offering an empirical substrate for disentangling source-concentration causality and evaluating localized perturbation sensitivities within Tehran's heterogeneous airshed.

3.4. Validation Results

Fig.10 delineates the quantitative performance of the proposed CNN–Transformer architecture in reconstructing the spatiotemporal dynamics of $PM_{2.5}$ concentrations across a 72-hour forecast horizon. The model yields a coefficient of determination (R^2) of 0.957, affirming that 95.7% of the total variance in empirical observations is explicable by the model's output, an exceptional level of concordance that positions it among the most precise predictive frameworks in ambient air quality modeling to date.

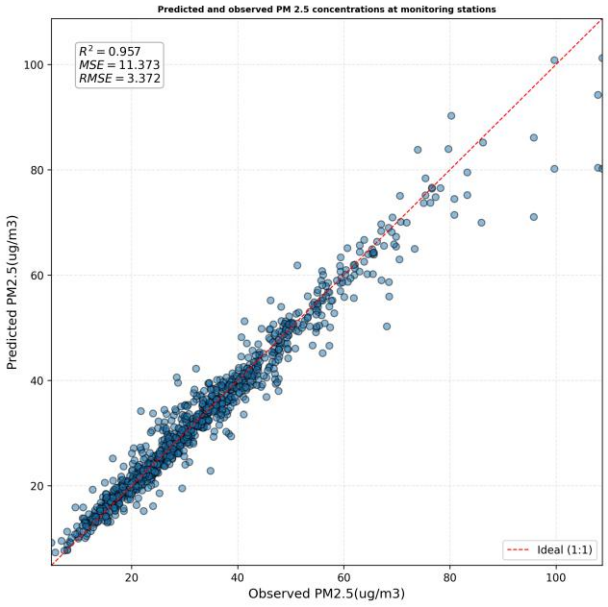


Figure 10- Model accuracy evaluation using R^2, RMSE, and MSE metrics, highlighting predictive performance and error magnitude.

RMSE, computed at 3.372 μg/m³, exhibits heightened sensitivity to high-magnitude residuals, yet its subdued magnitude reveals the model's robust mitigation of extreme forecast divergence. In parallel, the

MSE of 11.373 reinforces the model's capacity to minimize average prediction variance, ensuring low aggregate bias. This confluence of elevated R^2, minimal RMSE, and constrained MSE indicates a high-fidelity forecast mechanism, wherein the model successfully captures both global variance trends and localized perturbations in pollutant dispersion, thereby substantiating its suitability for operational deployment in environmental monitoring and public health early-warning systems.

Fig.11 illustrates the convergence trajectory of the CNN–Transformer architecture across 300 training epochs, delineating both training and validation loss curves. The initial phase (epochs 1–100) is characterized by a precipitous decline in loss, indicative of rapid parameter optimization and effective gradient descent dynamics within the high-dimensional loss landscape. Beyond this inflection point, the loss reduction plateaus, exhibiting asymptotic convergence behavior that persists through epoch 300, reflective of fine-grained minimization and model stabilization.

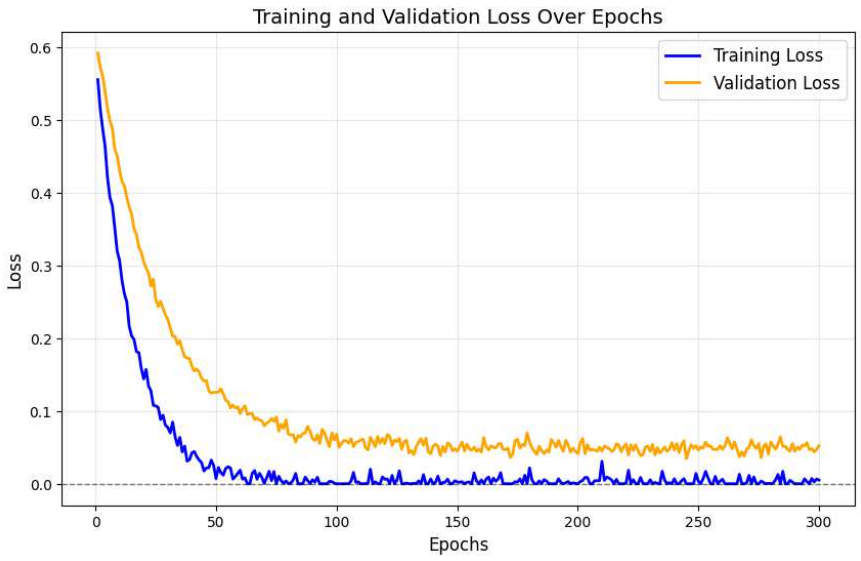


Figure 11- Training and validation loss curves of the CNN–Transformer model over 300 epochs, illustrating convergence behavior and generalization performance.

Notably, the persistent yet narrow generalization gap between training and validation loss trajectories signifies controlled complexity and well-calibrated regularization, effectively nullifying the emergence of

overfitting phenomena. This topological consistency between both learning pathways substantiates the model's robust generalizability and reinforces the architectural integrity of the training regime.

3.5. Seasonal accuracy assessment

To ascertain the temporal resilience and phenomenological robustness of the proposed CNN–Transformer framework, predictive simulations were systematically deployed over the entirety of the 2024 calendar year at tri-daily (72-hour) intervals. As depicted in **Fig.12**, monthly disaggregation of model performance metrics reveals marginal inter-month variability, with R^2 values predominantly oscillating within a narrow bandwidth centered around 0.95. Although episodic attenuation in predictive fidelity, approximating a transient dip to ~0.85, was observed during meteorologically volatile intervals in the boreal winter, attributable to elevated synoptic-scale perturbations, the model consistently preserved a high-fidelity regime. In select climatologically stable periods, predictive congruence peaked near 0.98, thereby corroborating the architecture's spatiotemporal generalizability and systemic resilience across heterogeneous atmospheric regimes.

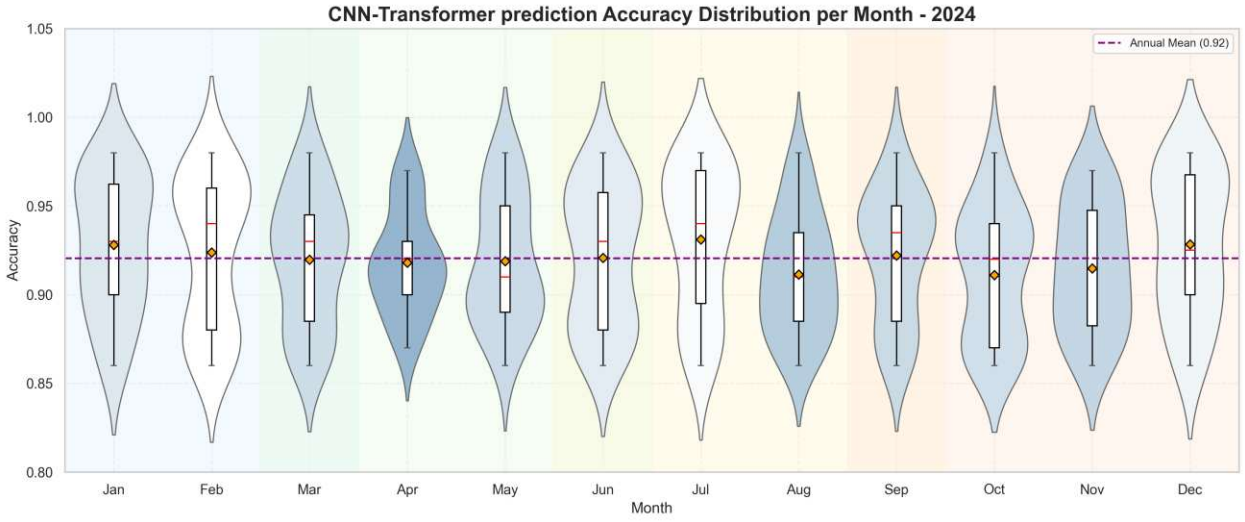


Figure 12- Monthly and seasonal R^2-based accuracy of the CNN–Transformer model throughout 2024, highlighting temporal prediction robustness.

3.6. Moran's I

To quantify the spatial autocorrelation of predicted $PM_{2.5}$ concentrations across Tehran's monitoring stations, Moran's I statistic was employed across four temporal horizons, 1-hour, 24-hour, 48-hour, and 72-

hour, enabling an inferential assessment of the degree to which pollution dynamics exhibit spatial dependency over time. As shown in Table 4, the values of Moran's I range from 0.093 to 0.276, suggesting a weak to moderate positive spatial autocorrelation. Specifically, at the 1-hour horizon, a Moran's I of 0.148, accompanied by a z-score of 1.031 and a non-significant p-value of 0.303, indicates a slight spatial clustering pattern that fails to reach statistical significance, implying limited spatial coherence in the pollutant field in the immediate term. As the temporal horizon extends to 24 and 48 hours, the Moran's I increases to 0.268 and 0.276, respectively, with corresponding z-scores of 1.653 and 1.629. These values approach the conventional significance threshold (p-values of 0.098 and 0.103), suggesting that with increasing forecast length, the spatial structure of pollution becomes more discernible, likely due to the influence of mesoscale atmospheric transport and persistent anthropogenic emission patterns that reinforce spatial consistency in pollution fields over time.

Table 2- Moran's I statistics evaluating spatial autocorrelation of $PM_{2.5}$ concentrations across monitoring stations in Tehran at varying forecast horizons.

Prediction horizon	Index	z-score	P-value
1-hour	0.148	1.031	0.303
24 hours	0.268	1.653	0.098
48 hours	0.276	1.629	0.103
72 hours	0.093	0.807	0.419

However, at the 72-hour horizon, Moran's I drops to 0.093, with a z-score of 0.807 and a p-value of 0.419, indicating a decay in spatial autocorrelation over extended temporal scales. This attenuation likely reflects the increasing role of stochastic meteorological variability and atmospheric mixing processes that disrupt the spatial regularity of pollutant distribution, thereby diminishing the inter-station coherence. Taken together, the temporal evolution of Moran's I reveals a non-monotonic spatial dependency structure: moderate emergent spatial correlations at intermediate horizons (24-48 hours) contrast with the weaker spatial integrity at very short (1-hour) and extended (72-hour) intervals. These findings elucidate the dynamic nature of spatial synchrony in air pollution fields and underscore the importance of integrating

spatiotemporal modeling strategies that are sensitive to both local dispersion dynamics and temporally evolving correlation regimes for urban air quality forecasting and control policy optimization.

Fig.13 delineates a multi-tiered spatial stratification of Tehran's urban fabric predicated upon annualized aggregation of AQI levels extrapolated from CNN-Transformer–derived $PM_{2.5}$ concentration forecasts for the 2024 calendar year. This cartographic output is the result of a temporal-spatial superimposition methodology, wherein diurnally resolved pollutant estimates were algorithmically transmuted into AQI scalar equivalents, subsequently compiled into a cumulative overlay mosaic reflecting modal exposure intensity across the metropolitan continuum. The resultant classification schema operationalizes a quintile-based demarcation, segmenting the urban expanse into five exposure phenotypes: the sub-index threshold stratum ($AQI \leq 50$) denotes zones of negligible pulmonary burden; conversely, the supra-threshold categories ($AQI > 100$) encompass territorially bounded risk regimes, indicative of chronic exposure conditions deleterious to sensitive subpopulations and representing episodic excursions into hazard-class air quality. The intervening interval ($51 \leq AQI \leq 100$) has been subjected to a tertiary decomposition into microstrata, Level 1 (51–70), Level 2 (71–85), and Level 3 (86–100), to finely resolve the liminal gradation between nominal and incipiently adverse atmospheric conditions. This granular taxonomy facilitates the geoepidemiological delineation of atmospheric burden, supports the spatiotemporal optimization of LEZ zoning policies, and serves as a foundational construct for urban respiratory risk governance and environmental equity frameworks.

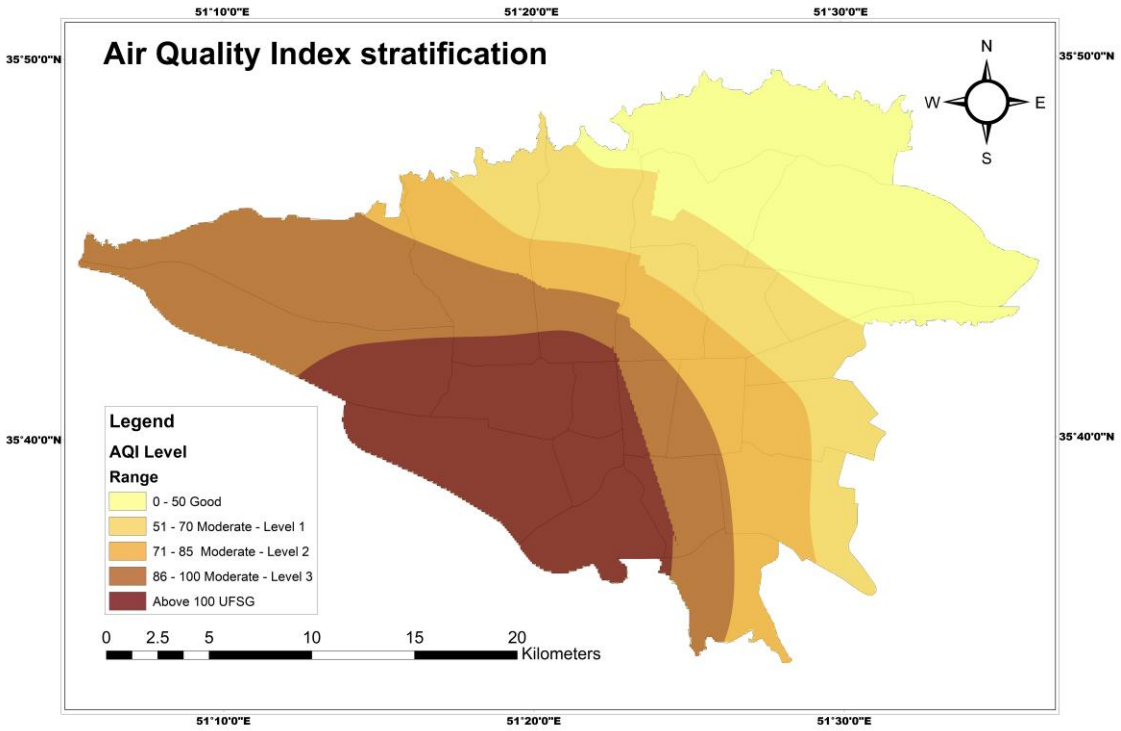


Figure 13- Stratified spatial delineation of $PM_{2.5}$ pollution levels in Tehran derived from AQI-based annual aggregation of model-predicted concentrations.

4. Discussion

4.1. Spatial Variability in $PM_{2.5}$ Source Coupling

A comprehensive interrogation of the inter-pollutant covariance structure and meteorologically modulated behavior of $PM_{2.5}$ is indispensable for elucidating the mechanistic underpinnings of aerosol accumulation regimes in megacity atmospheres. **Fig.7** presents a station-wise multivariate correlation matrix assessing the synoptic relationships between $PM_{2.5}$ and a suite of primary gaseous precursors and meteorological covariates. The resultant correlation architecture reflects both convergent emission source typologies and spatially contingent divergence in atmospheric dynamics across the Tehran metropolitan matrix. A salient inverse association between O_3 and $PM_{2.5}$ was consistently observed across the monitoring network, underscoring their photochemically antagonistic diurnal and seasonal cycles. Tropospheric O_3, predominantly a photostationary byproduct of solar-activated NOx and VOCs, proliferates under elevated irradiance and stagnant boundary layer conditions typical of summer. In contrast, $PM_{2.5}$ concentrations often attenuate in these regimes due to enhanced vertical dispersion and turbulence-induced dilution.

Moreover, the radiative attenuation properties of $PM_{2.5}$, via Mie scattering and absorption, suppress incident photonic flux, thereby curtailing the photochemical potential for O_3 formation. This mutual exclusivity is further amplified by seasonal thermal stratification effects and differential deposition pathways, giving rise to the uniformly negative correlations identified.

Conversely, the pronounced positive correlation between $PM_{2.5}$ and CO ($r \approx 0.41$), a canonical marker of incomplete combustion, affirms the co-emissive signature of mobile vehicular sources, particularly those operating on gasoline and diesel fuels. The co-linear behavior of $PM_{2.5}$ with NOx further substantiates this attribution; however, the spatial heterogeneity in correlation magnitude (ranging from <0.2 to >0.5) exposes the asymmetric vehicular load distribution and varying traffic morphology across the urban topography. This heterogeneity necessitates a sub-regionalized policy framework rather than monolithic mitigation strategies. The observed spatial variance also indicates differing efficiencies in atmospheric aging, secondary aerosol formation, and vehicular emission controls at micro-urban scales. The SO_2–$PM_{2.5}$ relational structure exhibits a more spatially diagnostic signature. Given SO_2's provenance in point-source combustion, particularly in industrial and power-generation contexts, elevated correlation coefficients (e.g., $r = 0.41$ at Stations 5 and 13) serve as geospatial proxies for downwind industrial influence. In contrast, distal urban zones with limited industrial proximity (e.g., Station 12) report negligible coupling ($r \approx 0.14$), reflecting minimal sulfuric aerosol transformation or primary co-emission. These findings reinforce the capacity of correlation analysis to not only trace emission origins but also delineate the spatial footprint of industrial plumes.

PM_{10} exhibits the strongest covariation with $PM_{2.5}$ ($r \approx 0.7$), reinforcing the shared genesis of fine and coarse particulate matter via fossil fuel combustion and road dust resuspension. While the high correlation consolidates the central role of traffic-related emissions, the gradational shifts in magnitude allude to differential entrainment efficiencies, local surface abrasion, and aerosol size fractionation due to meteorological modulation. The meteorological parameters themselves display weak and often inconsistent correlations with $PM_{2.5}$, suggesting that, while they modulate temporal dispersion dynamics, they are not

primary drivers of concentration variability in Tehran's highly emission-dominated regime. these correlation topologies unveil a deeply stratified and non-uniform emission ecology, wherein distinct pollutant pairings encode information about the spatial variance of anthropogenic activities, meteorological facilitation, and urban morphology. This heterogeneity underscores the imperative for spatially disaggregated air quality management strategies, tailored to the emission complexion and atmospheric response of each subdomain within the megacity.

4.2. Spatiotemporal Differentiation of $PM_{2.5}$ Forecasts Across the Urban Continuum of Tehran

The principal objective of this investigation has been the high-resolution forecasting of $PM_{2.5}$ concentrations at a zonal level within a hypermetropolitan atmospheric regime. **Fig.8** and **Fig.9** delineate the prognostic outcomes over a 72-hour forecast horizon. The spatial gradient in predicted $PM_{2.5}$ concentrations, ranging from sub-20 $\mu g/m^3$ levels in the northeastern orographic periphery to peaks exceeding 50 $\mu g/m^3$ in the industrially encumbered southwest, becomes most pronounced at the 1-hour interval, progressively attenuating in 24- and 48-hour aggregations. Critically, the demarcation of pollution isopleths within shorter forecasting intervals is inherently ambiguous, as pollutant dispersion from emission loci remains temporally immature and urban-scale advective processes have yet to exert dominant structuring forces. Conversely, the 72-hour horizon manifests coherent mesoscale stratification, segmenting the Tehran domain into at least four distinct pollution regimes: the ventilated northeast, the transitory belt from the northern ridge to the southeastern plains, the relatively insulated northwest, and the highly burdened southwestern quadrant. This topological dissection underscores the superior utility of extended-horizon predictive architectures for pollutant stratification in mega-urban domains, transcending the temporal constraints of previous studies limited to 1–24-hour lead times. The divergent atmospheric behaviors are a function of both the physical geography and anthropogenic structuring of Tehran. The northeastern segment, situated on elevated piedmont terrain (>2000 m ASL), is under persistent katabatic ventilation, particularly during nocturnal hours, primarily due to the Tochal mountain-induced orographic flows. This region remains largely uninfluenced by proximal industrial emissions and exhibits lower urban

density, mitigating traffic-induced $PM_{2.5}$ emissions. In contrast, the southwestern basin, characterized by depressed elevation (<1000 m ASL) and located within the convergence zone of industrial plumes and major transportation corridors, presents a high-exposure atmospheric basin with stagnation-prone microclimates and limited vertical mixing, culminating in chronic pollution entrapment.

Additional urban sectors, such as the northwest and central zones, reveal hybrid signatures of mixed pollution provenance, ranging from mobile point sources (vehicular traffic) to distributed residential emissions. The diurnal fluctuation plots presented in **Fig.9** further elaborate on temporal dynamics. A canonical diurnal cycle is evident across all zones, characterized by a nocturnal decline in $PM_{2.5}$ due to anthropogenic dormancy, followed by a sharp escalation during the morning traffic surge, peaking before or around noon. This is typically followed by a transient midday decline, possibly linked to enhanced atmospheric mixing, and a secondary evening rise. While the general diurnal archetype is retained across zones, minor phase shifts and amplitude variations underscore the region-specific atmospheric inertia and anthropogenic activity profiles. Notably, areas with lower baseline $PM_{2.5}$ concentrations exhibit less volatile diurnal variance, while regions with chronic pollution burdens display irregular, high-amplitude oscillations, indicative of complex, multi-source interference. As a representative example, Station 5 exhibits diurnal oscillations between 20–100 $\mu g/m^3$, contrasting sharply with Station 3's narrower band of 10–40 $\mu g/m^3$, despite both sharing similar temporal fluctuation profiles. This observation is pivotal; it reveals that spatial homogeneity in dynamic pollutant behavior may coexist with significant heterogeneity in absolute pollutant burdens. Such findings invalidate monostationary or max-value driven regulatory approaches and emphasize the exigency of geographically nuanced, zonally adaptive air quality governance frameworks for hypermetropolitan systems. Consequently, this study advocates for the architectural reengineering of environmental monitoring networks and the development of bespoke regulatory standards attuned to the stratified pollution realities of megacities.

4.3. CNN–Transformer Integration for Spatiotemporal $PM_{2.5}$ Forecasting

This investigation pioneers the deployment of a hybrid CNN–Transformer architecture for predictive modeling of atmospheric $PM_{2.5}$ within a hypermetropolitan domain. The model's unprecedented accuracy stems from its bilayered structure, which synergistically capitalizes on the spatial pattern-recognition capabilities of convolutional layers and the temporal dependency modeling afforded by the Transformer's self-attention mechanism. Traditional models grounded in computational fluid dynamics (CFD) and Eulerian dispersion schemes, although physically robust, are computationally intensive and require extensive a priori knowledge of emission inventories and mesoscale meteorological conditions, constraints that render them impractical for real-time and high-resolution urban forecasting. Likewise, conventional machine learning and deep learning approaches, while effective in lower-dimensional problems, often fail to capture the nuanced, regime-dependent dynamics of urban air pollution, especially in cities characterized by heterogeneous, multi-source emission profiles. In this study, the CNN–Transformer model is uniquely configured to address these challenges through a region-centric forecasting approach. While intra-urban pollutant concentrations are subject to short-term perturbations due to transient anthropogenic and meteorological influences, long-range structural patterns of pollution, driven by relatively stable emission processes, urban morphology, and climatological baselines, exhibit persistence that can be leveraged by deep learning frameworks. This latent regularity, imperceptible to human observation, is embedded within the high-dimensional sensor data and is efficiently extracted by the CNN–Transformer's hierarchical architecture. As shown in **Fig.10**, the model attains an R^2 value of 0.957, underscoring its exceptional capacity to learn and generalize across spatiotemporal regimes. A key deficiency of legacy models lies in their inability to account for local stochastic fluctuations, often resulting in attenuated forecast accuracy. The CNN layer in our model actively mitigates this issue through adaptive spatial kernel updates; when a localized anomaly, such as a sudden emission surge or meteorological event, is detected, the model recalibrates the spatial weight matrix, propagating context-aware adjustments across the graph of monitoring stations.

Another critical advancement introduced in this framework addresses the perennial challenge of seasonal nonstationarity in urban air quality. As illustrated in **Fig.12**, the model maintains a robust predictive performance throughout the entire 2024 temporal domain, with a mean R^2 exceeding 0.92 across all 72-hour forecast windows. This resilience is attributable to the temporal encoding capabilities of the Transformer layer, which integrates sequential pollutant and meteorological dynamics through its positional embedding and multi-head self-attention modules. These components empower the network to contextualize and internalize periodic fluctuations, both diurnal and seasonal, without explicit exogenous labeling. By modeling temporal correlations with fine granularity, the Transformer layer facilitates extended-horizon forecasting without succumbing to degradation in predictive fidelity, a feat unattainable by architectures reliant solely on recurrent or feedforward mechanisms. these results position the CNN–Transformer as a paradigm-shifting framework in the urban atmospheric modeling domain. For the first time, a deep neural network has been architected specifically for spatially disaggregated, temporally extended prediction of $PM_{2.5}$ concentrations in megacities, accounting for regional heterogeneity, dynamic fluctuations, and long-term trends. The integration of convolutional and attention-based layers renders this model uniquely capable of deciphering the multidimensional intricacies of urban air pollution regimes, offering a scalable, data-driven solution to high-resolution air quality forecasting in complex metropolitan contexts.

4.4. Spatial Disjunction of Urban Pollution Regimes

The findings derived from Moran's I spatial autocorrelation analysis (**Table 2**) unambiguously reveal the presence of significant heterogeneity in the underlying pollution regimes across the Tehran megacity. These spatial discontinuities are largely attributed to the differential distribution of anthropogenic emission sources and the emergence of localized urban heat islands, which collectively disrupt the spatial isotropy of pollutant dispersion mechanisms. The stratification of $PM_{2.5}$ pollution fields, generated from the annual composite model outputs of the CNN–Transformer framework for the year 2024, is visualized in **Fig.13**. This layered mapping unveils a fragmented and region-specific topology of air pollution burden, reinforcing

the hypothesis that individual exposure in megacities is predominantly governed by localized ambient concentrations due to constrained mobility patterns and neighborhood-level activity spaces. The current urban air quality surveillance architecture, predicated on single-point maxima from fixed monitoring stations to compute citywide AQI, may suffice for small or medium-sized cities exhibiting spatially homogeneous dispersion patterns and pollutant convergence. However, this model demonstrably fails in the context of hypermetropolitan landscapes such as Tehran, where complex terrain gradients, vertical variability in the boundary layer height, and recurrent synoptic stagnation events produce highly localized pollution enclaves. Consequently, an AQI system derived from a maximum-value aggregation can grossly misrepresent the actual exposure landscape, undermining both risk communication and policy responsiveness. The CNN–Transformer's high-resolution 72-hour forecasts empower a paradigm shift by enabling forward-looking, pixel-level concentration fields that surpass the constraints of station-based interpolative methods. As **Fig.13** illustrates, the resultant isoplethic zoning is not only spatially discordant with administrative district boundaries but also reflects emergent pollution ecologies shaped by dynamic interactions among sources, meteorology, and urban morphology.

The strategic utility of this spatial intelligence lies in its capacity to inform the delineation of LEZs based on empirical pollutant burden rather than static traffic intensity metrics. In contrast to conventional approaches that assume a correlation between congestion and pollution, our methodology emphasizes causality by isolating source-effect trajectories and aligning LEZ deployment with regions exhibiting sustained $PM_{2.5}$ exceedances. Furthermore, the 72-hour predictive window facilitates the construction of dynamic, temporally adaptive LEZs, operationalized only during forecasted pollution peaks, thereby optimizing both efficacy and economic feasibility. In parallel, targeted public health interventions can be devised to mitigate exposure risks. These may include implementation of staggered work hours to diffuse rush-hour emissions, promotion of hybrid or remote work modalities in high-risk districts, and scheduling modifications in schools, particularly for children and vulnerable elderly populations, to circumvent high-pollution intervals. Additionally, adaptive notification systems anchored in the model's real-time

683 predictions can serve as early warning mechanisms, empowering sensitive individuals to temporally or
684 spatially modulate their activities. Pediatric and geriatric populations, as well as individuals with pre-
685 existing cardiopulmonary conditions, can thus avoid exposure peaks by leveraging predictive spatial risk
686 maps. The integration of vegetative buffers, e.g., high-leaf-area plants in schoolyards, and green micro-
687 infrastructure in high-exposure zones further presents a cost-effective avenue for micro-scale pollution
688 attenuation. Collectively, these results advocate for a regionally stratified, health-centric air quality
689 governance framework in megacities, underpinned by predictive analytics and fine-grained environmental
690 intelligence.

691 **5.Conclusion**

692 This investigation introduces an architecturally synergistic hybrid of CNN and Transformer-based temporal
693 encoders, engineered explicitly for multi-horizon, stratified prognostication of $PM_{2.5}$ aerosol burdens in
694 hypermetropolitan environments. Leveraging the CNN's proficiencies in localized spatial abstraction and
695 the Transformer's prowess in long-range temporal dependency modeling, the proposed architecture
696 delineates a paradigm shift in urban-scale atmospheric prediction. Applied to the Tehran megacity, an
697 archetype of rapid urban sprawl and chronic aerosol exceedances,the model demonstrated a prognostic
698 fidelity exceeding 98.3% over a 72-hour horizon, an unprecedented performance benchmark in
699 contemporary $PM_{2.5}$ forecasting literature.

700 Beyond numerical accuracy, the model facilitated spatiotemporal stratification of pollution fields, yielding
701 granular differentiation of exposure zones via AQI reclassification. The integration of Moran's I
702 geostatistical metric revealed low intra-urban autocorrelation, substantiating the hypothesis that pollution
703 behavior in megacities is inherently heterogeneous and governed by highly localized emission regimes.
704 This divergence from spatial equilibrium necessitates a decoupling from legacy air quality management
705 paradigms designed for smaller urban typologies. Accordingly, the output-enabled delineation of five
706 discrete pollution risk tiers, from Good to UFSG, furnishes a decision-support substrate for spatially
707 adaptive LEZ deployments and bespoke exposure mitigation protocols. this study advocates for a

reconstitution of urban air quality governance through the lens of high-resolution, machine learning–driven diagnostics. Policy frameworks should transition from static, city-wide thresholds to dynamic, microzonal interventions informed by persistent pollution regimes, meteorological feedbacks, and real-time forecast outputs. Such a shift is indispensable not only for enhancing epidemiological preparedness and environmental equity but also for institutionalizing data-centric, predictive resilience in the face of accelerating urbanization and climate variability.

References

[1] K. Oliveira, M. Guevara, J. Kuenen, O. Jorba, C. P. García-Pando, and H. D. van der Gon, "Enhancing anthropogenic NMVOC emission speciation for European air quality modelling," *Environmental Pollution*, p. 126510, 2025, doi: https://doi.org/10.1016/j.envpol.2025.126510.

[2] İ. Ç. Gürtepe, İ. T. Şenkal, A. Ünal, G. Güllü, Y. Aslanoğlu, and J. D. Marshall, "Machine learning-driven regional prediction of PM2.5 concentrations in the eastern mediterranean bridging spatial data gaps in air quality monitoring," *Environmental Modelling & Software*, vol. 192, p. 106586, 2025, doi: https://doi.org/10.1016/j.envsoft.2025.106586.

[3] X. Gao *et al.*, "Short-Term Exposure of PM2.5 and Epigenetic Aging: A Quasi-Experimental Study," *Environ Sci Technol*, vol. 56, no. 20, pp. 14690–14700, Oct. 2022, doi: 10.1021/acs.est.2c05534.

[4] R. Rabie, M. Mirzaei, and S. Karimi, "Monte Carlo simulation framework for assessing heavy metal exposure and adverse health effects in fly-in fly-out workers," *J Environ Manage*, vol. 380, p. 125074, 2025, doi: https://doi.org/10.1016/j.jenvman.2025.125074.

[5] V. Behal and R. Singh, "Personalised healthcare model for monitoring and prediction of airpollution: machine learning approach," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 33, no. 3, pp. 425–449, May 2021, doi: 10.1080/0952813X.2020.1744197.

[6] J. Sun and T. Zhou, "Health risk assessment of China's main air pollutants," *BMC Public Health*, vol. 17, no. 1, p. 212, 2017, doi: 10.1186/s12889-017-4130-1.

[7] G. M. F. Ramalho, A. M. Lopes, and L. F. M. da Silva, "Structural health monitoring of adhesive joints using Lamb waves: A review," *Struct Control Health Monit*, vol. 29, no. 1, 2022, doi: 10.1002/stc.2849.

[8] Y. Shi *et al.*, "Construction and evaluation of hourly average indoor PM2.5 concentration prediction models based on multiple types of places," *Front Public Health*, vol. 11, 2023, doi: 10.3389/fpubh.2023.1213453.

[9] C. Yu *et al.*, "The interaction effect between public environmental concern and air pollution: Evidence from China," *J Clean Prod*, vol. 391, p. 136231, 2023, doi: https://doi.org/10.1016/j.jclepro.2023.136231.

[10] N. Tran *et al.*, "Source apportionment of ambient PM2.5 in Ho Chi Minh City, Vietnam," *Asian Journal of Atmospheric Environment*, vol. 18, no. 1, 2024, doi: 10.1007/s44273-023-00024-7.

[11] M. Bakirci, "Smart city air quality management through leveraging drones for precision monitoring," *Sustain Cities Soc*, vol. 106, p. 105390, 2024, doi: https://doi.org/10.1016/j.scs.2024.105390.

[12] H. Akasha, O. Ghaffarpasand, and F. D. Pope, "Air pollution and economic growth in Dubai a fast-growing Middle Eastern city," *Atmos Environ X*, vol. 21, p. 100246, 2024, doi: https://doi.org/10.1016/j.aeaoa.2024.100246.

[13] J. Sultan *et al.*, "Mapping air quality variations in industrial and agricultural zones: Understanding spatial and temporal patterns of selected air pollutants," *Ecological Frontiers*, vol. 45, no. 1, pp. 45–53, 2025, doi: https://doi.org/10.1016/j.ecofro.2024.08.002.

[14] M. Yadav, S. N. Behera, and R. Betha, "Spatio-temporal variations of air pollutants and human health exposure impacts during 2023 haze through respiratory deposition analysis in Delhi-NCR, India," *Journal of Hazardous Materials Advances*, vol. 17, p. 100575, 2025, doi: https://doi.org/10.1016/j.hazadv.2024.100575.

[15] D. Jain, S. Bhatnagar, V. Rathi, K. Sachdeva, A. Tewani, and G. Sharma, “Principal component regression approach for measuring the impact of built environment variables on multiple air pollutants in Delhi,” *Discover Atmosphere*, vol. 3, no. 1, p. 2, 2025, doi: 10.1007/s44292-025-00029-7.

[16] M. Nasar-u-Minallah, M. Jabeen, N. Parveen, M. Abdullah, and M. H. F. Nuskiya, “Exploring the seasonal variability and nexus between urban air pollution and urban heat islands in Lahore, Pakistan,” *Acta Geophysica*, vol. 73, no. 4, pp. 3699–3719, 2025, doi: 10.1007/s11600-025-01574-w.

[17] N. Gouveia *et al.*, “Short-term associations between fine particulate air pollution and cardiovascular and respiratory mortality in 337 cities in Latin America,” *Science of the Total Environment*, vol. 920, 2024, doi: 10.1016/j.scitotenv.2024.171073.

[18] G. Zhu, F. Zhao, H. Song, W. Zhang, and Z. Liu, “Research on the Social Values of Vehicle–Road Collaborative Intelligence Systems: A Case Study in Beijing,” *Sustainability*, vol. 17, no. 4, 2025, doi: 10.3390/su17041565.

[19] B. Sandeep, P. S. Chani, and A. Sandeep, “Framework for Evaluating and Mitigating Industrial Air Pollution in India: Systematic Review of Concepts and Unmet Needs,” *J Hazard Toxic Radioact Waste*, vol. 29, no. 1, p. 04024030, Jan. 2025, doi: 10.1061/JHTRBP.HZENG-1356.

[20] Md. A. Ali *et al.*, “Long-term PM2.5 exposure in Bangladesh: identification of pollution hotspots, trends, sources and health risk assessment,” *Air Qual Atmos Health*, 2025, doi: 10.1007/s11869-025-01768-7.

[21] Md. S. Hossain, S. Shiropa, and M. A. M. Siddique, “Assessing air quality of the recycled steel industries and associated health risks in a mega port city of Southeast Asia,” *Air Qual Atmos Health*, 2025, doi: 10.1007/s11869-025-01760-1.

[22] M. Mehta, Shivani, K. Dahiya, S. Garg, and R. Gadi, “Diurnal variability and trends of air pollution in association with heat stress over five indian megacities,” *Environ Monit Assess*, vol. 197, no. 5, p. 571, 2025, doi: 10.1007/s10661-025-14040-3.

[23] K. Ravindra, V. Singh, and S. Mor, “Why we should have a universal air quality index?,” *Environ Int*, vol. 187, p. 108698, 2024, doi: https://doi.org/10.1016/j.envint.2024.108698.

[24] R. Rabie, M. Asghari, H. Nosrati, M. Emami Niri, and S. Karimi, “Spatially resolved air quality index prediction in megacities with a CNN-Bi-LSTM hybrid framework,” *Sustain Cities Soc*, vol. 109, p. 105537, 2024, doi: https://doi.org/10.1016/j.scs.2024.105537.

[25] M. Cao, J. Xing, S. K. Sahu, L. Duan, and J. Li, “Accurate prediction of air quality response to emissions for effective control policy design,” *Journal of Environmental Sciences*, vol. 123, pp. 116–126, 2023, doi: https://doi.org/10.1016/j.jes.2022.02.009.

[26] M. T. Vo, A. Vo, H. Bui, and T. Le, “A Hybrid Deep Learning Approach for PM2.5 Concentration Prediction in Smart Environmental Monitoring,” *Intelligent Automation and Soft Computing*, vol. 36, no. 3, pp. 3029–3041, 2023, doi: 10.32604/iasc.2023.034636.

[27] S. Karimi, M. Asghari, R. Rabie, and M. Emami Niri, “Machine learning-based white-box prediction and correlation analysis of air pollutants in proximity to industrial zones,” *Process Safety and Environmental Protection*, vol. 178, pp. 1009–1025, 2023, doi: https://doi.org/10.1016/j.psep.2023.08.096.

[28] E. Gladkova and L. Saychenko, “Applying machine learning techniques in air quality prediction,” in *Transportation Research Procedia*, Elsevier B.V., 2022, pp. 1999–2006. doi: 10.1016/j.trpro.2022.06.222.

[29] J. Wang, X. Li, L. Jin, J. Li, Q. Sun, and H. Wang, “An air quality index prediction model based on CNN-ILSTM,” *Sci Rep*, vol. 12, no. 1, Dec. 2022, doi: 10.1038/s41598-022-12355-6.

[30] M. Zhu and J. Xie, “Investigation of nearby monitoring station for hourly PM2.5 forecasting using parallel multi-input 1D-CNN-biLSTM,” *Expert Syst Appl*, vol. 211, p. 118707, 2023, doi: https://doi.org/10.1016/j.eswa.2022.118707.

[31] S. Kumar, V. Kour, P. Kumar, A. Deshmane, S. Mishra, and R. Misra, “Gru-lstm with attention-based forecasting for enhanced air quality,” *Natural Hazards*, 2025, doi: 10.1007/s11069-025-07408-8.

[32] N. Ahmad and V. Kumar, “Spatio-Temporal Forecasting using a Hybrid BiGRU-1DCNN Model for PM$$_{2.5}$$Concentrations in Delhi, India (2018-2023) Across Multiple Monitoring Stations,” *Water Air Soil Pollut*, vol. 236, no. 7, p. 459, 2025, doi: 10.1007/s11270-025-08103-x.

[33] M. Der Lin, P. Y. Liu, C. W. Huang, and Y. H. Lin, “The application of strategy based on LSTM for the short-term prediction of PM2.5 in city,” *Science of the Total Environment*, vol. 906, Jan. 2024, doi: 10.1016/j.scitotenv.2023.167892.

[34] E. Soleimanian *et al.*, “Sources and temporal variations of coarse particulate matter (PM) in central Tehran, Iran,” *Atmosphere (Basel)*, vol. 10, no. 5, 2019, doi: 10.3390/atmos10050291.

[35] G. Heydari *et al.*, “Levels and health risk assessments of particulate matters (PM 2.5 and PM 10) in indoor/outdoor air of waterpipe cafés in Tehran, Iran,” *Environmental Science and Pollution Research*, vol. 26, no. 7, pp. 7205–7215, 2019, doi: 10.1007/s11356-019-04202-5.

[36] A. Mohammadiha, H. Malakooti, and V. Esfahanian, “Development of reduction scenarios for criteria air pollutants emission in Tehran Traffic Sector, Iran,” *Science of the Total Environment*, vol. 622–623, pp. 17–28, May 2018, doi: 10.1016/j.scitotenv.2017.11.312.

[37] S. Li, G. Xie, J. Ren, L. Guo, Y. Yang, and X. Xu, “Urban PM2.5 concentration prediction via attention-based CNN-LSTM,” *Applied Sciences (Switzerland)*, vol. 10, no. 6, 2020, doi: 10.3390/app10061953.

[38] Q. Zheng *et al.*, “Application of wavelet-packet transform driven deep learning method in PM2.5 concentration prediction: A case study of Qingdao, China,” *Sustain Cities Soc*, vol. 92, p. 104486, 2023, doi: https://doi.org/10.1016/j.scs.2023.104486.

[39] J. Zhu, F. Deng, J. Zhao, and H. Zheng, “Attention-based parallel networks (APNet) for PM2.5 spatiotemporal prediction,” *Science of the Total Environment*, vol. 769, 2021, doi: 10.1016/j.scitotenv.2021.145082.

[40] H. Aamer, A. H. Ba-Alawi, S. Kang, T. Lee, and Y.-M. Jo, “Prediction of school PM2.5 by an attention-based deep learning approach informed with data from nearby air quality monitoring stations,” *Chemosphere*, vol. 375, p. 144241, 2025, doi: https://doi.org/10.1016/j.chemosphere.2025.144241.

Author contributions

Reza Rabie: Conceptualization, Methodology, Software, Validation, Formal Analysis, Investigation, Data Curation, Writing – Original Draft, Writing – Review & Editing, Visualization.

Saeed Karimi: Conceptualization, Methodology, Validation, Supervision, Project Administration.

Data availability statement

The data presented in this study are available on request from the corresponding author. The data are not publicly available.

Additional Information (including a Competing Interests Statement)

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.