Time Series and Transformer based Analysis and Prediction of Air Quality Index in Lahore

Hamza Haider[1], Rabia Tehseen[1], Muhammad Chaudhry[1], Uzma Omer[2]
[1]Department of Computer Science, University of Central Punjab, Lahore, Pakistan
[2]Department of Information Sciences, University of Education, Lahore, Pakistan
*Corresponding | rabia.tehseen@ucp.edu.pk

Citation | Haider. H, Tehseen. R, Chaudhry. M, Omer. U, "Time Series and Transformer based Analysis and Prediction of Air Quality Index in Lahore", Magna Carta: Contemporary Social Science, Vol. 05 Issue 4 pp 291-307, October 2026
Received | Sep 02, 2026 **Revised |** Sep 25, 2026 **Accepted |** Sep 28, 2026**Published |** Oct 02, 2026.

Air pollution is a growing environmental and health problem. Urbanization, industry and seasonal weather conditions create periodic spikes of Particulate Matter 2.5 (PM 2.5) that can be extremely harmful. Traditional statistical analysis is not appropriate for measuring this type of data, which is characterized by being highly volatile and having nonlinear dependencies that lead to poor predictive performance. To address the above issues, we developed a comparative study of performance between the SARIMA model, which represents the traditional statistical baseline, and a new transformer-based deep-learning model for forecasting PM 2.5 in Lahore, by following the same procedure for processing all datasets (80/20 train test split, monthly averaging and same data preprocessing). In order to find the best modelling solution for forecasting the historical PM2. 5 concentrations in Lahore, all datasets were pre-processed utilizing imputation for missing values, treatment for outliers, and min-max scaling. The empirical results show that the Transformer model performed much better than the SARIMA model, achieving an overall accuracy of 83% compared with only 58% for the SARIMA. The SARIMA was limited by its linear constraints on the classification tasks and therefore underpredicted the amount of PM_2.5 found in the real world. In turn, this resulted in the SARIMA generating a widening 95% confidence interval that eventually extended into negative values over lengthy time horizons. By contrast, through the use of multi-head self-attention and a rolling window of 12 months to look back, the Transformer model more closely matched ground-truth trajectories while never producing unrealistic structural variations from the actual data. The Transformer model also had perfect 100% precision for the AQI_Increase class; it successfully captured complex seasonal patterns without generating false positives. Although the absence of spatial components throughout distributed tracking networks limited this research, it addressed an important research gap for urban centers in South Asia.

Keywords: Air Quality Index (AQI), Particulate Matter PM 2.5 Forecasting, Transformer Architecture, SARIMA Model, Self-Attention Mechanism, Urban Smog, Time-Series Analysis, Lahore, Pakistan, Predictive Analytics.

Introduction:

Pollution in the atmosphere has become an major environmental and public health concern among most developing nations worldwide; especially in urban areas with higher amounts of people moving around than other regions (cities) of the world. Of the various sources of air pollutants, the most dangerous are those classified as particulate matter; specifically, those with an average diameter of less than 2.5 micrometers (PM2.5), because these particles can penetrate deep into the lungs (respiratory system) and even possibly get into the bloodstream. When this happens, it can potentially create serious (or life-threatening)

medical consequences for individuals. For example, Lahore is the largest city in Pakistan (after Karachi); and has also been exposed to elevated levels (high concentrations) of PM2.5, due mainly to rapid increases in urbanization, overall industrial activity, traffic, construction, as well as seasonal smog, which has continually degraded the air quality in this city.

As a result, the increasing and sustained deterioration of air quality in Lahore raises serious concerns over the impacts that degraded quality of air may have on public health, environment sustainability and overall quality of life for Lahore's citizens. Therefore, to effectively respond to these issues, there is an urgent need to conduct accurate analyses and forecasts of PM2.5 levels, to support and inform current environmental monitoring systems, assist policymakers and community leaders/managers in finding timely strategies for preventing/reducing the occurrence of hazardous pollution materials, as well as provide citizens with tangible means of protecting themselves against hazardous pollution exposure.

The prediction of PM2.5 levels is difficult due to the existence of time relations, seasonal patterns and non-linear characteristics in the relevant data. In order to model patterns like these, many researchers use time-series forecasting methods. One of the oldest and most widely used types of statistical models used for identifying seasonal time series patterns and temporal correlations is called SARIMA (Seasonal Auto-regressive Integrated Moving Average). Unfortunately, most of the traditional time-series forecasting methods struggle with the complexities and/or long-term dependencies associated with environmental datasets.

The recent emergence of deep learning based upon Transformers has resulted in the introduction of models that incorporate self-attention mechanisms that better learn the complex relationships between points in time as well as how to use long-term dependencies more effectively. In this study, we compare the performance of the SARIMA model and the Transformer-based forecasting method for predicting PM2.5 levels in Lahore. The goal of this work is to develop a valid and reliable framework for predicting air quality and making decisions regarding the environment.

The growing levels of PM2.5 pollutants in Lahore present a major environmental and health risk, especially during the seasonal occurrence of smog events, where they cause hazardous air quality conditions. Accurate predictions of PM2.5 levels are needed to allow timely interventions, as well as the development of supporting environmental policies and the protection of public health. Traditional time series forecasting methods like ARIMA and SARIMA are useful for identifying seasonal behavior and time-dependent relationships in air quality datasets, but have difficulty modeling complicated non-linear relationships between variables[1],[2], [3],[4].

New technological developments have resulted in deep learning models built on a Transformer architecture that can successfully model long-term time-dependent relationships and surpass traditional methods (i.e., statistical and recurrent neural network) for forecasting accuracy [5],[6],[7],[8],[9],[10]. In addition, there have been a number of studies that have successfully used Transformer models for predicting PM2.5 and air quality, with many reporting increased accuracy and strength when used to model complicated environmental datasets[11], [12] [13] [14] [15] [16] [17] [18] [19], [20], [21] [22], [23],[24]. However, to date there has not been much research examining the relative effectiveness of traditional seasonal time-series models versus modern Transformer architecture for predicting PM2.5 levels in Lahore. As such, the goal of this paper is to compare the performance of SARIMA and Transformer-based models using a dataset of PM2.5 levels in Lahore against a dataset of PM 2.5 levels in Lahore and then determine which model provides the best forecast for use in air quality management.

Research Objectives:

This research aims to create an accurate forecasting framework for predicting PM 2.5 concentrations in Lahore through the use of historical air quality data. In addition, this study

will analyze the historical PM2.5 concentrations in Lahore to identify both temporal patterns and seasonal variations and evaluate how traditional statistical analyses compare with modern deep learning approaches in terms of their ability to forecast PM2.5 concentrations. A comparison of the SARIMA and Transformer models will also occur to evaluate the forecasting effectiveness of each. Finally, this research attempts to identify which model uses the most effective method for modelling the complex temporal dependencies associated with air quality data. The results of this study will assist in supporting the development of a suitable environmental monitoring system in Lahore and supporting evidence-based policymaking to address air pollution management issues in the city.

The specific objectives of this research are:

- To preprocess and prepare the PM2.5 dataset for time-series forecasting.
- To develop and implement a SARIMA model for PM2.5 prediction.
- To design and train a Transformer-based model for PM2.5 forecasting.
- To evaluate the performance of SARIMA and Transformer models using Accuracy, Precision, Recall and F1-score.

Research Gap and Novelty:

The performance of Transformer-based architectures for air quality forecasting and PM2.5 prediction is well documented, with studies showing that these models outperform traditional deep learning methods in their ability to leverage long-term temporal relationships when forecasting air quality [6] [7] [9] [10] [12] [13] [14] [15] [16] [17]. In addition, statistical time-series models like ARIMA and SARIMA are still frequently used for analyzing time series environmental data because they can capture both seasonal and temporal patterns in the data [1],[2],[4]. The vast majority of previous research has focused on air quality data from China, Europe, and North America; however, there has been little research on air quality data from Lahore (where PM2.5 levels are particularly problematic), particularly with the goal of comparing how traditional SARIMA models perform compared to modern Transformer-based models on the same Lahore PM2.5 dataset using a standardized evaluation mechanism.

This research is unique because it is the first to compare two different techniques of PM2.5 forecasting (SARIMA and Transformer) using the historical air quality data for Lahore, whereas previous studies only looked at one type of technique. The comparison is important because it shows how both statistical (SARIMA) and deep learning methods (Transformers) can be applied to accurately predict seasonal trends and long-term dependencies. Additionally, this research provides evidence of the potential use of advanced Transformer architectures in predicting air quality in Pakistan, as well as providing a point of reference for those who want to use transformers for PM2.5 forecasting in Lahore. By addressing an environmental issue specific to a region, and completing a thorough comparative analysis of the two techniques, the results will assist other researchers in developing reliable urban air quality forecasting systems.

Contributions:

A comparative framework approach is used to compare two different methods for predicting the concentration of PM2.5 particles within Lahore, namely by using the more traditional method of implementing a statistical time series model (SARIMA) along with a newer model based on a deep-learning algorithm (Transformer). By assessing the historical changes over time of PM2.5 concentrations in Lahore, trends of temporal, seasonal, and longitudinal behavior can be investigated that could provide an understanding of how these factors affect the ambient air quality of the city. The relative performance to forecasted PM2.5 concentrations of both models was measured through traditional error metrics so that an appropriate comparison could be made between the statistical and deep-learning methodologies. The results of this study provide insight into the appropriateness of using

either a statistical or deep-learning method for forecasting PM2.5 concentrations within Lahore and help to contribute to the establishment of reliable environmental-monitoring and decision-support systems for the city.

Literature Review:

As the health risks associated with PM2.5 pollution become more pronounced, air quality prediction has been the focus of developing a wide variety of methods to make accurate forecasts of PM2.5 concentrations for use by governments and environmental agencies to mitigate PM2.5concentrations and issue public health warnings. While both statistical approaches and deep learning techniques have been used to model air quality data, it has been found that their performance depends on the characteristics of the dataset and the forecasting horizon [22], [4],[25].

Traditional timeseries forecasting techniques such as ARIMA and SARIMA have an established history of use for air quality prediction due to their ability to capture time correlated trends and seasonal effects[14]. SARIMA has demonstrated an ability to model the seasonal component of PM2.5 in a number of recent studies [2].Combining SARIMA models with deep learning methods has demonstrated some improvement in forecasting performance [1].Similarly, integrating statistical techniques with neural network methods has produced positive outcomes for air quality prediction tasks [4].

Deep learning has advanced rapidly by developing Transformer architectures as excellent options for time-series forecasting. To overcome the drawbacks of recurrent neural networks, Informer [6], Auto former[7] and FED former [8] were created using self-attention mechanisms to effectively achieve long-term dependency modeling. All of these models have produced state-of-the-art results across numerous forecasting applications and laid the groundwork for countless other recent environmental prediction studies. Recent studies have also looked into the application of transformer models for predicting PM2.5 levels, confirming their usefulness for more general projects on the country and environmental level [26].

Specifically, several research efforts have utilized Transformer-based methods to successfully forecast PM2.5 and air quality data. Using sparse attention-based Transformer networks has enabled significant advancements compared to traditional methods of forecasting [12].Combining Transformer neural networks with data embedding methods has led to accurate PM2.5 forecasting [13] while utilizing multi-site Transformer architectures along with meteorological features has further increased forecasting accuracy [14].

Therefore, these results demonstrate the capabilities of attention-based modeling methods when applied to environmental monitoring. Recent studies of the deep learning time-series models have highlighted the need to evaluate their forecasting performance for various prediction horizons, specifically in the case of PM2.5 forecasting [27]. Reformer architectures are also utilized for forecasting air pollution data, making use of efficient Transformer variants for forecasting environmental time series [28].

Researchers have analyzed several Hybrid and Decomposition-based Models to improve forecasting accuracy. One excellent example that has achieved outstanding performance predicting long term PM2.5 levels is a CEEMDAN-Decomposed Dataset with a Deep-Transformer Model [15]. An additional example of improved air quality prediction using Time Series Decompositions integrated with Sparse-Self Attention Mechanisms is available [17]. When assessed against other models, Hybrid CNN-LSTM networks have been found to outperform traditional recurrent models based on Attention Mechanisms when applied to Air Quality datasets exhibiting High Complexity [16].

Methods based on CNN-LSTM have been studied for the prediction of PM2.5 and AQI in terms of the strength of the combination of the two architecture types for learning complicated patterns in the air quality data[29]. It has also been suggested to use hybrid

approaches, which entail integrating classical time series models with deep learning techniques in air quality predictions [30].

Moreover, recent research has used Spatial & Meteorological data within forecasting models to allow researchers to be able to accurately model the Environmental Dynamics associated with Poor Air Quality. For example, Hybrid Models have been developed that utilize Graph Neural Networks & Attention Mechanisms along with strong performance in the Spatiotemporal prediction of PM2.5 [25]. The use of Graph Neural Networks and Transformative Approaches modelling Spatial Relationships between Monitoring Stations can also be found to be successful [20]. Collectively, there is now an extensive amount of research showing that the addition of Contextual Information will substantially increase the predictive capabilities of forecasting models.

There is still a significant gap in research related to forecasting PM2.5 levels in Lahore, Pakistan, despite considerable advancements being made in the field. Most studies on PM2.5 use data from China, Europe, and North America; there are relatively few studies on using more advanced Transformer architectures to predict PM2.5 in Pakistan [12][13][14][15][16][17], [20],[21]. Additionally, there are very few studies comparing the results of SARIMA models with Transformer-based models when using the same dataset of Lahore PM2.5 levels. The goal of this research is to bridge the existing research gaps by evaluating both SARIMA and Transformer-based models' ability to predict PM2.5 levels in Lahore; therefore, increasing knowledge regarding their use for forecasting urban air quality levels.

Methodology:

The proposed methodology assesses the forecasting of PM2.5 concentrations through both traditional time series techniques and deep learning methods. The overall method consists of collecting and preparing the data through a series of stages that include identifying and handling any missing, duplicate or outlier data so that the data will be in good condition; the cleaned dataset will then be converted to a suitable time series format and split into two separate groups, training and testing.

The data will be used to develop and train 2 forecasting models, called SARIMA (Seasonal Autoregressive Integrated Moving Average) and Transformer, in order to conduct a fair comparison between the two model results. The SARIMA model was designed to model seasonal and linear temporal patterns while the Transformer model was built using self-attention mechanisms to learn data's long-term dependencies. After the two models have been trained, their performance will be assessed by using their accuracy, precision, recall and their f1 score to ultimately determine which model produces the best PM2.5 prediction for Lahore City. The full workflow of the proposed methodology is illustrated in Figure 1.

Figure 1 shows the framework process for the collection of PM2.5 and meteorological data, followed by preprocessing of the PM2.5 and meteorological data. Preprocessing of the PM2.5 and meteorological data includes duplicate removal, missing value handling, outlier treatment, and the aggregation of daily observations to create monthly averages. Processed PM2.5 values will then be converted to an AQI category and divided into training and testing datasets. AQI prediction will be performed using two models in parallel; a Seasonal Auto-Regressive Integrated Moving Average (SARIMA) model, which will analyze seasonal patterns exhibited in time series, and a Transformer model, which will analyze temporal dependencies using self-attention mechanisms. Both models will be evaluated using accuracy, precision, recall, and F1-score to determine the preferred approach to AQI prediction. The main variable to be predicted is PM2.5, although the Transformer employs past PM2.5 data to categorize the following month in terms of either AQI_Increase or AQI_Decrease.

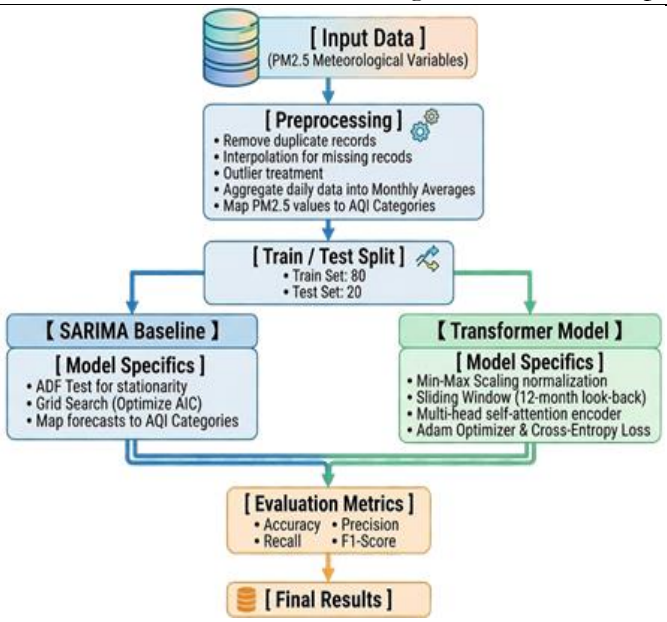


Figure 1. Methodology

Data Preprocessing and Normalization:

In order to establish the integrity of the PM2.5 and weather data collected for this analysis, there were several preprocessing steps. Initial data discrepancies (duplicate records) were identified and eliminated. Missing data was corrected via interpolation, having been forward/backward filled where appropriate. Any outliers, or abnormal population metrics, were examined and handled appropriately, to reduce possible adverse effects on model performance.

To avoid time-related data leakage, the data set was first divided into chronological training and test sets of 80% and 20% respectively. In addition, all parameters relating to preprocessing such as handling missing values, outliers, and the Min-Max method were obtained from the training set only. Afterward, the same methods of preprocessing were applied to the previously unseen test data.

After preprocessing the data, normalization occurred via the Min-Max Scaling approach, to improve numerical stability and thereby enhance efficiency during model training. Min-Max normalization takes feature data and transforms it into a specified range (0 to 1), while maintaining the original feature's relative relationship to all other original features. This process is easily illustrated with any of the numerous deep learning models (e.g., Transformers) and is critical, because it alleviates (or minimizes) excessive proportions of larger magnitude feature data (relative to smaller proportion features), and subsequently quickens the time to (or, lower the quantity of time needed for) convergence for any optimization process. The resultant normalized dataset was utilized to develop training/testing samples of SARIMA and Transformer-based models to predict the AQI.

SARIMA Model:

The Seasonal Auto Regressive Integrated Moving Average (SARIMA) model is the statistical model that serves as the baseline for AQI predictions because of its ability to accurately depict both temporal and seasonal changes within a time-series of data. In this study, the parameters for the SARIMA model were determined by analyzing the stationarity of the time series and using the Akaike Information Criterion (AIC) in modeling. In order to establish the degree of differencing needed and whether the time series was stationary, the Augmented Dickey-Fuller (ADF) test was employed. Several combinations of non-seasonal (p, d, q) and seasonal parameters (P, D, Q) were tested using grid search and using AIC as the parameter. As the data is monthly, the seasonal period was defined on the basis of seasonal pattern.

Once the optimal configuration of the parameters was determined, the SARIMA model was developed using the previously observed PM2.5 data from the training set. The SARIMA model learned how to make future forecasts by capturing the seasonal and temporal dependencies contained in the air quality data. After making predictions, we compared these predictions to actual observations in the testing dataset; thus, we were able to assess how well the SARIMA model could model AQI changes and establish a baseline against which to evaluate the Transformer model.

Transformer Model:

The research developed a predictive modeling approach based on a Transformer deep learning architecture to forecast AQI categories from historical PM2.5 measurements. Unlike standard time series models, the Transformer model employs a self-attention mechanism to model both short- and long-term dependencies in sequential data. The model consists of an input projection layer, positional encoders, multiple encoder layers and an output classification layer. Positional encoders provide a means of preserving temporal order, and the model is able to recognize significant associations between earlier observations through multi-head self-attention.

The Transformer model implemented a 12-month lookback period in order to forecast air quality index class for the following month. The structure comprises an input projection layer, positional encoding, transformer encoder layers, and finally an output classification layer. The application of Adam optimizer and categorical cross entropy loss was used in the process of training. The model was trained using an Adam optimizer with a categorical cross-entropy loss function through iterative training. This developed a robust feature representation of the historical PM2.5 time series, allowing the model to predict AQI classification and the ability to capture complex pollution trends not captured in traditional statistical models.

Model Training and Prediction:

To ensure a fair and consistent evaluation, both models use identical datasets for both training and testing purposes to evaluate fairness and consistency of results from both models. PM2.5 observations were split into separate training and testing sets, with 80% allocated for training, and the remaining 20% allocated for testing. Each model learned the fundamental patterns of AQI data during training. The SARIMA method uses statistical relationships to perform this function and the Transformer method uses neural network optimization. The PM2.5 readings were translated into AQI trend categories by examining the monthly values one after another. A rise of AQI value was categorized into AQI_Increase class while a decrease was put in AQI_Decrease category.

While testing was being performed on the unseen test dataset by the respective models, each model was able to predict the PM2.5 values based on their technique's inherent characteristics. The predicted PM2.5 values were compared against the actual recorded values for each AQI category in order to determine the success rate of each model predicting AQI. Detailed analysis of the predicted data provided insight into the advantages and disadvantages of each model and will be helpful in developing methods of using the models to predict AQI in Lahore.

Mathematical Formulations:

The SARIMA and Transformer mathematical formulations that provide the mathematical basis of their mechanisms used in AQI prediction have been explained in this section. The SARIMA model is able to model both seasonal and non-seasonal patterns that exist in time series data, and the Transformer model utilizes self-attention mechanisms to learn complex temporal relationships. The evaluation metrics for the proposed models have also been discussed.

The SARIMA model is represented as:

$$\text{SARIMA}(p,d,q)\ (P,D,Q)s$$

where p denotes the autoregressive order, d represents the degree of differencing, q indicates the moving average order, and P, D, Q correspond to the seasonal autoregressive, differencing, and moving average orders, respectively. The parameter s denotes the seasonal period.

For the Transformer model, the self-attention mechanism is computed as:

$$\text{Attention}(Q,K,V) = \text{Softmax}((QK^T / \sqrt{d_k}))V$$

where Q represents the query matrix, K denotes the key matrix, V represents the value matrix, and d_k is the dimension of the key vectors. This mechanism enables the model to capture relationships between different time steps in the sequence.

The Min-Max normalization technique used for data scaling is defined as:

$$X' = (X - Xmin) / (Xmax - Xmin)$$

where X is the original value, $Xmin$ is the minimum value, $Xmax$ is the maximum value, and X' is the normalized value.

The evaluation metrics employed in this study are expressed as:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN)$$

$$\text{Precision} = TP / (TP + FP)$$

$$\text{Recall} = TP / (TP + FN)$$

$$\text{F1-Score} = 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$$

where TP, TN, FP, and FN denote True Positives, True Negatives, False Positives, and False Negatives, respectively. These metrics are used to evaluate and compare the classification performance of the SARIMA and Transformer models for AQI prediction.

Experimental Setup:

The Experimental Setup and Model Description with all the details is presented in Table 1:

Table 1. Experimental Setup and Model Description

Category	Description
Programming Language	Python 3.13
Development Environment	Jupyter Notebook
Libraries & Tools	Python, Pandas, NumPy, PyTorch, Scikit-learn, MinMaxScaler, Transformer Encoder, Adam Optimizer, Matplotlib, Statsmodels (SARIMAX), Augmented Dickey-Fuller (ADF) Test
Operating System	Windows based system with GPU
Hardware	AMD Ryzen 5 8645HS, NVIDIA RTX 4050, RAM = 32, Storage = 1TB
Validation Technique	Train-Test Split (80:20), Accuracy Evaluation,Precision Score, Recall Score, F1-Score Evaluation,

Dataset Description:

Data used in this study include historical PM2.5 air quality data collection of Lahore, Pakistan. PM2.5 data contains time-dependent observations of PM2.5 concentrations and corresponding meteorological data that affect air quality conditions. One of the main goals is to analyze PM2.5 which includes temperature, humidity, dew point, wind speed, and pressure. A total number of 221 missing data for PM2.5 but not for meteorological parameters is given in Table 2. Over multiple years, researchers have made enough observations to allow them to examine pollution patterns, seasonal changes, and long-term trends. As the most harmful of the air pollutants, researchers selected PM2.5 as the primary indicator of air quality and the basis for assessing the impact of air quality on public health.

Table 2. Dataset Description

Parameter	Description

Data Type	Tabular Time-Series
File Format	.csv Files
Total Entries	1644
Aqi-pm2.5	31-503

Results:

The following area includes the data gathered through SARIMA and Transformer models to study how AQI can be predicted with Lahore's PM2.5 report from the time period used for each of these models. Both methods were evaluated using different measures of accuracy, such as Precision, Recall and F1. Then the comparative analysis was performed to understand how well both of these methods could determine the temporal patterns in AQI and classify the AQI estimates. The results provide insight into both methods, including the limitations and benefits of the two approaches to AQI forecasting, and will allow for further analysis of the best possible models to support air quality measurements and aid in the prediction and/or decision-making regarding air quality management in Lahore.

SARIMA:

The performance is presented in Table 3:

Table 3. Performance of Proposed Framework

	Accuracy	Precision	Recall	F1-score
Aqi_Decrease	0.58	0.57	0.67	0.62
Aqi_Increase	0.58	0.60	0.50	0.55

As indicated by the data in table 3, the model exhibited 58% accuracy of both AQI increase and AQI decrease predictions. The precision, recall, and F1 score for AQI_Decrease prediction was 57%, 67%, and 62%, respectively, which shows an acceptable level of performance of identifying improving air quality instances. However, the precision, recall and F1 Score for AQI_Increase prediction was 60%, 50% and 55%, respectively, which indicates reasonable accuracy in identifying instances of increasing AQI levels but there are a number of actual positive cases that were incorrectly classified. The greater recall for AQI_Decrease compared to the AQI_Increase indicates that predictive accuracy for pollution reduction is higher than predictive accuracy for pollution increase. Moderate F1 scores reflect moderate, stable performance, with a substantial predictive ability for predicting reductions in end-of-period pollution that exceeds the predictive capacity for predicting increases in pollution for the period prior to the end of the period.

SARIMA Forecast vs. Actual Analysis:

Figure 2 illustrates how the statistical baseline SARIMA model's performance mirrors the historical training data (January 15 - December 31, 2022) and the validation data (June 1, 2023). The in-sample fit (green dotted line) captures broad seasonal cycles, but does not capture local historical volatility. In the 2023 test period (June - December), the SARIMA forecasts (red dashed line) consistently under-predict actual (observed) PM2.5 levels and fail to capture short-term fluctuations accurately. The 95% CI (pink shaded area) expands quickly over time, dips below 0 (impossible) and indicates highly variable long-term statistical uncertainty. The model detects a moderate increase in the trend at the end of the year; however, Due to the linear constraints present, the model cannot demonstrate a precise mapping ability for air quality classifications that exhibit highly variable patterns in real life.

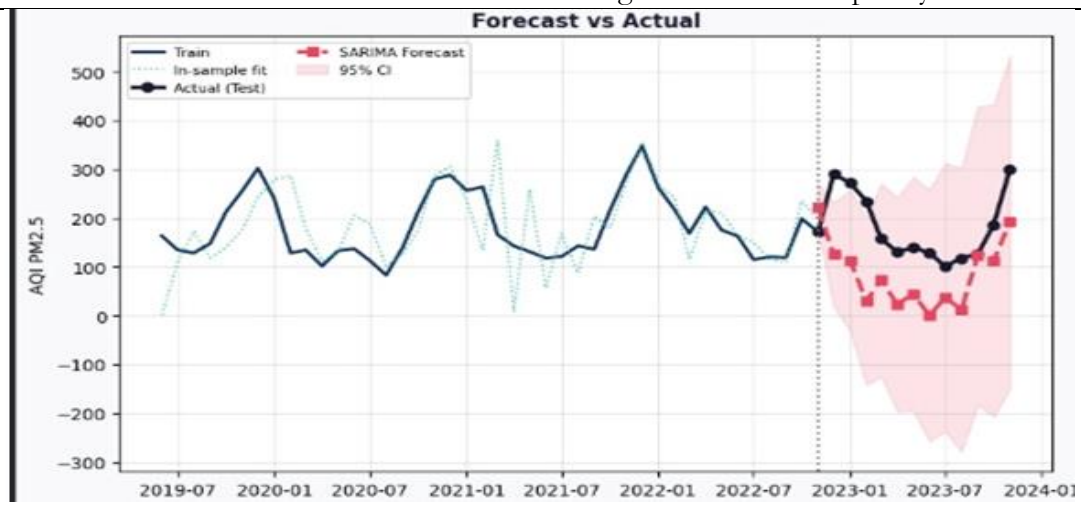


Figure 2. Forecast vs. Actual Analysis

Transformer:

The performance is presented in Table 4

Table 4. Performance of Proposed Framework

	Accuracy	Precision	Recall	F1-score
Aqi_Decrease	0.83	0.75	1.00	0.86
Aqi_Increase	0.83	1.00	0.67	0.80

Table 4 indicates the accuracy of the Transformer Model is high, 83%, and shows that this model is well suited to predict the trends of AQI (Air Quality Index) correctly. For the AQI_Decrease class, the model has produced correct precision, recall and F1-scores of 75%, 100% and 86%, respectively. Thus, it has successfully achieved predicting 100% of true negative AQI_Decrease classes correctly. For the AQI_Increase class, the precision is 100%, which indicates 100% of predicted AQI_Increase classes are correct, and the recall of 67% indicates that some actual AQI_Increase cases were incorrectly classified. However, the F1-score of 80% indicates that this model provides a reasonable balance for True Positives and True Negatives; therefore, the Transformer model has reliably and effectively provided good classification performance in predicting AQI trends.

Transformer Forecast vs. Actual Analysis:

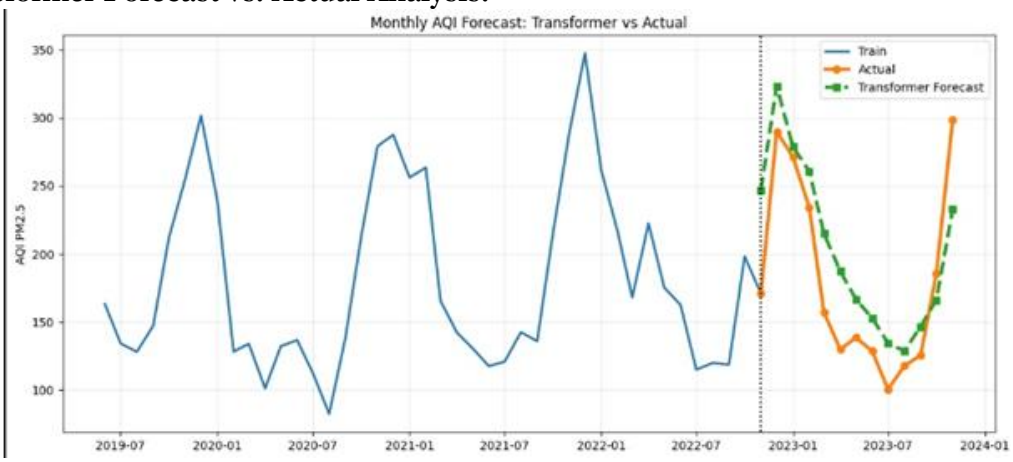


Figure 3. Forecast vs. Actual Analysis

Figure 3 shows how well the Transformer-based deep learning architecture can predict the monthly category values and actual Ground Truths (AQI PM2.5) for the past 12 months. The green dashed line (predictions) almost perfectly matches the orange line (actual) throughout the year and indicates that the multi head self-attention capabilities of the model

and the 12-month sliding lookback window were utilized effectively to enable high-resolution predictions of the future. The model accurately predicted the sharp drop in PM2.5 levels that occurred in mid-2023 and closely followed the sharp increase that occurred in late 2023. In contrast to traditional statistical baselines, the Transformer model has a very tight and focused margin of error with no unrealistic negative predicted values and no extreme deviations from the original values. The close correlation of the two lines during the entire 2023 verification period demonstrates the Transformer's ability to effectively capture complex, nonlinear temporal dependencies to classify long-term AQI levels.

Heatmap:

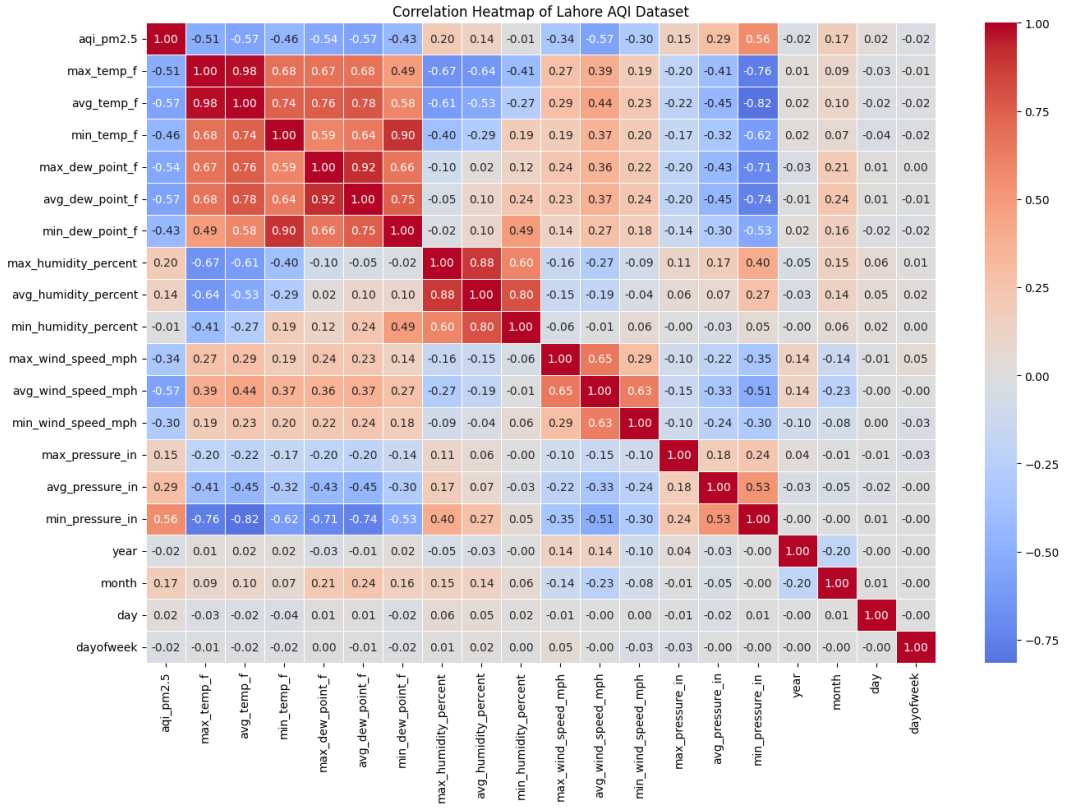


Figure 4. Heatmap

As seen in Figure 4, the correlation heat map shows the Pearson correlation coefficients for the Lahore AQI dataset and how they relate to the linear relationship between the target feature (aqi_pm2.5) and a total of 17 different weather variables and time indices. The color scheme uses deep blue to represent a strong negative correlation and bright red to show a strong positive correlation. As such, aqi_pm2.5 has its largest negative correlation with temperature and dew point variables (i.e., avg_temp_f = -0.57, avg_dew_point_f = -0.57, max_dew_point_f = -0.54). This negative correlation is consistent with winter inversion patterns in which colder, drier air traps particulate material near the surface. On the other hand, there is a strong positive correlation between aqi_pm2.5 and min_pressure_in (0.56) indicating that higher levels of atmospheric pressure create a corresponding increase in particulate matter concentrations. Additionally, the matrix shows a high degree of multicollinearity for certain weather-related features, such as max_temp_f/avg_temp_f (0.98) and max_humidity_percent/avg_humidity_percent (0.88); by comparison, the day (0.02) and day of week (-0.02) indices have negligible linear predictive value.

Discussion:

A side-by-side assessment of the traditional baseline SARIMA and the newer Transformer deep learning architecture sheds light on some important aspects of accurately

modeling complicated environmental time series. The mathematical characteristics of SARIMA are historically used as a very effective statistical benchmark for capturing seasonal cyclical behavior, but they also limit SARIMA's usefulness when it comes to working with highly erratic atmospheric datasets such as Lahore PM2.5. The air quality index is inherently affected by complex combinations of localized regional emission factors in combination with rapidly changing meteorological conditions. Because traditional statistical models assume a relatively constant structural pattern through time, they can't easily be generalized to account for extreme localized spikes in air quality or large-scale drops in air quality.

The difference in performance between these two analytical approaches supports a much stronger case for using data-based neural network architectures to model environmental risk in urban areas. Upon further examination, a significant difference in the two models' performances is apparent when looking at the experimental metrics. Specifically, the Transformer model had an overall accuracy of 83%, whereas SARIMA had an overall accuracy of 58%. The limitations of SARIMA's predictive power can be attributed primarily to its inability to identify non-linear dependencies.

From a classification perspective, SARIMA's F1-scores for AQI decreases and AQI increases were moderate (62% and 55%, respectively). While SARIMA demonstrated a precision of 60% for predicting worsening pollution events, the low recall of 50% indicates that it failed to correctly forecast approximately one-half of the dangerous spikes in pollution. In contrast, the Transformer model had an outstanding precision rate of 100% for the AQI_Increase classification, verifying that every predicted spike was accurate. Additionally, the balanced F1-score of 80% for the Transformer indicates its superior ability to serve as an early warning system for public health concerns.

The visual forecast trajectories further reveal how both methodologies share limitations regarding their core behaviors. Specifically, within the 2023 verification timeline, the SARIMA forecast underpredicted the amount of real life PM2.5, leading to a flattening of acute changes in trends. Additionally, the long-term horizon showed a massive expansion in its 95% confidence interval, which resulted in mathematically impossible negative values. This expansion demonstrates an inherent flaw within traditional statistical models whereby, as the forecast horizon grows larger, the uncertainty associated with the prediction grows exponentially, making the long-term bounds unusable for creating policies.

The Transformer Model circumvented the limitations of the SARIMA method by maintaining a narrower, more constant margin of error and matching the actual peaks and valleys of PM2.5, while not generating extreme, unrealistic overshoots.

The predictive superiority of this architecture is due to both the structural design and the way in which it is developed. SARIMA models use fixed linear lag components to process the same set of input data over a fixed period of time.

The Transformer model, however, uses multi-head self-attention layers in conjunction with 12 months of sliding window length so that the entire annual cycle can be processed at the same time. Additionally, the Transformer uses self-attention mechanisms to assign higher degrees of attention to non-consecutive months of historical data that connect with current conditions. Finally, combining Min-Max normalization with positional encoding to maintain the significant temporal sequence of the data being used to train the network helps ensure that the very large variances in scale do not prevent the gradient optimization process from continuing for all training variables, while still allowing the model to effectively decompose Lahore's complex seasonal smog patterns and multiple meteorological influences at various locations.

From a practical perspective, these findings provide an effective and practical operational framework for supporting urban air quality management. In general, predictive research has focused on specific regions of North America, Europe and East Asia; therefore,

there is a significant gap in research for densely populated cities experiencing high levels of air pollution in the South Asian region, such as Lahore (where predominantly high-density populations live). This study builds a common testing environment with the same training/testing splits in order to establish a local benchmark for use to build future environmental research. The demonstrated high reliability of the Transformer model provides a solid technical basis for building real-time environmental decision support systems that local governments can use to effectively implement targeted traffic control, enforce industrial mitigation policies and issue early advisories to protect at-risk populations.

Implications of the Proposed Methodology:

This new single benchmark for comparing urban air pollution management around the world has allowed us to move forward a great deal in managing the air quality of cities like Lahore, Pakistan which struggles heavily with pollution, and has very high population densities. Thus far, most of the sophisticated environmental models that have been developed have used North American, European, and East Asian datasets. Therefore, there was a significant lack of localized research regarding the unique seasonal smog patterns that exist in the country of Pakistan. By creating a traditional statistical model and an advanced deep-learning model based on identical train/test splits and the same data cleaning process, we have established a new benchmark for future environmental research that relies on a valid and trustworthy method of collecting localized data.

The advancement from traditional time series models (which use linear assumptions) towards architectures that use self-attention shows that deep-learning models are able to accurately model environmental dynamics that are both highly variable and non-linear, and for which traditional methods are unable to provide accurate predictions.

The differences in classification metrics are very noticeable and have implications for the protection of public health, and pro-active governance of cities from a practical and policy perspective.

The excellent performance of the Transformer model, with 100% precision for predicting the AQI_increase class, leads to a complete elimination of false alarms related to worsening pollution occurrences, while the high level of overall accuracy (83%) and an F1-score that indicates a high level of reliability of diagnostic results for implementation provides the necessary basis for the use of these models for actual application in the real-world; further validation is required before operational deployment for early health warnings for urban municipalities (as opposed to waiting until there is enough uncertainty or doubt in the baseline SARNMA model to develop a long-term planning model). In this regard, urban municipalities can provide early health warning, with confidence, due to the highly predictable performance of these models, thereby allowing environmental agencies, in conjunction with city policymakers, to have a solid technical basis upon which to implement timely\, targeted strategies for mitigating the impacts of further industrial emissions, reduced vehicular traffic on the road, and enacted plans for "smart zones" during periods of heightened risk for seasonal smog.

Limitations:

While the Transformer model has shown superior performance and predictive accuracy, it is essential to recognize various limitations attributable to the model's structure. The model is dependent upon data from one source (historical records for Lahore, Pakistan) and therefore may limit its immediate geographic generalizability to other areas (e.g., similar cities with varying climate patterns or urban topography). In addition, missing data that was treated with interpolation before beginning analysis as part of the Data Preparation Phase and duplicate entries in the database were removed to enhance the quality of analysis.

However, the data use is contingent upon continually monitoring the data being recorded through a sensor over time. Should either a localized sampling site fail for an

extended period of time due to equipment malfunction or data dropout, then the use of interpolation to fill in gaps of missing information may lead to an artificial smoothness in the overall dataset that obscures some of the very volatile nature of many environmental anomalies and extreme pollution peak events found within the dataset. These issues could potentially be problematic for accurate training of deep learning architecture(s) and therefore continue to highlight the need for improvements in both the quality and quantity of data being recorded at monitoring stations.

In addition to this, there exists a key structural limitation in the existing framework being used which currently cannot model spatial relationships. The Transformer uses multiple heads through its self-attention mechanism to properly model complex temporal dependencies and also seasonal behaviors of smog. This occurs over a twelve-month time period of looking back. However, the urban area within the model is treated as an entire aggregation of an area and has no consideration of the distributed nature of the geographical dimensions that would distinguish it from the model. In real life, localized spatiotemporal correlations have a much stronger influence on air pollution levels (example: transboundary pollutants moved by wind, industrial clustering, and vehicle traffic density) than modelling the overall region as one entity. As a result, since there are no spatial graph network components or contextual geographical embeddings in the current data model, it will not be able to forecast how each specific sub-region/neighborhood will experience different AQI category levels over time.

Conclusion:

The purpose of this research is to analyze the relationship between seasonality and weather patterns in Lahore's air quality index via the comparison of forecasting methods. The study also examined whether these two variables interact in predictable ways or merely exist together with no clear causation. Forecasting methods include: Seasonal Auto Regression Integrated Moving Average (SARIMA) and Transformer deep learning model architecture. Our approach to this research includes establishing an integrated evaluation framework using identical data preprocessing, aggregate monthly observations and train-test splits of 80/20 to juxtapose traditional statistics with self-attention mechanisms.

The results of our experimentation indicate that while SARIMAs can be used to accurately model short-term, long-range trends in seasonal patterns, they are restricted by the linearity of their models and therefore continually underestimate the actual PM2.5 measurements of pollutant emissions. Furthermore, the confidence intervals produced by SARIMA are continually expanding and thus yield highly variable results over long-range forecasting horizons. In contrast, the Transformer model performs significantly better than SARIMA in its ability to predict air pollution levels by accurately classifying AQI levels long-term through its high predictive performance when analyzing highly variable environmental conditions.

The performance metrics illustrate that the deep learning models for AQI forecasting have a distinct advantage over the SARIMA forecasting model. The engineering and performance metrics for this study indicate that the Transformer deep learning model achieved 83% overall forecasting accuracy, compared to only 58% for the SARIMA baseline model. In addition, the Transformer deep learning model achieved a perfect 100% precision in predicting the occurrence of events that increase the AQI. Nonetheless, these results must be confined to the dataset or test set at hand and cannot be interpreted in terms of overall superiority.

Further validation on independent, long-term, and multi-station datasets is necessary to evaluate the level of generalizability and practicality of the findings observed in the real world. Although the current approach to modelling does not include spatial components to support the distributed nature of tracking stations, it at least fills a significant regional research gap for densely populated and heavily polluted urban areas, such as Lahore, South Asia. This

study establishes a localized computational basis to provide confidence for the creation of real-time environmental decision-support systems, enable policymakers to implement timely strategies to mitigate exposure to pollution, and protect the health of citizens from the adverse effects of smog.

Future Directions

To further develop the findings of this study, as well as to rectify the limitations of this study, several possible directions for future research exist. First, future work should integrate spatial modeling components with the existing temporal network by developing hybrid spatiotemporal deep learning models. The integration of Transformer layers with Graph Neural Networks (GNNs) or Spatiotemporal Graph Convolutional Networks (STGCNs) will enable models to capture both long-range temporal dependencies and the complex geographic relationships of distributed air quality monitoring stations simultaneously.

Enhancement of these capabilities will allow for tracking micro-level spatial differences associated with a variety of neighborhoods in Lahore, such as localized differences created by traffic bottlenecks, single-source industrial discharge, and cross-border transport of contaminants. Furthermore, increasing the number of features in the data may provide another way to enhance model performance for the long term and increase predictive stability. Moving forward, future versions must include additional exogenous data sources, such as: real-time traffic density trends; satellite-based aerosol optical depths; agricultural burning schedules; and regional climate indices spanning multiple years.

The incorporation of these multiple forms of data within multimodal fusion networks would improve the ability of deep learning models to predict sudden, abnormal pollution spikes in comparison to only using historical interpolation. Lastly, this process should also be utilized to compare multiple cities within other South Asian metropolitan areas that experience elevated levels of air pollution across the region, thus supporting evidence for the geographical generalizability of Transformer-based models and providing evidence to support the development of an integrated regional early warning system.

References:

[1] "Hybrid SARIMA and Deep Learning Framework for Air... - Google Scholar." Accessed: Sep. 15, 2026. [Online]. Available: https://scholar.google.com.pk/scholar?q=Hybrid+SARIMA+and+Deep+Learning+Framework+for+Air+Pollution+Forecasting,&hl=en&as_sdt=0&as_vis=1&oi=scholart

[2] "Seasonal Time Series Analysis of PM2.5 Concentrations... - Google Scholar." Accessed: Sep. 15, 2026. [Online]. Available: https://scholar.google.com.pk/scholar?q=Seasonal+Time+Series+Analysis+of+PM2.5+Concentrations+Using+SARIMA+Models,&hl=en&as_sdt=0&as_vis=1&oi=scholart

[3] M. U. Gondal, H. ul Qudous, A. A. Farhan, and S. Alamri, "Interpretable PM2.5 Forecasting for Urban Air Quality: A Comparative Study of Operational Time-Series Models," *cornell Univ.*, Mar. 2026, Accessed: Sep. 07, 2026. [Online]. Available: https://arxiv.org/pdf/2603.25495

[4] "Hybrid ARIMA-LSTM Model for Air Quality Forecasting, - Google Scholar." Accessed: Sep. 15, 2026. [Online]. Available: https://scholar.google.com.pk/scholar?q=Hybrid+ARIMA-LSTM+Model+for+Air+Quality+Forecasting,&hl=en&as_sdt=0&as_vis=1&oi=scholart

[5] S. Wang, Y. Li, J. Zhang, Q. Meng, L. Meng, and F. Gao, "PM2.5-GNN: A Domain Knowledge Enhanced Graph Neural Network For PM2.5 Forecasting," *GIS Proc. ACM Int. Symp. Adv. Geogr. Inf. Syst.*, vol. 11, pp. 163–166, Oct. 2021, doi: 10.1145/3397536.3422208.

[6] Haoyi Zhou, Shanghang Zhang, Jieqi Peng, Shuai Zhang, Jianxin Li, Hui Xiong, Wancai Zhang, "Informer: Beyond Efficient Transformer for Long Sequence Time-

Series Forecasting," *arXiv:2012.07436*, 2020, [Online]. Available: https://arxiv.org/abs/2012.07436

[7] Haixu Wu, Jiehui Xu, Jianmin Wang, Mingsheng Long, "Autoformer: Decomposition Transformers with Auto-Correlation for Long-Term Series Forecasting," *arXiv:2106.13008*, 2021, [Online]. Available: https://arxiv.org/abs/2106.13008

[8] T. Zhou, Z. Ma, Q. Wen, X. Wang, L. Sun, and R. Jin, "FEDformer: Frequency Enhanced Decomposed Transformer for Long-term Series Forecasting," *Proc. Mach. Learn. Res.*, vol. 162, pp. 27268–27286, Jan. 2022, Accessed: Sep. 07, 2026. [Online]. Available: https://arxiv.org/pdf/2201.12740

[9] G. Woo, C. Liu, D. Sahoo, A. Kumar, and S. Hoi, "ETSformer: Exponential Smoothing Transformers for Time-series Forecasting," Feb. 2022, Accessed: Sep. 07, 2026. [Online]. Available: https://arxiv.org/pdf/2202.01381

[10] Yuqi Nie, Nam H. Nguyen, Phanwadee Sinthong, Jayant Kalagnanam, "A Time Series is Worth 64 Words: Long-term Forecasting with Transformers," *arXiv:2211.14730*, 2023, [Online]. Available: https://arxiv.org/abs/2211.14730

[11] T. C. Bui *et al.*, "STAR: Spatio-Temporal Prediction of Air Quality Using a Multimodal Approach," *Adv. Intell. Syst. Comput.*, vol. 1251 AISC, pp. 389–406, 2021, doi: 10.1007/978-3-030-55187-2_31/SAVE-RESEARCH.

[12] Z. Zhang & S. Zhang, "Modeling air quality PM2.5 forecasting using deep sparse attention-based transformer networks," *Int. J. Environ. Sci. Technol.*, vol. 20, pp. 13535–13550, 2023, [Online]. Available: https://link.springer.com/article/10.1007/s13762-023-04900-1

[13] J. Limperis, W. Tong, F. Hamza-Lup, and L. Li, "PM2.5 forecasting based on transformer neural network and data embedding," *Earth Sci. Informatics*, vol. 16, no. 3, pp. 2111–2124, Sep. 2023, doi: 10.1007/S12145-023-01002-X.

[14] H. Wang, L. Zhang, and R. Wu, "MSAFormer: A Transformer-Based Model for PM2.5 Prediction Leveraging Sparse Autoencoding of Multi-Site Meteorological Features in Urban Areas," *Atmos. 2023, Vol. 14, Page 1294*, vol. 14, no. 8, p. 1294, Aug. 2023, doi: 10.3390/ATMOS14081294.

[15] Q. Zeng, L. Wang, S. Zhu, Y. Gao, X. Qiu, and L. Chen, "Long-term PM2.5 concentrations forecasting using CEEMDAN and deep Transformer neural network," *Atmos. Pollut. Res.*, vol. 14, no. 9, p. 101839, Sep. 2023, doi: 10.1016/J.APR.2023.101839.

[16] B. Cui, M. Liu, S. Li, Z. Jin, Y. Zeng, and X. Lin, "Deep learning methods for atmospheric PM2.5 prediction: A comparative study of transformer and CNN-LSTM-attention," *Atmos. Pollut. Res.*, vol. 14, no. 9, p. 101833, Sep. 2023, doi: 10.1016/J.APR.2023.101833.

[17] W. Cao, W. Qi, and P. Lu, "Air Quality Prediction Based on Time Series Decomposition and Convolutional Sparse Self-Attention Mechanism Transformer Model," *IEEE Access*, vol. 12, pp. 155340–155350, 2024, doi: 10.1109/ACCESS.2024.3484579.

[18] "Transformer-Based Air Quality Prediction Using Multi-Sour... - Google Scholar." Accessed: Sep. 15, 2026. [Online]. Available: https://scholar.google.com.pk/scholar?q=Transformer-Based+Air+Quality+Prediction+Using+Multi-Source+Environmental+Data,"&hl=en&as_sdt=0&as_vis=1&oi=scholart

[19] "Attention-Driven Transformer for Urban PM2.5 Forecasting," - Google Scholar." Accessed: Sep. 15, 2026. [Online]. Available: https://scholar.google.com.pk/scholar?q=AttentionDriven+Transformer+for+Urban+PM2.5+Forecasting,"&hl=en&as_sdt=0&as_vis=1&oi=scholart

[20] Y. Ye, Y. Cao, Y. Dong, and H. Yan, "A graph neural network and Transformer-based model for PM2.5 prediction through spatiotemporal correlation," *Environ. Model. Softw.*, vol. 191, p. 106501, Jun. 2025, doi: 10.1016/J.ENVSOFT.2025.106501.

[21] Y. Xia and K. Cai, "iTransformer-MDC for short-term PM2.5 forecasting: A case study of Beijing," *Atmos. Pollut. Res.*, vol. 17, no. 7, p. 102988, Jul. 2026, doi: 10.1016/J.APR.2026.102988.

[22] A. Rashid and R. M. Atif, "Assessing meteorological influences and predicting PM2.5 using deep learning approach: A case of Asian metropolitan city, Lahore," *Atmos. Environ.*, vol. 381, p. 122110, Sep. 2026, doi: 10.1016/J.ATMOSENV.2026.122110.

[23] K. C. Lee and S. Q. Chen, "Transformer encoder for one-step-ahead AQI forecasting using multivariate air-quality monitoring data," *Front. Environ. Sci.*, vol. 14, p. 1815865, May 2026, doi: 10.3389/FENVS.2026.1815865/TEXT.

[24] A. Bongale, S. Kumar, C. H. Kumar, P. Birwadkar, and M. Harish Kumar, "Neural space alignment for enhanced multi-horizon air quality time series forecasting using transformers," *Discov. Appl. Sci. 2026 87*, vol. 8, no. 7, pp. 731-, May 2026, doi: 10.1007/S42452-026-08771-2.

[25] "(PDF) A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short-term memory." Accessed: Sep. 15, 2026. [Online]. Available: https://www.researchgate.net/publication/330818076_A_hybrid_model_for_spatiotemporal_forecasting_of_PM25_based_on_graph_convolutional_neural_network_and_long_short-term_memory

[26] W. Kristjanpoller, A. Lopez, F. Mora, and M. C. Minutolo, "PM2.5 forecasting: A comparative study of transfer learning with transformer models and panel data approaches for country-level predictions," *Expert Syst. Appl.*, vol. 317, p. 131971, Jun. 2026, doi: 10.1016/J.ESWA.2026.131971.

[27] L. Zeng, R. Dong, M. Yuan, L. Jing, and S. Jiao, "Evaluating deep learning time series models for PM2.5 forecasting across diverse horizons," *iScience*, vol. 29, no. 2, p. 114770, Feb. 2026, doi: 10.1016/J.ISCI.2026.114770.

[28] "Reformer-Based Air Quality Forecasting Model - Google Scholar." Accessed: Sep. 15, 2026. [Online]. Available: https://scholar.google.com.pk/scholar?q=Reformer-Based+Air+Quality+Forecasting+Model&hl=en&as_sdt=0&as_vis=1&oi=scholart

[29] Z. Guo, S. Wu, M. Zhu, and H. Guandi, "Air Quality PM2.5 Index Prediction Model Based on CNN-LSTM," Aug. 2025, Accessed: Sep. 07, 2026. [Online]. Available: https://arxiv.org/pdf/2508.11215

[30] S. Pahari and S. C. Kumain, "Air Quality Prediction Using LOESS-ARIMA and Multi-Scale CNN-BiLSTM with Residual-Gated Attention," Oct. 2025, Accessed: Sep. 07, 2026. [Online]. Available: https://arxiv.org/pdf/2510.22818



Copyright © by authors and 50Sea. This work is licensed under Creative Commons Attribution 4.0 International License.