BREATHE: A GeoAI-Powered Air Quality Monitoring and Forecasting System for Urban Sustainability

Siva Durga Adduri[1], Tarig Ali[2]

[1] Department of Computer Science and Engineering, American University of Sharjah, Sharjah, United Arab Emirates - g00095118@aus.edu
[2] Department of Civil Engineering, American University of Sharjah, Sharjah United Arab Emirates - atarigali@aus.edu

Keywords: Air Quality Monitoring, Deep Learning, Time-Series Forecasting, Urban Sustainability, Disaster Risk Reduction.

Abstract

Air pollution is one of the most important environmental and public health challenges of our time, and it uniquely impacts urban areas, especially in the rapidly developing countries of the UAE. Traditional air quality monitoring systems lack the predictive capabilities needed for proactive intervention and sustainable urban planning. The research proposes BREATHE—a system of integrated real-time monitoring and machine learning-based forecasting to address air pollution challenges in urban environments by using GeoAI. The research implements deep learning algorithms alongside geographical data to establish a scalable system for air quality management. BREATHE system features four essential aspects that include (1) real-time AQI checks at 12 locations across the UAE territory (2) predictive models for AQI forecasting through climatic and historical data analyses (3) interactive dashboards with mapping visuals and alert features and (4) AI-powered chatbot assistance along with non-specialist user-friendly accessibility. The data processing together with model deployment operates without interruptions through Python-based automation. By bridging the gap between monitoring and predictive analytics, this study presents a replicable framework for large-scale air quality management in urban environments worldwide.

1. Introduction

Air pollution is a key challenge to urban sustainability, especially in rapidly developing countries such as the United Arab Emirates (UAE). The rapid urbanization and industrialization in these countries have caused air pollutants to increase exponentially, presenting serious concerns to public health, disaster risk management, and urban planning. Major air pollutants influencing urban air quality are particulate matter (PM10 and PM2.5), nitrogen dioxide (NO_2), sulphur dioxide (SO_2), ozone (O_3), and carbon monoxide (CO) (Chen et al., 2022); Akomolafe et al., 2024; Vitaliano, 2024). They are mainly discharged from vehicular traffic, industrial operations, and construction activities, compounded by the distinct climate of the UAE, which has the potential to confine pollutants near ground level (Jena et al., 2023; Patel, 2020).

The effects of air pollution on public health are significant, and it plays a role in respiratory disease, cardiovascular disease, as well as premature death (Yan, 2023; Ramírez, 2024). For example, research has indicated that heightened exposure to PM2.5 is correlated with increased cases of respiratory symptoms among children, indicating the sensitivity of certain populations (Ramírez, 2024). In addition to this, air pollution makes disaster risk management even more challenging as it adds to the extremity of health emergencies, during times of extreme weather conditions (such as heatwaves), which are prevalent in the UAE (Jena et al., 2023). Urban planning therefore needs to accommodate measures to cushion against air pollution, such as by expanding green areas and enhancing public transportation networks to lower the usage of private automobiles (Wang & Xu, 2024; Vitaliano, 2024).

Current air quality monitoring networks usually utilize static sensors at fixed points to record pollutant concentrations. Although such systems are useful, they possess fundamental limitations such as the absence of spatial coverage and the incapability of recording real-time air quality dynamics in different urban zones (Fattoruso et al., 2020). Further, most monitoring systems lack predictive features that are necessary for forecasting pollution events and taking early interventions (Bachechi et al., 2020; Deroubaix, 2024).

Predictive air quality monitoring is increasingly required, rather than just real-time observations. Predictive models are able to use historical data to predict air quality levels, with proactive action being taken before pollution levels become dangerous (Bachechi et al., 2020; Deroubaix, 2024). Such a transition to predictive monitoring is essential to urban sustainability, so that it enables improved resource allocation along with enhanced public health response. Additionally, the incorporation of cutting-edge technologies like machine learning and IoT has the potential to increase the precision and timeliness of air quality monitoring, which can lead to healthier cities (Guo et al., 2022; Correia et al., 2023; Fattoruso et al., 2020).

The BREATHE Air Quality Monitoring and Prediction system is specially built to overcome these significant limitations by incorporating real-time monitoring, GeoAI-enhanced forecasting, interactive dashboard, and AI-mediated chatbot. By enhancing user accessibility and carrying out a large-scale model validation comparing two current forecasting techniques, BREATHE provides a scalable and reproducible framework for urban air quality management. This study seeks to close the gap between surveillance and actionable information, ultimately informing more sustainable city planning, proactive public health interventions, and data-driven decision-making in fast-growing cities.

2. Literature Review

Air quality prediction has evolved over the years from simple statistical models to more advanced machine learning (ML) and deep learning methods, the evolution mirrors how air quality prediction progressed during different years.

Statistical methods such as ARIMA and Multiple Linear Regression were mainly used with early models to forecast air quality indices, gaining insight into the dynamics of air pollutants and their interactions with meteorological variables. This field was pioneered by Kumar & Jain (2009) who used ARIMA to forecast O_3, NO, NO_2, and CO highlighting the application effectiveness of time-series based models for depicting air quality trends. Likewise, Nimesh et al. (2014) claimed ARIMA and its derivatives: ARFIMA and Holt-Winters smoothing are useful for predicting air quality indices and concluded that historical data is essential for best results. These initial statistical models served as a basis to develop benchmark methods for air quality prediction studies.

While useful, these conventional statistical models found it challenging to model complex non-linear trends present in the air quality data. For instance, Liu et al. (2022) indicated that while regression-based models predict reasonably well, they are not very efficient in mapping complex relationships. This limitation pushed researchers to explore machine learning techniques. For instance, M. Liu et al. (2023) reported a hybrid model with Random Forest and Neural Networks to represent the dependency between meteorological conditions and air quality, and achieved a better predictive performance.

Yazdi et al. (2020) utilized an ensemble method of different machine learning techniques to predict PM2.5 in the Greater London Area. Their model achieved a determination coefficient (R^2) of 0.59, with moderate predictive competence (Yazdi et al., 2020). Zhang et al. (2022) presented a Temporal Difference-Based Graph Transformer Networks system as a new method for PM2.5 prediction in China. They investigated how deep learning could link with transfer learning procedures to develop advanced cross-city air quality prediction results for urban areas.

The authors W. Wang et al. (2021) developed a recursive model that used Convolutional Long Short-Term Memory (ConvLSTM) neural networks to successfully predict air quality concentrations in Beijing. Through its specific design the model acquired insights from both spatial patterns and temporal relations within the dataset which produced superior prediction results. Ong et al. (2015) investigated Dynamically Pre-trained Deep Recurrent Neural Networks (DRNN) for PM2.5 prediction because these networks demonstrated their ability to learn from environmental monitoring data. The research findings showed that DRNN surpassed traditional models because deep learning operates exceptionally well in time-series forecasting scenarios.

Focusing on the GCC region, particularly the UAE, a few studies have described challenges and suitable methods for air quality forecasting. Ramadan et al. (2024) conducted extensive investigations about air quality prediction in Abu Dhabi using ARIMA models which produced an RMSE of 2.16 $\mu g/m^3$. The authors underlined the necessity of combining real-time data to boost air quality management efficiency (Ramadan et al., 2024). W. Wang and Yang (2020) created a BP neural network that predicts UAE metropolitan air quality. The researchers demonstrated how their neural network prediction model utilized historical data to predict air quality changes along with pollutant concentrations successfully (W. Wang & Yang, 2020).

3. Materials and Methods

3.1. Study Area and Dataset

This study assesses air quality at different stations throughout the United Arab Emirates (UAE) which span metropolitan, industrial zones, rural expanses and desert domains. Monitoring stations throughout Abu Dhabi Emirate and Al Ain Region, along with Dubai Emirate cover the entire geographic regions of air quality conditions in the country (Refer to Figure 1). The established monitoring stations provide essential data that helps researchers understand pollution origins as well as air quality trends throughout time.

The multiple monitoring stations positioned across Abu Dhabi provide sophisticated insights into the different pollution systems of urban zones and industrial regions, along with coastal areas. The air quality data in Abu Dhabi capital city is tracked at the representative urban site located at the US Embassy Abu Dhabi City. The industrial setting of Al Mafraq has manufacturing and transportation activities defining its primary pollutant sources. The coastal station Bain Al Jessrain experiences both urban and marine atmospheric processes in addition to urban pollution. Sweihan operationally serves as a rural monitoring station that functions to help determine natural air quality standards. The western industrial area consists of Ruwais which experienced severe environmental impact due to oil refinery emissions as well as Habshan South which functions as an essential location in the oil and gas industry.

Al Ain provides distinct air quality information through its desert and urban locations because of its arid climate and minimal industrial development. The urban and residential monitoring station Al Tawia Al Ain operates alongside the Islamic Institute Al Ain that functions at the city center for measuring urban air pollution variations. The air quality conditions throughout Al Ain City are most accurately measured at Al Ain City station. Natural levels of particulate matter can be studied in Al Quaa because this desert zone contains low anthropogenic pollution. The air quality trends in a semi-urban setting are tracked by Zakher Al Ain.

US Embassy Dubai stands as a vital urban metrology site in Dubai because it is located in an area facing severe traffic congestion and construction activities and industrial pollution.

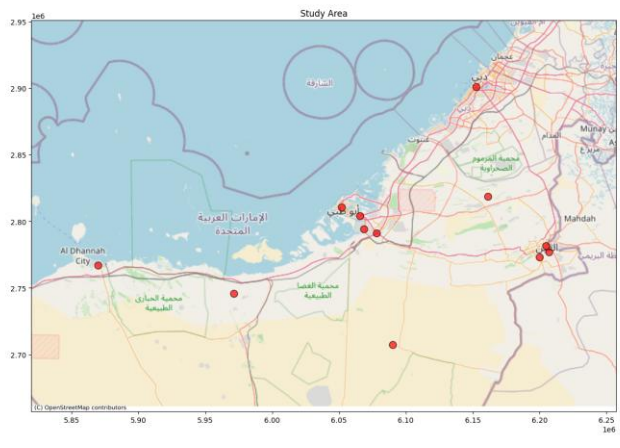

Figure 1. Study Area Locations on the UAE Map.

The air quality dataset used in this study is taken from the World Air Quality Index Project (aqicn.org/historical/), ensuring globally recognized data collection. The monitoring period varies by site from 2015 to 2025, offering a valid long-term observation of air pollution trends. Data provided include particulate matter ($PM_{2.5}$, PM_{10}), ozone (O_3), nitrogen dioxide (NO_2), and sulphur dioxide (SO_2), with the pollutant profile for each site as characterized in Table 1.

Emirate	Location Name	Date Range	Pollutants Monitored
Abu Dhabi	US Embassy Abu Dhabi City	2016-2025	PM2.5, PM10, O_3, NO_2, SO_2
	Al Mafraq	2024-2025	PM2.5, PM10, NO_2, SO_2
	Bain Al Jessrain	2016-2025	PM2.5, PM10, O_3, NO_2, SO_2
	Sweihan (School)	2026-2025	PM2.5, PM10, O_3, NO_2, SO_2
	Ruwais	2016-2025	PM2.5, PM10, O_3, NO_2, SO_2
	Habshan South	2016-2025	PM2.5, PM10, O_3, NO_2, SO_2
	Station Mussafah	2024-2025	PM2.5, PM10, O_3, NO_2, SO_2
Al Ain (Abu Dhabi)	Al Tawia Al Ain	2016-2025	PM2.5, PM10, O_3, NO_2, SO_2
	Islamic Institute Al Ain	2016-2025	PM2.5, PM10, O_3, NO_2, SO_2
	Al Quaa	2016-2025	PM2.5, PM10, O_3, NO_2, SO_2
	Zakher Al Ain	2016-2025	PM2.5, PM10, NO_2, SO_2
Dubai	Dubai US Embassy	2018-2025	PM2.5, O_3

Table 1. Study Area Locations and Information

3.2. Methods and Models Used

3.2.1. KNN Imputer: The K-Nearest Neighbours (KNN) imputer serves as a well-known technique for managing missing data across various domains that range from machine learning to bioinformatics. KNN imputation functions by relying on the data point similarities to estimate empty cell values. A KNN imputer locates 'k' nearest neighbours for target data points with missing entries by using distance metrics typically Euclidean distance to predict values by averaging neighbour data points as weights (Wei et al., 2018; Liao et al., 2014; Fadlil et al., 2022).

When working mathematically the KNN imputer calculates distances of data point X_j to every other data point inside the dataset. The K nearest neighbours $\mathrm{N_k}$ correspond to those points selected by the minimum distance computations. The calculated missing entry value gets computed through the following formula (Liao et al., 2014; Fadlil et al., 2022):

$$\widehat{X_J} = \frac{1}{k}\sum_{i=1}^{k} X_{N_i}, \tag{1}$$

where X_{N_i} are the observed nearest neighbour values

Choosing the 'k' value plays a vital role because it maintains equilibrium between bias and variance throughout the imputation process. The insufficient choice of 'k' leads to increased data variability, while selecting a larger 'k' value results in biased outcomes because of averaging dissimilar points (Gautam & Latifi, 2023; Magnussen & Tomppo, 2014). Studies indicate a k value of 5 stands out as the most appropriate choice for typical cases since it strikes an optimal balance between retrieving nearby zone knowledge and maintaining result stability (Gautam & Latifi, 2023).

3.2.2. GRU: Gated Recurrent Units (GRUs) represent a specific recurrent neural network architecture which addresses sequential data dependencies and overcomes conventional RNN’s drawbacks, like the vanishing gradient problem. GRU implementation surpasses LSTM complexity without compromising performance outcomes in text processing and time-series prediction, due to their less complex architecture (Choi et al., 2015; Nosouhian et al., 2021).

A GRU model contains two gate mechanisms which function as update gate z_t and reset gate r_t. At each time step t, the gates function to determine which information enters the hidden state h_t.

The flexible design of GRU allows it to identify different time-related correlations automatically therefore making it suitable for various sequential operations (Ravanelli et al., 2017; Nosouhian et al., 2021). The model exhibits efficient computing power and high performance thus it has gained wide adoption in deep learning research and applications (Choi et al., 2015; Nosouhian et al., 2021).

3.2.3. Transformer: In 2017, Vaswani et al. launched the Transformer model that completely eliminated recurrent architectures from NLP while introducing self-attention for parallel processing and better handling of sequential dependencies in text (Vaswani et al., 2017). The Transformer operates through all three essential components which include multistep attention processing and position-based encoding: alongside feed-forward neural networks to extract deep interrelations within the dataset.

Multi-Head Attention: Inside the model the multi-head attention mechanism allows parallel processing of various positional directions across the input sequence. The linear transformation distributes the information across multiple spaces where each space corresponds to its own "head." The basis of multi-head attention houses a mathematical formulation known as the scaled dot-product attention.

1. Input Representation: For an input sequence represented as matrices Q (queries), K (keys), and V (values), the attention scores are calculated by the dot product (Vaswani et al., 2017):

$$\mathrm{Attention}(Q, K, V) = \mathrm{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \tag{6}$$

where d_k is the keys’ dimension
and the softmax function normalizes the scores to a probability distribution.

2. Multi-Head Attention: The multi-head attention mechanism concatenates the results of several attention heads:

$$\mathrm{MultiHead}(Q, K, V) = \mathrm{Concat}(\mathrm{head}_1, \dots, \mathrm{head}_h)W^O, \tag{7}$$

where each head is computed as:

$$\mathrm{head}_i = \mathrm{Attention}\left(QW_i^Q, KW_i^K, VW_i^V\right), \tag{8}$$

where W_i^Q,W_i^K,W_i^V are learned weight matrices for each head
$W^O$is the output projection matrix

Positional Encoding: As the Transformer model architecture does not capture the sequence order of the input inherently, positional encoding is added to give information about the position of each token. Positional encoding is added to the input embeddings and is specified using sine and cosine functions (Zheng, 2021):

$$PE_{(pos,2i)} = \sin\left(\frac{pos}{10000^{2i/d_{model}}}\right), \quad (9)$$

$$PE_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{2i/d_{model}}}\right), \quad (10)$$

where pos is the position
i is the dimension index

This encoding enables the model to capture the relative positions of tokens, which is important for contextual understanding in sequences.

The Transformer contains two main components known as encoder and decoder stacks. Each encoder layer begins with a multi-head attention mechanism which applies feed-forward neural networks through normalization and residual connections throughout each process. The decoder contains the same structural design as the encoder yet adds an additional attention mechanism to consider the encoder output (Vaswani et al., 2017). The architecture provides efficient execution during training and inference stages so it can support various tasks beyond NLP such as image handling and time-series pattern analysis (Ahmed et al., 2022).

3.3. Data Preprocessing and Model Development

The process of data preprocessing and model development ensures that air quality predictions are accurate to changing environmental conditions. Before training begins the dataset (Figure 2) receives preprocessing operations that consist of KNN Imputer-based value imputation and Min-Max Scaling-based numeric feature normalization.

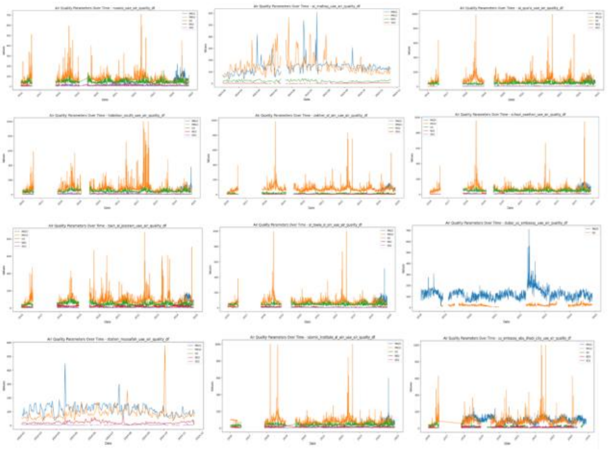

Figure 2. Initial Data collected from 12 stations across UAE

The research applies two deep learning approaches, including GRU-based and Transformer models for model development. The GRU model tracks extended dependencies across air quality sequences through its 64 units per layer section, which leads to the correlated output layer that predicts the next time step. The model implementation uses the Mean Squared Error (MSE) and Adam acts as the optimizer. Table 2 presents the detailed structure of the GRU model.

Layer Type	Units	Activation	Input Shape
GRU	64	tanh	(timesteps, features)
GRU	64	tanh	(timesteps, 64)
Dense	Features	Linear	(64,)

Table 2. GRU model architecture used

Self-attention functions within Transformers allows the model to dynamically find important time steps when making predictions. This framework incorporates dense embedding before implementing positional encoding, which helps maintain time sequence relations. Multi-head self-attention layers help detect distant relationships between elements while feedforward layers smooth features for better representation. The modelling application uses Mean Squared Error (MSE) loss together with Adam optimization. The Transformer model architecture can be found in Table 3.

Layer Type	Units	Activation	Input Shape
Input Layer	-	-	(timesteps, features)
Dense	128	Linear	(timesteps, features)
Positional Encoding	-	-	(timesteps, 128)
Multi-Head Attention	4 Heads	-	(timesteps, 128)
Dropout	-	-	(timesteps, 128)
Layer Normalization	-	-	(timesteps, 128)
Dense (Feed Forward)	128	ReLU	(timesteps, 128)
Dropout	-	-	(timesteps, 128)
Layer Normalization	-	-	(timesteps, 128)
Dense (Output)	Features	Linear	(128,)

Table 3. Transformer model architecture used

The preprocessing data pipeline delivers refined input data to the models. The GRU model suits sequential relationships, but the Transformer model is advantageous because of its self-attention capability to identify key time points dynamically. The output from both techniques generates multiple output variables which match the number of pollutants being estimated per location. The results of model comparison enable researchers to identify the most suitable forecasting approach for air quality assessment across the United Arab Emirates.

After training, the models are stored in Keras format for every location so inference can be performed without requiring any new retraining processes, unless mandatory. Nevertheless, model retraining becomes necessary for changing environmental conditions, and the process starts whenever concept drift is found. When a target variable's statistical attributes transform over time, it leads to a deterioration of model performance which we call concept drift. The processes of urbanization, traffic pattern modifications and regulatory changes make concept drift applicable for air quality modelling. Baier et al. (2020) state that performance tracking must continue with corresponding model retraining, for concept drift management strategies to work efficiently. Regular model assessment and training procedures start when we detect a substantial drift, to maintain accurate modelling of present-day air quality patterns.

3.4. System Architecture

The air quality forecasting system uses four essential modules which start with data acquisition followed by preprocessing and prediction along with output integration. The system retrieves data automatically through Selenium WebDriver and uses KNN to normalize and standardize it before data entry. Predictions from TensorFlow and Keras deep learning systems produce week-long estimates which the system transforms into usable results for external dashboards. Organizing the forecasting system in modules results in greater real-time capabilities as well as scalability and enhanced prediction accuracy.

3.4.1. Data Acquisition and Processing: The monitoring system operates through its built framework to automatically gather data for real-time processing and forecast evaluation leading to an updated air quality assessment. The system runs scheduled automation that combines efficient data processing with state-of-the-art predictive modelling to help decision-makers.

Selenium WebDriver controls automatic data extraction during the Automated Data Collection module to retrieve air quality data from web-based sources through timed operations. The system operates through time-based scheduling (once a week). While Selenium imitates browser functions, downloads the most recent data file and maintains them within a specified storage area. The data processing system employs dynamic file handling approaches to detect and handle new download files while preventing unnecessary repetition of processes.

A system referred to as Continuous Data Processing and Prediction functions through auto detection of new records to initiate real-time normalizing and cleaning of data with imputation steps. During preprocessing the module converts timestamps into a single datetime structure while applying KNN imputation techniques to replace missing values along with normalizing all column formats. The deep learning models receive processed data immediately after preprocessing completes so they can make a seven-day air quality forecast which prepares instant analysis opportunities for decision-making.

3.4.2. Dashboard and Alert System: The decision-making support mechanism integrates Historical and Forecast Data through a single dashboard that combines past trends with future analytical predictions. Historical trends can be compared with predicted AQI values, allowing for an evaluation of the likely effect of air pollution and encouraging the adoption of pre-emptive action where required.

The Alerting and Dashboard Integration mechanism facilitates the efficient distribution of real-time air quality analysis. The system arranges forecast and processed information in structure directories which enables visualization tools to retrieve them. The system has an automated warning procedure which organizes future air quality index (AQI) predictions into clearly defined templates from "Good" to "Hazardous" according to established health risk categories. Each alert group displays its respective information according to the AQI category to provide immediate health risk assessments.

The Inverse Distance Weighting (IDW) interpolation method predicts AQI values for points which do not receive direct monitoring station measurements. Through the search interface a user can specify a position, and the system calculates a forecasted AQI value from weighted sensor measurements of the nearest monitoring stations. Such methodology enhances air quality prediction resolution across territorial domains, so the entire area achieves accurate coverage.

3.4.3. AI-Powered Chatbot Integration: This platform uses an AI-trained chatbot interface which allows users to interact by asking questions about the air quality data and forecasting information. The chatbot system creates easy system-access by using natural language command queries which enhances accessibility to users of all backgrounds. Users receive proactive guidance and recommendations from the chatbot system beside its basic questioning capability. Progressively the AI system will locate major air quality incidents while giving users explanations about observed trends and prevention advice based on pollution level forecasts.

4. Results and Discussion

4.1. Model Performance: GRU vs. Transformer

The deep learning model performances for air quality prediction were evaluated by monitoring training and validation loss curves as well as Mean Absolute Error (MAE) from various air quality monitoring stations for both Gated Recurrent Units (GRU) and Transformer models. The main goal was to determine how well the models generalized to new unseen data and how well they captured temporal dependencies in concentrations.

4.1.1. GRU Model Performance: The validation loss outcome and the MAE for the GRU Model demonstrates a variable behaviour in air quality data. The training loss decreased steadily but the validation loss presented significant variations according to Figure 3.

Such oscillations in the performance suggest that the GRU model has excessive sensitivity to variation in air quality data which leads to overfitting during training while reducing its ability to maintain consistent results across different stations. The model demonstrates competency in capturing air quality dynamics through its low loss values even though its generalization capability remains limited.

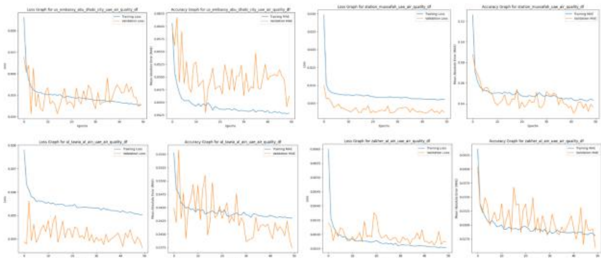

Figure 3. GRU Model’s training and validation loss and MAE

4.1.2. Transformer Model Performance: The Transformer model maintains higher stability when evaluating training along with validation loss. Generalization performance improves according to the validation MAE curves since Figure 4 shows significantly less fluctuations than the GRU model. The self-attention mechanism of Transformers enhances model performance by effectively capturing long air quality patterns and thus improves its noise-resilience properties.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-39-2025

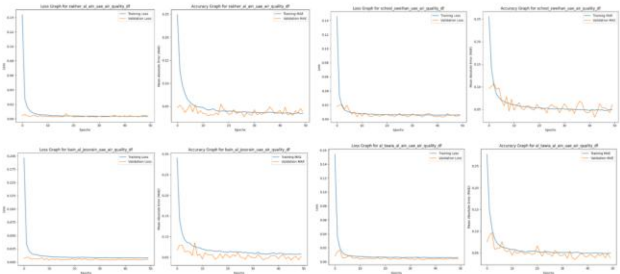

Figure 4. Transformer Model's training and validation loss and MAE

Station	Loss	MAE	Validation Loss	Validation MAE
Ruwais	0.0059	0.0495	0.0029	0.0398
Al Mafraq	0.0206	0.1027	0.0125	0.0982
Al Quaa	0.0057	0.049	0.0029	0.0328
Habshan	0.0043	0.0422	0.0059	0.0549
Zakher	0.0033	0.0352	0.0033	0.0381
School Sweihan	0.0061	0.0505	0.0057	0.0607
Bain Al Jessrain	0.0079	0.0574	0.0054	0.0519
Al Tawia	0.0063	0.0505	0.0037	0.0383
Dubai US Embassy	0.004	0.0416	0.009	0.0509
Station Mussafah	0.0113	0.075	0.005	0.0509
Islamic Institute Al Ain	0.0072	0.0544	0.0058	0.0575
US Embassy	0.0054	0.048	0.0041	0.0437
AVERAGE	**0.0073**	**0.055**	**0.0055**	**0.051**

Table 4. Numerical values of Training and Validation Loss and MAE for each station

4.1.3. Comparative Analysis and Implications: The Transformer produces smooth loss curves during training which demonstrates a more stable learning procedure thus making it the better choice for real-world air quality prediction applications. The GRU model produces high levels of instability which suggests that deployment issues could affect the quality of air quality warning predictions. The Transformer model should be chosen for practical applications needing consistent results instead of GRU.

4.2. Dashboard Functionality and User Experience

The air quality monitoring dashboard provides users with an interactive system featuring real-time as well as historical data analysis coupled with easy-to-use interfaces. The platform provides multiple integral features including real-time AQI reporting, air quality warning notifications, location-targeted searches, prediction projections and automated AI chat support functions. The system delivers better accessibility while supporting improved decision-making capabilities to general users combined with parties who work with air pollution and environmental health issues.

One of the key highlights of the dashboard is the real-time air quality monitoring and alert system, which enables users to see AQI values for various locations at a glance. The main dashboard gives a holistic overview of air pollution levels for several monitoring stations, offering insights into pollutant concentrations, as seen in Figure 5. Whenever a user clicks on a specific location, an alert pop-up is activated with more detailed information regarding pollutant concentrations and corresponding health implications. Further, a specialized "Air Quality Alerts" tab consolidates current alerts, allowing users to follow pollution episodes effectively and remain updated on dangerous air quality conditions.

The main interface displays complete air pollution information across multiple monitoring sites through figures like the one presented in Figure 5. A user can trigger an alert box that shows pollution measurements and their medical effects by selecting any particular location on the interface. Users can monitor pollution episodes more effectively through the "Air Quality Alerts" tab because it presents all current alerts in one location.

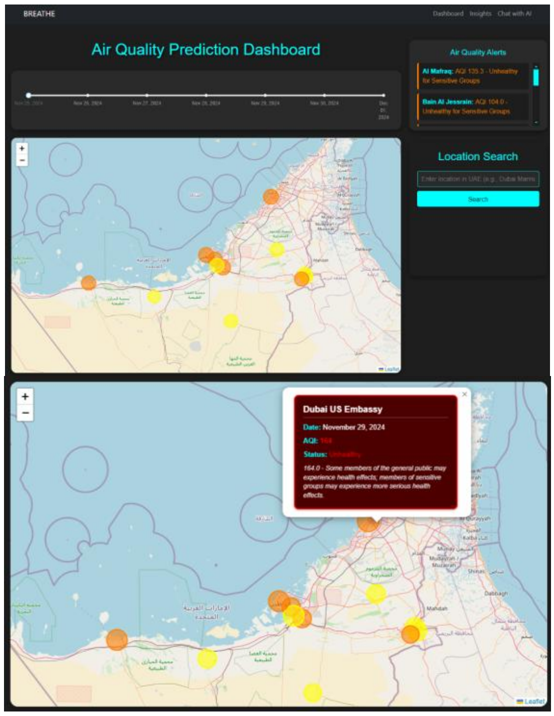

Figure 5: BREATHE Air Quality Prediction Dashboard's a) Home page b) Alert pop-up.

The dashboard's user experience is improved through its location-based search function (Figure 6). Users can benefit from this feature since it lets them type in any desired location to receive instantaneous updated readings about AQI and pollutant concentrations.

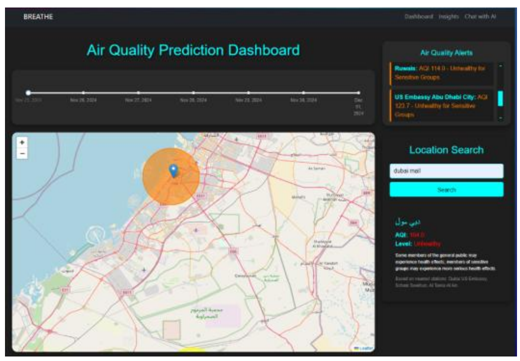

Figure 6: Location Search Option Demonstration.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-39-2025

The dashboard provides real-time monitoring capabilities through its "Insights" tab in Figure 7 which provides current and projected air quality information. User-selected air quality stations give them access to historical pollution data alongside model-based predictions of pollutant levels through deep learning algorithms. This capability serves researchers and policymakers and health practitioners who seek environmental planning evidence and public health intervention support based on data evidence.

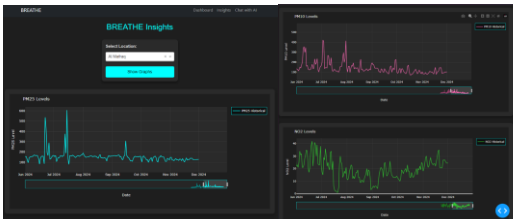

Figure 7: Historical and Predicted Values Graph Display.

The dashboard enhances user experience by adding an interactive chatbot accessible in Figure 8 which provides air quality information together with pollutant details as well as health-based advice.

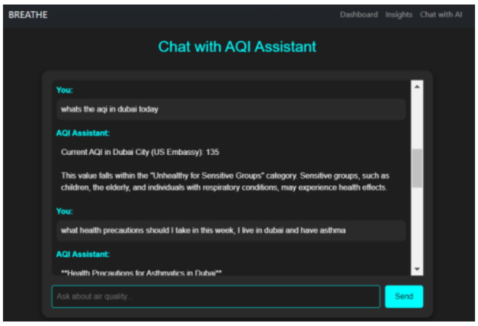

Figure 8: Chatbot Demonstration.

The dashboard serves as an interactive system that combines current environment monitoring with data visualizations of historical data alongside analysis predictions. Users experience a smooth interface because of diverse interactive elements which allows them to stay updated about air quality and make informed health-related decisions for outdoor activities. Future development ought to focus on implementing meteorological data and pollutant dispersion models into the system to develop more precise prediction capabilities for a comprehensive analysis of air pollution variations.

5. Conclusion

This research investigated air quality in the UAE across different monitoring stations ranging from urban to industrial stations and rural and desert sites. Air pollution patterns across the country were accurately depicted through the selection of monitoring stations which spanned throughout the three major regions Abu Dhabi Emirate and Al Ain Region and Dubai Emirate. The data was extracted from the World Air Quality Index Project, ranging from 2015 to 2025, allowing strong monitoring of air quality trends.

The research utilized KNN imputation as part of its predictive enhancements through the implementation of Gated Recurrent Units and Transformer models. The real-world air quality prediction system showed better stability and generalization from Transformer models making them the most dependable tool for air quality prediction. The developed air quality forecasting system contained four major components which consisted of data acquisition, preprocessing, prediction and output integration modules to deliver real-time monitoring, automated alerts, location-based search functionality, historical trend visualization and an AI-powered chatbot to users. These technological advancements enabled better decision-making opportunities for researchers together with policymakers and members of the public.

This research is of great use for air quality control, helping to implement preventive pollution reduction strategies. Future developments may include predictive model refinement, sensor network extension, and incorporation of other environmental parameters to increase forecast precision and user interaction even more.

References

Ahmed, S., Nielsen, I. E., Tripathi, A., Siddiqui, S., Rasool, G., & Ramachandran, R. P. (2022). *Transformers in Time-Series Analysis: A Tutorial*. https://doi.org/10.48550/arxiv.2205.01138

Akomolafe, O. O., Olorunsogo, T., Anyanwu, E. C., Osasona, F., Ogugua, J. O., & Daraojimba, O. H. (2024). Air Quality and Public Health: A Review of Urban Pollution Sources and Mitigation Measures. *Engineering Science & Technology Journal*, *5*(2), 259–271. https://doi.org/10.51594/estj.v5i2.751

Bachechi, C., Desimoni, F., Po, L., & Casas, D. M. (2020). *Visual Analytics for Spatio-Temporal Air Quality Data*. 460–466. https://doi.org/10.1109/iv51561.2020.00080

Baier, L., Reimold, J., & Kühl, N. (2020). *Handling Concept Drift for Predictions in Business Process Mining*. 76–83. https://doi.org/10.1109/cbi49978.2020.00016

Chen, Z., Liu, N., Tang, H., Gao, X., Zhang, Y., Kan, H., Deng, F., Zhao, B., Zeng, X., Sun, Y., Qian, H., Liu, W., Mo, J., Zheng, X., Huang, C., Sun, C., & Zhao, Z. (2022). Health Effects of Exposure to Sulfur Dioxide, Nitrogen Dioxide, Ozone, and Carbon Monoxide Between 1980 and 2019: A Systematic Review and Meta-analysis. *Indoor Air*, *32*(11). https://doi.org/10.1111/ina.13170

Choi, E., Bahadori, M. T., Schuetz, A., Stewart, W. F., & Sun, J. (2015). *Doctor AI: Predicting Clinical Events via Recurrent Neural Networks*. https://doi.org/10.48550/arxiv.1511.05942

Correia, C., Martins, V., Matroca, B., Santana, P., Mariano, P., Almeida, A., & Almeida, S. M. (2023). A Low-Cost Sensor System Installed in Buses to Monitor Air Quality in Cities. *International Journal of Environmental Research and Public Health*, *20*(5), 4073. https://doi.org/10.3390/ijerph20054073

Deroubaix, A. M. (2024). *Air Quality Forecasts With Observation-Based Scaling of Anthropogenic Emissions for Urban Agglomerations*. https://doi.org/10.22541/essoar.171352084.46269313/v1

Fadlil, A., Herman, H., & M, D. P. (2022). K Nearest Neighbor Imputation Performance on Missing Value Data Graduate User Satisfaction. *Jurnal Resti (Rekayasa Sistem Dan Teknologi*

Informasi), *6*(4), 570–576. https://doi.org/10.29207/resti.v6i4.4173

Fattoruso, G., Nocerino, M., Toscano, D., Pariota, L., Sorrentino, G. P., Manna, V., Vito, S. D., Cartenì, A., Fabbricino, M., & Francia, G. D. (2020). Site Suitability Analysis for Low Cost Sensor Networks for Urban Spatially Dense Air Pollution Monitoring. *Atmosphere*, *11*(11), 1215. https://doi.org/10.3390/atmos11111215

Gautam, R., & Latifi, S. (2023). Comparison of Simple Missing Data Imputation Techniques for Numerical and Categorical Datasets. *Journal of Research in Engineering and Applied Sciences*, *8*(1), 468–475. https://doi.org/10.46565/jreas.202381468-475

Guo, R., Qi, Y., Zhao, B., Pei, Z., Wen, F., Wu, S., & Zhang, Q. (2022). High-Resolution Urban Air Quality Mapping for Multiple Pollutants Based on Dense Monitoring Data and Machine Learning. *International Journal of Environmental Research and Public Health*, *19*(13), 8005. https://doi.org/10.3390/ijerph19138005

Jena, M. C., Mishra, S. K., & Moharana, H. S. (2023). Challenges and Way Forward to Maintain Air Quality Standard in Urban Areas. *The Global Environmental Engineers*, *10*, 33–43. https://doi.org/10.15377/2410-3624.2023.10.4

Kumar, U., & Jain, V. K. (2009). ARIMA Forecasting of Ambient Air Pollutants (O3, NO, NO2 and CO). *Stochastic Environmental Research and Risk Assessment*, *24*(5), 751–760. https://doi.org/10.1007/s00477-009-0361-8

Liao, S. G., Lin, Y., Kang, D., Chandra, D., Bon, J., Kaminski, N., Sciurba, F. C., & Tseng, G. C. (2014). Missing Value Imputation in High-Dimensional Phenomic Data: Imputable or Not, and How? *BMC Bioinformatics*, *15*(1). https://doi.org/10.1186/s12859-014-0346-6

Liu, M., Hu, H., Zhang, L., Zhang, Y., & Li, J. (2023). Construction of Air Quality Level Prediction Model Based on STEPDISC-PCA-BP. *Applied Sciences*, *13*(14), 8506. https://doi.org/10.3390/app13148506

Liu, X., Lü, D., Zhang, A., Liu, Q., & Jiang, G. (2022). Data-Driven Machine Learning in Environmental Pollution: Gains and Problems. *Environmental Science & Technology*, *56*(4), 2124–2133. https://doi.org/10.1021/acs.est.1c06157

Magnussen, S., & Tomppo, E. (2014). Thek-Nearest Neighbor Technique With Local Linear Regression. *Scandinavian Journal of Forest Research*, *29*(2), 120–131. https://doi.org/10.1080/02827581.2013.878744

Nimesh, R., Arora, S., Mahajan, K. K., & Gill, A. N. (2014). Predicting Air Quality Using ARIMA, ARFIMA and HW Smoothing. *Model Assisted Statistics and Applications*, *9*(2), 137–149. https://doi.org/10.3233/mas-130285

Nosouhian, S., Nosouhian, F., & Khoshouei, A. K. (2021). *A Review of Recurrent Neural Network Architecture for Sequence Learning: Comparison Between LSTM and GRU*. https://doi.org/10.20944/preprints202107.0252.v1

Ong, B. T., Sugiura, K., & Zettsu, K. (2015). Dynamically Pre-Trained Deep Recurrent Neural Networks Using Environmental Monitoring Data for Predicting PM2.5. *Neural Computing and Applications*, *27*(6), 1553–1566. https://doi.org/10.1007/s00521-015-1955-3

Patel, S. (2020). The Potential for Urban Vegetation to Mitigate Ambient Air Pollution Threats to Public Health. *Topophilia*, 53–62. https://doi.org/10.29173/topo28

Ramadan, M. S., Abuelgasim, A., & Al Hosani, N. (2024). Advancing air quality forecasting in Abu Dhabi, UAE using time series models. *Frontiers in Environmental Science*, *12*, 1393878. https://doi.org/10.3389/fenvs.2024.1393878

Ramírez, C. (2024). Impact of Air Quality on Respiratory Health in Children in Mexico. *Global Journal of Health Sciences*, *9*(2), 1–10. https://doi.org/10.47604/gjhs.2575

Ravanelli, M., Brakel, P., Omologo, M., & Bengio, Y. (2017). *Improving Speech Recognition by Revising Gated Recurrent Units*. https://doi.org/10.21437/interspeech.2017-775

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention Is All You Need*. https://doi.org/10.48550/arxiv.1706.03762

Vitaliano, S. (2024). Mitigating Built Environment Air Pollution by Green Systems: An in-Depth Review. *Applied Sciences*, *14*(15), 6487. https://doi.org/10.3390/app14156487

Wang, W., Mao, W., Tong, X., & Xu, G. (2021). A Novel Recursive Model Based on a Convolutional Long Short-Term Memory Neural Network for Air Pollution Prediction. *Remote Sensing*, *13*(7), 1284. https://doi.org/10.3390/rs13071284

Wang, W., & Yang, S. (2020). Analysis and Forecast of Urban Air Quality Based on BP Neural Network. *International Journal of Advanced Network Monitoring and Controls*, *5*(3), 57–64. https://doi.org/10.21307/ijanmc-2020-028

Wang, X., & Xu, B. (2024). *Evaluating the Impact of Urban Green Spaces on Air Quality in Beijing a Review of Studies From 2015 to 2023*. https://doi.org/10.31219/osf.io/74ua9

Wei, R., Wang, J., Su, M., Jia, E., Chen, S., Chen, T., & Ni, Y. (2018). Missing Value Imputation Approach for Mass Spectrometry-Based Metabolomics Data. *Scientific Reports*, *8*(1). https://doi.org/10.1038/s41598-017-19120-0

Yan, W. (2023). Effect of PM2.5 Air Pollution on the Incidence of Respiratory Diseases: A Python-Based Data Analysis. *Theoretical and Natural Science*, *8*(1), 70–75. https://doi.org/10.54254/2753-8818/8/20240361

Yazdi, M. D., Kuang, Z., Dimakopoulou, K., Barratt, B., Süel, E., Amini, H., Lyapustin, A., Katsouyanni, K., & Schwartz, J. (2020). Predicting Fine Particulate Matter (PM2.5) in the Greater London Area: An Ensemble Approach Using Machine Learning Methods. *Remote Sensing*, *12*(6), 914. https://doi.org/10.3390/rs12060914

Zhang, Z., Zhang, S., Zhao, X., Chen, L., & Ye, J. (2022). Temporal Difference-Based Graph Transformer Networks for Air Quality PM2.5 Prediction: A Case Study in China. *Frontiers in Environmental Science*, *10*. https://doi.org/10.3389/fenvs.2022.924986

Zheng, J. (2021). *Rethinking Positional Encoding*. https://doi.org/10.48550/arxiv.2107.02561

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-G-2025-39-2025