Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://asr.nsps.org.ng/index.php/asr

African Scientific Reports 5 (2026) 598

African Scientific Reports

A comparative study of deep learning architectures for particulate matter prediction in Lugbe, FCT Abuja, Nigeria

Amarachukwu Enumah[a,*], Abdullahi Mohammed Evuti[a], Kamoru Adio Salam[a], Yahaya Sayyadi Mohammad[b], Samuel Shamaye Silas[c], Babatunde Abdulraheem Atoyebi[d]

[a]*Chemical Engineering Department, University of Abuja, Abuja, Nigeria*
[b]*Department of Food Science and Technology, Ibrahim Badamasi Babangida University, Lapai, Nigeria*
[c]*Chemical Engineering Department, University of Calabar, Cross River State, Nigeria*
[d]*Department of Microbiology, University of Calabar, Calabar, Cross River State, Nigeria*

Abstract

This study developed and evaluated three deep learning architectures: a recurrent neural network (RNN), a long short-term memory (LSTM) network and a convolutional neural network–bidirectional LSTM (CNN–BiLSTM) for predicting fine ($PM_{2.5}$) and coarse (PM_{10}) particulate matter concentrations. The models used 2-h (60-observation) sliding-window inputs and air quality and meteorological data comprising 1,211,810 entries. Hyperparameters were optimized using GridSearchCV with time-series cross-validation. Performance was assessed using mean absolute error (MAE), root mean square error (RMSE), relative absolute error (RAE), mean absolute percentage error (MAPE) and the coefficient of determination (R^2). All three models demonstrated strong $PM_{2.5}$ predictive capability ($R^2 > 0.90$), explaining more than 90% of the observed variability. Model 2 showed the most balanced performance, with MAE of 2.898 $\mu g/m^3$, RAE of 0.1697, RMSE of 6.453 $\mu g/m^3$, MAPE of 7.7013% and R^2 of 0.9166; Model 3 achieved a marginally higher R^2 of 0.9188. Model 2 was considered the most reliable $PM_{2.5}$ estimator because its error metrics were generally lower than those of Model 3. Prediction was weaker for PM_{10} across all architectures (R^2 = 0.7037–0.7348), reflecting greater temporal variability and complexity. For PM_{10}, Model 2 achieved the lowest RMSE (12.7097 $\mu g/m^3$) and highest R^2 (0.7348), indicating superior generalization. The results highlight the effectiveness of hybrid deep learning approaches for air quality forecasting, particularly for fine particulate matter.

DOI: 10.46481/asr.2026.5.3.598

Keywords: Deep learning, Air quality, Neural networks, Prediction, Particulate matter

Article History:
Received: 02 July 2026
Received in revised form: 16 September 2026
Accepted for publication: 08 October 2026
Available online: 09 November 2026

© 2026 The Author(s). Published by the Nigerian Society of Physical Sciences under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

1. Introduction

Air pollution is a major global problem and organizations such as the World Health Organization (WHO), the African Union (AU) and national agencies such as the National Environmental Standards Regulation and Enforcement Agency (NESREA) are actively investigating its health effects. These organizations, with varying degrees of influence, have advocated and enacted legislation to keep air quality within certain limits and prioritize public health. The WHO reports that approximately 4.2 million deaths are due to

*Corresponding Author Tel. No.: +234-805-884-2507.
Email address: enumahcorp@gmail.com (Amarachukwu Enumah)

air pollution-related diseases each year, prompting a review of pollution guidelines in 2016 [1, 2]. According to the report, more than 90% of the world's population breathes air that does not comply with WHO standards. In response, in 2016 WHO member states developed an action plan for an enhanced global approach to combating the health effects of air pollution. The United Nations (UN) then set goals to reduce global air pollution by 2030 [2, 3]

In 2019, Nigeria adopted a National Action Plan to Reduce Short-Term Climate Pollutants (SLCP) with the aim of reducing black carbon by 83% by 2030. The aim of the plan is to reduce various types of air pollution and envisage a 22% reduction in air pollution exposure and prevent about 7000 premature deaths by 2030 [4]. A major threat is air pollution, which significantly increases mortality and morbidity. Small particles ($PM_{2.5}$, PM_{10}) are particularly dangerous, causing 2.9 million premature deaths worldwide in 2017, which is approximately 9% of all deaths [5]. Nigeria, ranked 4th globally for poor air quality and 150th out of 180 countries in the Environmental Impact Index, is in a critical situation, especially in Abuja, FCT [6, 7]. Despite being one of the fastest growing megacities in the world, FCT Abuja has limited air quality monitoring stations and relies on sparse, short-term data [8].

As there are still concerns about air pollution in FCT Abuja, there is currently no credible air pollution impact assessment or comprehensive air pollution control plan. To solve this critical problem, it is necessary to develop a comprehensive air quality monitoring, forecasting and modeling system. Therefore, the purpose of the study is to monitor, and predict $PM_{2.5}$ and PM_{10} using three Deep learning Model architecture.

Several studies have shown that conventional statistical approaches are less accurate than AI technologies for forecasting [9, 10]. Artificial Neural Networks (ANN) stand out due to their ability to handle large datasets, identify complex non-linear relationships and detect unusual interactions. While traditional AI methods outperform statistical approaches, hybrid models perform better [11]. Researchers often use ANN or hybrid models, such as Artificial-Intelligence System (AIS) and fuzzy neural network (FNN) for accuracy and simplicity of design. Although choosing the best deep learning method is difficult, hybrid models can increase predictive performance by improving stability, reliability, and accuracy of air pollution predictions. Hybrid models, despite their importance, increase computational costs due to complexity and long processing time [11]. Table 1 summarizes deep learning (DL) and machine learning (ML) methods and their reported performance in air pollution prediction.

This study advances particulate matter forecasting in Nigeria by comparatively evaluating three deep learning architectures (RNN, LSTM, and CNN–BiLSTM) thereby providing a comprehensive assessment of their predictive capabilities beyond the conventional machine learning approaches predominantly reported in the Nigerian literature. Previous research focuses on $PM_{2.5}$, this study simultaneously predicts $PM_{2.5}$ and PM_{10}, enabling a more comprehensive evaluation of particulate matter dynamics. The models were developed using 2-min resolution air quality data collected over a 13-month period and implemented within a time-series-aware modelling framework employing chronological training, validation, and testing datasets together with time-series-aware model development and hyperparameter optimization, thereby preserving temporal dependencies, preventing information leakage, and improving the robustness and reliability of the forecasting results.

Table 1: Approaches to deep learning and machine learning for air pollution prediction.

Study	Data / pollutant	Model / comparator	Performance	Findings
1: 2017	AQ, M; Daily; $PM_{2.5}$, PM_{10}; Abruzzo region, Italy	RNN / ANN, MLR	R^2: 0.8855	The study shows correctly 95% forecasts of $PM_{2.5}$, where is not measured, using PM_{10} concentrations, as input parameter in the model. Adding CO concentration improves the outcome of the MLR model, and also the increase the performances of the neural network models. [12]
2: 2020	M and Historical AQ data from 384 China monitoring station.; Daily; $PM_{2.5}$; Beijing, China	CNN-LSTM. / LSTM, MLP	RMSE: 3.855; MAE: 3.007; MAPE: 0.027	The authors concluded that a mutual information estimator combined with a CNN-LSTM model will predict a more accurate daily average concentration of $PM_{2.5}$. [13]
3: 2021	$PM_{2.5}$ concentration data sets; Hourly (24h); $PM_{2.5}$ (7days); Beijing, China	Hybrid CNNLSTM / Bi-LSTM, GRU, Bi-GRU, CNN	RMSE: 16.625; MAE: 9.034; R^2: 0.979	The Hybrid model captures both temporal and spatial features using CNN and LSTM, offering superior accuracy and stability. It delivers more precise predictions compared to the traditional models mentioned. [14]

Table 1: Approaches to deep learning and machine learning for air pollution prediction (continued).

Study	Data / pollutant	Model / comparator	Performance	Findings
4: 2021	AQ , $PM_{2.5}$; Hourly; $PM_{2.5}$; Beijing, China	EMD-BiLSTM / BiLSTM, LSTM	RMSE: 6.86; MAE: 4.92; MAPE: 10.66; R^2: 0.989	The EMD-BiLSTM model simplifies structure and reduces computational demands by addressing hidden temporal factors with minimal auxiliary data, achieving accurate and stable $PM_{2.5}$ predict. [15]
5: 2021	AQ, M; hourly; PM2.5	PM10; south Korea	CNN; RMSE: 15.70; R^2: 0.97	The findings of the study revealed that the coefficient correlation (R^2) for $PM_{2.5}$ and PM_{10} CNN models is almost similar, which is about 0.9, indicating the validity of the model [16].
6: 2023	M, AQ; Daily; $PM_{2.5}$; China (Beijing, Lanzhou, Jiaozuo and Guangzhou)	ARIMA-DBO-CNN-LSTM; CEEMDAN-CNN-LSTM; CEEMDAN-LSTM / ARIMA, LSTM, SVM	RMSE: 7.594; 14.94; 7.841; 5.496; MAE: 5.285; 10.94; 5.12; 3.77; R^2: 0.989; 0.962; 0.953; 0.953	Overall, ARIMA-DBO-CNN-LSTM model has higher prediction accuracy for the four cities and better adaptability. [17]
7: 2023	AQ, M; Daily; $PM_{2.5}$; Johannesburg, South Africa	CBR,	RMSE: 25.72	The study concluded that a hybrid model is more accurate than a sole model. [18]
7: 2023	AQ, M; Daily; $PM_{2.5}$; Johannesburg, South Africa	XGBR	RMSE: 27.64	The study concluded that a hybrid model is more accurate than a sole model. [18]
8: 2023	M, AQI; Hourly; $PM_{2.5}$, PM_{10}; Nigeria, Port Harcourt	BiLSTM	RMSE: 7.28; MAE: 3.4; R^2: 0.97	The study concluded that BILSTM model forecast AQI seamless. [19]
9: 2024	AQ, M; Hourly; SO2, CO, $PM_{2.5}$, PM_{10}, NO2, O3; Jinan, China	LSTM / RF, ARIMA, WEIGHTED, LR, BP, LGBM	RMSE: 16.5248; MAE: 11.2485; R^2: 0.9046	The LSTM model demonstrated to be the most effective, representing high degree of goodness-of-fit. It achieved above 91.37% and 90.46% prediction, for AQI, O3 prediction, and a perfect fit for forecasting concentrations of primary pollutants in the test set. [20]
10: 2024	Indoor Air Quality; Daily; $PM_{2.5}$, PM_{10},; Northern China	BP-ANN, MLNN, LSTM	R^2: Between 0.78 and 0.81	MLNN model predicted $PM_{2.5}$ and PM_{10} concentrations in an office building more accurately. It has the advantages of high accuracy, real-time monitoring, and multi-source data integration. [21]
11: 2025	M, AQ; Daily; $PM_{2.5}$ (7days); Rwanda, Kigali	LSTM; CBR	RMSE: 0.66 ; 0.61; Normalized value; R^2: 0.501; 0.582	CatBoost achieved the best performance (R2=0.582; RMSE= 0.61; MSE=0.43) over RF, GB, LSTMAdaBost, Ridge, Lasso ElasticNet. [22]
12: 2025	M, AQ; 2 min; $PM_{2.5}$; Nigeria, Kogi state Anyigba	RF	RMSE: 3.55; MAE: 2.25; R^2: 0.89	The study reported that RF outperformed SVR, GB, LR and MLR. [23]
12: 2025	M, AQ; 2 min; $PM_{2.5}$; Benin City	MLR	RMSE: 12.43; R^2: 0.68	The study reported that RF outperformed SVR, GB, LR and MLR. [23]

Table 1: Approaches to deep learning and machine learning for air pollution prediction (continued).

Study	Data / pollutant	Model / comparator	Performance	Findings
13: 2025	M and Historical AQ data (PM_1, $PM_{2.5}$, PM_{10}) from 1 May 2021–30 April 2023; 2 min; $PM_{2.5}$; Lagos, Nigeria	RF, GB, SVM	RMSE: 10.84; 11.89; 22.89; MAE: 6.62; 17.79; 7.73; R^2: 0.77; 0.72; 0.24	The authors concluded that Random Forest model achieved the best predictive performance, explaining 77% of the variance in $PM_{2.5}$ levels (R^2 = 0.77, RMSE = 10.84). the outperformance is as a RF's ability to model complex, nonlinear relationships through ensemble average. [24]

2. Materials and methods

2.1. Description of the study area

Abuja Municipal Area Council (AMAC) is the urban centre of the Federal Capital Territory (FCT) of Nigeria. It is located between latitude 8°40'N and 9°20'N of the equator and longitude 6°40'E and 7°40'E of the Greenwich meridian. The Federal Capital City (FCC) is situated within AMAC and it is the largest and most developed of the six area councils. The major infrastructures and developmental projects are concentrated within AMAC. It contains the bulk of the built-up areas, designed into phases and districts. The main districts are Asokoro, Maitama, Garki, Wuse, Kado, Jabi, Utako, Gwarinpa I, Gwarinpa II, Central Area and other newly developed districts including Gaduwa, Mabushi, Gudu, Lokogoma, Lugbe, Nbora, Kaura, Kagini, Karsana, Karmo, Guzape, Katampe, Kafe, Karmo, Jahi, Galadimawa, Durumi, Dape, etc. have shown recent developmental growth in the FCC. The FCT has a land mass of approximately 8,000 km^2, of which AMAC occupies about 250 km^2 with a population of 778,298 people [25]. This population was projected to 2,368,558 in 2016 [25, 26].

The available information extrapolated from adjacent weather stations, showed that the study area has temperatures ranging from 21°C–26.7°C yearly and a total annual rainfall of approximately 1,650 mm. About 60% of the annual rains fall during the month of July, August and September [27]. The temperatures are higher during the dry season months, when the mean minimum and maximum temperatures are 30°C and 35°C. During the rainy seasons on the other hand, the mean temperature ranges between 21°C and 33°C [27]. A crucial climatic characteristic of this area is the frequent occurrence of squall lines heralded by thunderstorms, lightning, strong winds and rainfall of high intensity [28].

Figure 1is a map of FCT Abuja showing Lugbe (coordinates: 8. 9909° N, 7. 3551° E) a fast-growing suburban area along the Airport Road in Nigeria's Federal Capital Territory (FCT) [8]. Lugbe has expanded greatly because of its proximity to important infrastructure. The area has residential estates (e.g., River Park Estate, Trademore Estate, FHA Estate), commercial areas, and public buildings including the National Space Research and Development Agency (NASRDA). The area's appearance has changed due to ongoing construction, improved road infrastructure, and growing economic activity. Its rapid urbanization has presented difficulties, including traffic congestion, air pollution, insufficient waste management, and deficits in public services [8].

2.2. Data collection

Air quality data was collected using a PurpleAir sensor installed at Lugbe, Abuja, Nigeria, along Airport Road (coordinates: 8.9909° N, 7.3551° E). The sensor was mounted at a height of 1.85 m above ground level to minimize ground-level interference while ensuring temporal measurements of ambient air quality. The monitoring device has dimensions of 85 × 85 × 125mm and is powered via a 5V Micro-USB supply, requiring 0.18A during continuous operation and 600 mA at peak load. Wireless data transmission is supported through an ESP-WROOM-02D module operating on IEEE 802.11b/g/n (2.4GHz) networks with an integrated 2dBi PCB antenna. The device complies with FCC, CE, TELEC, and SRRC certification standards [29].

Particulate matter (PM) concentrations were measured using dual Plantower PMS5003 laser-scattering sensors, which provide redundant measurements to improve data reliability and enable continuous assessment of sensor performance. The sensors detect particles with aerodynamic diameters ranging from 0.3 to 10μm, with counting efficiencies of 50% at 0.3μm and 98% for particles ≥0.5μm. The effective measurement range for $PM_{2.5}$ is 0–500$\mu g/m^3$, with a maximum measurable concentration of ≥1000$\mu g/m^3$. The sensors sample a standard air volume of 0.1L, with a single response time of ≤1second(s) and an overall response time of ≤10s.

Meteorological parameters, including ambient temperature, relative humidity, and atmospheric pressure, were simultaneously measured using a Bosch BME280 sensor. The sensor operates over temperature and pressure ranges of -40 to 85°C and 300–1100hPa, respectively. Relative humidity measurements have an accuracy of ±3% RH, a hysteresis of ≤2% RH, and a response time of approximately 1s. The PurpleAir monitor incorporates factory-calibrated Plantower optical particle sensors, and no post-manufacturing recalibration of the sensing elements was performed [29].

The dataset comprises meteorological data (Temperature; pressure; relative humidity, wind speed) three particulate matter indicators: PM_1, $PM_{2.5}$ and PM_{10}. Data were continuously monitored over a thirteen-month period from 1 January 2024 to 31 January

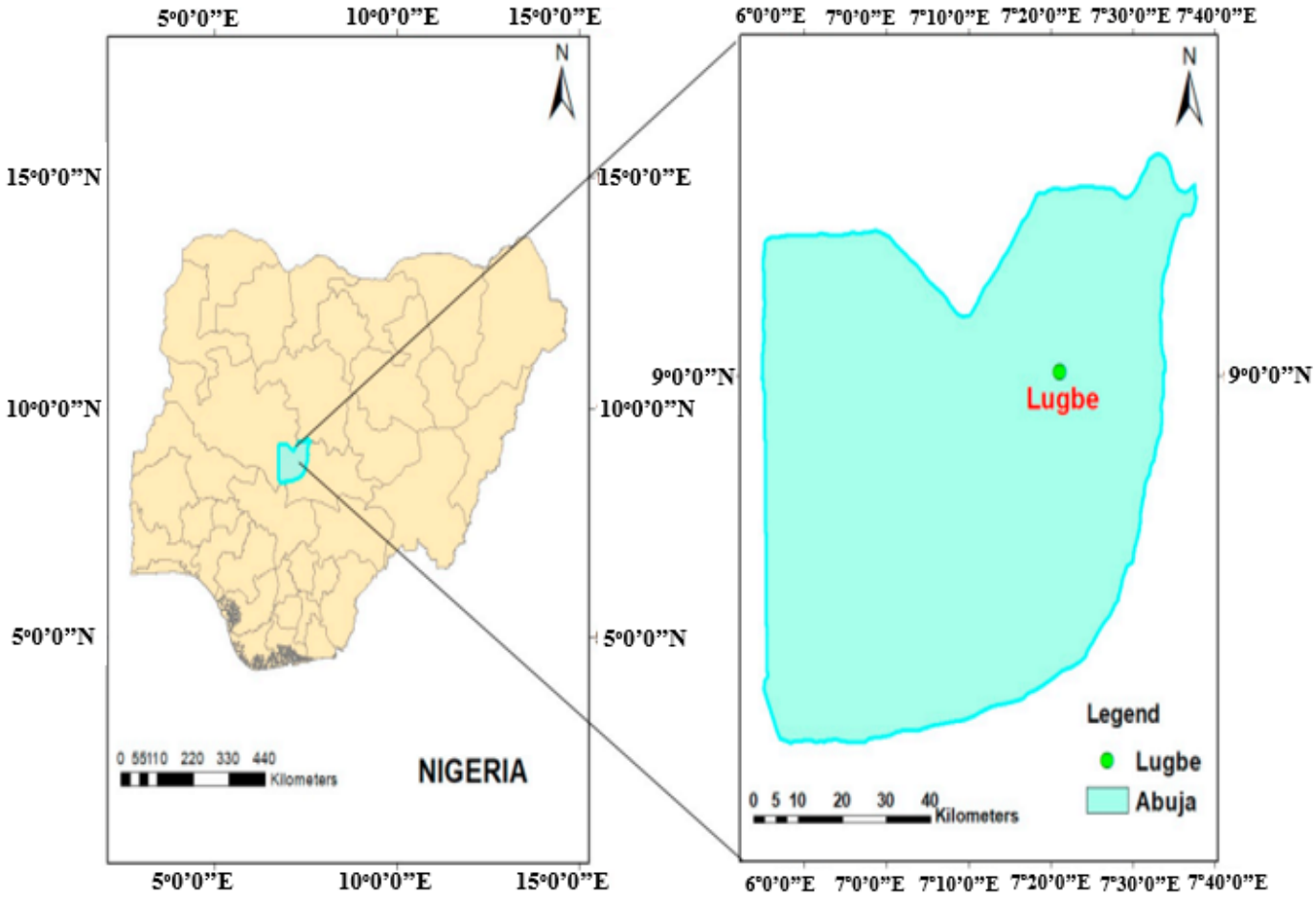


Figure 1: Map of Nigeria showing the study location [8].

2025 at an interval of 2 min to ensure sufficient temporal coverage. After data acquisition, the dataset was preprocessed, including feature engineering, treatment of missing values, and outlier detection and removal to improve data quality and model reliability. The final processed dataset comprised 1,211,810 observations across all variables. The particulate matter variables (PM_1, $PM_{2.5}$ and PM_{10}) and timestamps each contained 165,350 observations, while the meteorological variables of temperature, relative humidity, and atmospheric pressure contained 163,471 observations after quality control. Wind speed contained 59,997 observations because it was obtained from an external source with a lower sampling frequency and intermittent data availability. Consequently, only synchronized observations with complete records were retained for model development and evaluation.

2.3. Training, testing and validation of the dataset

This study partitioned the available dataset into three subsets: training, validation, and testing datasets. The training dataset was utilized for the development of predictive models by letting the algorithms learn the inherent patterns and relationships between the input variables and target variable. A separate validation dataset was employed during the development process to fine-tune model architecture and optimize hyperparameters, so that overfitting could be prevented and generalization capability improved. Lastly, the test dataset represents data that is absolutely unseen during both training and validation, and this is used for unbiased evaluation of the model in terms of predictive performance. Such three-stage partitioning enhances the robustness of the developed models, helping them to generalize well and perform reliably on new unseen data.

The preparation of the training, validation, and testing datasets for deep learning models was split 70 %–15 %–15 %, with 70 % of the data for training, 15 % for validation, and the remaining 15 % for testing. the training set consisted of observations from 1 January 2024 to 31 October 2024, the validation set included observations from 1 November 2024 to 31 December 2024, and the testing set comprised observations from 1 January 2025 to 31 January 2025. This chronological partitioning ensured that the models were trained using historical data and evaluated on previously unseen future observations, thereby providing a realistic assessment of predictive performance.

To eliminate data leakage, all preprocessing steps, including feature engineering and feature scaling, were performed using the training dataset only. The parameters required for preprocessing, such as scaling statistics, were estimated exclusively from the training data and subsequently applied unchanged to the validation and testing datasets. This approach ensured that no information from the validation or testing sets influenced model training or feature transformation, thereby preserving the integrity of model development and evaluation.

2.4. Models

To efficiently analyze the temporal variability of particulate matter concentration data ($PM_{2.5}$, PM_{10}), the input data sequences for training the models were created using a sliding window technique. Each input sequence contained past 60 observations, which meant a lookback period of 2h due to observations were made with an interval of 2min. The target variable was $PM_{2.5}$ and PM_{10} level at time t+1 constituting one-step-ahead forecasting with a 2-min prediction horizon. A stride of 1 was used, which means that the sliding window moves forward by one observation every time. In order to maintain time dependency and avoid leaking of information, input sequences were formed separately for each of training, validation, and test datasets after chronological splitting of the data. Three deep learning architectures were used in this study. They include: Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM) network and a Convolutional Neural Network-Bidirectional Long Short-Term Memory (CNN-BiLSTM). All these three models were selected due to their ability to address problems with environmental data that are both sequentially dependent as well as nonlinear.

The Recurrent Neural Network (RNN) is a basic type of Deep Learning model that is tailored for the processing of sequential data. Unlike conventional Feed Forward Neural Networks, the RNN model uses feedback connections that enable the influence of previous time steps to impact the predictions [13].This enhances the RNN model for time series forecasting techniques like air quality forecasting, where the air quality at a particular time depends on the past air quality patterns. Eqs. (1) and (2) define how the RNN updates its memory and computes its output in each time step [30]. Nonetheless, the RNN model was faced with the vanishing or exploding gradient problem [13].

$$h_t = \tanh(W^{(hf)} x_t + W^{(hx)} h_t + x_h), \tag{1}$$

$$R_t = W_{hy} h_t + x_y, \tag{2}$$

where $W^{(hf)}$, $W^{(hx)}$ and W_{hy} are weights; h_{t-1} is the previous hidden state, x_h and x_y are bias terms; and tanh is the activation function that maps values to the interval $[-1, 1]$.

To mitigate these challenges, a Long Short-Term Memory (LSTM) network was applied in the proposed model. The LSTM network is a particular RNN designed with memory cells, together with input gates, forget gates, and output gates (Eqs. (3)–(7)) [31], controlling the information flow [13]. The strengths of the system lie in the capability of the LSTM model not only in designing a complex decision system that interprets the AQ data, taking into consideration the complex long-term dependencies of the data, especially when it contains seasonal trends, daily variability, and sudden changes caused by weather systems and human activities, respectively [12, 17].

$$i_t = a_{\text{sigmoid}}(W^{(xi)} x_t + W^{(hi)} h_{t-1} + b_i), \tag{3}$$

$$f_t = a_{\text{sigmoid}}(W^{(xf)} x_t + W^{(hf)} h_{t-1}) + b_z, \tag{4}$$

where $W^{(xf)}$ and $W^{(hf)}$ are network weights, h_{t-1} is the previous state, b_z is a bias term, and a_{sigmoid} is a sigmoid activation function. Eqs. (1) and (2) are combined to yield C_t as follows:

$$C_t = f_t * C_{t-1} + i_t * a_{\tanh}(W^{(xC)} x_t + W^{(hC)} h_{t-1} b_C). \tag{5}$$

The final output of the LSTM model is calculated from the hidden state as follows:

$$O_t = a_{\tanh}(W^{(xo)} x_t + W^{(ho)} h_{t-1} + b_o), \tag{6}$$

$$h_t = O_t * a_{\tanh}(C_t). \tag{7}$$

The Convolutional Neural Network-Bidirectional Long Short-Term Memory (CNN-BiLSTM) is an integrated deep learning model consisting of a Convolutional Neural Network (CNN) and a Bidirectional Long Short Term Memory (BiLSTM). In the CNN, the convolution layer extracts local characteristics and spatial relationships from the sequence input data [15]. On the other hand, the BiLSTM layer is used to extract temporal dependencies in both directions [14]. Therefore, due to its unique ability to extract both temporal and spatial relations, this model is best suited for sequence modelling tasks.

By comparing the three deep learning architectures, the analysis will enable the selection of the best model capable of making predictions for the desired outcomes within the study area. Figure 2 shows the flow chart for the architecture of the models.

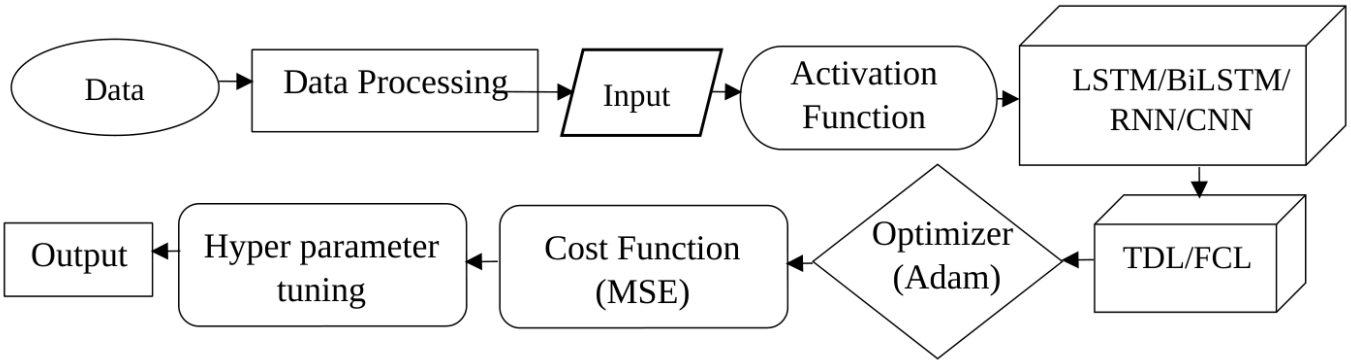


Figure 2: Deep learning architecture.

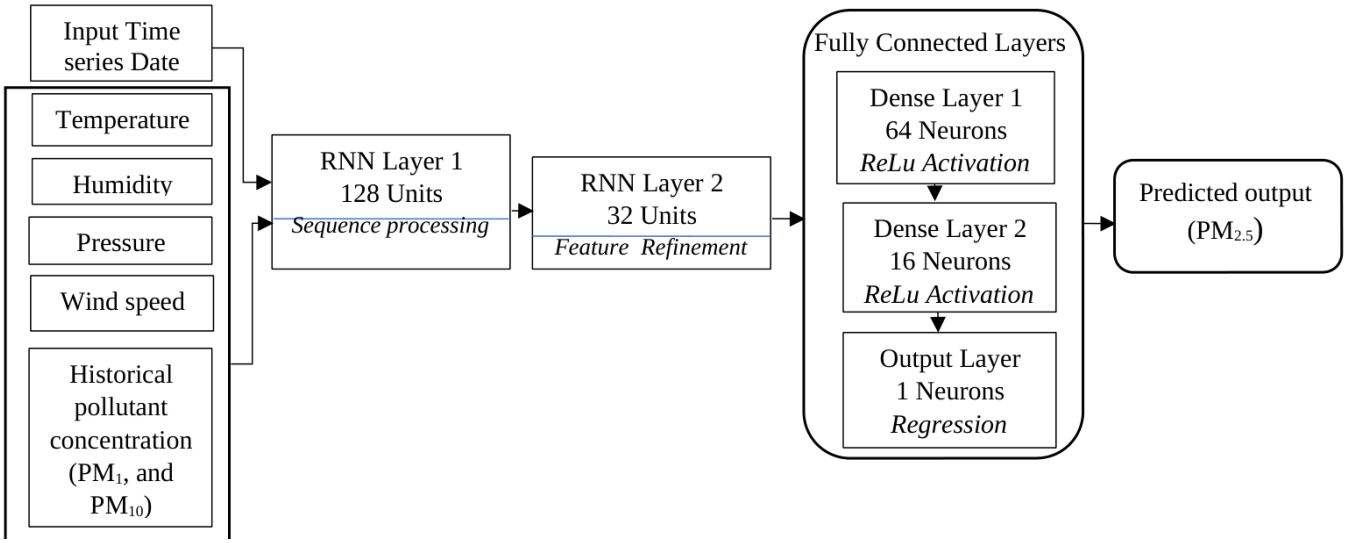


Figure 3: Overview of the architecture of Model 1.

2.5. Model architecture

2.5.1. Model 1: RNN (128), RNN (32) and FCL (64, 16, 1)

The model is composed of a stacked Recurrent Neural Network (RNN) structure architecture with fully connected layers (FCL), designed for time series forecasting of particulate matter concentration. This network is made up of two RNN layers having 128 and 32 units in their respective hidden layers. This is followed by three dense layers with 64, 16, and 1 neuron. In this architecture, the RNN stack is used for extracting dependencies within the data, and then a fully connected layer were used to map the extracted features to the desired output.

The first RNN layer with 128 units extracts complex temporal features in air quality data, while the second layer with 32 units continues to enhance temporal features while performing dimensionality reduction to address overfitting issues. The next stage involves using a feed-forward neural network with three fully connected layers, where the first two layers have 64 and 16 neurons, respectively, utilizing a nonlinear activation function (ReLU). The last layer features one neuron with a linear activation function to generate air quality values of $PM_{2.5}$, or PM_{10} as continuums in their continuous prediction output. Figure 3 illustrate the overview of the architecture the model.

2.5.2. Model 2: LSTM (128), TDL (8), LSTM (64) and FCL (32, 16, 1)

Model 2 combines Long Short-Term Memory (LSTM) networks and a Time Distributed Layer (TDL) were created to analyze particulate matter concentration over thirteen (13) months. The deep learning network starts with a layer of LSTM with 128 hidden units, a TDL layer using eight neurons, a second layer of LSTM with 64 hidden units, and a fully connected network using 32, 16, and 1 neuron(s), respectively.

the architecture is in figure 4.

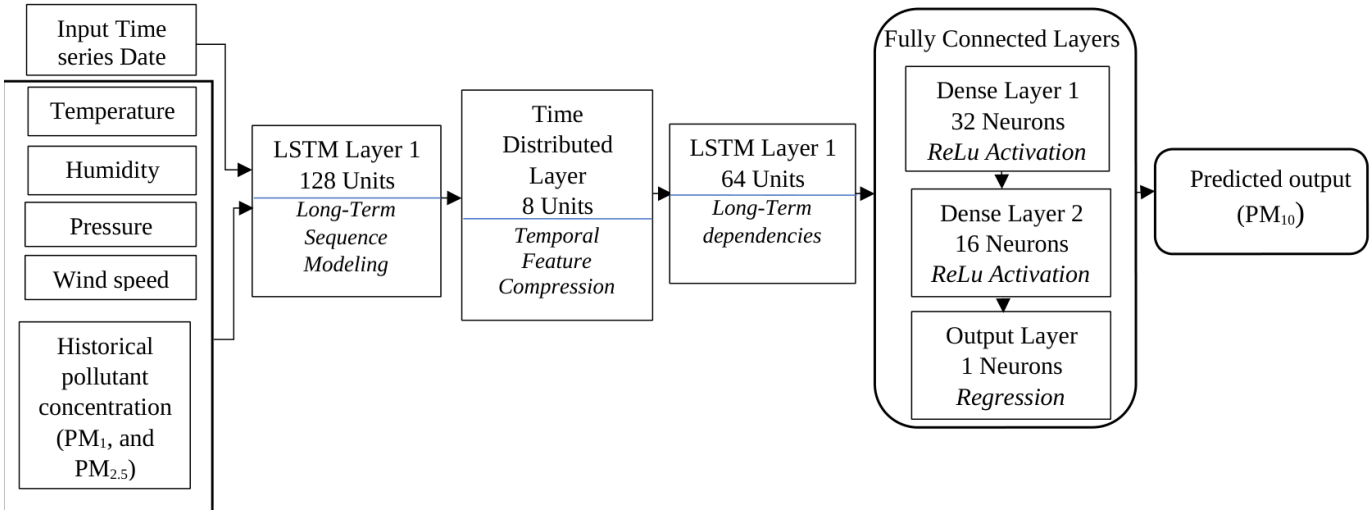


Figure 4: Overview of the architecture of Model 2.

The first LSTM layer was used for extracting long-term temporal dependencies from the input time series data for air quality, containing daily patterns, seasons, as well as complex relationships between future and past concentrations for different pollutants. The output from this layer is a sequence feature representation with temporal structure.

In order to reduce the dimensionality of features while still retaining the temporal information, a Time Distributed Layer with eight neurons was applied at each time step to the output of the LSTM layer. The Time Distributed Layer, a point-wise layer applied to the temporal data maintained the sequential characteristics of the data while efficiently reducing features by simply filtering out noises in the data.

The transformed sequence was passed through a second LSTM layer with 64 hidden units that further refine temporal representations by integrating information across multiple time steps. The final hidden state of this layer can be interpreted as a compact feature embedding summarizing the complex dynamics underlying particulate matter variability.

For the final prediction, the temporal features learned were mapped using a fully connected network with two hidden layers, each containing 32 and 16 neurons, respectively, using ReLU for activation. The output layer was made up of a single neuron that uses the linear activation function to provide a continuous estimate of the concentration of $PM_{2.5}$, or PM_{10}. The architecture is shown in Figure 4.

2.5.3. Model 3: CNN (64 filters), max pooling (2), BiLSTM (64) and FCL (32, 16, 1)

To model the temporal variability of particulate matter concentrations, a hybrid model involving the structure of a one-dimensional Convolutional Neural Network, Bidirectional LSTM networks, dropout regularization, as well as the multilayer fully connected structure were used. It starts with a convolutional layer that is one dimensional and comprises 64 filters with a kernel size of three. This layer is responsible for local feature extraction of the input sequence along its temporal axis and helps detect short-term variations and localized patterns of particulate matter concentrations. This convolutional output is then fed into a max-pooling layer with a pooling size of two. This helps reduce the temporal dimensions of the feature maps while minimizing computational complexity.

The combined features are then passed to a Bidirectional LSTM layer with 64 units in it. Unlike normal LSTMs, the Bidirectional LSTMs consider the input features in two ways: in the forward direction and in the backward direction. Each input sequence was explicitly built using only historical observations that precede the target timestamp. For a prediction at time *t*, the input window includes observations from *t-60* to *t-1*. The BiLSTM then goes through the historical input in reverse order during the backward pass of its computations based on past values only. This construction prevents any future information from leaking into the prediction, making the model deployable in operational settings.

For preventing overfitting and enabling generalization, there is a dropout layer with dropout rate of 0.2 applied to the result of the BiLSTM layer. It randomly sets 20% of the neurons to zero. Then, this feature extraction result is fed into a fully connected network comprising two hidden layers with 32 and 16 neurons, respectively, and rectified linear unit activation function. Lastly, there is an output layer with a single neuron and linear activation function for predicting the concentrations of $PM_{2.5}$, or PM_{10}.

The architecture of Model 3 is shown in Figure 5.

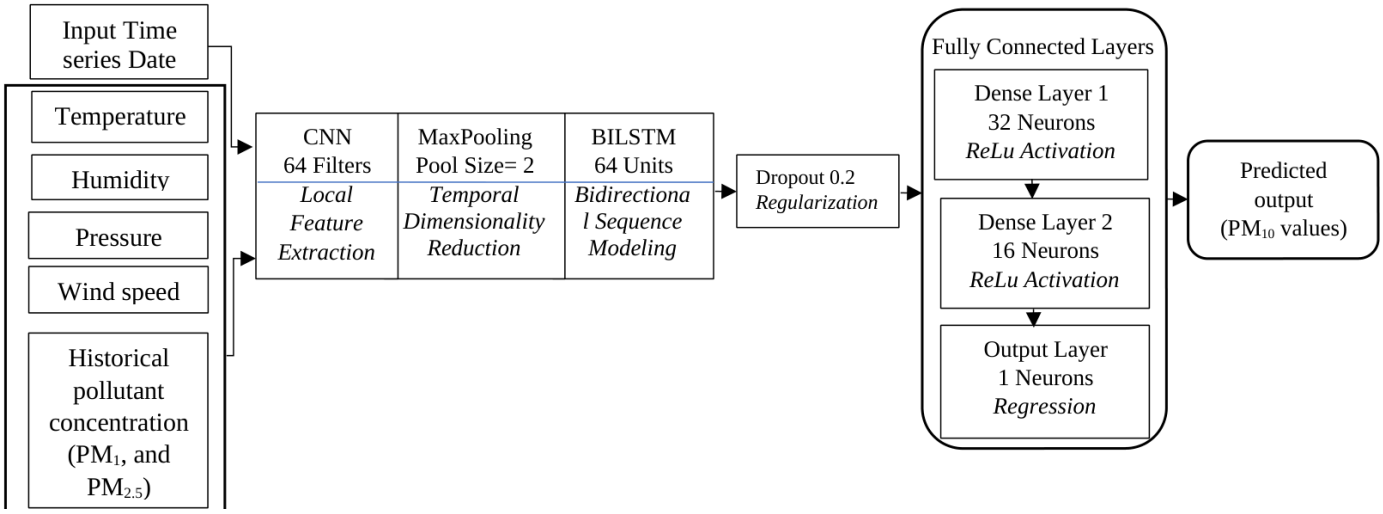


Figure 5: Overview of the architecture of Model 3.

2.6. Justification

Three different architecture were developed to predict the PMs concentrations in FCT Abuja. Techniques such as supervised deep learning models including RNN, LSTM, BiLSTM, CNN were employed. The hyperparameters of the deep learning models were optimized using a grid search implemented with GridSearchCV in conjunction with TimeSeriesSplit cross-validation. Unlike conventional shuffled k-fold cross-validation, TimeSeriesSplit preserves the chronological order of observations by ensuring that each validation fold contains only observations occurring after the corresponding training fold, thereby preventing information leakage from future data into the training process [17].

The hyperparameter search combined number of hidden units (32, 64, 128), batch size (32, 64, 128), learning rate (0.001, 0.0005, 0.0001), dropout rate (0.2), optimizer (Adam), and number of training epochs (50, 100, and 150). Hyperparameter selection was based on the lowest Root Mean Square Error (RMSE) obtained across the validation folds. The optimal hyperparameter configuration identified by the grid search was subsequently used to train the final RNN, LSTM, and CNN–BiLSTM models.

GridSearchCV was employed solely for hyperparameter optimization and was not used for feature selection. Feature engineering, including the construction of time-series input sequences, was performed before model training, while preprocessing operations requiring parameter estimation, such as feature scaling, were fitted exclusively on the training dataset and subsequently applied unchanged to the validation and testing datasets. Following hyperparameter optimization, the final models were retrained using the complete training dataset and their predictive performance was evaluated on the independent test dataset, which was not involved in either model training or hyperparameter tuning.

The selection of DL models is based on relevance to the air quality prediction, interpretability, and wide usage. Each of these models offers distinct advantages. The choice of RNN, CNN, LSTM, and BiLSTM models was based on previous research studies related to air quality prediction, where the performance of these models had been exceptionally good. These models can efficiently handle non-linear relations present in the weather and pollutant data.

Additionally, RNN is good for modelling sequential and time-dependent data while CNN is efficient at extracting local patterns and reducing noise through features. LSTM overcomes vanishing gradient problem and capture long-term dependencies, and resistant to overfitting. BiLSTM learn from both past and future context, improve understanding of temporal relationship.

These models were chosen for comparison based on different deep learning strategies to test their ability as predictive tools for air quality forecasting. On the whole, RNN is extremely effective for time-series analysis, CNN and BiLSTM framework combines local feature learning and bidirectional temporal information effectively, offering a potent approach to high-resolution air quality modelling while LSTM-TDL-LSTM framework effectively balances temporal feature extraction, dimensionality reduction, and non-linear regression.

2.7. Evaluation metrics

In the air quality research, the models were implemented in Python 3.10.16 using TensorFlow 2.15.0 (Keras API). The data preprocessing and feature engineering were performed using Pandas 2.2.2 and NumPy 1.26.4, respectively, while the hyperparameter tuning and time-series cross-validation were performed using Scikit-learn 1.4.2. Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Relative Absolute Error (RAE), Coefficient of Determination (R^2), and Mean Absolute Percentage Error (MAPE), were employed to test how well predicted and actual concentrations of PM are related. Before calculating MAPE, the dataset was

examined for zero values. No zero valuesd observations were present after data preprocessing; therefore, no additional correction for division by zero was required. The evaluation metrics were calculated by using Eqs. (8)–(12) [23–27].

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n}|x_i - \widehat{x_i}|, \tag{8}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \widehat{x_i})^2}{N}}, \tag{9}$$

$$\mathrm{RAE} = \frac{\sum_{i=1}^{n}|x_i - \widehat{x_i}|}{\sum_{i=1}^{n}|x_i - \overline{x}|}, \tag{10}$$

$$\mathrm{MAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{x_i - \widehat{x_i}}{x_i}\right|, \tag{11}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(x_i - \widehat{x_i})^2}{\sum_{i=1}^{n}(x_i - \overline{x})^2}, \tag{12}$$

where n (or N) is the number of evaluated observations, x_i is the observed value for the ith data point, $\widehat{x_i}$ is the model prediction for that data point, $\overline{x}$ is the mean of the observed values, and $\sum$ denotes summation.

Diebold-Mariano test (DM) was employed to determine the accuracy of the models. The test is designed to evaluate the accuracy of forecast provided by different models. Eqs. (13)–(17) define how DM is computed [32].

$$\mathrm{DM} = \frac{\overline{d}}{\sqrt{\widehat{\Upsilon}_0/n}}, \tag{13}$$

where $\overline{d}$ is the mean loss differential, defined as

$$\overline{d} = \frac{1}{n}\sum_{i=1}^{n}\mathbb{d}_i, \tag{14}$$

$$\mathbb{d}_i = |e_i| - |r_i|, \tag{15}$$

where n is the total number of forecasts; $\mathbb{d}_i$ denotes the loss differential; and $|e_i|$ and $|r_i|$ denote losses from the two competing models at time i.

$$\widehat{\Upsilon}_0 = \Upsilon_0 + 2\sum_{k=1}^{h-1}\Upsilon_k, \tag{16}$$

where $\widehat{\Upsilon}_0$ estimates the long-run variance of the loss differential, Υ_0 is its sample variance, h is the forecast horizon, and Υ_k is the autocovariance at lag k. The latter is defined as

$$\Upsilon_k = \frac{1}{n}\sum_{i=k+1}^{n}(\mathbb{d}_i - \overline{d})(\mathbb{d}_{i-k} - \overline{d}). \tag{17}$$

The null and alternative hypotheses are $H_0 = E(\mathbb{d}_i) = 0$ and $H_1 = E(\mathbb{d}_i) \neq 0$, respectively.

Under the null hypothesis, the two models have equal predictive accuracy. Rejection of indicate that the forecasting accuracy of the two models differs statistically significantly [32].

3. Results and discussion

3.1. $PM_{2.5}$ and PM_{10} performance evaluation

The performance of the three predictive models for $PM_{2.5}$ and PM_{10} concentrations is compared based on the five standard statistical metrics: Mean Absolute Error (MAE), Relative Absolute Error (RAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and the coefficient of determination (R^2). Prior to model training, the predictor variables were normalized using Min–Max scaling to improve numerical stability and accelerate model convergence. The normalized values were computed by equation 18.

$$x_{\mathrm{norm}} = \frac{x - x_{\mathrm{min}}}{x_{\mathrm{max}} - x_{\mathrm{min}}}, \tag{18}$$

Table 2: Performance evaluation metrics of the three models.

	$PM_{2.5}$					PM_{10}				
Model	MAE	RAE	RMSE	MAPE	R^2	MAE	RAE	RMSE	MAPE	R^2
Model 1	3.0415	0.1785	6.741	7.8326	0.9089	7.3013	0.4007	13.453	15.2874	0.7037
Model 2	2.898	0.1697	6.453	7.7013	0.9166	7.5718	0.4153	12.7097	17.1473	0.7348
Model 3	2.980	0.1748	6.371	8.1625	0.9188	7.6168	0.4175	13.1604	16.5556	0.7158

MAE and RMSE are in $\mu g/m^3$; MAPE is in percent.

Table 3: Diebold–Mariano test for differences in prediction accuracy.

Model comparison	DM statistic	p-value	Decision ($\alpha = 0.05$)	Interpretation
$PM_{2.5}$				
Model 1 vs ; Model 2	16.395	<0.001	Reject H0	Prediction accuracies differ significantly
Model 1; Vs; Model 3	17.817	<0.001	Reject H0	Prediction accuracies differ significantly
Model 2; vs; Model 3	13.707	<0.001	Reject H0	Prediction accuracies differ significantly
PM_{10}				
Model 1 vs ; Model 2	12.20	<0.001	Reject H0	Prediction accuracies differ significantly
Model 1; Vs; Model 3	0.768	0.442	Fail to Reject H0	No significant prediction difference
Model 2; vs; Model 3	-9.264	<0.001	Reject H0	Prediction accuracies differ significantly

where x is the original feature value and $x_{\min}$ and $x_{\max}$ denote the minimum and maximum values estimated from the training dataset only. The same transformation parameters were subsequently applied to the validation and testing datasets to prevent information leakage.

Following model prediction, both the predicted and reference PM concentrations were inverse transformed to their original scale before performance evaluation. Consequently, the reported MAE and RMSE values are expressed in micrograms per cubic metre ($\mu g/m^3$), while RAE and the coefficient of determination (R^2) were computed using the inverse-transformed concentrations. Reporting the evaluation metrics in the original physical units facilitates direct interpretation of prediction errors and comparison with regulatory air quality standards.

All three models achieve a strong predictive ability in $PM_{2.5}$ estimation, which shows an R^2 greater than 0.90 to reflect that over 90% of the variation in $PM_{2.5}$ concentrations is captured by the models. This aligns with the study of Ref. [16] in South Korea, although their CNN model recorded R^2 of 97.6% while this study CNN hybrid model has an R^2 of 91.88%. However, Ref. [14] reported R^2 of 97.9 in their classical study in Beijing China using Hybrid CNNLSTM. From Table 2, Model 2 is the most balanced and has the best performance for the lowest MAE of 2.898 $\mu g/m^3$, RAE of 0.1697, RMSE of 6.453 $\mu g/m^3$, and MAPE of 7.7013 along with an R^2 as high as 0.9166. Model 3 recorded an R^2 a little higher at 0.9188, but it is relatively less precise in prediction than Model 2 owing to a relatively higher value of error metrics. Model 2, therefore, is the most reliable for the estimation of $PM_{2.5}$.

On the other hand, The performance of the models for PM_{10} forecasting varied across the evaluation metrics used to assess the models. While Model 1 had the lowest MAE (7.3013 $\mu g/m^3$) and MAPE (15.2874) and hence was the best in terms of average and relative prediction errors, Model 2 had the lowest RMSE (12.7097 $\mu g/m^3$) and the highest R^2 (0.7348), which indicates good accuracy and agreement between predicted and observed concentrations. Although Model 1 shows a small error of MAE (7.3013 $\mu g/m^3$), the corresponding smaller R^2 (0.7037) value. This implies that the performance of the models is dependent on the criterion used, and there was no particular model that performed better than the other models in all the criteria used.

The improvement in forecast accuracy obtained in $PM_{2.5}$ as compared to PM_{10} can mainly be attributed to the better temporal characteristics and autocorrelation of fine particulate matter, which enable DL models in learning temporal patterns. In contrast, PM_{10} is influenced by both fine and coarse matter where the coarse fraction is significantly dependent on sporadic sources like dust from roads, construction sites, and soil blown by the wind, leading to higher variability and prediction uncertainty. Additionally, meteorological parameters considered in this study have more impact on $PM_{2.5}$ formation and transport as opposed to PM_{10}. Therefore, the models obtained higher forecasting accuracy for $PM_{2.5}$ consistent with previous air quality predictions studies.

3.2. Particulate matter prediction

From the three graphs (Figure 6), All models exhibit fluctuations in their values as the data samples increase across time. Descriptively, the observed and predicted values overlap in all graphs. Table 4 shows a sample of measured and predicted values for all the models.

The Diebold–Mariano test results are summarized in Table 3.

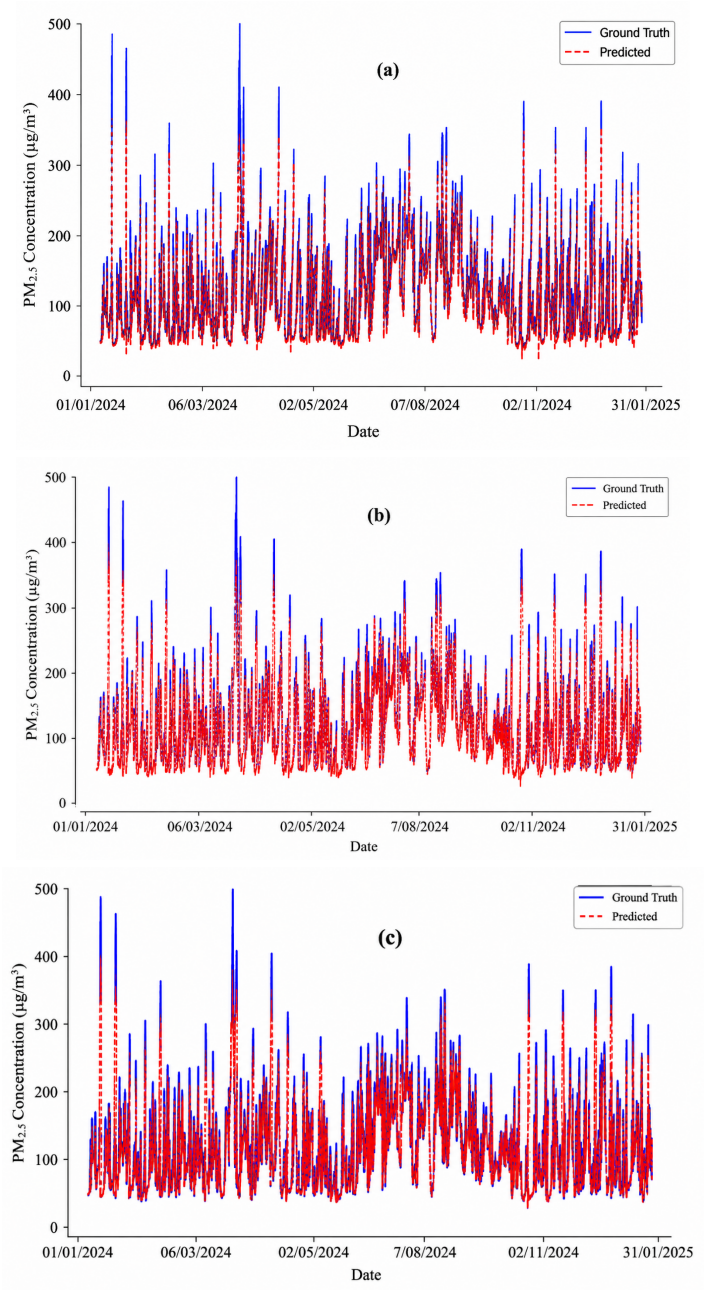


Figure 6: Observed and predicted $PM_{2.5}$ values for Model 1 (panel a), Model 2 (panel b) and Model 3 (panel c).

Table 4: Measured and predicted particulate matter concentrations.

Timestamp	$PM_{2.5}$		PM_{10}	
	Measured	Predicted	Measured	Predicted
Model 1: measured and predicted concentrations ($\mu g/m^3$)				
2024-07-06 04:36:11 +00:00	45.475	63.391	60.321	53.62
2024-07-06 04:38:11 +00:00	45.275	64.630	58.520	54.247
2024-07-06 04:40:10 +00:00	45.409	65.453	58.890	54.202
2024-07-06 04:42:11 +00:00	45.761	61.736	60.410	52.515
2024-07-06 04:44:11 +00:00	46.130	59.656	62.211	52.265
2025-01-30 23:55:51 +00:00	84.775	87.961	85.769	86.671
2025-01-30 23:57:51 +00:00	80.999	79.569	84.630	80.397
2025-01-30 23:59:51 +00:00	74.205	75.967	78.620	76.642

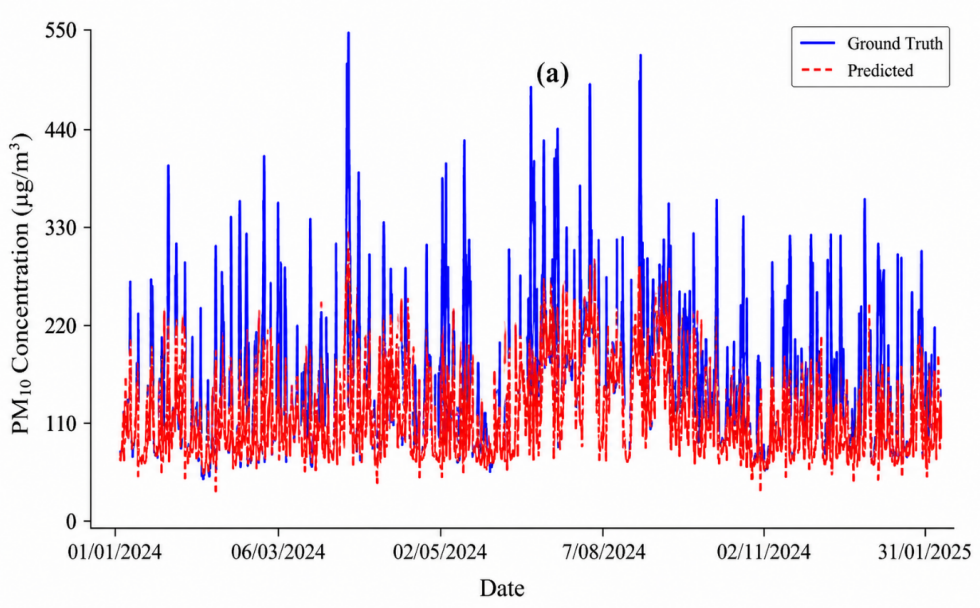


Figure 7: Observed and predicted PM_{10} values for Model 1 (panel a).

Table 4: Measured and predicted particulate matter concentrations (continued).

Timestamp	$PM_{2.5}$		PM_{10}	
	Measured	Predicted	Measured	Predicted
Model 3 (continued)				
Model 2: measured and predicted concentrations ($\mu g/m^3$)				
2024-07-06 04:36:11 +00:00	45.475	60.364	60.321	56.785
2024-07-06 04:38:11 +00:00	45.275	61.031	58.520	57.081
2024-07-06 04:40:10 +00:00	45.409	60.645	58.890	56.897
2024-07-06 04:42:11 +00:00	45.761	62.239	60.410	58.424
2024-07-06 04:44:11 +00:00	46.130	63.413	62.211	59.397
2025-01-30 23:55:51 +00:00	84.775	85.380	85.769	85.797
2025-01-30 23:57:51 +00:00	80.899	77.892	84.630	78.893
2025-01-30 23:59:51 +00:00	74.205	75.146	78.620	75.200
Model 3: measured and predicted concentrations ($\mu g/m^3$)				
2024-07-06 04:36:11 +00:00	45.475	62.944	60.321	58.743
2024-07-06 04:38:11 +00:00	45.275	63.787	58.520	58.861
2024-07-06 04:40:10 +00:00	45.409	63.151	58.890	58.772
2024-07-06 04:42:11 +00:00	45.761	61.925	60.410	57.917
2024-07-06 04:44:11 +00:00	46.130	61.695	62.211	57.581
2025-01-30 23:55:51 +00:00	84.775	82.266	85.769	91.524
2025-01-30 23:57:51 +00:00	80.999	74.786	84.630	83.020
2025-01-30 23:59:51 +00:00	74.205	72.489	78.620	80.622

From Table 3 The pairwise comparison results for $PM_{2.5}$ show statistically significant differences in predictive accuracy ($p < 0.001$). This is because the Diebold-Mariano test (DM) results of the comparison between Model 1 and Model 2 (16.395), Model 1 and Model 3 (17.817), and Model 2 and Model 3 (13.707) showed a rejection of the null hypothesis. These findings shows that the models exhibit significant different forecasting accuracy, demonstrating that the architecture differences between the Model 1, Model 2, and Model 3 made them produce different forecasting accuracy for $PM_{2.5}$. For PM_{10}, significant differences were noted in Model 1 and Model 2 (12.200, $p < 0.001$) and Model 2 and Model 3 (-9.264, $p < 0.001$), implying that there were indeed significant differences between the Model 2 and both competing models in terms of PM_{10} forecasting. Comparing Model 1 and Model 3 shows a non-significant difference (0.768, $p = 0.442$). This demonstrates that the forecasting performance of Model 1 and Model 3 was statistically similar even though there may have been numerical differences in their respective performance metrics.

Generally, the DM test shows that although the three deep learning techniques had significantly different forecasting performances

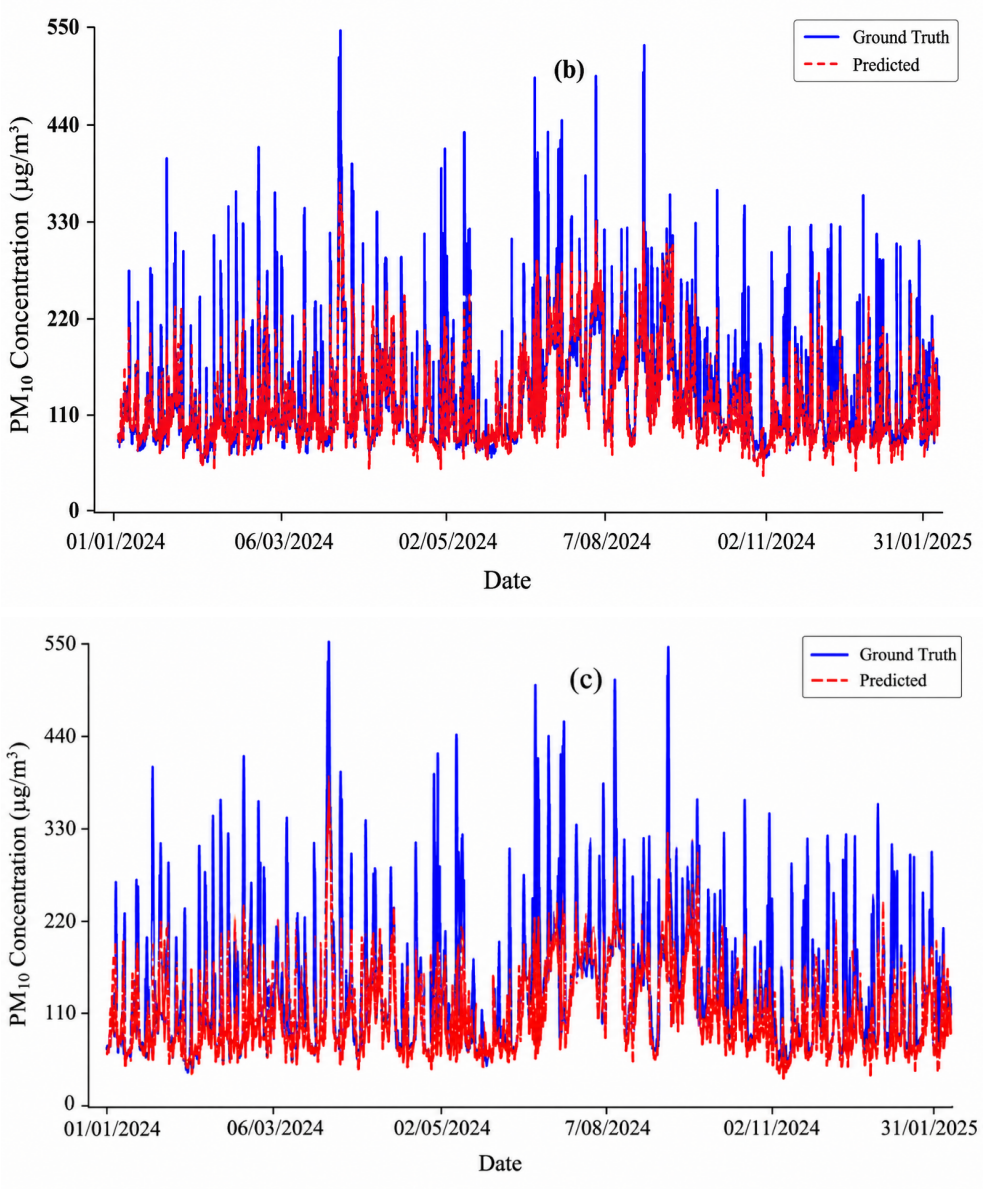


Figure 7: Observed and predicted PM_{10} values for Model 2 (panel b) and Model 3 (panel c).

in the case of $PM_{2.5}$, it is not true that superiority in terms of forecasting of PM_{10} was evident in all models relative to each other. The significant similarity in performance of Model 1 and Model 3 in the case of PM_{10} underscores the necessity of incorporating statistical significance test. Therefore, conclusion regarding model superiority should not be based on metrics alone rather on both predictive performance indicators and statistical evidence.

Examples of measured and predicted values are presented in Table 4.

Figure 7 shows similar fluctuations in PM_{10} concentrations across all three models. Lower predicted PM_{10} values were recorded in July and August, whereas higher values were observed in October.

4. Conclusion

The performance of three deep learning models architecture in predicting particulate matter concentrations ($PM_{2.5}$ and PM_{10}) using air quality and meteorological data was evaluated in this study. $PM_{2.5}$ demonstrated high prediction efficiency for the three models as their R^2 values surpassed 0.90. This shows that the models predicted measured values (0.9166) of above 90%. Model 2 presented the best overall performance among the three models due to its lowest MAE (2.898 $\mu g/m^3$), RAE (0.1697), RMSE (6.453 $\mu g/m^3$), and MAPE (7.7013%).

The accuracy of the prediction for PM_{10} is slightly lower for all models, as evident in the elevated error values and the moderate R^2 values varying between 0.7037 and 0.7348. This decline in accuracy clearly shows that the concentration of PM_{10} exhibits a higher level of temporal variability rather than $PM_{2.5}$. Again, the best generalization performance was shown by Model 2, as it recorded the lowest value of RMSE (12.7097 $\mu g/m^3$) with the highest R^2 value of 0.7348. Although the value of the MAE for Model 1 was slightly less for 7.3013 $\mu g/m^3$, the lower R^2 value signifies less explanatory power.

Finally, this study emphasizes the capability of deep learning models in PM predictive systems, with a special focus on the accuracy of predictive modelling of particulate matter. Model 2 has proven to be more robust in predicting the concentrations of $PM_{2.5}$ and PM_{10} respectively. The higher level of accuracy identified in the results concerning $PM_{2.5}$ concentrations over predictions made for PM_{10} is consistent with the established literature in the existing body of work concerning air quality. The level of moderate accuracy concerning the predictions made for PM_{10} concentrations indicates that further work is required in this area.

Data availability

Data will be made available upon reasonable request from the corresponding author.

Declaration of competing interest

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

Funding

The authors received no specific funding from any public, commercial, or not-for-profit organisation for this research.

References

[1] World Health Organization, "WHO global air quality guidelines for particulate matter($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide and sulfur dioxide", 2021. https://www.who.int/publications/i/item/9789240034433.

[2] United Nations Environment Programme, "Cleaner fuels and vehicle emission standards proposed for West Africa", UN Environment (2018). Available online: https://www.unep.org/news-and-stories/story/cleaner-fuels-and-vehicle-emission-standards-proposed-west-africa.

[3] A. M. Evuti, "Sustainable environmental management in Nigeria: an appraisal", The Journal of Social Sciences Research, Special Issue **2** (2018) 257. https://doi.org/10.32861/jssr.spi2.257.263.

[4] Climate & Clean Air Coalition, "Nigeria - National action plan (NAP) to reduce short lived climate pollutant planning on short-lived climate pollutants(SLCP)", 2023. https://www.ccacoalition.org/sites/default/files/policy-documents/nigeria-s-national-action-plan-nap-to-reduce-short-lived-climate-pollutants-slcps-.pdf.

[5] J. Adetayo, R. Yusuf, M. Amole, L. Jimoda & J. Sonibara, "Air quality impact of diesel back-up generators (BUGs) in Nigeria's mobile telecommunication base transceiver stations (BTS)", Management of Environmental Quality: An International Journal **28** (2017) 723. https://doi.org/10.1108/MEQ-09-2015-0168.

[6] O. Adama, "Urban imaginaries: funding mega infrastructure projects in Lagos, Nigeria", GeoJournal **83** (2018) 257. https://doi.org/10.1007/s10708-016-9761-8.

[7] J. S. Apte, J. D. Marshall, A. J. Cohen & M. Brauer, "Addressing global mortality from ambient $PM_{2.5}$", Environmental Science & Technology **49** (2015) 8057. https://doi.org/10.1021/acs.est.5b01236.

[8] A. Enumah, A. M. Evuti, K. A. Salam, Y. S. Mohammad & S. S. Silas, "Temporal variations of PM_{10} and $PM_{2.5}$ in Lugbe, FCT Abuja Nigeria", Journal CleanWAS **9** (2025) 98. https://www.researchgate.net/publication/399528918_Temporal_Variations_Of_Pm_And_Pm_In_Lugbe.

[9] J. Wang, T. Niu & R. Wang, "Research and application of an air quality early warning system based on a modified least squares support vector machine and a cloud model", International Journal of Environmental Research and Public Health **14** (2017) 249. https://doi.org/10.3390/ijerph14030249.

[10] S. Wu, X. Li, W. Dong, S. Wang, X. Zhang & Z. Xu, "Multi-source and heterogeneous marine hydrometeorology spatio-temporal data analysis with machine learning: a survey", World Wide Web **26** (2023) 1115. https://doi.org/10.1007/s11280-022-01069-4.

[11] S. Subramaniam, N. Raju, A. Ganesan, N. Rajavel, M. Chenniappan, C. Prakash, A. Pramanik, A. K. Basak & S. Dixit, "Artificial intelligence technologies for forecasting air pollution and human health: a narrative review", Sustainability **14** (2022) 9951. https://doi.org/10.3390/su14169951.

[12] F. Biancofiore, M. Busilacchio, M. Verdecchia, B. Tomassetti, E. Aruffo, S. Bianco, S. Di Tommaso, C. Colangeli, G. Rosatelli & P. Di Carlo, "Recursive neural network model for analysis and forecast of PM_{10} and $PM_{2.5}$", Atmospheric Pollution Research **8** (2017) 652. https://doi.org/10.1016/j.apr.2016.12.014.

[13] U. Pak, J. Ma, U. Ryu, K. Ryom, U. Juhyok, K. Pak & C. Pak, "Deep learning-based $PM_{2.5}$ prediction considering the spatiotemporal correlations: a case study of Beijing, China", Science of the Total Environment **699** (2020) 133561. https://doi.org/10.1016/j.scitotenv.2019.07.367.

[14] A. Bekkar, B. Hssina, S. Douzi & K. Douzi, "Air-pollution prediction in smart city, deep learning approach", Journal of Big Data **8** (2021) 161. https://doi.org/10.1186/s40537-021-00548-1.

[15] L. Zhang, P. Liu, L. Zhao, G. Wang, W. Zhang & J. Liu, "Air quality predictions with a semi-supervised bidirectional LSTM neural network", Atmospheric Pollution Research **12** (2021) 328. https://doi.org/10.1016/j.apr.2020.09.003.

[16] S. Chae, J. Shin, S. Kwon, S. Lee, S. Kang & D. Lee, "PM10 and PM2.5 real-time prediction models using an interpolated convolutional neural network", Scientific Reports **11** (2021) 11952. https://doi.org/10.1038/s41598-021-91253-9.

[17] J. Duan, Y. Gong, J. Luo & Z. Zhao, "Air-quality prediction based on the ARIMA-CNN-LSTM combination model optimized by dung beetle optimizer", Scientific Reports **13** (2023) 12127. https://doi.org/10.1038/s41598-023-36620-4.

[18] T. D. Morapedi & I. C. Obagbuwa, "Air pollution particulate matter ($PM_{2.5}$) prediction in South African cities using machine learning techniques", Frontiers in Artificial Intelligence **6** (2023) 1230087. https://doi.org/10.3389/frai.2023.1230087.

[19] O. E. Taylor & P. S. Ezekiel, "A model for forecasting air quality index in Port Harcourt Nigeria using Bi-LSTM algorithm", arXiv (2023). https://doi.org/10.48550/arXiv.2302.03930.
[20] Q. Liu, B. Cui & Z. Liu, "Air quality class prediction using machine learning methods based on monitoring data and secondary modeling", Atmosphere **15** (2024) 553. https://doi.org/10.3390/atmos15050553.
[21] S. Gaowa, Z. Zhang, J. Nie, L. Li, H. A-ru & Z. Yu, "Using artificial neural networks to predict indoor particulate matter and TVOC concentration in an office building: model selection and method development", Energy and Built Environment **6** (2025) 750. https://doi.org/10.1016/j.enbenv.2024.03.001.
[22] D. Faye, R. Lguensat, F. Kaly, A. Sudmant, A. Gaye & E. Kalisa, "Machine learning for air quality forecasting: insights from five provinces of Rwanda", Scientific African **30** (2025) e02959. https://doi.org/10.1016/j.sciaf.2025.e02959.
[23] O. A. Falaiye, O. F. Odubanjo & M. Sanni, "Leveraging machine learning for accurate $PM_{2.5}$ concentration prediction in selected Nigerian locations", Journal of Atmospheric and Solar-Terrestrial Physics **271** (2025) 106515. https://doi.org/10.1016/j.jastp.2025.106515.
[24] O. F. Odubanjo, T. O. Aluko, M. Sanni & O. E. Jimoh, "Machine learning prediction of $PM_{2.5}$ in Lagos using embedded real-time environmental monitoring", FUDMA Journal of Sciences **9** (2025) 218. https://doi.org/10.33003/fjs-2025-0909-3932.
[25] National Population Commission, "Nigeria population projections and demographic indicators", 2020, pg. 172. Available online: https://factcheckhub.com/wp-content/uploads/2022/07/national-population-commission-Projection_2022.pdf.
[26] National Bureau of Statistics, 2012, pg. 65. Available online: https://nigerianstat.gov.ng/pdfuploads/annual_abstract_2012.pdf.
[27] Nigerian Meteorological Agency, "State of the climate in Nigeria", 2023, pp 5-12. Available online: https://nimet.gov.ng/document.
[28] E. Abdullahi, L. Okobia & S. Hassan, "Assessment of ambient atmospheric concentration of volatile organic compounds", J. Chem. Bio. Phy. Sci. Sec. D. **2** (2012) 1637. https://www.researchgate.net/publication/236612106_Correspondence_Author_Assessment_of_Ambient_Atmospheric_Concentration_of_Volatile_Organic_Compounds_in_Abuja-Nigeria
[29] PurpleAir, "Classic Plus air quality monitor" (2026). Available online: https://www.purpleair.com/products/classic-plus-air-quality-monitor.
[30] C. Yi, J. Chen & L. Li, "An adaptive robust gradient-based recurrent neural network for solving time-varying linear matrix equation and its application", Journal of the Franklin Institute **362** (2025) 107991. https://doi.org/10.1016/j.jfranklin.2025.107991.
[31] Y. Huang, X. Zhang & Y. Li, "A novel hybrid model for $PM_{2.5}$ concentration forecasting based on secondary decomposition ensemble and weight combination optimization", IEEE Access **11** (2023) 119748. https://doi.org/10.1109/ACCESS.2023.3327707.
[32] E. Gross, R. Kruger & F. Toerien, "A comparative analysis of option pricing models: Black–Scholes, Bachelier, and artificial neural networks", Risk Management **27** (2025) 8. https://doi.org/10.1057/s41283-025-00160-0.