Microwave Review
December 2025, Vol. 31, No. 2, pp. 35-44
DOI: 10.18485/mtts_mr.2025.31.2.5

Forecasting Air Pollution at Construction Sites: An Extended Comparison of Machine Learning Approaches

Eliezer Zahid Gill[1], Leonardo Cangelmi[1], Paola Cellini[1], Alessandro Zaldei[2], Ivo R. Draganov[3], Daniela Cardone[1] and Alessia Amelio[1]

Abstract – **In this paper, we introduce a new approach designed to forecast up to 24 hours in advance five air pollutants at construction sites given the temperature as an exogenous variable. The considered pollutants are two particulate matter (PM2.5, PM10) and three gaseous pollutants (NO_2, CO and O_3). The proposed approach employs and compares three well-known predictive model types on forecasting each air pollutant: random forest, feed-forward artificial neural network and transformer architecture. Their performances are evaluated in terms of Root Mean Squared Error on the single pollutants with single-output models. Experimental results, based on data collected from AirQino sensor stations at a real construction site, demonstrate that random forest continues to be a reliable and frequently top-performing choice for long-term forecasting in this context. However, the transformer architecture emerges as a strong alternative, especially well-suited for short-term prediction tasks and scenarios with limited data availability.**

Keywords – **Ensemble learning, Artificial neural network, Transformer architecture, Air pollution, Construction site**

I. INTRODUCTION

In recent time, the problem of air pollution has become a serious matter, which is influencing the environment and the health of inhabitants. Specifically, the metropolitan areas are the major sources of air pollution, given the multiple tasks occurring in the different fields. One of them is the activity of the workers in the construction site, which is particularly dangerous for workers, environment and people living in the neighbourhood of the construction site, since it spreads various pollutants in the air. Indeed, each step of the working site project, from demolition to construction, produces multiple types of air pollutants. They include: (i) suspended particulate (PM), (ii) carbon monoxide (CO), (iii) nitric oxide (NO) and nitrogen dioxide (NO_2), (iv) volatile organic compounds (VOCs), (v) ozone (O_3), (vi) Polycyclic Aromatic Hydrocarbons (PAHs), and (vii) sulphur dioxide (SO_2). Among the different generated pollutants, the particulate matter (PM2.5 and PM10), caused by the vehicles during the construction and demolition steps, is the most critical one [1], since it causes serious illnesses at the respiratory and cardiovascular system, infertility, damages to the nervous system and cancer. Multiple studies have showed that global annual mortality linked to air pollution from PM2.5 fine particles alone is more than double that calculated previously and amounts to almost 9 million deaths [2]. The main sources of emissions in the construction site include: (i) earthmoving activities, (ii) use of heavy machinery, (iii) combustion processes, and (iv) inadequate waste management. Accordingly, it is essential to pinpoint the sources of pollution on the construction sites and implement corrective measures to promote a healthy work environment and minimize environmental impact.

Due to the significant risks posed by air pollution, the EU has established directives such as 2008/50/EC, 2004/107/EC, and 2004/37/EC to set pollutant limits and quality objectives, requiring Member States to monitor and reduce exposure [3]. However, continuous monitoring and strict enforcement remain essential to ensure compliance.

For this reason, different solutions have flourished for monitoring and predicting the diffusion of air pollutants in the environment, with reference to construction sites. They include spectroscopy, chromatography, and spectrometry. Among the different platforms, the AirQino sensor was introduced for monitoring the emission of both particulate matter and gaseous pollutants, together with the tracking of other meteorological variables [4]. In addition, different machine learning models, including Long Short-Term Memory (LSTM) network, Artificial Neural Network (ANN), linear regression and ensemble learning methods, have been adopted to predict the diffusion of the air pollutants in the construction sites at multiple stations, given the knowledge of various environmental and meteorological variables, like temperature, pressure, dew point, rain, wind direction and wind speed [5-12]. Some recent works employed the LSTM model to make a 6 or 12-hours forecast of six different pollutants (PM2.5, PM10, SO_2, NO_2, CO and O_3) at a time [13,14]. Another work compared different single and multi-output machine learning and statistical models, like ANN, LSTM, Random Forest (RF), Gradient Boosting (GB), Autoregressive Integrated Moving Average (ARIMA), Vector Autoregressive (VAR) and Vector Error Correction (VECM) models, for the same task of making a 12-hours prediction of the six pollutants [15]. Finally, the work in [16] extended the previous works by comparing RF and ANN regressors to forecast five of the six air pollutants, i.e. PM2.5, PM10, O_3,

Article history: Received November 25, 2025; Accepted December 03, 2025. This paper is an expanded version of the article "Air Pollution Forecasting at Construction Sites: An Intelligent Comparative Framework" presented at 60th International Scientific Conference on Information, Communication and Energy Systems and Technologies, (ICEST 2025), Ohrid, North Macedonia, June 26 – 28, 2025. [DOI: 10.1109/ICEST66328.2025.11098263].

[1]Eliezer Z. Gill, Leonardo Cangelmi, Paola Cellini, Daniela Cardone and Alessia Amelio are with the Department InGeo, University "G. d'Annunzio" Chieti-Pescara, Viale Pindaro 42, 65127 Pescara, Italy, E-mail: alessia.amelio@unich.it

[2]Alessandro Zaldei is with National Research Council of Italy, Institute for Bioeconomy, Via Madonna del Piano 10, 50019 Firenze

[3]Ivo R. Draganov is with Technical University of Sofia, Studentski Kompleks, ul."Professor Georgi Bradistilov" 11, 1756 Sofia, Bulgaria

NO_2, and CO, with 12 hours in advance, given the temperature as exogenous variable. Differently from the previous works, the adopted data were collected from AirQino sensor stations located at different positions of a real construction site.

In this paper, we further advance in the direction of the work in [16] by proposing a new study which extends the prediction time of the five pollutants from 12 to 24 hours. We still adopt the ANN and RF models, but also compare the results with those of transformer architectures on the same AirQino data. Indeed, they are particularly robust to capture long-range dependencies and handle large and complex datasets efficiently. In such a context, the novelty of our work is the use of transformer models on AirQino data leveraging with self-attention mechanisms to try modelling intricate patterns and contextual relationships for 12 and 24-hours forecast.

The paper is organized as follows. Section II presents the related literature. Section III introduces the proposed approach in terms of data preprocessing, training and testing of the models. Sections IV and V show, respectively, the results obtained from the experiments and discuss them. Finally, Section VI draws conclusions and outlines future work directions.

II. RELATED WORKS

Air pollution has been investigated for many years now with the explicit role of evaluating the risk factors for humans and the environment. In the recent years, lots of machine learning algorithms have been tested and some prove to be efficient enough.

Berkani et al. [17] applied Support Vector Regression, XGB, CatBoost, LightGBM and RF. Especially appropriate tend to be the CatBoost and LightGBM, with a Mean Squared Error (MSE) of 14.0387 (PM10) and 6.7501 (PM2.5), respectively. More than just aiming accurate prediction at a single location, the authors paid attention to the overall spatio-temporal distribution of the pollution. Still, the tested predictors are more complex and time consuming than others which could be optimized.

Rakholia et al. [18] investigated NO_2, SO_2, O_3, and CO concentrations with regard of World Health Organization norms using a multi-step multi-output multivariate model. Meteorological conditions are also embedded into this model. The average mean absolute percentage error observed is between 0.16 and 0.23 for the different pollutants. There are no details about the adaptability of the proposed framework for different control environments.

RF, XGB and LSTM were used in the study of Zhang et al. [19] for 1, 2 and 3d forecasts of PM10, NO_x and O_3. The prediction rate for NO_x reached 60% while the rates for PM10 and O_3 remained lower. The authors describe in-depth the set of hyperparameters of their framework so it could be rearranged for different sites.

Another framework for air pollution prediction was proposed by Kalajdjieski et al. [20] including basic LSTM, and two data augmentation models - ANEncoded and AAN. The loss after training varies from 8.92 for the LSTM, through 5.91 for ANEncoded to 5.55 for AAN. The test MSE changes from 3.41 to 1.27 for 3 datasets, one of which is without augmentation. Augmenting the raw data increased the accuracy of the model.

Jairi et al. [21] tried to overcome the training of multiple models when predicting the levels of air pollutants. They proposed generalizing models scattered at different locations for the separate particulate matters. Starting with PM2.5, they continued with fine-tuning for the rest types of pollutants. Especially effective seems to be the multilayer perceptron with transfer learning when forecasting the time series. The MSE at the testing phase is 0.0649 versus 0.0653 – for a simple Recurrent Neural Network (RNN), 0.0681 – for Gated Recurrent Unit (GRU) and 0.0656 – for LSTM.

The long-term forecasting is the focus of the study of Zhang et. al [22], especially for the PM2.5 pollutant. For a 30 days forecasting, the CatBoost was the optimal solution, while for 90 and 180 days – Bi-LSTM and Bi-GRU, respectively. The authors can establish a connection between the CO concentrations and the influence of PM10 over PM2.5.

FastAI tabular model is considered successfully applicable for the forecasting of NO_2, PM10 and PM2.5 in the traffic-related air pollution. Akinosho et al. [23] implemented lagged variables and categorical embeddings, which led to an increased accuracy of the model. At the same time, it is important to consider the highway data into the rest of the registered values for a broader coverage of the prediction process. Weather data on its own also proved important to generate accurately enough predictions.

A typical case study from the region of Houston, Texas in the US about the air pollution with O_3 reveals the applicability of deep graph neural networks [24]. The spatio-temporal analysis at the base of GraphSAGE paradigm is leading to a successful model capable of maintaining 1, 3 and 6 hours forecast horizons. The forecasting rate changes from 33.7%, through 48.7% to 57.1% for these time limits. The solar radiation is found to have a significant impact on the O_3 generation with the increase of the time horizon.

Some authors such as Handhayani [25] relied on the integrated analysis over the air pollution together with the meteorological conditions. It turns out that PM10, SO_2 and O_3 pollutants could be well forecasted together with the wind speed using LSTM and GRU with some advantage for the first. The resulting models are useful in calculating the air quality index.

Deep learning solutions could be improved according to Jimenez-Navarro [26] by the introduction of a temporal selection layer. Not only the increase of accuracy, but also the feature selection process, interpretability and efficiency could be enhanced. The time selected feed-forward ANN achieved the lowest RMSE of 16.740 among 9 popular classifiers and was the 4^{th} as per the forecasting time.

A further step in the improvement of the forecasters for air pollution includes genetic algorithms. Praveenchandar et al. [27] utilized the Adaptive Network Fuzzy Inference System (ANFIS) to process data for CO, SO_2, O_3 and NO_2. Adding an Improved Genetic Algorithm (IGA) capability, they formed an ANFIS-IGA framework and achieved an RMSE of 0.052658.

Nguyen et al. [28] took other directions of expanding proven solutions for the same purpose as they developed an

LSTM Bayesian neural network. A new approach allowing kernel density estimation and subdivision helps into getting higher accuracy and lower execution time. Also, spatially-adjusted multivariate imputation was introduced which allowed a better forecasting on a local scale given global scale data. PM2.5 and O_3 are the main pollutants of interest. A decrease of 30% for the RMSE is achieved and 50% in inferring time.

Baniasadi et al. [29] tried another enhancement over a classic LSTM network introducing a binary chimp optimization algorithm. The primary aim of their study is the PM2.5 pollution. The new forecasting structure achieves 96.41% accuracy. Eight other methods, including a pure LSTM, RNN and ANN, prove to be less accurate and providing lower R^2 results.

RNNs apart from LSTM are also subject to a wide testing for air pollutant forecasting. In the study of Fauzi et al. [30], an additional step was done by mitigating the influence of volatile environmental factors. Correlation between NO_2 and CO was found as well as the overall influence of the COVID-19 pandemic on the pollutant spreading in Jakarta, Indonesia. Concentrations of PM10 and O_3 are also well predicted by the RNN. On the other hand, SO_2 and CO quantities are better predicted by an LSTM.

On prolonged registration intervals as a monthly basis, various deep learning models appear to be adequate for air pollution forecasting. He and Guo [31] applied them for the Dezhou city in China. PM2.5 is the main pollutant in this study, with a dataset covering 10 years. A forecaster CNN-GRU-LSTM achieves 34.28% higher accuracy than a simple ANN, among of all 8 tested models. At the same time, R^2 and RMSE are 0.9686 and 4.6491, respectively.

Whale algorithm was found appropriate by Luo and Gong [32] to optimize another neural structure known as ARIMA-WOA-LSTM forecasting air pollutants. The ARIMA part extracts the linear part of the processed data and then the WOA-LSTM processes the non-linear part. The whale algorithm selects the most suitable hyperparameters of the LSTM. Both the accuracy and stability of the prediction are higher than a single LSTM.

III. The Proposed Approach

We propose a new approach to forecast the temporal diffusion of air pollutants at certain positions of a construction site. It consists of three main stages: (i) data preprocessing, (ii) model training, and (iii) model evaluation. The approach builds upon the work in [16], which provided 12-hours-ahead forecasts, comparing the performances of RF and ANN models for each pollutant.

This study extends the previous work by also considering a 24-hours prediction, which provides a more realistic basis for taking corrective measures on the construction site. Additionally, it also introduces a new model of encoder-based transformer for time series forecasting, which is compared to the RF and ANN models for each pollutant. In the proposed approach, single-output ANN, RF and transformer models are designed for predicting the following five air pollutants at construction sites: (i) PM2.5, (ii) PM10, (iii) O_3, (iv) NO_2, and (v) CO with the temperature as an exogenous variable. Comparing ANN, RF and transformer models in the context of air pollutant forecasting at construction sites is relevant because each model type captures different aspects of time series behaviour. ANNs are effective for modelling nonlinear relationships with relatively low computational cost, RFs provide robustness to noisy data, while transformers excel at learning complex temporal dependencies. Pollutants such as PM2.5, PM10, O_3 , NO_2 , and CO exhibit diverse dynamics influenced by both short-term activities and longer-term environmental cycles. Accordingly, comparing the models helps identify the most suitable machine learning approach for each pollutant. This comparison ensures more accurate forecasts and supports realistic corrective actions on-site. For testing, we rely on data collected from sensor stations deployed at an actual construction site. All models are trained from scratch using a subset of this data, without pre-training or fine-tuning. The subsequent sections provide a detailed description of each step of the proposed approach.

A. Data Preprocessing

Data preprocessing is characterized by the following steps: (i) feature selection, (ii) handling missing values, (iii) making visual inspection, (iv) deleting negative values, (v) computing the cross-correlation and autocorrelation functions, (vi) verifying the time series' stationarity, (vii) applying the log transformation and min-max scaling.

Feature selection consisted of explicitly dropping irrelevant or redundant features. This step focused purely on the target pollutants (PM2.5, PM10, O_3, NO_2, and CO) and the input feature (temperature). This improves computational efficiency and model performance by removing noise.

Missing spots for each feature were imputed with the mean. Next, aggregation was applied for computing daily average pollutant levels to visually spot long-term trends. Then, the raw dataset was carefully cleaned by eliminating unnecessary measurements and removing any illogical negative readings. The next step involved exploring how strongly each pollutant is linked to temperature (cross-correlation) and how much it depends on its own history (autocorrelation) over the last 50 time steps.

A major technical hurdle, ensuring the data was statistically stationary, meaning its properties don't change over time, was addressed with the Dickey-Fuller test.

Finally, before feeding the data to the prediction algorithms, any heavy skewing was fixed using a log transformation to stabilize the variance. It makes the variance of the data more consistent over time and normalizes skewed data by making their distribution closer to normal, which helps artificial models to perform better.

Finally, min-max scaling was applied to uniformly fit all values into the [0, 1] interval.

B. Training of the Model

We experiment three different machine learning models: (i) RF for regression [33], (ii) raw feed-forward ANN [34], and (iii) encoder-based transformer for time series forecasting [35].

An RF for regression is an ensemble learning structure, constituted of several independent decision trees for regression, each of which processes a subset of the sample data. The input subsets of the decision trees that compose the RF for regression are chosen at random, possibly with repetition. The prediction of the RF for regression is the average of the predictions of its trees. The use of randomly chosen data subsets prevents overfitting and improves the final prediction. A single decision tree predicts the value of a real target variable in function of the values of some feature variables. The prediction mechanism consists in partitioning the domain of features and assigning to each part a target value, derived from the sample data. A sample datum is a sequence of features components with a corresponding target component. The domain of features, and consequently the data sample, is recursively split in two by a tree-like algorithm, which terminates according to some stopping conditions. At each node of the tree a feature is chosen, together with a splitting of the domain of that feature. The choice criterion is to minimize a certain function of the target values. The set of leaves of the tree correspond to the resulting partition of the domain of features, which naturally provides also a partition of the sample data. For each feature subdomain in the partition, the prediction is the average of the target values in the corresponding sample part. The RF for regression is mathematically defined as:

$$\hat{y}_{RF}(x) = \frac{1}{K}\sum_{k=1}^{K} T_k(x) \qquad (1)$$

where:

- $\hat{y}_{RF}(\mathbf{x})$ is the final RF prediction for input $\mathbf{x}$,
- K is the number of estimators,
- $T_k(\mathbf{x})$ is the prediction of the k-th individual decision tree for input $\mathbf{x}$.

Figure 1 shows the functioning of the RF regressor.

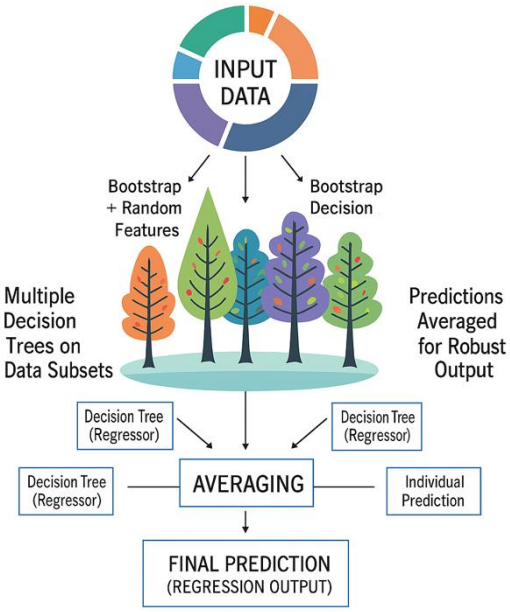

Fig. 1. Flowchart of the RF for regression

A feed-forward ANN is the simplest kind of ANN. It is constituted by a sequence of interconnected layers, an input layer, possibly followed by one or more hidden intermediate layers, and an output layer. Each layer consists of mutually independent computational units, called nodes or neurons, and any two adjacent layers are fully connected each with the other, i.e., all the neurons of a layer are connected to all the neurons of the subsequent layer. Information is propagated forward, from a layer to the next one, up to the output layer, by affine linear functions and possibly by some nonlinear activation functions. Those affine linear functions are the learning parameters. The learning mechanism is the backpropagation algorithm. Mathematically, one layer of the feed-forward ANN is defined as:

$$\mathbf{h} = f(\mathbf{W}\mathbf{x}+\mathbf{b}) \qquad (2)$$

where:

- $\mathbf{x}$ is the input vector,
- $\mathbf{W}$ is the weight matrix connecting the previous layer to the current layer,
- $\mathbf{b}$ is the bias vector,
- $\mathbf{W}\mathbf{x} + \mathbf{b}$ is the linear transformation,
- f is the activation function,
- $\mathbf{h}$ is the output vector of the current layer.

Figure 2 shows a feed-forward ANN for regression with 3 dense hidden layers and one output.

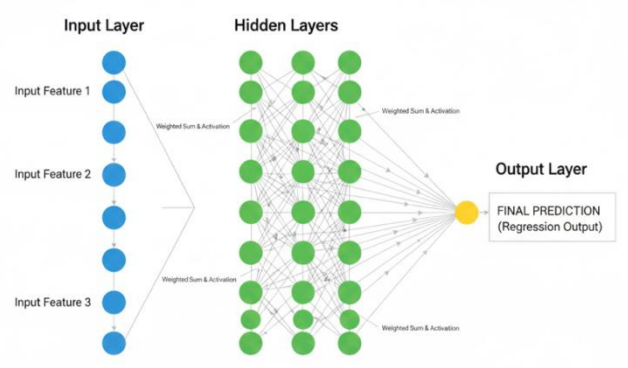

Fig. 2. Feed-forward ANN with 3 dense hidden layers and one output

Finally, an encoder-based transformer model for time series prediction consists in an encoder to which a prediction head is added. It can handle long-range dependencies well, which is vital for recognising recurring pollutant patterns (e.g., daily/weekly cycles). An encoder network is composed by an embedding layer followed by a stack of encoder layers. The embedding layer converts the input data into a suitable input for the encoder stack. The input of the network is a sequence of input units. In the embedding stage each input unit $\mathbf{X}$ is transformed into a real embedding vector $\mathrm{E}(\mathbf{X})$ of fixed dimension. Moreover, a positional encoding vector $\mathrm{P}(\mathbf{X})$ is summed to $\mathrm{E}(\mathbf{X})$, to consider the temporal arrangement of the units. The value $\mathrm{P}(\mathbf{X})$ is a function of the sole position of $\mathbf{X}$, say t, and is of type $(\sin\theta_1(t), \cos\theta_1(t) \ldots \sin\theta_h(t), \cos\theta_h(t))$, where h is half the dimension, and $\theta_i(t) = t/r^i$, for $i = 1...h$, with r a fixed very large number. The vectors $\mathrm{E}(\mathbf{X}) + \mathrm{P}(\mathbf{X})$, arranged in a matrix, are the input of the encoder stack. Each encoder layer consists of a set of attention heads which work

in parallel, each of which computes a self-attention matrix. The concatenation of those attention matrices is then passed to the feed-forward layer. The feed-forward neural (FFN) networks used in transformers are two layers perceptrons which act independently of the position. This implies that each FFN layer acts on its input matrix **H** by a transformation of type $L_2(\varphi(L_1(\mathbf{H}))$, with L_1 and L_2 of type $L_i(\mathbf{Y}) = \mathbf{Y}\mathbf{W}_i + \mathbf{B}_i$, where the matrices $\mathbf{W}_i$, $\mathbf{B}_i$, (i = 1, 2) are learning parameters, and φ is an activation function. Here, $\varphi(x)$ is the ReLU function *max* (0, x) applied pointwise. Both the self-attention and feed-forward sublayers are enriched by a residual connection and normalization module. The final output embedding is then passed to the prediction head to forecast future sequence values.

The adopted model includes the input embedding/projection, the positional encoding and an encoder layer, consisting of multi-head self-attention and feed-forward networks. The positional input embedding **K** is mathematically defined as:

$$\mathbf{K} = \mathbf{E} + \mathbf{P}$$
$$\mathbf{E} = \mathrm{ReLU}\,(\mathbf{X}\mathbf{W}_{in} + \mathbf{b}_{in}) \quad (3)$$

where:

- **X** is the flattened input window,
- $\mathbf{W}_{in}$ is the input projection weight matrix,
- $\mathbf{b}_{in}$ is the input projection bias vector.

Figure 3 shows the architecture of the adopted encoder-based transformer.

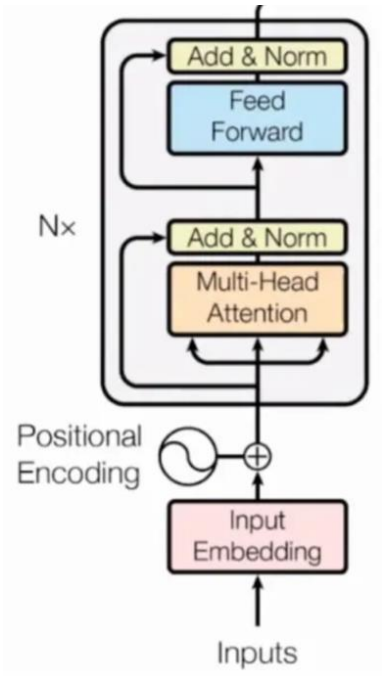

Fig. 3. Architecture of the encoder-based transformer

C. Testing of the Model

Environmental data for the experiment were collected from two AirQino stations, referred to as Station 1 and Station 2. AirQino is a low-cost air quality monitoring platform developed by the Italian National Research Council (CNR-IBE) in collaboration with industrial partners. The system is designed to complement regulatory monitoring networks by providing high-resolution environmental observations through dense spatial deployment. Each unit integrates modular sensors for gaseous pollutants (NO_2, O_3, CO), carbon dioxide (CO_2), particulate matter (PM2.5, PM10), and meteorological parameters (temperature, relative humidity). Data acquisition and preprocessing are managed by a microcontroller, with communication modules (Wi-Fi, GSM, Ethernet) enabling transmission to a centralized web portal. The devices are typically powered via mains supply and include backup battery options. Particulate matter is measured using a laser-scattering optical particle counter, which estimates particle concentrations from light scattering intensity and temporal patterns. AirQino returns the pollutant values in micrograms per cubic meter (μg/m³). The raw data collected by the stations are transmitted to a web server, where calibration coefficients are applied to convert raw counts into calibrated values (μg/m³). Because low-cost sensors are sensitive to environmental factors such as humidity and aerosol composition, AirQino incorporates laboratory calibration against reference instruments and field validation through co-location with regulatory monitors. These procedures improve linearity, bias, and reproducibility, yielding data quality that approaches indicative-level accuracy for environmental assessment and research applications. Figure 4 shows the AirQino platform.

The two AirQino sensor stations (Station 1 and 2) were installed at the same construction site to facilitate a direct comparison. One dataset was collected for each station, capturing hourly values from February 13, 2021, to December 31, 2023, resulting in a total of 20,891 data points per station. To evaluate the robustness of the ANN, RF and transformer architecture models against variations in data size and source, all models were trained using two different data periods. For Station 1, data from one year (August 13, 2021 – August 13, 2022) were used, and for Station 2, six months of data (February 13, 2022 – August 13, 2022) were used. The models were tested using the next six months of data (August 14, 2022 – February 14, 2023) from both stations. The 10% of the training data was reserved for validation purposes.

During the preprocessing stage, the focus was on extracting and filtering air pollutants expressed in counts over a period. Furthermore, any instances of negative values were deleted in both datasets. To ensure data quality, the Dickey-Fuller test was also applied to assess the stationarity of the time series, identifying some time series as non-stationary. This led us to explore the performance of RF, ANN, and transformer models in handling non-stationary data.

The datasets from both stations were split into 48-hour time windows, which previous studies have shown to be the optimal duration for similar tasks [13-16]. This two-day window helped create input data that would allow the models to predict pollutant levels for both 12-hours and 24-hours forecasts. It must be noted that after prediction, the process is reversed. The model's output is reshaped from the batch structure back into time steps.

To maximise the training effectiveness and avoid overfitting, the default hyperparameters' values were used for each model, which are typically suitable for regression tasks.

(a) (b)

(c)

Fig. 4. AirQino platform

In particular, the main values for all three models include:

- Adam optimiser: a standard, efficient optimisation algorithm for deep learning with MSE as the loss function (learning rate=0.0001).
- Regularisation: prevents overfitting by penalising large weights (weight decay=1e-05).
- Learning rate scheduler: adjusts the learning rate dynamically (reduces it by a factor of 0.5 if the loss plateaus for 10 epochs), aiding convergence.
- Early stopping: halts training if the loss doesn't improve after a set number of epochs (patience=20), preventing overfitting and saving time and computational cost.
- Gradient clipping: used to prevent the exploding gradients during the training. It rescales gradients if their norm exceeds a specified threshold.
- Number of estimators: the number of decision trees in the forest. In the analysis, 100 trees were adopted.

The designed ANN architecture was characterized by three dense layers connected in sequence with (128, 64, 32) neurons. It maps the flattened input directly to the 12 and 24-hours predicted output. It is computationally lighter but might struggle to capture complex temporal dependencies. About the transformer architecture, the flattened input window contained 48 values. The input and output projection spaces were set to 128 dimensions, with 8 attention heads and 3 layers in the encoder. The hidden state of the transformer was mapped to the 12-hours and 24-hours prediction horizons using a multi-layer perceptron.

For the prediction of each air pollutant, the final model performance was quantified using the Root MSE, a robust metric for regression defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(Actual_i - Predicted_i)^2} \qquad (4)$$

where $Actual_i$ and $Predicted_i$ are, respectively, the measured and predicted values of pollutant at time step i and N is the length of the pollutant time series. Lower RMSE values indicate a better model performance.

The experiments were run on a laptop computer with a CPU i5-12500H, 16GB RAM and a GPU NVIDIA GeForce 3050 RTX. The Python 3.13 was adopted as the programming language. To implement the ANN and RF models, Tensorflow Keras and Scikit-learn libraries were utilized, respectively. Additionally, the PyTorch library was adopted to implement the transformer architecture. Additional libraries included Pandas and NumPy for data manipulation and mathematical computation.

IV. RESULTS

This section presents the prediction results in terms of RMSE for air pollutant concentrations obtained from the three distinct models: RF, ANN, and the transformer architecture. Results are computed on the test data from the two monitoring stations (Station 1 and Station 2) for two prediction horizons: (i) the previously studied 12-hours ahead prediction, and (ii) a newly extended 24-hours ahead prediction. For clarity, the best result for each pollutant and prediction horizon is bolded in Table 1 (for Station 1) and Table 2 (for Station 2), and the mean RMSE values for each model are highlighted in red. Additionally, Figs. 5 and 6 report for each pollutant the difference computed between RMSE of ANN (resp. RF) and RMSE of the transformer architecture at Station 1 (resp. Station 2) for 12-hours and 24-hours prediction. Higher positive values in the charts mean that transformer models achieve a better performance than the competing models.

For the Station 1 forecasting of the pollutants for 12-hours (see Table 1 and Fig. 5), the RF model outperforms both ANN and transformer models in most of the pollutants. In particular, the RF shows the lowest RMSE for most pollutants: O_3 (0.107), NO_2 (0.087), and CO (0.070). Also, the transformer model outperforms the ANN model for all pollutants. It is also seen that the RMSE for NO_2 is much higher for the ANN model (0.121) than for the transformer (0.089). Finally, it can be noted that the transformer model obtains the best RMSE for particulate matter prediction.

In general, for the 12-hours prediction task at Station 1, which aligns with our previous research, the ensemble and transformer models demonstrate competitive performance. On average, the RF model (mean RMSE = 0.115) outperforms the ANN model (mean RMSE = 0.129), while it is comparable with the transformer architecture (mean RMSE = 0.117), corroborating our earlier findings. The introduction of the transformer architecture reveals a more nuanced picture. The

transformer achieves a mean RMSE of 0.117, positioning it as a strong contender, outperforming ANN but not surpassing the established RF model for this shorter-term forecast.

TABLE 1
RESULTS OBTAINED FROM RF, ANN AND TRANSFORMER MODELS IN TERMS OF RMSE ON THE TEST SET OF STATION 1 FOR 12 HOURS AND 24 HOURS IN ADVANCE PREDICTION

Data source	Station 1					
Air pollutant	12-hrs prediction			24-hrs prediction		
	ANN	RF	Transf.	ANN	RF	Transf.
PM2.5	0.147	0.148	**0.146**	0.878	**0.850**	0.870
PM10	0.160	0.161	**0.159**	0.990	**0.960**	0.991
O_3	0.120	**0.107**	0.109	**0.090**	**0.090**	0.102
NO_2	0.121	**0.087**	0.089	0.120	**0.117**	0.125
CO	0.095	**0.070**	0.085	0.082	**0.060**	0.077
Avg. PM	**0.154**	**0.155**	**0.153**	**0.934**	**0.905**	**0.931**
Avg. gaseous	**0.112**	**0.088**	**0.094**	**0.097**	**0.089**	**0.101**
Tot. Avg.	**0.129**	**0.115**	**0.117**	**0.432**	**0.415**	**0.434**

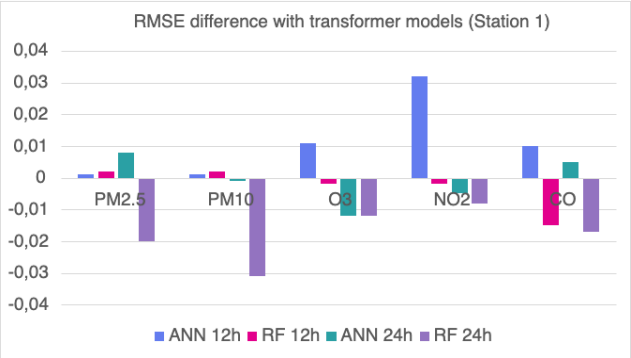

Fig. 5. Difference between RMSE of ANN (resp. RF) and RMSE of transformer models in 12-hours prediction (ANN 12h and RF 12h), and in 24-hours prediction (ANN 24h and RF 24h) for each pollutant at Station 1. Higher values mean better performances of the transformer model

The results from Station 1 for the 24-hours prediction horizon (see Table 1 and Fig. 5) reveal a significant shift in model performance and a general increase in forecasting difficulty, as evidenced by the substantially higher mean RMSE values across all models. In this more challenging scenario, the RF model emerges as the clear frontrunner with a mean RMSE of 0.415, outperforming both the ANN model (0.432) and the transformer model (0.434). Notably, the transformer, which was competitive at 12 hours, now shows the highest average error. This is particularly evident for PM10 and PM2.5, where the RMSE is significantly higher than that of RF (see Fig. 5). This indicates that while the transformer's attention mechanism is powerful, it may require more data or specific architectural tuning to maintain its advantage over longer forecasting horizons where long-range dependencies become even more critical. RF's ensemble strategy appears more robust to this increased temporal complexity. A detailed analysis shows that the transformer model secured the best performance for O_3 and CO, albeit by a very narrow margin under RF (e.g., 0.102 vs. 0.09 for O_3 and 0.077 vs. 0.06 for CO). This suggests that its self-attention mechanism begins to capture the complex, non-linear temporal patterns of these gaseous pollutants more effectively but no better than the ensemble models. Indeed, for both particulate matter and gaseous pollutants, the RF model remains dominant (see Fig. 5), indicating its robust feature learning from tabular data for these species.

The results from Station 2 provide critical insights into model robustness and data efficiency (see Table 2 and Fig. 6).

For the 12-hours prediction, a dramatic performance shift is observed with the introduction of the transformer. It achieves a substantially lower mean RMSE (0.108) compared to both RF (0.134) and ANN (0.139). This superiority is consistent across all individual pollutants, with particularly significant improvements for O_3, NO_2, and CO (see Fig. 6). This strongly suggests that the transformer architecture is remarkably data efficient in this context. Its ability to model long-range dependencies directly, even with limited data, allows it to learn the underlying temporal dynamics of air pollution more effectively than the sequential processing of ANN or the bagging approach of RF.

For the 24-hours prediction at the Station 2, the performance hierarchy changes again, mirroring the trend observed at Station 1. The RF model delivers the best average performance (mean RMSE = 0.393), followed closely by the transformer model (0.398) and then ANN model (0.403). The RF and transformer models show similar performance in most cases, with slight differences. For PM2.5 and PM10, the RF model performs best, while for O_3, NO_2, and CO, transformer and RF models are quite comparable (see Fig. 6). The transformer model has a slight disadvantage in predicting PM2.5 (0.797) and PM10 (0.861) compared to RF (0.788 and 0.849, respectively). This reinforces the finding that RF models exhibit superior robustness for extended forecasting periods. However, it is notable that the transformer's performance is much more competitive than it was at Station 1 for the 24-hours task, nearly matching RF. It further underscores its data-efficient nature, as it manages to generalize well for a longer temporal horizon despite the smaller training set.

From all above, it is worth noting that the transformer model's performance tends to be closer to that of RF in both stations, especially for 12-hours predictions, but struggles more when extended to 24-hours predictions. Station 1 data, with more complex temporal patterns in pollutants, shows a sharper decline in the transformer model's performance, particularly for particulate matter pollutants. Station 2 data, with a smaller training set, sees the transformer outperforming both ANN and RF models for 12-hours prediction, but transformer does worse in specific pollutants (such as PM2.5 and PM10) for 24-hours prediction (see Figs 5 and 6).

The most striking difference between the two stations is the performance of the transformer on the 12-hours prediction. At Station 1, it is the second-best model, but at Station 2, it is unequivocally the best (see Tables 1 and 2). This inversion highlights the key finding that the transformer architecture demonstrates exceptional data efficiency, achieving top performance even when training data is scarce, a scenario

where traditional ANN and RF models typically struggle. In contrast, for the 24-hours prediction window, the RF model performs the best over transformer and ANN models, also due to its robustness to noisy variables. Indeed, it is seen that the ANN and transformers struggle to make a successful long windows predictive analysis. Finally, while the RF remains a robust and often best-performing model for long-term forecasting, the transformer architecture establishes itself as a powerful alternative in this context, particularly for short-term predictions and in data-scarce environments. Its superior ability to capture temporal dependencies with high data efficiency makes it a highly promising model for air quality prediction at construction sites, warranting further investigation and potential architectural optimization for longer-horizon forecasts.

TABLE 2
RESULTS OBTAINED FROM RF, ANN AND TRANSFORMER MODELS IN TERMS OF RMSE ON THE TEST SET OF STATION 2 FOR 12 HOURS AND 24 HOURS IN ADVANCE PREDICTION

Data source	Station 2					
Air pollutant	12-hrs prediction			24-hrs prediction		
	ANN	RF	Transf.	ANN	RF	Transf.
PM2.5	0.136	0.133	**0.129**	0.809	**0.788**	0.797
PM10	0.150	0.153	**0.148**	0.867	**0.849**	0.861
O_3	0.110	0.118	**0.068**	0.109	**0.109**	0.113
NO_2	0.148	0.123	**0.077**	0.107	0.107	**0.105**
CO	0.151	0.147	**0.116**	0.122	0.122	**0.115**
Avg. PM	**0.143**	**0.143**	**0.139**	**0.838**	**0.819**	**0.829**
Avg. gaseous	**0.136**	**0.129**	**0.087**	**0.113**	**0.113**	**0.111**
Tot. Avg.	**0.139**	**0.134**	**0.108**	**0.403**	**0.393**	**0.398**

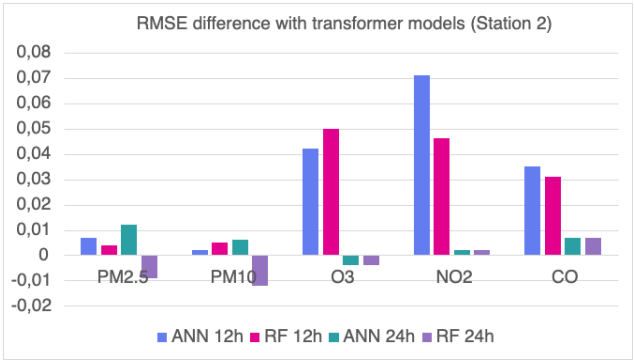

Fig. 6. Difference between RMSE of ANN (resp. RF) and RMSE of transformer models in 12-hours prediction (ANN 12h and RF 12h), and in 24-hours prediction (ANN 24h and RF 24h) for each pollutant at Station 2. Higher values mean better performances of the transformer model

V. DISCUSSION

The comparison of all the results, which involve the amount of data used for training, the prediction horizon, the pollutant types, and the machine learning methods, clearly shows that the RF model and the transformer architecture give the best results in all cases, even if the ANN model gives comparable results in some cases. In particular, limiting our attention to the two particulate matter pollutants, we have very similar results for all three models; however, the short-term prediction results are very good, while the long-term ones are not valid at all. As regards the three gaseous pollutants, the situation is quite diversified. For short-time prediction, the RF model and the transformer architecture are superior to the ANN model in all cases. However, on Station 2 (which was trained with data of only six months), the transformer architecture gives excellent results, and the other two models are almost equivalent. On the other side, on Station 1 (trained with data of one year), the transformer architecture and the RF model give similar results, which are better than those of the ANN model, except for CO, in which case the RF model is superior. For long-time prediction, the results are very similar for all three models, with a slight superiority of the RF model and the transformer architecture. We note that the result of the RF model for the pollutant CO is excellent on Station 1, as it is in the short-term case. This different behaviour of the models, also depending on the time range of training data, deserves some further investigation in the future, through a detailed analysis of the data and context where data were acquired.

VI. CONCLUSION

This work presented a new approach aimed to forecast up to 24-hours in advance five air pollutants (PM2.5, PM10, NO_2, CO and O_3) in construction sites. Three distinct models, such as RF, ANN, and the transformer architecture were employed, and their performances were evaluated in terms of RMSE. The results demonstrated that the RF model was the best performing in case of 24-hours prediction, followed by the transformer architecture. Although the RF model is consistently durable and frequently excels in long-term forecasting, the transformer architecture emerges as a formidable alternative, especially for short-term predictions and in data-scarce settings. Although preliminary, these results demonstrate the feasibility of the approach. Future work is thus forwarded to increase the sample size of the measurements and to improve the model performance by applying transfer learning or testing other ensemble strategies with some explainability techniques and/or by including other meteorological variables, such as relative humidity, as input for the models.

ACKNOWLEDGEMENT

This work has been partially realized at the High Performance Computing Laboratory of the InGeo Department of the University of Chieti-Pescara. The authors would like to thank the Research and Development Sector at the Technical University of Sofia for the financial support.

REFERENCES

[1] H. Yan, Q. Li, K. Feng and L. Zhang, "The Characteristics pf PM Emissions from Construction Sites During the Earthwork and Foundation Stages: An Empirical Study Evidence", *Environ. Sci. Pollut. Res. Int*, vol. 30, no. 22, pp. 62716-62732, 2023, DOI: 10.1007/s11356-023-26494-4

[2] Y. Gu, D. K. Henze, M. O. Nawaz, H. Cao and U. J. Wagner, "Sources of PM2.5-Associated Health Risks in Europe and Corresponding Emission-Induced Changes During 2015", *GeoHealth*, vol. 7, 2023, DOI: 10.1029/2022GH000767

[3] J. López, "REPORT on the Implementation of the Ambient Air Quality Directives: Directive 2004/107/EC and Directive 2008/50/EC", Report - A9-0037/2021, 2021

[4] Institute for Bioeconomy, CNR, AirQino. [Online]. Available at: https://www.AirQino.it/

[5] J. Hong, H. Kang, S. Jung, S. Sung, T. Hong, H.S. Park and D.E. Lee, "An Empirical Analysis of Environmental Pollutants on Building Construction Sites for Determining the Real-Time Monitoring Indices", *Building and Environment*, vol. 170, 2020, DOI: 10.1016/j.buildenv.2019.106636

[6] K. Liao, X. Huang, H. Dang, Y. Ren, S. Zuo and C. Duan, "Statistical Approaches for Forecasting Primary Air Pollutants: A Review", *Atmosphere*, vol. 12, no. 6, 2021, DOI: 10.3390/atmos12060686

[7] J. Krall, "Predictive Learning for Air Pollution at Construction Sites Using Multimodal Data", *i3CE'24, Conference Proceedings*, Pittsburgh, Pennsylvania, USA, 2024, pp.1-10

[8] A. A. Wieser, M. Scherz, A. Passer and H. Kreiner, "Challenges of a Healthy Built Environment: Air Pollution in Construction Industry", *Sustainability*, vol. 13, no. 18, 2021, DOI: 10.3390/su131810469

[9] A. Biloshchytskyi, A. Kuchansky, Y. Andrashko, A. Neftissov, V. Vatskel, D. Yedilkhan and M. Herych, "Building a Model for Choosing a Strategy for Reducing Air Pollution Based on Data Predictive Analysis", *Eastern-European Journal of Enterprise Technologies*, vol. 117, no. 4, 2022, DOI: 10.15587/1729-4061.2022.259323

[10] L. Milivojević, S. Mrazovac Kurilić, Z. Božilović, S. Koprivica and O. Krčadinac, "Study of Particular Air Quality and Meteorological Parameters at a Construction Site", *Atmosphere*, vol. 14, no. 8, 2023, DOI: 10.3390/atmos14081267

[11] Z. Xu, Y. Ran and Z. Rao, "Design and Integration of Air Pollutants Monitoring System for Emergency Management in Construction Site Based on BIM and Edge Computing", *Building and Environment*, vol. 211, 2022, DOI: 10.1016/j.buildenv.2021.108725

[12] E. Z. Gill, D. Cardone and A. Amelio, "Revolutionizing the Construction Industry by Cutting Edge Artificial Intelligence Approaches: A Review", *Frontiers in Artificial Intelligence*, vol. 7, 2024, DOI: 10.3389/frai.2024.1474932

[13] M. Mastromatteo and A. Amelio, "A Deep Learning Approach for Predicting Air Pollutants on the Construction Site", *MIPRO'24, Conference Proceedings*, Opatija, Croatia, 2024, pp. 2089-2094, DOI: 10.1109/MIPRO60963.2024.10569266

[14] E. Z. Gill, A. Amelio, D. Cardone, M. Mastromatteo, P. Cellini, L. Cangelmi, "Integrating Meteorological Data with Artificial Intelligence for Air Quality Prediction on Construction Sites," *24th International Symposium INFOTEH-JAHORINA. INFOTEH 2025*, East Sarajevo, Bosnia and Herzegovina, 2025, pp. 1-6, DOI: 10.1109/INFOTEH64129.2025.10959186

[15] E. Z. Gill, M. Mastromatteo, L. Cangelmi, P. Cellini, D. Cardone and A. Amelio, "A Comparative Analysis of Artificial Intelligence Methods for Air Quality Prediction on Construction Sites," *MIPRO 48th ICT and Electronics Convention*, Opatija, Croatia, 2025, pp. 2127-2132, DOI: 10.1109/MIPRO65660.2025.11131894

[16] E. Z. Gill, L. Cangelmi, P. Cellini, A. Zaldei, I. R. Draganov, D. Cardone, "Air Pollution Forecasting at Construction Sites: An Intelligent Comparative Framework",*60th International Scientific Conference on Information, Communication and Energy Systems and Technologies(ICEST)*, Ohrid, North Macedonia, 2025, pp. 1-4, DOI: 10.1109/ICEST66328.2025.11098263

[17] S. Berkani, I. Gryech, M. Ghogho, B. Guermah and A. Kobbane, "Data Driven Forecasting Models for Urban Air Pollution: MoreAir Case Study", *IEEE Access*, vol. 11, pp. 133131-133142, 2023, DOI: 10.1109/ACCESS.2023.3331565

[18] R. Rakholia, Q. Le, B. Q. Ho, K. Vu and R. S. Carbajo, "Multi-Output Machine Learning Model for Regional Air Pollution Forecasting in Ho Chi Minh City, Vietnam", *Environment International*, vol. 173, 2023, DOI:10.1016/j.envint.2023.107848

[19] Z. Zhang, C. Johansson, M. Engardt, M. Stafoggia and X. Ma, "Improving 3-day Deterministic air Pollution Forecasts using Machine Learning Algorithms", *Atmospheric Chemistry and Physics*, vol. 24, no. 2, pp. 807-851, DOI: 10.5194/acp-24-807-2024

[20] J. Kalajdjieski, K. Trivodaliev, G. Mirceva, S. Kalajdziski and S. Gievska, "A Complete Air Pollution Monitoring and Prediction Framework", *IEEE Access*, vol. 11, pp. 88730-88744, DOI: 10.1109/ACCESS.2023.3251346

[21] I. Jairi, S. Ben-Othman, L. Canivet and H. Zgaya-Biau, "Enhancing Air Pollution Prediction: A Neural Transfer Learning Approach Across Different Air Pollutants", *Environmental Technology & Innovation*, vol. 36, 2024, DOI: 10.1016/j.eti.2024.103793

[22] Y. Zhang, Q. Sun, J. Liu and O. Petrosian, "Long-Term Forecasting of Air Pollution Particulate Matter (PM2.5) and Analysis of Influencing Factors", *Sustainability*, vol. 16, no. 1, 2024, DOI: 10.3390/su16010019

[23] T. D. Akinosho, M. Bilal, E. T. Hayes, A. Ajayi, A. Ahmed and Z. Khan, "Deep Learning-Based Multi-Target Regression for Traffic-Related Air Pollution Forecasting", *Machine Learning with Applications*, vol. 12, 2023, DOI: 10.1016/j.mlwa.2023.100474

[24] V. Oliveira Santos, P. A. C. Rocha, J. Scott, J. Van Griensven Thé and B. Gharabaghi, "Spatiotemporal Air Pollution Forecasting in Houston-TX: A Case Study for Ozone Using Deep Graph Neural Networks", *Atmosphere*, vol. 14, 2023, DOI: 10.3390/atmos14020308

[25] T. Handhayani, "An Integrated Analysis of Air Pollution and Meteorological Conditions in Jakarta", *Scientific Reports*, vol. 13, no. 1, 2023, DOI: 10.1038/s41598-023-32817-9

[26] M. J. Jiménez-Navarro, M. Lovrić, S. Kecorius, E. K. Nyarko and M. Martínez-Ballesteros, "Explainable Deep Learning on Multi-Target Time Series Forecasting: An Air Pollution use case", *Results in engineering*, vol. 24, 2024, DOI: 10.1016/j.rineng.2024.103290

[27] J. Praveenchandar, K. Venkatesh, B. Mohanraj, M. Prasad and R. Udayakumar, "Prediction of Air Pollution Utilizing an Adaptive Network Fuzzy Inference System with the aid of Genetic Algorithm", *Natural and engineering sciences*, vol. 9, no. 1, pp. 46-56, 2024, DOI: 10.28978/nesciences.1489228

[28] H. A. Nguyen, Q. P. Ha, H. Duc, M. Azzi, N. Jiang, X. Barthelemy and M. Riley, "Long Short-Term Memory Bayesian Neural Network for Air Pollution Forecast", *IEEE Access*, vol. 11, pp. 35710-35725, 2023, DOI: 10.1109/ACCESS.2023.3265725

[29] S. Baniasadi, R. Salehi, S. Soltani, D. Martín, P. Pourmand and E. Ghafourian, "Optimizing Long Short-Term Memory Network for Air Pollution Prediction using a Novel Binary Chimp Optimization Algorithm", *Electronics*, vol. 12, no. 18, 2023, DOI: 10.3390/electronics12183985

[30] F. Fauzi, R. Wasono and I. Kharisudin, "Evaluating Recurrent Neural Networks and Long Short-Term Memory for Air Pollution Forecasting: Mitigating the Impact of Volatile Environmental Factors", *Commun. Math. Biol. Neurosci.*, 2023, Article ID 129

[31] Z. He and Q. Guo, "Comparative Analysis of Multiple Deep Learning Models for Forecasting Monthly Ambient PM2.5 Concentrations: A case Study in Dezhou City, China", *Atmosphere*, vol. 15, no. 12, 2024, DOI: 10.3390/atmos15121432

[32] J. Luo and Y. Gong, "Air Pollutant Prediction Based on ARIMA-WOA-LSTM Model", *Atmospheric Pollution Research*, vol. 14, no. 6, 2023, DOI: 10.1016/j.apr.2023.101761

[33] W. Ding and X. Qie, "Prediction of Air Pollutant Concentrations via RANDOM Forest Regressor Coupled with Uncertainty Analysis—A Case Study in Ningxia", *Atmosphere*, vol. 13, no. 6, 2022,DOI: 10.3390/atmos13060960

[34] G. Corani, "Air Quality Prediction in Milan: Feed-Forward Neural Networks, Pruned Neural Networks and Lazy Learning", *Ecological Modelling*, vol. 185, no. 2–4, pp. 513-529, 2005, DOI: 10.1016/j.ecolmodel.2005.01.008

[35] T. Jayanth, A. Manimaran, V. R. K. Reddy and N. Rajashekar, "Enhancing Air Quality Prediction Through Holt–Winters Smoothing and Transformer-BiGRU with Bayesian Optimization", *IEEE Access*, vol. 13, pp. 180756-180780, 2025, DOI:10.1109/ACCESS.2025.3621231