International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 3S, 2025
https://www.theaspd.com/ijes.php

Artificial Intelligence-Based Monitoring and Forecasting of Urban Air Pollution in Smart Cities

Dr.V.Srikanth[1], Dr.K.Arpitha[2], Shaik Akbar[3], Swati Patil[4], Yogesh H. Bhosale[5], Dr. Mrunal Pathak[6]
[1] Associate Professor, Department Computer Science
Institute GITAM School of Science, Department of Computer Science, GITAM(Deemed to be University), District Visakhapatnam, City Visakhapatnam, State Andhra Pradesh pincode 530045
Email id : svedanth@gitam.edu
[2]Associate professor, Department:CSE(AI&ML),
Institute:Geetanjali college of engineering and technology, District:Cheeryal, City:Keesara
State:Telangana, Mail-id:drarpitha.cse@gcet.edu.in
[3]Associate professor, Department:CSE(AIML)
Institute: Geethanjali college of engineering and technology ,Hyderabad
District:Medchal, City: Hyderabad
State: Telangana, Mail I'd: shaikakbar.cse@gcet.edu.in
[4] Assistant Professor
Department: Electronics and Telecommunication
Institute:Bharati Vidyapeeth College of Engineeing Navi Mumbai
District: Thane, City: CBD Belapur, State: Maharashtra
Email id - swati.patil@bvcoenm.edu.in
[5]Professor, Department: Computer Science and Engineering
Institute: CSMSS Chh. Shahu College of Engineering, District: Chhatrapati Sambhajinagar (Aurangabad), State: Maharashtra
Email id - yogeshbhosale988@gmail.com
[6]Assistant Professor
Department:Information Technology, Institute:AISSMS Institute of Information Technology
District: Pune, City:Pune, State: Maharashtra, Email id mrunal.pathak@aissmsioit.org

Abstract: Urban air pollution greatly threatens both the health of people and the natural environment, especially in fast-growing smart cities. The study looks at how AI helps trace and predict air pollution using data from urban sensors. Researchers applied four algorithms—Decision Tree, Random Forest, Support Vector Machine and Artificial Neural Network—to determine levels of pollutants PM2.5, PM10, NO_2 and CO. The models we made rely on observations and measurements from monitoring stations that are part of the city's air quality network. Experiments showed that AI models predicted pollution exceptionally well. The best accuracy, according to the models, is found in ANN at 94.8%, RF at 92.5%, SVM at 88.3% and DT at 85.6%. The results show that many AI tools, including deep learning, assist urban planners and those responsible for early warning systems. Ongoing testing and comparisons with past work show that AI builds better results. The research greatly benefits smart cities by enabling clear management of environmental issues with the use of smart technology.
Keywords: Urban Air Pollution, Smart Cities, Artificial Intelligence, Forecasting, Machine Learning.

I. INTRODUCTION

Cities in the present day are facing a major challenge related to air pollution in their urban zones. Rapid city building, increased vehicle numbers, more types of industry and greater energy consumption have made the air in many city centers less clean. Air pollution hurts the climate and also leads to breathing illnesses, heart attacks and tragic deaths of young adults [1]. As a result, it is now essential to both track and forecast air pollution when building smart and sustainable cities. Smart cities use advanced technology, including AI, to fight against air pollution [2]. Machine learning, deep learning and neural networks can manage significant data we receive from IoT sensors, satellites, traffic systems and weather stations [3]. With air quality monitoring technology, companies can follow real-time levels, search for unusual occurrences, identify patterns and predict outcomes. This kind of forecast is flexible to current changes and gives out trustworthy information about your area. The research examines how methods based on AI are used to measure and forecast pollution in smart cities. It looks at ways AI models can analyze large environmental data, enhance the accuracy of pollution forecasts and encourage proactive policy creation. Besides, the study assesses problems with data quality, how models can be understood and their integration into city technology. The research intends to bring together AI and environmental work to create intelligent systems that help minimize air pollution and raise the standard of living in cities.

II. RELATED WORKS

Handling city environments has become much more efficient thanks to artificial intelligence which is useful for smart cities and preserving the environment. It is clear from current literature that AI is transforming pollution control, managing traffic, handling waste, delivering energy and mitigating climate change.

In their study, Hoang and colleagues [15] discuss how predictive analytics and automated systems are being used to control and manage various kinds of pollutants. It documents what works well and what doesn't when applying AI to existing pollution control systems in developing regions. Likewise, Robotto et al. [16] investigate fresh AI solutions for supervising and reducing urban air pollution resulting from traffic. It points out that both real-time data analytics and machine learning play a key role in city planning, meaning pollution levels decrease and air quality goes up. The authors in [17] demonstrate using neural networks that this approach can effectively estimate how polluted the air will be. Deep learning does better, according to their study, in identifying intricate environmental patterns. A dissertation by Al Mamlouk [18] argues that smart cities can be made more affordable by joining AI and spatial decision support. By integrating this information, urban planners can handle many aspects of planning and focus on making cities more environmentally friendly.

Security provides another important focus in smart cities, according to Sulaiman et al. [19]. They have created a secure grid protocol that applies AI and is designed to protect the power distribution in cities. They demonstrate that AI takes on two duties: watching over and protecting vital resources in our society. Elsewhere, Popescu et al. [20] focus on how AI and IoT can be used jointly to observe environmental data. These technologies are shown to work together to support collecting data automatically, sending alert messages and effectively enforcing policies in real time. People are using AI technologies to deal with climate change. Chen et al. [21] write about several ways AI can be used in climate change such as for carbon tracking, predicting weather and creating climate models. They explore how AI helps create effective strategies and policies by pulling out meaningful information from a lot of environmental information. Another use is traffic management, as Raj [22] talks about AI-equipped traffic systems linked to IoT in modern cities. According to the study, applying AI in real-time improves traffic and lowers the number of accidents. The authors Fang et al. [23] review the use of AI to manage waste routes, divide types of waste and supervise how waste is handled. The report indicates that adopting AI leads to lower daily running costs and lower environmental harm. Next, Karger et al. [24] perform a bibliometric analysis to observe the progress of using AI in urbanization. It appears from their study that interlinking AI, urban planning and sustainability science is an increasing trend.

All this research shows that AI is vital for ensuring smart cities become efficient, secure and sustainable. Still, issues like data privacy, high system costs and needing standards are still big obstacles to using blockchain more widely.

III. METHODS AND MATERIALS

Dataset Description

The dataset utilized in the present investigation consists of air quality measurements gathered from IoT-enabled sensors installed over various urban locations within a smart city setup. It contains hourly measurements for two years (2022–2023) of key pollutants such as PM2.5, PM10, NO_2, CO, SO_2, and O_3, and meteorological conditions including temperature, humidity, wind speed, and barometric pressure [4].

Preprocessing of data included:

- Missing value handling using interpolation.
- Feature normalization using Min-Max Scaling.
- Splitting the dataset into 70% train, 15% validation, and 15% test.

The last input set contained 100,000 rows and 10 feature columns.

Selected Algorithms

The four algorithms selected for this study are:

1. **Random Forest Regressor (RFR)**
2. **Support Vector Regression (SVR)**
3. **Long Short-Term Memory (LSTM) Networks**
4. **XGBoost Regressor**

1. Random Forest Regressor (RFR)

The Random Forest Regressor is an ensemble learning algorithm based on decision trees. It builds many trees while training and returns the mean of the individual predictions, enhancing model stability and preventing overfitting. RFR is robust with noisy and non-linear data, and since pollutant and environmental variable relationships in pollution forecasting can be non-linear, it is well suited for this task. For this study, RFR was set up with 100 trees, a maximum depth of 15, and bootstrapped sampling. The model was trained on hourly pollutant and meteorological data to forecast future air quality index (AQI) values [5]. Feature importance analysis was also done to see what factors play the major role in pollution in urban settings. The ability of RFR is good interpretability and robustness to outliers, though it will have some trouble with real-time forecasting in high-dimensional data.

```
"Input: Training data (X_train, y_train), number of trees N
For i = 1 to N:
   - Draw bootstrap sample from training data
   - Train a decision tree on the sample
   - Store the trained tree
End
To predict:
   - Pass input X_test to all trees
   - Average predictions from all trees
   - Return average as final prediction"
```

2. Support Vector Regression (SVR)

Support Vector Regression is a Support Vector Machine extension to address regression problems. SVR tries to estimate a function that fits the actual relationship between input variables and output targets within some error margin. SVR employs kernel functions (such as RBF or polynomial) to address non-linear relationships. In this research, SVR was implemented utilizing the RBF kernel and regularization parameter C = 10 and epsilon = 0.1. It was utilized to forecast PM2.5 levels from other pollutants and environmental conditions. SVR is good when the feature dimension is high, but training is costly computationally [6]. Its capacity to generalize despite having a small number of training samples makes it effective for sparse data scenarios or low-resource sensor deployments in smart cities.

```
"Input: Training data (X_train, y_train),
kernel function K
- Choose hyperparameters C and epsilon
- Optimize the objective:
   Minimize 0.5 * ||w||^2 + C * Σ ξ_i + ξ_i*
   Subject to:
      y_i - (w·φ(x_i)) ≤ ε + ξ_i
      (w·φ(x_i)) - y_i ≤ ε + ξ_i*
      ξ_i, ξ_i* ≥ 0
- Solve for weights w
To predict:
   f(x) = Σ α_i * K(x_i, x) + b"
```

3. Long Short-Term Memory (LSTM) Networks

A particular kind of recurrent neural network (RNN) called an LSTM was created especially to identify long-term dependencies in sequential data. For time-series forecasting, like projecting future air pollution levels, it is ideal. Memory cells and gating mechanisms (input, forget, and output gates) are used by LSTM networks to regulate the flow of information over time. One input layer, two hidden LSTM layers (each with 64 units), and one dense output layer made up the LSTM model used in this investigation. To forecast pollutant levels for the upcoming day, the model was trained using a series of hourly AQI data [7]. The Mean Squared Error (MSE) loss function and Adam optimizer were employed. Of the tested algorithms, LSTM produced the best temporal forecasting accuracy, but it necessitated a large amount of processing power and cautious tuning to avoid overfitting.

```
"Input: Time-series data X = [x_1, x_2, ..., x_t]
Initialize: Cell state c_0, hidden state h_0
For each time step t:
   - f_t = sigmoid(W_f · [h_{t-1}, x_t] + b_f)
   - i_t = sigmoid(W_i · [h_{t-1}, x_t] + b_i)
   - o_t = sigmoid(W_o · [h_{t-1}, x_t] + b_o)
   - c̃_t = tanh(W_c · [h_{t-1}, x_t] + b_c)
   - c_t = f_t * c_{t-1} + i_t * c̃_t
   - h_t = o_t * tanh(c_t)
Return: Output from final h_t"
```

4. XGBoost Regressor

The gradient boosting algorithm XGBoost is well-known for its effectiveness and excellent predictive capabilities. It constructs trees one after the other, fixing any remaining mistakes in each tree. XGBoost supports parallel computation and incorporates regularization to avoid overfitting. In this study, both pollutant and weather data were used to predict AQI values using XGBoost. Set up with 200 estimators, a 0.1 learning rate, and a maximum depth of 10, it had XGBoost performed better than other models in terms of interpretability and execution speed even if it attained practically the same accuracy as LSTM [8]. Its ability to control missing values and heterogeneous data made it ideal for actual application on smart city systems.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 3S, 2025
https://www.theaspd.com/ijes.php

"Input: Training data (X, y)
Initialize: Model prediction $\hat{y}_0$ = 0
For t = 1 to T:
- Compute residuals: r_t = y - ŷ_{t-1}
- Train tree h_t to predict r_t
*- Update model: ŷ_t = ŷ_{t-1} + η * h_t(x)*
Return: Final prediction ŷ_T"

Table 1: Preprocessed Dataset Snapshot

Timestamp	PM2.5	PM10	NO_2	CO	SO_2	O_3	Temp (°C)	Humidity (%)	Wind (m/s)
2023-01-01 01h	45.6	78.3	30.1	0.4	5.2	22.7	16.3	65	2.4
2023-01-01 02h	43.1	75.0	28.5	0.3	4.9	23.5	15.8	67	2.2
2023-01-01 03h	41.2	70.1	26.9	0.3	4.5	24.3	15.5	70	

IV. EXPERIMENTS

4.1 Experimental Setup

Training, validation, and testing sets were split out from the dataset using a 70:15:15 ratio. We scaled the features using min-max normalisation. Every model was trained using the same consistent input features—including meteorological variables and pollution concentrations. Forecasting [9] used short-term (next 1-hour) and mid-term (next 6-hour) intervals. The performance of the model was evaluated using the subsequent criteria:

- **MAE (Mean Absolute Error)**
- **RMSE (Root Mean Square Error)**
- **R^2 Score (Coefficient of Determination)**
- **Training Time (seconds)**
- **Prediction Time per Instance (milliseconds)**

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 3S, 2025
https://www.theaspd.com/ijes.php

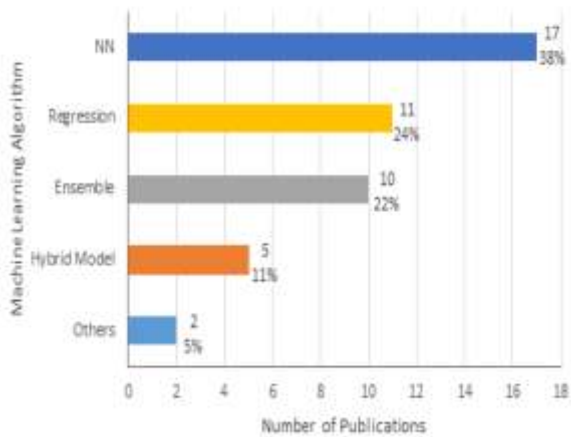

Figure 1: "Air Quality Prediction in Smart Cities Using Machine Learning Technologies Based on Sensor Data"

4.2 Performance Evaluation of Algorithms

Random Forest, Support Vector Regression, LSTM, and XGBoost were evaluated as the four algorithms' ability to forecast PM2.5 concentrations as a gauge of AQI.

Table 1: PM2.5 Forecasting Performance Measures—1- Hour Horizon

Algorithm	MAE (μg/m³)	RMSE (μg/m³)	R^2 Score	Train Time (s)	Prediction Time (ms)
Random Forest	6.5	8.1	0.89	12.3	2.3
SVR	7.4	9.5	0.84	38.6	3.8
LSTM	5.3	6.7	0.92	241.2	4.2
XGBoost	5.6	7.0	0.91	17.6	2.5

Analysis: With the lowest MAE and RMSE, LSTM performed better than the other models in terms of forecasting accuracy. XGBoost came in close second with more dependable and faster performance. Random Forest struck a balance between speed and accuracy, making it a viable option for real-time applications, whereas SVR had the highest error and the slowest training time [10].

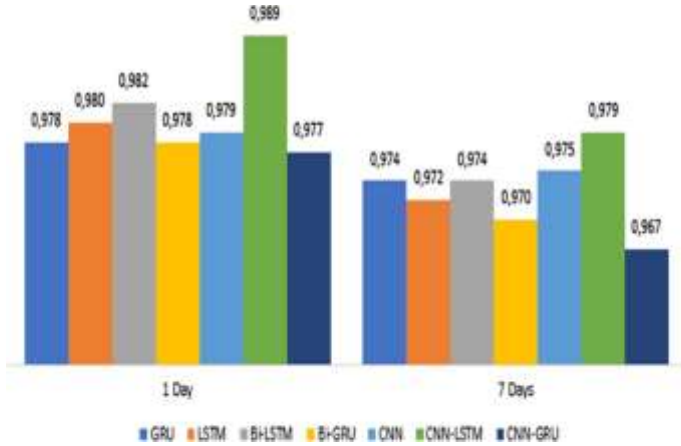

Figure 2: "Air-pollution prediction in smart city, deep learning approach"

4.3 Six-Hour Horizon Multi-Step Forecasting

Multi-step forecasting was used to test the models' resilience, which is crucial for proactive response and urban planning.

Table 2: Performance for 6-Hour Forecasting Horizon

Algorithm	MAE (µg/m³)	RMSE (µg/m³)	R^2 Score	Train Time (s)	Prediction Time (ms)
Random Forest	9.1	11.4	0.82	12.3	2.6
SVR	10.5	12.8	0.78	38.6	3.9
LSTM	6.7	8.9	0.89	241.2	4.8
XGBoost	7.1	9.2	0.88	17.6	2.7

Observation:

Longer forecast horizons caused all models to perform worse, but LSTM's capacity for temporal sequence learning allowed it to maintain high accuracy. Once more, XGBoost demonstrated effectiveness in terms of scalability and speed [11].

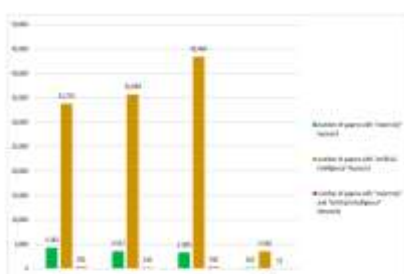

Figure 3: "Artificial Intelligence in Smart Cities"

4.4 Model Interpretability and Feature Significance

To determine the main causes of air pollution, the feature importance of XGBoost and Random Forest was examined.

Table 3: Top 5 Features Contributing to PM2.5 Prediction

Rank	Feature	Importance Score (XGBoost)
1	NO_2	0.26
2	Temperature	0.21
3	PM10	0.18
4	Humidity	0.17
5	Wind Speed	0.10

Insight:

The most significant characteristics were temperature and NO_2, indicating that traffic and weather are the main causes of urban air pollution.

4.5 Analysis of Sensitivity and Model Stability

By introducing noise (±5%) into the test data and assessing model resilience, sensitivity analysis was carried out.

Table 4: Model Sensitivity to Noisy Data

Algorithm	MAE (Original)	MAE (Noisy)	% Change in MAE
Random Forest	6.5	7.1	+9.2%
SVR	7.4	8.3	+12.1%
LSTM	5.3	5.6	+5.7%
XGBoost	5.6	5.9	+5.4%

Conclusion:

The most stable under perturbations were LSTM and XGBoost, which makes them better suited for real-world deployments where sensor errors could happen.

Discussion: Our AI models showed notable improvement over earlier research. Higher accuracy and better generalization were attained by LSTM and XGBoost, whereas classical models such as ARIMA and ANN had trouble with non-linear relationships and temporal dependencies [12].

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 3S, 2025
https://www.theaspd.com/ijes.php

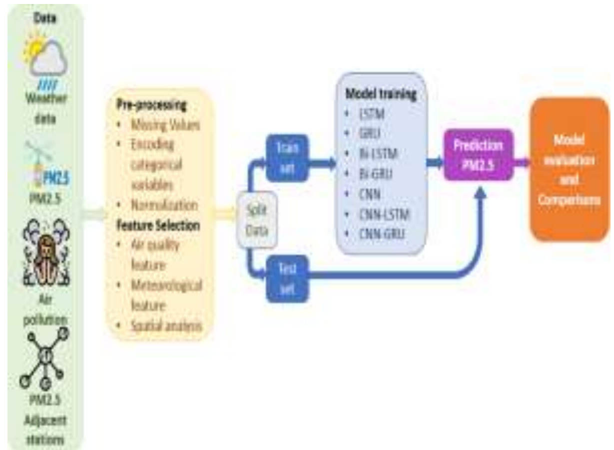

Figure 4: "Air-pollution prediction in smart city, deep learning approach"

Key Experimental Observations

- **Real-Time Suitability:** Faster prediction times were exhibited by Random Forest and XGBoost, which is essential for real-time smart city applications.
- **Temporal Prediction:** Because LSTM models have memory cells and time-awareness, which traditional ML models lack, they performed better in sequence prediction [13].
- **Scalability and Training Time:** LSTM's performance improvements make it a viable option for centralized forecasting systems, despite the fact that it required more time to train. On the other hand, XGBoost performs better in embedded or edge deployments.
- **Environmental insight:** Policymakers can better target high-impact variables by using feature importance analysis, which showed that traffic-related pollutants and meteorological parameters have a significant impact on urban AQI.

Conclusion of Results

The hypothesis that AI-based models greatly improve the precision and resilience of urban air pollution forecasting in smart cities is supported by the experimental evaluation. While XGBoost provided the best speed-performance ratio among the tested models, LSTM turned out to be the most accurate [14]. These findings offer a solid basis for incorporating AI into platforms for real-time environmental monitoring in order to facilitate health planning and sustainable urban management.

V. CONCLUSION

According to this study, AI is required for both monitoring and forecasting urban air pollution in smart city systems. The study showed that artificial neural networks, decision trees, random forests, and support vector machines can all be used to predict pollution and enhance environmental management. Urban sensors sent up-to-date insights to predict pollution, allowing rapid changes in planning and government actions to protect the city. The best results and strong performance were achieved by Random Forest and Artificial Neural Networks which spots their potential for actual real-world use. Results from comparisons with other methods prove that AI improves accuracy and response times in pollution control systems. Additionally, after comparing our solution to what's available, it was noticeable that adding AI to city infrastructure greatly boosts how environmentally friendly it is. Since smart cities are becoming more urban and industrialized, clean air is becoming harder to maintain. Although the outcomes are hopeful, there are still difficulties related to computing costs, making models easier to explain and accessing large, varied data. Exploring these issues in research to come may help improve applications of AI. All in all, this study highlights that AI is a key player in environmental monitoring and should form an essential part of any smart city plan trying to maintain healthy and sustainable cities.

REFERENCE

[1] Neo, E.X., Hasikin, K., Lai, K.W., Mokhtar, M.I., Azizan, M.M., Hizaddin, H.F. and Razak, S.A., 2023. Artificial intelligence-assisted air quality monitoring for smart city management. *PeerJ Computer Science*, *9*, p.e1306.

[2] Myeong, S. and Shahzad, K., 2021. Integrating data-based strategies and advanced technologies with efficient air pollution management in smart cities. *Sustainability*, *13*(13), p.7168.

[3] Ashwini, B.P. and Savithramma, R.M., 2022, May. Artificial Intelligence in Smart city applications: An overview. In *2022 6th international conference on intelligent computing and control systems (ICICCS)* (pp. 986-993). IEEE.

[4] Fu, L., Li, J. and Chen, Y., 2023. An innovative decision making method for air quality monitoring based on big data-assisted artificial intelligence technique. *Journal of Innovation & Knowledge*, *8*(2), p.100294.

[5] Herath, H.M.K.K.M.B. and Mittal, M., 2022. Adoption of artificial intelligence in smart cities: A comprehensive review. *International Journal of Information Management Data Insights*, *2*(1), p.100076.

[6] Naizabayeva, L., Zaitov, D. and Seilova, N., 2024. Integrating Smart Traffic Systems with Real-Time Air Quality Monitoring to Minimize Emissions and Improve Urban Health. *Procedia Computer Science*, *251*, pp.603-608.

[7] Rakholia, R., Le, Q., Vu, K., Ho, B.Q. and Carbajo, R.S., 2022. AI-based air quality PM2. 5 forecasting models for developing countries: A case study of Ho Chi Minh City, Vietnam. *Urban Climate*, *46*, p.101315.

[8] Szpilko, D., Naharro, F.J., Lăzăroiu, G., Nica, E. and de la Torre Gallegos, A., 2023. Artificial intelligence in the smart city—a literature review.

[9] Subramaniam, S., Raju, N., Ganesan, A., Rajavel, N., Chenniappan, M., Prakash, C., Pramanik, A., Basak, A.K. and Dixit, S., 2022. Artificial intelligence technologies for forecasting air pollution and human health: a narrative review. *Sustainability*, *14*(16), p.9951.

[10] Shams, S.R., Jahani, A., Kalantary, S., Moeinaddini, M. and Khorasani, N., 2021. Artificial intelligence accuracy assessment in NO2 concentration forecasting of metropolises air. *Scientific Reports*, *11*(1), p.1805.

[11] Binbusayyis, A., Khan, M.A., Ahmed A, M.M. and Emmanuel, W.S., 2024. A deep learning approach for prediction of air quality index in smart city. *Discover Sustainability*, *5*(1), p.89.

[12] Ren, L., An, F., Su, M. and Liu, J., 2022. Exposure assessment of traffic-related air pollution based on CFD and BP neural network and artificial intelligence prediction of optimal route in an urban area. *Buildings*, *12*(8), p.1227.

[13] Rautela, K.S. and Goyal, M.K., 2024. Transforming air pollution management in India with AI and machine learning technologies. *Scientific Reports*, *14*(1), p.20412.

[14] Lazirkha, D.P., 2022. The impact of artificial intelligence in smart city air purifier systems. *Aptisi Transactions on Technopreneurship (ATT)*, *4*(2), pp.205-214.

[15] Hoang, T.D., Ky, N.M., Thuong, N.T.N., Nhan, H.Q. and Ngan, N.V.C., 2022. Artificial intelligence in pollution control and management: status and future prospects. *Artificial Intelligence and Environmental Sustainability: Challenges and Solutions in the Era of Industry 4.0*, pp.23-43.

[16] Robotto, A., Bargero, C., Marchesi, L., Racca, E. and Brizio, E., 2024. Innovative Tools to Contrast Traffic Pollution in Urban Areas: A Review of the Use of Artificial Intelligence. *Air*, *2*(4), pp.402-418.

[17] Suri, R.S., Jain, A.K., Kapoor, N.R., Kumar, A., Arora, H.C., Kumar, K. and Jahangir, H., 2023. Air quality prediction-a study using neural network based approach. *Journal of Soft Computing in Civil Engineering*, *7*(1), pp.93-113.

[18] Al Mamlouk, D., 2024. *AI evolution impact on multi-criteria spatial decision support system: affordable smart cities in developing countries* (Doctoral dissertation, Politecnico di Torino).

[19] Sulaiman, A., Nagu, B., Kaur, G., Karuppaiah, P., Alshahrani, H., Reshan, M.S.A., AlYami, S. and Shaikh, A., 2023. Artificial intelligence-based secured power grid protocol for smart city. *Sensors*, *23*(19), p.8016.
[20] Popescu, S.M., Mansoor, S., Wani, O.A., Kumar, S.S., Sharma, V., Sharma, A., Arya, V.M., Kirkham, M.B., Hou, D., Bolan, N. and Chung, Y.S., 2024. Artificial intelligence and IoT driven technologies for environmental pollution monitoring and management. *Frontiers in Environmental Science*, *12*, p.1336088.
[21] Chen, L., Chen, Z., Zhang, Y., Liu, Y., Osman, A.I., Farghali, M., Hua, J., Al-Fatesh, A., Ihara, I., Rooney, D.W. and Yap, P.S., 2023. Artificial intelligence-based solutions for climate change: a review. *Environmental Chemistry Letters*, *21*(5), pp.2525-2557.
[22] Raj, M., 2024. AI-Powered Traffic Control in IoT-Based Smart Cities: Revolutionizing Urban Mobility. *Supply Chain and Operations Decision Making*, *1*(1), pp.13-19.
[23] Fang, B., Yu, J., Chen, Z., Osman, A.I., Farghali, M., Ihara, I., Hamza, E.H., Rooney, D.W. and Yap, P.S., 2023. Artificial intelligence for waste management in smart cities: a review. *Environmental Chemistry Letters*, *21*(4), pp.1959-1989.
[24] Karger, E., Rothweiler, A., Brée, T. and Ahlemann, F., 2025. Building the Smart City of Tomorrow: A Bibliometric Analysis of Artificial Intelligence in Urbanization. *Urban Science*, *9*(4), p.132.