



Integrating Satellite Data and Artificial Intelligence for Air Quality Prediction and Mitigation Strategies: A Case Study of Tehran

Mohammad Amin Javadi¹ | Armin Mollaei Rudsary² | Zahra Amiri³ | Maryam Zare Shahne^{1✉} | Zahra Barzegar⁴ | Naghmeh Mobarghaee Dinan⁵

1. Department of Civil and Environmental Engineering, K.N. Toosi University of Technology, Tehran, Iran

2. Department of Civil and Environmental Engineering, Sharif University of Technology, Tehran, Iran

3. Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario, Canada

4. Director of Environmental Studies, Tehran Urban Research and Planning Centre (TURPC), Tehran 1964635611, Iran

5. Department of Environmental Planning and Design, Environmental Sciences Research Institute, Shahid Beheshti University, Tehran, Iran

Article Info

Article type:

Research Article

Article history:

Received: 14 April 2025

Revised: 20 October 2025

Accepted: 11 January 2026

Keywords:

Air pollution

Remote sensing

Google earth engine

Machine learning

ABSTRACT

Air pollution is a critical challenge in urban centers like Tehran, where over 8 million residents are exposed to pollutants from transportation, industry, and energy use. To address this, researchers combine satellite observations (e.g., Sentinel-5P) with AI models to monitor and predict concentrations of pollutants such as PM_{2.5}, PM₁₀, O₃ and NO₂. By integrating remote sensing data with ground measurements, machine learning methods—including neural networks, decision trees, and regression models—establish links between meteorological conditions and pollution levels. This hybrid approach overcomes the limitations of traditional monitoring systems while benefiting from tools like Google Earth Engine for efficient analysis of Tehran's air quality (2019–2024). The resulting forecasts provide policymakers with actionable insights for pollution control, urban planning, and public health strategies.

Cite this article: Javadi, M.A., Mollaei Rudsary, A., Amiri, Z., Zare Shahne, M., Barzegar, Z., & Mobarghaee Dinan, N. (2026). Integrating Satellite Data and Artificial Intelligence for Air Quality Prediction and Mitigation Strategies: A Case Study of Tehran. *Pollution*, 12(1), 58-70.

<https://doi.org/10.22059/poll.2025.393453.2885>



© The Author(s).

Publisher: The University of Tehran Press.

DOI: <https://doi.org/10.22059/poll.2025.393453.2885>

INTRODUCTION

Air pollution poses a major global threat to public health, the environment, and the economy, making air quality monitoring essential (Heger, Sarraf & Heger, 2018). With the world population projected to reach 9.7 billion by 2050 (United Nations, 2022), the risks associated with pollution are expected to intensify. Air pollution levels vary widely even within small areas, making ground station-based models insufficient for accurately capturing dispersion in cities. While adding more stations could improve accuracy, data collection is limited by high costs and other constraints (Järvi et al., 2023).

Ground-based monitoring provides accurate local data but faces major drawbacks, including limited coverage, high costs, and vulnerability to equipment issues (Javan, Samadzadegan & Toosi, 2025). Statistical models and interpolation partly address these gaps but often misrepresent reality (Akima, 1970). Remote sensing overcomes these limitations by offering

*Corresponding Author Email: maryam_Zare@kntu.ac.ir

wide spatial coverage, frequent observations, and access to hard-to-reach regions. Integrating ground and satellite data creates a more accurate and cost-effective system, improving spatial and temporal resolution for informed environmental policies and public health decisions (Ibrahim et al., 2024; Kafi et al., 2024; Javadi, Shahne & Amiri, 2025). Ground-based measurements offer high temporal resolution but limited spatial coverage, particularly in megacities with diverse microclimates (Bourbour, Zare Shahne & Safarkhanloo, 2025). Satellites provide broad coverage but lack precision for local estimates without ground validation. To address these limitations, this study applies a hybrid model integrating satellite variables with ground-based meteorological and air quality data. Building on prior research, this approach combines the strengths of both sources, creating a more accurate and scalable framework for predicting pollution in Tehran, where monitoring infrastructure is sparse and environmental variability is high (Kafi et al., 2024; Bourbour, Zare Shahne & Safarkhanloo, 2025).

Conventional environmental monitoring faces major challenges, chiefly high costs of manual sampling and laboratory analysis, which require skilled staff, equipment, and chemicals (Dressing & Meals, 2010; Ditria et al., 2022). These constraints often limit programs to small samples and narrow focus, preventing a full picture of environmental conditions. Traditional methods are also slow, with results taking weeks or months, delaying decisions in crises (Thomson et al., 2011; Ceccato et al., 2014). Human observation adds subjectivity and potential error, reducing accuracy (Hameed et al., 2017). Furthermore, advanced technologies and trained personnel are often unavailable due to high costs and resource limitations, making regular monitoring difficult, especially in regions lacking infrastructure (Cordier et al., 2021; Delbari, Zare Shahne & Hosseini, 2024).

Environmental monitoring involves systematic observation, measurement, and evaluation of natural conditions to detect harmful changes affecting ecosystems or public health (Artiola, Pepper & Brusseau, 2004). Traditional methods—statistical analysis, laboratory testing, and manual sampling—are limited by high costs, time demands, and lower accuracy (Sarker, 2021b, 2021a; Waymond, 2020). Artificial intelligence (AI) now plays a key role in overcoming these challenges by enhancing objectivity, analysing large datasets, and improving access in resource-limited areas.

AI and remote sensing (RS) have become key tools for modeling and predicting air pollution at global and regional scales (Koçak, 2025; Chadalavada et al., 2025). Machine learning (ML) methods such as random forests, support vector machines, and deep learning (DL) effectively estimate pollutant concentrations by combining satellite imagery with meteorological data. Yet, applying these techniques in Iran—especially in Tehran’s complex urban setting—faces challenges, including limited and unevenly distributed monitoring stations, restricted data access, urban heterogeneity, and complex topography, all of which limit model accuracy and generalizability (Kafi et al., 2024; Amiri & Zare Shahne, 2025; Maryam Zare, Amir & Fatemeh, 2022; Alotaibi & Nassif, 2024; Bourbour, Zare Shahne & Safarkhanloo, 2025).

Tehran faces severe air pollution driven by heavy traffic, outdated public transport, and nearby industrial zones (Arhami et al., 2017, 2018; Zare Shahne et al., 2025; Maryam Zare, Amir & Fatemeh, 2022). These anthropogenic sources are intensified by the city’s basin location, surrounded by the Alborz Mountains, which restrict pollutant dispersion. Climatic factors such as weak winds, frequent temperature inversions, and low rainfall further trap pollutants, making air quality management especially difficult (Ali-Taleshi et al., 2025; Mohammadi et al., 2012; Farr et al., 2007). RS and ML have been widely applied in air quality monitoring, with hybrid models showing strong potential (Kafi et al., 2024). For instance, ANFIS achieved over 80% accuracy in predicting O_3 in Serbia (Savić et al., 2014), while PCA-ANN models in India predicted NO_2 with R^2 values up to 0.91 (Mishra & Goyal, 2015). In Tehran, (Shams et al., 2021) found MLP neural networks ($R^2=0.89$) outperformed MLR ($R^2=0.81$) for NO_2 , highlighting the role of green spaces and time-lag effects. (Noori et al., 2010) also applied ANNs and ANFIS for

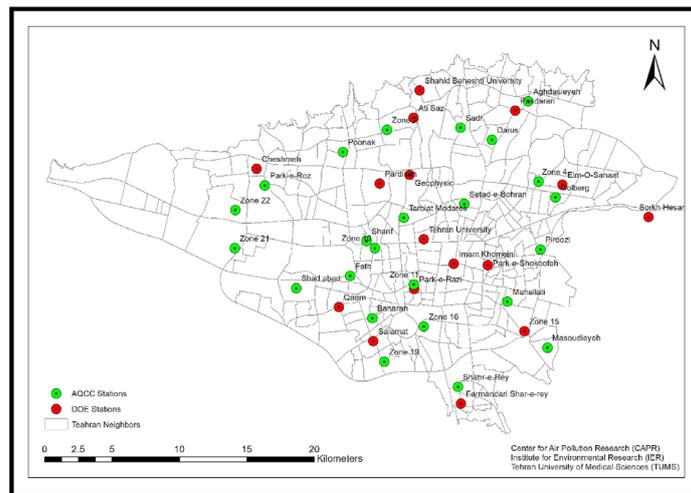


Fig. 1. The layout of air quality monitoring stations managed by the AQCC (green points) and the DOE (red points).

CO prediction, reducing computational costs. (Yeganeh et al., 2018) showed ANFIS ($R^2=0.81$) outperformed SVM and ANNs in $PM_{2.5}$ estimation for sparsely populated areas. More recently, (Bagheri, 2022) used MAIAC AOD data for $PM_{2.5}$ mapping in Tehran, achieving $R^2=0.74$ for daily predictions (Bagheri, 2022; Amiri & Zare Shahne, 2025).

Although many studies use ground-based data to study air pollution, limited spatial coverage of monitoring stations is a major challenge in large, complex cities like Tehran (Anggraini et al., 2024). Few studies in Iran have successfully combined satellite and surface data to enhance the spatial and temporal resolution of air quality assessments. Despite global advances in air quality monitoring, domestic research often relies on traditional methods or standalone ML models. This highlights the need for hybrid approaches that integrate ground, meteorological, and satellite data with AI. International studies show that models like LSTM, SVM, and Random Forest effectively capture complex atmospheric interactions and improve pollution predictions, especially when combined with high-resolution satellite imagery.

This study presents a hybrid framework combining multi-source datasets with ML and DL to model and predict air pollution in Tehran. By improving accuracy and scalability, it addresses methodological gaps in Iran and provides evidence-based insights for urban pollution management and sustainable development.

MATERIALS AND METHODS

Tehran's complex topography and geological conditions, including surrounding mountains and prevailing winds, promote pollutant accumulation by trapping photochemical smog and directing emissions from industrial areas toward the city center (Mohammadi et al., 2012).

The study was conducted in Tehran, a densely populated city with severe air pollution from traffic and industry. Its basin location surrounded by mountains traps pollutants, worsening air quality. Data from 2019–2024 were collected from DOE and AQCC ground stations, which, despite their importance, have limited and uneven coverage. Integrating these measurements with satellite observations allows for a more complete assessment of pollutant distribution, intensity, and trends.

Ground-based air quality data were obtained from Tehran's DOE and AQCC monitoring stations, measuring $PM_{2.5}$, PM_{10} , NO_2 , CO, and O_3 . Only stations with valid daily records (data on more than half of the research period) were included. These measurements were synchronized with meteorological data from the Iranian Meteorological Organization (2019–2024), including

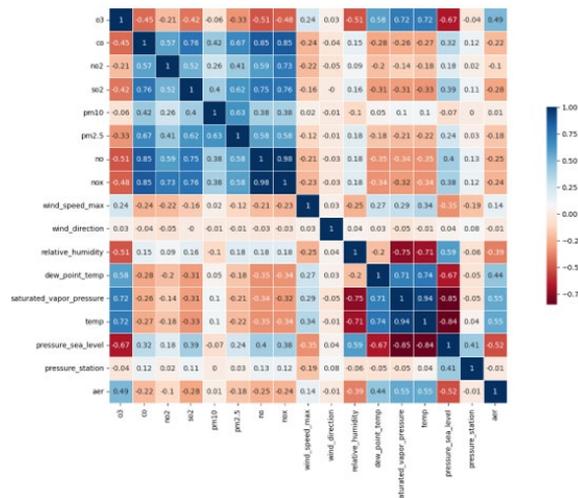


Fig. 2. Pearson's Correlation Matrix for Tarbiat Modares Station

wind speed, solar radiation, precipitation, humidity, and temperature.

Satellite data from the Sentinel-5P mission, providing ~1.1 km resolution, were used to obtain Aerosol Index, CO, NO₂, and O₃ concentrations via the Copernicus Open Access Hub. Data were preprocessed with Google Earth Engine, and daily median values were calculated. Spatial interpolation aligned satellite measurements with ground-based station locations for analysis.

ML algorithms were used to predict pollutant concentrations from integrated datasets. Missing data, however, can complicate predictions and affect result interpretation (Zhou, 2020). Missing data can also reduce the effectiveness of air quality-related public services (Yu et al., 2020). Missing data is common in air pollution measurements and occurs in other fields like medicine, energy, and traffic (Austin et al., 2021; Laña et al., 2018; Ma et al., 2020b). Data loss can result from sensor or system failures, maintenance, human error, or other common issues in the field (Pena, Ortega & Orellana, 2019; Ma et al., 2020a). Depending on the cause, missing air pollution data can occur over short intervals or extended periods (Moshenberg, Lerner & Fishbain, 2015). Short gaps in air quality data often result from routine maintenance, while sensor failures or major events can cause extended data loss. Random Forest (RF), Support Vector Regression (SVR), and Long Short-Term Memory (LSTM) networks were applied to model complex relationships between ground-level pollution and meteorological factors. The models produced accurate predictions of pollutant dispersion. The dataset was split for training and validation, and performance was evaluated using RMSE, MAE, and R², ensuring reliable predictions under varying conditions. Literature supports LSTM and SVR as effective algorithms for reliable air pollution prediction across diverse spatio-temporal contexts.

RESULTS AND DISCUSSION

This section presents the research findings, starting with statistical descriptions of meteorological and air pollution data, followed by correlations and identification of modeling variables. Results include regression analyses and the best-fit model, with comparisons between predicted and observed pollutant concentrations using metrics such as Epoch-Error, Frequency-Error, and residual plots.

Figure 2 illustrates correlations of O₃ with meteorological parameters at Piruzi Station. O₃ negatively correlates with relative humidity and positively with temperature, reflecting

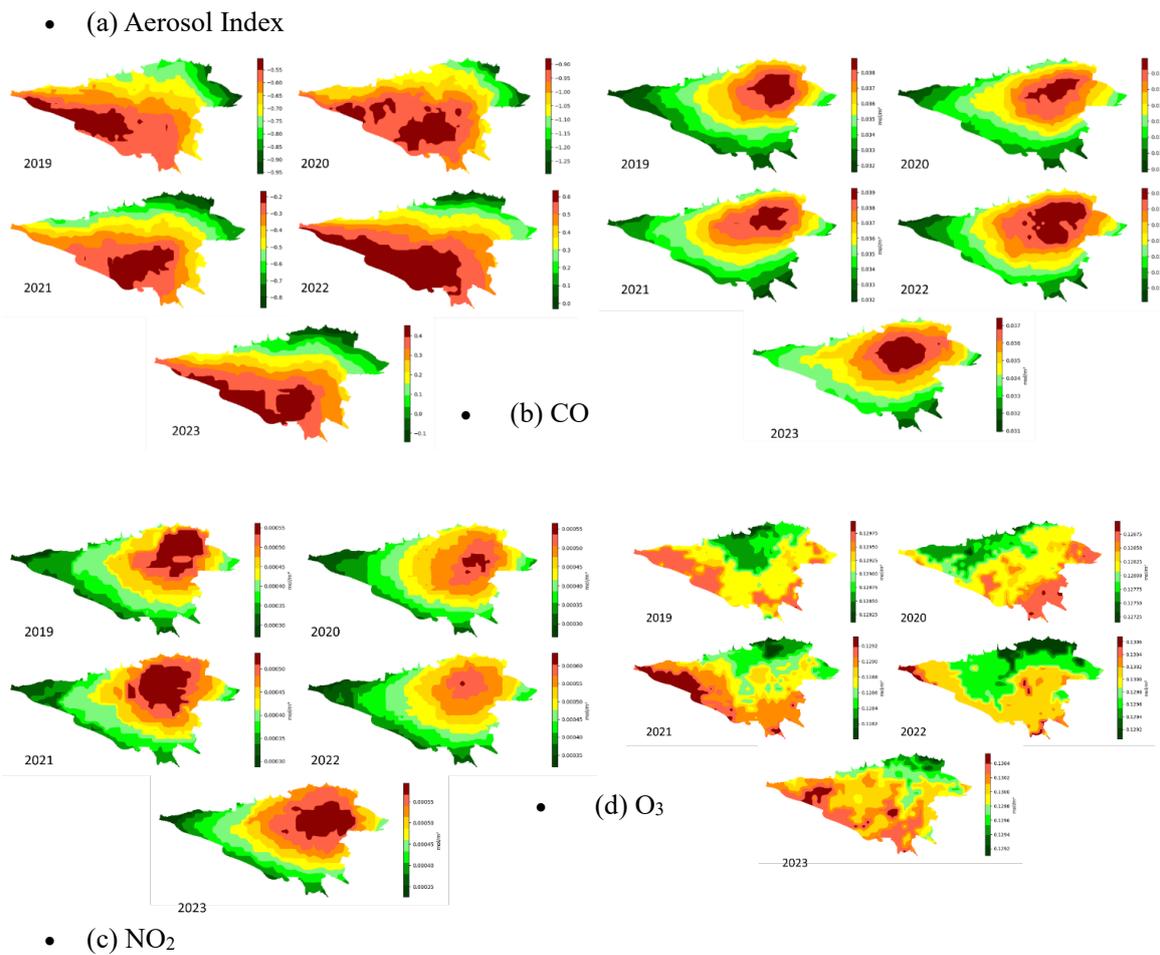


Fig. 3. Spatial distribution maps for aerosol index, CO, NO₂, and O₃.

photochemical formation influenced by solar radiation and cloud cover. Wind has minor effects on O₃, while higher pressure is associated with increased O₃. Additionally, wind direction correlates positively with humidity and negatively with temperature, and temperature and humidity are strongly inversely related.

The negative correlation between O₃ and humidity indicates that higher humidity, likely from increased cloud cover, reduces solar radiation and photochemical O₃ formation, while higher temperatures enhance O₃ production. Weak correlations with wind suggest limited pollutant dispersal, highlighting temperature and humidity as key drivers of O₃ variability in Tehran and informing targeted mitigation strategies during stagnant meteorological conditions.

Figure 3 presents processed satellite imagery of annual mean pollutant concentrations. PM_{2.5} and NO₂ peaked in 2022, with the highest aerosol index in southwestern Tehran and central/eastern regions showing elevated NO₂. CO levels were relatively uniform, slightly higher in central to northeast areas. The aerosol index indicates that southern Tehran experienced the greatest particulate pollution.

AI Modeling for Ground Data Prediction

This section illustrates AI models, including ML and DL, used to predict key pollutant concentrations. Preprocessing steps, such as imputing missing data and removing outliers, were performed before modeling. Data quality critically influences AI model accuracy. Missing values were handled using statistical or ML-based imputation, and noise was reduced through filtering

and preprocessing. Representativeness was ensured by covering all regions, timeframes, and key factors like weather, traffic, and industry. Consistency across sources was achieved via normalization, and sufficient, diverse data volume supported effective learning for both ML and DL models. A review by (Méndez, Merayo & Núñez, 2023) of major scientific databases found that LSTM, SVR, and RF are the most commonly used algorithms for air pollution prediction, supporting their selection in this study. LSTM was chosen because it effectively captures temporal dependencies in air quality and meteorological data and models the nonlinear relationships common in environmental systems. SVR was chosen for its robustness with moderate-sized datasets, ability to avoid overfitting, and capacity to model complex nonlinear trends using the RBF kernel without extensive feature engineering. SVR and RF are widely adopted due to their flexibility with diverse datasets, with RF particularly strong in handling variable types. Linear regression remains popular as a foundational method, while standalone decision trees are less used because RF offers superior performance. ARIMA is less common, as its effectiveness declines with many predictors.

Feature selection is essential for accurate prediction. Relevant meteorological variables—temperature, humidity, wind speed, and direction—were included, while unavailable features like sunlight, visibility, and cloud cover were excluded. Mutual Information was used to retain the most informative features, refining the dataset and improving model performance. In addition to feature selection, model parameters were carefully tuned. LSTM architecture was optimized for complexity and generalization, while SVR parameters were adjusted to enhance predictive performance and reduce overfitting. Training incorporated relevant variables—pollutant concentrations, wind, temperature, dew point, pressure, vapor pressure, and humidity—selected via Mutual Information. Standard optimization techniques were applied to improve model performance despite computational constraints. AI model performance can be affected by sudden environmental changes, such as storms or wildfires. To address this, models were designed to adapt to temporal variability, with adaptive algorithms considered to enhance prediction under dynamic conditions.

The model architecture used multiple layers with dropout and early stopping to prevent overfitting. Hyperparameters—including learning rate, dropout rate, optimizer type, and neuron count (32–256)—were tuned using GridSearchCV and KerasTuner. Adam and SGD optimizers were applied, and early stopping monitored validation performance to halt training when improvements ceased, ensuring robust and efficient model evaluation. To prevent overfitting, several techniques like regularization, dropout and batch normalization were employed.

Model Evaluation

Several common error indicators have been used to examine the predictive performance of the models. This research employs R^2 , along with MAE and RMSE, as evaluation metrics. The MAE and RMSE are reported in $\mu\text{g}/\text{m}^3$ for particulate matter and in ppb for NO_2 and O_3 .

The effectiveness of the ML models was assessed through comparative analysis across pollutants ($\text{PM}_{2.5}$, PM_{10} , O_3 , NO_2). Performance metrics— R^2 , RMSE, and MAE—were calculated for each model at multiple monitoring stations, highlighting their strengths and limitations under varying conditions. For $\text{PM}_{2.5}$, LSTM performed best at the Piruzi station ($R^2 = 0.83$), capturing temporal trends effectively. SVR and RF were also competitive, with RF showing strong overall robustness but slightly higher RMSE in noisy stations, while SVR was stable in cleaner areas (e.g., Aghdasiyeh, RMSE = 5.40) but less effective in highly polluted environments. For PM_{10} , RF models consistently outperformed LSTM and SVR, achieving R^2 values of 0.76 (Sharif) and 0.70 (Shad Abad), showing robustness in modeling coarse particulate matter. LSTM performed moderately, struggling in high-variability areas (e.g., Shahr-dari 21st district, $R^2 = 0.70$, RMSE = 30.52), while SVR was reliable in stable air quality stations but less effective under abrupt fluctuations. For O_3 , SVR performed best, achieving R^2 values up to 0.80 (Golbarg) and 0.77

Table 1. Statistical error indicators for PM2.5, PM10, O3, and NO2

Station	Model	R ²	RMSE	MAE	R ²	RMSE	MAE
Pollutant			PM _{2.5}			PM ₁₀	
Tarbiat Modares	SVR	0.7	7.17	4.87	0.62	17.96	11.27
	RF	0.82	6.83	4.87	0.69	16.06	10.76
	LSTM	0.81	6.43	4.78	0.62	17.34	12.11
Aghdasiyeh	SVR	0.71	5.4	3.83	0.59	14.72	9.73
	RF	0.71	5.38	3.56	0.69	12.31	8.99
	LSTM	0.72	5.45	3.81	0.52	15.62	10.81
Sharif	SVR	0.76	7.52	5.27	0.74	17.53	12.71
	RF	0.73	7.9	5.63	0.76	17.32	12.87
	LSTM	0.65	8.63	6.06	0.71	18.67	12.16
Piruzi	SVR	0.8	6.14	4.48	0.66	19.08	12.53
	RF	0.78	6.31	4.36	0.67	18.53	11.43
	LSTM	0.83	6.44	4.62	0.7	19.32	12.68
Setad Bohran	SVR	0.43	9.85	6.69	0.61	16.3	11.84
	RF	0.62	8.34	5.9	0.59	16.54	11.44
	LSTM	0.6	7.76	5.38	0.53	18.77	13.27
Shahrdari 21 dist	SVR	0.7	7.2	5.15	0.57	24.19	16.84
	RF	0.73	6.77	5.12	0.61	27.11	17.69
	LSTM	0.73	7.97	5.67	0.7	30.52	16.96
Shad Abad	SVR	0.74	7.14	5.34	0.61	21.34	13.66
	RF	0.78	7.16	5.11	0.7	18.19	12.17
	LSTM	0.78	6.88	5.28	0.66	18.08	13.55
Masudieh	SVR	0.57	5.46	4.02	0.45	18.89	12.96
	RF	0.52	6.14	4.34	0.53	17.66	12.76
	LSTM	0.42	6.51	6.80	0.48	17.63	13.12
Golbarg	SVR	0.56	5.03	3.62	0.45	15.96	10.59
	RF	0.66	4.62	3.39	0.5	15.81	10.63
	LSTM	0.58	5.23	3.68	0.46	14.84	10.55
Pollutant			O ₃			NO ₂	
Tarbiat Modares	SVR	0.6	7.98	5.57	0.98	2	0.91
	RF	0.65	7.88	5.09	0.9	4.45	2.73
	LSTM	0.62	7.17	5.35	0.95	3.04	1.51
Aghdasiyeh	SVR	0.72	6.5	4.65	1	0.59	0.35
	RF	0.74	6.13	4.4	0.93	5.37	0.87
	LSTM	0.75	6.16	4.37	0.99	2.08	1.07
Sharif	SVR	0.71	5.31	3.6	1	1.19	0.65
	RF	0.83	4.01	2.94	0.92	5.09	0.338
	LSTM	0.68	5.68	4.08	0.99	2.10	1.30
Piruzi	SVR	0.77	5.3	4.02	0.99	0.85	0.43
	RF	0.72	5.97	4.2	0.85	4.66	2.61
	LSTM	0.72	5.9	4.48	0.99	1.43	0.90
Setad Bohran	SVR	0.68	7.37	5.01	1	1.4	0.74
	RF	0.74	6.36	4.64	0.87	8.33	3.59
	LSTM	0.73	6.44	4.99	0.99	2.36	1.84
Shahrdari 21 dist	SVR	0.74	4.75	3.56	0.89	4.19	2.57
	RF	0.73	4.92	3.62	0.76	5.41	3.81
	LSTM	0.69	5.24	3.96	0.91	3.74	2.7
Shad Abad	SVR	0.72	5.16	3.95	0.79	5.17	3.84
	RF	0.71	5.54	3.9	0.75	5.43	4.05
	LSTM	0.72	5.19	4.07	0.7	6.74	5.26
Masudieh	SVR	0.6	6.54	4.72	0.71	6.08	4.59
	RF	0.63	6.69	4.77	0.76	5.78	4.38
	LSTM	0.67	6.49	4.82	0.77	5.67	4.29
Golbarg	SVR	0.8	5.47	4.09	1	0.4	0.22
	RF	0.8	5.34	3.88	0.96	2.4	1.6
	LSTM	0.77	5.77	4.43	1	0.53	0.41

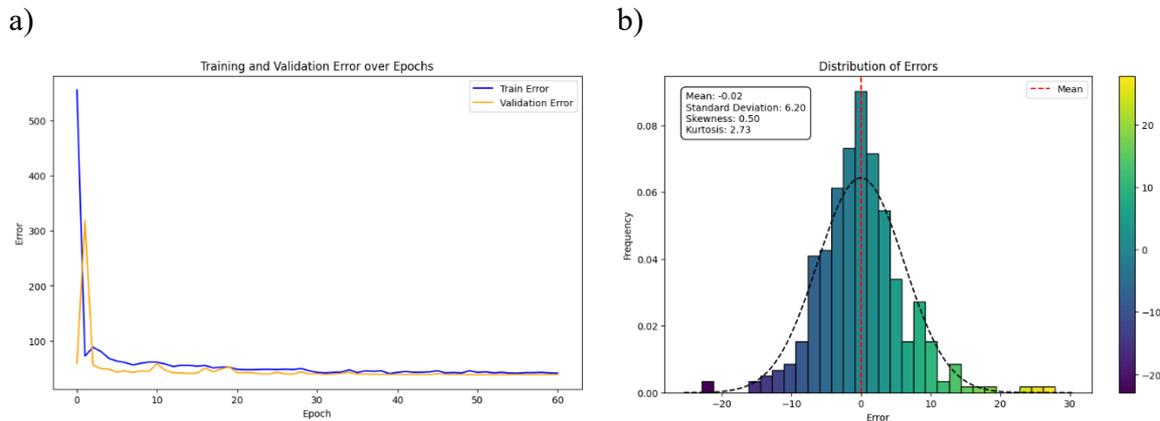


Fig. 4. a) Error versus epoch graph for training the neural network at the Tarbiat Modares station for the PM_{2.5} pollutant. b) Histogram of error distribution at the Tarbiat Modares station for the PM_{2.5} pollutant.

(Piruzi), reflecting its strength in modeling smooth nonlinear relationships. RF and LSTM were also competitive, with RF showing slightly better generalization, while LSTM was sensitive to fluctuations and required careful tuning. Overall, SVR balanced accuracy and computational efficiency for O₃ prediction. For NO₂, SVR outperformed RF and LSTM, achieving near-perfect R² (1.00) at Aghdasiyeh, Sharif, and Golbarg, showing its effectiveness for pollutants with stable temporal patterns. LSTM performed well overall (R² = 0.99) but declined in volatile stations like Shad Abad (R² = 0.70). RF showed moderate performance with higher RMSE and MAE in stations with irregular NO₂ profiles.

The Epoch-Error chart evaluates model performance during training by showing error changes across iterations. It reveals learning progress, convergence, and potential underfitting or overfitting. A consistent error reduction indicates effective learning, while high or fluctuating errors signal underfitting or overfitting. The chart also guides hyperparameter selection, helping optimize learning rate and number of epochs for efficient, reliable training.

The Frequency-Error chart examines the distribution of prediction errors, often resembling a Gaussian curve. It shows whether errors are consistently centered around the mean or if extreme values occur frequently. Metrics like mean, standard deviation, skewness, and kurtosis provide further insight, helping identify systematic biases and assess how well the model handles data variability. The Frequency-Error chart analyzes the distribution of prediction errors, typically forming a Gaussian curve. It reveals whether errors are centered around the mean or skewed, highlighting consistency and potential biases. Metrics like mean, standard deviation, skewness, and kurtosis help evaluate model performance and identify areas needing improvement. The Frequency-Error chart evaluates the distribution of prediction errors, often forming a Gaussian curve. It shows whether errors are centered around the mean or skewed, revealing consistency and potential biases. Metrics like mean, standard deviation, skewness, and kurtosis help assess model performance and highlight areas needing improvement. Figure 5 compares model performance using only ground-based data (5-a) versus combined ground and satellite data (5-b). Integrating satellite observations improved overall accuracy, with Piruzi station showing the highest R² and Golbarg the lowest, likely due to sudden pollutant spikes. Despite these fluctuations, the model captured overall trends, and including variables like NO and NO_x enhanced NO₂ predictions.

Our study confirms a strong correlation between meteorological parameters and air pollutant concentrations using both ground-based and satellite data. Importantly, it demonstrates the reliability of satellite RS for urban air quality monitoring. The high agreement between satellite-derived and ground measurements supports the optimization of monitoring infrastructure,

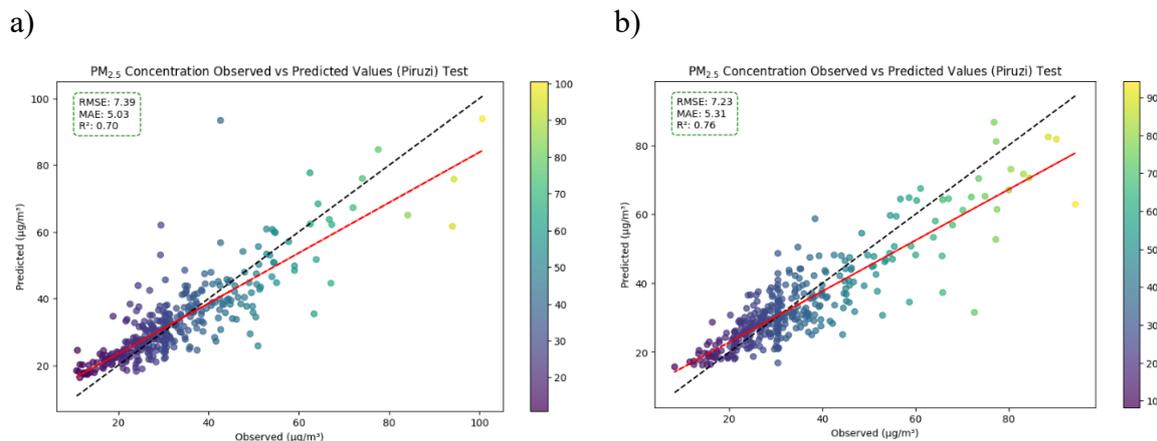


Fig. 5. a) Graph of observed versus predicted values for PM_{2.5} at the Piruzi station using only ground-based data. b) Graph of observed versus predicted values for PM_{2.5} at the Piruzi station using a combination of ground-based and satellite data.

reducing dependence on dense ground-station networks, especially in resource-limited settings. This study demonstrates the value of RS as a cost-effective, scalable, and reliable component of modern air quality management. It highlights the potential to reduce the number of physical monitoring stations without sacrificing data quality, encouraging investment in integrated, smart monitoring systems. While the hybrid model performs well overall, it struggles to capture abrupt pollution spikes caused by short-term meteorological changes or sudden emissions. Future research should integrate real-time emissions data, higher-frequency satellite observations, and near-real-time auxiliary datasets (e.g., traffic, industrial activity). Additionally, adaptive ML techniques could improve the model's responsiveness and accuracy under rapidly changing conditions.

CONCLUSION

This study provides a comprehensive analysis of the interactions between meteorological parameters and air pollution dynamics in Tehran by integrating ground-based and satellite-derived datasets. Results confirm that temperature, wind speed, relative humidity, and atmospheric pressure play decisive roles in pollutant concentrations. Strong correlations between satellite and ground-based data validate the reliability of satellite observations as a cost-effective supplement to conventional monitoring, addressing spatial coverage and maintenance limitations. By adopting a hybrid approach, this research enhances both spatial and temporal resolution in areas with sparse ground stations. Advanced ML models—LSTM, SVR, and Random Forest—effectively captured nonlinear relationships between meteorology and pollutant behavior. LSTM excelled in PM_{2.5} prediction ($R^2 = 0.83$ at Piruzi), SVR performed best for NO₂ ($R^2 = 1.00$ at Aghdasiyeh and Sharif), and Random Forest showed strong results across O₃ and PM_{2.5}. Diagnostic analyses highlighted that model performance is pollutant- and location-specific, emphasizing the need for flexible and adaptive modeling strategies. From a practical perspective, this study offers guidance for policymakers: incorporating real-time and forecasted meteorological data allows targeted interventions during stagnation events, while satellite data can expand monitoring coverage in underserved areas. The results also inform optimal ground station placement, identification of reliable data sources, and potential model application to nearby cities with similar conditions (e.g., Karaj, Qods), reducing uncertainty in regional air quality assessments.

Overall, the integration of diverse datasets with advanced ML offers a scalable, evidence-

based framework for air quality assessment and management. These findings provide urban authorities with actionable tools to identify pollution hotspots, simulate interventions, issue timely alerts, and implement cost-effective strategies to safeguard public health, thereby supporting sustainable urban planning and policymaking.

GRANT SUPPORT DETAILS

The present research did not receive any financial support.

CONFLICT OF INTEREST

The authors declare that there is not any conflict of interests regarding the publication of this manuscript. In addition, the ethical issues, including plagiarism, informed consent, misconduct, data fabrication and/ or falsification, double publication and/or submission, and redundancy has been completely observed by the authors.

LIFE SCIENCE REPORTING

No life science threat was practiced in this research.

REFERENCES

- Akima, H. (1970) A New Method of Interpolation and Smooth Curve Fitting Based on Local Procedures. *Journal of the ACM (JACM)*. 17 (4). doi:10.1145/321607.321609.
- Ali-Taleshi, M.S., Bakhtiari, A.R., Liu, N. & Hopke, P.K. (2025) Characterization and Transport Pathways of High PM_{2.5} Pollution Episodes During 2015–2021 in Tehran, Iran. *Aerosol and Air Quality Research*. 25 (5), 1–18. doi:10.1007/S44408-025-00016-Y/METRICS.
- Alotaibi, E. & Nassif, N. (2024) Artificial intelligence in environmental monitoring: in-depth analysis. *Discover Artificial Intelligence*. 4 (1). doi:10.1007/S44163-024-00198-1.
- Amiri, Z. & Zare Shahne, M. (2025) Modeling PM_{2.5} concentration in tehran using satellite-based Aerosol optical depth (AOD) and machine learning: Assessing input contributions and prediction accuracy. *Remote Sensing Applications: Society and Environment*. 38, 101549. doi:10.1016/J.RSASE.2025.101549.
- Anggraini, T.S., Irie, H., Sakti, A.D. & Wikantika, K. (2024) Machine learning-based global air quality index development using remote sensing and ground-based stations. *Environmental Advances*. 15, 100456. doi:10.1016/J.ENVADV.2023.100456.
- Arhami, M., Hosseini, V., Zare Shahne, M., Bigdeli, M., Lai, A. & Schauer, J.J. (2017) Seasonal trends, chemical speciation and source apportionment of fine PM in Tehran. *Atmospheric Environment*. 153. doi:10.1016/j.atmosenv.2016.12.046.
- Arhami, M., Shahne, M.Z., Hosseini, V., Roufigar Haghghat, N., Lai, A.M. & Schauer, J.J. (2018) Seasonal trends in the composition and sources of PM_{2.5} and carbonaceous aerosol in Tehran, Iran. *Environmental Pollution*. 239. doi:10.1016/j.envpol.2018.03.111.
- Artiola, J.F., Pepper, I.L. & Brusseau, M.L. (2004) *Environmental Monitoring and Characterization*. doi:10.1016/B978-0-12-064477-3.X5000-0.
- Austin, P.C., White, I.R., Lee, D.S. & van Buuren, S. (2021) Missing Data in Clinical Research: A Tutorial on Multiple Imputation. *Canadian Journal of Cardiology*. 37 (9). doi:10.1016/j.cjca.2020.11.010.
- Bagheri, H. (2022) A machine learning-based framework for high resolution mapping of PM_{2.5} in Tehran, Iran, using MAIAC AOD data. *Advances in Space Research*. 69 (9). doi:10.1016/j.asr.2022.02.032.
- Bourbour, S., Zare Shahne, M. & Safarkhanloo, H. (2025) Non-Gaussian Misalignment Regression Model Development Based on On-Site and Satellite Meteorological and Air Pollutant Data in Tehran, Iran. *International Journal of Environmental Research*. 19 (4), 1–22. doi:10.1007/S41742-025-00797-5/METRICS.
- Ceccato, P., Fernandes, K., Ruiz, D. & Allis, E. (2014) Climate and environmental monitoring for

- decision making. *Earth Perspectives*. 1 (1). doi:10.1186/2194-6434-1-16.
- Chadalavada, S., Faust, O., Salvi, M., Seoni, S., Raj, N., Raghavendra, U., Gudigar, A., Barua, P.D., Molinari, F. & Acharya, R. (2025) Application of artificial intelligence in air pollution monitoring and forecasting: A systematic review. *Environmental Modelling & Software*. 185, 106312. doi:10.1016/J.ENVSOFT.2024.106312.
- Cordier, T., Alonso-Sáez, L., Apothéloz-Perret-Gentil, L., Aylagas, E., Bohan, D.A., Bouchez, A., Chariton, A., Creer, S., Frühe, L., Keck, F., Keeley, N., Laroche, O., Leese, F., Pochon, X., Stoeck, T., Pawlowski, J. & Lanzén, A. (2021) Ecosystems monitoring powered by environmental genomics: A review of current strategies with an implementation roadmap. In: *Molecular Ecology*. 2021 p. doi:10.1111/mec.15472.
- Delbari, S.H., Zare Shahne, M. & Hosseini, V. (2024) An Analysis of Primary Contributing Sources to the PM_{2.5} Composition in a Port City in Canada Influenced by Traffic, Marine, and Wildfire Emissions. *Atmospheric Environment*. 334, 120712. doi:10.1016/J.ATMOSENV.2024.120712.
- Ditria, E.M., Buelow, C.A., Gonzalez-Rivero, M. & Connolly, R.M. (2022) Artificial intelligence and automated monitoring for assisting conservation of marine ecosystems: A perspective. *Frontiers in Marine Science*. 9. doi:10.3389/fmars.2022.918104.
- Dressing, S.A. & Meals, D.W. (2010) Monitoring and evaluating nonpoint source watershed projects. *United States Environmental Protection Agency Office of Water*. (May).
- Duan, Y., Edwards, J.S. & Dwivedi, Y.K. (2019) Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*. 48. doi:10.1016/j.ijinfomgt.2019.01.021.
- Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D. & Alsdorf, D.E. (2007) The shuttle radar topography mission. *Reviews of Geophysics*. 45 (2). doi:10.1029/2005RG000183.
- Hameed, M., Sharqi, S.S., Yaseen, Z.M., Afan, H.A., Hussain, A. & Elshafie, A. (2017) Application of artificial intelligence (AI) techniques in water quality index prediction: a case study in tropical region, Malaysia. *Neural Computing and Applications*. 28. doi:10.1007/s00521-016-2404-7.
- Heger, M., Sarraf, M. & Heger, M.P. (2018) Air Pollution in Tehran: Health Costs, Sources, and Policies. *Air Pollution in Tehran*. doi:10.1596/29909.
- Ibrahim, A., Wayayok, A., Shafri, H.Z.M. & Toridi, N.M. (2024) Remote Sensing Technologies for Unlocking New Groundwater Insights: A Comprehensive Review. *Journal of Hydrology X*. 23, 100175. doi:10.1016/J.HYDROA.2024.100175.
- Järvi, L., Kurppa, M., Kuuluvainen, H., Rönkkö, T., Karttunen, S., Balling, A., Timonen, H., Niemi, J. V. & Pirjola, L. (2023) Determinants of spatial variability of air pollutant concentrations in a street canyon network measured using a mobile laboratory and a drone. *Science of the Total Environment*. 856. doi:10.1016/j.scitotenv.2022.158974.
- Javadi, M.A., Shahne, M.Z. & Amiri, Z. (2025) Integration of Sentinel-5P Satellite Data and Machine Learning for Spatiotemporal Prediction of NO₂ in Delhi: Impacts of COVID-19 Lockdown. *Atmospheric Pollution Research*. 102702. doi:10.1016/J.APR.2025.102702.
- Javan, F.D., Samadzadegan, F. & Toosi, A. (2025) Air pollution observation—bridging spaceborne to unmanned airborne remote sensing: a systematic review and meta-analysis. *Air Quality, Atmosphere & Health* 2025. 1–69. doi:10.1007/S11869-025-01771-Y.
- Jolliffe, I.T. & Cadima, J. (2016) Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 374 (2016). doi:10.1098/rsta.2015.0202.
- Kafi, A.M., Hosseinipoor, M., Zare Shahne, M., Jamaat, A., Kafi, A.M., Hosseinipoor, M., Zare Shahne, M. & Jamaat, A. (2024) Integrating Sentinel-5P Satellite Data and Machine Learning Algorithms for Air Quality Index Prediction in Tehran: A Comprehensive Study on Factors Influencing Air Quality. *EGUGA*. 4506. doi:10.5194/EGUSPHERE-EGU24-4506.
- Kamali, N., Zare Shahne, M. & Arhami, M. (2015) Implementing spectral decomposition of time series data in artificial neural networks to predict air pollutant concentrations. *Environmental Engineering Science*. 32 (5). doi:10.1089/ees.2014.0350.
- Kim, M. kyu, Chang, J.W., Park, K. & Yang, D.R. (2022) Comprehensive assessment of the effects of operating conditions on membrane intrinsic parameters of forward osmosis (FO) based on principal component analysis (PCA). *Journal of Membrane Science*. 641. doi:10.1016/j.memsci.2021.119909.

- Koçak, E. (2025) Comprehensive evaluation of machine learning models for real-world air quality prediction and health risk assessment by AirQ+. *Earth Science Informatics*. 18 (3), 1–17. doi:10.1007/S12145-025-01941-7/FIGURES/4.
- Laña, I., Olabarrieta, I. (Iñaki), Vélez, M. & Del Ser, J. (2018) On the imputation of missing data for road traffic forecasting: New insights and novel techniques. *Transportation Research Part C: Emerging Technologies*. 90. doi:10.1016/j.trc.2018.02.021.
- Lever, J., Krzywinski, M. & Altman, N. (2017) Points of Significance: Principal component analysis. *Nature Methods*. 14 (7). doi:10.1038/nmeth.4346.
- Ma, J., Cheng, J.C.P., Jiang, F., Chen, W., Wang, M. & Zhai, C. (2020a) A bi-directional missing data imputation scheme based on LSTM and transfer learning for building energy data. *Energy and Buildings*. 216. doi:10.1016/j.enbuild.2020.109941.
- Ma, J., Li, Z., Cheng, J.C.P., Ding, Y., Lin, C. & Xu, Z. (2020b) Air quality prediction at new stations using spatially transferred bi-directional long short-term memory network. *Science of the Total Environment*. 705. doi:10.1016/j.scitotenv.2019.135771.
- Maryam Zare, S., Amir, S. & Fatemeh, N. (2022) A hybrid deep learning model to forecast air quality data based on COVID-19 outbreak in Mashhad, Iran. *Annals of Civil and Environmental Engineering*. 6 (1). doi:10.29328/journal.acee.1001035.
- Méndez, M., Merayo, M.G. & Núñez, M. (2023) Machine learning algorithms to forecast air quality: a survey. *Artificial Intelligence Review*. 56 (9). doi:10.1007/s10462-023-10424-4.
- Mishra, D. & Goyal, P. (2015) Development of artificial intelligence based NO₂ forecasting models at Taj Mahal, Agra. *Atmospheric Pollution Research*. 6 (1). doi:10.5094/APR.2015.012.
- Mohammadi, H., Cohen, D., Babazadeh, M. & Rokni, L. (2012) The Effects of Atmospheric Processes on Tehran Smog Forming. *Iranian Journal of Public Health*. 41 (5), 1. <https://pmc.ncbi.nlm.nih.gov/articles/PMC3468992/>.
- Moshenberg, S., Lerner, U. & Fishbain, B. (2015) Spectral methods for imputation of missing air quality data. *Environmental Systems Research*. 4 (1). doi:10.1186/s40068-015-0052-z.
- Noori, R., Hoshyaripour, G., Ashrafi, K. & Araabi, B.N. (2010) Uncertainty analysis of developed ANN and ANFIS models in prediction of carbon monoxide daily concentration. *Atmospheric Environment*. 44 (4). doi:10.1016/j.atmosenv.2009.11.005.
- Pena, M., Ortega, P. & Orellana, M. (2019) A novel imputation method for missing values in air pollutant time series data. In: *2019 IEEE Latin American Conference on Computational Intelligence, LA-CCI 2019*. 2019 p. doi:10.1109/LA-CCI47412.2019.9037053.
- Ringnér, M. (2008) What is principal component analysis? *Nature Biotechnology*. 26 (3). doi:10.1038/nbt0308-303.
- Sarker, I.H. (2021a) Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective. *SN Computer Science*. 2 (5). doi:10.1007/s42979-021-00765-8.
- Sarker, I.H. (2021b) Machine Learning: Algorithms, Real-World Applications and Research Directions. *SN Computer Science*. 2 (3). doi:10.1007/s42979-021-00592-x.
- Savić, M., Mihajlović, I., Arsić, M. & Živković, Ž. (2014) Adaptive-network-based fuzzy inference system (ANFIS) modelbased prediction of the surface ozone concentration. *Journal of the Serbian Chemical Society*. 79 (10). doi:10.2298/JSC140126039S.
- Shahne, M.Z., Arhami, M., Hosseini, V. & Al Haddad, E. (2022a) Particulate emissions of real-world light-duty gasoline vehicle fleet in Iran. *Environmental Pollution*. 292. doi:10.1016/j.envpol.2021.118303.
- Shahne, M.Z., Sezavar, A., Najibi, F. & Khoshand, A. (2022b) *Automatic Image-based Air Quality Detection based on Deep Convolutional Neural Networks and Histogram Analysis*. 255. doi:10.34894/VQ1DJA.
- Shams, S.R., Jahani, A., Kalantary, S., Moeinaddini, M. & Khorasani, N. (2021) Artificial intelligence accuracy assessment in NO₂ concentration forecasting of metropolises air. *Scientific Reports*. 11 (1). doi:10.1038/s41598-021-81455-6.
- Subramaniam, S., Raju, N., Ganesan, A., Rajavel, N., Chenniappan, M., Prakash, C., Pramanik, A., Basak, A.K. & Dixit, S. (2022) Artificial Intelligence Technologies for Forecasting Air Pollution and Human Health: A Narrative Review. *Sustainability (Switzerland)*. 14 (16). doi:10.3390/su14169951.
- Thomson, M.C., Connor, S.J., Zebiak, S.E., Jancoes, M. & Mihretie, A. (2011) Africa needs climate data to fight disease. *Nature*. 471 (7339). doi:10.1038/471440a.
- United Nations (2022) United Nations Department of Economic and Social Affairs, Population Division

- (2022). World Population Prospects 2022: Summary of Results. *Department of Economic and Social Affairs Population Division*. 3 (UN DESA/POP/2022/TR/NO. 3).
- Waymond, R. (2020) Artificial Intelligence in a Throughput Model. *Artificial Intelligence in a Throughput Model*. doi:10.1201/9780429266065.
- Yeganeh, B., Hewson, M.G., Clifford, S., Tavassoli, A., Knibbs, L.D. & Morawska, L. (2018) Estimating the spatiotemporal variation of NO₂ concentration using an adaptive neuro-fuzzy inference system. *Environmental Modelling and Software*. 100. doi:10.1016/j.envsoft.2017.11.031.
- Yu, Y., Yu, J.J.Q., Li, V.O.K. & Lam, J.C.K. (2020) A Novel Interpolation-SVT Approach for Recovering Missing Low-Rank Air Quality Data. *IEEE Access*. 8. doi:10.1109/ACCESS.2020.2988684.
- Zare Shahne, M., Haghghat, N.R., Hosseini, V., Uzu, G., Taheri, A., Darfeuil, S., Ginot, P., Besombes, J.L., Pin, M., Jaffrezo, J.L. & Shamloo, A. (2025) Seasonal variation of the chemical content and source identification of PM_{2.5} in a mixed landuse in Iran. *International Journal of Environmental Science and Technology*. 22 (6), 4157–4172. doi:10.1007/S13762-024-06138-X/METRICS.
- Zhou, X.H. (2020) Challenges and strategies in analysis of missing data. *Biostatistics and Epidemiology*. 4 (1). doi:10.1080/24709360.2018.1469810.