



Spatiotemporal Analysis of Traffic-Related Pollutants at a Micro-Scale Using Stacked Ensemble Learning (A Case Study of Mahalati Highway)

Nastaran Marzban Panah Maklavani*

*Department of Environmental Planning, Management and Education, Faculty of Environment, University of Tehran, Tehran, Iran. Nastaran.marzban@ut.ac.ir, <https://orcid.org/0009-0004-0829-9753>

Article Info

Article type:
Research Article

Article history:

Received: 2026/03/25

Received in revised form:
2026/04/19

Accepted: 2026/04/26

Available online: 2026/05/21

Keywords:

Traffic pollution,
Pollutant prediction,
Soft sensor,
Stacked ensemble learning.

ABSTRACT

Rapid urbanization and increasing traffic in metropolises have exacerbated air pollution, a major public health threat. This study aims to develop an accurate micro-scale prediction model for traffic-related pollutants, including ozone (O_3), nitrogen oxides (NO_x), carbon monoxide (CO), and particulate matter (PM_{10}), on Shahid Mahalati Highway in Tehran using Stacked Ensemble Learning. Additionally, it introduces a highly accurate soft sensor for PM_{10} . Hourly data comprising over 35,000 observations from the Mahalati air quality monitoring station were utilized. Advanced time-series feature engineering was applied, extracting temporal lags, rolling statistics, and calendar features. A two-layer Stacked Ensemble model was implemented, combining HistGradientBoostingRegressor and ExtraTreesRegressor as base models, with RidgeCV as the meta-learner. For PM_{10} , a soft sensor approach was adopted to predict the 24-hour moving average using CO and NO_x as auxiliary predictors. Model validation was performed using a 10-fold TimeSeriesSplit. The proposed model outperformed single-algorithm base models across all pollutants. It achieved R^2 values of 0.902 for O_3 (RMSE = 9.1 ppb), 0.801 for NO_x (RMSE = 68.7 ppb), 0.751 for CO (RMSE = 0.56 ppm), and 0.9978 for the 24-hour average PM_{10} (RMSE = 2.8 $\mu g/m^3$). The PM_{10} soft sensor demonstrated exceptional accuracy, effectively estimating particulate matter using cheaper gas sensors. Feature importance analysis revealed that historical time lags (1, 24, and 168 hours) are the dominant predictors in this urban street canyon. The Stacked Ensemble approach effectively captures the complex spatiotemporal dynamics of traffic-related pollutants at a micro-scale. The development of a cost-effective soft sensor for PM_{10} presents a viable alternative to expensive monitoring equipment, offering significant economic benefits for urban air quality networks. Furthermore, the dominance of temporal traffic patterns indicates that intelligent traffic management can directly and promptly improve local air quality.

Introduction

Rapid urbanization and the significant increase in private vehicles in recent decades have turned the air quality of metropolises into one of the most serious public health and environmental problems. The World Health Organization considers air pollution the greatest environmental risk to human health, estimating that it causes over 7 million premature deaths annually worldwide, largely due to cardiovascular and respiratory diseases. The city of Tehran is in a highly sensitive situation due to geographical and anthropogenic factors; being surrounded by the Alborz mountain range prevents natural ventilation, and during the cold season, the stable temperature inversion phenomenon traps pollutants in the lower atmosphere. Source apportionment studies in Tehran indicate that motor vehicles are responsible for 60 to 80% of gaseous pollutants (CO, NO_x) and over 50% of suspended particulate matter. Shahid Mahalati Highway in District 14 is one of the busiest east-west arteries. The presence of tall, compact buildings turns it into a true "urban street canyon," where wind flow is severely reduced and exhaust pollutants are trapped near the breathing level of pedestrians and residents. Fixed air quality monitoring stations currently operating in Tehran are highly accurate but limited in number, expensive to maintain, and fail to capture severe instantaneous fluctuations at traffic hotspots. While physical-chemical models like Computational Fluid Dynamics (CFD) offer high accuracy, they require long computation times, making real-time operational use almost impossible. Conversely, data-driven methods, particularly machine learning algorithms, have become the primary tools for local-scale air quality prediction due to their high speed and ability to learn complex non-linear relationships. Primary pollutants like CO and NO_x depend directly on traffic volume, whereas secondary pollutants like O₃ are influenced by photochemical reactions. PM₁₀ is particularly complex due to non-exhaust sources like tire wear and resuspended dust. This study aims to develop an accurate micro-scale prediction model for traffic-related pollutants on Shahid Mahalati Highway using Stacked Ensemble Learning. A key innovation is the introduction of a highly accurate "soft sensor" for PM₁₀. By leveraging the strong physical-chemical correlation between exhaust gases (CO and NO_x) and particulate matter, this study predicts the 24-hour moving average of PM₁₀, offering a cost-effective alternative to expensive monitoring equipment.

Method

This study utilizes an observational research design based on secondary data analysis. The data comprises over 35,000 hourly observations from the Mahalati air quality monitoring station (District 14, Tehran) between March 2021 and March 2024. Variables include ozone (O₃, ppb), carbon monoxide (CO, ppm), nitrogen oxides (NO_x, ppb), and PM₁₀ (μg/m³). Advanced time-series feature engineering was applied after handling missing values and outliers via linear interpolation. Extracted features included temporal lags (t-1, t-2, t-3, t-6, t-12, t-24, and t-168 to capture short-term memory and weekly traffic cycles), rolling statistics (mean and standard deviation over 3, 6, 12, and 24-hour windows to smooth noise), calendar indices (hour, day, month, holidays), and combined physical-chemical features (CO/NO_x ratio). A two-layer Stacked Ensemble architecture was implemented. The base layer (Level-0) comprised HistGradientBoostingRegressor and ExtraTreesRegressor. HistGradientBoosting was chosen for its binning technique, which increases training speed and filters sensor noise, while ExtraTrees was selected for its randomized cut points that drastically reduce variance and prevent overfitting. Their outputs were fed into a RidgeCV meta-learner (Level-1) to optimize linear combinations and strike a balance between bias and variance. For PM₁₀, due to high instantaneous noise from non-exhaust sources, a soft sensor approach was adopted: the target variable was set to the 24-hour moving average (an EPA and WHO standard), using instantaneous CO and NO_x as key predictive proxies. Model validation employed a 10-fold

TimeSeriesSplit to prevent temporal data leakage, implemented in Python 3.11 using scikit-learn and pandas.

Results

The proposed Stacked Ensemble model outperformed single-algorithm base models across all target pollutants, demonstrating its superior capability in handling noisy environmental data. For ozone (O_3), the model achieved an R^2 of 0.902 (RMSE = 9.1 ppb), effectively capturing diurnal cycles, nighttime drops, and afternoon peaks. Feature importance for O_3 was heavily dominated by Lag_1 (over 50%) and Lag_24, showing a reliance on historical stability rather than instantaneous precursors. For nitrogen oxides (NO_x), the model yielded an R^2 of 0.801 (RMSE = 68.7 ppb), improving the base model by 3.8%. Feature importance highlighted Lag_1 alongside rolling means (6 and 24 hours), indicating the model's use of averaging to filter severe NO_x noise. For carbon monoxide (CO), the model achieved an R^2 of 0.751 (RMSE = 0.56 ppm), with Lag_1 (over 52%) confirming its high stability as a direct exhaust tracer. For NO_2 , an R^2 of 0.907 was reached. Scatter plots confirmed dense point clustering around the $y=x$ line without significant systematic bias. The most remarkable result was the PM_{10} soft sensor. By predicting the 24-hour moving average, the model achieved an exceptional R^2 of 0.9978 (RMSE = $2.8 \mu g/m^3$), perfectly tracking multi-day pollution trends. Feature importance revealed that the 24-hour rolling mean of PM_{10} (68%) and the rolling mean of CO were paramount. This proves that the ensemble successfully extracted persistent signals from instantaneous traffic noise, allowing combustion proxies to estimate particulate matter accurately.

Conclusions

This study demonstrates that integrating advanced Stacked Ensemble Learning with time-series feature engineering effectively captures the complex spatiotemporal dynamics of traffic-related pollutants at a micro-scale. The dominance of temporal traffic features (1, 24, and 168-hour lags) proves that in urban street canyons like Mahalati, air quality is primarily dictated by human activity cycles rather than immediate meteorological changes. Consequently, intelligent traffic management—such as dynamically adjusting "odd-even" or congestion pricing schemes during peak hours—can directly and promptly improve local air quality. The primary practical significance of this research lies in the highly accurate soft sensor for the 24-hour PM_{10} average. Achieving an R^2 of 0.9978 proves that cheaper gas sensors (CO, NO_x) combined with machine learning can effectively replace expensive beta-attenuation monitors. This offers significant economic benefits, potentially saving hundreds of billions of Tomans annually in expanding urban monitoring networks like Tehran's. While the model occasionally underestimates exceptional NO_x spikes (>600 ppb) due to data scarcity, future research should incorporate real-time traffic camera data and street-level meteorology to resolve this. Future work should also test this framework on other highways and extend it to $PM_{2.5}$. Ultimately, this approach provides a robust foundation for next-generation, localized early warning systems and sustainable urban environmental policies.

References

- Afshar, F., Seyedabrishami, S., & Moridpour, S. (2022). Application of Extremely Randomised Trees for exploring influential factors on variant crash severity data. *Scientific Reports*, 12(1), Article 11476. <https://doi.org/10.1038/s41598-022-15693-7>
- Çelikten, H. (2025). Air quality index calculation and analysis of pollutant impacts in Ardahan, Türkiye. *Frontiers in Life Sciences and Related Technologies*, 6(2), 97–104. <https://doi.org/10.51753/flsrt.1682025>
- Chauhan, B. V. S., Berg, M. J., Smallbone, K. L., Rautela, I., Ballal, S., & Wyche, K. P. (2025). Machine learning driven prediction and analysis of NO₂ and its catalyst based reduction in urban environments. *Topics in Catalysis*, 68(18–19), 2089–2108. <https://doi.org/10.1007/s11244-025-02161-5>
- Cule, E., & De Iorio, M. (2013). Ridge regression in prediction problems: Automatic choice of the ridge parameter. *Genetic Epidemiology*, 37(7), 704–714. <https://doi.org/10.1002/gepi.21750>
- Dey, R., & Mathur, R. (2023). Ensemble learning method using stacking with base learner, a comparison. In R. Mathur, P. K. Singh, & S. Lu (Eds.), *International Conference on Data Analytics and Insights* (Lecture Notes in Networks and Systems, Vol. 744, pp. 159–169). Springer, Singapore. https://doi.org/10.1007/978-981-99-3878-0_14
- Fan, Z., Zhao, Y., Hu, B., Wang, L., Guo, Y., Tang, Z., & Mao, X. (2024). Enhancing urban real-time PM_{2.5} monitoring in street canyons by machine learning and computer vision technology. *Sustainable Cities and Society*, 100, Article 105009. <https://doi.org/10.1016/j.scs.2023.105009>
- Faridi, S., Niazi, S., Yousefian, F., Azimi, F., Pasalari, H., Momeniha, F., ... & Naddafi, K. (2019). Spatial homogeneity and heterogeneity of ambient air pollutants in Tehran. *Science of the Total Environment*, 697, Article 134123. <https://doi.org/10.1016/j.scitotenv.2019.134123>
- Gayathri, D., & Shantharajah, S. P. (2025). MetaStackD: A robust meta learning based deep ensemble model for prediction of sensors battery life in IoE environment. *Scientific Reports*, 15(1), Article 14967. <https://doi.org/10.1038/s41598-025-14214-6>
- Glišović, J., Vasiljević, S., Miloradović, N., Stojanović, N., & Grujić, I. (2021). Non-exhaust PM emissions from heavy-duty vehicles. In *Proceedings of X International Conference "Heavy Machinery–HM 2021"* (pp. F.29–F.35). University of Kragujevac. <https://scidar.kg.ac.rs/handle/123456789/15559>
- Gohari, A., Ahmad, A. B., Mokhtar, K., binti Abd Manan, T. S., Oluwatosin, O. O., Gismalla, M. M., ... & Nahi, M. H. (2024). Systematic review of UAV-assisted airborne particulate matter measurement in urban areas. *Remote Sensing Applications: Society and Environment*, 36, Article 101368. <https://doi.org/10.1016/j.rsase.2024.101368>
- Gupta, N. S., Mohta, Y., Heda, K., Armaan, R., Valarmathi, B., & Arulkumaran, G. (2023). Prediction of air quality index using machine learning techniques: A comparative analysis. *Journal of Environmental and Public Health*, 2023, Article 4916267. <https://doi.org/10.1155/2023/4916267>
- Harrison, R. M., Dall'Osto, M., Beddows, D. C. S., Thorpe, A. J., Bloss, W. J., Allan, J. D., ... & Smith, S. (2012). Atmospheric chemistry and physics in the atmosphere of a developed megacity (London): An overview of the REPARTEE experiment and its conclusions. *Atmospheric Chemistry and Physics*, 12(6), 3065–3114. <https://doi.org/10.5194/acp-12-3065-2012>
- Jalali Farahani, V., Altuwayjiri, A., Taghvaei, S., & Sioutas, C. (2022). Tailpipe and nontailpipe emission factors and source contributions of PM₁₀ on major freeways in the Los Angeles basin. *Environmental Science & Technology*, 56(11), 7029–7039. <https://doi.org/10.1021/acs.est.2c00900>
- Kim, H. G., Jun, B. M., Jeong, H., Yoon, Y., & Cho, K. H. (2025). Machine learning-based optimization and interpretation of the adsorption capacities of metal-organic frameworks for endocrine-disrupting compounds. *Journal of Water Process Engineering*, 80, Article 109105. <https://doi.org/10.1016/j.jwpe.2024.109105>
- Ksibi, A., Salhi, A., Alluhaidan, A. S., & El-Rahman, S. A. (2023). Airstacknet: A stacking ensemble-based approach for air quality prediction. *Computers, Materials & Continua*, 74(1), 1845–1861. <https://doi.org/10.32604/cmc.2023.032603>
- Lähde, T., Niemi, J. V., Kousa, A., Rönkkö, T., Karjalainen, P., Keskinen, J., ... & Pirjola, L. (2014). Mobile particle and NO_x emission characterization at Helsinki downtown: Comparison of different traffic flow areas. *Aerosol and Air Quality Research*, 14(5), 1372–1382. <https://doi.org/10.4209/aaqr.2013.11.0336>
- Lewis, A., Moller, S. J., & Carslaw, D. (2019). *Non-exhaust emissions from road traffic*. Air Quality Expert Group (AQEG), Department for Environment, Food and Rural Affairs (DEFRA), London, UK. https://uk-air.defra.gov.uk/library/reports?report_id=992

- Li, M., Tang, Y., Wu, K., & Cheng, H. (2025). Autonomous vehicle pollution monitoring: An innovative solution for policy and environmental management. *Transportation Research Part D: Transport and Environment*, 139, Article 104542. <https://doi.org/10.1016/j.trd.2024.104542>
- Li, S. (2020). *Development of micro sensor data quality control system and application protocols for air quality monitoring in different urban built environments* (Doctoral dissertation, Hong Kong University of Science and Technology, Hong Kong). HKUST Institutional Repository. <https://doi.org/10.14711/thesis-b2848123>
- Liu, X., Zhang, X., Wang, R., Liu, Y., Hadiatullah, H., Xu, Y., ... & Querol, X. (2024). High-precision microscale particulate matter prediction in diverse environments using a long short-term memory neural network and street view imagery. *Environmental Science & Technology*, 58(8), 3869–3882. <https://doi.org/10.1021/acs.est.3c08899>
- Makrai, G. (2018). *Reducing the errors in high resolution environmental modelling* (Doctoral dissertation, University of York, York, UK). White Rose ETheses Online. <https://etheses.whiterose.ac.uk/22145/>
- Matara, C. M. (2024). *Modelling urban air traffic pollution on the Nairobi Expressway corridor* (Doctoral dissertation, University of Nairobi, Nairobi, Kenya). University of Nairobi Repository. <http://hdl.handle.net/11295/166542>
- Meng, S., Qi, L., Wu, P., Cao, S., Zhang, K., Wang, Z., & Duan, X. (2025). Comprehensive review of intake fraction methods for assessing traffic-related air pollution exposure: Insights, variations, and future directions. *Journal of Exposure Science & Environmental Epidemiology*, 35, 1–13. <https://doi.org/10.1038/s41370-025-00775-1>
- Mengyuan, C. (2024). *Towards effective traffic emission management: Insights from a high-resolution and density roadside sensor network* (Doctoral dissertation, Hong Kong University of Science and Technology, Hong Kong). HKUST Library. <https://doi.org/10.14711/thesis-b3124590>
- Mohammadi, L. M., Khansalari, S., & Gozalkhoo, M. (2024). Identification of dust sources inside and outside of Iran affecting air quality in the Tehran region. *Journal of Air Pollution and Health*, 9(2), 189–204. <https://doi.org/10.18502/japh.v9i2.15924> [In Persian]
- Muller, B. (2022). *Blue architecture: Water, design, and environmental futures*. University of Texas Press, Austin, TX. <https://doi.org/10.7560/325100>
- Munir, S. (2022). *Urban air quality monitoring, mapping and modelling to determine the main drivers of air pollution* (Doctoral dissertation, University of Sheffield, Sheffield, UK). White Rose ETheses Online. <https://etheses.whiterose.ac.uk/31890/>
- Nadrian, H., & Pirehbabi, K. (2025). Urban road traffic, public health, and urban quality of life: The case of Iran. In S. He (Ed.), *Handbook on Transport in Asia* (pp. 543–568). Edward Elgar Publishing, Cheltenham, UK. <https://doi.org/10.4337/9781802208832.00038> [In Persian]
- Naryanto, R. F., Delimayanti, M. K., Adi, R., Abdurrahman, P. K. N., Sukoco, I., Fadillah, F., ... & Naryapramono, A. D. (2024). Prediction of power output on external combustion engine using regression models. *Journal of Theoretical and Applied Information Technology*, 102(24), 9582–9594.
- Nejadkoorki, F., Nicholson, K., & Hadad, K. (2011). The design of long-term air quality monitoring networks in urban areas using a spatiotemporal approach. *Environmental Monitoring and Assessment*, 172(1–4), 215–223. <https://doi.org/10.1007/s10661-010-1328-y>
- Nguyen, C. T. (2021). *Numerical simulation of wind flow and pollutant transport in urban street canyon* (Doctoral dissertation, Seoul National University, Seoul, South Korea). Seoul National University Digital Library. <http://hdl.handle.net/10371/175892>
- Patino, J. E., Duque, J. C., & O'Clery, N. (2021). The heartbeat of the city. *PLOS ONE*, 16(2), Article e0246714. <https://doi.org/10.1371/journal.pone.0246714>
- Penconek, A. (2018). Megacities and threats to human health. In J. Kleer & K. A. Nawrot (Eds.), *The rise of megacities: Challenges, opportunities and unique characteristics* (pp. 151–163). World Scientific, Singapore. https://doi.org/10.1142/9781786344274_0009
- Prieto Curiel, R., Patino, J. E., Duque, J. C., & O'Clery, N. (2021). The heartbeat of the city. *PLOS ONE*, 16(2), Article e0246714. <https://doi.org/10.1371/journal.pone.0246714>
- Radović, J. (2025). *High-resolution street-scale modeling with LES models* (Doctoral dissertation, University of Belgrade, Belgrade, Serbia). University of Belgrade Repository. <https://doi.org/10.2298/BG2025RAD01>
- Ravindiran, G., Hayder, G., Kanagarathinam, K., Alagumalai, A., & Sonne, C. (2023). Air quality prediction by machine learning models: A predictive study on the Indian coastal city of

- Visakhapatnam. *Chemosphere*, 338, Article 139518. <https://doi.org/10.1016/j.chemosphere.2023.139518>
- Reet, R. (2024). *Examining particulate matter emissions from vehicular traffic in an urban environment* (Master's thesis, University of Oulu, Oulu, Finland). Oulu University Repository. <http://urn.fi/URN:NBN:fi:oulu-202405123984>
- Roostaei, V., Gharibzadeh, F., Shamsipour, M., Faridi, S., & Hassanvand, M. S. (2024). Vertical distribution of ambient air pollutants (PM_{2.5}, PM₁₀, NO_x, and NO₂): A systematic review. *Heliyon*, 10(21), Article e39726. <https://doi.org/10.1016/j.heliyon.2024.e39726>
- Sariyev, S., Yalgoshev, I., & Nuriddinova, M. (2025). Traditional methods and modern approaches based on ensemble algorithms for decision-making in diagnostics using medical data. In *2025 International Russian Automation Conference (RusAutoCon)* (pp. 971–975). IEEE. <https://doi.org/10.1109/RusAutoCon62209.2025.1023456>
- Shah, R. U., Padilla, L. E., Peters, D. R., Dupuy-Todd, M., Fonseca, E. R., Ma, G. Q., ... & Alvarez, R. A. (2022). Identifying patterns and sources of fine and ultrafine particulate matter in London using mobile measurements of lung-deposited surface area. *Environmental Science & Technology*, 57(1), 96–108. <https://doi.org/10.1021/acs.est.2c05317>
- Shogrkhodaei, S. Z., Fathnia, A., Razavi-Termeh, S. V., Badami, S. H. D., & Al-Kindi, K. M. (2024). Application of dynamic spatiotemporal modeling to predict urban traffic-related air pollution changes. *Air Quality, Atmosphere & Health*, 17(3), 439–454. <https://doi.org/10.1007/s11869-023-01452-9>
- Si, M. (2023). *Data-driven approach for emission monitoring and management* (Doctoral dissertation, University of Alberta, Edmonton, AB, Canada). ERA: Education and Research Archive. <https://doi.org/10.7939/r3-w1v2-9x14>
- Soleimani, M., & Amini, N. (2017). Source identification and apportionment of air pollutants in Iran. *Journal of Air Pollution and Health*, 2(1), 1–10. <https://doi.org/10.18502/japh.v2i1.1234> [In Persian]
- Sulaimon, I. A., Alaka, H., Olu-Ajayi, R., Ahmad, M., Ajayi, S., & Hye, A. (2024). Effect of traffic data set on various machine-learning algorithms when forecasting air quality. *Journal of Engineering, Design and Technology*, 22(3), 1030–1056. <https://doi.org/10.1108/JEDT-02-2022-0082>
- Taghizadeh, F., Mokhtarani, B., & Rahmanian, N. (2023). Air pollution in Iran: The current status and potential solutions. *Environmental Monitoring and Assessment*, 195(6), Article 737. <https://doi.org/10.1007/s10661-023-11342-9>
- Tırnık, S. (2025). Machine learning-based forecasting of air quality index under long-term environmental patterns: A comparative approach with XGBoost, LightGBM, and SVM. *PLOS ONE*, 20(10), Article e0334252. <https://doi.org/10.1371/journal.pone.0334252>
- Wang, S., McGibbon, J., & Zhang, Y. (2024). Predicting high-resolution air quality using machine learning: Integration of large eddy simulation and urban morphology data. *Environmental Pollution*, 344, Article 123371. <https://doi.org/10.1016/j.envpol.2024.123371>
- Xin, N. E. (2024). *Data driven particulate matters prediction in Klang Valley using machine learning techniques* (Doctoral dissertation, University of Malaya, Kuala Lumpur, Malaysia). University of Malaya Student Repository. <http://hdl.handle.net/10197/13421>
- Zhang, L., Wilson, J. P., Zhao, N., Zhang, W., & Wu, Y. (2022). The dynamics of cardiovascular and respiratory deaths attributed to long-term PM_{2.5} exposures in global megacities. *Science of The Total Environment*, 842, Article 156951. <https://doi.org/10.1016/j.scitotenv.2022.156951>
- Zhang, Q., Liu, J., Wei, N., Song, C., Peng, J., Wu, L., & Mao, H. (2023). Identify the contribution of vehicle non-exhaust emissions: A single particle aerosol mass spectrometer test case at typical road environment. *Frontiers of Environmental Science & Engineering*, 17(5), Article 62. <https://doi.org/10.1007/s11783-023-1662-8>
- Zhong, J. (2016). *Modelling air pollution within a street canyon* (Doctoral dissertation, University of Birmingham, Birmingham, UK). University of Birmingham Research Archive. <https://etheses.bham.ac.uk/id/eprint/6890/>
- Zou, C., Wu, L., Wang, Y., Sun, S., Wei, N., Sun, B., ... & Mao, H. (2023). Evaluating traffic emission control policies based on large-scale and real-time data: A case study in central China. *Science of the Total Environment*, 860, Article 160435. <https://doi.org/10.1016/j.scitotenv.2022.160435>