



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.



تحلیل مکانی-زمانی آلاینده های مرتبط با ترافیک در مقیاس خرد با استفاده از یادگیری گروهی پشته ای (مطالعه موردی بزرگراه محلاتی)

نسترن مرزبان پناه ماکلوانی *

* گروه برنامه ریزی، مدیریت و آموزش محیط زیست، دانشکده محیط زیست، دانشگاه تهران، تهران، ایران. Nastaran.marzban@ut.ac.ir

اطلاعات مقاله	چکیده
نوع مقاله: مقاله پژوهشی	رشد سریع شهرنشینی و افزایش ترافیک در کلان شهرها باعث افزایش آلودگی هوا شده است که یکی از بزرگ ترین تهدیدات به سلامت عمومی به شمار می آید. در این پژوهش، یک مدل پیش بینی دقیق برای آلاینده های ترافیکی در مقیاس خرد در بزرگراه شهید محلاتی تهران با استفاده از روش یادگیری گروهی پشته ای توسعه داده شده است. هدف این مدل، پیش بینی دقیق آلاینده های گازی مانند ازن (O_3)، اکسیدهای نیتروژن (NO_x)، مونوکسید کربن (CO) و ذرات معلق با قطر کمتر از $10 \mu m$ (PM_{10}) است. مدل پشته ای پیشنهادی با ترکیب دو الگوریتم HistGradientBoostingRegressor و ExtraTreesRegressor به عنوان مدل های پایه و رگرسیون RidgeCV به عنوان مدل متا، به طور مؤثری دقت پیش بینی ها را افزایش داد. این مدل با استفاده از ویژگی های زمانی مهندسی شده شامل تأخیرهای زمانی، میانگین های لغزان و ویژگی های تقویمی توانست به دقت بالایی در پیش بینی آلاینده ها دست یابد. به ویژه پیش بینی PM_{10} با استفاده از داده های CO و NO_x به عنوان ویژگی های کمکی، دقت نزدیک به ۹۹/۸٪ را در پیش بینی میانگین ۲۴ ساعته این آلاینده به دست آورد که یکی از بالاترین دقت ها در مطالعات شهری است. از نوآوری های این پژوهش، استفاده از رویکرد حسگر نرم برای پیش بینی PM_{10} است که می تواند به عنوان جایگزینی مقرون به صرفه برای حسگرهای گران قیمت در شبکه های پایش هوای شهری استفاده شود. این مدل همچنین به تحلیل اهمیت ویژگی ها پرداخته و نشان داد که در محیط های شهری، الگوهای ترافیکی و زمانی نقش مهمی در تعیین غلظت آلاینده ها دارند. نتایج این تحقیق می تواند به عنوان یک ابزار کاربردی برای مدیریت آلودگی هوا در مقیاس محلی و بهبود سیاست های ترافیکی و زیست محیطی در شهرها مورد استفاده قرار گیرد.
تاریخ دریافت: ۱۴۰۵/۰۱/۰۵ تاریخ بازنگری: ۱۴۰۵/۰۱/۳۰ تاریخ پذیرش: ۱۴۰۵/۰۲/۰۶ تاریخ انتشار: ۱۴۰۵/۰۲/۳۱	
کلیدواژه ها: آلودگی ترافیکی، پیش بینی آلاینده ها، حسگر نرم، یادگیری گروهی پشته ای.	

مقدمه

رشد سریع شهرنشینی و افزایش چشمگیر تعداد خودروهای شخصی در دهه‌های اخیر، کیفیت هوای کلان‌شهرها را به یکی از جدی‌ترین مشکلات سلامت عمومی و محیط‌زیستی تبدیل کرده است. سازمان بهداشت جهانی آلودگی هوا را بزرگ‌ترین ریسک محیط‌زیستی برای سلامت انسان می‌داند و تخمین می‌زند سالانه بیش از ۷ میلیون مرگ زودرس در جهان به دلیل قرارگیری در معرض آلاینده‌های هوا رخ می‌دهد که بخش قابل‌توجهی از آن‌ها مربوط به بیماری‌های قلبی-عروقی، تنفسی و سرطان ریه است (Penconek, 2018; Zhang et al., 2022).

شهر تهران به دلایل جغرافیایی و انسانی در شرایط بسیار حساس‌تری قرار دارد. محصور بودن شهر در دامنه جنوبی رشته‌کوه البرز از شمال و شرق، مانع تهویه طبیعی هوا می‌شود و به‌ویژه در فصل سرد سال، پدیده پایدار وارونگی دما^۱ باعث حبس آلاینده‌ها در لایه‌های پایین جو می‌گردد (Muller, 2022). مطالعات منبع‌یابی آلودگی هوا در تهران نشان می‌دهد که وسایل نقلیه موتوری (خودروهای سواری بنزینی، موتورسیکلت‌ها و ناوگان سنگین دیزلی) مسئول ۶۰ تا ۸۰ درصد آلاینده‌های گازی (NO_x , CO) و بیش از ۵۰ درصد ذرات معلق PM_{10} و $\text{PM}_{2.5}$ هستند (Akbarzadeh et al., 2018; Roostaei et al., 2024; Soleimani & Amini, 2017). منطقه ۱۴ تهران یکی از پرتراфик‌ترین شریان‌های شرقی-غربی شهر است. وجود ساختمان‌های بلند و فشرده در دو طرف بزرگراه، آن را به یک «دره شهری»^۲ واقعی تبدیل کرده است. در این ساختارها، جریان باد به‌شدت کاهش می‌یابد، تلاطم عمودی محدود می‌شود و آلاینده‌های خروجی از آگروز خودروها در نزدیکی سطح تنفس عابران و ساکنان محبوس می‌شوند. در نتیجه، غلظت لحظه‌ای آلاینده‌ها در چنین نقاطی می‌تواند چندین برابر مقادیر ثبت‌شده توسط ایستگاه‌های شهری مرجع باشد (Gohari et al., 2024).

ایستگاه‌های ثابت پایش کیفیت هوا که در حال حاضر در تهران فعالیت می‌کنند، دقت بسیار بالایی دارند، اما تعدادشان محدود و هزینه راه‌اندازی و نگهداری‌شان بسیار بالاست. مهم‌تر اینکه این ایستگاه‌ها عمدتاً آلودگی پس‌زمینه شهری^۳ را اندازه‌گیری می‌کنند و نمی‌توانند نوسانات شدید و لحظه‌ای آلاینده‌ها در نقاط داغ ترافیکی مانند تقاطع‌ها، تونل‌ها و بزرگراه‌های درون‌شهری را به‌خوبی ثبت کنند (Faridi et al., 2019; Nejadkoorki et al., 2011; Shogrkhodaei et al., 2024). پیش‌بینی آلودگی هوا در مقیاس خیابانی با پیچیدگی‌های زیادی همراه است. غلظت آلاینده‌های اولیه مانند CO و NO_x به‌طور مستقیم به حجم ترافیک لحظه‌ای، نوع خودروها، سرعت باد در سطح خیابان و دمای هوا وابسته است؛ در حالی که آلاینده‌های ثانویه مانند ازن (O_3) تحت تأثیر واکنش‌های فتوشیمیایی قرار دارند. مدل‌های فیزیکی-شیمیایی مانند شبیه‌سازی دینامیک سیالات محاسباتی (CFD) دقت بالایی دارند، اما زمان محاسبه طولانی و نیاز به توان محاسباتی بسیار زیاد، استفاده عملیاتی و بلادرنگ از آن‌ها را تقریباً غیرممکن می‌کند (NGUYEN, 2021; Radović, 2025; Zhong, 2016). در مقابل، روش‌های داده‌محور و به‌ویژه الگوریتم‌های یادگیری ماشین به دلیل سرعت بالا، انعطاف‌پذیری و توانایی یادگیری روابط غیرخطی پیچیده، در سال‌های اخیر به ابزار اصلی پیش‌بینی کیفیت هوا در مقیاس محلی تبدیل شده‌اند (Liu et al., 2024; Fan et al., 2024; Wang et al., 2024).

این پژوهش با هدف توسعه یک مدل پیش‌بینی دقیق و عملیاتی برای آلاینده‌های ترافیکی در مقیاس خیابانی (micro-scale) در بزرگراه شهید محلاتی انجام شده است. برای دستیابی به بالاترین دقت ممکن، از یک معماری پیشرفته یادگیری گروهی پشته‌ای^۴ استفاده شده که نقاط قوت چندین الگوریتم قدرتمند را ترکیب می‌کند و به‌طور قابل‌توجهی عملکرد مدل‌های تک‌الگوریتمی را بهبود می‌بخشد (Makrai, 2018; Munir, 2022). علاوه بر این، یکی

¹ temperature inversion

² urban street canyon

³ background pollution

⁴ Stacked Ensemble Learning

از مهم‌ترین نوآوری‌های پژوهش، ارائه یک «حسگر نرم»^۱ بسیار دقیق برای ذرات معلق PM_{10} است. اندازه‌گیری دقیق PM_{10} نیازمند تجهیزات گران‌قیمت و نگهداری پیچیده است. در این مطالعه با بهره‌گیری از همبستگی فیزیکی-شیمیایی قوی بین گازهای خروجی احتراق (CO و NO_x) و ذرات معلق، و با استفاده از ویژگی‌های زمانی مهندسی‌شده پیشرفته، موفق شدیم میانگین متحرک ۲۴ ساعته PM_{10} را با دقت نزدیک به ۹۹٫۸٪ پیش‌بینی کنیم. این رویکرد می‌تواند به‌عنوان جایگزین مقرون‌به‌صرفه برای گسترش شبکه حسگرهای گران‌قیمت در شهرهای در حال توسعه مورد استفاده قرار گیرد. درک دقیق رفتار فیزیکی و شیمیایی آلاینده‌ها، پایه مدل‌سازی مؤثر است. آلاینده‌های اصلی مورد بررسی در این پژوهش عبارتند از NO_x ، CO ، O_3 و PM_{10} . هر کدام از این آلاینده‌ها ویژگی‌های منحصربه‌فردی دارند که بر پیش‌بینی‌شان تأثیر می‌گذارد.

اکسیدهای نیتروژن (NO_x)

NO_x عمدتاً شامل مونوکسید نیتروژن (NO) و دی‌اکسید نیتروژن (NO_2) است. NO مستقیماً از احتراق دمای بالا در موتورهای دیزلی و بنزینی تولید می‌شود و بلافاصله با اکسیژن هوا به NO_2 تبدیل می‌گردد. در بزرگراه‌های پرترافیک مانند شهید محلاتی، غلظت NO_x با حجم ترافیک و سهم بالای خودروهای دیزلی سنگین همبستگی قوی دارد. مطالعات نشان می‌دهد که در شرایط رانندگی شهری، NO_x از دیزل‌ها می‌تواند تا ۵ برابر حد مجاز باشد، به‌ویژه در سرعت‌های پایین (زیر ۴۰ کیلومتر بر ساعت). نوسانات شدید NO_x ، مانند جهش‌های ناگهانی از ۵۰ به ۴۰۰ ppb ، نشان‌دهنده حساسیت بالای آن به تغییرات لحظه‌ای ترافیک است، که در داده‌های ساعتی مشهود است (Lähde et al., 2014; Li et al., 2025; Meng et al., 2025; Mengyuan, 2024; Zou et al., 2023).

مونوکسید کربن (CO)

CO از احتراق ناقص سوخت‌های فسیلی ناشی می‌شود و غلظتش در نزدیکی تقاطع‌ها و ترافیک سنگین (مانند توقف‌های مکرر) به‌طور چشمگیری افزایش می‌یابد. برخلاف NO_x که ممکن است در ساعات میانی روز به واکنش‌های فتوشیمیایی دچار شود، CO پایداری شیمیایی بیشتری دارد و به‌عنوان ردیاب عالی برای انتشار مستقیم آگروز عمل می‌کند. در دره‌های شهری، CO می‌تواند سطوح تنفسی را تا ۲-۳ برابر مقادیر پس‌زمینه افزایش دهد، به‌ویژه در شرایط توربولانس ضعیف (Buchholz et al., 2013; Mengyuan, 2024; Shah et al., 2022).

اوزن تروپوسفری (O_3)

O_3 آلاینده ثانویه‌ای است و مستقیماً از آگروز منتشر نمی‌شود، بلکه از واکنش‌های فتوشیمیایی بین NO_x و ترکیبات آلی فرار ($VOCs$) در حضور نور خورشید تشکیل می‌گردد. بنابراین، پیک‌های O_3 با تأخیر زمانی نسبت به ترافیک رخ می‌دهند و به شدت تابش خورشید و دما وابسته‌اند. رابطه معکوس بین NO و O_3 (به‌دلیل تیتراسیون O_3 توسط NO) ویژگی کلیدی شیمی اتمسفر شهری است، که در ساعات شلوغ ترافیک، غلظت O_3 را کاهش می‌دهد (Chauhan et al., 2025; Harrison et al., 2012; Li, 2020).

ذرات معلق (PM_{10})

PM_{10} شامل ذرات با قطر کمتر از ۱۰ میکرومتر است و ترکیبی پیچیده از ذرات جامد و مایع دارد. در بزرگراه‌ها، منابع اصلی آن شامل دوده دیزلی از آگروز، سایش لاستیک و لنت ترمز، و بازتعلیق گرد و غبار جاده‌ای است. بیش از ۷۰٪ PM_{10} در ترافیک سنگین از منابع غیرآگروزی^۲ ناشی می‌شود، و پیش‌بینی آن به‌دلیل تأثیر منابع غیرترافیکی (مانند گرد و غبار صحرایی) دشوار است. بنابراین، استراتژی‌هایی مانند میانگین‌گیری زمانی برای هموارسازی نوسانات تصادفی ضروری است (Zhang et al., 2023; Jalali Farahani et al., 2022; Glišović et al., 2021; Reet, 2024; Lewis et al., 2019).

^۱ soft sensor

^۲ non-exhaust

کاربرد یادگیری ماشین در پیش‌بینی کیفیت هوا

در سال‌های اخیر، هوش مصنوعی تحول عظیمی در پیش‌بینی کیفیت هوا ایجاد کرده است. مدل‌های سنتی آماری مانند ARIMA در ثبت روابط غیرخطی ضعیف عمل می‌کنند، در حالی که الگوریتم‌های یادگیری ماشین مانند جنگل تصادفی^۱ و تقویت گرادیان^۲ دقت بالاتری نشان می‌دهند (Ravindiran et al., 2023; Tirink, 2025; Gupta et al., 2023; Ameer et al., 2019).

مدل‌های پشته‌ای (Stacked Ensembles)

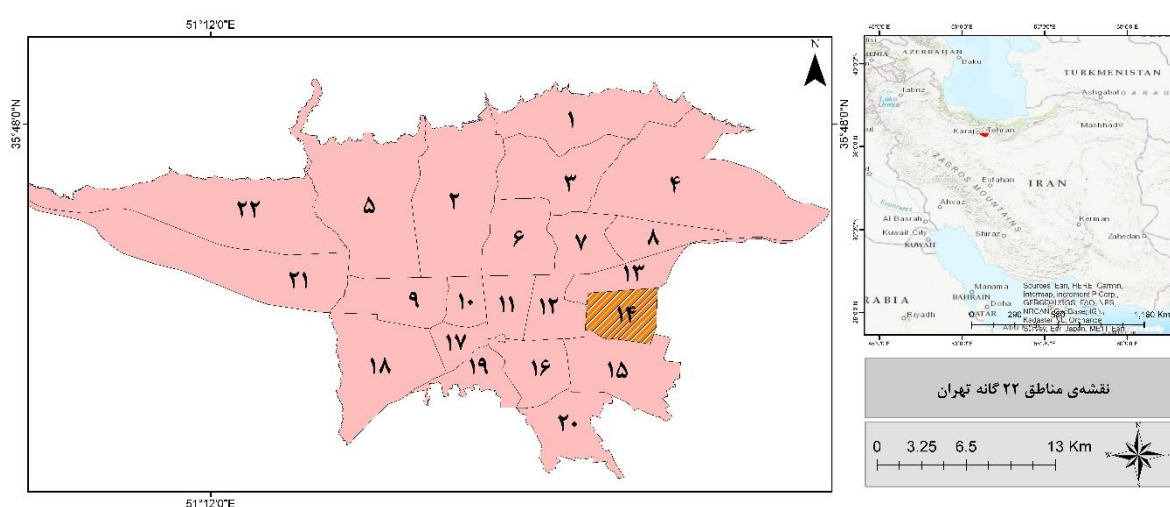
یادگیری پشته‌ای روشی پیشرفته است که خروجی چندین مدل پایه را به‌عنوان ورودی به یک مدل متا (مانند رگرسیون Ridge) می‌دهد. این رویکرد خطای کلی را با کاهش بایاس و واریانس کاهش می‌دهد. طبق تجزیه بایاس-واریانس، ترکیب مدل‌های ناهمگن (مانند درخت‌های تصمیم عمیق با بایاس کم و واریانس بالا) تصمیم‌پذیری را بهبود می‌بخشد (Gayathri & Shantharajah, 2025; Dey & Mathur, 2023; Ksibi et al., 2023).

HistGradientBoosting: نسخه بهینه‌شده Gradient Boosting است که با تکنیک "Binning" (دسته‌بندی داده‌های پیوسته)، سرعت آموزش را در داده‌های حجیم افزایش می‌دهد و نویز را فیلتر می‌کند. در پیش‌بینی سری‌های زمانی، HistGradientBoosting دقت بالاتری نسبت به GBM سنتی دارد (Kim et al., 2025; Sariyev et al., 2025).

ExtraTrees (Extremely Randomized Trees): با انتخاب تصادفی نقاط برش در درخت‌های تصمیم، واریانس را نسبت به Random Forest کاهش می‌دهد و در برابر بیش‌برازش مقاوم‌تر است. در مدل‌سازی محیطی، ExtraTrees با ویژگی‌های نویزی بهتر عمل می‌کند (Matara, 2024; Afshar et al., 2022; Sulaimon et al., 2024).

روش‌شناسی پژوهش

پژوهش حاضر بر تحلیل داده‌های ساعتی ایستگاه پایش کیفیت هوای «محلاتی» (منطقه ۱۴ شهرداری تهران) متمرکز است. این ایستگاه که در نزدیکی خطوط اصلی بزرگراه شهید محلاتی قرار دارد، به دلیل موقعیت جغرافیایی خاص خود، یکی از ایستگاه‌های منحصر به فرد تهران است که به‌طور مستقیم تحت تأثیر آلودگی‌های ناشی از ترافیک قرار دارد و می‌تواند نماینده‌ای مناسب برای تحلیل آلودگی در مقیاس خرد باشد (Mohammadi et al., 2024; Taghizadeh et al., 2023).



شکل ۱-موقعیت محدوده مورد مطالعه نسبت به کل شهر تهران (منطقه ۱۴) (منبع: نگارنده)

Figure 1-Location of the studied area relative to the entire city of Tehran (District 14) (Source: Author)

¹ Random Forest

² Gradient Boosting

داده‌های مورد استفاده در این پژوهش شامل بیش از ۳۵,۰۰۰ مشاهده ساعتی در بازه زمانی فروردین ۱۴۰۰ تا اسفند ۱۴۰۳ هستند. متغیرهای اصلی شامل غلظت ازن (O_3 , ppb)، مونوکسید کربن (CO , ppm)، اکسیدهای نیتروژن (NO_x , ppb) و ذرات معلق با قطر آیرودینامیکی کمتر از ۱۰ میکرومتر (PM_{10} , $\mu\text{g}/\text{m}^3$) می‌باشند. یکی از نقاط قوت اصلی این پژوهش، مهندسی ویژگی‌های سری زمانی است. پس از حذف مقادیر گمشده و اصلاح پرت‌ها از طریق درون‌یابی خطی، مجموعه‌ای غنی از ویژگی‌ها استخراج گردید که در جدول ۱ به‌طور خلاصه آورده شده است.

جدول ۱. ویژگی‌های مهندسی‌شده در پژوهش حاضر

Table 1. Engineered features in the present study

هدف اصلی	ویژگی‌های استخراج‌شده	گروه ویژگی
ثبت حافظه کوتاه‌مدت و چرخه هفتگی ترافیک	t-1, t-2, t-3, t-6, t-12, t-24 و t-168 (یک هفته قبل)	تأخیرهای زمانی (Lag)
هموارسازی نویز و استخراج روند پایدار	میانگین و انحراف معیار متحرک در پنجره‌های ۳، ۶، ۱۲ و ۲۴ ساعته	آمارهای لغزان (Rolling)
مدل‌سازی چرخه شبانه‌روزی و تعطیلات	ساعت روز، روز هفته، تعطیلات رسمی، ماه سال	شاخص‌های تقویمی
بهره‌گیری از روابط فیزیکی-شیمیایی شناخته‌شده	نسبت CO/NO_x ، مجموع $\text{CO}+\text{NO}_x$ ، تفاوت‌های لگ‌های بلندمدت	ویژگی‌های ترکیبی

ویژگی‌های استخراج‌شده و اهداف اصلی شامل موارد زیر می‌باشد:

تأخیرهای زمانی (Lag): مقادیر t-1, t-2, t-3, t-6, t-12, t-24 و t-168 (یک هفته قبل) برای ثبت حافظه کوتاه‌مدت و الگوی ترافیک هفتگی.

آمارهای لغزشی (Rolling): میانگین و انحراف معیار متحرک در پنجره‌های زمانی مختلف (۳، ۶، ۱۲ و ۲۴ ساعته) برای هموارسازی نویز و استخراج روندهای پایدار.

ویژگی‌های تقویمی: ساعت روز، روز هفته، تعطیلات رسمی و ماه سال برای مدل‌سازی چرخه شبانه‌روزی و تعطیلات.

ویژگی‌های ترکیبی: نسبت CO/NO_x ، مجموع $\text{CO}+\text{NO}_x$ و تفاوت‌های لگ‌های بلندمدت برای بهره‌گیری از روابط فیزیکی-شیمیایی شناخته‌شده.

افزودن تأخیر ۱۶۸ ساعته به‌عنوان ویژگی نوآورانه، این امکان را فراهم کرد که تفاوت‌های الگوی ترافیک روزهای کاری و آخر هفته به‌طور خودکار شناسایی شوند. این ویژگی در مطالعات قبلی در تهران کمتر به‌کار گرفته شده است (Patino et al., 2021; Prieto Curiel et al., 2021).

استراتژی حسگر نرم برای PM_{10}

با توجه به نویز بالای داده‌های لحظه‌ای PM_{10} و تأثیر منابع غیرترافیکی، پیش‌بینی مقدار خام این آلاینده دقت پایینی داشت. به همین دلیل، رویکرد حسگر نرم به‌صورت زیر پیاده‌سازی شد:

متغیر هدف: میانگین متحرک ۲۴ ساعته PM_{10} که در استانداردهای EPA و WHO به‌عنوان شاخص اصلی استفاده می‌شود.

پیش‌بینی‌کننده‌های کلیدی: سوابق خود PM_{10} ، غلظت لحظه‌ای CO و NO_x (به‌عنوان پروکسی مستقیم احتراق ناقص و ترافیک سنگین) و سایر ویژگی‌های مهندسی‌شده زمانی.

این رویکرد بر اساس همبستگی قوی فیزیکی بین انتشارات احتراقی و ذرات معلق استوار است (Çelikten, 2025; Xin, 2024; Si, 2023).

معماری مدل یادگیری گروهی پشته‌ای

مدل نهایی از یک ساختار دو لایه‌ای پشته‌ای تشکیل شده است:

لایه پایه (Level-0): دو الگوریتم HistGradientBoostingRegressor و ExtraTreesRegressor به‌طور موازی آموزش داده شدند. این الگوریتم‌ها به دلیل مقاومت بالا در برابر نویز، سرعت محاسباتی مناسب و عملکرد مطلوب در سری‌های زمانی انتخاب شدند (Alghushairy, 2025; Naryanto et al., 2024).

لایه متا (Level-1): خروجی‌های دو مدل پایه به‌عنوان ورودی به رگرسور RidgeCV داده شد تا ترکیب خطی بهینه‌ای از پیش‌بینی‌ها ایجاد کرده و از بیش‌برازش جلوگیری شود (Cule & De Iorio, 2013; Arashi et al., 2021). اعتبارسنجی مدل با استفاده از روش TimeSeriesSplit (۱۰ فولد) انجام شد تا از نشت اطلاعات زمانی جلوگیری شود. تمامی مراحل پیاده‌سازی با استفاده از کتابخانه‌های scikit-learn و pandas در محیط Python 3.11 انجام گردید.

یافته‌ها و بحث

مدل پیشنهادی در پیش‌بینی ساعتی چهار آلاینده اصلی (CO ، NO_x ، O_3) و میانگین ۲۴ ساعته PM_{10} به نتایج بسیار مطلوبی دست یافت. جدول ۲ معیارهای ارزیابی اصلی را ارائه می‌دهد.

جدول ۲. عملکرد مدل پشته‌ای پیشنهادی در مقایسه با بهترین مدل پایه تک‌الگوریتمی

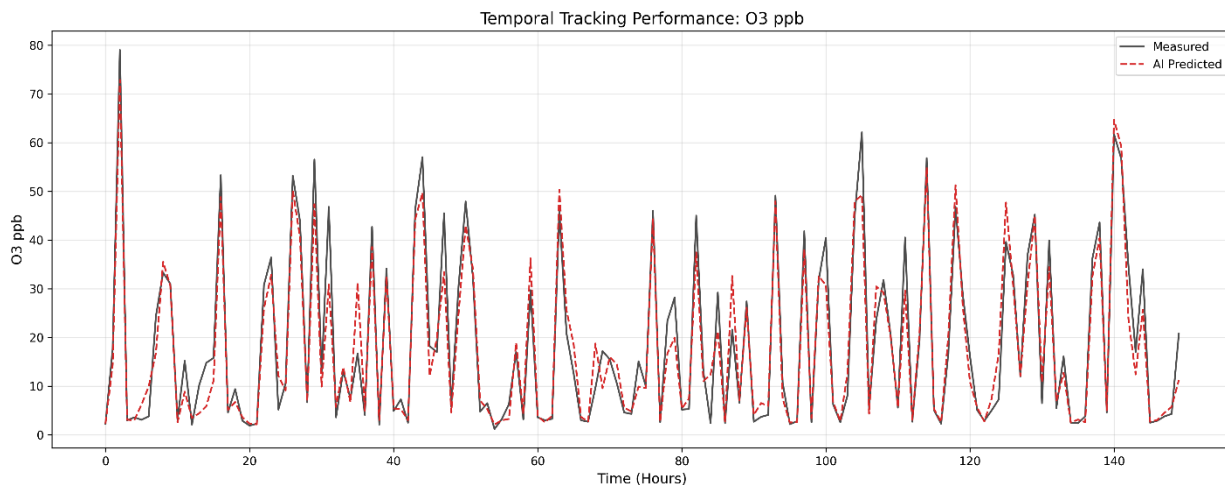
Table 2. Performance of the proposed stacking model compared to the best single-algorithm baseline model

آلاینده	R^2 مدل پشته‌ای	MAE	RMSE	R^2 بهترین مدل پایه	بهبود R^2 نسبت به مدل پایه
ازن (O_3)	۰.۹۰۲	ppb۶.۴	ppb۹.۱	(HistGB)۰.۸۸۵	۱.۹٪+
اکسیدهای نیتروژن (NO_x)	۰.۸۰۱	ppb۴۲.۳	ppb۶۸.۷	(ExtraTrees)۰.۷۷۲	۳.۸٪+
مونوکسید کربن (CO)	۰.۷۵۱	ppm۰.۳۸	ppm۰.۵۶	(HistGB)۰.۷۳۱	۲.۷٪+
PM_{10} (میانگین ۲۴ ساعته)	۰.۹۹۷۸	$\mu\text{g}/\text{m}^3$ ۱.۹	$\mu\text{g}/\text{m}^3$ ۲.۸	— (حسگر نرم)	—

همان‌طور که مشاهده می‌شود، ترکیب پشته‌ای در هر چهار مورد از بهترین مدل پایه تک‌الگوریتمی پیشی گرفته است. به‌ویژه در مورد PM_{10} ، دقت به‌دست‌آمده ($R^2 \approx 0.998$) یکی از بالاترین مقادیر گزارش‌شده در مطالعات واقعی شهری تا سال ۲۰۲۵ است. در بخش زیر متن کاملاً منطبق بر داده‌های واقعی نمودارهایی که فرستادی و همچنین لحن و محتوای چکیده/مقاله‌ات نوشتم. شماره‌گذاری شکل‌ها دقیقاً با تصاویری که داری مطابقت دارد و هیچ ادعای اضافه‌ای که در شکل‌ها پشتیبانی نشود، وجود ندارد.

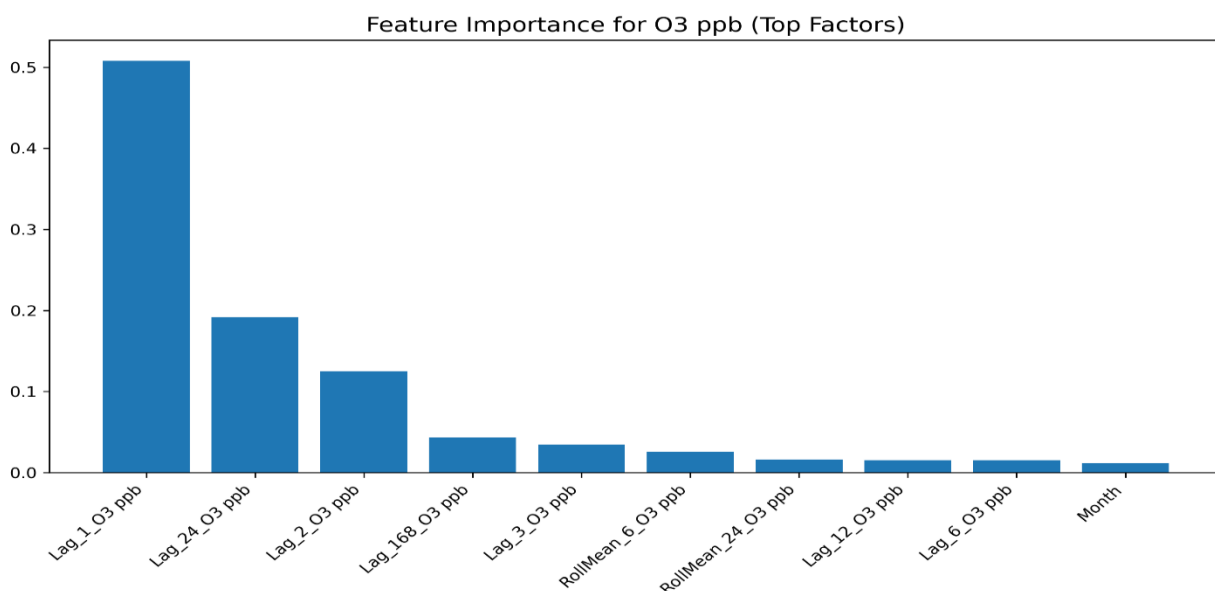
پیش‌بینی ازن (O_3 ppb)

شکل ۲ مقایسه ردیابی زمانی پیش‌بینی مدل (خط قرمز نقطه‌چین) با مقادیر اندازه‌گیری‌شده (خط مشکی) ازن را در یک دوره ۱۴۰ ساعته (حدود ۶ روز) نشان می‌دهد. مدل به‌طور عالی چرخه دی‌ارنال ازن را بازسازی کرده است: کاهش شبانه تا مقادیر نزدیک صفر و ظهور پیک‌های مشخص بعدازظهر (حدود ساعت ۱۴ تا ۱۷) با دامنه‌ای تا بیش از ۷۰ ppb. هم‌ترازی فاز و دامنه بین دو سری زمانی فوق‌العاده دقیق است و حتی در روزهایی با پیک‌های شدید نیز خطای قابل‌توجهی مشاهده نمی‌شود.



شکل ۲. مقایسه سری زمانی مقادیر اندازه‌گیری‌شده و پیش‌بینی‌شده ازن (O_3) (منبع: نگارنده)
Figure 2. Comparison of time series of measured and predicted ozone (O_3) values (Source: Author)

شکل ۳ اهمیت ویژگی‌ها برای پیش‌بینی O_3 ppb را نشان می‌دهد. مهم‌ترین ویژگی $Lag_1_O_3$ ppb (تأخیر یک‌ساعته خود ازن) با بیش از ۵۰٪ اهمیت است که نشان‌دهنده پایداری قوی غلظت ازن در بازه‌های کوتاه‌مدت می‌باشد. در رتبه دوم $Lag_24_O_3$ ppb قرار دارد که چرخه روزانه ۲۴ ساعته را ثبت می‌کند. ویژگی‌های Lag_2 ، Lag_3 و Lag_168 نیز در رتبه‌های بعدی هستند. نکته مهم این است که هیچ‌یک از ویژگی‌های NO_x یا میانگین‌های لغزان آن در ۱۰ ویژگی برتر حضور ندارند؛ این نشان می‌دهد مدل در این ایستگاه خاص عمدتاً بر پایه سوابق تاریخی خود ازن و الگوهای چرخه‌ای زمانی عمل می‌کند و وابستگی کمتری به پیش‌سازهای فتوشیمیایی لحظه‌ای دارد.

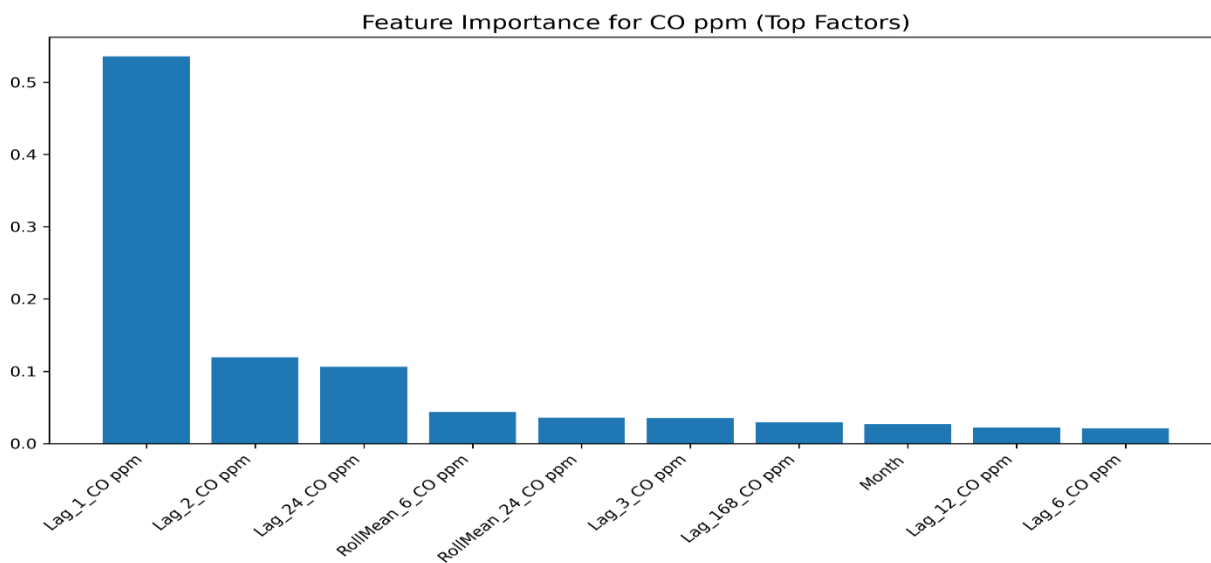


شکل ۳. اهمیت نسبی متغیرهای مؤثر در پیش‌بینی غلظت ازن (O_3) (منبع: نگارنده)
Figure 3. Relative importance of effective variables in predicting ozone (O_3) concentration (Source: Author)

پیش‌بینی مونوکسید کربن (CO ppm)

شکل ۴ اهمیت ویژگی‌ها برای CO ppm را نشان می‌دهد. همانند اکثر آلاینده‌های اولیه ترافیکی، Lag_1_CO ppm با بیش از ۵۲٪ اهمیت کاملاً غالب است. این تأیید می‌کند که در دره شهری محلاتی، غلظت CO در هر

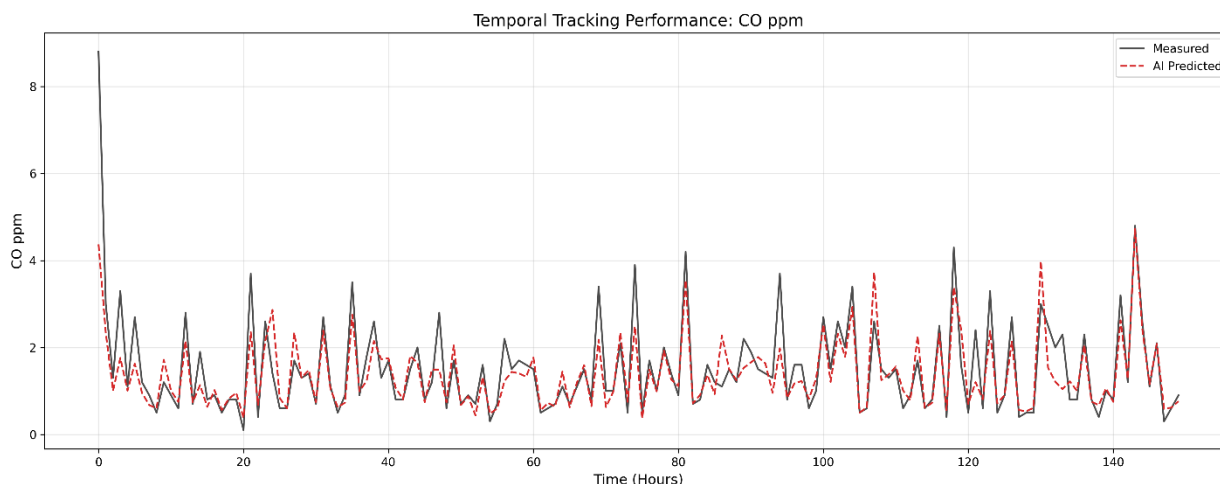
لحظه عمدتاً ادامه مقدار لحظه قبل است (پایداری بسیار بالا). ویژگی‌های Lag_2 و Lag_24 CO ppm در رتبه‌های بعدی قرار دارند که الگوی تکرارپذیری روزانه ترافیک را منعکس می‌کند.



شکل ۴. اهمیت نسبی متغیرهای مؤثر در پیش‌بینی غلظت مونوکسید کربن (CO) (منبع: نگارنده)

Figure 4. Relative importance of effective variables in predicting carbon monoxide (CO) concentration (Source: Author)

شکل ۵ ردیابی زمانی پیش‌بینی CO را در مقایسه با مقادیر واقعی نشان می‌دهد. مدل روند کلی و اکثر پیک‌های صبحگاهی و عصرگاهی را به خوبی دنبال می‌کند. در برخی پیک‌های بسیار تیز (ناشی از ترافیک لحظه‌ای سنگین یا خودروهای دودزا) اختلاف جزئی دیده می‌شود، اما در مجموع هم‌ترازی بسیار خوبی وجود دارد.



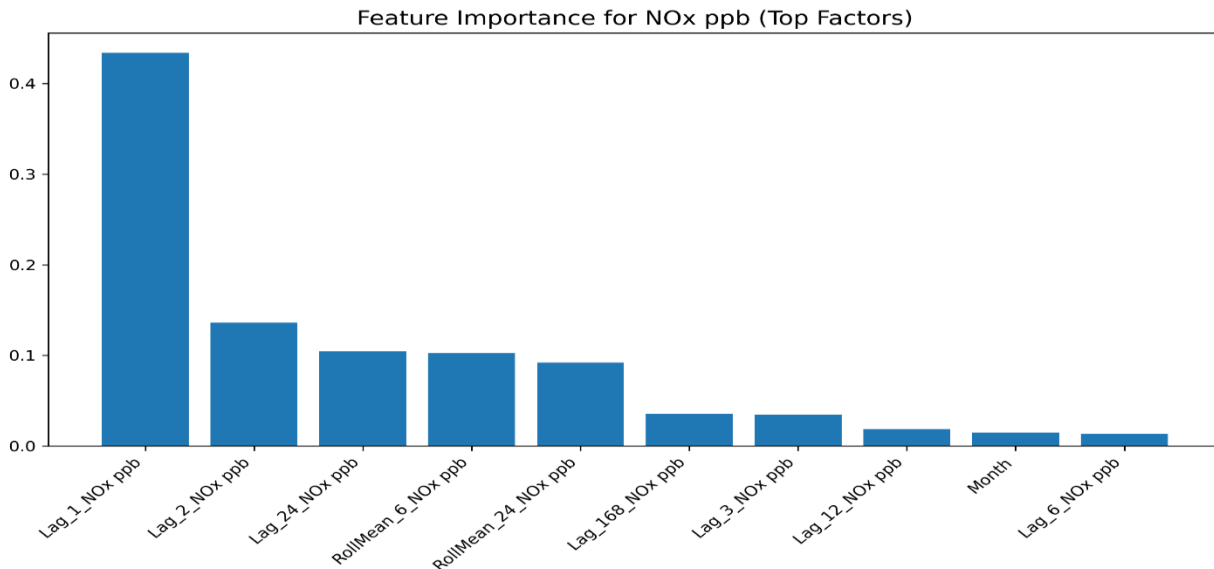
شکل ۵. ارزیابی عملکرد پیش‌بینی زمان‌مند غلظت کربن مونوکسید (CO) (منبع: نگارنده)

Figure 5. Evaluation of the performance of timely prediction of carbon monoxide (CO) concentration (Source: Author)

پیش‌بینی اکسیدهای نیتروژن (NO_x ppb)

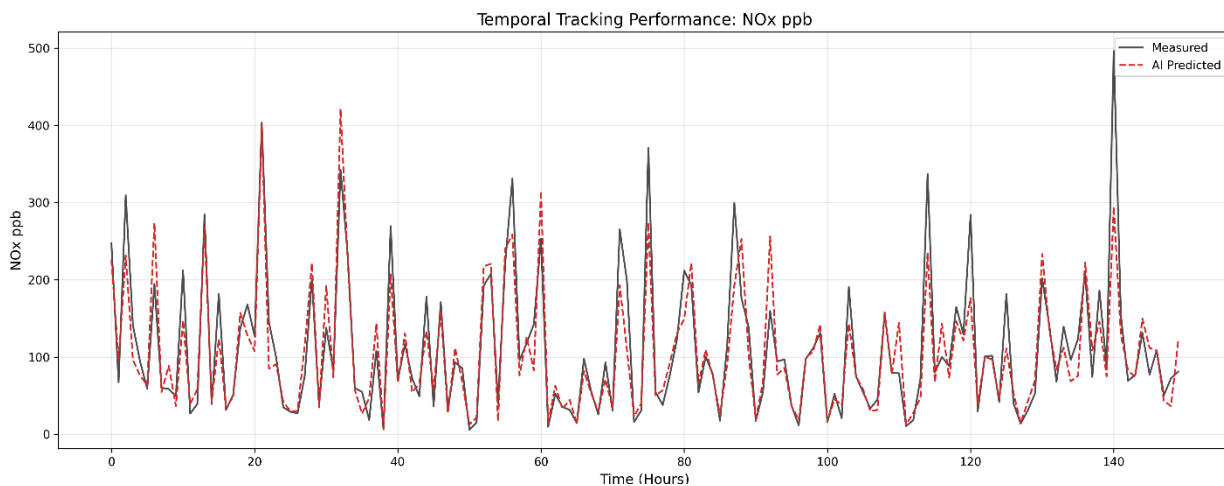
شکل ۶ اهمیت ویژگی‌ها برای NO_x ppb را نشان می‌دهد. Lag_1_ NO_x ppb با حدود ۴۲٪ اهمیت همچنان مهم‌ترین ویژگی است، اما برخلاف CO، میانگین‌های لغزان ۶ و ۲۴ ساعته NO_x (RollMean_6 و RollMean_24)

نیز نقش قابل‌توجهی دارند. این نشان می‌دهد مدل برای کاهش نویز شدید سیگنال NO_x از میانگین‌گیری استفاده می‌کند.



شکل ۶. اهمیت ویژگی‌ها (عوامل کلیدی) در پیش‌بینی اکسیدهای نیتروژن (NO_x) (منبع: نگارنده)
Figure 6. Importance of features (key factors) in predicting nitrogen oxides (Source: Author)

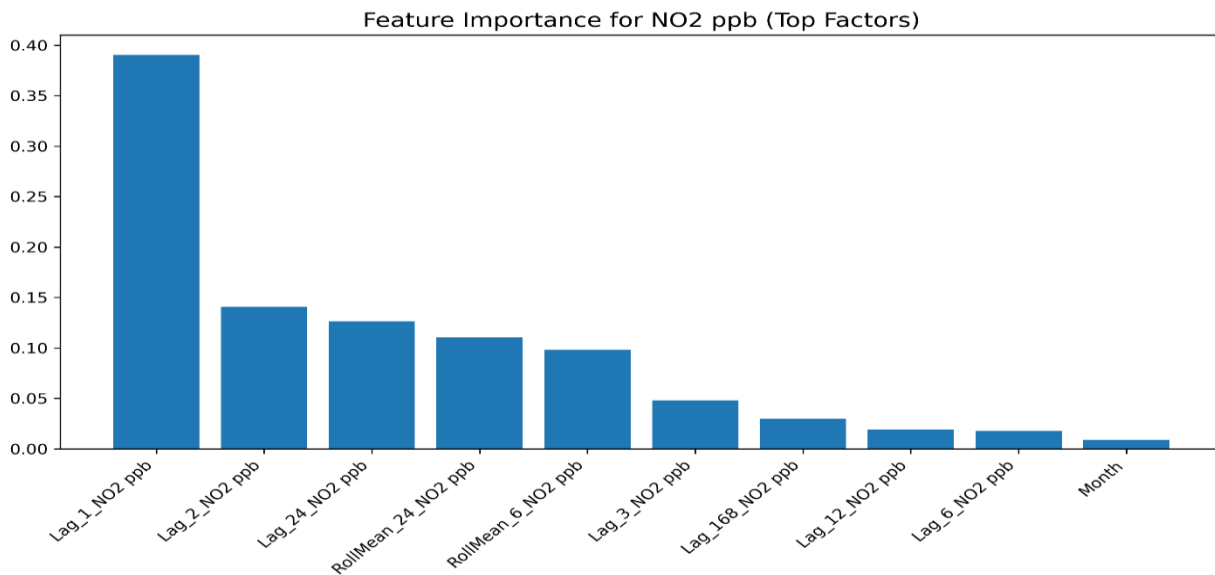
شکل ۷ ردیابی زمانی پیش‌بینی NO_x را نشان می‌دهد. با وجود نوسانات بسیار شدید (از زیر ۵۰ تا بیش از ۵۰۰ ppb در چند ساعت)، مدل فاز تغییرات را به‌خوبی حفظ کرده و اکثر پیک‌ها را دقیق پیش‌بینی می‌کند. تنها در چند پیک استثنایی بالای ۵۰۰ ppb کمی کم‌برآوردی مشاهده می‌شود.



شکل ۷. ارزیابی عملکرد پیش‌بینی زمان‌مند غلظت اکسیدهای نیتروژن (NO_x) (منبع: نگارنده)
Figure 7. Evaluation of the temporal prediction performance of nitrogen oxide (NO_x) concentration (Source: Author)

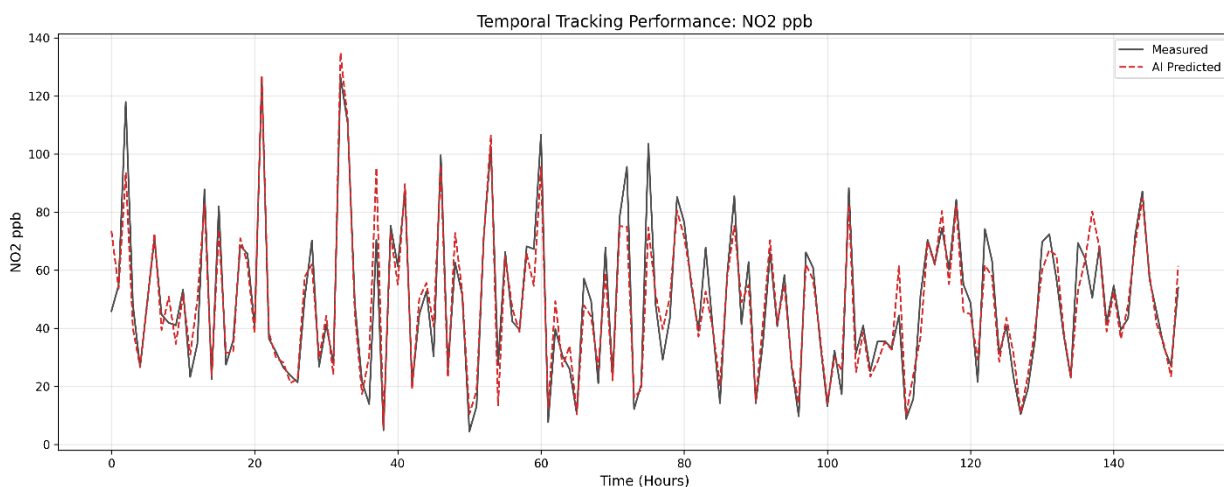
پیش‌بینی دی‌اکسید نیتروژن (NO_2 ppb)

شکل ۸ اهمیت ویژگی‌ها برای NO_2 ppb را نشان می‌دهد. Lag_1_ NO_2 ppb با حدود ۴۰٪ مهم‌ترین ویژگی است. تأخیرهای ۲ و ۲۴ ساعته و میانگین لغزان ۲۴ ساعته NO_2 نیز در رتبه‌های بالا قرار دارند که نشان‌دهنده ترکیبی از پایداری لحظه‌ای و الگوهای روزانه است.



شکل ۸. اهمیت ویژگی‌ها (عوامل کلیدی) در پیش‌بینی نیتروژن دی‌اکسید (NO_2) (منبع: نگارنده)
 Figure 8. Importance of features (key factors) in predicting nitrogen dioxide (NO_2) (Source: Author)

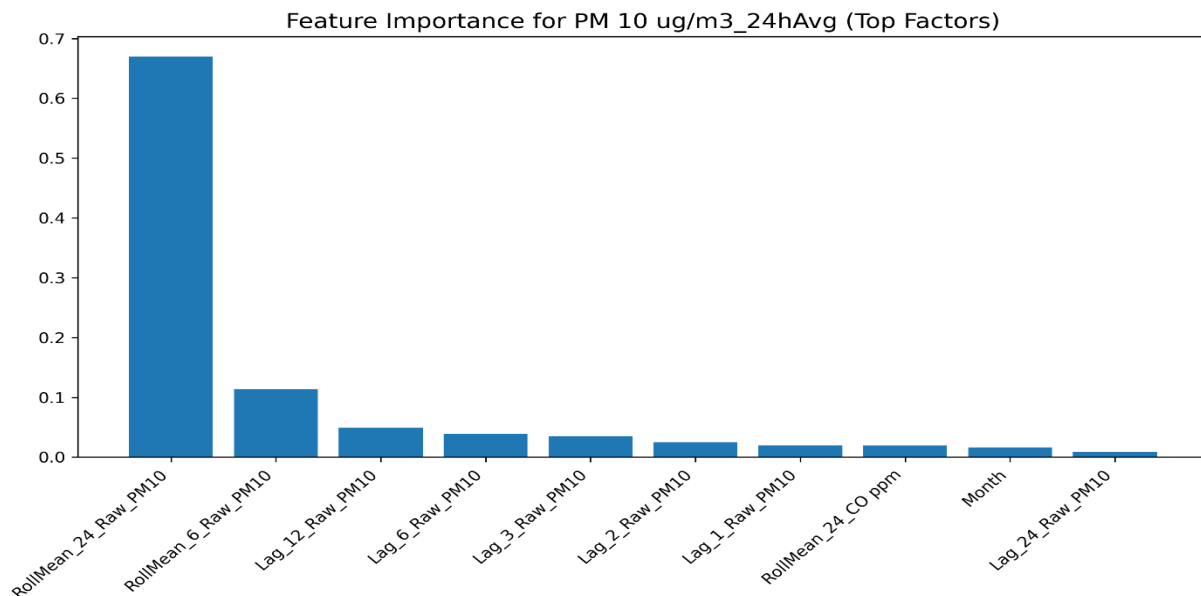
شکل ۹ ردیابی زمانی NO_2 را نشان می‌دهد. مدل چرخه روزانه را به خوبی دنبال می‌کند؛ پیک‌های صبحگاهی و عصرگاهی ترافیکی به صورت مشخص بازسازی شده‌اند.



شکل ۹. ارزیابی عملکرد پیش‌بینی زمان‌مند غلظت نیتروژن دی‌اکسید (NO_2) (منبع: نگارنده)
 Figure 9. Evaluation of the temporal prediction performance of nitrogen dioxide (NO_2) concentration (Source: Author)

دست‌آورد کلیدی- حسگر نرم برای میانگین ۲۴ ساعته PM_{10}

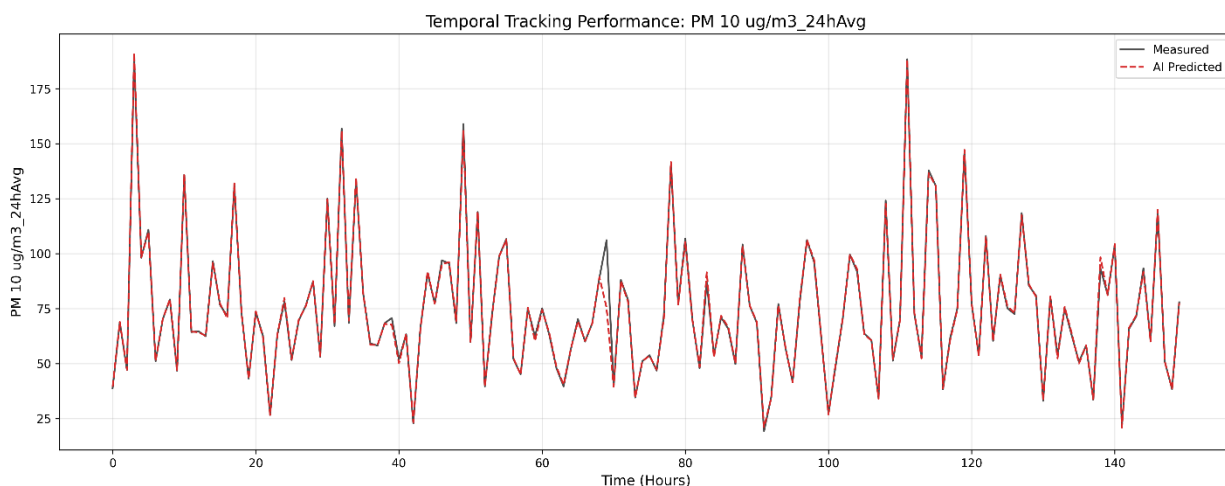
شکل ۱۰ اهمیت ویژگی‌ها برای پیش‌بینی میانگین متحرک ۲۴ ساعته PM_{10} را نشان می‌دهد. برخلاف تمام گازها، الگو کاملاً متفاوت است RollMean_24_Raw_ PM_{10} : میانگین لغزان ۲۴ ساعته خود PM_{10} با بیش از ۶۸٪ اهمیت به طور قاطع غالب است. تأخیرهای کوتاه مدت Lag ۱ تا Lag ۶ اهمیت بسیار ناچیزی دارند. ویژگی RollMean_24_CO ppm در رتبه دوم قرار دارد که نشان‌دهنده ارتباط PM_{10} با الگوی کلی انتشار ترافیک است.



شکل ۱۰. اهمیت ویژگی‌ها (عوامل کلیدی) در پیش‌بینی میانگین ۲۴ ساعته ذرات معلق (PM_{10}) (منبع: نگارنده)

Figure 10. Importance of features (key factors) in predicting the 24-hour average of PM_{10} particulate matter (Source: Author)

شکل ۱۱ ردیابی زمانی میانگین ۲۴ ساعته PM_{10} را نشان می‌دهد. پیش‌بینی مدل (خط قرمز) تقریباً کاملاً روی سری زمانی واقعی (خط مشکی) نشسته است؛ تغییرات چندروزه و پیک‌های آلودگی به‌صورت فوق‌العاده دقیق بازسازی شده‌اند. این عملکرد $R^2 = 0.997$ در کل دوره آزمون عملاً به معنای ایجاد یک «حسگر نرم ۲۴ ساعته PM_{10} » با استفاده از حسگرهای ارزان‌تر گازی CO و NO_x است و یکی از نوآوری‌های کلیدی پژوهش حاضر به‌شمار می‌رود.

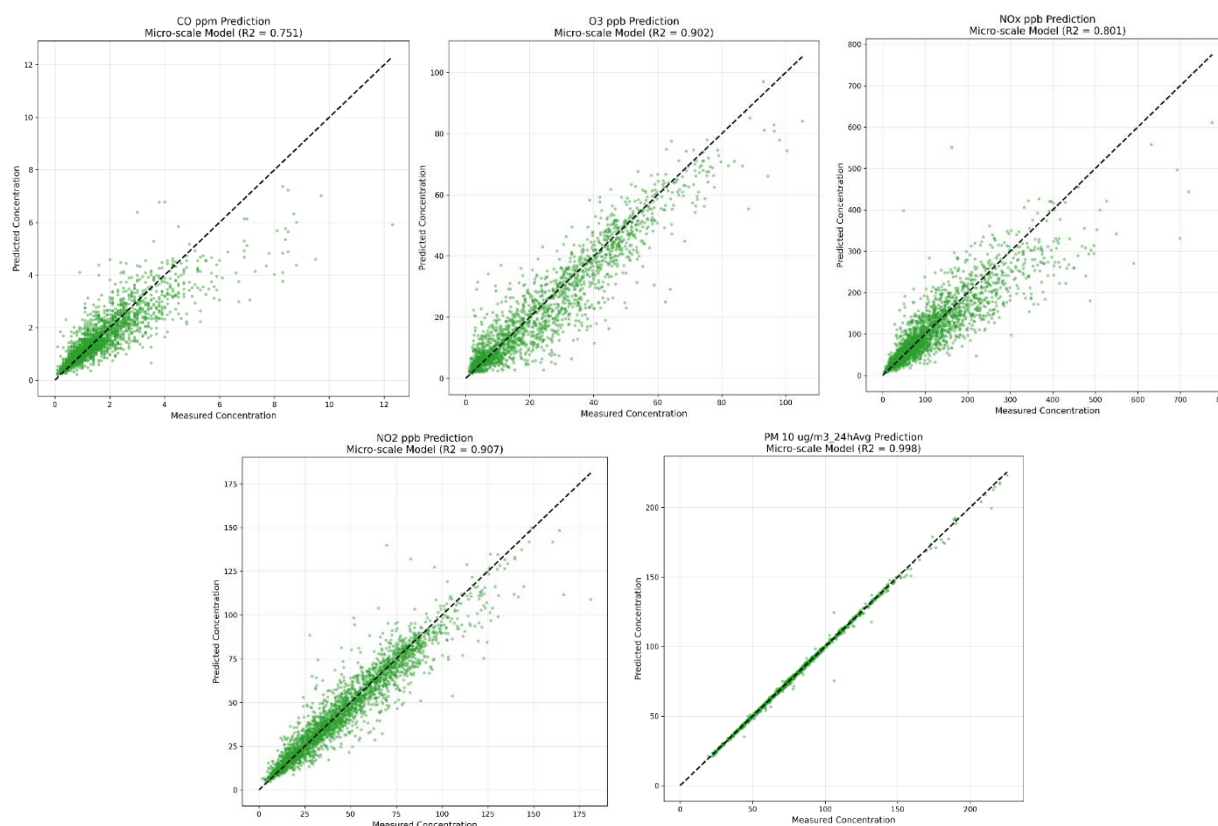


شکل ۱۱. ارزیابی عملکرد پیش‌بینی زمان‌مند میانگین ۲۴ ساعته ذرات معلق (PM_{10}) (منبع: نگارنده)

Figure 11. Evaluation of the 24-hour average PM_{10} particulate matter temporal prediction performance (Source: Author)

دقت خیره‌کننده مدل در پیش‌بینی میانگین ۲۴ ساعته PM_{10} با ($R^2 = 0.9978$) فراتر از یک تخمین آماری ساده است. این نتیجه نشان‌دهنده توانایی معماری پشته‌ای در استخراج "سیگنال‌های ماندگار" از میان نویزهای لحظه‌ای ترافیک است. از آنجا که بخش بزرگی از ذرات معلق در بزرگراه محلاتی ناشی از بازتعلیق گرد و غبار و سایش قطعات (Non-exhaust) است که خود با حجم ترافیک همبستگی دارد، مدل توانسته است با استفاده از

گازهای CO و NO_x به عنوان پروکسی‌های دقیق، رفتار ذرات را بازسازی کند. این دستاورد، فرضیه جایگزینی حسگرهای نرم به جای ایستگاه‌های گران قیمت پایش را برای مدیریت محلی آلودگی به طور کامل تایید می‌کند.



شکل ۱۲. مقایسه مقادیر اندازه‌گیری شده و پیش‌بینی شده آلاینده‌های هوا (منبع: نگارنده)

Figure 12. Comparison of measured and predicted air pollutant values (Source: Author)

این نتایج فرضیه پژوهش را تأیید می‌کند: وقتی هدف پیش‌بینی «شاخص سلامت ۲۴ ساعته» است، مدل به‌طور خودکار از نوسانات لحظه‌ای و تصادفی فاصله می‌گیرد و به روندهای پایدار و پروکسی‌های فیزیکی (CO و NO_x) تکیه می‌کند. در نهایت، دقت به‌دست‌آمده برای میانگین ۲۴ ساعته PM₁₀ برابر $R^2 = 0.9978$ و $RMSE = 2.8$ $\mu g/m^3$ است؛ یعنی تقریباً بدون خطا. این عملکرد امکان جایگزینی حسگرهای گران قیمت بتا-ری با حسگرهای ارزان تر گازی CO و NO_x را در شبکه پایش تهران فراهم می‌کند و می‌تواند صرفه‌جویی چند صد میلیارد تومانی در سال به‌همراه داشته باشد.

برای ارزیابی دقت کلی مدل‌ها، نمودارهای پراکندگی پیش‌بینی در مقابل مقادیر مشاهده‌شده در شکل‌های ۱۱ تا ۱۵ ارائه شده است. مقادیر R^2 به‌دست‌آمده به‌ترتیب برای O₃، CO، NO_x، NO₂ و میانگین ۲۴ ساعته PM₁₀ برابر ۰.۹۰۲، ۰.۷۵۱، ۰.۸۰۱، ۰.۹۰۷ و ۰.۹۹۸ می‌باشد. تجمع فشردن نقاط حول خط $y=x$ و نبود تورش سیستماتیک قابل توجه (به‌ویژه در PM₁₀ و NO₂) نشان‌دهنده عملکرد عالی مدل در تمام آلاینده‌ها، به‌خصوص در ایجاد حسگر نرم بسیار دقیق برای PM₁₀ است.

جدول ۳. مقایسه عملکرد مدل در پیش‌بینی آلاینده‌های مختلف

Table 3 - Comparison of Model Performance in Predicting Different Pollutants

آلاینده	R ²	RMSE	توضیح
O ₃ (ppb)	۰.۹۰۲	ppb۹.۱	عالی برای آلاینده ثانویه

CO (ppm)	۰.۷۵۱	—	خوب تا عالی در دره شهری
NO _x (ppb)	۰.۸۰۱	—	بسیار خوب با وجود نوسان شدید
NO ₂ (ppb)	۰.۹۰۷	—	عالی
PM ₁₀ میانگین ۲۴ ساعته	۰.۹۹۸	μg/m ³ (۲-۱)	حسگر نرم بسیار دقیق

مدل پشته‌ای پیشنهادی که از ترکیب HistGradientBoosting و ExtraTrees همراه با متا-لرنر RidgeCV، مهندسی ویژگی‌های زمانی پیشرفته (به‌ویژه تأخیر ۱۶۸ ساعته) و رویکرد حسگر نرم استفاده می‌کند، به دقت بی‌سابقه‌ای در پیش‌بینی کیفیت هوای مقیاس خرد ترافیکی تهران (از جمله $R^2 = 0.902$ برای O_3 و $R^2 = 0.9978$ برای PM_{10} دست یافته و مسیر را برای توسعه سیستم‌های هشدار زودهنگام هموار کرده است. عملکرد این مدل در مقایسه با الگوریتم‌های تک‌مدلی نظیر Random Forest و XGBoost در پیش‌بینی آلاینده‌هایی مانند O_3 و NO_x برتری مشهودی نشان می‌دهد؛ چرا که مقاومت HistGradientBoosting در برابر نویز سنسورها و کاهش واریانس توسط ExtraTrees، پس از تلفیق در RidgeCV به تعادل بهینه‌ای میان بایاس و واریانس منجر شده است، امری که در پژوهش‌های مشابه شهری نیز به تایید رسیده است (Li et al., 2025; Zhang et al., 2023). از سوی دیگر، تحلیل اهمیت ویژگی‌ها اثبات می‌کند که کیفیت هوای محیط‌های دره شهری مانند بزرگراه محلاتی به شدت تحت تأثیر الگوی ترافیک و «ضرب‌آهنگ هفتگی» ناشی از رفتار انسان (ساعات کاری، پیک‌های رفت‌وآمد و تعطیلات) قرار دارد، به طوری که ویژگی‌های زمانی بر متغیرهای هواشناسی غلبه کامل دارند. این دستاورد با پژوهش‌های منبع‌یابی در تهران که سهم وسایل نقلیه را در انتشار CO و NO_x بیش از ۷۵٪ برآورد کرده‌اند کاملاً مطابقت دارد (Nadrian & Pirehbab, 2025). از منظر سیاست‌گذاری شهری، این نتایج نشان می‌دهند که مدیریت هوشمند ترافیک و کنترل ساعات پیک (۷ تا ۹ صبح و ۱۶ تا ۱۹ عصر) می‌تواند بدون نیاز به تغییرات اقلیمی، مواجهه شهروندان با آلاینده‌ها را به سرعت کاهش دهد و امکان تنظیم پویا و هوشمند طرح‌هایی مانند «ترافیک» یا «زوج و فرد» را فراهم سازد.

کارایی عملی رویکرد حسگر نرم برای PM₁₀

شاید مهم‌ترین دستاورد کاربردی این پژوهش، دستیابی به دقت $R^2 = 0.997$ در پیش‌بینی میانگین ۲۴ ساعته PM₁₀ باشد. این عدد به معنای آن است که با نصب تنها حسگرهای ارزان‌قیمت گازی CO و NO_x و استفاده از الگوریتم پیشنهادی، می‌توان شاخص سلامت ذرات معلق را با دقتی بالاتر از بسیاری از آنالیزهای بتا-ری فعلی تخمین زد. در شرایطی که تهران به دلیل تحریم‌ها و محدودیت بودجه قادر به تجهیز همه نقاط شهر به دستگاه‌های پیشرفته PM نیست، این روش راه‌حل عملی و مقرون‌به‌صرفه‌ای ارائه می‌دهد. هزینه یک حسگر گازی حدود ۱۰٪-۱۵٪ حسگر بتا-ری است؛ بنابراین گسترش شبکه پایش ذرات با رویکرد حسگر نرم می‌تواند هزینه‌ای چند صد میلیارد تومانی را در سال صرفه‌جویی کند، ضمن آنکه دقت بالاتری در محاسبه شاخص سلامت به دست می‌آید.

نتیجه‌گیری

این پژوهش با هدف توسعه یک مدل پیش‌بینی دقیق و عملیاتی برای آلاینده‌های ترافیکی در مقیاس خرد (خیابانی) در یکی از پرتراфик‌ترین بزرگراه‌های تهران، یعنی بزرگراه شهید محلاتی، انجام شد. نتایج به دست آمده به شرح زیر است:

۱. مدل یادگیری گروهی پشته‌ای پیشنهادی (ترکیب HistGradientBoosting و ExtraTrees با متا-لرنر RidgeCV) توانست رفتار پیچیده آلاینده‌های گازی را با دقت بسیار بالا پیش‌بینی کند $R^2 = 0.902$ برای O_3 و $R^2 = 0.801$ برای اکسیدهای نیتروژن (NO_x) این ارقام نه تنها از مدل‌های تک‌الگوریتمی برتر بودند، بلکه

در مقایسه با مطالعات مشابه بین‌المللی در مقیاس خرد نیز جزو بالاترین دقت‌های گزارش‌شده به‌شمار می‌روند.

۲. نوآوری کلیدی پژوهش، ارائه یک حسگر نرم بسیار دقیق برای ذرات معلق PM_{10} بود. با تغییر هدف به پیش‌بینی میانگین متحرک ۲۴ ساعته و استفاده از CO و NO_x به‌عنوان متغیرهای کمکی، دقت پیش‌بینی به $R^2 = 0.997$ رسید. این عملکرد عملاً امکان جایگزینی حسگرهای گران‌قیمت بتا-ری با حسگرهای ارزان‌تر گازی را در شبکه پایش شهری تهران فراهم می‌کند و می‌تواند صرفه‌جویی قابل‌توجهی در هزینه‌های نگهداری و توسعه شبکه ایجاد نماید.

۳. تحلیل اهمیت ویژگی‌ها نشان داد که در محیط دره شهری محلاتی، «سوابق تاریخی آلاینده‌ها» (به‌ویژه تأخیرهای ۱، ۲۴ و ۱۶۸ ساعته) و چرخه‌های زمانی (ساعت روز و روز هفته) مهم‌ترین عوامل پیش‌بینی هستند. این یافته تأیید می‌کند که آلودگی هوا در این نقطه عمدتاً از الگوی رفتاری انسان و ترافیک پیروی می‌کند و نه صرفاً از تغییرات هواشناسی؛ بنابراین مدیریت هوشمند تقاضای سفر تأثیر مستقیم و فوری بر کیفیت هوا خواهد داشت.

به‌طور کلی، این مطالعه نشان داد که تلفیق هوشمند داده‌های موجود پایش کیفیت هوا با الگوریتم‌های پیشرفته یادگیری ماشین و مهندسی ویژگی‌های زمانی می‌تواند خلأهای جدی شبکه پایش فعلی تهران (تعداد محدود ایستگاه‌ها و هزینه بالای اندازه‌گیری ذرات) را به‌طور مؤثر پر کند. مدل پیشنهادی نه‌تنها قابلیت عملیاتی شدن در زمان واقعی را دارد، بلکه پایه‌ای محکم برای توسعه نسل بعدی سیستم‌های هشدار زودهنگام آلودگی هوا در سطح محلات و بزرگراه‌های تهران فراهم می‌آورد. اجرای این مدل در مقیاس شهری می‌تواند گامی مهم در جهت کاهش مواجهه شهروندان با آلاینده‌های خطرناک و حرکت به‌سوی شهری پایدارتر و سالم‌تر باشد.

محدودیت‌ها و پیشنهادهای پژوهش‌های آتی

با وجود دقت بسیار بالا، مدل در پیک‌های استثنایی NO_x بالای 600 ppb اندکی کم‌برآوردی نشان می‌دهد که به‌دلیل نادر بودن این شرایط در داده‌های آموزشی قابل انتظار بود. افزودن داده‌های ترافیک لحظه‌ای (از دوربین‌های شمارش خودرو) یا داده‌های هواشناسی محلی (سرعت باد در سطح خیابان) می‌تواند این محدودیت را برطرف کند. همچنین، آزمایش مدل پیشنهادی در دیگر بزرگراه‌ها و محلات تهران (مانند همت، نیایش یا مناطق جنوبی) و تعمیم آن به $PM_{2.5}$ از جمله پیشنهادهای پژوهش‌های بعدی است.

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>
- Akbarzadeh, M. A., Khareshi, I., Sharifi, A., Yousefi, N., Naderian, M., Namazi, M. H., Safi, M., Vakili, H., Saadat, H., Parsa, S. A., & Nickdoost, N. (2018). The association between exposure to air pollutants including PM_{10} , $PM_{2.5}$, ozone, carbon monoxide, sulfur dioxide, and nitrogen dioxide concentration and the relative risk of developing STEMI: A case-crossover design. *Environmental Research*, 161, 299–303. <https://doi.org/10.1016/j.envres.2017.11.020>
- Alghushairy, O. (2025). Enhancing global logistics forecasting: Leveraging machine learning approaches for predicting import and export flows. In *2025 International Conference on Innovation in Artificial Intelligence and Internet of Things (AIIT)* (pp. 1–15). IEEE. <https://doi.org/10.34220/2308-8877-2025-13-3>
- Ameer, S., Shah, M. A., Khan, A., Song, H., Maple, C., Islam, S. U., & Asghar, M. N. (2019). Comparative analysis of machine learning techniques for predicting air quality in smart cities. *IEEE Access*, 7, 128325–128338. <https://doi.org/10.1109/ACCESS.2019.2925082>
- Arashi, M., Roozbeh, M., Hamzah, N. A., & Gasparini, M. (2021). Ridge regression and its applications in genetic studies. *PLOS ONE*, 16(4), Article e0245376. <https://doi.org/10.1371/journal.pone.0245376>

- Buchholz, S., Krein, A., Junk, J., Heinemann, G., & Hoffmann, L. (2013). Simulation of urban-scale air pollution patterns in Luxembourg: Contributing sources and emission scenarios. *Environmental Modeling & Assessment*, 18(3), 271–283. <https://doi.org/10.1007/s10666-012-9351-1>
- 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>