



## A Meta-Analysis and Systematic Review of Integrating Satellite-Derived Aerosol Optical Depth Data with Machine Learning for Estimating Fine Particulate Matter (PM<sub>2.5</sub>) Concentrations

Rasool Taherian Esfahani<sup>a</sup>, Mehdi Ghanbarzadeh Lak<sup>a\*</sup>

<sup>a</sup>Civil Engineering Department, School of Engineering, Urmia University, Urmia, Iran

Article Info	Abstract
<p><b>Article history</b>  <b>Received</b> :13 July 2024  <b>Received in revised form:</b>            4 December 2024  <b>Accepted:</b> 25 December 2024  <b>Available online:</b>            21 March 2025</p>	<p>Exposure to fine particulate matter (PM<sub>2.5</sub>) significantly impacts public health, particularly in regions where annual average levels of PM<sub>2.5</sub> exceed the World Health Organization (WHO) guidelines. According to the literature, in Iran, elevated fine particulate matter levels contribute substantially to mortality among adults. The spatial coverage limitations and intermittent data gaps of ground PM<sub>2.5</sub> monitoring stations pose challenges for effective air quality management.</p>
<p><b>Keywords:</b></p> <p>PM<sub>2.5</sub>, AEROSOL OPTICAL DEPTH (AOD), MACHINE LEARNING, MODIS, ENSEMBLE LEARNING, HYBRID MODELS, SYSTEMATIC REVIEW, PUBLIC HEALTH IMPACT</p>	<p>The products of remote sensing technologies, such as Aerosol Optical Depth (AOD) from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensors, offer a promising alternative for fine particulate matter estimation. This study reviews previous research on using machine learning algorithms to predict PM<sub>2.5</sub> ground concentrations based on AOD data. A structured analysis of 127 selected studies reveals varying correlations between AOD and PM<sub>2.5</sub> (with the resultant coefficient of determination, R<sup>2</sup>, between ground PM<sub>2.5</sub> concentrations and AOD data ranging from 48 to 99%), influenced by auxiliary variables like meteorological conditions and environmental factors. Integrating these variables enhances prediction accuracy, though it may increase complexity and potential errors in machine learning models. The hybrid machine learning models demonstrate superior performance compared to individual algorithms, leveraging their adaptability, parallel processing capabilities, and ability to handle missing data. Despite advancements, challenges persist due to data uncertainty and meteorological dynamics. In conclusion, while machine learning offers robust tools for PM<sub>2.5</sub> forecasting using AOD data, ongoing research is essential to address existing limitations and optimize model performance amidst environmental variability.</p>

\*.Corresponding author: Ghanbarzadeh Lak

E-mail address: m.ghanbarzadehlak@urmia.ac.ir

**How to cite this article:** Taherian Esfahani, R., & Ghanbarzadeh Lak, M. (2025). A Meta-Analysis and Systematic Review of Integrating Satellite-Derived Aerosol Optical Depth Data with Machine Learning for Estimating Fine Particulate Matter (PM<sub>2.5</sub>) Concentrations. *Journal of Geography and Environmental Hazards*, 14(1), pp.151-186, DOI:https://doi.org/10.22067/geoeh.2024.88696.1497



## Extended Abstract

### Introduction

Air pollution, an inevitable consequence of industrialization, climate change, and increased fossil fuel usage, has emerged as a critical environmental concern, especially in urban areas. Globally, air pollution is recognized as one of the leading environmental health risks, contributing significantly to premature mortality and morbidity, with Iran ranking particularly high in terms of annual deaths linked to air pollution. Recent data place air pollution as the eighth leading global risk factor for mortality and the seventh in Iran, underlining its severe public health implications. The health effects range from respiratory and cardiovascular diseases to heightened risks of cancer. Emerging evidence also highlights the psychological and cognitive effects of air pollution, linking it to anxiety, depression, reduced cognitive performance, and even increased criminal tendencies.

Among air pollutants, fine particulate matter (PM<sub>2.5</sub>) is especially concerning due to its ability to penetrate deep into the respiratory system and bloodstream, causing widespread systemic harm. PM<sub>2.5</sub> particles, with diameters of 2.5 microns or less, are associated with increased risks of heart attacks, strokes, and chronic lung diseases. In Iran, approximately 75,000 deaths annually are attributed to PM<sub>2.5</sub> exposure. The average population-weighted PM<sub>2.5</sub> concentration in the country stands at 48 µg/m<sup>3</sup>—substantially higher than the World Health Organization’s recommended maximum of 10 µg/m<sup>3</sup>. Such alarming figures underscore the critical need for effective air quality management strategies, including accurate and consistent monitoring of PM<sub>2.5</sub> concentrations.

Conventional ground-based air quality monitoring stations are valuable for tracking PM<sub>2.5</sub> levels but are limited by their sparse spatial coverage, especially in smaller cities and rural regions. To complement ground-based monitoring, remote sensing technologies have gained prominence as a means of providing broader spatial coverage. Aerosol Optical Depth (AOD), derived from satellite observations, is a key parameter for assessing atmospheric aerosols and estimating PM<sub>2.5</sub> concentrations. Sensors such as the Moderate Resolution Imaging Spectroradiometer (MODIS) have been instrumental in providing AOD data through retrieval algorithms like Dark Target (DT) and Deep Blue (DB). However, traditional approaches for converting AOD data into PM<sub>2.5</sub> concentrations often fall short in accuracy and adaptability, necessitating the exploration of advanced methodologies like machine learning.

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) method was employed in the present paper to evaluate relevant studies and summarize the findings of previous research. The main advantage of this method compared to others lies in its high accuracy in extracting information, reducing bias, and providing a comprehensive and well-documented perspective on previous research findings. This approach, particularly in studies based on quantitative data, yields stronger and more reliable results. The focus of the analysis was centered on three main axes:

1. Identifying factors that significantly influence the establishment of a statistically meaningful correlation between AOD data obtained from satellite imagery and PM<sub>2.5</sub> data derived from ground-based monitoring;
2. Assessing the success rate of previous research in establishing such correlations using statistical indicators; and

3. Examining the machine learning algorithms employed in these studies.

## **Material and Methods**

This study systematically reviews machine learning algorithms used for estimating PM<sub>2.5</sub> concentrations based on AOD data, focusing on their performance, scalability, and adaptability across diverse environmental settings. The primary objectives are to identify the strengths and limitations of existing methodologies, highlight gaps in current research, and propose avenues for improvement.

To ensure a comprehensive analysis, the study employed a systematic review and meta-analytical approach based on the PRISMA guidelines. Reputable scientific databases, including PubMed, Google Scholar, and Science Direct, were queried using keywords (Fig. 1) such as "PM<sub>2.5</sub> estimation," "Aerosol Optical Depth," and "machine learning." An initial pool of 977 documents was narrowed down to 127 highly relevant articles through a rigorous screening process (Fig. 2). Key extraction parameters included the correlation between AOD and PM<sub>2.5</sub>, machine learning models employed, and performance metrics such as R<sup>2</sup> values, root mean square error (RMSE), and mean absolute error (MAE).

## **Results and Discussion**

Correlation between AOD and PM<sub>2.5</sub>

Numerous studies demonstrate a strong correlation between AOD values and ground-level PM<sub>2.5</sub> concentrations, although the strength of this relationship varies based on geographical, meteorological, and land-use factors (Fig. 3 and Table 1). Meteorological conditions such as temperature, relative humidity, and wind speed significantly influence the correlation by affecting aerosol properties and atmospheric dispersion. Additionally, land-use characteristics—such as urban density, vegetation cover, and proximity to industrial zones—modulate local PM<sub>2.5</sub> levels.

Incorporating these auxiliary variables (see Fig. 4) into predictive models has proven effective in enhancing the accuracy of PM<sub>2.5</sub> estimations. For example, regression models integrating meteorological and traffic data achieved significantly improved performance compared to models relying solely on AOD data. However, the inclusion of too many auxiliary variables can lead to overfitting, reducing the model's generalizability. Balancing model complexity with predictive accuracy remains a key challenge in this domain.

## **Machine Learning Algorithms for PM<sub>2.5</sub> Estimation**

Machine learning (ML) methods have revolutionized the estimation of PM<sub>2.5</sub> concentrations from AOD data by effectively capturing the complex, non-linear relationships between variables. Among these, ensemble learning models such as Random Forest (RF) and XGBoost have consistently outperformed traditional linear regression models due to their ability to handle high-dimensional data and mitigate biases. Deep learning techniques, including Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, have also shown promise, particularly in time-series predictions of PM<sub>2.5</sub> levels. Studies comparing ML algorithms highlight the superior performance of ensemble models, especially when combined with feature

selection techniques to reduce input data redundancy. For instance, hybrid approaches that integrate RF and XGBoost have achieved  $R^2$  values exceeding 0.9 in several case studies, indicating exceptional predictive power. Furthermore, the use of advanced optimization techniques, such as Bayesian optimization, has enhanced the performance of these models by fine-tuning hyperparameters.

### **Limitations and Challenges**

Despite the advancements in ML-based  $PM_{2.5}$  estimation, several challenges persist. The accuracy of these models is highly dependent on the quality of input data, which can be compromised by factors such as cloud cover, sensor limitations, and temporal mismatches between satellite observations and ground measurements. Additionally, the variability in meteorological conditions introduces uncertainties that are difficult to account for, particularly in regions with complex topography. Another limitation is the reliance on high-resolution satellite imagery, which is often expensive and not readily available for all regions. Addressing these challenges requires integrating data from multiple sources, including ground-based sensors, satellite datasets, and meteorological models. The development of robust data fusion techniques is essential for improving the scalability and reliability of  $PM_{2.5}$  estimation models.

### **Potential Applications and Future Directions**

Future research should focus on developing hybrid algorithms that combine the strengths of multiple ML techniques, such as deep learning and ensemble learning. These algorithms should be capable of handling diverse datasets, enhancing temporal and spatial resolution, and addressing data uncertainty. Additionally, integrating long-term climate change scenarios into these models could provide more comprehensive insights into the dynamics of air pollution.

### **Conclusion**

This study underscores the utility of AOD as a proxy for estimating  $PM_{2.5}$  concentrations and highlights the transformative potential of machine learning in enhancing the accuracy and scalability of air quality monitoring (please refer to Figs. 5 and 6). Ensemble learning models, particularly hybrid approaches, offer significant advantages in capturing the complex interactions between AOD and  $PM_{2.5}$ . However, addressing challenges related to data quality, meteorological variability, and scalability is crucial for realizing the full potential of these methods.



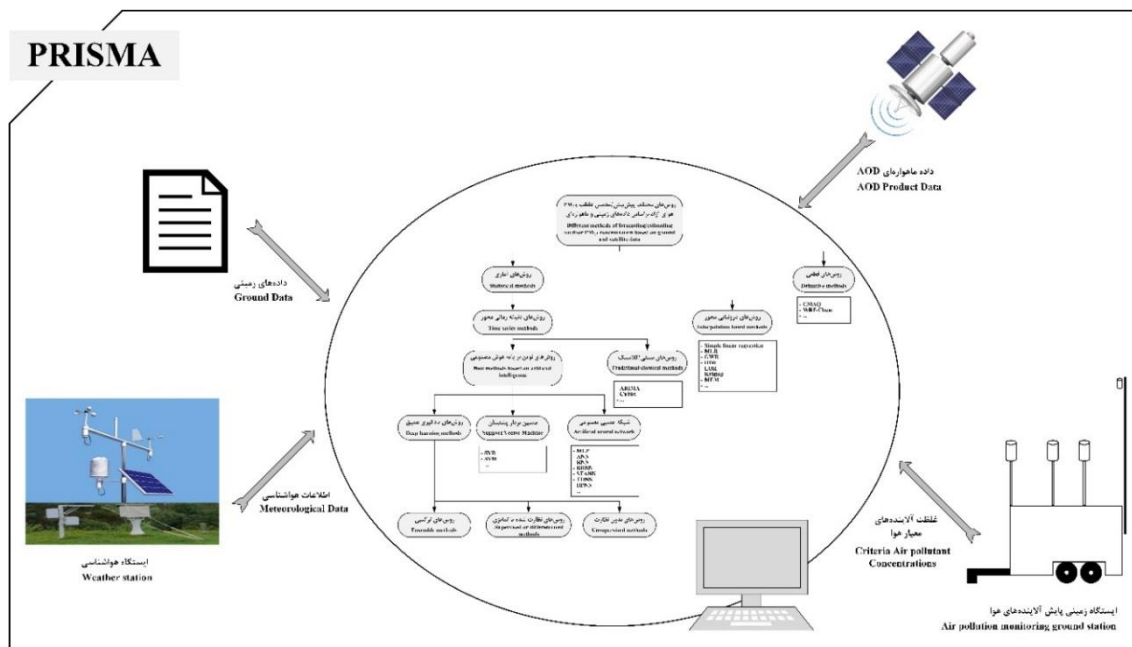
## فراکاوای و مرور سیستماتیک ادغام داده‌های عمق نوری آئروسول برگرفته از ماهواره با یادگیری ماشین در برآورد غلظت ذرات معلق ریز ( $PM_{2.5}$ )

رسول طاهریان اصفهانی<sup>۱</sup>، مهدی قنبرزاده لک<sup>\*۱</sup>

<sup>۱</sup>گروه مهندسی عمران، دانشکده فنی و مهندسی، دانشگاه ارومیه، ارومیه، ایران

اطلاعات مقاله	چکیده
تاریخچه مقاله	
تاریخ دریافت: ۱۴۰۳/۴/۲۳	
تاریخ بازنگری: ۱۴۰۳/۹/۱۱	
تاریخ پذیرش: ۱۴۰۳/۱۰/۵	
<b>کلمات کلیدی:</b>	
ذرات $PM_{2.5}$	
عمق نوری آئروسول (AOD)	
یادگیری ماشین	
سنجنده MODIS	
یادگیری گروهی	
مدل‌های ترکیبی	
مرور سیستماتیک	
تأثیر بر سلامت عمومی	
	قرار گرفتن در معرض ذرات ریز ( $PM_{2.5}$ ) آلاینده هوا، به ویژه در مناطقی که میانگین سالانه آن از دستورالعمل‌های سازمان بهداشت جهانی فراتر است، به طور قابل توجهی بر سلامت عمومی تأثیر می‌گذارد. در ایران، افزایش سطح مواجهه با $PM_{2.5}$ ، به میزان قابل توجهی بر مرگ و میر درمیان بزرگسالان نقش داشته است. محدودیت‌های پوشش مکانی و شکاف مقطعی داده‌های ایستگاه‌های پایش زمینی، چالش‌هایی را در مدیریت مؤثر کیفیت هوا ایجاد نموده است. محصولات فناوری‌های سنجش از دور، مانند عمق‌نوری آئروسول (AOD) برگرفته از حسگرهای MODIS، جایگزین امیدوارکننده‌ای در برآورد $PM_{2.5}$ به شمار می‌روند. در مقاله حاضر، تحقیقات پیشین در زمینه استفاده از الگوریتم‌های یادگیری ماشین در برآورد غلظت $PM_{2.5}$ بر اساس داده‌های AOD، مرور شده است. تجزیه و تحلیل ساختاری ۱۲۷ مطالعه انتخاب شده، همبستگی‌های متفاوتی را بین AOD و $PM_{2.5}$ نشان داد ( $R^2$ در محدوده ۴۸ تا ۹۹٪)، که می‌تواند با بهره‌گیری از متغیرهای کمکی نظیر شرایط هواشناسی و عوامل محیطی، بهبود یابد. اگرچه ادغام این متغیرها، دقت برآورد را افزایش می‌دهد، ولی در عین حال، پیچیدگی و خطاهای بالقوه را در مدل‌های یادگیری ماشین افزایش خواهد داد. مدل‌های یادگیری ماشین ترکیبی عملکرد بهتری را در مقایسه با الگوریتم‌های فردی نشان می‌دهند، چراکه قابلیت انطباق، پردازش موازی و مدیریت داده‌های از دست رفته را دارند. علی‌رغم پیشرفت‌های اخیر، هنوز چالش‌هایی به دلیل عدم قطعیت داده‌ها و دینامیک بودن پدیده‌های هواشناسی، باقی مانده‌اند. الگوریتم‌های یادگیری ماشین، اگرچه ابزاری قوی در برآورد $PM_{2.5}$ بر اساس شاخص AOD ارائه می‌دهد، لیکن تحقیقات آتی در راستای رفع محدودیت‌ها و بهینه‌سازی عملکرد مدل متناسب با تغییرات محیطی، ضروری است.

چکیده تصویری



مقدمه

امروزه بحث آلودگی هوا به عنوان پیامدی از توسعه فرآیندهای صنعتی، تغییرات اقلیمی و افزایش استفاده از سوخت‌های فسیلی، به یکی از معضلات جدی زیست‌محیطی خصوصاً در شهرهای بزرگ تبدیل شده است (Abbas, Ekowati, Suhariadi & Fenitra, 2023; Bai, Wang, Ma & Lu, 2018; Bono et al., 2015; Izah, Iyiola, Yarkwan & Richard, 2023; Jia, Li, Chen & Yang, 2023; West et al., 2016; Lu, Li, Li & Al-Barakani, 2021). بگونه‌ای که در سطح جهانی، به عنوان هشتمین عامل خطر و در کشور ایران به عنوان هفتمین عامل خطر با بالاترین تعداد مرگ و میر سالانه، شناخته می‌شود (Institute for Environmental Research, 2019). طرح‌ریزی و اجرای اقدامات حفاظتی کیفیت هوای آزاد، با پیش‌گیری از به مخاطره افتادن سلامتی افراد در معرض (به عنوان نیروی محرکه توسعه صنعتی) و در نتیجه کاهش غیبت افراد در محیط کاری و جلوگیری از کاهش بهره‌وری فردی کارکنان، منجر به رشد و شکوفایی اقتصادی جوامع خواهد شد. از دیدگاه روان‌شناختی، آلودگی هوا یکی از دلایل اصلی ایجاد اضطراب در رابطه با سلامتی و آینده، کاهش شادکامی، افزایش افسردگی و بروز اختلالات روانی همچون اسکیزوفرنی و اوتیسم می‌باشد. در عین حال، به لحاظ شناختی، آلودگی هوا می‌تواند عملکرد تصمیم‌گیری افراد را تحت تأثیر قرار داده و از نقطه نظر اجتماعی، منجر به تشدید رفتارهای مجرمانه و غیر اخلاقی گردد (Lu, 2020). به بیان دیگر، آلودگی هوا نه تنها سلامت افراد، بلکه سلامت جامعه را نیز به مخاطره می‌اندازد.

دسته‌بندی آلاینده‌های هوا (قابل شناسایی در هوای آزاد یا در گاز خروجی از منابع ساکن و متحرک) براساس دیدگاه‌های مختلفی همچون ترکیب شیمیایی گونه‌ها، ویژگی‌های واکنش‌پذیری، خصوصیات پراکنش در هوا، پایداری در محیط، توانایی انتقال در مسافت‌های کوتاه یا طولانی و اثرات تدریجی بر سلامتی انسان و/یا حیوانات، انجام می‌شود. با این وجود و با در نظر گرفتن برخی خصوصیات مشابه، آلاینده‌های هوا را می‌توان در چهار طبقه کلی (۱) آلاینده‌های گازی (مانند دی‌اکسید گوگرد، منوکسید کربن، اکسیدهای نیتروژن، اُزن و ترکیبات آلی فرار)؛ (۲) آلاینده‌های آلی دیرپا<sup>۱</sup> (همچون دی‌اکسین‌ها و فیورن‌ها)<sup>۲</sup>؛ (۳) فلزات سنگین (نظیر سرب و جیوه)؛ و (۴) ذرات جامد معلق (که در ادبیات فنی در دو دسته کلی  $PM_{2.5}$  و  $PM_{10}$  طبقه‌بندی می‌شوند)؛ دسته‌بندی نمود (Kampa & Castanas, 2008). در این بین، آلاینده‌های معیار کیفیت هوای آزاد، ترکیباتی را شامل می‌شوند که اولاً حضور آنها در هوای آزاد شهری متداول بوده و ثانیاً سطوح قابل قبولی از قرار گرفتن در معرض آنها توسط استانداردهای کیفیت هوای آزاد، تعیین شده است. براساس استانداردهای سازمان حفاظت از محیط‌زیست ایالات متحده آمریکا، این آلاینده‌ها شامل اُزن، منوکسید کربن، سرب، دی‌اکسید گوگرد، دی‌اکسید نیتروژن و ذرات معلق، می‌باشند (U.S. EPA, 2024). ذرات معلق با قطر آئرودینامیکی مساوی یا کمتر از  $2.5 \mu m$  ( $PM_{2.5}$ )، از آلاینده‌های مهم کیفیت هوای آزاد محسوب می‌گردند و قادرند از طریق دستگاه تنفسی انسان، وارد جریان خون شوند. اثرات استنشاق  $PM_{2.5}$  از ساده‌ترین حالات آلرژی تا بیماری‌های قلبی-عروقی و همچنین افزایش خطر مبتلا شدن به بیماری‌های سرطانی، متغیر است (Bell, Dominici, Ebisu, Zeger & Samet, 2007; Chen et al., 2021; Chudnovsky et al., 2013; Feng, Li, Wang & Du, 2020; Li, Yu, Huang, Sun & Jia, 2023; Moryani, Kong, Du & Bao, 2020; Nabavi, Haimberger & Abbasi, 2019; Osimobi, Yorkor & Nwankwo, 2019; Perrone et al., 2013; Pope III & Dockery, 2006; Saeed, Hussain, Awan & Idris, 2017; Xiao et al., 2017; Xue et al., 2022; Yan, Zang, Luo, Jiang & Li, 2020; Kelly & Fussell, 2012; Institute for Environmental Research, 2019; He & Huang, 2018).

مطالعات اخیر نشان دادند که در ایران، سالانه حدود ۷۵ هزار مرگ و میر (بین ۶۲ تا ۸۶ هزار) به دلیل قرار گرفتن افراد در معرض ذرات معلق محیطی  $PM_{2.5}$  است (یعنی چیزی حدود ۲۴ درصد از کل مرگ و میر سالانه افراد بالای ۲۵ سال در ایران). بر اساس میانگین وزنی-جمعیتی سالانه، در ایران غلظت آلاینده  $PM_{2.5}$  در هوای آزاد (که افراد در معرض آن می‌باشند)، حدود  $48 \mu g m^{-3}$  برآورد می‌شود که نسبت به میانگین جهانی آن (یعنی  $51 \mu g m^{-3}$ ) کمی پایین است؛ اما با این وجود، همچنان قریب به  $4/8$  برابر بیشتر از سطح توصیه شده توسط سازمان بهداشت جهانی (WHO) یعنی  $10 \mu g m^{-3}$  و تقریباً چهار برابر بیشتر از استاندارد ملی است (Institute for Environmental Research, 2019). بنابراین ضروری است که میزان این ذرات به طور مستمر اندازه‌گیری و در خصوص کنترل غلظت آن، مراقبت شود (Xu, Chen, Yang, Tang & Dong, 2021).

1 Persistent organic pollutants

2 Dioxins and Furans

3 Particulate Matter less than 2.5 micrometers in aerodynamic diameter ( $PM_{2.5}$ )4 Particulate Matter less than 10 micrometers in aerodynamic diameter ( $PM_{10}$ )

5 World Health Organization (WHO)

اندازه‌گیری ساعتی و روزانه غلظت ذرات معلق  $PM_{2.5}$  در هوای آزاد شهری توسط ایستگاه‌های زمینی پایش کیفیت هوا و پیش‌بینی شرایط آتی آن، از اهمیت قابل توجهی برخوردار است. با این حال، پوشش مکانی ایستگاه‌های پایش زمینی محدود بوده و در مواقعی از سال، ممکن است بدلیل نیاز به واسنجی تجهیزات یا مسائل نگهداری، از مدار خارج شوند. علاوه بر آن، تعداد اندک ایستگاه‌های زمینی پایش کیفیت هوای نصب شده در شهرهای کوچک، امکان پاسخگویی مناسب قبل از وقوع رخداد آلودگی هوا را تا حد زیادی محدود نموده است. در سالیان اخیر تلاش‌هایی در راستای بهره‌گیری از تکنولوژی‌های سنجش از دور به عنوان ابزاری جایگزین، صورت گرفته و بدین ترتیب می‌توان پوشش گسترده‌تر فضایی و زمانی از غلظت آلاینده‌های هوا جهت اقدامات پیشگیرانه در دسترس داشت (Chudnovsky et al., 2013; Hu, Liebens & Rao, 2011).

یکی از پارامترهای اصلی برآورد غلظت آلاینده‌های  $PM$  در هوای آزاد براساس تکنولوژی‌های سنجش از دور، عمق نوری آئروسول (AOD)<sup>۱</sup> است که از طریق تصاویر ماهواره‌ای بدست می‌آید و در مطالعات پیشین به طور گسترده‌ای از آن بهره‌برداری شده است (Li, Ma, Xu & Song, 2018b; Meng et al., 2016). شاخص AOD معیار مناسبی در برآورد میزان ذرات معلق موجود در جو است که بر اساس نسبت امواج خورشیدی وارد شده و جذب شده توسط اتمسفر (به علت تعامل با ذرات) و طول مسیر طی شده، محاسبه می‌گردد. AOD بطور معمول در یک ستون عمودی از هوا تعیین شده و داده‌های تولیدی توسط سنجنده MODIS<sup>۲</sup> به طور گسترده‌ای، بدین منظور مورد استفاده قرار گرفته است (Mao, Qiu, Taheri Shahraini & Sodoudi, 2016). الگوریتم اهداف تیره (DT)<sup>۳</sup> و الگوریتم آبی عمیق (DB)<sup>۴</sup>، دو الگوریتم متداول در بازیابی ویژگی‌های آئروسول از محصول تصحیح اتمسفری سنجنده MODIS به شمار می‌روند (Nabavi et al., 2019).

الگوریتم DT توسط کافمن و همکاران (Kaufman et al., 1997) معرفی شد. DT از این مفهوم فیزیکی الهام گرفته است که ذرات معلق در هوا از سطوح تاریک‌تر زمین مانند جنگل‌ها و سطح اقیانوس‌ها، روشن‌تر هستند. به دلیل محدودیت این الگوریتم در شناسایی ذرات معلق خصوصاً در مناطق شهری و بیابانی که دارای سطوح روشن می‌باشند، بازیابی دقیق ذرات معلق، یک چالش بزرگ بوده است. همین موضوع سبب شد تا هسو و همکاران (Hsu, Tsay, King & Herman, 2004)، الگوریتم دیگری به نام DB را بر روی تصاویر ماهواره‌ای MODIS با تفکیک‌پذیری ۱۰ کیلومتری پیاده‌سازی کنند. در الگوریتم اخیر از باند طول موجی ۴۱۲ نانومتر استفاده شد که آبی عمیق نام دارد و در آن آئروسول‌ها از هر نوع سطح زمینی، روشن‌تر هستند (Nabavi et al., 2019).

هدف اصلی در مقاله مروری حاضر، بررسی الگوریتم‌های مختلف یادگیری ماشین بکار رفته در برآورد غلظت ذرات معلق ریز با استفاده از داده‌های AOD و میزان تطابق آنها با نتایج پایش‌های زمینی است. با توجه به نوپا

1 Aerosol Optical Depth (AOD)

2 Moderate Resolution Imaging Spectroradiometer (MODIS)

3 Dark Target (DT)

4 Deep Blue (DB)



بودن کاربرد الگوریتم‌های یادگیری ماشین در این شاخه از علوم و با عنایت به این موضوع که برخی روش‌های سنتی نیز در برآورد غلظت PM<sub>2.5</sub> از داده‌های AOD تاکنون مورد استفاده قرار گرفته اند، دامنه مطالعه مروری حاضر قدری گسترده‌تر انتخاب شد و مقالات و مستندات<sup>۱</sup> که به بررسی همبستگی ما بین غلظت PM<sub>2.5</sub> زمینی و داده‌های AOD با یا بدون بهره‌گیری از یادگیری ماشین پرداخته بودند، مورد توجه قرار گرفتند.

لازم بذکر است در مطالعات مروری، روش‌های مختلفی جهت گردآوری و تحلیل اطلاعات وجود دارد. مرور سنتی<sup>۱</sup> به مرور کلی مطالعات می‌پردازد و معمولاً فاقد ساختار دقیق و روش‌شناسی مشخص است؛ این روش بیشتر برای مطالعات مقدماتی مناسب بوده و ممکن است شامل سوگیری شود. مرور سیستماتیک<sup>۲</sup> با استفاده از رویکردی ساختاریافته و منظم، به شناسایی و ارزیابی مطالعات مرتبط با یک سؤال پژوهشی خاص می‌پردازد و برای ارائه نتایج دقیق‌تر و جامع، طراحی شده است. مرور دامنه‌ای<sup>۳</sup> نیز برای شناسایی شکاف‌های تحقیقاتی و ارزیابی گستردگی موضوعی پژوهش‌ها به کار می‌رود، اما برخلاف مرور سیستماتیک، بر کیفیت مطالعات تأکید کمتری دارد. علاوه بر این، متا-آنالیز<sup>۴</sup> یا فراکاووی یک روش کمی است که داده‌های آماری مطالعات مختلف را ترکیب کرده و نتایج جامع و قابل اعتمادی ارائه می‌دهد.

بدین منظور و در مطالعه مروری حاضر، از روش مرور ساختاریافته و فراکاووی (PRISMA)<sup>۵</sup> بهره‌گیری گردید. این رویکرد به دلیل ساختار دقیق و منظم خود، امکان شناسایی و ارزیابی مطالعات مرتبط را فراهم می‌کند و با بهره‌گیری از تکنیک‌های آماری، نتایج مطالعات پیشین را تحلیل و خلاصه‌سازی می‌نماید. مزیت اصلی این روش نسبت به روش‌های دیگر، دقت بالا در استخراج اطلاعات، کاهش سوگیری و ارائه یک دیدگاه جامع و مستند از یافته‌های پژوهش‌های پیشین است. این ترکیب، به ویژه در پژوهش‌های مبتنی بر داده‌های کمی، نتایجی قوی‌تر و قابل اعتمادتر فراهم می‌آورد. کانون توجه در استخراج نتایج از مقالات و مستندات بر سه محور (۱) تعیین عوامل مؤثر بر ایجاد همبستگی معنی‌دار از نقطه نظر آماری بین داده‌های AOD، که از طریق تصاویر ماهواره‌ای بدست می‌آید و داده‌های PM<sub>2.5</sub> حاصل از پایش زمینی؛ (۲) میزان موفقیت تحقیقات پیشین در ایجاد همبستگی فوق براساس شاخص‌های آماری؛ و نهایتاً (۳) الگوریتم‌های یادگیری ماشین بکار رفته؛ متمرکز بوده است.

## روش تحقیق

مقالات مبتنی بر روش مرور ساختاریافته و فراکاووی، از این جهت که متخصصان با بهره‌گیری از آنها قادر خواهند بود تا در زمینه تخصصی خود به‌روز باشند، دارای اهمیت قابل توجهی است. اساس کار در این روش، بررسی سؤال یا مسأله‌ای مشخص و معین بوده و بدین منظور از الگوریتم‌های ساختار یافته و صریح در شناسایی، انتخاب و ارزیابی منصفانه مطالعات مرتبط استفاده می‌نماید. در نهایت، با گردآوری و تجزیه و تحلیل نتایج

1 Traditional Review

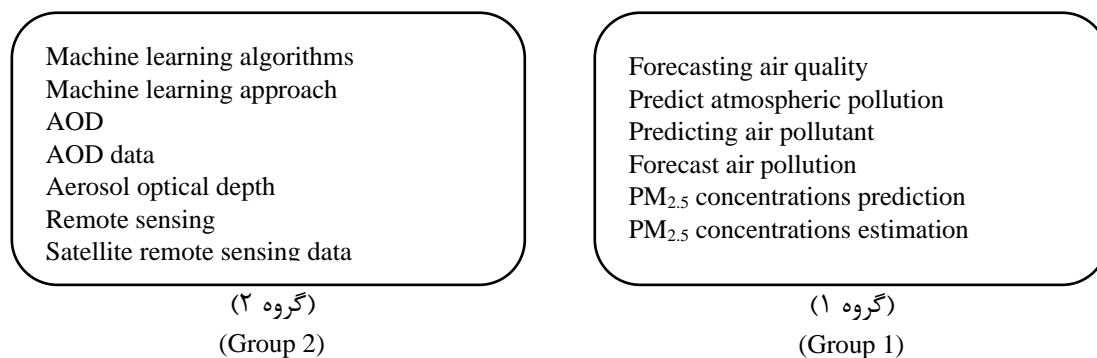
2 Systematic Review

3 Scoping Review

4 Meta-Analysis

5 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)

مطالعات پیشین طی جداول و اشکال (فراکاو)، خلاصه‌ای از پیشرفت‌های پیشین حاصل شده در شاخه مورد تحقیق، به خوانندگان ارائه خواهد شد. به بیان دیگر، فراکاو با استفاده تکنیک‌های آماری به تجزیه و تحلیل و خلاصه‌سازی نتایج مطالعات پیشین می‌پردازد (Moher, Liberati, Tetzlaff, Altman & PRISMA Group, 2009). در این مطالعه، از رویکرد PRISMA برای شناسایی، غربالگری و انتخاب مقالات مرتبط استفاده شده است. برای جستجوی مطالعات پیشین، مستندات منتشر شده در پایگاه‌های اطلاعاتی PubMed، Google Scholar و Science Direct مورد بررسی قرار گرفت. جستجوی مدارک با ترکیب دو گروه از کلمات کلیدی مرتبط با موضوع، پی‌گیری گردید. بگونه‌ای که گروه اول شامل کلید واژه‌های توصیف کننده برآورد آلودگی هوا بوده و گروه دوم، کلید واژه‌های مربوط به یادگیری ماشین و سنجش از دور را شامل می‌شدند. جستجوی مدارک با استفاده از کلید واژه‌های انگلیسی و با ترکیب احتمالی کلمات مهم، اصلی و حساس، انجام شد. بدین منظور، کلید واژه‌های منتخب به همراه عملگرهای AND و OR، در بخش‌های عنوان، کلمات کلیدی و چکیده مدارک، جستجو شدند. واژگان کلیدی که در جستجوی مقالات مورد استفاده قرار گرفتند، در شکل (۱) آورده شده است.

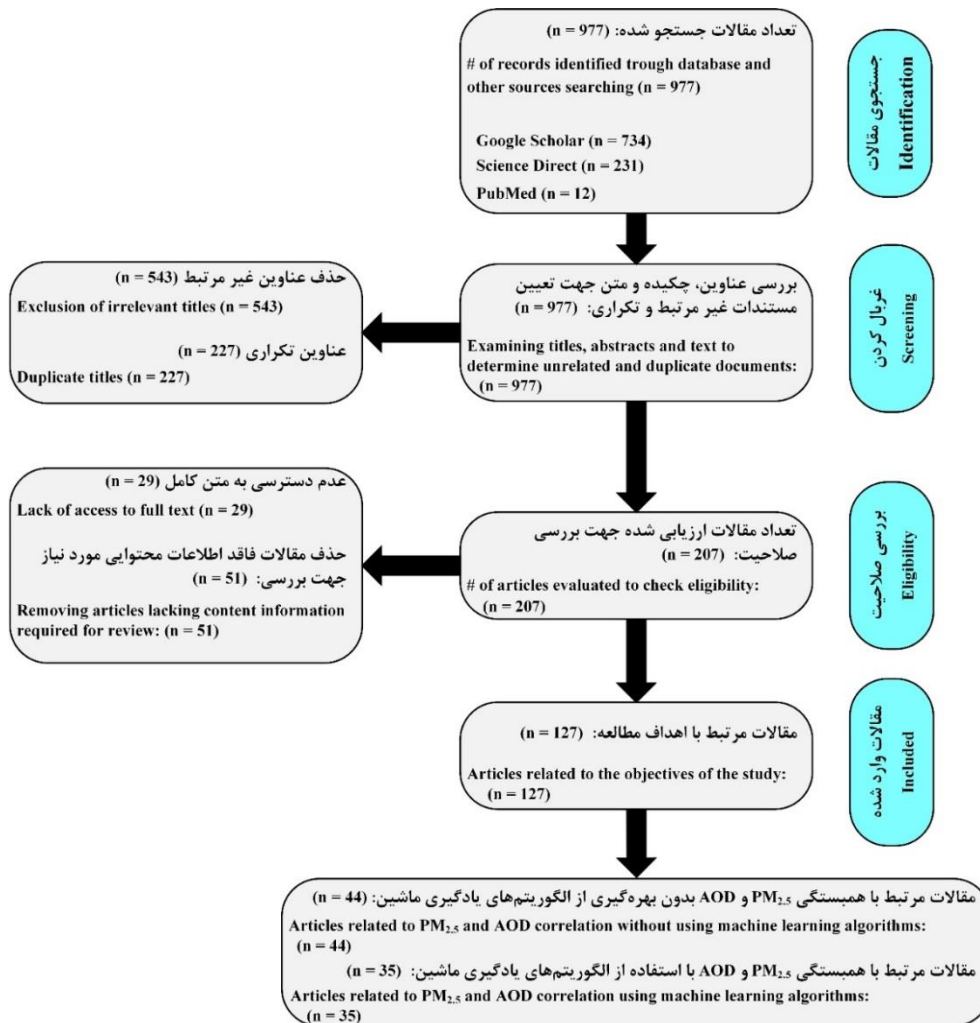


شکل ۱- واژگان کلیدی مورد استفاده در جستجوی مقالات

**Fig.1. Keywords Used in Article Searches**

در مرحله اول عملیات جستجو، مقالاتی انتخاب شدند که در آنها وجود همبستگی و شدت آن بین داده‌های AOD ماهواره‌ای و مقادیر پایش شده زمینی PM<sub>2.5</sub>، با/بدون استفاده از الگوریتم‌های یادگیری ماشین، مورد مطالعه قرار گرفته بود. در بررسی اولیه سه پایگاه داده فوق الذکر، در کل ۹۷۷ مقاله حائز شرایط این تحقیق تا تاریخ ۱۹ آوریل ۲۰۲۴ (۳۱ فروردین ۱۴۰۳)، به دست آمد. سپس عنوان و چکیده تمامی مقالات، مستندات و گزارش‌های منتخب در مرحله اول، استخراج و با بررسی محتوای آن‌ها، موارد تکراری حذف شدند. بر اساس بررسی عناوین و چکیده‌ها، تعداد زیادی از منابع حذف گردیدند. محتوای مقالات باقی‌مانده به طور دقیق مورد مطالعه قرار گرفت و منابع کاملاً مرتبط، انتخاب شدند. در نهایت بر اساس عنوان، نویسنده اصلی، سال انتشار، نوع الگوریتم انتخابی و مشخصات دیگر، مستندات به صورت خلاصه در جدولی طبقه‌بندی گردیدند. بدین ترتیب

فرایند جمع‌آوری و بررسی منابع پژوهشی به طور سیستماتیک انجام پذیرفت تا اطلاعات مفید برای پاسخ به سؤال پژوهش به دست آید. فرایند جستجو، ورود و خروج مقالات، در شکل (۲) آورده شده است.



شکل ۲- فرایند جستجو، ورود و خروج مقالات

Fig.2. The Process of Searching, Entering and Exiting Articles

## نتایج و بحث

توسعه مدل‌های پیش‌بینی/ برآورد غلظت آلاینده‌های هوا در یک مکان خاص، ابزاری قدرتمند در طرح‌ریزی اقدامات پیشگیرانه همچون کنترل منابع پراکنش و هشداردهی به شهروندان در شرایط اضطراری را در اختیار مدیران شهری قرار می‌دهد (Cortina-Januchs, Quintanilla-Dominguez, Vega-Corona & Andina, 2015). به طور کلی، راهکارهای پیش‌بینی/ برآورد غلظت یک آلاینده در یک محیط براساس داده‌های پایش

پیشین همان آلاینده یا سایر آلاینده‌های معیار و اطلاعات محیطی، در دو دسته راهکارهای قطعی و آماری، تقسیم‌بندی می‌شوند (Bai et al., 2018).

مدل‌های کیفیت هوای چند مقیاسی جامعه (CMAQ)<sup>۱</sup> و پیش‌بینی عددی وضع هوای جفت شده با مباحث شیمی (WRF-Chem)<sup>۲</sup>، از جمله پُرکاربردترین مدل‌های قطعی پیش‌بینی غلظت آلاینده‌های هوا به شمار می‌روند (Kong, Xin, Zhang & Wang, 2016; Li et al., 2018b). این مدل‌ها با در نظر گرفتن عملکردهای فیزیکی شامل گسیل، پراکنش، انتقال و پخش آلاینده؛ درهم‌کنش‌های شیمیایی به وقوع پیوسته در جو را نیز شبیه‌سازی می‌نمایند (Van Donkelaar et al., 2010; Geng et al., 2015). با این وجود، از آنجا که بیشتر این مدل‌ها به دانش پیچیده‌ای درباره حالت پخش آلاینده و مسیرهای واکنش شیمیایی نیاز دارند، برآورد غلظت آلاینده PM<sub>2.5</sub> از طریق آنها بسیار پیچیده و از لحاظ محاسباتی پُرهزینه خواهد بود (Stern et al., 2008).

در راستای رفع محدودیت‌های راهکارهای قطعی، از مدل‌های آماری در پیش‌بینی/ برآورد آلودگی هوا بهره‌گیری شده است. این مدل‌ها خود در دو طبقه مدل‌های فضایی (مکانی) و مدل‌های دنباله زمانی، دسته‌بندی می‌شوند. در مدل‌های فضایی (مکانی) راهکارهایی که بر مبنای درونیابی عمل می‌کنند، تحت عنوان راهکارهای زمین آماری در مدل‌سازی آلودگی هوا شناخته شده و نمونه‌هایی از آنها شامل روش‌های رگرسیون خطی ساده (Wang & Christopher, 2003; Koelemeijer, Homan & Matthijsen, 2006)، رگرسیون خطی چندگانه (MLR)<sup>۳</sup> (Liu, Sarnat, Kilaru, Jacob & Koutrakis, 2005)؛ Paciorek, Liu, Moreno-Macias & Fotheringham, Charlton & Brunson, 2008)، رگرسیون وزن دار جغرافیایی (GWR)<sup>۴</sup> (Kondragunta, 2008)، درونیابی وزنی وارونه (IDW)<sup>۵</sup> (Masroor et al., 2020)، رگرسیون کاربری اراضی (LUR)<sup>۶</sup> (Vicedo-Cabrera, Biggeri, Grisotto, Barbone & Catelan, 2013)، روش کریجینگ (Lee et al., 2015) و مدل اثر ترکیبی (MEM)<sup>۷</sup> (Lee, Coull, Bell, Schwartz & Koutrakis, 2011) می‌باشند که در تحقیقات گذشته در برآورد مکانی آلودگی در یک زمان مشخص در سطح شهر، به کار رفته‌اند.

بخش دوم یعنی راهکارهای دنباله زمانی محور با هدف پیش‌بینی/ برآورد آلودگی هوا، شامل روش‌های سنتی و روش‌های نوین می‌باشند. متداول‌ترین روش سنتی، مدل رگرسیونی میانگین متحرک یکپارچه خود هم‌بسته (ARIMA)<sup>۸</sup> است که در برآورد غلظت PM<sub>2.5</sub> مورد استفاده قرار گرفته است (Hosseini & Shahbazi, 2016). از جمله محدودیت‌های روش‌های سنتی می‌توان به عدم عملکرد مؤثر این مدل‌ها در تجزیه و تحلیل و ادغام داده‌های چند منبعی مانند داده‌های هواشناسی، کاربری زمین و ترافیک، اشاره داشت که منجر به دقت اندک

1 the Community Multiscale Air Quality (CMAQ)

2 Weather Research and Forecasting model coupled to Chemistry (WRF-Chem)

3 Multi Linear Regression (MLR)

4 Geographically Weighted Regression (GWR)

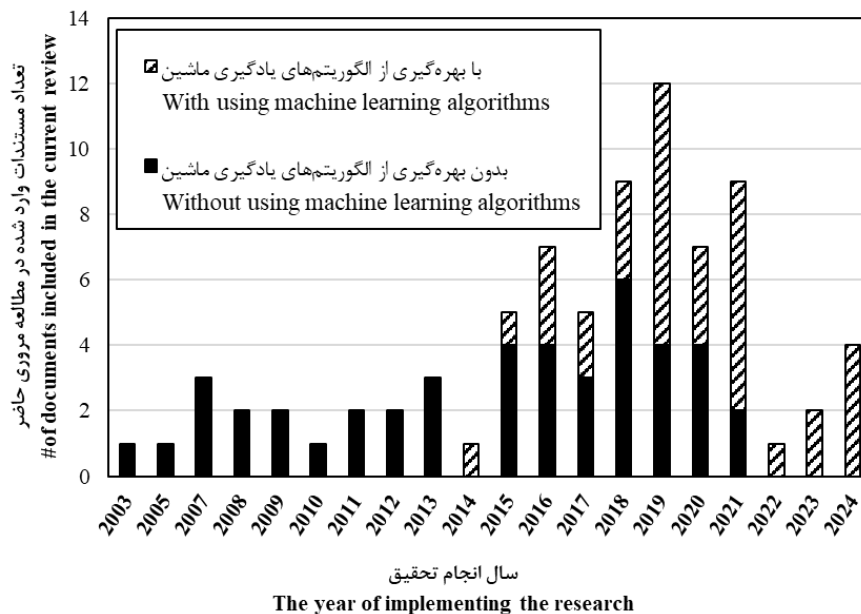
5 Inverse Distance Weighted (IDW)

6 Land Use Regression (LUR)

7 Mixed Effect Model (MEM)

8 AutoRegressive Integrated Moving Average (ARIMA)

آنها در برآوردها شده است. علاوه بر آن، روش‌های سنتی مدل‌سازی مبتنی بر دنباله زمانی نیازمند پیش‌فرض‌هایی همچون انتخاب یک توزیع خاص آماری و پذیرش خطی بودن مدل هستند که در صورت عدم برقراری این پیش‌فرض‌ها، نتایج مدل با خطای قابل توجهی همراه خواهد بود (Song, Qin, Qu & Liu, 2015). روش‌های نوین مبتنی بر دنباله زمانی در ادامه مورد بررسی بیشتر قرار خواهند گرفت. در شکل (۳) فراوانی مقالات وارد شده در مطالعه مروری حاضر براساس سال انتشار آنها آورده شده است. همانطور که از این شکل برمی‌آید، روند انتشار تحقیقات مرتبط با برآورد غلظت‌های زمینی PM<sub>2.5</sub> براساس داده‌های ماهواره‌ای، در سالیان اخیر یک روند صعودی است.

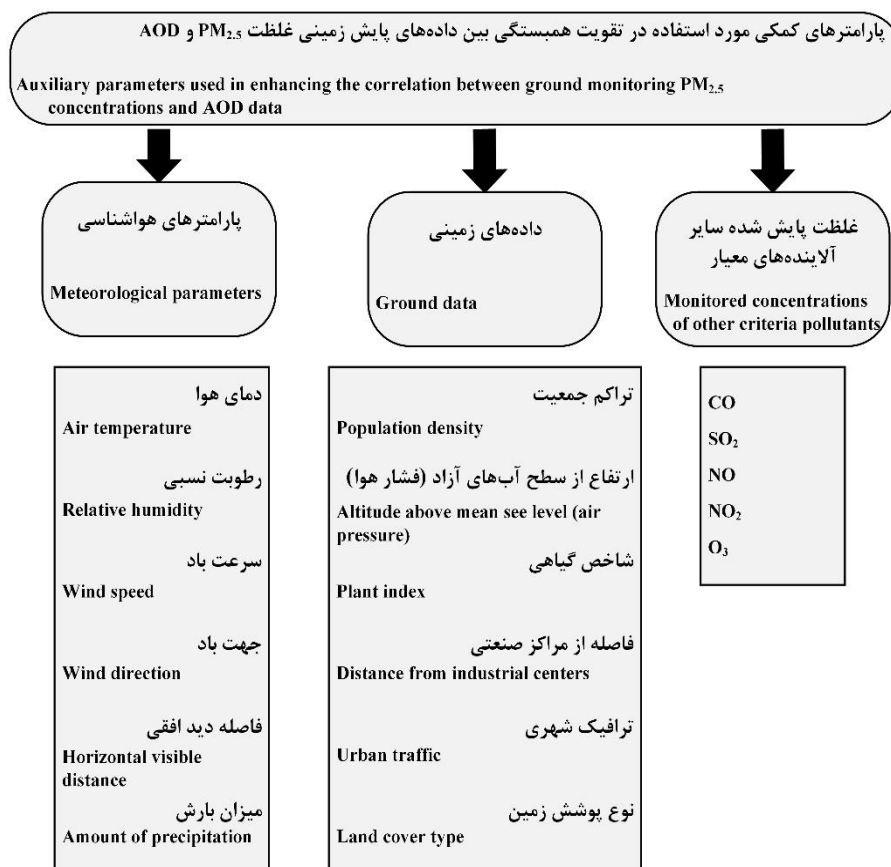


شکل ۳- فراوانی مقالات وارد شده در مطالعه مروری حاضر براساس سال انتشار

Fig.3. Frequency of Articles Included in the Current Review by Publication Year

در تعدادی از مطالعات بررسی شده در تحقیق حاضر، مقادیر بالای همبستگی بین غلظت PM<sub>2.5</sub> زمینی و داده‌های AOD، مشاهده شده است. این امر قابلیت بالای استفاده از داده‌های AOD به عنوان برآوردگر مناسبی در تخمین غلظت PM<sub>2.5</sub> زمینی را نشان می‌دهد (Beckerman et al., 2013; De Hoogh et al., 2016; Xue, Zhang, Zhong, Ji & Huang, 2020). با این وجود، برخی مطالعات نیز وجود همبستگی معنی‌دار بین PM<sub>2.5</sub> و AOD را رد کرده یا شدت پایینی از همبستگی میان آنها را نشان داده‌اند. این مستندات، حصول همبستگی قوی‌تر بین PM<sub>2.5</sub> و AOD را منوط به استفاده از متغیرهای دیگری همچون پارامترهای هواشناسی، دانسته‌اند. به بیان دیگر، از نظر تعدادی از محققان، برقراری ارتباط مطمئن میان PM<sub>2.5</sub> و AOD، مستلزم بهره‌گیری از داده‌های کمکی بیشتری است (Nabavi et al., 2016; Han & Tong, 2019; Jiang et al., 2021).

Yang et al., 2019؛ al., 2019). در شکل (۴) پارامترهای کمکی مورد استفاده در تقویت همبستگی بین داده‌های غلظت  $PM_{2.5}$  پایش شده زمینی و داده‌های AOD آورده شده است.



شکل ۴- پارامترهای کمکی در تقویت همبستگی بین داده‌های غلظت  $PM_{2.5}$  و AOD

Fig.4. Auxiliary Parameters Enhancing the Correlation Between  $PM_{2.5}$  Concentration and AOD Data

هو و همکاران (Hu et al., 2013)، بر این باور بودند که استفاده صرف از محصولات ماهواره‌ای ممکن است منجر به خطاهای برآورد در نتایج شود. ایشان در جهت رفع این خطاها پیشنهاد نمودند که علاوه بر AOD، از پارامترهای محلی نظیر متغیرهای پوشش زمین و پارامترهای هواشناسی، به عنوان ورودی‌های کمکی در مدل رگرسیون استفاده شود. آنها به این نتیجه رسیدند که ورود این پارامترهای کمکی می‌تواند عملکرد مدل رگرسیون خطی در برآورد غلظت  $PM_{2.5}$  را بهبود بخشد. یانگ و همکاران (Yang et al., 2019)، مشاهده کردند که در مناطق خشک مانند شمال چین، همبستگی بین  $PM_{2.5}$  و AOD بیشتر است؛ در حالی که در مناطق ساحلی، این همبستگی ضعیف‌تر بدست آمد. آنها عواملی همچون توپوگرافی سطحی و پارامترهای هواشناسی را بر تغییرات زمانی و مکانی رابطه بین  $PM_{2.5}$  و AOD مؤثر دانستند. در مطالعه‌ای دیگر، وای و همکاران (Yi et al., 2019)، به برآورد غلظت  $PM_{2.5}$  در استان‌های شمالی چین پرداختند. آن‌ها دریافتند که شرایط آب و

هوایی گوناگون در منطقه مورد مطالعه، تغییرات مختلفی در شدت همبستگی غلظت  $PM_{2.5}$  و AOD ایجاد نموده است و نتایج با مدل‌های بدست‌آمده از مناطق دیگر، تفاوت دارد. هوانگ و همکاران (Huang et al., 2019)، با استفاده از پارامترهای هواشناسی و ارتفاع از سطح دریاهای آزاد، ابتدا مقادیر گمشده AOD با تفکیک‌پذیری<sup>۱</sup> یک کیلومتر در شهر نیویورک را برآورد نموده و سپس با بهره‌گیری از داده‌های کمکی پوشش سطح، غلظت  $PM_{2.5}$  را تخمین زدند. میری و همکاران (Miri, Ghassoun, Dovlatabadi, Ebrahimnejad & Löwner, 2019)، در برآورد غلظت  $PM_{2.5}$  در شهر سبزوار، از پارامترهای کمکی پوشش زمین، فاصله از خیابان و مناطق صنعتی و همچنین ارتفاع، استفاده کردند. برخلاف مطالعات قبلی، در این تحقیق از مقادیر AOD در برآورد غلظت  $PM_{2.5}$  استفاده نشد. چن و همکاران (Chen, Wu, Tu & Cao, 2020)، از یک مدل رگرسیون خطی شامل پارامترهای هواشناسی، داده‌های مربوط به ترافیک، انواع پوشش زمین، جمعیت و داده‌های ارتفاعی در بازه زمانی زمستان ۲۰۱۲ تا ۲۰۱۵ در کشور چین، به منظور برآورد غلظت  $PM_{2.5}$  استفاده کردند. فنگ و همکاران (Feng et al., 2020)، در دو شهر پکن و چانگدو جهت برآورد غلظت  $PM_{2.5}$  از مقادیر AOD ماهواره‌ای، از پارامترهای کمکی هواشناسی و متغیرهای پوشش سطح استفاده نمودند. آنها در این مطالعه موفق به دستیابی به ضریب تعیین  $R^2$  برابر با ۰/۷۱ شدند. ژانگ و همکاران (Zhang et al., 2021)، عدم اطمینان موجود در برآورد  $PM_{2.5}$  از نتایج مبتنی بر AOD را یک چالش جدی دانسته و با استفاده از پارامترهای هواشناسی و متغیرهای پوشش زمین، سعی نمودند این مشکل را حل کنند. در پژوهشی دیگر، جیانگ و همکاران (Jiang et al., 2021)، بیان داشتند وقتی غلظت آئروسول در هوای آزاد بالا است، برآورد  $PM_{2.5}$  صرفاً براساس AOD دچار خطا می‌شود. آنها برای حل این مشکل از پارامترهای هواشناسی علاوه بر داده‌های AOD استفاده نمودند. استفاده از داده‌های کمکی در تحقیقات دیگر محققان نیز مورد اشاره قرار گرفته است (Guo, Tang, Li et al., 2018 a, b; Liu, Franklin, Kahn & Koutrakis, 2007; Gong & Zhang, 2017; Liu, Paciorek & Koutrakis, 2009; Meng et al., 2016; Moore, Jerrett, Mack & Künzli, 2007; Qi, Li, Karimian & Liu, 2019; Ross, Jerrett, Ito, Tempalski & Thurston, 2007; Shi, Ho, Xu & Ng, 2018; Su et al., 2009; Wu et al., 2017; Zou, Fang, Feng & Zhou, 2021).

(۱)، پارامترهای کمکی مورد استفاده توسط محققان در برآورد غلظت  $PM_{2.5}$  از داده‌های AOD به همراه دقت ارزیابی، ارائه شده است. همانطور که در جدول ۱ مشاهده می‌شود، مطالعات پیشین با استفاده از پارامترهایی نظیر دما، رطوبت نسبی، فشار سطحی و داده‌های تراکم جمعیت، به دقت‌های متنوعی دست یافته‌اند که نشان‌دهنده اهمیت ترکیب مناسب این پارامترها در بهبود دقت مدل‌ها است. همچنین، برخی مطالعات (به‌ویژه مطالعاتی که از داده‌های دقیق‌تر و متنوع‌تر استفاده کرده‌اند) موفق به کسب دقت بالاتری شده‌اند که نقش کلیدی انتخاب متغیرهای ورودی مناسب را تأیید می‌کند. این جدول علاوه بر نشان دادن تنوع متغیرهای به‌کاررفته، بر اهمیت استفاده از داده‌های محلی و تنظیمات خاص منطقه‌ای جهت بهبود پیش‌بینی  $PM_{2.5}$  تأکید دارد.

## جدول ۱- پارامترهای کمکی مورد استفاده در برآورد غلظت PM<sub>2.5</sub> براساس داده‌های AOD به‌همراه دقت نتایج

Table 1- Auxiliary Parameters Used in Estimating PM<sub>2.5</sub> Concentrations from AOD Data with Accuracy Results

مرجع Reference	% R <sup>2</sup>	AOD	پارامترهای هواشناسی Meteorological parameters										داده‌های زمینی Ground data							منطقه مورد مطالعه Study area
			شاخص کیفیت هوا Air quality index	فاصله دید افقی Horizontal visible distance	آب قابل بارش Precipitable water	دید Visibility	سرعت و جهت باد Wind speed and direction	رطوبت نسبی Relative humidity	بارش Precipitation	فشار Pressure	دما Temperature	PM <sub>2.5</sub>	PM <sub>10</sub>	نوع پوشش زمین Land cover type	نوع کاربری زمین Land use type	تراکم شهر Urban traffic	فاصله از مراکز صنعتی Distance from industrial centers	شاخص گیاهی Plant index	ارتفاع از سطح دریا Height above sea level	
Ayus et al., 2023	۹۹	✓	✓																	چین (China)
Reddy et al., 2018	۹۶	✓																		چین (China)
Shoghrhodaei et al., 2021	۹۵	✓																		تهران (Tehran)
Ren et al., 2024	۹۵	✓																		چین (China)
Miri et al., 2019	۹۲																			سوزور (Sabzevar)
Yang et al., 2024	۹۲	✓																		چین (China)
Feng et al., 2020	۹۱	✓																		چین (China)
Qi et al., 2019	۹۱	✓	✓																	چین (China)
Mohammadi et al., 2024	۹۰	✓																		تهران (Tehran)
Chen et al., 2021	۸۹	✓																		چین (China)
Liu et al., 2019	۸۶	✓																		چین (China)
Huang et al., 2019	۸۵	✓																		نیویورک (New York)
Jiang et al., 2021	۸۵	✓		✓																چین (China)
Obodoze et al., 2021	۸۵	✓																		چین (China)
Enebish et al., 2021	۸۲	✓																		مغولستان (Mongolia)
Li et al., 2018b	۸۱																			پکن (Beijing)
Zhang et al., 2021	۸۱	✓	✓																	چین (China)
Ghaemi et al., 2016	۸۱	✓																		تهران (Tehran)
Zamani Joharestani et al., 2019	۸۱	✓		✓																تهران (Tehran)
Kariman et al., 2019	۸۰	✓																		تهران (Tehran)
Aguilera et al., 2023	۷۸	✓																		آمریکا (USA)
Chen et al., 2024	۷۸	✓																		چین (China)
Yang et al., 2019	۷۷	✓																		منطقه ساحلی چین (Coastal region of China)
Cortina-Jamachs et al., 2015	۷۷	✓																		مکزیک (Mexico)
Yi et al., 2019	۷۶	✓																		شمال چین (North China)
Li et al., 2018a	۷۵	✓																		شاندونگ (Shandong)
Chen et al., 2020	۷۱																			جنوب چین (South China)
Zhang et al., 2018	۶۹	✓		✓																تگزاس (Texas)
Nabavi et al., 2019	۶۸	✓		✓																تهران (Tehran)
Yan et al., 2020	۶۲	✓																		چین (China)
Hu et al., 2013	۶۱	✓																		آمریکا (USA)
Kong et al., 2016	۵۸	✓																		پکن (Beijing)
Xu et al., 2018	۴۸	✓																		کانادا (Canada)

نکته دیگری که در اینجا حائز اهمیت است، احتمال بروز خطای بیش‌برازش<sup>۱</sup> در صورت استفاده از تعداد زیادی از پارامترهای کمکی در یک مدل است. هرچه تعداد پارامترهای قابل تنظیم در یک مدل بیشتر باشد، مدل پیچیده‌تر می‌شود و توانایی بیشتری برای یادگیری الگوهای بسیار خاص و جزئی در داده‌های آموزشی پیدا می‌کند. این امر باعث می‌شود مدل به شدت به داده‌های آموزشی وابسته شده و در مواجهه با داده‌های جدید که ممکن است اندکی متفاوت باشند، عملکرد ضعیفی از خود نشان دهد. به عبارت دیگر، یک مدل با پارامترهای کمکی زیاد، می‌تواند به راحتی داده‌های آموزشی را حفظ کند، اما این حفظ کردن به معنای یادگیری واقعی نیست. در واقع، مدل به جای یادگیری قوانین کلی حاکم بر داده‌ها، صرفاً جزئیات تصادفی و نویزهای موجود در داده‌های آموزشی را یاد می‌گیرد. این پدیده شبیه به این است که یک دانش‌آموز به جای درک مفاهیم اصلی یک درس، فقط به حفظ کردن متن کتاب بسنده کند. در نتیجه، زمانی که با سؤالی روبرو می‌شود که کمی

1 Overfitting



متفاوت از متن کتاب باشد، قادر به پاسخگویی نخواهد بود. بنابراین، افزایش تعداد پارامترهای کمکی بدون کنترل، می‌تواند به طور مستقیم به افزایش خطر بیش برآزش منجر شود. برای جلوگیری از این مشکل، روش‌های مختلفی مانند منظم‌سازی<sup>۱</sup>، کاهش ابعاد داده‌ها و انتخاب ویژگی‌های مهم و استفاده از تکنیک‌های توقف زودهنگام، به کار می‌روند (Karimian et al., 2019).

همانطور که قبلاً گفته شد، در سالیان اخیر از روش‌های نوین مبتنی بر دنباله زمانی در برآورد غلظت PM<sub>2.5</sub> از داده‌های AOD بهره‌برداری شده است. این روش‌ها شامل شبکه‌های عصبی مصنوعی (ANN)<sup>۲</sup>، الگوریتم‌های مختلف یادگیری ماشین همچون رگرسیون بردار پشتیبان (SVR)<sup>۳</sup> و یادگیری عمیق (Lai, Li & Pan, 2021 ; Peng, Wang, Xia & Gao, 2022 ; Li et al., 2023) و روش‌های یادگیری جمعی نظیر جنگل تصادفی (RF)<sup>۴</sup> و تقویت گرادیان، می‌شوند. با توجه به وجود رابطه غیرخطی بین PM<sub>2.5</sub> و متغیرهای پیش‌بینی کننده، روش‌های مذکور به دلیل توانایی قابل توجه در انتخاب و مدیریت تعداد زیادی متغیر مستقل و امکان تحلیل مسائل پیچیده‌تر، بسیار مورد توجه محققان بوده اند. در شکل (۵) نمودار فراوانی استفاده از روش‌های یادگیری ماشین در برآورد غلظت PM<sub>2.5</sub> براساس داده‌های AOD، آورده شده است. لازم بذکر است مبنای استخراج شکل (۵)، کلیه الگوریتم‌های بهره‌گیری شده توسط محقق در مستند انتشار یافته بوده؛ لیکن از نقطه نظر معرفی بهترین الگوریتم جهت انجام برآوردها، روش‌های ترکیبی یا Ensemble Model با فراوانی نه، جنگل تصادفی با فراوانی هفت؛ و XGboost با فراوانی سه؛ به ترتیب در جایگاه‌های اول تا سوم قرار می‌گیرند. استفاده از الگوریتم‌های یادگیری ماشین، انجام پیش‌بینی‌های دقیق در مسائل تحلیل پیشرفته داده‌ها را امکان‌پذیر نموده است (Vapnik, 2013). این روش‌ها مبتنی بر کرنل<sup>۵</sup> بوده و در تحقیقات با هدف پیش‌بینی دنباله زمانی (Das Chagas Moura, Zio, Lins & Droguett, 2011; Quan, Liu & Liu, 2010; Sapankevych & Ghaemi, Farnaghi & Alimohammadi, ) و (Sankar, 2009; Tay & Cao, 2001) و پیش‌بینی آلودگی هوا (2016)، با موفقیت به کار گرفته شده‌اند.

همانطور که از شکل ۵ برمی‌آید، الگوریتم‌های متنوعی در زمینه ایجاد همبستگی بین داده‌های زمینی پایش PM<sub>2.5</sub> و اطلاعات برگرفته از ماهواره (AOD) به کار گرفته شده است، اما میزان استفاده از هر الگوریتم به شدت متغیر است. الگوریتم‌هایی مانند Random Forest (RF) و XGBoost با بیشترین فراوانی در صدر جدول قرار دارند. این امر نشان می‌دهد که این الگوریتم‌ها به دلیل دقت و کارایی بالای خود در برآورد پارامترهای پیچیده و محیطی مانند غلظت PM<sub>2.5</sub>، بیشتر مورد توجه قرار گرفته‌اند. در مقابل، برخی الگوریتم‌ها با فراوانی کمتر استفاده شده‌اند که ممکن است به دلیل محدودیت‌های این الگوریتم‌ها در مدل‌سازی داده‌های غیرخطی و پیچیده باشد. به طور کلی، نمودار حاکی از روند رو به رشد استفاده از روش‌های پیشرفته‌تر و ترکیبی در

1 Regularization

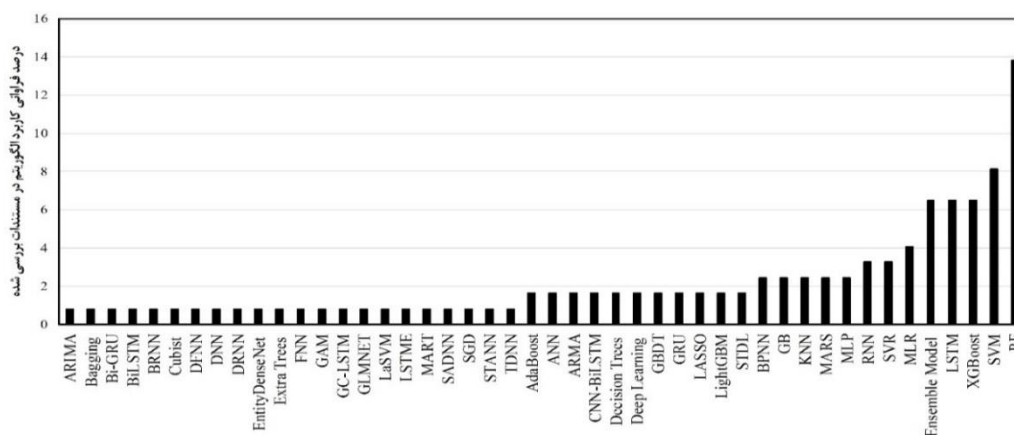
2 Artificial Neural network (ANN)

3 Support Vector Regression (SVR)

4 Random Forest (RF)

5 Kernel

مدل‌سازی پدیده‌های محیطی است که به مرور جایگزین روش‌های ساده‌تر شده‌اند. این توزیع همچنین نشان‌دهنده علاقه پژوهشگران به استفاده از الگوریتم‌های انعطاف‌پذیر و توانمند در پردازش داده‌های حجیم و پیچیده می‌باشد.



AutoRegressive Integrated Moving Average (ARIMA)	Auto Regression Moving Average (ARMA)
Bidirectional Gated Recurrent Unit (Bi-GRU)	Convolutional Neural Network BiLSTM (CNN-BiLSTM)
Bidirectional Long Short-Term Memory (BiLSTM)	Gradient Boosting Decision Tree (GBDT)
Bayesian Regularized Neural Networks (BRNN)	Gated Recurrent Unit (GRU)
Deep Feedforward Neural Networks (DFNN)	Least Absolute Shrinkage and Selection Operation (LASSO)
Deep Neural Networks (DNN)	Light Gradient Boosting Model (LightGBM)
Deep Recurrent Neural Networks (DRNN)	SpatioTemporal Deep Learning (STDL)
Feed-forward Neural Networks (FNN)	Back-Propagation Neural Networks (BPNN)
Generalized Additive Model (GAM)	Gradient Boosting (GB)
Graph Convolutional Long Short-Term Memory (GC-LSTM)	K-Nearest-Neighbors (KNN)
Generalized Linear Model with elastic NET penalties (GLMNET)	Multivariate Adaptive Regression Splines (MARS)
Large-scale approximate Support Vector Machine (LaSVM)	MultiLayer Perceptron (MLP)
Long Short-Term Memory Extended (LSTIME)	Recurrent Neural Network (RNN)
Multiple Additive Regression Trees (MART)	Support Vector Regression (SVR)
Self-Adaptive Deep Neural Network (SADNN)	Multi Linear Regression (MLR)
Stochastic Gradient Descent (SGD)	Long Short-Term Memory (LSTM)
SpatioTemporal Artificial Neural Network (STANN)	Extreme Gradient Boosting (XGBoost)
Time Delay Neural Network (TDNN)	Support Vector Machine (SVM)
Adaptive Boosting (AdaBoost)	Random Forest (RF)
Artificial Neural network (ANN)	

شکل ۵- فراوانی کاربرد الگوریتم‌های یادگیری ماشین در برآورد غلظت  $PM_{2.5}$

Fig.5. Frequency of Machine Learning Algorithm Applications in Estimating  $PM_{2.5}$  Concentrations

مدل ماشین بردار پشتیبان نه تنها به صورت مستقل در پیش‌بینی‌های مبتنی بر دنباله زمانی استفاده شده است، بلکه در ترکیب با روش‌های دیگر نیز می‌تواند سری‌های زمانی پیچیده و نامان را مدل‌سازی نماید. ژو و همکاران (Zhou et al., 2019)، این مدل را در ترکیب با روش‌های دیگر بکار برده و با استفاده از متغیرهای غلظت آلاینده‌های هوا، پارامترهای هواشناسی و داده‌های ترافیکی، به پیش‌بینی مکانی-زمانی آلودگی در چند ساعت آینده، پرداختند. با این حال، شبکه‌های ماشین بردار پشتیبان با وجود توانایی بالا در حل مسائل پیچیده، با محدودیت‌هایی نیز روبرو می‌باشند. به طوری که با افزایش حجم داده‌های ورودی، زمان پردازش نیز افزایش یافته و همچنین قادر به حفظ اطلاعات در دنباله زمانی نیستند و در پیش‌بینی‌هایی در چند گام جلوتر، شبکه با کاهش دقت مواجه می‌شود (Han, Zhao, Leung, Ma & Wang, 2019).

اوبودوئز و همکاران (Obodoeze, Nwabueze & Akaneme, 2021)، جهت برآورد غلظت  $PM_{2.5}$  در فاصله زمانی سال‌های ۲۰۰۸ تا ۲۰۱۳، از نه روش رگرسیون خطی چندگانه شامل رگرسیون بردار پشتیبان، رگرسیون خطی چندگانه، شبکه عصبی مصنوعی پرسپترون چند لایه، درخت تصمیم، درخت تصمیم تقویت شده با AdaBoost<sup>۱</sup>، جنگل تصادفی، XGBoost<sup>۲</sup>، Extra Trees و رگرسیون LASSO<sup>۳</sup>، در شهر پکن، چین، استفاده کردند. نتایج نشان دهنده عملکرد بهتر مدل XGBoost نسبت به سایر مدل‌ها بود.

شبکه‌های عصبی یکی دیگر از مدل‌های متداول در پیش‌بینی‌های مبتنی بر دنباله زمانی هستند. این شبکه‌ها با پیشینه‌ای حدود ۷۰ ساله، با موفقیت در حل مسائل مربوط به رگرسیون، طبقه‌بندی و استخراج ویژگی‌ها استفاده شده‌اند. وانگ و همکاران (Wang, Zeng & Chen, 2015)، لچترمچر و فولر (Lachtermacher & Fuller, 1995)، ملیت و پاون (Mellit & Pavan, 2010) و میشر و گوپال (Mishra & Goyal, 2015)، برای پیش‌بینی دنباله‌های زمانی و آلودگی هوا، از این روش استفاده کردند. در مطالعه‌ای دیگر، هان و همکاران (Han et al., 2019)، متوجه ضعف عملکرد شبکه عصبی در برابر مسائل پیچیده با حجم انبوه ورودی شدند که می‌تواند منجر به کاهش عملکرد شبکه و عدم توانایی در تقریب توابع پیچیده شود.

یادگیری عمیق، روشی جدید در حوزه یادگیری ماشین است که در سالیان اخیر در بخش هوش مصنوعی معرفی شده و قادر است الگوهای مرتبط با خصوصیات بسیار پیچیده را با استفاده از حجم بزرگی از داده‌های ورودی تشخیص دهد (Deng & Yu, 2014؛ Yamins & DiCarlo, 2016). در مقایسه با روش‌های سنتی، یادگیری عمیق می‌تواند به طور کارآمد، مدلی را از داده‌های آموزشی استخراج کرده و پیش‌بینی دقیق‌تری را ارائه دهد (Zhang, Zheng & Qi, 2017؛ Meyer, Kühnlein, Appelhans & Nauss, 2016). با گسترش استفاده از یادگیری عمیق در موضوعات متنوع، بررسی پیش‌بینی غلظت آلاینده‌های هوای شهری با استفاده از یادگیری عمیق در سال‌های اخیر مورد توجه ویژه محققان بوده است (Kuremoto, Kimura, Kobayashi & Obayashi, 2014؛ Zuo, Guo, Shi & Zhang, 2020). یادگیری عمیق می‌تواند از حجم بالای داده بهره برده و رابطه‌های پیچیده بین داده‌ها را از طریق آموزش کارآمد، شناسایی کند (Schmidhuber, 2015).

راهکارهای یادگیری عمیق در حوزه پیش‌بینی، مبتنی بر دنباله زمانی، شامل سه دسته اصلی (۱) بدون نظارت (مانند شبکه خود رمزنگار)؛ (۲) نظارت شده یا تمایزی (مانند شبکه‌های کانولوشنی و بازگشتی)؛ و (۳) یادگیری ترکیبی است (Sarker, 2021). لو و همکاران (Lv, Duan, Kang, Li & Wang, 2014) و لی و همکاران (Li, Peng, Hu, Shao & Chi, 2016)، از شبکه‌های خود رمزنگار جهت پیش‌بینی دنباله زمانی مانند پیش‌بینی جریان ترافیک و پیش‌بینی مصرف انرژی در ساختمان، استفاده کردند. در یک پژوهش در کشور ژاپن با هدف پیش‌بینی غلظت  $PM_{2.5}$  با استفاده از شبکه‌های خود رمزنگار عمیق، یک روش نوین پیش‌آموزش شبکه پویا برای داده‌های سری زمانی توسعه یافت. در این تحقیق، از داده‌های آلودگی هوا با دقت بالا در ۵۲ شهر کشور ژاپن و داده‌های هواشناسی استفاده شده است. مقایسه مدل پیشنهادی با داده‌های سازمان مطالعات محیط‌زیست

1 Adaptive Boosting (AdaBoost)

2 Extreme Gradient Boosting (XGBoost)

3 Least Absolute Shrinkage and Selection Operation (LASSO)

کشور ژاپن نشان‌دهنده عملکرد مناسب مدل پیشنهادی بوده است (Ong, Sugiura & Zettsu, 2016). در پژوهشی دیگر در کشور چین، پیش‌بینی میانگین غلظت آلاینده در تمام ایستگاه‌ها به طور همزمان برای چهار فصل آینده سال، توسط الگوریتم‌های ARMA<sup>۱</sup>، LSTM<sup>۲</sup>، LSTME<sup>۳</sup>، STDL<sup>۴</sup>، SVR و TDNN<sup>۵</sup> مورد بررسی قرار گرفت. در این مطالعه با استفاده از داده‌های ۱۲ ایستگاه سنجش آلودگی هوا در شهر پکن، برای در نظر گرفتن ارتباطات ذاتی مکانی-زمانی، از ترکیب دو مدل خود رمزنگار عمیق و رگرسیون منطقی استفاده شد. در این روش، خروجی آخرین لایه پنهان خود رمزنگار به عنوان ورودی برای لایه رگرسیون منطقی استفاده شده است. مدل ارائه شده در این تحقیق، براساس درصد میانگین مطلق خطای (MAPE)<sup>۶</sup> معادل ۲۵ درصد، در مقایسه با سایر روش‌های شبکه عصبی مصنوعی ارائه شده، دقت بالاتری داشته است (Li et al., 2017).

هوچریتز و اسچمیدبر (Hochreiter & Schmidhuber, 1997)، برای اولین بار نوع خاصی از شبکه‌های بازگشتی تحت عنوان LSTM را ارائه کردند که یکی از روش‌های پر کاربرد در زمینه پیش‌بینی آلودگی هوا به شمار می‌آید (Greff, Srivastava, Koutník, Steunebrink & Schmidhuber, 2016). به عنوان مثال در کشور چین، Reddy و همکاران (Reddy, Yedavalli, Mohanty & Nakhat, 2018)، با استفاده از روش حافظه طولانی کوتاه مدت (LSTM) در بازه سال‌های ۲۰۱۵ تا ۲۰۱۷، به پیش‌بینی غلظت PM<sub>2.5</sub> پرداختند. در مطالعه‌های دیگر، کریمیان و همکاران (Karimian et al., 2019)، با استفاده از داده‌های نه ایستگاه سنجش آلودگی هوا در شهر تهران در بازه سال‌های ۲۰۱۳ تا ۲۰۱۷ با استفاده از سه روش به پیش‌بینی غلظت PM<sub>2.5</sub> پرداختند که روش LSTM بهترین نتایج را به دست داد.

چانگ و همکاران (Chung, Gulcehre, Cho & Bengio, 2014)، شبکه بازگشتی GRU<sup>۴</sup> را ارائه کردند که مدل ساده‌سازی شده شبکه LSTM محسوب می‌شود. در مطالعه‌ای که آسیرا و همکاران (Athira, Geetha, Vinayakumar & Soman, 2018)، در کشور چین انجام دادند، از سه الگوی یادگیری عمیق شامل شبکه عصبی بازگشتی، شبکه LSTM و شبکه GRU جهت پیش‌بینی غلظت PM<sub>2.5</sub> از داده‌های ۱۴۹۸ سنجنده آلودگی استفاده شد. نتایج حاصله عملکرد بهتر شبکه GRU را اثبات می‌کند.

در پژوهشی که ژو و همکاران (Xu et al., 2018) در استان بریتیش کلمبیا در غرب کانادا انجام دادند، جهت برآورد غلظت ذرات PM<sub>2.5</sub> از روش‌های یادگیری ماشین استفاده گردید. براساس نتایج این مطالعه، روش‌های XGBoost، Cubist و جنگل تصادفی، عملکرد بهتری نسبت به سایر روش‌ها داشتند. در مطالعه‌ای دیگر، لیو و همکاران (Liu, Weng, Li & Cribb, 2019)، در برآورد غلظت PM<sub>2.5</sub> با استفاده از روش جنگل تصادفی (RF)، به مقدار ۰/۸۶ برای R<sup>2</sup> رسیدند. آهنی و همکاران (Ahani, Salari & Shadman, 2019) در برآورد غلظت PM<sub>2.5</sub> هوای آزاد شهر مشهد، از دو روش یادگیری ماشین شبکه عصبی چندلایه (MLP)<sup>۵</sup> و میانگین متحرک

1 Auto Regression Moving Average (ARMA)

2 Long Short-Term Memory (LSTM)

3 Long Short-Term Memory Extended (LSTME)

4 SpatioTemporal Deep Learning (STDL)

5 Time Delay Neural Network (TDNN)

6 Mean Absolute Percentage Error (MAPE)

خود همبسته یکپارچه (ARIMA)، استفاده نمودند. پارامترهای هواشناسی و کیفیت هوا بعنوان ورودی کمکی مدل‌ها استفاده شده و نتایج نشان داد که روش ARIMA بهترین عملکرد را در برآورد غلظت  $PM_{2.5}$  داشت. زمانی و همکاران (Zamani Joharestani, Cao, Ni, Bashir & Talebiesfandarani, 2019) برای پیش‌بینی غلظت آلاینده  $PM_{2.5}$  در شهر تهران، از سه روش یادگیری ماشین یعنی جنگل تصادفی، XGBoost و یادگیری عمیق استفاده کردند. در مقایسه با روش‌های جنگل تصادفی و یادگیری عمیق، روش XGBoost با داشتن مقدار  $R^2$  برابر با ۰/۸۱ و  $RMSE^1$  برابر با  $13/85 \mu g m^{-3}$ ، عملکرد بهتری را از خود نشان داد. پنگ و همکاران (Peng et al., 2022)، در استان هونان چین در مجموع ۱۷۵۲۰ داده را با بهره‌گیری از دو الگوریتم یادگیری ماشین (تقویت گرادیان XGBoost و یادگیری عمیق)، شش پارامتر هواشناسی (فشار اتمسفر، رطوبت نسبی، سرعت باد، جهت باد، بارندگی و دما) و غلظت  $PM_{2.5}$  که به طور مستقل از دو سایت پایش در طول سال ۲۰۲۱ اندازه‌گیری شده بود، مورد بررسی قرار دادند. نتایج حاکی از عملکرد بهتر شبکه XGBoost بوده است. در پژوهش انجام‌شده توسط انبیش و همکاران (Enebish, Chau, Jadamba & Franklin, 2021) در شهر اولان باتور در کشور مغولستان، محققان در بازه زمانی سال‌های ۲۰۱۰ تا ۲۰۱۸ از شش الگوریتم یادگیری ماشین استفاده کردند. نتایج این مطالعه نشان داد که الگوریتم‌های جنگل تصادفی و XGBoost عملکرد بهتری در مقایسه با سایر الگوریتم‌ها داشتند. در مطالعه شکرخدایی و همکاران (Shogrkhodaei, Razavi-Termeh & Fathnia, 2021) در شهر تهران، برای برآورد غلظت  $PM_{2.5}$  از سه روش یادگیری ماشین استفاده شده است: AdaBoost، جنگل تصادفی و گرادیان کاهشی تصادفی (SGD)<sup>۲</sup>. نتایج این پژوهش نشان داد که در میان این سه روش، روش جنگل تصادفی عملکرد بهتری داشته است. در مطالعه‌ای دیگر، آیوس و همکاران (Ayus, Natarajan & Gupta, 2023)، در بازه زمانی هفت ساله از سال ۲۰۱۳ تا ۲۰۲۱ برای ده شهر بزرگ در سراسر چین با استفاده از شش تکنیک مختلف یادگیری عمیق شامل شبکه عصبی مکرر (RNN)<sup>۳</sup>، واحد برگشتی دروازه‌ای دو جهته (Bi-GRU)<sup>۴</sup>، حافظه طولانی کوتاه مدت دو جهته (BiLSTM)<sup>۵</sup>، شبکه عصبی کانولوشنالی حافظه طولانی کوتاه مدت دو جهته (CNN-BiLSTM)<sup>۶</sup>، شبکه کانولوشنالی حافظه طولانی کوتاه مدت دو جهته (Conv1D-BiLSTM)<sup>۷</sup> و مدل تقویت گرادیان (XGBoost)، استفاده کردند. نتایج نشان داد Conv1D-BiLSTM و CNN-BiLSTM در بین مدل‌های یادگیری عمیق در برآورد شاخص کیفیت هوا (AQI)<sup>۸</sup>، عملکرد خوبی دارند؛ در حالی که مدل یادگیری ماشین (XGBoost)، بهتر از مدل‌های یادگیری عمیق است. لی و همکاران (Li, Huo & Liu, 2022)

1 Root Mean Squared Error (RMSE)

2 Stochastic Gradient Descent (SGD)

3 Recurrent Neural Network (RNN)

4 Bidirectional Gated Recurrent Unit (Bi-GRU)

5 Bidirectional Long Short-Term Memory (BiLSTM)

6 Convolutional Neural Network BiLSTM (CNN-BiLSTM)

7 Convolutional BiLSTM (Conv1D-BiLSTM)

8 Air Quality Index (AQI)

در شهر شینبی کشور تایوان با استفاده روش پیش‌بینی ترکیبی LSTM-XGBoost-SVR بر اساس ویژگی‌های زمانی-مکانی، توانستند دقت پیش‌بینی را بهبود بخشند.

محمدی و همکاران (Mohammadi, Teiri, Hajizadeh, Abdolahnejad & Ebrahimi, 2024)، از داده‌های هواشناسی نه سال گذشته در شهر اصفهان، برای پیش‌بینی غلظت PM<sub>2.5</sub> با استفاده از چهار الگوریتم یادگیری ماشین شامل شبکه‌های عصبی مصنوعی (ANN)، K-همسایگان نزدیک (KNN)<sup>۱</sup>، ماشین‌های بردار پشتیبان (SVM)<sup>۲</sup> و جنگل تصادفی (RF)، استفاده کردند. نتایج نشان داد که بر اساس درصد دقت، مدل ANN و RF در مقایسه با سایر مدل‌ها برای پیش‌بینی غلظت PM<sub>2.5</sub> عملکرد بهتری داشت.

در مطالعه‌ای دیگر یانگ و همکاران (Yang et al., 2024) برای برآورد غلظت PM<sub>2.5</sub>، یک مدل شبکه عصبی حافظه بلندمدت و کوتاه‌مدت (LSTM) و رگرسیون جنگل تصادفی (RF)، استفاده کردند. در نهایت، با توجه به شاخص‌های ارزیابی مدل، مدل رگرسیون جنگل تصادفی ( $RMSE = 4.59 \mu g m^{-3}$ ,  $R^2 = 0.93$ )، به عنوان مدل بهینه انتخاب شد. رن و همکاران (Ren, Zhang & Fan, 2024)، جهت برآورد غلظت ذرات معلق PM<sub>2.5</sub> از الگوریتم‌های جنگل تصادفی، XGBoost، MLR و مدل تلفیقی<sup>۳</sup> این الگوریتم‌ها استفاده کردند. نتایج نشان دهنده دقت بالای برآورد مدل تلفیقی نسبت به سایر مدل‌ها بود. در مطالعه‌ای دیگر چن و همکاران (Chen et al., 2024)، برای برآورد غلظت PM<sub>2.5</sub> از مدل تلفیقی توسعه یافته استفاده کردند؛ این الگوریتم براساس برآورد انباشتگی با ساختار دو لایه و تلفیق چندین مدل یادگیری ماشین، ساخته شده است. مدل تلفیقی توسعه یافته در این مطالعه در مقایسه با سایر مدل‌های مستقل (جنگل تصادفی، LightGBM<sup>۴</sup> و XGBoost) دقت بالاتری را نشان می‌دهد ( $R^2 = 0.87$ ).

در پژوهش انجام شده توسط آگیلرا و همکاران (Aguilera et al., 2023)، سه الگوریتم یادگیری ماشین (یعنی یادگیری عمیق، جنگل تصادفی (RF) و گرادیان تقویتی) در برآورد غلظت PM<sub>2.5</sub> مورد استفاده قرار گرفت. نتایج نشان داد که تلفیق بهینه این سه الگوریتم، بهترین عملکرد را با ضریب همبستگی  $R^2 = 0.78$  به دست خواهد داد. این یافته نشان می‌دهد که استفاده از یک ترکیب چند الگوریتمی (در مقایسه با استفاده از هر الگوریتم به تنهایی) می‌تواند به طور مؤثرتری غلظت PM<sub>2.5</sub> را برآورد نماید. رویکرد ترکیبی فوق مزایایی همچون بهبود دقت، پوشش وسیع‌تر ویژگی‌ها و مقاومت در برابر نقاط ضعف هر الگوریتم را به همراه خواهد داشت.

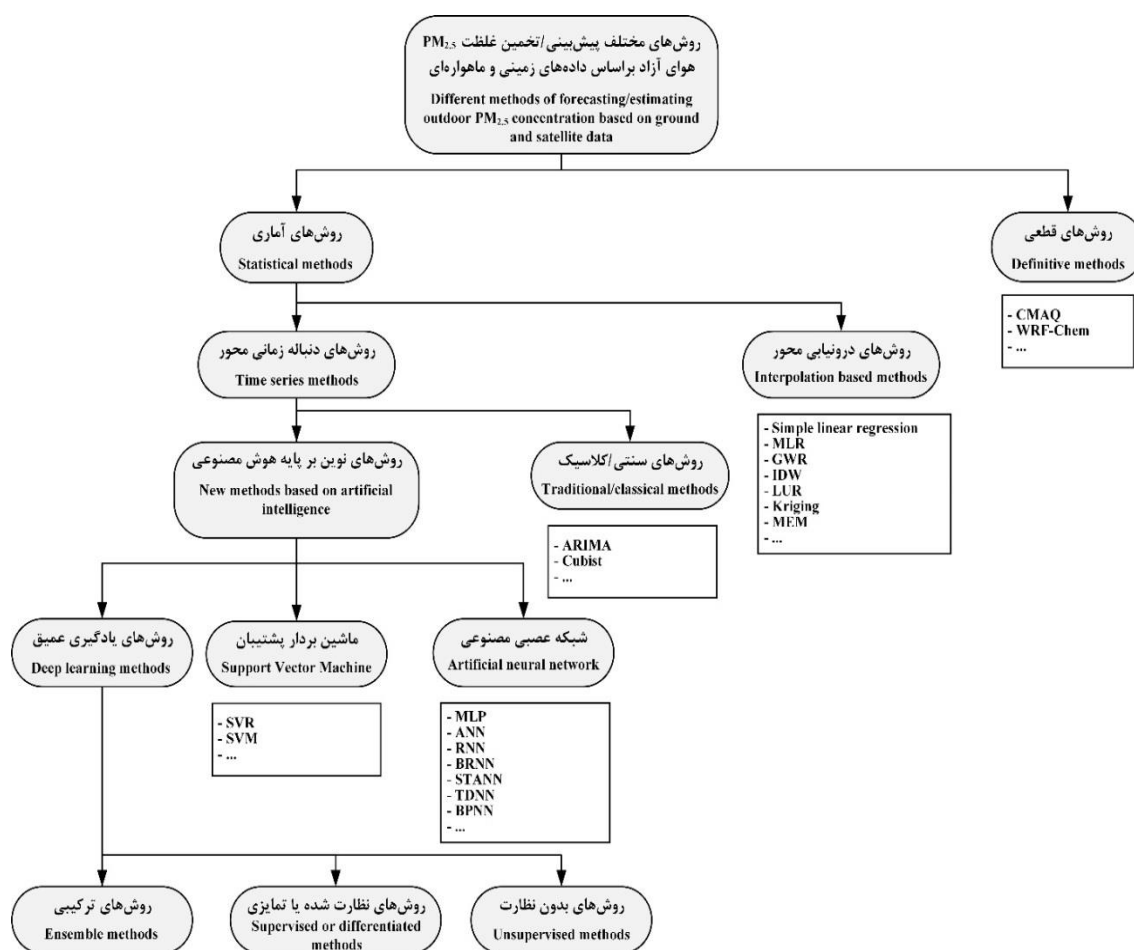
همانطور که در متن اشاره گردید، الگوریتم‌های یادگیری ماشین بخش نوینی از روش‌های آماری مبتنی بر دنباله زمانی بوده (شکل ۶) و روند استفاده از آنها در برآورد غلظت PM<sub>2.5</sub> زمینی براساس AOD، افزایشی است.

1 K-Nearest-Neighbors (KNN)

2 Support Vector Machine (SVM)

3 Ensemble Model

4 Light Gradient Boosting Model (LightGBM)

شکل ۶- روش‌های مختلف پیش‌بینی / برآورد غلظت  $PM_{2.5}$  براساس داده‌های زمینی و ماهواره‌ایFig.6. Various Methods for Predicting/Estimating  $PM_{2.5}$  Concentrations Based on Ground and Satellite Data

## نتیجه‌گیری

در برآورد غلظت  $PM_{2.5}$  زمینی، شاخص AOD معیار مناسبی ارائه می‌دهد. مطالعات بیانگر آن است که استفاده از داده‌های کمکی نظیر شرایط اتمسفری (رطوبت نسبی هوا، دمای سطحی، فاصله دید افقی و سرعت و جهت باد)، بر تقویت همبستگی  $PM_{2.5}$  و AOD تأثیرگذار خواهد بود. با این وجود، افزایش تعداد پارامترهای مورد استفاده، منجر به پیچیدگی روابط و افزایش خطای الگوریتم‌های یادگیری ماشین خواهد شد. استفاده از الگوریتم‌های تلفیقی یادگیری در چنین حالاتی، به دلیل توانایی در انتخاب و مدیریت تعداد پارامترهای زیاد، توصیه شده است. مجموعه داده‌های بکار رفته در برقراری همبستگی قوی بین غلظت  $PM_{2.5}$  زمینی و شاخص AOD توسط محققان عبارتند از:

(الف) داده‌های ماهواره‌ای: شامل عمق نوری آئروسول (AOD) که از سنسورهای MODIS و OMI استخراج

شده‌اند.

(ب) داده‌های ایستگاه‌های هواشناسی: شامل متغیرهای دما، رطوبت نسبی، سرعت باد، جهت باد و بارش.  
(ج) داده‌های زمینی: شامل شاخص پوشش گیاهی (NDVI)<sup>۱</sup>، تراکم جمعیت، تراکم شبکه جاده‌ای، ارتفاع و تعداد ساختمان‌ها.

(د) داده‌های آلودگی هوا: شامل غلظت PM<sub>2.5</sub> و سایر آلاینده‌ها مانند CO، O<sub>3</sub>، NO، NO<sub>2</sub> و SO<sub>2</sub>.  
مطابق با تحقیقات انجام گرفته، با توجه به حجم بالای داده‌ها و استفاده همزمان از چند پارامتر و همچنین به دلیل وجود رابطه غیرخطی بین PM<sub>2.5</sub> و متغیرهای پیش‌بینی کننده، الگوریتم‌های یادگیری ماشین بواسطه توانایی برتر در انتخاب و مدیریت تعداد زیادی متغیر مستقل، که قادر به پیش‌بینی متغیر وابسته در مسائل تحلیل پیشرفته داده‌ها و حل کردن مسائل پیچیده هستند، در سالیان اخیر در کانون توجه محققان قرار گرفته اند.

در بین الگوریتم‌های یادگیری ماشین، مدل‌های ترکیبی به دلیل برآوردهای دقیقتر، انعطاف‌پذیر بودن، پردازش موازی و سرعت بالا، توجه به داده‌های گمشده و صرف زمان کمتر برای آموزش و برآورد، در برآورد غلظت PM<sub>2.5</sub> زمینی عملکرد مناسبتری نسبت به عملکرد هر یک از الگوریتم‌ها به تنهایی، داشته اند. الگوریتم‌های ترکیبی که چندین مدل (همچون RF، XGboost و CNN<sup>۲</sup>) را ترکیب می‌کنند، سوگیری<sup>۳</sup> را کاهش داده، استحکام پیش‌بینی‌ها را افزایش می‌دهند و در بهبود دقت پیش‌بینی موثر هستند. با وجود پیشرفت‌های اخیر، هنوز چالش‌ها و محدودیت‌های ذاتی به دلیل عدم قطعیت در داده‌های ورودی و تغییرات جوی و هواشناسی، در این زمینه وجود داشته و اثربخشی مدل‌ها برای افق‌های پیش‌بینی طولانی‌تر (بیش از ۱۲ ساعت در برخی موارد)، کاهش می‌یابد.

از جمله این چالش‌ها که تحقیقات آتی باید به آن‌ها بپردازد می‌توان به عدم اطمینان در داده‌های ورودی، پیچیدگی‌های ناشی از تغییرات زمانی و مکانی و نیاز به ترکیب داده‌های چندمنبعی، اشاره کرد. برای رفع این چالش‌ها، تحقیقات آینده باید بر توسعه الگوریتم‌های ترکیبی پیشرفته‌تر متمرکز شوند که قابلیت مدیریت داده‌های گمشده، پردازش حجم بالای داده‌ها و تطبیق با تغییرات محیطی را داشته باشند. همچنین، استفاده از داده‌های ماهواره‌ای با وضوح مکانی و زمانی بالاتر، بهبود کیفیت متغیرهای کمکی (مانند داده‌های هواشناسی) و تلفیق داده‌های ایستگاه‌های زمینی و سنجش از دور می‌تواند دقت پیش‌بینی‌ها را افزایش دهد. علاوه بر این، طراحی مدل‌هایی که توانایی پیش‌بینی تأثیرات بلندمدت تغییرات آب‌وهوایی را داشته باشند، می‌تواند مسیر جدیدی برای پژوهش‌های آینده ایجاد کند.

1 Normalized Difference Vegetation Index (NDVI)

2 Convolutional Neural Network

3 Bias



## References

- Abbas, A., Ekowati, D., Suhariadi, F., & Fenitra, R. M. (2023). Health implications, leaders societies, and climate change: a global review. *Ecological footprints of climate change: Adaptive approaches and sustainability*, 653-675. [https://doi.org/10.1007/978-3-031-15501-7\\_26](https://doi.org/10.1007/978-3-031-15501-7_26)
- Aguilera, R., Luo, N., Basu, R., Wu, J., Clemesha, R., Gershunov, A., & Benmarhnia, T. (2023). A novel ensemble-based statistical approach to estimate daily wildfire-specific PM<sub>2.5</sub> in California (2006–2020). *Environment International*, 171, 107719. <https://doi.org/10.1016/j.envint.2022.107719>
- Ahani, I. K., Salari, M., & Shadman, A. (2019). Statistical models for multi-step-ahead forecasting of fine particulate matter in urban areas. *Atmospheric Pollution Research*, 10(3), 689-700. <https://doi.org/10.1016/j.apr.2018.11.006>
- Athira, V., Geetha, P., Vinayakumar, R., & Soman, K. P. (2018). DeepAirNet: Applying recurrent networks for air quality prediction. *Procedia Computer Science*, 132, 1394-1403. <https://doi.org/10.1016/j.procs.2018.05.068>
- Ayus, I., Natarajan, N., & Gupta, D. (2023). Comparison of machine learning and deep learning techniques for the prediction of air pollution: a case study from China. *Asian Journal of Atmospheric Environment*, 17(1), 4. <https://doi.org/10.1007/s44273-023-00005-w>
- Bai, L., Wang, J., Ma, X., & Lu, H. (2018). Air pollution forecasts: An overview. *International Journal of Environmental Research and Public Health*, 15(4), 780. <https://doi.org/10.3390/ijerph15040780>
- Beckerman, B. S., Jerrett, M., Serre, M., Martin, R. V., Lee, S. J., Van Donkelaar, A., ... & Burnett, R. T. (2013). A hybrid approach to estimating national scale spatiotemporal variability of PM<sub>2.5</sub> in the contiguous United States. *Environmental Science & Technology*, 47(13), 7233-7241. <https://doi.org/10.1021/es400039u>
- Bell, M. L., Dominici, F., Ebisu, K., Zeger, S. L., & Samet, J. M. (2007). Spatial and temporal variation in PM<sub>2.5</sub> chemical composition in the United States for health effects studies. *Environmental Health Perspectives*, 115(7), 989-995. <https://doi.org/10.1289/ehp.962>
- Bono, R., Tassinari, R., Bellisario, V., Gilli, G., Pazzi, M., Pirro, V., ... & Piccioni, P. (2015). Urban air and tobacco smoke as conditions that increase the risk of oxidative stress and respiratory response in youth. *Environmental Research*, 137, 141-146. <https://doi.org/10.1016/j.envres.2014.12.008>
- Chen, B., You, S., Ye, Y., Fu, Y., Ye, Z., Deng, J., ... & Hong, Y. (2021). An interpretable self-adaptive deep neural network for estimating daily spatially-continuous PM<sub>2.5</sub> concentrations across China. *Science of The Total Environment*, 768, 144724. <https://doi.org/10.1016/j.scitotenv.2020.144724>
- Chen, L., Wu, Z., Tu, W., & Cao, Z. (2020). Applying LUR model to estimate spatial variation of PM<sub>2.5</sub> in the Greater Bay Area, China. *In Spatiotemporal Analysis of Air Pollution and Its Application in Public Health*, 207-215. <https://doi.org/10.1016/B978-0-12-815822-7.00010-8>

- Chen, X., Zhang, W., He, J., Zhang, L., Guo, H., Li, J., & Gu, X. (2024). Mapping PM<sub>2.5</sub> concentration from the top-of-atmosphere reflectance of Himawari-8 via an ensemble stacking model. *Atmospheric Environment*, 120560. <https://doi.org/10.1016/j.atmosenv.2024.120560>
- Chudnovsky, A., Tang, C., Lyapustin, A., Wang, Y., Schwartz, J., & Koutrakis, P. J. A. C. (2013). A critical assessment of high-resolution aerosol optical depth retrievals for fine particulate matter predictions. *Atmospheric Chemistry and Physics*, 13(21), 10907-10917. <https://doi.org/10.5194/acp-13-10907-2013>
- Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. *arXiv preprint arXiv:1412.3555*. <https://doi.org/10.48550/arXiv.1412.3555>
- Cortina-Januchs, M. G., Quintanilla-Dominguez, J., Vega-Corona, A., & Andina, D. (2015). Development of a model for forecasting of PM<sub>10</sub> concentrations in Salamanca, Mexico. *Atmospheric Pollution Research*, 6(4), 626-634. <https://doi.org/10.5094/APR.2015.071>
- Das Chagas Moura, M., Zio, E., Lins, I. D., & Droguett, E. (2011). Failure and reliability prediction by support vector machines regression of time series data. *Reliability Engineering & System Safety*, 96(11), 1527-1534. <https://doi.org/10.1016/j.res.2011.06.006>
- De Hoogh, K., Gulliver, J., van Donkelaar, A., Martin, R. V., Marshall, J. D., Bechle, M. J., ... & Hoek, G. (2016). Development of West-European PM<sub>2.5</sub> and NO<sub>2</sub> land use regression models incorporating satellite-derived and chemical transport modelling data. *Environmental Research*, 151, 1-10. <https://doi.org/10.1016/j.envres.2016.07.005>
- Deng, L., & Yu, D. (2014). Deep learning: methods and applications. *Foundations and Trends® in signal processing*, 7(3-4), 197-387. <http://dx.doi.org/10.1561/20000000039>
- Enebish, T., Chau, K., Jadamba, B., & Franklin, M. (2021). Predicting ambient PM<sub>2.5</sub> concentrations in Ulaanbaatar, Mongolia with machine learning approaches. *Journal of Exposure Science & Environmental Epidemiology*, 31(4), 699-708. <https://doi.org/10.1038/s41370-020-0257-8>
- Feng, L., Li, Y., Wang, Y., & Du, Q. (2020). Estimating hourly and continuous ground-level PM<sub>2.5</sub> concentrations using an ensemble learning algorithm: The ST-stacking model. *Atmospheric Environment*, 223, 117242. <https://doi.org/10.1016/j.atmosenv.2019.117242>
- Fotheringham, A. S., Charlton, M. E., & Brunsdon, C. (1998). Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environment and planning A*, 30(11), 1905-1927. <https://doi.org/10.1068/a301905>
- Geng, G., Zhang, Q., Martin, R. V., Van Donkelaar, A., Hong, H., Che, H., ... & He, K. (2015). Estimating long-term PM<sub>2.5</sub> concentrations in China using satellite-based aerosol optical depth and a chemical transport model. *Remote Sensing of Environment*, 166, 262-270. <https://doi.org/10.1016/j.rse.2015.05.016>
- Ghaemi, Z., Farnaghi, M., & Alimohammadi, A. (2016). An Online Approach for Spatio-Temporal Prediction of Air Pollution in Tehran using Support Vector Machine. *Engineering Journal of Geospatial Information Technology*, 3(4), 43-63.[In Persian] <http://jgit.kntu.ac.ir/article-1-305-en.html>

- Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R., & Schmidhuber, J. (2016). LSTM: A search space odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, 28(10), 2222-2232. <https://doi.org/10.1109/TNNLS.2016.2582924>
- Gulliver, J., Morley, D., Dunster, C., McCrea, A., van Nunen, E., Tsai, M. Y., ... & Kelly, F. J. (2018). Land use regression models for the oxidative potential of fine particles (PM<sub>2.5</sub>) in five European areas. *Environmental Research*, 160, 247-255. <https://doi.org/10.1016/j.envres.2017.10.002>
- Guo, Y., Tang, Q., Gong, D. Y., & Zhang, Z. (2017). Estimating ground-level PM<sub>2.5</sub> concentrations in Beijing using a satellite-based geographically and temporally weighted regression model. *Remote Sensing of Environment*, 198, 140-149. <https://doi.org/10.1016/j.rse.2017.06.001>
- Han, W., & Tong, L. (2019). Satellite-based estimation of daily ground-level PM<sub>2.5</sub> concentrations over urban agglomeration of Chengdu Plain. *Atmosphere*, 10(5), 245. <https://doi.org/10.3390/atmos10050245>
- Han, Z., Zhao, J., Leung, H., Ma, K. F., & Wang, W. (2019). A review of deep learning models for time series prediction. *IEEE Sensors Journal*, 21(6), 7833-7848. <https://doi.org/10.1109/JSEN.2019.2923982>
- He, Q., & Huang, B. (2018). Satellite-based high-resolution PM<sub>2.5</sub> estimation over the Beijing-Tianjin-Hebei region of China using an improved geographically and temporally weighted regression model. *Environmental Pollution*, 236, 1027-1037. <https://doi.org/10.1016/j.envpol.2018.01.053>
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780. <https://doi.org/10.1162/neco.1997.9.8.1735>
- Hosseini, V., & Shahbazi, H. (2016). Urban air pollution in Iran. *Iranian Studies*, 49(6), 1029-1046. <https://doi.org/10.1080/00210862.2016.1241587>
- Hsu, N. C., Tsay, S. C., King, M. D., & Herman, J. R. (2004). Aerosol properties over bright-reflecting source regions. *IEEE Transactions on Geoscience and Remote Sensing*, 42(3), 557-569. <https://doi.org/10.1109/tgrs.2004.824067>
- Hu, X., Waller, L. A., Al-Hamdan, M. Z., Crosson, W. L., Estes Jr, M. G., Estes, S. M., ... & Liu, Y. (2013). Estimating ground-level PM<sub>2.5</sub> concentrations in the southeastern U.S. using geographically weighted regression. *Environmental Research*, 121, 1-10. <https://doi.org/10.1016/j.envres.2012.11.003>
- Hu, Z., Liebens, J., & Rao, K. R. (2011). Merging satellite measurement with ground-based air quality monitoring data to assess health effects of fine particulate matter pollution. *In Geospatial Analysis of Environmental Health* (pp. 395-409). Dordrecht: Springer Netherlands. [https://doi.org/10.1007/978-94-007-0329-2\\_20](https://doi.org/10.1007/978-94-007-0329-2_20)
- Huang, K., Bi, J., Meng, X., Geng, G., Lyapustin, A., Lane, K. J., ... & Liu, Y. (2019). Estimating daily PM<sub>2.5</sub> concentrations in New York City at the neighborhood-scale: Implications for integrating non-regulatory measurements. *Science of the Total Environment*, 697, 134094. <https://doi.org/10.1016/j.scitotenv.2019.134094>
- Institute for Environmental Research, Tehran University of Medical Sciences and National Institute of Health Research, (I.R. Iran). (2019). Air quality in Iran and its effects on health

- in 2017. Available at (accessed: May 2024): <https://enier.tums.ac.ir/Center-for-Air-Pollution-Research-News/Air-quality-in-Iran-its-effects-on-health-in-2017>
- Izah, S. C., Iyiola, A. O., Yarkwan, B., & Richard, G. (2023). Impact of air quality as a component of climate change on biodiversity-based ecosystem services. In *Visualization techniques for climate change with machine learning and artificial intelligence*, 123-148. <https://doi.org/10.1016/B978-0-323-99714-0.00005-4>
- Jia, N., Li, Y., Chen, R., & Yang, H. (2023). A review of global PM<sub>2.5</sub> exposure research trends from 1992 to 2022. *Sustainability*, 15(13), 10509. <https://doi.org/10.3390/su151310509>
- Jiang, T., Chen, B., Nie, Z., Ren, Z., Xu, B., & Tang, S. (2021). Estimation of hourly full-coverage PM<sub>2.5</sub> concentrations at 1-km resolution in China using a two-stage random forest model. *Atmospheric Research*, 248, 105146. <https://doi.org/10.1016/j.atmosres.2020.105146>
- Kampa, M., & Castanas, E. (2008). Human health effects of air pollution. *Environmental Pollution*, 151(2), 362-367. <https://doi.org/10.1016/j.envpol.2007.06.012>
- Karimian, H., Li, Q., Wu, C., Qi, Y., Mo, Y., Chen, G., ... & Sachdeva, S. (2019). Evaluation of different machine learning approaches to forecasting PM<sub>2.5</sub> mass concentrations. *Aerosol and Air Quality Research*, 19(6), 1400-1410. <https://doi.org/10.4209/aaqr.2018.12.0450>
- Kaufman, Y. J., Wald, A. E., Remer, L. A., Gao, B. C., Li, R. R., & Flynn, L. (1997). The MODIS 2.1- $\mu\text{m}$  channel-correlation with visible reflectance for use in remote sensing of aerosol. *IEEE transactions on Geoscience and Remote Sensing*, 35(5), 1286-1298. <https://doi.org/10.1109/36.628795>
- Kelly, F. J., & Fussell, J. C. (2012). Size, source and chemical composition as determinants of toxicity attributable to ambient particulate matter. *Atmospheric Environment*, 60, 504-526. <https://doi.org/10.1016/j.atmosenv.2012.06.039>
- Koelemeijer, R. B. A., Homan, C. D., & Matthijsen, J. (2006). Comparison of spatial and temporal variations of aerosol optical thickness and particulate matter over Europe. *Atmospheric Environment*, 40(27), 5304-5315. <https://doi.org/10.1016/j.atmosenv.2006.04.044>
- Kong, L., Xin, J., Zhang, W., & Wang, Y. (2016). The empirical correlations between PM<sub>2.5</sub>, PM<sub>10</sub> and AOD in the Beijing metropolitan region and the PM<sub>2.5</sub>, PM<sub>10</sub> distributions retrieved by MODIS. *Environmental Pollution*, 216, 350-360. <https://doi.org/10.1016/j.envpol.2016.05.085>
- Kuremoto, T., Kimura, S., Kobayashi, K., & Obayashi, M. (2014). Time series forecasting using a deep belief network with restricted Boltzmann machines. *Neurocomputing*, 137, 47-56. <https://doi.org/10.1016/j.neucom.2013.03.047>
- Lachtermacher, G., & Fuller, J. D. (1995). Back propagation in time-series forecasting. *Journal of Forecasting*, 14(4), 381-393. <https://doi.org/10.1002/for.3980140405>
- Lai, X., Li, H., & Pan, Y. (2021). A combined model based on feature selection and support vector machine for PM<sub>2.5</sub> prediction. *Journal of Intelligent & Fuzzy Systems*, 40(5), 10099-10113. <https://doi.org/10.3233/JIFS-202812>
- Lee, H. J., Coull, B. A., Bell, M. L., & Koutrakis, P. (2012). Use of satellite-based aerosol optical depth and spatial clustering to predict ambient PM<sub>2.5</sub> concentrations. *Environmental Research*, 118, 8-15. <https://doi.org/10.1016/j.envres.2012.06.011>

- Lee, H. J., Liu, Y., Coull, B. A., Schwartz, J., & Koutrakis, P. (2011). A novel calibration approach of MODIS AOD data to predict PM<sub>2.5</sub> concentrations. *Atmospheric Chemistry and Physics*, 11(15), 7991-8002. <https://doi.org/10.5194/acp-11-7991-2011>
- Lee, J. H., Wu, C. F., Hoek, G., de Hoogh, K., Beelen, R., Brunekreef, B., & Chan, C. C. (2015). LUR models for particulate matters in the Taipei metropolis with high densities of roads and strong activities of industry, commerce and construction. *Science of the Total Environment*, 514, 178-184. <https://doi.org/10.1016/j.scitotenv.2015.01.091>
- Li, H., Yu, Y., Huang, Z., Sun, S., & Jia, X. (2023). A multi-step ahead point-interval forecasting system for hourly PM<sub>2.5</sub> concentrations based on multivariate decomposition and kernel density estimation. *Expert Systems with Applications*, 226, 120140. <https://doi.org/10.1016/j.eswa.2023.120140>
- Li, L., Zhang, J., Meng, X., Fang, Y., Ge, Y., Wang, J., ... & Kan, H. (2018a). Estimation of PM<sub>2.5</sub> concentrations at a high spatiotemporal resolution using constrained mixed-effect bagging models with MAIAC aerosol optical depth. *Remote Sensing of Environment*, 217, 573-586. <https://doi.org/10.1016/j.rse.2018.09.001>
- Li, R., Ma, T., Xu, Q., & Song, X. (2018b). Using MAIAC AOD to verify the PM<sub>2.5</sub> spatial patterns of a land use regression model. *Environmental Pollution*, 243, 501-509. <https://doi.org/10.1016/j.envpol.2018.09.026>
- Li, X., Huo, H., & Liu, Z. (2022). Analysis and prediction of PM<sub>2.5</sub> concentration based on LSTM-XGBoost-SVR model. <https://doi.org/10.21203/rs.3.rs-2158285/v1>
- Li, X., Peng, L., Hu, Y., Shao, J., & Chi, T. (2016). Deep learning architecture for air quality predictions. *Environmental Science and Pollution Research*, 23, 22408-22417. <https://doi.org/10.1007/s11356-016-7812-9>
- Li, X., Peng, L., Yao, X., Cui, S., Hu, Y., You, C., & Chi, T. (2017). Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental Pollution*, 231, 997-1004. <https://doi.org/10.1016/j.envpol.2017.08.114>
- Liu, J., Weng, F., Li, Z., & Cribb, M. C. (2019). Hourly PM<sub>2.5</sub> estimates from a geostationary satellite based on an ensemble learning algorithm and their spatiotemporal patterns over central east China. *Remote Sensing*, 11(18), 2120. <https://doi.org/10.3390/rs11182120>
- Liu, Y., Franklin, M., Kahn, R., & Koutrakis, P. (2007). Using aerosol optical thickness to predict ground-level PM<sub>2.5</sub> concentrations in the St. Louis area: A comparison between MISR and MODIS. *Remote Sensing of Environment*, 107(1-2), 33-44. <https://doi.org/10.1016/j.rse.2006.05.022>
- Liu, Y., Paciorek, C. J., & Koutrakis, P. (2009). Estimating regional spatial and temporal variability of PM<sub>2.5</sub> concentrations using satellite data, meteorology, and land use information. *Environmental Health Perspectives*, 117(6), 886-892. <https://doi.org/10.1289/ehp.0800123>
- Liu, Y., Sarnat, J. A., Kilaru, V., Jacob, D. J., & Koutrakis, P. (2005). Estimating ground-level PM<sub>2.5</sub> in the eastern United States using satellite remote sensing. *Environmental Science & Technology*, 39(9), 3269-3278. <https://doi.org/10.1021/es049352m>

- Lu, J. G. (2020). Air pollution: A systematic review of its psychological, economic, and social effects. *Current Opinion in Psychology*, 32, 52-65. <https://doi.org/10.1016/j.copsyc.2019.06.024>
- Lu, J., Li, B., Li, H., & Al-Barakani, A. (2021). Expansion of city scale, traffic modes, traffic congestion, and air pollution. *Cities*, 108, 102974. <https://doi.org/10.1016/j.cities.2020.102974>
- Lv, Y., Duan, Y., Kang, W., Li, Z., & Wang, F. Y. (2014). Traffic flow prediction with big data: A deep learning approach. *IEEE Transactions on Intelligent Transportation Systems*, 16(2), 865-873. <https://doi.org/10.1109/TITS.2014.2345663>
- Mao, L., Qiu, Y., Kusano, C., & Xu, X. (2012). Predicting regional space-time variation of PM<sub>2.5</sub> with land-use regression model and MODIS data. *Environmental Science and Pollution Research*, 19, 128-138. <https://doi.org/10.1007/s11356-011-0546-9>
- Masroor, K., Fanaei, F., Yousefi, S., Raeesi, M., Abbaslou, H., Shahsavani, A., & Hadei, M. (2020). Spatial modelling of PM<sub>2.5</sub> concentrations in Tehran using Kriging and inverse distance weighting (IDW) methods. *Journal of Air Pollution and Health*, 5(2), 89-96. <https://doi.org/10.18502/japh.v5i2.4237>
- Mellit, A., & Pavan, A. M. (2010). A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy. *Solar Energy*, 84(5), 807-821. <https://doi.org/10.1016/j.solener.2010.02.006>
- Meng, X., Fu, Q., Ma, Z., Chen, L., Zou, B., Zhang, Y., ... & Liu, Y. (2016). Estimating ground-level PM<sub>10</sub> in a Chinese city by combining satellite data, meteorological information and a land use regression model. *Environmental Pollution*, 208, 177-184. <https://doi.org/10.1016/j.envpol.2015.09.042>
- Meyer, H., Kühnlein, M., Appelhans, T., & Nauss, T. (2016). Comparison of four machine learning algorithms for their applicability in satellite-based optical rainfall retrievals. *Atmospheric Research*, 169, 424-433. <https://doi.org/10.1016/j.atmosres.2015.09.021>
- Miri, M., Ghassoun, Y., Dovlatabadi, A., Ebrahimnejad, A., & Löwner, M. O. (2019). Estimate annual and seasonal PM<sub>1</sub>, PM<sub>2.5</sub> and PM<sub>10</sub> concentrations using land use regression model. *Ecotoxicology and Environmental Safety*, 174, 137-145. <https://doi.org/10.1016/j.ecoenv.2019.02.070>
- Mishra, D., & Goyal, P. (2015). Development of artificial intelligence based NO<sub>2</sub> forecasting models at Taj Mahal, Agra. *Atmospheric Pollution Research*, 6(1), 99-106. <https://doi.org/10.5094/APR.2015.012>
- Mohammadi, F., Teiri, H., Hajizadeh, Y., Abdolahnejad, A., & Ebrahimi, A. (2024). Prediction of atmospheric PM<sub>2.5</sub> level by machine learning techniques in Isfahan, Iran. *Scientific Reports*, 14(1), 2109. <https://doi.org/10.1038/s41598-024-52617-z>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group, T. (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Annals of Internal Medicine*, 151(4), 264-269. <https://doi.org/10.7326/0003-4819-151-4-200908180-00135>

- Moore, D. K., Jerrett, M., Mack, W. J., & Künzli, N. (2007). A land use regression model for predicting ambient fine particulate matter across Los Angeles, CA. *Journal of Environmental Monitoring*, 9(3), 246-252. <https://doi.org/10.1039/B615795E>
- Moryani, H. T., Kong, S., Du, J., & Bao, J. (2020). Health risk assessment of heavy metals accumulated on PM<sub>2.5</sub> fractioned road dust from two cities of Pakistan. *International Journal of Environmental Research and Public Health*, 17(19), 7124. <https://doi.org/10.3390/ijerph17197124>
- Nabavi, S. O., Haimberger, L., & Abbasi, E. (2019). Assessing PM<sub>2.5</sub> concentrations in Tehran, Iran, from space using MAIAC, deep blue, and dark target AOD and machine learning algorithms. *Atmospheric Pollution Research*, 10(3), 889-903. <https://doi.org/10.1016/j.apr.2018.12.017>
- Obodoeze, F. C., Nwabueze, C. A., & Akaneme, S. A. (2021). Comparative Evaluation of Machine Learning Regression Algorithms for PM<sub>2.5</sub> Monitoring. *American Journal of Engineering Research*, 10(12), 19-33.
- Ong, B. T., Sugiura, K., & Zettsu, K. (2016). Dynamically pre-trained deep recurrent neural networks using environmental monitoring data for predicting PM<sub>2.5</sub>. *Neural Computing and Applications*, 27, 1553-1566. <https://doi.org/10.1007/s00521-015-1955-3>
- Osimobi, O. J., Yorkor, B., & Nwankwo, C. A. (2019). Evaluation of daily pollutant standard index and air quality index in a university campus in Nigeria using PM<sub>10</sub> and PM<sub>2.5</sub> particulate matter. *Journal of Science, Technology and Environment Informatics*, 7(2), 517-532. <https://doi.org/10.18801/jstei.070219.54>
- Paciorek, C. J., Liu, Y., Moreno-Macias, H., & Kondragunta, S. (2008). Spatiotemporal associations between GOES aerosol optical depth retrievals and ground-level PM<sub>2.5</sub>. *Environmental science & technology*, 42(15), 5800-5806. <https://doi.org/10.1021/es703181j>
- Peng, L., Wang, L., Xia, D., & Gao, Q. (2022). Effective energy consumption forecasting using empirical wavelet transform and long short-term memory. *Energy*, 238, 121756. <https://doi.org/10.1016/j.energy.2021.121756>
- Perrone, M. G., Gualtieri, M., Consonni, V., Ferrero, L., Sangiorgi, G., Longhin, E., ... & Camatini, M. (2013). Particle size, chemical composition, seasons of the year and urban, rural or remote site origins as determinants of biological effects of particulate matter on pulmonary cells. *Environmental Pollution*, 176, 215-227. <https://doi.org/10.1016/j.envpol.2013.01.012>
- Pope III, C. A., & Dockery, D. W. (2006). Health effects of fine particulate air pollution: lines that connect. *Journal of the Air & Waste Management Association*, 56(6), 709-742. <https://doi.org/10.1080/10473289.2006.10464485>
- Qi, Y., Li, Q., Karimian, H., & Liu, D. (2019). A hybrid model for spatiotemporal forecasting of PM<sub>2.5</sub> based on graph convolutional neural network and long short-term memory. *Science of the Total Environment*, 664, 1-10. <https://doi.org/10.1016/j.scitotenv.2019.01.333>
- Quan, T., Liu, X., & Liu, Q. (2010). Weighted least squares support vector machine local region method for nonlinear time series prediction. *Applied Soft Computing*, 10(2), 562-566. <https://doi.org/10.1016/j.asoc.2009.08.025>

- Reddy, V., Yedavalli, P., Mohanty, S., & Nakhat, U. (2018). Deep air: forecasting air pollution in Beijing, China. *Environmental Science*, 1564.
- Ren, Y., Zhang, Y., & Fan, S. (2024). PM<sub>2.5</sub> Inversion Based on XGBoost And LightGBM Integrated Models. *Proceedings of the 4<sup>th</sup> International Conference on Environment Resources and Energy Engineering (ICEREE 2024)*. <https://doi.org/10.1051/e3sconf/202452002023>
- Ross, Z., Jerrett, M., Ito, K., Tempalski, B., & Thurston, G. D. (2007). A land use regression for predicting fine particulate matter concentrations in the New York City region. *Atmospheric Environment*, 41(11), 2255-2269. <https://doi.org/10.1016/j.atmosenv.2006.11.012>
- Saeed, S., Hussain, L., Awan, I. A., & Idris, A. (2017). Comparative analysis of different statistical methods for prediction of PM<sub>2.5</sub> and PM<sub>10</sub> concentrations in advance for several hours. *IJCSNS International Journal of Computer Science and Network Security*, 17(11), 45-52. [http://ijcsns.org/07\\_book/html/201711/201711006.html](http://ijcsns.org/07_book/html/201711/201711006.html)
- Sapankevych, N. I., & Sankar, R. (2009). Time series prediction using support vector machines: a survey. *IEEE Computational Intelligence Magazine*, 4(2), 24-38. <https://doi.org/10.1109/MCI.2009.932254>
- Sarker, I. H. (2021). Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, 2(6), 420. <https://doi.org/10.1007/s42979-021-00815-1>
- Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, 85-117. <https://doi.org/10.1016/j.neunet.2014.09.003>
- Shi, Y., Ho, H. C., Xu, Y., & Ng, E. (2018). Improving satellite aerosol optical depth-PM<sub>2.5</sub> correlations using land use regression with microscale geographic predictors in a high-density urban context. *Atmospheric Environment*, 190, 23-34. <https://doi.org/10.1016/j.atmosenv.2018.07.021>
- Shogrkhodaei, S. Z., Razavi-Termeh, S. V., & Fathnia, A. (2021). Spatio-temporal modeling of PM<sub>2.5</sub> risk mapping using three machine learning algorithms. *Environmental Pollution*, 289, 117859. <https://doi.org/10.1016/j.envpol.2021.117859>
- Song, Y., Qin, S., Qu, J., & Liu, F. (2015). The forecasting research of early warning systems for atmospheric pollutants: A case in Yangtze River Delta region. *Atmospheric Environment*, 118, 58-69. <https://doi.org/10.1016/j.atmosenv.2015.06.032>
- Stern, R., Builtjes, P. J. H., Schaap, M., Timmermans, R., Vautard, R., Hodzic, A., ... & Kerschbaumer, A. (2008). A model inter-comparison study focussing on episodes with elevated PM<sub>10</sub> concentrations. *Atmospheric Environment*, 42(19), 4567-4588. <https://doi.org/10.1016/j.atmosenv.2008.01.068>
- Su, J. G., Jerrett, M., Beckerman, B., Wilhelm, M., Ghosh, J. K., & Ritz, B. (2009). Predicting traffic-related air pollution in Los Angeles using a distance decay regression selection strategy. *Environmental Research*, 109(6), 657-670. <https://doi.org/10.1016/j.envres.2009.06.001>
- Taheri Shahraini, H., & Sodoudi, S. (2016). Statistical modeling approaches for PM<sub>10</sub> prediction in urban areas; A review of 21st-century studies. *Atmosphere*, 7(2), 15. <https://doi.org/10.3390/atmos7020015>



- Tay, F. E., & Cao, L. (2001). Application of support vector machines in financial time series forecasting. *Omega*, 29(4), 309-317. [https://doi.org/10.1016/S0305-0483\(01\)00026-3](https://doi.org/10.1016/S0305-0483(01)00026-3)
- U.S. EPA. (2024). Criteria air pollutants, the National Ambient Air Quality Standards (NAAQS) Table. Available at (accessed: May 2024): <https://www.epa.gov/criteria-air-pollutants/naaqs-table>
- Van Donkelaar, A., Martin, R. V., Brauer, M., Kahn, R., Levy, R., Verduzco, C., & Villeneuve, P. J. (2010). Global estimates of ambient fine particulate matter concentrations from satellite-based aerosol optical depth: development and application. *Environmental Health Perspectives*, 118(6), 847-855. <https://doi.org/10.1289/ehp.0901623>
- Vapnik, V. (2013). *The nature of statistical learning theory*. Springer science & business media. New York: Springer.
- Vicedo-Cabrera, A. M., Biggeri, A., Grisotto, L., Barbone, F., & Catelan, D. (2013). A Bayesian kriging model for estimating residential exposure to air pollution of children living in a high-risk area in Italy. *Geospatial Health*, 8(1), 87-95. <https://doi.org/10.4081/gh.2013.57>
- Wang, J., & Christopher, S. A. (2003). Intercomparison between satellite-derived aerosol optical thickness and PM<sub>2.5</sub> mass: Implications for air quality studies. *Geophysical Research Letters*, 30(21). <https://doi.org/10.1029/2003GL018174>
- Wang, L., Zeng, Y., & Chen, T. (2015). Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. *Expert Systems with Applications*, 42(2), 855-863. <https://doi.org/10.1016/j.eswa.2014.08.018>
- West, J. J., Cohen, A., Dentener, F., Brunekreef, B., Zhu, T., Armstrong, B., ... & Wiedinmyer, C. (2016). What we breathe impacts our health: improving understanding of the link between air pollution and health. *Environmental Science & Technology*, 50(10), 4895-4904. <https://doi.org/10.1021/acs.est.5b03827>
- World Health Organization. (2016). Ambient air pollution: A global assessment of exposure and burden of disease. *Clean Air J*, 26(2). <https://iris.who.int/handle/10665/250141>
- Wu, C. D., Chen, Y. C., Pan, W. C., Zeng, Y. T., Chen, M. J., Guo, Y. L., & Lung, S. C. C. (2017). Land-use regression with long-term satellite-based greenness index and culture-specific sources to model PM<sub>2.5</sub> spatial-temporal variability. *Environmental Pollution*, 224, 148-157. <https://doi.org/10.1016/j.envpol.2017.01.074>
- Xiao, Q., Wang, Y., Chang, H. H., Meng, X., Geng, G., Lyapustin, A., & Liu, Y. (2017). Full-coverage high-resolution daily PM<sub>2.5</sub> estimation using MAIAC AOD in the Yangtze River Delta of China. *Remote Sensing of Environment*, 199, 437-446. <https://doi.org/10.1016/j.rse.2017.07.023>
- Xu, Q., Chen, X., Yang, S., Tang, L., & Dong, J. (2021). Spatiotemporal relationship between Himawari-8 hourly columnar aerosol optical depth (AOD) and ground-level PM<sub>2.5</sub> mass concentration in mainland China. *Science of the Total Environment*, 765, 144241. <https://doi.org/10.1016/j.scitotenv.2020.144241>
- Xu, X., & Zhang, C. (2020). Estimation of ground-level PM<sub>2.5</sub> concentration using MODIS AOD and corrected regression model over Beijing, China. *PLoS One*, 15(10), e0240430. <https://doi.org/10.1371/journal.pone.0240430>

- Xu, Y., Ho, H. C., Wong, M. S., Deng, C., Shi, Y., Chan, T. C., & Knudby, A. (2018). Evaluation of machine learning techniques with multiple remote sensing datasets in estimating monthly concentrations of ground-level PM<sub>2.5</sub>. *Environmental Pollution*, 242, 1417-1426. <https://doi.org/10.1016/j.envpol.2018.08.029>
- Xue, Q., Tian, Y., Liu, X., Wang, X., Huang, B., Zhu, H., & Feng, Y. (2022). Potential risks of PM<sub>2.5</sub>-bound polycyclic aromatic hydrocarbons and heavy metals from inland and marine directions for a marine background site in North China. *Toxics*, 10(1), 32. <https://doi.org/10.3390/toxics10010032>
- Xue, W., Zhang, J., Zhong, C., Ji, D., & Huang, W. (2020). Satellite-derived spatiotemporal PM<sub>2.5</sub> concentrations and variations from 2006 to 2017 in China. *Science of the Total Environment*, 712, 134577. <https://doi.org/10.1016/j.scitotenv.2019.134577>
- Yamins, D. L., & DiCarlo, J. J. (2016). Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, 19(3), 356-365. <https://doi.org/10.1038/nn.4244>
- Yan, X., Zang, Z., Luo, N., Jiang, Y., & Li, Z. (2020). New interpretable deep learning model to monitor real-time PM<sub>2.5</sub> concentrations from satellite data. *Environment International*, 144, 106060. <https://doi.org/10.1016/j.envint.2020.106060>
- Yang, Q., Yuan, Q., Yue, L., Li, T., Shen, H., & Zhang, L. (2019). The relationships between PM<sub>2.5</sub> and aerosol optical depth (AOD) in mainland China: About and behind the spatio-temporal variations. *Environmental Pollution*, 248, 526-535. <https://doi.org/10.1016/j.envpol.2019.02.071>
- Yang, Y., Wang, Z., Cao, C., Xu, M., Yang, X., Wang, K., ... & Shi, Z. (2024). Estimation of PM<sub>2.5</sub> Concentration across China Based on Multi-Source Remote Sensing Data and Machine Learning Methods. *Remote Sensing*, 16(3), 467. <https://doi.org/10.3390/rs16030467>
- Yi, L., Mengfan, T., Kun, Y., Yu, Z., Xiaolu, Z., Miao, Z., & Yan, S. (2019). Research on PM<sub>2.5</sub> estimation and prediction method and changing characteristics analysis under long temporal and large spatial scale-A case study in China typical regions. *Science of the Total Environment*, 696, 133983. <https://doi.org/10.1016/j.scitotenv.2019.133983>
- Zamani Joharestani, M., Cao, C., Ni, X., Bashir, B., & Talebiesfandarani, S. (2019). PM<sub>2.5</sub> prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data. *Atmosphere*, 10(7), 373. <https://doi.org/10.3390/atmos10070373>
- Zhang, J., Zheng, Y., & Qi, D. (2017). Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 31(1). <https://doi.org/10.1609/aaai.v31i1.10735>
- Zhang, T., He, W., Zheng, H., Cui, Y., Song, H., & Fu, S. (2021). Satellite-based ground PM<sub>2.5</sub> estimation using a gradient boosting decision tree. *Chemosphere*, 268, 128801. <https://doi.org/10.1016/j.chemosphere.2020.128801>
- Zhang, X., Chu, Y., Wang, Y., & Zhang, K. (2018). Predicting daily PM<sub>2.5</sub> concentrations in Texas using high-resolution satellite aerosol optical depth. *Science of the Total Environment*, 631, 904-911. <https://doi.org/10.1016/j.scitotenv.2018.02.255>

- Zhou, Y., Chang, F. J., Chang, L. C., Kao, I. F., Wang, Y. S., & Kang, C. C. (2019). Multi-output support vector machine for regional multi-step-ahead PM<sub>2.5</sub> forecasting. *Science of the Total Environment*, 651, 230-240. <https://doi.org/10.1016/j.scitotenv.2018.09.111>
- Zou, B., Fang, X., Feng, H., & Zhou, X. (2021). Simplicity versus accuracy for estimation of the PM<sub>2.5</sub> concentration: A comparison between LUR and GWR methods across time scales. *Journal of Spatial Science*, 66(2), 279-297. <https://doi.org/10.1080/14498596.2019.1624203>
- Zuo, X., Guo, H., Shi, S., & Zhang, X. (2020). Comparison of six machine learning methods for estimating PM<sub>2.5</sub> concentration using the Himawari-8 aerosol optical depth. *Journal of the Indian Society of Remote Sensing*, 48(9), 1277-1287. <https://doi.org/10.1007/s12524-020-01154-z>