Contemporary Mathematics
https://ojs.wiserpub.com/index.php/CM/

Review

A Systematic Literature Review on the Estimation of High Air Pollution Periods Using Machine Learning Approaches

Nawwal Dhwaiher N Alrasheedi[1,2*], **Nurulkamal Masseran**[1], **Razik Ridzuan Mohd Tajuddin**[1]

[1]Department of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600, UKM Bangi, Selangor, Malaysia
[2]Department of Statistics and Operation Research, College of Science, AL-Qassim University, Buraida, Saudi Arabia
E-mail: math.word@gmail.com

Received: 20 September 2024; **Revised:** 13 January 2025; **Accepted:** 21 January 2025

Abstract: Air pollution remains a pressing global issue, presenting serious threats to environmental sustainability and public health. This systematic review examines the application of machine learning (ML) techniques for predicting periods of elevated air pollution between 2016 and 2023, guided by the preferred reporting items for systematic reviews and meta-analyses (PRISMA) framework. The review encompasses a range of ML approaches-including classical algorithms, deep learning, ensemble methods, and hybrid models-evaluating their performance in improving prediction accuracy. Particular attention is given to the role of key pollutants, such as nitrogen oxides (NO_x), and localized emission sources in influencing model outputs. While deep learning techniques effectively capture complex temporal patterns, ensemble and hybrid models demonstrate superior robustness and adaptability, especially in modeling spatial heterogeneity. Despite these advancements, notable challenges persist, including high computational requirements, limited generalizability across regions, and issues with data quality. By tracing the progression of ML applications in this field, the review synthesizes current achievements, outlines existing limitations, and offers strategic recommendations to enhance model scalability, interpretability, and practical deployment. The findings aim to guide researchers, policymakers, and environmental practitioners in advancing accurate and actionable air pollution forecasting through machine learning.

Keywords**:** machine learning, air pollution estimation, systematic literature review, forecasting, environmental healths

MSC: 68T07, 62P12

1. Introduction

Air pollution stands as a complex and urgent global concern, originating from a diverse array of anthropogenic activities that encompass industrial emissions, vehicular traffic, agricultural practices, and energy production [1, 2]. This intricate interplay of human activities contributes to the dynamic spatial and temporal distribution of air pollutants significantly, posing formidable challenges for effective air quality management. The extent and complexity of this issue are underscored by the complex relationship of meteorological conditions, geographical features, and atmospheric chemistry, each playing a distinct role in shaping the patterns of air pollution [3, 4].

DOI: https://doi.org/10.37256/cm.6320255760

The pervasive nature of air pollution unfolds into far-reaching consequences for both human health and the environment, creating a complex nexus of interconnected challenges. Human exposure to pollutants such as fine particulate matter (PM2.5), ozone (O_3), and nitrogen dioxide (NO_2) has been linked to a spectrum of health problems, encompassing respiratory diseases, cardiovascular complications, and premature mortality [5–7]. Beyond its direct impact on human health, air pollution significantly contributes to broader environmental challenges, playing a pivotal role in climate change, biodiversity loss, and disruptions in ecosystems. This emphasizes the multifaceted and global nature of the challenge, requiring comprehensive and integrated approaches for effective mitigation and management [1, 8, 9].

In response to the urgent imperative to address the adverse effects of air pollution, a substantial body of research has been dedicated to assessing, modeling, and forecasting air pollution levels. This research becomes particularly crucial during high pollution periods, driven by dynamic factors such as weather conditions and pollutant sources, which contribute to the variability and intensity of air pollution events [3, 10]. Within this intricate context, machine learning emerges as a promising and innovative approach, showcasing its capability to capture intricate patterns and dynamic interactions within air pollution data. The adoption of machine learning techniques holds the promise of significantly enhancing the accuracy of air quality predictions, providing valuable insights for effective mitigation strategies and policy development [11–13].

The significance of accurate air quality predictions is underscored by their potential to mitigate the health risks associated with air pollution. High pollution periods, exacerbated by factors such as weather conditions and pollutant sources, necessitate precise forecasting to minimize health risks and environmental damage. In this critical context, the integration of machine learning techniques becomes instrumental, given its ability to unravel complex patterns within air pollution data, providing a sophisticated understanding of the underlying dynamics. Machine learning, thus, stands as a valuable tool for improving predictions and interventions during these crucial periods, contributing to more effective strategies for public health and environmental management [11, 12]. The synthesis of interdisciplinary research efforts, merging environmental science with advanced computational methodologies, marks a transformative step toward addressing the multifaceted challenges posed by air pollution on a global scale.

1.1 *Background and motivation*

The pervasive and detrimental impacts of air pollution on human health and the environment underscore an urgent need for comprehensive research and effective strategies. As highlighted in the existing literature, air pollution not only poses significant health risks, contributing to respiratory diseases, cardiovascular problems, and various adverse health outcomes, but also exerts adverse environmental effects, including contributions to climate change, biodiversity loss, and ecological imbalances. This complex and urgent global issue necessitates a thorough exploration of its multifaceted consequences, requiring interdisciplinary efforts to understand and address the intricate dynamics leading to high air pollution events [3, 4, 8].

The recognition of air pollution as a complex challenge has prompted a multidisciplinary approach, acknowledging the interconnectedness of human activities, meteorological conditions, and geographical features that contribute to its occurrence. Scholars have underscored the need for robust methodologies and innovative techniques to navigate the intricacies of air quality management, emphasizing the importance of integrating diverse perspectives and expertise in addressing this global challenge [14, 15].

Within the realm of environmental research, there is a notable trend towards the increasing prominence of computational methods in understanding and addressing complex challenges. Machine learning techniques, in particular, emerge as a promising avenue in this endeavor. These techniques offer the potential to enhance the accuracy of air quality forecasts and identify key factors driving air pollution. The integration of machine learning and diverse data sources provides significant potential for forecasting air pollutant levels, with direct implications for informed decision-making and policy development. This aligns with broader efforts to harness technological advancements for more effective environmental management and protection [15, 16].

A critical examination of the existing body of work reveals notable gaps characterized by diverse machine learning methods, data sources, and geographic scopes. The current state of knowledge in air pollution estimation using machine learning reflects the complex nature of the phenomenon and the diverse contexts in which it occurs. Discerning the most effective machine learning techniques for air pollution prediction and understanding how factors like weather conditions and pollutant sources impact model accuracy become crucial aspects addressed by this systematic literature review. Additionally, the identification of limitations in current research is fundamental for guiding future studies and refining methodologies. This systematic review aims to provide a cohesive and critical analysis, offering insights that contribute to the advancement of knowledge in air pollution estimation and the application of machine learning in this domain.

1.2 *Research problem and objectives*

The central research problem addressed in this literature review is to identify and assess the most effective machine learning techniques for predicting high air pollution periods and to explore how these techniques can be applied to enhance air quality forecasts.

To address this problem, the following specific research objectives have been defined:

• To identify and categorize the machine learning techniques utilized in the literature for predicting high air pollution periods.

• To evaluate the accuracy and effectiveness of these machine learning techniques in air quality forecasting.

• To understand the impact of pollutant sources on the performance of machine learning models in predicting high air pollution events.

• To identify the limitations and gaps in the existing research and suggest directions for future studies to improve air pollution predictions using machine learning techniques.

1.3 *Structure of the paper*

This paper is organized into several sections to comprehensively address the research problem and objectives. The subsequent sections include a review of the related literature, categorizing machine learning techniques for air quality prediction and assessing their strengths and weaknesses. The literature review also discusses the impact of factors such as pollutant sources on model accuracy. Furthermore, it highlights the limitations in existing research and provides suggestions for future studies.

The paper concludes with a summary of the main findings and their implications for enhancing air quality management. The systematic nature of this literature review ensures a rigorous examination of the current state of research in machine learning applications for air pollution prediction, offering valuable insights for researchers, policymakers, and environmental practitioners.

2. Air pollution and its impact

Air pollution, a multifaceted environmental challenge, has been extensively explored through diverse studies focusing on distinct facets of its definition, types of pollutants, health and environmental repercussions, and the pivotal aspect of predicting high pollution periods. A comprehensive review of selected studies provides valuable insights into the complex nature of air pollution.

2.1 *Definition and categorization of air pollutants*

Air pollution is characterized by the presence of deleterious substances in the atmosphere. As posited by Bai et al. [3], air pollutants manifest in various forms, broadly categorized into particulate matter (PM), nitrogen oxides (NO_x), sulfur dioxide (SO_2), carbon monoxide (CO), ozone (O_3), and volatile organic compounds (VOCs). These pollutants emanate from diverse sources, including vehicular emissions, industrial processes, and natural origins (see Table 1).

Table 1. Air pollutant categories, specific pollutants, major sources, and relevant references

Air pollutant category	Specific pollutants	Sources	References
Particulate matter (PM)	PM2.5, PM10, Particle surface area (SA)	Vehicular emissions, industrial processes, natural origins	[17]
Nitrogen oxides (NO_x)	NO_2	Vehicular emissions, industrial processes, natural origins	[18]
Sulfur dioxide (SO_2)	SO_2	Vehicular emissions, industrial processes, natural origins	[18]
Carbon monoxide (CO)	CO	Vehicular emissions, industrial processes, natural origins	[18, 19]
Ozone (O_3)	O_3	Vehicular emissions, industrial processes, natural origins	[17]
Volatile organic compounds (VOCs)	-	Vehicular emissions, industrial processes, natural origins	[20]

Table 2. Health and environmental impacts in air pollution

Aspect	Study	Effect	Key findings
Health impacts	He et al. [10]	Respiratory and cardiovascular diseases	PM2.5 exposure linked to respiratory and cardiovascular diseases.
	Lee et al. [6]	Asthma-related hospital visits and admissions	NO_2 and CO significant contributors to asthma-related outpatient visits and hospital admissions.
	Zou et al. [21]	Public health impacts during wildfires	Machine learning models predict PM2.5 levels during wildfires, addressing public health impacts.
	Lee et al. [22]	Attribution of PM2.5 sources and seasonal variability	Attribution of PM2.5 sources in Ulsan, South Korea, providing a nuanced understanding of health impacts and seasonal variability.
	Zhang et al. [23]	Ozone pollution	Exploration of summer surface ozone pollution patterns in China, offering insights into the health impacts of ozone pollution and the need for nuanced strategies.
Environmental impacts	Yang et al. [24]	Ecosystems, climate change, and air quality	Proposal of environmental impacts, including effects on ecosystems, climate change, and overall air quality.
	Tuluri [5]	Dynamic nature of air pollution during lockdown	Changes in air quality during COVID-19 lockdown, demonstrating fluctuations in pollutant concentrations and highlighting air pollution's dynamic nature.
	Essamlali et al. [25]	Climate change, air pollution, and urban sustainability	Inter-correlation of climate change, air pollution, and urban sustainability. Machine learning and spatial information science used to underscore the importance of climate variables in air quality prediction.
	Ngom et al. [26]	High air pollution in large cities	Proposal of an approach combining system observations, multi-agent simulation, and machine learning to address high air pollution in large cities, emphasizing environmental impacts in densely populated urban areas.
	Mahalingam et al. [18]	Urban air quality management	Emphasis on the significance of air quality in urban management, introducing a machine learning model for predicting the air quality index (AQI).

2.2 *Health and environmental consequences of air pollution*

Air pollution has been extensively studied, with scholarly literature highlighting its deleterious effects on both human health and the environment. Various studies employ diverse methodologies, including machine learning models, to comprehend the multifaceted impacts of air pollution. This overview synthesizes key findings from recent research, providing insights into the health and environmental consequences associated with air pollution as displayed in Table 2.

2.3 *Importance of predicting high pollution periods*

In light of the detrimental impacts associated with air pollution, the anticipation of elevated pollution periods assumes critical importance in the realms of public health and environmental stewardship. Razavi-Termeh et al. [27] posit that precise forecasting enables the timely implementation of pollution control measures. The methodology proposed by Hu et al. [28], known as WRFC-XGB, for gauging ozone concentrations, underscores the potential advantages of predictive models in evaluating air quality.

Furthermore, Gocheva-Ilieva et al. [15] emphasize the importance of forecasting by amalgamating independent ex-ante data sources with machine learning techniques. Their investigation in Pernik, Bulgaria, illustrates that an averaging strategy for forecasts outperforms singular models, suggesting the potential of advancements in air pollution forecasting.

Son et al. [4] propose a deep learning-based method to estimate surface PM2.5 concentrations utilizing data from the TROPOMI instrument. The study successfully captures spatiotemporal variations, underscoring the significance of predicting elevated pollution periods, particularly during events such as fires that exert a substantial impact on air pollution. The research highlights the role of carbon monoxide and the repercussions of fires on air pollution, contributing to a comprehensive understanding of high pollution events.

Gu et al. [29] introduce a hybrid machine learning model for predicting particulate matter 2.5 (PM2.5), emphasizing enhanced prediction accuracy, especially during peak values. The study underscores the importance of predicting elevated pollution periods for specific air pollutant types, offering insights into factors contributing to peak pollution events. Predicting high pollution times is vital for public health and the environment. The studies highlight the importance of precise forecasting, especially during events like fires. Improved prediction methods play a crucial role in implementing effective pollution control measures. The collective findings stress the need for ongoing efforts to enhance our ability to anticipate and address elevated pollution periods.

2.4 *Machine learning in air pollution estimation*

This review delves into the current landscape of machine learning applications in the estimation of air pollution, synthesizing insights from a range of recent studies. The exploration is organized into three main sections: an overview of machine learning, applications of machine learning in environmental science, and the specific relevance of machine learning in predicting high air pollution periods.

2.5 *Overview of machine learning*

Machine learning has emerged as a powerful tool in environmental science, offering novel approaches for understanding and predicting air pollution. The works of Bai et al. [3] and He et al. [10], among others, contribute to a comprehensive understanding of the diverse machine learning techniques employed in this domain, as shown in Table 3.

2.6 *Applications of machine learning in environmental science*

In the realm of environmental science, machine learning has become a valuable tool for understanding and tackling air pollution challenges. This exploration dives into a set of studies from various global locations, showcasing the diverse applications of machine learning in the context of air quality research (see Table 3).

Table 3. Overview of machine learning applications in air quality research

Study	ML application	Key findings	Practical implication	Geographical focus
[3]	Air pollution forecasting	Comprehensive overview of forecasting methods; lacks specific ML recommendations	Enhancing understanding of diverse forecasting approaches	-
[10]	Daily adaptive modeling framework for PM2.5 pollution	High accuracy in mapping PM2.5; limited geographical focus on China	Improved mapping and assessment of PM2.5 pollution	China
[27]	Ensemble ML algorithms for asthma-prone areas	AdaBoost superiority in spatial modeling; restricted focus on Tehran	Identifying influential factors in asthma-prone areas	Tehran
[28]	WRFC-XGB method for ozone concentrations	High accuracy in Beijing-Tianjin-Hebei region; limited regional applicability	Accurate estimation of near-surface ozone concentrations	Beijing-Tianjin-Hebei region
[24]	Analysis of atmospheric pollution in Shanghai	O_3 identified as primary pollutant; exclusive focus on Shanghai	Coordinating control measures for atmospheric pollution	Shanghai
[30]	Two-dimensional visibility estimation model	Innovative random forest (RF) and geostatistics combination; emphasizes input variables' significance	Enhancing visibility estimation using RF and geostatistics	-
[5]	Impact of COVID-19 lockdown on air quality	Effective use of air quality and weather data; incomplete content with missing information	Assessing the impact of mobility restrictions on air quality	Mississippi, USA
[15]	Novel approach for forecasting air pollutants	RF and Arc-x4 combination; averaging strategy outperforms single models	Advancements in air pollution forecasting using novel approach	Pernik, Bulgaria
[20]	PM2.5 concentration prediction model	Convolutional neural network (CNN)-RF ensemble for accurate predictions; lack of in-depth exploration of potential limitations	Stable and accurate PM2.5 concentration prediction model	Kaohsiung, Taiwan
[6]	Association between air pollutants and hospital visits	NO_2 and CO identified as significant pollutants; time lags highlighted	Identifying significant pollutants affecting hospital visits	Seoul, South Korea
[17]	Estimating reduction in PM2.5 during COVID-19 lockdown	RF, support vector regression (SVR), and artificial neural network (ANN) used for PM2.5 prediction; significant decrease during lockdown	Demonstrating the potential of ML for air pollution exposure assessment	Yangtze River Delta during the COVID-19 pandemic
[31]	Ensemble model for daily PM2.5 prediction	Effective in predicting day-to-day changes; spatial variability predictions less accurate	Creating pollutant proxies with prediction uncertainty assessments	Greater London
[32]	Developing air pollutant proxies using ML	Bayesian neural network for creating pollutant proxies; lacks explicit disadvantages	Creating pollutant proxies with prediction uncertainty assessments	-
[21]	Wildfire smoke and PM2.5 prediction	Public health impacts of wildfire smoke; effectiveness of random forest model; significant health impacts emphasized	Utility of analytical tools for estimating air quality changes during natural hazards	-
[11]	Forecasting air pollution in Taiwan	Superiority of ML models over traditional methods for PM2.5 concentration forecasting; valuable insights into air pollution prediction	Improved air pollution prediction using ML models	Taiwan
[33]	ML for predicting traffic-related air pollution in Toronto	Comparison of ML models (ANN, gradient boost) with LUR approaches; ML models outperform LUR; potential of ML in air quality prediction	Potential of ML in air quality prediction	Toronto, Canada

Table 3. (cont.)

Study	ML application	Key findings	Practical implication	Geographical focus
[34]	Physics-informed multi-task deep neural networks (DNN) for satellite-based air pollutant retrieval	Simultaneous estimation of multiple air pollutants; improved efficiency and accuracy compared to single-pollutant methods	Improved modeling efficiency and accuracy using multi-task DNN	Mainland China (2019)
[4]	Deep learning for surface PM2.5 estimation in Thailand	Estimation of surface PM2.5 using TROPOMI data; outperformance of other ML models; influence of carbon monoxide and fires on air pollution highlighted	Better estimation of high PM2.5 concentrations using deep learning	Thailand (2018-2020)
[29]	Hybrid ML model for PM2.5 prediction	Improved prediction accuracy, especially for peak values; emphasis on interpretability; outperformance of other models	Improved accuracy and interpretability in PM2.5 prediction	-
[35]	ML for predicting PM2.5 concentrations in China	Use of RF models to predict PM2.5 concentrations; identification of significant contributing factors to air pollution events	Insights into factors contributing to air pollution events	China
[7]	Spatial-temporal changes in air pollution in the Pearl River Delta	Analysis of air pollutants and meteorological factors; identification of seasonal variations and areas susceptible to meteorological factors	Insights into urban air pollution control based on spatial-temporal analysis	Pearl River Delta, China (2006-2019)
[23]	Co-pollution of O_3 and PM2.5 in Hainan Province	Machine learning to explain NO_x role in co-pollution; shift in O_3 formation due to NO_x concentrations; impact of local industrial activities highlighted	Reduction of NO_x emissions could help control O_3 and PM2.5 pollution	Hainan Province, China (wintertime)
[22]	Sources of high PM2.5 in Ulsan, South Korea	Identification of sources during winter and summer; influence of local emissions, long-range transport, and secondary formation emphasized	Insights for improving PM2.5 management policies based on source attribution	Ulsan, South Korea
[36]	GD-GRU for repairing environmental data in Guilin	Use of GD-GRU for data repair; higher accuracy in predicting extreme values; better prediction of PM2.5 concentrations compared to other models	Improved accuracy in predicting extreme PM2.5 values using GD-GRU	Guilin, China
[37]	Improvement in air quality in Beijing	Analysis of ABC decline in Beijing from 2015 to 2020; attribution to local emission reductions and regional transport reduction	Emphasis on regional collaboration for controlling air pollution	Beijing, China (2015-2020)
[38]	Daily NO_2 concentrations in Chinese urban agglomerations	LME model using TROPOMI products for NO_2 estimation; outperformance of other models; identification of pollution hotspots and anomalies during COVID-19 lockdown	Identification of pollution hotspots and anomalies during COVID-19	Chinese urban agglomerations (2018-2020)
[39]	Non-linear relationship among O_3, SOA, and VOCs in Tianjin	ML to identify impact of specific VOCs on O_3 and SOA formation; analysis of photochemical consumption of VOCs	Insights into O_3 and SOA formation based on specific VOCs impact	Tianjin, China
[12]	Impact of surface water bodies on PM2.5 in Delhi-NCT	ML models predict PM2.5 levels; significant effect of water bodies within 3 km in reducing PM2.5 concentration	Actionable measures for air quality enhancement	Delhi-NCT, India
[40]	Air quality analysis in Hanoi during winter	Classification of winter days based on local winds and AOD values; nuanced understanding of cold surges' impact on PM2.5 levels	Nuanced understanding of cold surges' impact on PM2.5 levels	Hanoi, Vietnam

Table 3. (cont.)

Study	ML application	Key findings	Practical implication	Geographical focus
[41]	Air pollution in Beijing during COVID-19 and Olympics	Identification of non-linear effects; need for collaborative emission reduction strategies	Insights into complex interactions of meteorological conditions and emission reductions on PM2.5 concentration	Beijing, China
[42]	Long-term particle surface area concentrations in China	Insights into spatial heterogeneity and variations in SA concentrations; focus on heavy pollution periods	Valuable insights into spatial heterogeneity and variations in SA concentrations	China
[1]	Evolution of summer surface ozone pollution patterns	Identification of dominant patterns and their associations in China	Insights into ozone pollution patterns in China	China
[43]	Pollution characteristics and source origins in Xinxiang	Detailed analysis of seasonal and diurnal variations; valuable information for designing strict control measures	Valuable information for designing strict control measures	Xinxiang, North China
[44]	Inter-correlation of climate change, air pollution, and urban sustainability	Importance of climate variables in air quality prediction; promising for early warning mechanisms	Importance of climate variables in air quality prediction	-
[45]	Air quality prediction for urban environments	Accuracy of hybrid models in predicting air quality index (AQI); significant advantage in model performance	Accuracy of hybrid models in predicting AQI	-
[46]	Air pollution levels in Sakarya, Turkey	Performance of deep learning models (long short-term memory (LSTM) and bidirectional long short-term memory (Bi-LSTM)) in estimating daily CO concentration	Reliable results for air pollution modeling using deep learning models	Sakarya, Turkey
[47]	Air pollution estimation and regional characteristics	Deep learning models for region classification and AQI estimation; valuable tools for accurate air quality assessment	Accurate air quality assessment using deep learning models	-
[48]	Low-cost sensor devices for monitoring particulate matter	Integration of monitoring technologies and role of ML in device calibration; reliable screening of particulate matter	Reliable screening of particulate matter using low-cost sensors	-
[49]	MLP neural network to estimate aerosol acidity for PM2.5	Effectiveness of MLP in estimating aerosol acidity with high accuracy	Insights into aerosol acidity estimation using MLP	-
[50]	Low-cost portable sensors for air quality monitoring	Advantages and challenges of crowd sensing and low-cost sensors; importance of local pollutant concentration and mobility data	Emphasizes the usefulness of portable monitors	Urban and indoor environments
[51]	Time-series analysis of SO_2 in Tehran, Iran	Box-Jenkins autoregressive integrated moving average (ARIMA) approach provides valuable insights into SO_2 air pollution factors	Valuable insights into SO_2 air pollution factors	Tehran, Iran
[52]	Air pollution estimation in Chennai	K-nearest neighbor method and ML algorithms effectiveness; emphasis on deep learning methods	Emphasis on the effectiveness of deep learning methods	Chennai, India
[53]	Predicting AQI in Delhi	Comparison of various ML techniques for AQI prediction; well-suited methods identified	Identifying well-suited methods for predicting air quality	Delhi, India
[54]	Metropolitan air pollution estimation in Sydney	HazeEst machine learning model; application of support vector regression (SVR); high spatial resolution estimates	High spatial resolution estimates using HazeEst model	Sydney, Australia

Table 3. (cont.)

Study	ML application	Key findings	Practical implication	Geographical focus
[55]	Industrial pollution impact on human health	Prediction of emission rates using various ML algorithms; identification of Multi-layer perceptron model as least error-prone	Identification of the least error-prone ML model	-
[56]	SVR-based model for air pollution in Sydney	SVR model utilizing historical data and wireless sensor network; more accurate air pollution estimations than ANN	Potential benefits for researching air pollution-related diseases	Sydney, Australia
[57]	Predicting ground-level Ozone levels in Gauteng province	Models based on big data analytics and cognitive computing; computationally efficient and insightful without requiring massive computational power	Computational efficiency without massive computational power	Gauteng province, South Africa
[58]	PM2.5 pollution in China using internet of things (IoT) sensors	Establishment of a fine particulate matter network (FPMN) using IoT sensors; reconstruction of reliable regional field of PM2.5 concentration	Reconstruction of a reliable regional field of PM2.5 concentration	China
[59]	Urban big data for identifying causal pathways	Introduction of "pg-Causality", a pattern-aided graphical causality analysis approach; efficient identification of causal pathways	Efficient identification of causal pathways using "pg-Causality"	-
[8]	Deep-AIR model for air pollution estimation	Introduction of "Deep-AIR" machine learning model with hybrid CNN-LSTM structure; outperforms baseline models in estimating and forecasting various pollutants	Identification of key predictors and outperformance of baseline models	Metropolitan areas
[60]	Supervised ML for predicting AQI in densely populated areas	Comparison of six ML algorithms for predicting AQI; highlighting strengths and weaknesses of each algorithm	Determining the most accurate model for predicting AQI	Densely populated areas
[18]	Neural networks and SVM for predicting AQI in urban areas	Introduction of ML model using neural networks and support vector machines for predicting AQI; improved prediction accuracy for Delhi AQI and other smart cities	Improved prediction accuracy for urban AQI	Delhi and other smart cities
[26]	Optimizing pollution monitoring networks in large cities	Combination of system observations, multi-agent simulation, and ML for assimilating PM10 pollution data; optimization of pollution monitoring networks	Optimization of pollution monitoring networks in large cities	Dakar, Senegal
[61]	Spatiotemporal distribution of PM2.5 in Shaanxi, China	Combination of three-dimensional landscape pattern index (TDLPI) with eXtreme gradient boosting (XGBoost) and ML for optimizing LUR modeling; accurate estimation of PM2.5 pollution	Accurate estimation of PM2.5 pollution using the "LTX" approach	Shaanxi, China

2.7 *The relevance of machine learning in predicting high air pollution periods*

Predicting high air pollution periods is a critical aspect of air quality management, and machine learning techniques offer unique advantages in this regard. The studies of Gocheva-Ilieva et al. [15] and Abubakar et al. [62] exemplify the potential of machine learning in forecasting such as air pollutant levels with high accuracy. Gocheva-Ilieva et al. [15] emphasize the efficacy of combining independent ex-ante data sources with machine learning techniques, showcasing advancements in air pollution forecasting. Chen et al. [20] present a novel PM2.5 concentration prediction model that integrates convolutional neural network (CNN) feature extraction with random forest (RF) regression, demonstrating stability and accuracy in predicting day-to-day variations in PM2.5 levels and many other relevant studies as displayed in Table 3.

Machine learning has become an indispensable tool in air pollution estimation, offering innovative solutions to the challenges posed by the dynamic and intricate nature of environmental data. The reviewed studies collectively showcase the diverse applications and strengths of machine learning in predicting high air pollution periods (see Table 3). However, the existing literature highlights the need for a more comprehensive discussion of limitations and opportunities for future research to ensure the robustness and generalizability of machine learning models in the context of air pollution estimation.

3. Methodology

In this section, we outline the systematic methodology used to identify, select, and analyze the relevant literature in the field of estimating high air pollution periods through machine learning approaches. We have adopted the PRISMA framework, a widely recognized guideline for conducting systematic reviews, to ensure the rigor and comprehensiveness of our study.

3.1 *Research questions and their descriptions*

This section introduced the research questions and their descriptions in Table 4 below.

Table 4. Research questions and their corresponding descriptions

Research question	Description
RQ1: What are the most effective machine learning techniques for predicting high air pollution periods, and how can they be applied to enhance air quality forecasts?	This question seeks to identify the most suitable machine learning techniques for predicting high pollution events. It focuses on exploring advanced forecasting methods, enhancing model accuracy and reliability, and providing insights into optimal implementation practices.
RQ2: How do specific factors, such as weather conditions and pollutant sources, impact the accuracy of machine learning models in predicting high air pollution events?	This question examines how meteorological variables and emission sources influence the performance of predictive models. It aims to understand the interactions between environmental factors and pollutants and how these affect model precision.
RQ3: What are the limitations in the current research, and how can future studies address these gaps to improve air pollution predictions?	This question highlights the existing shortcomings in machine learning-based air pollution studies. It identifies methodological gaps and proposes directions for future research to strengthen the accuracy and applicability of predictive approaches.

3.2 *Inclusion and exclusion criteria*

To ensure the relevance and consistency of the reviewed literature, specific inclusion and exclusion criteria were defined according to the PRISMA framework. These criteria guided the selection of studies for inclusion in this review. The following Table 5 outlines the criteria used for the selection of studies.

Table 5. Inclusion and exclusion criteria

Criteria	Inclusion criteria	Exclusion criteria
Publication source	Peer-reviewed journals or conference proceedings	Studies not published in peer-reviewed sources
Research focus	Estimation of high air pollution periods using machine learning techniques	Studies not related to air pollution estimation or machine learning
Methodology detail	Clear description of machine learning algorithms used	Studies lacking sufficient information on methodology
Language	Written in English for accessibility and comprehension	Studies in languages other than English
Time frame	Publication within 2016 to 2023	Publication outside 2016 to 2023

3.3 *Search and selection process*

The following steps were taken, following the PRISMA guidelines, to identify and select relevant literature.

3.3.1 *Database search*

A comprehensive search was conducted on reputable academic databases and search engines, including Springer Link, IEEE Xplore, Elsevier, and MDPI as displayed in Figure 1. The search string used included "air pollution" or "polluted air" or "air quality" or "environmental pollution" and "Machine Learning" or "Artificial Intelligence" and "Estimation" or "Prediction".

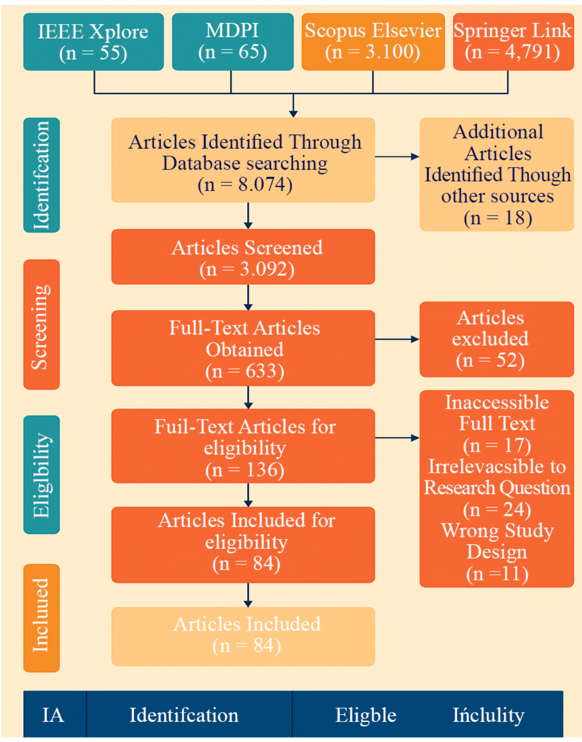

Figure 1. Search databases used for the literature review

4. Literature review

4.1 *Machine learning techniques for air quality prediction*

Example 1: Artificial neural networks (ANNs) are widely used for air quality prediction [17, 33]. Lu et al. focused on estimating the reduction in PM2.5 concentration in the Yangtze River Delta during the COVID-19 pandemic lockdown, utilizing ANN and comparing it with random forest (RF) and support vector regression (SVR) [17]. In contrast, Wang et al. [33] explored machine learning for predicting traffic-related air pollution, comparing ANN and gradient boosting models with traditional land use regression (LUR) approaches. Both studies found that ANN is highly effective in predicting air quality, with robust performance in various contexts.

Example 2: Most air quality prediction studies focus on PM2.5 concentrations due to its significant health impact [4, 10]. Several studies, such as those conducted by A, B, and C, demonstrate that PM2.5 serves as a critical predictor for air quality, especially in urban settings where its levels fluctuate significantly due to industrial, vehicular, and meteorological factors.

The study by Zhang [8] contributes notably to the field of air quality prediction by introducing the "Deep-AIR" model, which combines domain-specific features and a hybrid CNN-LSTM structure. This model outperforms baseline models, emphasizing the potential of advanced machine learning techniques for more accurate and fine-grained air pollution estimation [8].

He et al. [10] proposed a two-stage daily adaptive modeling framework tailored for PM2.5 pollution mapping, which demonstrated high accuracy. However, the study's geographical focus on China raises questions about the broader applicability of the framework, suggesting that further research could explore its generalizability to other regions. Additionally, the paper could benefit from a more thorough discussion of potential limitations and avenues for future improvements.

Wang et al. [39] investigated the complex non-linear relationships between ozone (O_3), secondary organic aerosol (SOA), and volatile organic compounds (VOCs) in Tianjin, China, using machine learning models to pinpoint the impact of specific VOCs on O_3 and SOA levels.

Bamrah et al. [47] focused on regional air pollution estimation and AQI prediction, emphasizing the importance of incorporating local environmental factors. Their use of deep learning models for region classification and AQI estimation showcases the utility of advanced machine learning in improving regional air quality assessments.

Chen et al. [20] presented a novel approach for PM2.5 concentration prediction by integrating convolutional neural network (CNN) feature extraction with random forest (RF) regression. Their model demonstrated both stability and accuracy, outperforming conventional methods in predicting air pollution levels.

Razavi-Termeh et al. [27] used ensemble machine learning algorithms to identify asthma-prone areas and influential factors, concluding that the AdaBoost algorithm was particularly effective in spatial modeling.

Lee et al. [6] investigated the relationship between air pollutants and hospital visits by asthma patients in Seoul, South Korea, applying both statistical and machine learning techniques to analyze healthcare data and air quality levels.

Dong et al. [38] utilized satellite-based data to estimate daily NO_2 concentrations in Chinese urban agglomerations, introducing a linear mixed-effects model (LME) that surpassed other models in predictive accuracy.

Zaidan et al. [32] employed mutual information and a Bayesian neural network to develop air pollutant proxies, specifically for ozone concentration estimation, highlighting the model's efficacy in atmospheric monitoring.

Doreswamy et al. [11] applied machine learning regression models to predict PM2.5 concentrations using Taiwan Air Quality Monitoring data, providing valuable insights into the use of machine learning in air pollution forecasting.

Wang et al. [33] focused on predicting traffic-related air pollution by comparing machine learning models like ANN and gradient boosting with traditional land use regression (LUR), emphasizing the potential of machine learning to refine air quality prediction in urban areas.

Yang et al. [24] introduced a multi-task deep neural network that is physics-informed for satellite-based retrieval of multiple air pollutants. This approach demonstrated enhanced modeling efficiency and accuracy, marking a significant step forward in air pollution modeling Son et al. [4] proposed a deep learning-based approach for estimating surface

PM2.5 concentrations by incorporating atmospheric gas species, demonstrating superior performance compared to other machine learning models.

Caquilpán et al. [48] explored low-cost sensor devices for particulate matter monitoring, focusing on integrating machine learning techniques to improve the calibration of these sensors for more accurate pollution measurements.

Tao et al. [49] utilized a multilayer perceptron (MLP) neural network to estimate aerosol acidity for fine particles (PM2.5), showcasing the model's effectiveness in accurately predicting aerosol properties and their relationship with air quality.

These studies highlight the growing reliance on machine learning techniques, particularly deep learning and ensemble methods, to address the challenges in air quality prediction, especially for PM2.5 concentration estimation. As the field continues to evolve, further exploration of hybrid models and the incorporation of region-specific factors will likely enhance prediction accuracy across diverse environmental settings.

4.2 *Application of machine learning in air quality forecasting*

The integration of machine learning (ML) techniques into air quality forecasting has garnered significant attention in recent years. Various studies have explored and developed models to enhance the accuracy and granularity of air pollution predictions.

Bai et al. [3] provided a comprehensive overview of air pollution forecasting methods, categorizing them into classical approaches. However, the study lacks specificity in its recommendations and comparisons of machine learning techniques, which limits its practical utility for researchers and policymakers.

Zhang et al. [8] introduced "Deep-AIR", a hybrid framework combining convolutional neural networks (CNN) and long short-term memory (LSTM) networks. This model captures both spatial and temporal features of air pollution, achieving higher accuracy in fine-grained hourly estimations and forecasts in metropolitan cities like Hong Kong and Beijing. The incorporation of urban dynamics, such as road density and street canyon effects, into the model underscores the importance of contextual urban factors in air quality predictions.

Gocheva-Ilieva et al. [15] proposed a novel framework for forecasting air pollutant levels, integrating independent ex-ante data sources with machine learning models. Their use of random forest (RF) and Arc-x4 models to predict PM10, SO_2, and NO_2 in Pernik, Bulgaria, shows promise. Nonetheless, the study would benefit from a deeper analysis of limitations and future research directions.

Choi et al. [30] developed an innovative two-dimensional visibility estimation model by integrating random forest (RF) with geostatistical methods. This approach not only enhances the accuracy of visibility predictions but also demonstrates the potential of combining machine learning with geostatistics in environmental modeling.

Tuluri [5] examined the impact of mobility restrictions during the COVID-19 lockdown on air quality in Mississippi, USA. By analyzing air quality and weather data, the study provides insights into how reduced human activity influences pollutant concentrations, offering valuable information for future urban planning and public health policies.

Gayen et al. [12] investigated the effect of surface water bodies on PM2.5 concentrations in Delhi-NCT, India. Utilizing machine learning models, they found that proximity to water bodies significantly influences pollutant dispersion, highlighting the need to consider geographical features in air quality assessments.

Yazdi et al. [31] developed an ensemble model to predict daily PM2.5 levels in Greater London. Their approach combines multiple machine learning algorithms to enhance prediction accuracy, demonstrating the effectiveness of ensemble methods in air quality forecasting.

Srivastava et al. [53] compared various machine learning techniques to predict the air quality index (AQI) in Delhi. Their assessment provides a benchmark for selecting appropriate models based on performance metrics, aiding in the development of more reliable forecasting systems.

Hernández-Gordillo et al. [50] conducted a systematic review on the use of low-cost portable sensors for urban and indoor air quality monitoring. Emphasizing crowd sensing, their work underscores the potential of integrating community-driven data collection with machine learning for comprehensive air quality assessments.

Nelgadevi and Grasha [52] addressed air pollution estimation in Chennai by applying the K-Nearest Neighbor method alongside other machine learning algorithms. Their study contributes to the growing body of research focusing on localized air quality prediction models.

Nuraeni et al. [54] introduced the "HazeEst" model, which combines data from fixed-station monitors and mobile sensors to estimate metropolitan air pollution in Sydney. This integration of diverse data sources enhances the spatial resolution of pollution estimates.

Simu et al. [55] explored industrial pollution and its impact on human health by employing various machine learning algorithms to predict emission rates. Their work highlights the applicability of ML in assessing industrial contributions to air quality degradation.

Hu et al. [56] developed a support vector regression (SVR) model for estimating air pollution in Sydney with fine spatial granularity. This approach emphasizes the importance of spatial detail in pollution modeling.

Chiwewe and Ditsela [57] focused on predicting ground-level ozone levels in Gauteng province, South Africa, using cross-correlation and spatial-correlation of air pollutants. Their study illustrates the utility of statistical correlations in enhancing prediction models.

Li et al. [58] addressed the issue of fine particulate matter (PM2.5) pollution in China by establishing a fine particulate matter network (FPMN) using IoT sensors. This network facilitates real-time monitoring and data collection, crucial for timely interventions.

Zhu et al. [59] discussed the use of urban big data, air quality, and meteorological data to identify spatiotemporal causal pathways for air pollutants. Their approach leverages large datasets to uncover underlying patterns and relationships in pollution dynamics.

Yarragunta et al. [60] explored the use of supervised machine learning algorithms to predict the air quality index (AQI) in densely populated areas. Their findings contribute to the development of predictive models tailored for urban environments.

Mahalingam et al. [18] emphasized the significance of air quality in urban management and its impact on human health and sustainable development. They introduced a machine learning model for predicting AQI, highlighting the role of technology in environmental management.

Ngom et al. [26] addressed high air pollution in large cities such as Dakar by proposing an approach that combines system observations with multi-agent simulation and machine learning to assimilate PM10 pollution data. This multidisciplinary method offers a comprehensive framework for pollution analysis.

Zhang et al. [61] focused on estimating the spatiotemporal distribution of PM2.5 concentration in Shaanxi, China. By combining a three-dimensional landscape pattern index (TDLPI) with machine learning (XGBOOST), they optimized land use regression (LUR) modeling, demonstrating the integration of landscape metrics in pollution prediction.

These studies collectively underscore the evolving role of machine learning in air quality forecasting, highlighting the importance of integrating diverse data sources, considering urban dynamics, and employing advanced algorithms to enhance predictive accuracy.

4.3 *Impact of weather conditions on air quality prediction*

Weather conditions play a critical role in the dispersion, formation, and accumulation of air pollutants. Several studies have underscored the importance of meteorological influences on air quality modeling and prediction.

Ngoc et al. [40] analyzed air quality in Hanoi, Vietnam, during the winter season, with a particular focus on local meteorological factors. The study investigated the sources of air pollution and used wind regimes and aerosol optical depth (AOD) values to classify winter days. Their findings emphasized the dominant influence of local winds on the spatial distribution of pollutants and the effectiveness of incorporating AOD in air quality assessment models.

Hu et al. [28] proposed a novel hybrid method, WRFC-XGB, which integrates the Weather Research and Forecasting model with the eXtreme gradient boosting (XGBoost) algorithm to estimate near-surface ozone concentrations in the Beijing-Tianjin-Hebei region. Their approach achieved high predictive accuracy by effectively capturing the nonlinear interactions between meteorological variables and ozone formation.

Yang et al. [24] conducted an extensive evaluation of atmospheric pollution in Shanghai, identifying ozone (O_3) as the primary pollutant of concern. The study also highlighted the difficulty of implementing coordinated control measures due to the complex interplay of emission sources and meteorological conditions. Their results underscore the need for comprehensive policy integration across sectors to address the transboundary nature of ozone pollution.

Zhu et al. [37] documented a significant improvement in Beijing's air quality between 2015 and 2020, attributing these changes to both local emission reduction efforts and regional transport control strategies. Their temporal analysis revealed that meteorological variability alone could not account for the observed reductions, thereby emphasizing the effectiveness of coordinated environmental policies.

Tuluri [5] investigated the effect of mobility restrictions during the COVID-19 pandemic on air pollutant concentrations in Mississippi, USA. By incorporating meteorological data, the study provided nuanced insights into how reduced human activities-along with prevailing weather conditions-contributed to temporary improvements in air quality. This work highlights the potential of incorporating behavioral and meteorological data into predictive frameworks.

4.4 *Effect of pollutant sources on model accuracy*

The accuracy of air quality prediction models is closely tied to the understanding and integration of pollutant source information. Several recent studies have examined how source characteristics influence model performance.

Gocheva-Ilieva et al. [15] made a significant contribution by combining independent ex-ante data sources with machine learning techniques to forecast pollutant levels in Pernik, Bulgaria. Beyond methodological advancements, the study emphasized the need to account for specific pollutant sources (e.g., industrial emissions, vehicular traffic) to enhance model robustness. Their findings suggest that overlooking source-specific data may result in reduced model accuracy and misinformed policy decisions.

Razavi-Termeh et al. [27] used ensemble learning techniques to identify asthma-prone zones in Tehran. Their results showed the superior performance of the AdaBoost algorithm in spatial modeling. However, the study's geographical limitation raises concerns about the generalizability of its findings. Future research could expand the scope and incorporate more diverse environmental and demographic data to improve applicability across regions.

Zou et al. [21] examined the health impacts of wildfire smoke using machine learning models, with a particular emphasis on the Random Forest algorithm. Their work provided compelling evidence of the association between wildfire events and respiratory health outcomes, and demonstrated the potential of machine learning to forecast the public health risks of pollution events arising from natural sources.

Gu et al. [29] developed a hybrid deep learning model by integrating a deep neural network (DNN) with a nonlinear auto-regressive moving average model with exogenous inputs (NARMAX) for predicting PM2.5 concentrations. This architecture effectively captured both short-term dependencies and nonlinear relationships inherent in air quality data, showcasing the synergy of classical time-series methods with modern machine learning.

Li et al. [35] applied the random forest algorithm to model PM2.5 levels in China and identify dominant contributing factors. Their feature importance analysis revealed meteorological parameters, such as temperature and humidity, as key influencers alongside anthropogenic emissions, reinforcing the need for integrated data modeling.

Xu et al. [36] introduced a deep learning method known as GD-GRU for imputing and predicting PM2.5 concentrations. The model demonstrated superior performance in recovering extreme values, a challenging aspect in air quality modeling. This contributes to improving the reliability of datasets affected by missing or inconsistent entries.

Zhang et al. [8] proposed an innovative approach for modeling the spatiotemporal distribution of PM2.5 by integrating a three-dimensional landscape pattern index (TDLPI) with XGBoost. Their enhanced land use regression (LUR) model in Shaanxi, China, incorporated both topographical and land-use factors, significantly improving the model's explanatory power and accuracy.

These literature reviewed have shown considerable advancements in machine learning-based air quality prediction. Pioneering models such as Deep-AIR, WRFC-XGB, GD-GRU, and hybrid NARMAX-DNN frameworks have enhanced the precision of pollution forecasts. These efforts underscore the growing integration of deep learning, ensemble algorithms, and hybrid architectures in environmental modeling.

Nevertheless, a notable gap remains in the generalizability of findings, as many studies focus on specific regions-particularly China and Southeast Asia-limiting broader applicability. The inclusion of diverse data sources, such as low-cost sensors, geospatial features, and meteorological parameters, marks an interdisciplinary trend that improves prediction quality. Additionally, incorporating pollutant source characteristics has been shown to enhance model robustness.

To further elevate the field, future research should aim to address model limitations through more comprehensive error analyses, enhance transferability across regions, and consider real-world deployment scenarios. Continued refinement of hybrid and ensemble models, informed by rich contextual and source-specific data, holds great promise for advancing air quality forecasting and contributing to public health and policy planning.

5. Result analysis

Research Question 1 (RQ1): What are the most effective machine learning techniques for predicting high air pollution periods, and how can they be applied to enhance air quality forecasts.

Table 6. Summary of machine learning techniques and their applications

Approach/Category	Studies	Algorithms used	Application
Classical categories overview	[3]	-	Categorization of air pollution forecasting methods into classical categories.
Deep learning	[8, 36]	Deep-AIR, GD-GRU	Fine-grained forecasting; repairing atmospheric environmental data.
Ensemble models	[27, 31, 53]	AdaBoost, Ensemble methods	AQI prediction, daily PM2.5 levels, asthma-prone area modeling.
Random forest/Decision trees	[15, 30, 35]	RF, Arc-x4, Geostatistics	PM10, SO_2, NO_2 forecasting; visibility estimation; wildfire impact.
Support vector regression (SVR)	[56]	SVR	Fine-granular pollution estimation.
XGBoost	[8]	XGBoost, TDLPI	Estimation of spatiotemporal PM2.5 distribution.
K-nearest neighbor (KNN)	[54]	KNN	Estimating pollution in metropolitan areas.
IoT sensors	[58]	IoT-based models	IoT for fine particulate matter network (FPMN).
Multi-agent simulation	[26]	Multi-agent models	Urban air quality monitoring and simulation.
Review on monitoring techniques	[50]	-	Systematic review of portable low-cost sensors.
Impact analysis during COVID-19	[5]	-	Air quality changes due to lockdown restrictions.
Winter period analysis	[40]	-	Seasonal source and weather influence analysis.
Ozone concentration estimation	[28]	WRFC-XGB	Estimating near-surface ozone concentrations.
Comprehensive analysis of atmospheric pollution	[24]	-	Identified O_3 as dominant pollutant in Shanghai.
Improvement in air quality	[24]	-	Beijing air quality improvements (2015-2020).
Source consideration in forecasting	[15]	-	Role of source analysis in forecasting.
Asthma-prone areas modeling	[27]	AdaBoost	Identifying influencing factors in asthma-prone zones.
Wildfire smoke impact assessment	[21]	Random forest	Public health impact of wildfire smoke.
Hybrid deep learning model	[29]	DNN, NARX	Hybrid model for PM2.5 prediction.

Note: RF = random forest; DNN = deep neural network; NARX = nonlinear auto-regressive with exogenous input; TDLPI = three-dimensional landscape pattern index

Table 6 encapsulates the breadth of research in the field of air pollution forecasting, showcasing the multitude of machine learning approaches and algorithms. The varied techniques applied, from deep learning and ensemble models to spatial regression and hybrid methodologies, reflect the complexity of the task at hand. The delineation of applications further emphasizes the relevance of these studies in addressing real-world challenges, spanning from urban air quality management to the impact of mobility restrictions and the integration of low-cost sensors.

5.1 *Application of machine learning in air quality forecasting*

Based on the data provided in Table 6, several machine learning techniques are commonly applied in air quality forecasting. The most frequently observed techniques are displayed in Figure 2.

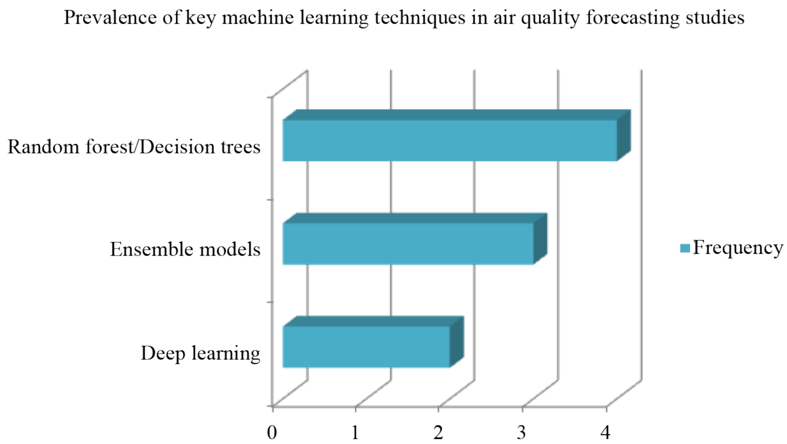

Figure 2. Common machine learning techniques in air quality forecasting

Research Question 2 (RQ2): How do pollutant sources impact the accuracy of machine learning models in predicting high air pollution events?

This investigation delves into the influence of pollutant sources on machine learning models predicting high air pollution events. Focused on various sources, Table 7 distills key insights, shedding light on the nuanced relationship between pollutant sources and model accuracy. This analysis contributes valuable insights to the broader understanding of air pollution prediction dynamics.

Table 7. Impact of pollutant sources on model accuracy in air pollution prediction studies

Study	Model/Method	Key variables studied	Pollutant source	Source impact on model accuracy
[10]	Two-stage daily adaptive	PM2.5 pollution	Air pollution control policies	Achieves high accuracy; limited to China; requires broader validation and discussion of limitations.
[27]	Ensemble machine learning	Asthma-prone areas, influential factors	Local sources	AdaBoost shows strong spatial modeling; limited applicability beyond Tehran.
[24]	Physics-informed multi-task learning	PM2.5, PM10, SO_2, NO_2, CO, O_3	Multiple sources	Efficient estimation of six pollutants; more accurate than single-pollutant models.
[4]	Deep learning	PM2.5 concentrations	Atmospheric gas species	Excels at high PM2.5 levels; CO and fires strongly influence model accuracy.

Table 7. (cont.)

Study	Model/Method	Key variables studied	Pollutant source	Source impact on model accuracy
[23]	Machine learning	O_3, PM2.5	NOx, industrial activities	Explains NO_x's role in co-pollution; notes changes in O_3 due to NO_x levels.
[36]	GD-GRU model	PM2.5 concentrations	GRU and diffusion modeling	Improved accuracy in extreme PM2.5 prediction; helpful for environmental data repair.
[38]	Linear mixed effect model	NO_2 concentrations	TROPOMI satellite data	Outperforms others; detects pollution hotspots and COVID-19-related anomalies.
[39]	Machine learning	O_3, SOA, VOCs	Specific VOCs	Nonlinear interactions emphasized; key VOCs drive O_3 and SOA formation.
[12]	Machine learning	PM2.5 concentrations	Surface water bodies	Water bodies reduce PM2.5 significantly; supports nature-based mitigation.
[44]	ML with spatial context	Climate change, urban sustainability	Climate variables	Links climate and pollution; promising for early warning systems.
[57]	Big data, cognitive computing	Ground-level O_3	Spatial and cross-correlations	Provides fast insights without heavy computation; efficient urban monitoring.
[35]	IoT and FPMN network	PM2.5 pollution	Sparse fixed stations	FPMN enhances spatial resolution; effective for regional PM2.5 mapping.
[59]	pg-Causality model	Air pollutants	Urban big data, meteorology	Detects causal air pollution paths; supports advanced urban analytics.
[8]	Deep-AIR ML model	Fine-grained pollution prediction	Urban data, meteorology	Outperforms baselines; accurately predicts key pollutants.
[26]	Multi-agent, ML	PM10 pollution	Not specified	Integrates ML and simulation; lacks thorough limitation discussion.
[48]	Low-cost sensors + ML	Particulate matter	Sensor and ML integration	Provides reliable screening; more practical applications needed.
[50]	Systematic review	Urban/indoor air quality	Crowd sensing, low-cost sensors	Identifies benefits of sensors; discusses monitoring trade-offs.
[51]	ARIMA modeling	SO_2 concentrations	Meteorology and particulates	Offers insights into SO_2 levels; calls for deeper policy implications.
[54]	HazeEst + SVR	Urban haze pollution	Monitor and mobile sensor data	High-resolution predictions; needs broader validation.
[63]	Various ML algorithms	Industrial pollution	Emission rates	Multi-layer perceptron is least error-prone; deeper analysis needed.
[56]	SVR model	Air pollution estimation	Historical data and WSN	SVR outperforms ANN; useful for pollution-related health studies.

5.2 *Common limitations in current research*

This study highlights the flexibility and adaptability of machine learning models in forecasting air pollution events, accounting for a diverse range of variables and pollutant sources. However, model accuracy tends to vary significantly, underscoring the importance of further exploration and the systematic addressing of model limitations. The predictive performance of these models is heavily influenced by the selection of variables, pollutant source identification, and the modeling techniques employed. To enhance prediction reliability and generalizability, it is crucial to address current methodological constraints. This becomes increasingly vital as air quality forecasting continues to evolve as a scientific discipline.

Research Question (RQ3): What are the limitations in the current research, and how can future studies fill these gaps to improve air pollution predictions?

Table 8. Impact of pollutant sources on model accuracy in air pollution prediction studies

Limitation	Summary	Freq.	Citations
Limited generalizability	Regional variation	7	[3, 10, 21, 24, 29, 33, 36]
Limited exploration of potential limitations	Inadequate analysis	9	[3, 10, 15, 20, 24, 27, 28, 30, 40]
Limited applicability beyond specific regions	Regional dependence	6	[17, 26–28, 44, 51]
Lack of in-depth discussion of limitations	Superficial analysis	14	[3, 5, 6, 10, 15, 17, 20, 24, 27, 28, 30, 32, 39, 43]
Lacks actionable recommendations	Actionable deficiency	8	[1, 4, 11, 24, 29, 33, 35, 60]
Limited discussion on potential limitations	Limited discourse	6	[8, 20, 24, 33, 36, 37]
Needs deeper discussion of limitations and scalability of the approach	Scalability depth	5	[15, 23, 40–42]
More in-depth discussion of potential limitations and their implications for policy development	Policy implications depth	5	[23, 44, 47–49]
Could strengthen its contribution by discussing potential limitations and broader implications for air quality management	Strengthening contribution	4	[8, 18, 46, 51]
More comprehensive discussion of limitations and practical applications could further enhance the paper's impact	Enhanced impact discussion	2	[45, 60]
Deeper exploration of potential limitations and practical implications for AQI prediction	AQI prediction depth	2	[18, 46]
Could delve deeper into potential limitations and practical implications for feasible air pollution assessment	Pollution assessment depth	3	[43, 44, 51]
A more detailed exploration of potential limitations and broader applications could further enhance its significance	Significance exploration	2	[58, 59]
A more comprehensive discussion of limitations and potential applications for urban air quality management	Urban management discussion	2	[8, 60]
A deeper exploration of limitations and practical implications for AQI prediction	AQI prediction depth	2	[18, 26]

In our pursuit of improving air pollution forecasts, this section critically examines the inherent limitations present in existing studies utilizing machine learning models. Given the pivotal role these models play in detecting and forecasting episodes of elevated pollution, it is essential to investigate the barriers that hinder their accuracy and practical application.

We present a comprehensive analysis of these challenges, identifying areas that require refinement and further research. This discussion provides a foundation for future advancements, aiming to strengthen predictive models' robustness, generalizability, and overall performance. For a detailed summary of these limitations, refer to Table 8 and Figure 3.

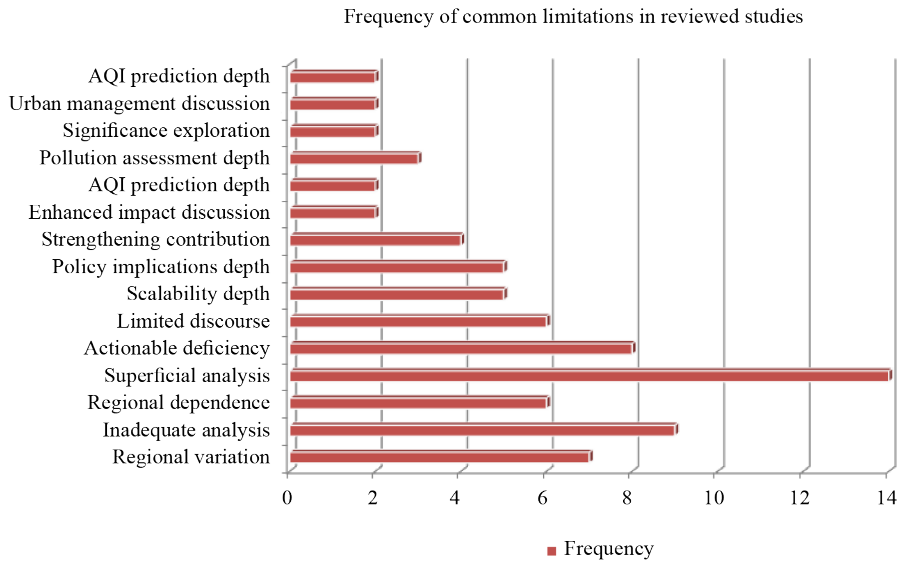

Figure 3. Common limitations

This investigation systematically identifies and categorizes prevailing limitations in current machine learning research on air pollution prediction (see Figure 3 and Table 8). The findings reveal challenges such as limited generalizability across regions, superficial or insufficient analysis of model shortcomings, and a lack of actionable or policy-oriented insights. These issues highlight the complexity and evolving nature of predictive environmental modeling. Addressing these gaps is crucial to strengthening the reliability, scalability, and practical relevance of future prediction models. By acknowledging and learning from these limitations, future studies can contribute to the development of more comprehensive, accurate, and policy-aligned approaches to air quality forecasting.

5.3 *Recommendations for future research*

Building upon the limitations identified in the current body of research, the following recommendations aim to guide future studies in enhancing the robustness and practical relevance of air pollution prediction models:

1. Enhancing generalizability

• Investigate approaches to improve the generalizability of predictive models across diverse geographic regions and demographic contexts.

• Examine regional determinants of air quality variation and develop adaptive modeling frameworks that can accommodate localized characteristics.

2. Deepening the exploration of study limitations

• Promote a culture of critical reflection by encouraging researchers to conduct comprehensive assessments of their study limitations.

• Develop standardized guidelines or checklists to facilitate the systematic reporting of limitations and their implications.

3. Promoting in-depth discussion and transparency

• Foster more rigorous and reflective discussions around methodological shortcomings, model performance constraints, and scalability issues.

• Offer training or editorial support to encourage nuanced presentations of limitations and their broader impacts.

4. Improving spatial prediction capabilities

• Employ advanced geostatistical and deep learning techniques to refine spatial resolution in air quality modeling.

• Leverage high-resolution environmental and meteorological datasets to enhance spatial variability detection.

5. Bridging research and policy with actionable outcomes

• Emphasize the inclusion of clear, actionable recommendations that can guide urban planning, environmental governance, and public health responses.

• Encourage interdisciplinary collaboration to ensure that research findings are directly translatable into real-world policy and practice.

6. Conclusion

This systematic review has offered a comprehensive synthesis of existing research on machine learning applications for predicting periods of elevated air pollution. Through the examination of various modeling techniques, from conventional regression-based approaches to state-of-the-art neural networks, the versatility and adaptability of machine learning in tackling complex environmental problems have been clearly demonstrated.

However, the review also reveals significant limitations in current methodologies, including constraints in model generalizability, superficial analysis of limitations, and a lack of actionable guidance for stakeholders. These challenges underscore the need for more nuanced, scalable, and policy-relevant research efforts.

Addressing these gaps is not merely a methodological refinement-it is a necessary evolution for advancing the precision, relevance, and social impact of air pollution prediction. Future research must prioritize the development of models that are adaptable to regional diversity, supported by rigorous analysis of their limitations, and equipped with recommendations that inform practical interventions.

In charting this path forward, the recommendations presented herein provide a clear roadmap for strengthening both the theoretical foundations and practical applications of this field. The integration of high-resolution data, advanced spatial modeling, and interdisciplinary collaboration holds the promise of transforming air quality prediction into a more precise, responsive, and impactful tool.

Ultimately, the path toward more effective air quality forecasting lies in the concerted efforts of researchers, practitioners, and policymakers. By embracing the insights and recommendations outlined in this review, the academic and applied communities can collaboratively shape a future where environmental health risks are more accurately anticipated and effectively mitigated.

Conflict of interest

The authors declare no competing financial interest.

References

[1] Zhang Y, Wang H, Lu X, Li L, He Y, Wang H, et al. Evolution of summer surface ozone pollution patterns in China during 2015-2020. *Atmospheric Research*. 2023; 291: 106836. Available from: https://doi.org/10.1016/j.atmosres.2023.106836.

[2] Li X, Sun M, Ma Y, Zhang L, Zhang Y, Yang R, et al. Using sensor network for tracing and locating air pollution sources. *IEEE Sensors Journal*. 2021; 21(10): 12162-12170. Available from: https://doi.org/10.1109/JSEN.2021.3063815.

[3] Bai L, Wang J, Ma X, Lu H. Air pollution forecasts: An overview. *International Journal of Environmental Research and Public Health*. 2018; 15(4): 780. Available from: https://doi.org/10.3390/ijerph15040780.

[4] Son R, Stratoulias D, Kim HC, Yoon JH. Estimation of surface PM2.5 concentrations from atmospheric gas species retrieved from TROPOMI using deep learning: Impacts of fire on air pollution over Thailand. *Atmospheric Pollution Research*. 2023; 14(10): 101875. Available from: https://doi.org/10.1016/j.apr.2023.101875.

[5] Tuluri F, Remata R, Walters WL, Tchounwou PB. Impact of regional mobility on air quality during COVID-19 lockdown in mississippi, USA using machine learning. *International Journal of Environmental Research and Public Health*. 2023; 20(11): 6022. Available from: https://doi.org/10.3390/ijerph20116022.

[6] Lee S, Ku H, Hyun C, Lee M. Machine learning-based analyses of the effects of various types of air pollutants on hospital visits by asthma patients. *Toxics*. 2022; 10(11): 644. Available from: https://doi.org/10.3390/toxics10110644.

[7] Hu M, Wang Y, Wang S, Jiao M, Huang G, Xia B. Spatial-temporal heterogeneity of air pollution and its relationship with meteorological factors in the Pearl River Delta, China. *Atmospheric Environment*. 2021; 254: 118415. Available from: https://doi.org/10.1016/j.atmosenv.2021.118415.

[8] Zhang P, Yang L, Ma W, Wang N, Wen F, Liu Q. Spatiotemporal estimation of the PM2.5 concentration and human health risks combining the three-dimensional landscape pattern index and machine learning methods to optimize land use regression modeling in Shaanxi, China. *Environmental Research*. 2022; 208: 112759. Available from: https://doi.org/10.1016/j.envres.2022.112759.

[9] Abubakar H, Shafiq M. Soil fertility classication via hybrid hopfield neural network with election algorithm for random boolean satisfiability reverse analysis. *Suranaree Journal of Science & Technology*. 2022; 29(4): 1.

[10] He W, Zhang S, Meng H, Han J, Zhou G, Song H, et al. Full-coverage PM2.5 mapping and variation assessment during the three-year blue-sky action plan based on a daily adaptive modeling approach. *Remote Sensing*. 2022; 14(15): 3571. Available from: https://doi.org/10.3390/rs14153571.

[11] Doreswamy, Harishkumar KS, Yogesh KM, Gad I. Forecasting air pollution particulate matter (PM2.5) using machine learning regression models. *Procedia Computer Science*. 2020; 171: 2057-2066. Available from: https://doi.org/10.1016/j.procs.2020.04.221.

[12] Gayen BK, Dutta D, Acharya P, Sreekesh S, Kulshrestha UC, Acharya N. Exploring the effect of water bodies coupled with other environmental parameters to model PM2.5 over Delhi-NCT in northwest India. *Atmospheric Pollution Research*. 2022; 13(12): 101614. Available from: https://doi.org/10.1016/j.apr.2022.101614.

[13] Ali GA, Abubakar H, Alzaeemi SAS, Almawgani AH, Sulaiman A, Tay KG. Artificial dragonfly algorithm in the Hopfield neural network for optimal Exact Boolean *k* satisfiability representation. *Plos One*. 2023; 18(9): e0286874. Available from: https://doi.org/10.1371/journal.pone.0286874.

[14] Abubakar H, Masanawa AS, Yusuf S, Boaku G. Optimal representation to high order random boolean ksatisability via election algorithm as heuristic search approach in hopeld neural networks. *Journal of the Nigerian Society of Physical Sciences*. 2021; 3(3): 201-208. Available from: https://doi.org/10.46481/jnsps.2021.217.

[15] Gocheva-Ilieva S, Ivanov A, Kulina H, Stoimenova-Minova M. Multi-step ahead ex-ante forecasting of air pollutants using machine learning. *Mathematics*. 2023; 11(7): 1566. Available from: https://doi.org/10.3390/math11071566.

[16] Abubakar H, Masanawa SA, Yusuf S. Neuro-symbolic integration of hopfield neural network for optimal maximum random ksatisfiability (maxrksat) representation. *Journal of Reliability and Statistical Studies*. 2020; 13(1): 199-220. Available from: https://doi.org/10.13052/jrss1550-4646.13110.

[17] Lu D, Mao W, Zheng L, Xiao W, Zhang L, Wei J. Ambient PM2.5 estimates and variations during the COVID-19 pandemic in the Yangtze River Delta using machine learning and big data. *Remote Sensing*. 2021; 13(8): 1423. Available from: https://doi.org/10.3390/rs13081423.

[18] Mahalingam U, Elangovan K, Dobhal H, Valliappa C, Shrestha S, Kedam G. A machine learning model for air quality prediction for smart cities. In: *2019 International Conference on Wireless Communications, Signal Processing and Networking*. Chennai, India: IEEE; 2019. p.452-457. Available from: https://doi.org/10.1109/WiSPNET45539.2019.9032734.

[19] Ambade B, Sankar TK, Gautam S, Mahato DK, Dumka UC, Mohammad F, et al. Black carbon vs carbon monoxide: Assessing the impact on Indian urban cities. *Water, Air, & Soil Pollution*. 2023; 234(11): 710. Available from: https://doi.org/10.1007/s11270-023-06706-w.

[20] Chen MH, Chen YC, Chou TY, Ning FS. PM2.5 concentration prediction model: A CNN-RF ensemble framework. *International Journal of Environmental Research and Public Health*. 2023; 20(5): 4077. Available from: https://doi.org/10.3390/ijerph20054077.

[21] Zou Y, O'Neill SM, Larkin NK, Alvarado EC, Solomon R, Mass C, et al. Machine learning-based integration of high-resolution wildfire smoke simulations and observations for regional health impact assessment. *International Journal of Environmental Research and Public Health*. 2019; 16(12): 2137. Available from: https://doi.org/10.3390/ijerph16122137.

[22] Lee SJ, Lee H, Kim S, Kang HJ, Kim H, Seo YK, et al. Pollution characteristics of PM2.5 during high concentration periods in summer and winter in Ulsan, the largest industrial city in South Korea. *Atmospheric Environment*. 2023; 292: 119418. Available from: https://doi.org/10.1016/j.atmosenv.2022.119418.

[23] Zhan J, Zheng F, Xie R, Liu J, Chu B, Ma J, et al. The role of NO_x in Co-occurrence of O_3 and PM2.5 pollution driven by wintertime east Asian monsoon in Hainan. *Environment International*. 2023; 179: 118645. Available from: https://doi.org/10.1016/j.jenvman.2023.118645.

[24] Yang J, Fu X, Qiao L, Yao L, Zhang F, Li W. Characteristics of atmospheric pollution in a chinese megacity: Insights from three different functional areas. *Sustainability*. 2023; 15(3): 2429. Available from: https://doi.org/10.3390/su15032429.

[25] Essamlali I, Nhaila H, El Khaili M. Supervised machine learning approaches for predicting key pollutants and for the sustainable enhancement of urban air quality: a systematic review. *Sustainability*. 2024; 16(3): 976. Available from: https://doi.org/10.3390/su16030976.

[26] Ngom B, Diallo M, Seyc MR, Drame MS, Cambier C, Marilleau N. PM10 data assimilation on real-time agent-based simulation using machine learning models: case of dakar urban air pollution study. In: *Proceedings of the 2021 IEEE/ACM 25th International Symposium on Distributed Simulation and Real Time Applications*. Valencia, Spain: IEEE; 2021. p.1-4. Available from: https://doi.org/10.1109/DS-RT52167.2021.9576143.

[27] Razavi-Termeh SV, Sadeghi-Niaraki A, Choi SM. Spatial modeling of asthma-prone areas using remote sensing and ensemble machine learning algorithms. *Remote Sensing*. 2021; 13(16): 3222. Available from: https://doi.org/10.3390/rs13163222.

[28] Hu X, Zhang J, Xue W, Zhou L, Che Y, Han T. Estimation of the near-surface ozone concentration with full spatiotemporal coverage across the Beijing-Tianjin-Hebei Region based on extreme gradient boosting combined with a WRF-chem model. *Atmosphere*. 2022; 13(4): 632. Available from: https://doi.org/10.3390/atmos13040632.

[29] Gu Y, Li B, Meng Q. Hybrid interpretable predictive machine learning model for air pollution prediction. *Neurocomputing*. 2022; 468: 123-136. Available from: https://doi.org/10.1016/j.neucom.2021.09.051.

[30] Choi W, Park J, Kim D, Park J, Kim S, Lee H. Development of two-dimensional visibility estimation model using machine learning: preliminary results for South Korea. *Atmosphere*. 2022; 13(8): 1233. Available from: https://doi.org/10.3390/atmos13081233.

[31] Yazdi MD, Kuang Z, Dimakopoulou K, Barratt B, Suel E, Amini H, et al. Predicting fine particulate matter (PM2.5) in the greater London area: An ensemble approach using machine learning methods. *Remote Sensing*. 2020; 12(6): 914. Available from: https://doi.org/10.3390/rs12060914.

[32] Zaidan MA, Dada L, Alghamdi MA, Al-Jeelani H, Lihavainen H, Hyvärinen A, et al. Mutual information input selector and probabilistic machine learning utilisation for air pollution proxies. *Applied Sciences*. 2019; 9(20): 4475. Available from: https://doi.org/10.3390/app9204475.

[33] Wang A, Xu J, Tu R, Saleh M, Hatzopoulou M. Potential of machine learning for prediction of traffic related air pollution. *Transportation Research Part D: Transport and Environment*. 2020; 88: 102599. Available from: https://doi.org/10.1016/j.trd.2020.102599.

[34] Yang Q, Yuan Q, Gao M, Li T. A new perspective to satellite-based retrieval of ground-level air pollution: Simultaneous estimation of multiple pollutants based on physics-informed multi-task learning. *Science of the Total Environment*. 2023; 857: 159542. Available from: https://doi.org/10.1016/j.scitotenv.2022.159542.

[35] Li T, Zhang Q, Peng Y, Guan X, Li L, Mu J, et al. Contributions of various driving factors to air pollution events: Interpretability analysis from Machine learning perspective. *Environment International*. 2023; 173: 107861. Available from: https://doi.org/10.1016/j.envint.2023.107861.

[36] Xu R, Deng X, Wan H, Cai Y, Pan X. A deep learning method to repair atmospheric environmental quality data based on Gaussian diffusion. *Journal of Cleaner Production*. 2021; 308: 127446. Available from: https://doi.org/10.1016/j.jclepro.2021.127446.

[37] Zhu Z, Tang G, Wu L, Wang Y, Liu B, Li Q, et al. Significant decline in aerosols in the mixing layer in Beijing from 2015 to 2020: Effects of regional coordinated air pollution control. *Science of The Total Environment*. 2022; 838: 156364. Available from: https://doi.org/10.1016/j.scitotenv.2022.156364.

[38] Dong J, Cai X, Tian L, Chen F, Xu Q, Li T, et al. Satellite-based estimates of daily NO_2 exposure in urban agglomerations of China and application to spatio-temporal characteristics of hotspots. *Atmospheric Environment*. 2023; 293: 119453. Available from: https://doi.org/10.1016/j.atmosenv.2022.119453.

[39] Wang F, Zhang Z, Wang G, Wang Z, Li M, Liang W, et al. Machine learning and theoretical analysis release the non-linear relationship among ozone, secondary organic aerosol and volatile organic compounds. *Journal of Environmental Sciences*. 2022; 114: 75-84. Available from: https://doi.org/10.1016/j.jes.2021.07.026.

[40] Ngoc BAP, Dieudonné E, Delbarre H, Deboudt K, Nguyen ST, Bui VH, et al. Intricate behavior of winter pollution in Hanoi over the 2006-2020 semi-climatic period. *Atmospheric Environment*. 2023; 300: 119669. Available from: https://doi.org/10.1016/j.atmosenv.2023.119669.

[41] Wang X, Duan W, Zhu J, Wei W, Cheng S, Mao S. Nonlinear influence of winter meteorology and precursor on PM2.5 based on mathematical and numerical models: A COVID-19 and Winter Olympics case study. *Atmospheric Environment*. 2022; 278: 119072. Available from: https://doi.org/10.1016/j.atmosenv.2022.119072.

[42] Qiu Y, Wu Z, Man R, Liu Y, Shang D, Tang L, et al. Historically understanding the spatial distributions of particle surface area concentrations over China estimated using a non-parametric machine learning method. *Science of The Total Environment*. 2022; 824: 153849. Available from: https://doi.org/10.1016/j.scitotenv.2022.153849.

[43] Liu H, Jia M, Tao J, Yao D, Li J, Zhang R, et al. Elucidating pollution characteristics, temporal variation and source origins of carbonaceous species in Xinxiang, a heavily polluted city in North China. *Atmospheric Environment*. 2023; 298: 119626. Available from: https://doi.org/10.1016/j.atmosenv.2023.119626.

[44] Balogun AL, Tella A, Baloo L, Adebisi N. A review of the inter-correlation of climate change, air pollution and urban sustainability using novel machine learning algorithms and spatial information science. *Urban Climate*. 2021; 40: 100989. Available from: https://doi.org/10.1016/j.uclim.2021.100989.

[45] Rahimpour A, Amanollahi J, Tzanis CG. Air quality data series estimation based on machine learning approaches for urban environments. *Air Quality, Atmosphere & Health*. 2021; 14: 191-201. Available from: https://doi.org/10.1007/s11869-020-00925-4.

[46] Feizi H, Sattari MT, Prasad R, Apaydin H. Comparative analysis of deep and machine learning approaches for daily carbon monoxide pollutant concentration estimation. *International Journal of Environmental Science and Technology*. 2023; 20(2): 1753-1768. Available from: https://doi.org/10.1007/s13762-022-04702-x.

[47] Bamrah SK, Srivatsan S, Gayathri KS. Region classification for air quality estimation using deep learning and machine learning approach. In: *Lecture Notes in Electrical Engineering*. Singapore: Springer; 2023. p.333-344. Available from: https://doi.org/10.1007/978-981-19-5868-7_25.

[48] Caquilpán PV, Aros GG, Elgueta AS, Díaz SR, Sepúlveda KG, Sierralta JC. Advantages and challenges of the implementation of a low-cost particulate matter monitoring system as a decision-making tool. *Environmental Monitoring and Assessment*. 2019; 191(11). Available from: https://doi.org/10.1007/s10661-019-7875-4.

[49] Tao M, Xu Y, Gong J, Liu Q. Estimation of aerosol acidity at a suburban site of Nanjing using machine learning method. *Journal of Atmospheric Chemistry*. 2022; 79(2): 141-151. Available from: https://doi.org/10.1007/s10874-022-09433-4.

[50] Hernández-Gordillo A, Ruiz-Correa S, Robledo-Valero V, Hernández-Rosales C, Arriaga S. Recent advancements in low-cost portable sensors for urban and indoor air quality monitoring. *Air Quality, Atmosphere and Health*. 2021; 14(12): 1931-1951. Available from: https://doi.org/10.1007/s11869-021-01067-x.

[51] Borhani F, Shafiepour Motlagh M, Rashidi Y, Ehsani AH. Estimation of short-lived climate forced sulfur dioxide in Tehran, Iran, using machine learning analysis. *Stochastic Environmental Research and Risk Assessment*. 2022; 36(9): 2847-2860. Available from: https://doi.org/10.1007/s00477-021-02167-x.

[52] Nelgadevi M, Jacob G. Air pollution-level estimation in smart cities using machine learning algorithms. In: *International Conference on Artificial Intelligence, Smart Grid and Smart City Applications*. Cham: Springer; 2020. p.343-352. Available from: https://doi.org/10.1007/978-3-030-24051-6_33.

[53] Srivastava C, Singh S, Singh AP. Estimation of air pollution in Delhi using machine learning techniques. In: *2018 International Conference on Computing, Power and Communication Technologies*. Greater Noida, India: IEEE; 2018. p.304-309. Available from: https://doi.org/10.1109/GUCON.2018.8675022.

[54] Hu K, Rahman A, Bhrugubanda H, Sivaraman V. HazeEst: Machine learning based metropolitan air pollution estimation from fixed and mobile sensors. *IEEE Sensors Journal*. 2017; 17(11): 3517-3525. Available from: https://doi.org/10.1109/JSEN.2017.2690975.

[55] Mehmood MS, Rehman A, Sajjad M, Song J, Zafar Z, Shiyan Z, et al. Evaluating land use/cover change associations with urban surface temperature via machine learning and spatial modeling: Past trends and future simulations in Dera Ghazi Khan, Pakistan. *Frontiers in Ecology and Evolution*. 2023; 11: 1115074. Available from: https://doi.org/10.3389/fevo.2023.1115074.

[56] Hu K, Sivaraman V, Bhrugubanda H, Kang S, Rahman A. SVR based dense air pollution estimation model using static and wireless sensor network. In: *2016 IEEE SENSORS*. Orlando, FL, USA: IEEE; 2016. p.1-3. Available from: https://doi.org/10.1109/ICSENS.2016.7808827.

[57] Chiwewe TM, Ditsela J. Machine learning based estimation of Ozone using spatio-temporal data from air quality monitoring stations. In: *2016 IEEE 14th International Conference on Industrial Informatics (INDIN)*. Poitiers, France: IEEE; 2016. p.58-63. Available from: https://doi.org/10.1109/INDIN.2016.7819134.

[58] Li X, Sun M, Ma Y, Zhang L, Zhang Y, Yang R, et al. Using sensor network for tracing and locating air pollution sources. *IEEE Sensors Journal*. 2021; 21(10): 12162-12170. Available from: https://doi.org/10.1109/JSEN.2021.3063815.

[59] Zhu JY, Zhang C, Zhang H, Zhi S, Li VOK, Han J, et al. pg-Causality: Identifying spatiotemporal causal pathways for air pollutants with urban big data. *IEEE Transactions on Big Data*. 2017; 4(4): 571-585. Available from: https://doi.org/10.1109/TBDATA.2017.2723899.

[60] Yarragunta S, Nabi MA, Jeyanthi P, Revathy S. Prediction of air pollutants using supervised machine learning. In: *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*. Madurai, India: IEEE; 2021. p.1633-1640. Available from: https://doi.org/10.1109/ICICCS51141.2021.9432078.

[61] Zhang Q, Han Y, Li VOK, Lam JCK. Deep-AIR: A hybrid CNN-LSTM framework for fine-grained air pollution estimation and forecast in metropolitan cities. *IEEE Access*. 2022; 10: 55818-55841. Available from: https://doi.org/10.1109/ACCESS.2022.3174853.

[62] Abubakar H, Danrimi ML. Hopfield type of artificial neural network via election algorithm as heuristic search method for random boolean ksatisfiability. *International Journal of Computing and Digital Systems*. 2021. Available from: https://doi.org/10.12785/ijcds/100163.

[63] Gupta V, Singh KP. Machine learning-driven predictive modeling of biodiesel characteristics with heterogeneous catalysts: a review. *Machine Learning*. 2023; 55(12): 4301-4315.