Environmental Modelling and Software 185 (2025) 106312

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: www.elsevier.com/locate/envsoft

Application of artificial intelligence in air pollution monitoring and forecasting: A systematic review

Sreeni Chadalavada [a], Oliver Faust [b], Massimo Salvi [c,*], Silvia Seoni [c], Nawin Raj [d], U. Raghavendra [e], Anjan Gudigar [e], Prabal Datta Barua [f], Filippo Molinari [c], Rajendra Acharya [d,g]

[a] School of Engineering, University of Southern Queensland, Springfield, Australia
[b] School of Computing and Information Science, Anglia Ruskin University Cambridge Campus, UK
[c] Biolab, PolitoBIOMedLab, Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy
[d] School of Mathematics, Physics and Computing, University of Southern Queensland, Springfield, Australia
[e] Department of Instrumentation and Control Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education, Manipal, 576104, India
[f] School of Business, University of Southern Queensland, Springfield, Australia
[g] Centre for Health Research, University of Southern Queensland, Australia

ARTICLE INFO

Keywords:
Air pollution
Artificial intelligence
Deep learning
Monitoring tools
Health
PRISMA guidelines

ABSTRACT

Air pollution poses a significant global health hazard. Effective monitoring and predicting air pollutant concentrations are crucial for managing associated health risks. Recent advancements in Artificial Intelligence (AI), particularly Machine Learning (ML) and Deep Learning (DL), offer the potential for more precise air pollution monitoring and forecasting models. This comprehensive review, conducted according to PRISMA guidelines, analyzed 65 high-quality Q1 journal articles to uncover current trends, challenges, and future AI applications in this field. The review revealed a significant increase in research papers utilizing ML and DL approaches from 2021 onwards. ML techniques currently dominate, with Random Forest being the most frequent method, achieving up to 98.2% accuracy. DL techniques show promise in capturing complex spatiotemporal relationships in air quality data. The study highlighted the importance of integrating diverse data sources to improve model accuracy. Future research should focus on addressing challenges in model interpretability and uncertainty quantification.

1. Introduction

Air pollution, exacerbated by environmental and climate change, is a pressing global issue affecting numerous cities and regional places. Exposure to Particulate Matter (PM), a key air pollutant, has been linked to approximately 4.2 million deaths, ranking it the fifth leading global health risk (WHO TEAM, 2016). While global efforts have led to decreasing pollutant concentrations in many areas, levels often remain above World Health Organization (WHO) guidelines and national thresholds in numerous regions. This persistent challenge underscores the need for effective solutions to address air pollution. Air pollution contributes to various health issues, including respiratory complications, heart and lung conditions, premature mortality, and adverse community impacts, which can lead to mental health problems (Nazmul Hoq et al., 2019), (Soleimani et al., 2019). Timely access to air quality information and preventive measures are essential for mitigating health risks. Consequently, monitoring air quality has become imperative (WHO TEAM, 2016). Hence, air pollution poses significant public health challenges, contributing to various adverse effects on human health and the environment. Long-term exposure to poor air quality leads to pulmonary and cardiovascular diseases, lung cancer, and strokes, resulting in a high global mortality rate (Zhang et al., 2014), (Yamamoto et al., 2014). Developing nations are particularly vulnerable, with a substantial number of premature deaths attributed to air pollution. Even developed countries are not immune to its impacts, with a considerable percentage of the population exposed to hazardous conditions (WHO TEAM, 2016). Short-term effects include sneezing, headaches, dizziness, and eye irritation. Moreover, emerging research has linked air pollution to infertility (Choe et al., 2018), (Gaskins et al., 2019).

The proliferation of cost-effective remote sensors and the availability of vast environmental and clinical data have led to a surge in pollution datasets for analysis. However, these large, complex datasets present

* Corresponding author.
E-mail address: massimo.salvi@polito.it (M. Salvi).

https://doi.org/10.1016/j.envsoft.2024.106312
Received 20 September 2024; Received in revised form 7 December 2024; Accepted 9 December 2024
Available online 29 December 2024

challenges that traditional epidemiological and environmental health models struggle to handle effectively. New analytical methods are required to gain deeper insights from these data. Data mining and Machine Learning (ML) techniques offer scalable and reliable approaches to analyze modern, large-scale air pollution datasets (Li et al., 2023a). However, analysing these extensive and complex datasets with traditional statistical regression models has limitations in capturing nonlinear relationships and achieving high prediction accuracy (Pielke et al., 2007). As such, more accurate and convenient methods are urgently needed for effective data analysis in atmospheric science.

Artificial Intelligence (AI), encompassing both ML and Deep Learning (DL) techniques, has emerged as a powerful solution to improve traditional epidemiological and environmental health models (Pielke et al., 2007). Furthermore, AI plays a crucial role in the real-time tracking of pollution hotspots, identifying trends in pollution levels, and modeling the impact of meteorological factors (Han and Wang, 2021). This technology empowers researchers and policymakers to make informed decisions for air quality management and public health protection (Jerrett et al., 2001). By offering fast, accurate, and reliable solutions, AI enables a deeper understanding of complex datasets in various scientific disciplines and supports data-driven decision-making processes. ML algorithms, such as Artificial Neural Networks (ANN), Random Forest (RF), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGBoost), have proven to be invaluable tools in addressing air pollution and their impact on public health (Li et al., 2023a), (Zheng et al., 2021). These algorithms excel at handling extensive and multidimensional air quality datasets, allowing for a deeper understanding of the complex relationships between air pollutants and their health effects (Wei et al., 2021). By leveraging ML, we can achieve higher prediction accuracy and robustness, even in scenarios where air pollution exhibits intricate and nonlinear patterns (Chen et al., 2022), (Joharestani et al., 2019). These advancements in ML empower us to make more informed decisions in environmental protection, assess the health risks associated with air pollution, and develop effective mitigation strategies to safeguard public health. Unsupervised learning methods, like K-means clustering, can improve data analysis further by revealing hidden patterns within unlabeled datasets, providing valuable insights for understanding complex phenomena (Ahmed et al., 2020). DL incorporates outcome driven feature engineering which leads to robust models that perform well in practical settings. Commonly used DL architectures are: Long Short-Term Memory (LSTM), Bidirectional Long Short-Term Memory (LSTM), and Recurrent Neural Networks (RNN). DL models have shown exceptional performance in predicting various air quality parameters, including fine particulate matter ($PM_{2.5}$) concentrations (Bekkar et al., 2021), (Zhu et al., 2018). Their ability to capture nuanced relationships within air quality data makes DL a powerful tool for forecasting and health risk assessment (Bekkar et al., 2021). Fig. 1 shows a typical AI-based framework for air pollution monitoring.

To the best of our knowledge, existing reviews have predominantly concentrated on either ML or DL tools for air pollution monitoring and forecasting. In contrast, our work offers a comprehensive review of ML and DL tools, encompassing most air pollutants. While previous reviews focused on air pollution prediction, our review addresses both forecasting and monitoring aspects. Monitoring typically refers to the real-time or near-real-time assessment of current air quality conditions, involving the collection, analysis, and interpretation of data about present air pollution levels. Forecasting, on the other hand, involves predicting future air quality conditions based on current data and historical patterns. Moreover, we delve into the factors contributing to air pollution trends. Our paper also considers hybrid approaches combining AI methods with other techniques like geospatial parameters and data mining. Our paper encompasses a broad spectrum of air pollution aspects, many of which have remained unexplored in existing reviews.

This paper provides a comprehensive review of AI-based air pollution monitoring and forecasting. The remainder of the manuscript is structured as follows: background discussion, methods employed, and a thorough exploration of air pollutants. The study analyses various AI algorithms and delves into the results and implications. Extensive discussion unfolds the complexities, followed by insights into potential areas for future research. We highlight the limitations of current approaches and opportunities to enhance predictive performance. The review synthesizes key trends, knowledge gaps, and new directions to guide further advancement in applying AI for air quality monitoring and forecasting.

2. Methods

We closely adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to select the most relevant articles on multi-modality in healthcare.

2.1. Related reviews

In recent years, there has been a surge in literature documenting efforts to harness the potential of ML and DL techniques for air pollution forecasting and monitoring. These efforts have resulted in notable improvements in prediction accuracy. This review paper explores the latest advancements in air pollution forecasting and real-time monitoring. To emphasize the unique and innovative aspects of our approach, we have summarized six existing review papers, facilitating a comparative analysis.

- Gugnani et al. (Gugnani and Singh, 2022) (2022) offers an insightful yet limited overview of various DL techniques for predicting air quality. It highlights the advantages and disadvantages of DL without covering other ML models.
- Bellinger et al. (2017) (2017) systematically review data mining and machine learning algorithms applied in air pollution epidemiology, focusing solely on their use for forecasting and prediction.
- Balogun et al. (2021) (2021) review the interplay of climate change, air pollution, and urban sustainability using novel ML and spatial techniques. However, the scope is restricted to these spatial and ML tools for prediction to inform stakeholder interventions.
- Li et al. (2023b) (2023) provide a bibliometric analysis of ML applications in air pollution research, concentrating exclusively on

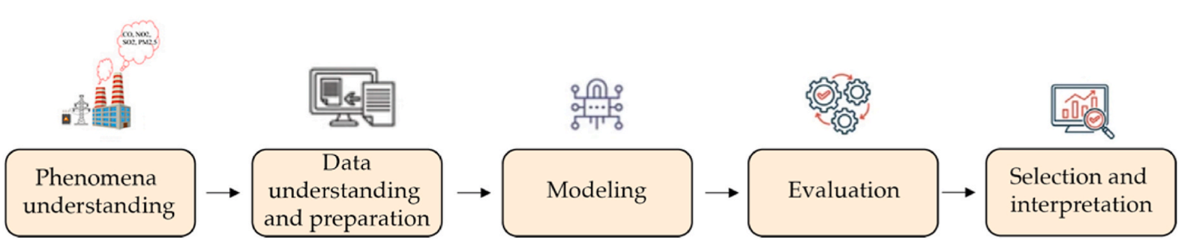

Fig. 1. Typical steps for an AI-based air pollution monitoring and prediction framework. Sensors are generally used to collect air quality and environmental data. The data is used to train ML or DL models to learn patterns from historical data. Model performance is evaluated on test data, and results are provided to users through a dashboard or mobile app.

chemical characterization, forecasting, detection, and emission control optimizations.

- Masood et al. (Masood and Ahmad, 2021) (2021) comprehensively review prominent AI techniques for air pollution forecasting, including ANN, DNN, SVM, and fuzzy logic, via a systematic literature review. However, the scope is limited to these standard AI methods.
- Subramaniam et al. (2022) (2022) discuss AI methodologies and ML algorithms for forecasting and early warning systems. Nonetheless, the focus remains narrowly on ML techniques.

While some recent reviews have emerged (Li et al., 2023b), (Subramaniam et al., 2022), they focused on specific aspects such as bibliometric analysis or forecasting applications. Unlike these prior reviews, our comprehensive systematic review synthesizes both ML and DL techniques, expanding beyond forecasting to encompass monitoring and identification of pollution drivers. We provide a holistic analysis of the rapid technological evolution from 2019 to 2023, a period that saw substantial improvements in AI model performance (Abu El-Magd et al., 2023). Moreover, our review goes beyond conventional approaches by exploring integrating AI techniques with geospatial data, data mining, and other related methodologies. Our comprehensive review of AI applications in air quality research allows for an in-depth examination of the complex, multifaceted factors influencing pollution. We assess a spectrum of pollutants and their health impacts using ML and DL models. This allows the identification of the most suitable techniques for specific pollutants based on spatial and temporal characteristics.

Our review offers researchers a comprehensive and holistic perspective on AI techniques across air pollution analysis domains. We assess the evolution of monitoring and prediction capabilities based on ML versus DL methods. Fig. 2 highlights our expansive scope compared to existing literature, spanning forecasting, monitoring, drivers, health impacts, ML, and DL approaches. This comprehensive vantage point empowers impactful solutions to this critical environmental and public health challenge.

2.2. Literature search strategy

The selection and review of articles on air pollution forecasting and monitoring using AI techniques in this paper comply with the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines. To ensure a comprehensive article search, we have used the following databases: PubMed, ScienceDirect, and Web of Science. The search was carried out based on the following keywords: "Air Pollution", "Human Health," "Machine Learning," "Artificial Intelligence," and "Deep Learning,". This article's examination encompasses the most recently accepted papers until June 20, 2023.

Following PRISMA guidelines, the initial literature search identified 412 research papers. After removing 193 duplicate papers, 219 papers with unique titles were screened further. Papers were excluded at this stage if they did not meet the inclusion criteria of 1) being peer-reviewed research articles, 2) being published in English, 3) being relevant to the application of AI techniques for air pollution analysis, forecasting, or prediction, and 4) accessible in full text. Based on these criteria, 99 papers were excluded. The remaining 120 articles underwent a detailed assessment of quality and relevance to identify papers published in Q1 journals per Scimago rankings. A final set of 65 high-quality relevant papers were selected for review. Fig. 3 illustrates the flow diagram documenting the systematic article selection process, including identification, screening, eligibility assessment, and papers included and excluded at each stage as per PRISMA guidelines.

Fig. 2. Comparison of our review paper with existing literature reviews.

2.3. Air pollutants

This section provides an overview of the chemical properties and associated health effects of the most significant air pollutants.

- Particulate matter (PM): PM consists of small solid or liquid particles suspended in air. These particles can vary in size and composition, with $PM_{2.5}$ referring to fine particles with a diameter of 2.5 μm or smaller, and PM_{10} includes larger particles up to 10 μm in diameter. PM can be composed of various materials, including dust, pollen, soot, organic compounds, and metals. Long-term exposure to $PM_{2.5}$ is associated with a range of health problems, including respiratory infections, aggravated asthma, decreased lung function, and premature death (Arden et al.). PM exposure has also been linked to cardiovascular diseases and lung cancer (Brunekreef and Holgate).
- Nitrogen Oxides (NOx): NOxs are a family of reactive gases that contain nitrogen and oxygen atoms. The two most common forms are nitrogen dioxide (NO_2) and nitric oxide (NO). NOx emissions from combustion processes, such as those in vehicles and industrial activities, primarily consist of nitric oxide (NO) (80–90%) with a smaller proportion of NO_2 (10–20%). However, NO_2 concentrations in ambient air increase rapidly, typically within minutes, due to the atmospheric conversion of NO to NO_2. Prolonged exposure to NOx is associated with respiratory problems and decreased lung function (E. C. C. and H. WHO TEAM Air quality and health, 2006). It can also contribute to cardiovascular diseases (Brook et al., 2010).
- Ground-level ozone (O_3): O_3 is a secondary pollutant formed when precursor pollutants, such as volatile organic compounds (VOCs) and nitrogen oxides (Nox), react in the presence of sunlight. It is a molecule composed of three oxygen atoms and is a key component of smog. O_3 can irritate the respiratory system and lead to coughing and shortness of breath (Bell et al., 2005). Long-term exposure to ozone is associated with reduced lung function and the development or exacerbation of respiratory diseases, including asthma and COPD (Jerrett et al., 2001).
- Carbon monoxide (CO): CO is a colorless, odorless gas consisting of one carbon and one oxygen atom. It is produced primarily by incomplete combustion of carbon-containing fuels, such as gasoline and wood. CO binds to hemoglobin in red blood cells, reducing their ability to carry oxygen. CO interferes with oxygen transport in the body, leading to symptoms like headaches and dizziness. While ambient CO levels monitored for air quality purposes are typically well below those causing acute poisoning, they serve as an important

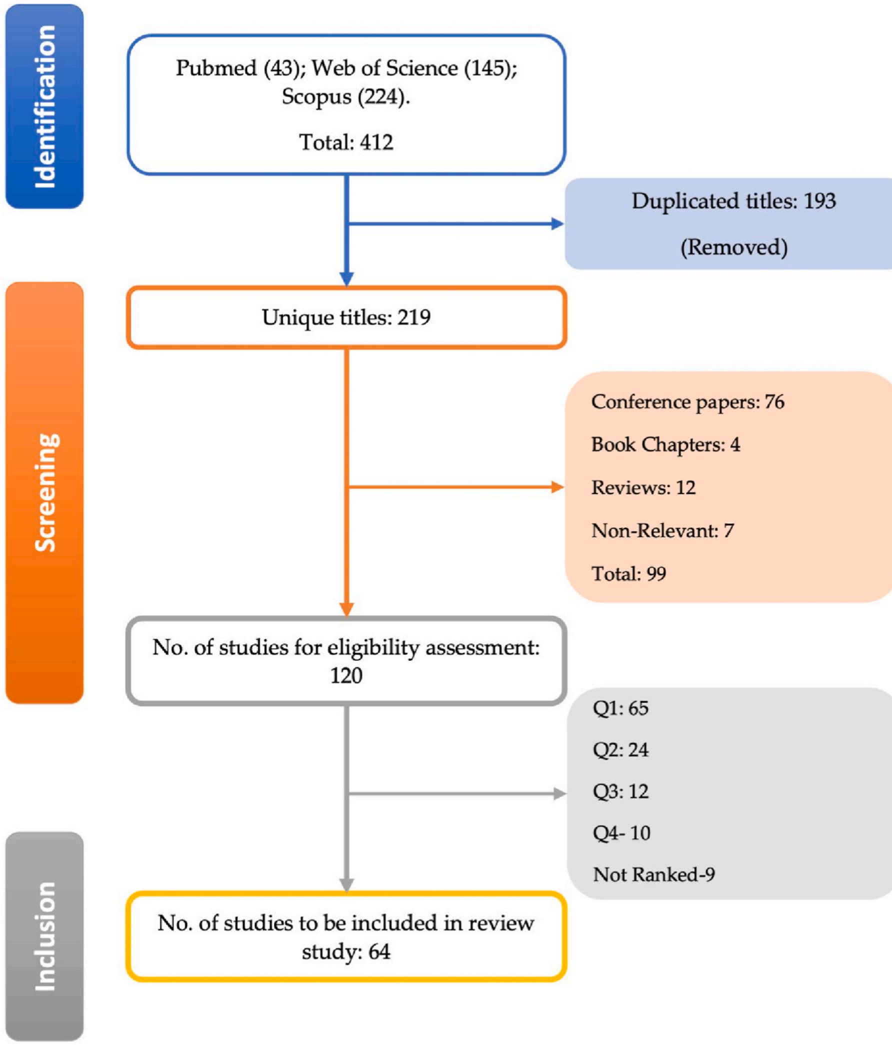

Fig. 3. Study search strategy by PRISMA guidelines. Papers published in non-Q1 journals were excluded from this review.

indicator of combustion-related pollution. Chronic exposure to elevated ambient CO levels, though much lower than those causing acute poisoning, has been linked to an increased risk of cardiovascular health issues (Brook et al., 2010). High levels of CO exposure can result in carbon monoxide poisoning, which can be fatal (Raub et al., 2000).

- Sulphur dioxide (SO_2): SO_2 is a pungent gas composed of one Sulphur atom and two oxygen atoms. It is generated primarily by burning fossil fuels containing Sulphur, such as coal and oil. SO_2 can react with other atmospheric substances to form sulfuric acid (H_2SO_4), contributing to acid rain. SO_2 irritates the respiratory system, causing symptoms like coughing and wheezing (Eftim et al., 2008). Long-term exposure to SO_2 can contribute to respiratory diseases and exacerbate existing conditions. SO_2 is also associated with cardiovascular problems (Brook et al., 2010).

Air pollutants can interact synergistically, resulting in cumulative adverse health effects. The severity of these health risks depends on factors including pollutant concentrations, length of exposure, and individual susceptibility. Effective regulatory measures and air quality monitoring are vital to mitigate air pollution and associated public health burdens (Bekkar et al., 2021). Fig. 4 shows the different types of contaminants, while Table 1 summarizes their impact on human health.

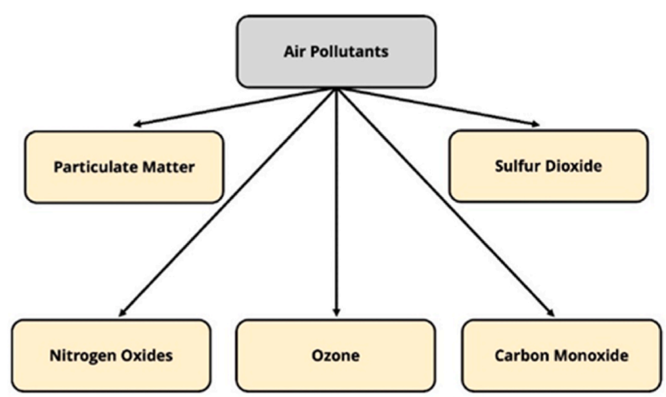

Fig. 4. Different types of air contaminants.

2.4. Artificial intelligence algorithms

AI represents a transformative field miming human intelligence in machines using techniques like ML and DL. ML employs an algorithm that learns patterns from data to make decisions, while DL leverages neural networks to extract complex features automatically. This enables DL to excel in complex tasks like image and speech recognition, setting it apart from traditional ML techniques. The following ML techniques were

Table 1
Air pollutants and their impact on human health.

Contaminant	Health impact
Particulate matter (PM)	Long-term exposure to PM can cause respiratory infections, reduce lung function, and lead to premature death (Arden et al.), (Brunekreef and Holgate).
Nitrogen Oxides (NOx)	NOx irritates the respiratory system and is associated with respiratory problems and cardiovascular diseases (E. C. C. and H. WHO TEAM Air quality and health, 2006), (Brook et al., 2010).
Ground-level ozone (O_3)	Long-term exposure to O3 can lead to reduced lung function and the development or worsening of respiratory diseases (Jerrett et al., 2001), (Bell et al., 2005).
Carbon monoxide (CO)	High levels of CO exposure can lead to carbon monoxide poisoning, which can be fatal (Brook et al., 2010), (Raub et al., 2000).
Sulphur dioxide (SO_2)	Long-term exposure to SO2 contributes to respiratory diseases and is associated with cardiovascular problems (Brook et al., 2010), (Eftim et al., 2008).

used for air pollution forecasting and monitoring.

- K-Nearest Neighbours (KNN) is a simple yet effective ML algorithm used for classification and regression tasks. It makes predictions by finding the K data points nearest to a given input and assigning a label based on the majority class (for classification) or computing a weighted average (for regression) (Cover and Hart, 1952). KNN is easy to implement, non-parametric, and can handle both categorical and numerical data. It's also beneficial for small to medium-sized datasets.
- Random Forest (RF) is an ensemble learning method that combines multiple decision trees to make predictions (Probst et al., 2019). It operates by constructing decision trees during training and outputs the class, which is the mode of the classes (classification) or mean prediction (regression) of individual trees. RF is robust against overfitting, handles high-dimensional data, and handles both categorical and numerical features well.
- Extreme Gradient Boosting (XGBoost) is a gradient-boosting algorithm widely used in classification and regression tasks (Chen and Guestrin, 2016). It sequentially builds an ensemble of decision trees, optimizing for residuals at each step. XGBoost offers high performance, handles missing data, and provides feature ranking.
- Regional Feature Selection-based ML (RFSML) is a ML approach that emphasizes selecting relevant features from different regions or parts of the data (Wu et al., 2017). It typically involves feature engineering and selection techniques tailored to the specific areas of the dataset. RFSML allows for fine-grained feature selection, potentially improving model performance and interpretability.

The following DL techniques were used for air pollution forecasting and monitoring.

- Neural Basis Expansion Analysis for Time Series forecasting (N-BEATS) is a DL architecture designed for time series forecasting tasks (Oreshkin et al., 2019). It consists of fully connected neural networks stacked in an ensemble-like fashion. N-BEATS is scalable to handle complex time series data and offers an interpretable and flexible architecture.
- LSTM with Bayesian optimizer combines two methods to achieve high classification and prediction performance. LSTM is a recurrent neural network (RNN) designed to model sequential data, making it suitable for time series and sequence prediction tasks (Smagulova and James, 2019). Bayesian optimization is a method for hyperparameter tuning to optimize the model's performance. LSTM captures temporal dependencies in data, while Bayesian optimization efficiently searches hyperparameters.
- Enhanced Long Short-Term Memory (ELSTM) extends the traditional LSTM architecture. It incorporates enhancements to capture longer-range dependencies and improve model performance in sequential data tasks. ELSTM addresses the vanishing gradient problem by introducing gating mechanisms and is well-suited for tasks where capturing long-term dependencies is crucial, such as natural language processing and time series analysis.
- Convolutional Long Short-Term Memory (Conv-LSTM) is a hybrid neural network architecture that combines the convolutional neural networks with the sequential modeling capabilities of LSTM. It is beneficial for tasks where spatial and temporal patterns must be captured (Shi et al.). Conv-LSTM can effectively learn spatial features from data with temporal dependencies, making it applicable to tasks like video analysis, weather forecasting, and image captioning.
- Bayesian Deep-Learning Model combines deep neural networks with Bayesian inference techniques. It introduces probabilistic interpretations into DL to quantify uncertainty and make more reliable predictions (Seoni et al., 2023), (Gal and Ghahramani, 2015). Bayesian deep-learning models offer probabilistic predictions, which are valuable in scenarios where uncertainty estimation is critical, such as medical diagnosis, autonomous driving, and financial forecasting.
- Hybrid Deep Learning-Driven Sequential Concentration Transport Emission Model (DL-CTEM) is a specialized DL model designed for modeling and forecasting concentration, transport, and emissions of pollutants in a sequential manner. It combines DL techniques with domain-specific knowledge (Kim et al., 2023). DL-CTEM leverages the power of deep learning to capture complex patterns in environmental data while incorporating domain-specific information, making it suitable for air quality prediction and environmental monitoring.

3. Results

Fig. 5 depicts the publication trends for automated air pollution monitoring and forecasting, specifically based on ML and DL approaches. The graph highlights a noticeable uptick in research papers from 2021 onwards, indicating these techniques have gained significant traction recently. There is a growth in papers on ML/DL for air pollution analysis starting in 2021, underscoring rapid advances in this field.

This reflects the maturation of AI capabilities and the pressing need for improved pollution modeling. In addition, the number of papers relying primarily on ML methods vastly exceeds those using DL. This suggests ML techniques currently dominate, likely due to greater interpretability and ease of implementation. However, DL adoption is accelerating, given its advantages in capturing complex spatiotemporal relationships. The dominance of ML thus far indicates that most published studies focus on forecasting and regression tasks where ML excels currently.

Fig. 6 displays the average accuracy of AI models for air pollution

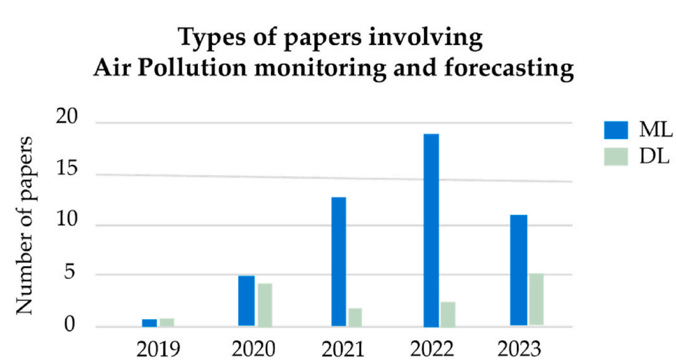

Fig. 5. Number of papers published on ML and DL for Air Pollution monitoring and forecasting.

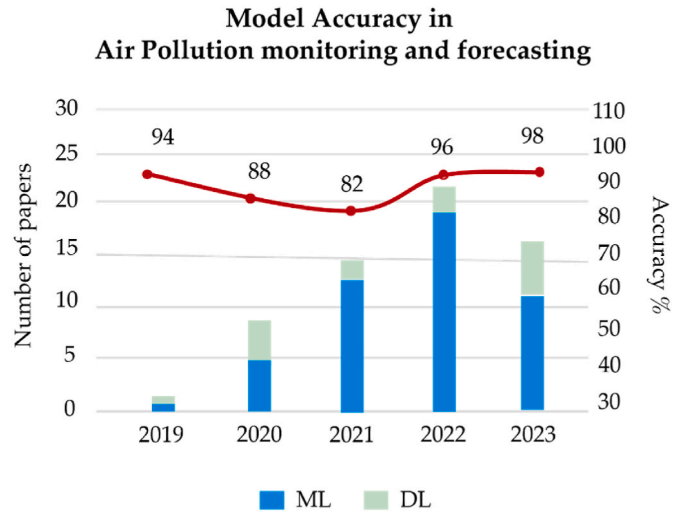

Fig. 6. Evolution of model accuracy and number of studies over time.

forecasting and monitoring over the past five years. The highest accuracy recorded was observed in ML-based study in 2023. The RF method is the most frequent and popular ML method, with 98.2 % accuracy for automated air pollution forecasting and monitoring conducted by El-Magd et al. (Abu El-Magd et al., 2023). This work considers only one contaminant (PM_{10}), rather than a mixture of contaminants, as in most of the studies conducted.

Table A1 provides a comprehensive summary of the studies on ML techniques for air pollution forecasting and monitoring, while Table A2 presents a detailed compilation of the studies that focus on DL methods. Both tables serve as valuable references, offering a consolidated view of the different ML-based and DL-based methods utilized in air pollution forecasting and monitoring. They provide researchers and practitioners with a comprehensive resource for exploring and comparing the diverse approaches employed in this field.

In our analysis, we classify AI applications in air pollution research into three distinct areas.

- Prediction (forecasting): This field focuses on forecasting future concentrations of air pollutants using historical data. To predict pollution levels hours or days in advance, these models usually use temporal sequences of historical pollution measurements, meteorological data, and other pertinent variables. For instance, Wang et al. (2022a) forecasted pollution levels 24–48 h in advance using historical $PM_{2.5}$ data and meteorological parameters.
- Monitoring: This area involves real-time or near-real-time assessment of current air quality conditions using contemporaneous data. in contrast to prediction models, monitoring applications estimate current pollution levels using measurements of co-variates (such as traffic density, weather, or satellite data). For example, Adams et al. (2020) created models to estimate real-time $PM_{2.5}$ concentrations across unmonitored areas using ground measurements and current satellite data.
- Impact Assessment: This area examines the relationships between air pollution and various outcomes, such as public health effects, economic consequences, or environmental impacts. These studies use AI to analyze historical data and identify correlations or causal relationships. For example, Meng et al. (2022) used ML to assess the relationship between $PM_{2.5}$ exposure and COPD exacerbations, while Zou et al. (2022) analyzed air pollution's impact on housing prices.

This classification helps organize the literature based on the primary objective and temporal focus of each study, though some research may span multiple areas. Fig. 7 shows the distribution of studies across these

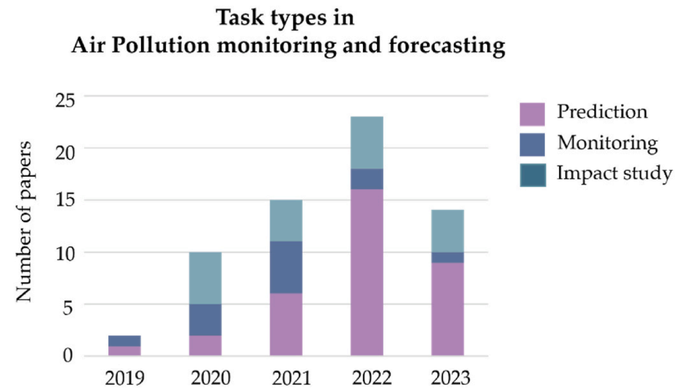

Fig. 7. Evolution of task trends (impact study, monitoring, prediction) in AI models for air pollutant analysis.

three areas, highlighting the evolution of research focus over time. In the early years, such as 2020, more studies emphasized monitoring current pollution levels and assessing the impacts of air quality on public health and the environment. However, beginning in 2021, there was a noticeable shift towards prediction-oriented research. Specifically, the figure indicates that in 2022, the majority (16) of related studies were focused on developing methods and models to predict future air pollutant concentrations based on factors like emissions, meteorology, and land use. Only two studies focused on sensor-based monitoring approaches, while 5 assessed the impacts of pollution exposure. This shift demonstrates the growing interest from researchers in applying ML and AI to develop predictive air quality systems.

Fig. 8 provides insight into the types of models used across the different tasks of prediction, monitoring, and impact assessment. It shows a clear prevalence of traditional ML algorithms over DL approaches. ML methods dominated all three tasks. Monitoring studies employed ML almost exclusively, with just one identified work leveraging a DL model. This single DL study focused on an integrated approach for air pollutant detection and sensor data prediction. Meanwhile, prediction and impact assessment saw a slightly higher application of DL techniques than monitoring, but ML still accounted for most models in these areas. This suggests that researchers have relied more heavily on traditional algorithms like regression, decision trees, and support vector machines that can effectively capture patterns in air quality datasets with fewer data requirements.

Fig. 9 provides insight into the distribution of air contaminant types studied across the surveyed literature. Most works (over 70%) investigated multiple pollutants simultaneously rather than focusing on a single contaminant. The most examined joint pollutants were particulate matter indicators like $PM_{2.5}$ and PM_{10} and gaseous pollutants such as CO_2, SO_2, and NOx. Among individual contaminants, particulate matter $PM_{2.5}$ and PM_{10} received the highest research focus, with over half of all

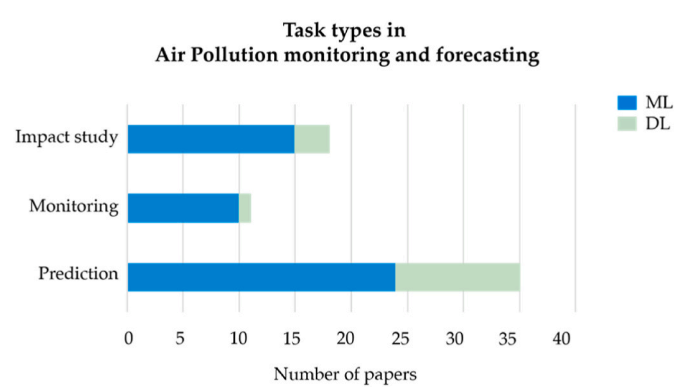

Fig. 8. Tasks executed by AI models divided by ML and DL techniques.

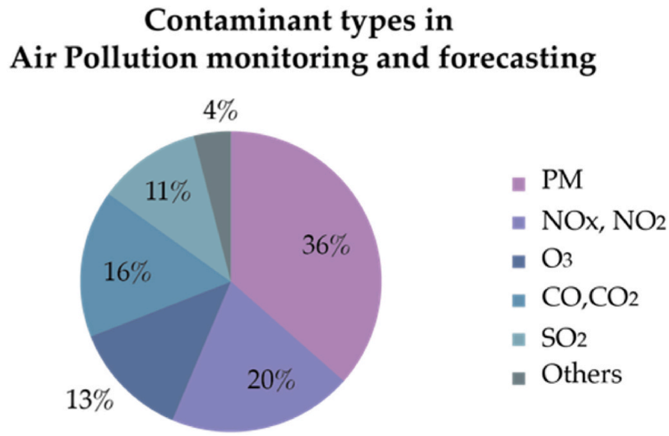

Fig. 9. Air pollutants studied in the works reviewed.

papers analyzing these air contaminants.

4. Discussion

The harmful health impact of air pollution has driven researchers worldwide to advance air quality monitoring and forecasting. AI methods have proven highly effective for those efforts. AI techniques have witnessed a remarkable rise, increasing from just two studies in 2019 to 22 in 2022, as shown in Fig. 5. By early 2023, 16 more publications contributed to this rapid growth, considering only relevant Q1 journal papers. The total number of publications has risen in the past three years, totaling nearly 120, with 65 Q1 publications featured in this review. These trends reflect the growing recognition that AI has the potential to transform global air pollution monitoring and forecasting. The rapid expansion underscores the appeal of this technology for tackling the challenge of understanding and forecasting air quality. Continued growth is expected as researchers leverage AI innovations, from ML interpretability to DL spatiotemporal modeling. Improved pollution forecasting could help protect even more individuals from the adverse health impacts of air pollution.

4.1. Air quality monitoring using AI methods

Monitoring air quality in urban areas is inherently challenging due to the complex interplay of weather, topology, and seasons that impact pollution levels. Advanced techniques that integrate wireless sensor networks with ML models have proven effective. Rosero-Montalvo et al. (Rosero-Montalvo et al. (2022) achieved 96% accuracy in monitoring NO and CO using optimally positioned sensors and a neural network algorithm. A comparable approach combining sensors and artificial neural networks (ANN) yielded an R^2 of 0.78 for CO and NO_2. However, model accuracy depends on the number of deployed sensors, requiring optimization between sensor networks and ML tools to enhance efficiency, as demonstrated by Zhao et al. (2021). Recent advances in data aggregation techniques using optimization algorithms (Heidari et al., 2024) could improve the integration and processing of data from distributed air quality sensor networks.

For risk monitoring and early warning, Xie et al. (2021) developed a hybrid deep learning sequential Concentration Transport Emission Model (DL-CTEM) to monitor $PM_{2.5}$, PM_{10}, SO_2, CO, NO_2 and O_3 with 95.6% accuracy. Satellite imagery integrated with AI and ground data can track pollution levels and seasonal impacts. Song et al. (2021a) effectively applied this with a RF model to monitor diverse pollutants. Geospatial techniques add flexibility, with Stacked Ensemble Modeling achieving high accuracy in Adams et al. (2020).

Assessing the influence of multiple factors using AI aids in real-world monitoring. Li et al. (2023a) applied RF on hourly data to evaluate pollution drivers related to fuel combustion with 94% accuracy for $PM_{2.5}$ prediction. Numerous studies leveraged AI to model the COVID-19 lockdown's pollution impact, including Wijnands et al. (2022) and Habeebullah et al. (2022), again demonstrating RF's reliability despite complex inputs. Beyond health, Zou et al. (2022) applied ML to examine air pollution's economic impact on housing prices.

In summary, multivariate AI techniques show tremendous promise for unraveling the intricacies of real-world air quality monitoring. Hybrid approaches that optimize sensor networks leverage satellite imagery, and evaluate diverse pollution factors enable robust, accurate systems for tracking hazardous pollution. Air quality modeling has long provided valuable insights to support public health policies and inform populations. However, as research continues, AI-enabled monitoring could offer additional actionable intelligence to strengthen these efforts and better protect public health.

4.2. Air quality forecasting using AI methods

Fine particulate matter exacerbates chronic obstructive pulmonary disease (COPD), significantly increasing mortality risk. A KNN model by Meng et al. (2022) (demonstrated a 74% accuracy in showing that $PM_{2.5}$ emission reductions could protect non-smoking COPD patients. Conibear et al. (Wijnands et al., 2022) also used ML emulators to predict over 99% of the variance in $PM_{2.5}$ and ozone concentrations, along with associated health impacts from emission changes across five key sectors.

Remote sensing and GIS data on factors like morphology and geo-information aid monitoring. El-Magd et al. (Abu El-Magd et al., 2023) leveraged Landsat spectral bands and land use indices as inputs, providing pollution distribution insights. Landsat bands and land use indexes assessed PM_{10} distribution. Song et al. (2021a) integrated satellite, reanalysis, and ground data into a ML model examining seasonal and annual pollution changes from 2018 to 2019. However, substantial data is needed to fully exploit deep learning possibilities, with RF proving a reliable geospatial forecasting model.

Huang et al. (Habeebullah et al., 2022) found shallow neural networks optimal for building morphology and pollutant concentration modeling, achieving a $R^2 = 0.94$ for $PM_{2.5}$. Mobile monitoring data tends to be less stable and challenging to model realistically. While mobile monitoring data tends to be less stable and challenging to model realistically, shallow neural networks demonstrated superior performance compared to conventional dispersion models, providing faster and more accurate predictions of pollution diffusion patterns in urban environments.

Traditional ML prioritizes overall accuracy, neglecting peak pollution values and complex factor interplay. Addressing this, Bai et al. (2022) developed a hybrid extreme learning machine with multi-objective optimization for prediction. This improved deterministic accuracy to 71.52% and enabled effective pollutant concentration interval forecasts.

In summary, multivariate techniques leveraging diverse data sources provide insights into spatiotemporal pollution distributions and associated health impacts. While shallow neural networks show promise for focused applications, RF excels for geospatial forecasting given sufficient data. Advances in hybrid AI now allow accurate peak pollution prediction while capturing intricate relationships. Such comprehensive, reliable modeling empowers targeted, effective interventions.

4.3. AI technique selection based on pollutant characteristics

Our analysis reveals that the effectiveness of AI techniques varies depending on the spatial and temporal characteristics of specific pollutants. This underscores the importance of tailoring the choice of AI method to the nature of the pollutant being studied.

For pollutants exhibiting strong seasonal patterns, such as $PM_{2.5}$ and PM_{10}, ensemble methods like RF have demonstrated efficacy (Abu El-Magd et al., 2023), (Zhang et al., 2023a). These methods excel at capturing non-linear relationships and handling the multivariate nature of seasonal pollution data. For instance, Abu El Magd et al. (Abu El-Magd et al., 2023) achieved 98.3% accuracy using RF for PM_{10} prediction.

In contrast, when dealing with pollutants characterized by complex temporal dynamics, such as O_3 and NO_2, DL approaches, particularly LSTM networks, have shown superior performance (Sun et al., 2022), (Dairi et al., 2021). Sun et al. (2022) found that a hybrid WRF-CMAQ LSTM model reduced prediction errors for $PM_{2.5}$ and O_3 by 40% and 20%, respectively, compared to traditional models. The ability of these models to capture long-term dependencies in time series data makes them well-suited for predicting pollutants with intricate temporal patterns.

For pollutants with significant spatial variability, such as ultra-fine particles, techniques that can incorporate geospatial data have proven advantageous. Convolutional Neural Networks (CNNs), especially when combined with satellite imagery, have shown promise in capturing the spatial distribution and variation of these pollutants (Liang et al., 2023), (Steininger et al., 2020). Liang et al. (2023) developed a multi-scale, attention-enhanced CNN architecture that achieved an R^2 of 0.94 for $PM_{2.5}$ prediction using geospatial features.

This observed relationship between pollutant characteristics and AI model performance highlights the need for a nuanced approach in air quality modeling. Researchers and practitioners should consider not only the overall performance metrics of AI models but also how well these models align with the specific spatial and temporal characteristics of the pollutants under investigation.

4.4. Health impacts

Air pollution can have far-reaching health impacts with both acute and chronic effects. The size and composition of particles determine toxicity and health consequences. Larger PM10 particles deposit in the upper airways while finer $PM_{2.5}$ penetrates deeper into the alveolar region, increasing cardiopulmonary risks. Absorption into the bloodstream can lead to ischemic heart disease, myocardial infarction, and cerebrovascular disease. Solubility also affects distribution; highly soluble SO_2 absorbs in the upper airways, exacerbating respiratory illness, while less soluble NO_2 reaches deeper regions, impairing lung development and increasing influenza susceptibility in children. High CO levels are fatal, and moderate exposure impairs vision and coordination.

Air pollution from various sources poses significant health risks across different regions. Dust storms in dry regions, forest fires, transboundary pollution events, and urban air pollution from industrial and automobile emissions are major contributors. Every source has different health risks. Increased respiratory and cardiovascular morbidity is linked to urban air pollution, especially from industrial and transportation sources (Chen et al., 2017). Long-distance movement of tiny particulate matter by dust storms can exacerbate pre-existing diseases like asthma and cause acute respiratory symptoms (Achilleos et al., 2014). $PM_{2.5}$, CO, and volatile organic compounds are among the complex mixture of contaminants found in wildfire smoke that can have both short-term and long-term negative health impacts (Reid et al., 2016). In order to track these many pollution sources and give susceptible populations early warnings, it is now essential to create AI-based monitoring and forecasting systems. More focused public health measures are now possible because of recent research showing how well machine learning algorithms anticipate pollution levels from a variety of sources (Agbehadji and Obagbuwa, 2024), (Zhang et al., 2024).

While many studies focus on single pollutants, the real-world scenario involves simultaneous exposure to multiple air pollutants. AI techniques can be used to understand these intricate multi-pollutant interactions and their combined health effects. For example, Usmani et al. (2023) employed a Long Short-Term Memory model to analyze the combined effects of multiple pollutants (CO, O_3, NO, NO_2, NO_x, SO_2, and PM_{10}) on cardiorespiratory mortality, achieving high prediction accuracy. Similarly, Sun et al. (2023) developed a ML-based early warning model for mixed exposure to multiple air pollutants, demonstrating 94.5% sensitivity in predicting respiratory disease mortality. These AI approaches can disentangle the relative contributions of different pollutants to specific health outcomes, enabling more targeted intervention strategies. Ravindra et al. (2023) utilized RF models to evaluate how combinations of primary and secondary pollutants affect respiratory hospital admissions, achieving an R^2 of 0.872. Such multi-pollutant analyses provide a more realistic assessment of environmental health risks and can better inform public health policies than single-pollutant studies.

In summary, AI techniques such as deep neural networks now offer new insights which can help with understanding the complex relationships between various pollution exposures and related health effects, enabling focused interventions. In this regard, risk assessment and vulnerable population protection may be enhanced by AI-enabled pollution-health impact modeling.

4.5. Challenges and limitations

Despite the significant advancements in AI-based air pollution monitoring and forecasting, several challenges persist. Data quality and availability remain major hurdles, particularly in developing countries where air quality monitoring infrastructure is limited. The heterogeneity of data sources, including ground-based sensors, satellite imagery, and meteorological data, poses integration challenges for AI models. Moreover, the interpretability of complex deep learning models is still a concern, especially when these models are used to inform policy decisions. The generalizability of AI models across different geographical regions and climate conditions is another ongoing challenge, as models trained on data from one area may not perform well in others due to varying pollution sources and weather patterns. Additionally, the computational resources required for training and deploying sophisticated AI models can be a barrier to widespread adoption, particularly in resource-constrained environments. Similar challenges in implementing distributed AI systems have been documented in other domains (Aminizadeh et al., 2024), highlighting common issues of data integration, system reliability, and resource optimization.

The environmental impact of implementing AI, especially for DL models, is another major limitation. Significant energy and computational resources are needed to train these models, which may exacerbate the environmental problems they are intended to solve. According to recent research, the lifetime carbon emissions of training a large deep learning model can equal the carbon emissions of five cars (Strubell et al., 2020). Implementing lightweight architectures and model compression strategies, making use of green computing infrastructure, and using transfer learning to eliminate the need for training from scratch are some strategies that can lessen these environmental costs (Howard et al., 2019-). These tactics can assist in striking a balance between the advantages of AI deployment for bettering air quality monitoring and management and its negative effects on the environment. The challenge lies in developing more energy-efficient algorithms while maintaining high prediction accuracy and model performance.

Data integration poses significant challenges in AI-based air pollution monitoring. While diverse data sources (ground sensors, satellite imagery, meteorological data, traffic monitors) can enhance model performance, integrating these heterogeneous data streams presents several technical hurdles. Different temporal and spatial resolutions among data sources require preprocessing techniques. For example, satellite data typically provides broad spatial coverage but limited temporal frequency, while ground-based sensors offer high temporal resolution but sparse spatial coverage. Researchers have addressed this through various approaches: Wang et al. (2022a) employed a multi-scale fusion network to combine different resolution data streams, achieving a 27.65% improvement in RMSE compared to single-source models. Time series alignment techniques and interpolation methods have harmonized data with different sampling frequencies. Li et al. (2023a) demonstrated success using a hierarchical data fusion approach that achieved 94% accuracy in $PM_{2.5}$ prediction by combining hourly ground measurements with daily satellite observations.

Despite these promising results, there are issues with managing missing data, resolving measurement uncertainties from various sources, and preserving data quality throughout integration. Additionally, the computational overhead of processing and integrating multiple data streams in real time can impact model deployment, especially in resource-constrained environments.

4.6. Real-world applications

Several notable implementations demonstrate the practical impact of AI in air quality monitoring and management. In Hong Kong, Li et al. (2021) created a comprehensive smart air quality monitoring system by fusing deep learning models with Internet of Things sensors. Through a mobile application, their Grid-LSTM framework enabled personalized health recommendations while achieving 82% accuracy in real-time pollution prediction. Better public health awareness and more efficient pollution exposure management resulted from the system's successful implementation. In Vietnam, Rakholia et al. (2023) used a multi-output ML model to forecast regional air pollution. Their system produced precise predictions for several pollutants at once by combining meteorological parameters with data from 18 monitoring stations. During periods of high pollution, the implementation assisted local authorities in modifying traffic management plans and promptly issuing air quality warnings. A particularly successful application in China by Wang et al. (2022b) demonstrated how supportAI can support policy decisions. Compared to conventional techniques, their hybrid GWO-LSTM model increased prediction accuracy by more than 30%, allowing authorities to enact focused emission control measures. Measurable improvements in urban air quality resulted from the system's predictions, which assisted in identifying the main sources of pollution and refining intervention tactics. In resource-constrained environments, Ali et al. (2021a) developed a cost-effective solution combining low-cost sensors with AI-based calibration in Malaysia. Their ANN-based system demonstrated how AI can improve the dependability of low-cost monitoring networks by achieving an R^2 of 0.78 for CO and NO_2 monitoring. Cities with inadequate monitoring infrastructure were given a scalable solution by this implementation.

4.7. Future research directions

As AI continues to revolutionize air pollution monitoring and forecasting, several key areas emerge as critical for future research. Future research should focus on developing AI models capable of integrating diverse data sources beyond traditional air quality sensors. This could include satellite imagery, social media data, traffic patterns, and citizen-reported observations. The challenge lies in harmonizing these heterogeneous data types and dealing with varying temporal and spatial resolutions. For instance, researchers could explore DL architectures that simultaneously process satellite imagery, ground-based sensor data, and text-based reports to provide a more holistic view of air quality. This multi-modal approach could reveal complex interactions between pollution sources, meteorological conditions, and human activities that are currently overlooked. Moreover, this research direction could benefit from advancements in natural language processing to interpret unstructured data from social media and citizen reports, potentially providing early warning signals of air quality issues before they're detected by traditional sensors.

Explainable AI (Seoni et al., 2023) is a potential area for further research in air quality assessment. While current AI models offer high predictive accuracy, their 'black box' nature often makes it difficult for the endusers to trust and act on their predictions (Loh et al., 2022). Several interpretability techniques have been successfully implemented in recent air pollution studies. For example, Li et al. (2023c) used SHAP (SHapley Additive exPlanations) analysis to identify meteorological conditions and the previous day's pollution levels as key predictors of $PM_{2.5}$ concentrations. Stirnberg et al. (2021a) showed how well LIME (Local Interpretable Model-agnostic Explanations) worked to identify local emission sources and explain particular cases of elevated PM1 concentrations. Additionally, Wang et al. (2022b) demonstrated in their LSTM-based forecasting system that attention mechanisms in deep learning models have yielded important insights into temporal patterns of pollution. Zhang et al. (2023b) have also found partial dependence plots to be helpful in visualizing the correlations between $PM_{2.5}$ levels and meteorological parameters. The goal of these approaches is to provide clear, intuitive explanations for why a model predicts high pollution levels on a given day, highlighting the relative importance of different factors (e.g., traffic patterns, industrial activities, weather conditions). Such explainable models could significantly enhance the actionability of AI predictions, allowing policymakers to implement targeted interventions and helping the public understand and mitigate their exposure to air pollution.

Uncertainty quantification (Seoni et al., 2023) in AI predictions represents another critical area for future research in air quality assessment. While current models demonstrate high accuracy in point estimates, the reliable quantification of prediction uncertainties remains challenging yet crucial for effective decision-making. Recent studies have shown promising directions: Bayesian deep learning approaches have successfully provided probability distributions rather than single-point predictions, with Han et al. (2022a) demonstrating error reductions up to 12.4% for $PM_{2.5}$ and PM_{10} predictions while quantifying prediction uncertainties. However, several challenges persist in distinguishing between different uncertainty sources: measurement errors in input data, model structural uncertainties, and parameter uncertainties. Future research should focus on developing computationally efficient methods that can quantify both epistemic uncertainty (model uncertainty) and aleatoric uncertainty (inherent data noise) in real-time applications. This is particularly important in air quality monitoring, where reliable uncertainty estimates could improve public health advisories and support more informed policy decisions.

Future research should also explore the development of dynamic, self-adapting AI models that can continuously learn and adjust to changing environmental conditions. In fact, climate change may alter weather patterns and potentially affect air pollution dynamics, causing static AI models to become less effective over time. The design of dynamic AI models could involve online learning algorithms that update in real-time as new data becomes available or meta-learning approaches that can quickly adapt to new scenarios with minimal additional training. These adaptive models could be crucial for maintaining accurate air quality predictions in the face of climate change-induced shifts in weather patterns and pollution dynamics. Researchers might also explore the use of reinforcement learning techniques to create AI agents that can proactively suggest optimal air quality management strategies under different climate scenarios. Recent advances in cloud-based IoT environments using optimization algorithms (Vakili et al., 2024), (Darbandi et al., 2018), (Darbandi, 2017) could enhance the processing efficiency and scalability of large-scale air quality monitoring systems.

Novel applications of AI in air pollution monitoring and forecasting present exciting opportunities for advancement. One unexplored area is the integration of AI with blockchain technology to create a decentralized system for air quality data collection and sharing. This could enhance data reliability and transparency, which is crucial for both research and policy-making. Another innovative application could be the development of AI-powered personal air quality assistants that combine real-time pollution data with individual health information to provide personalized recommendations for outdoor activities and route planning. To address current limitations, researchers could explore using federated learning techniques to develop robust models that can learn from distributed datasets without compromising data privacy or security. This approach could be particularly valuable in overcoming data-sharing barriers between different regions or countries. In public health, AI-driven air quality management systems could be integrated with healthcare databases to provide early warnings to vulnerable populations and help hospitals prepare for potential increases in respiratory-related admissions.

5. Conclusion

Rapid urbanization and industrialization have led to major air pollution problems worldwide. Air quality monitoring and forecasting are challenging due to complex interactions between diverse contributing elements. Recent advances in AI methodologies show promise in improving predictions fueled by extensive datasets, computing capabilities, and recognition of AI's advantages. Our analysis shows that the research was concentrated in India, China, and the US.

This systematic review has analyzed the state-of-the-art AI applications for air pollution monitoring and prediction, revealing key trends, challenges, and future directions. The field has shifted towards ML and DL techniques, with research papers increasing notably from 2021. Random Forest has emerged as the most effective method, achieving up to 98.2% accuracy. DL techniques, while less common, show promise in capturing complex spatiotemporal relationships.

Challenges include the 'black box' nature of many AI models, posing difficulties for interpretability and trust. Harmonizing heterogeneous data types and varying resolutions remain hurdles in developing comprehensive models. Adapting AI models to account for dynamic pollution patterns, especially considering climate change, is an ongoing challenge.

Future AI applications are promising, including multi-modal models integrating diverse data sources, explainable AI techniques enhancing trust and actionability, and dynamic, self-adapting models. Interdisciplinary collaboration between AI researchers, environmental scientists, and policymakers will be crucial in translating these advancements into tangible improvements in air quality and public health. To guide future research, we propose the following key questions and hypotheses.

1. How can AI models be designed to effectively capture the long-term impacts of climate change on air pollution patterns?
2. Can federated learning techniques be employed to develop globally applicable air quality models while respecting data privacy concerns?
3. Hypothesis: Integrating social media data and citizen science initiatives into AI models will significantly improve the accuracy of hyperlocal air quality predictions.
4. How can AI be leveraged to optimize the placement and operation of air quality monitoring sensors, especially in resource-constrained environments?

CRediT authorship contribution statement

Sreeni Chadalavada: Writing – original draft, Formal analysis, Data curation. **Oliver Faust:** Writing – original draft, Methodology, Formal analysis. **Massimo Salvi:** Writing – original draft, Visualization. **Silvia Seoni:** Writing – review & editing, Visualization. **Nawin Raj:** Writing – review & editing. **U. Raghavendra:** Writing – review & editing. **Anjan Gudigar:** Writing – review & editing. **Prabal Datta Barua:** Writing – review & editing. **Filippo Molinari:** Writing – review & editing, Supervision. **Rajendra Acharya:** Writing – review & editing, Supervision, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The open-access fee was supported by Politecnico di Torino through the CRUI-CARE & CNR agreement.

APPENDIX

Table A1
Summary of studies that ML approaches for air pollution forecasting and monitoring.

Author, year	Contaminants	Features	Classifier	Findings/Results (%)
(Meng et al., 2022)	PM2.5	Age, sex, smoking status, Group ABCD, FEV1, CAT score, and RNA profile	KNN	Sen: 46%, Acc: 74.3%
(Conibear et al., 2021)	PM2.5 and ozone (O3)	Mean PM2.5 concentrations and O3 exposure	ML Emulator	The emulators predicted 99.9% of the variance in PM2.5 and O3 concentrations.
(Li et al., 2023a)	PM2.5	PM2.5, NO2, SO2, CO, and O3	RF	Prediction of PM2.5 concentration: RMSE and MAE of 9.4 μg/m3 and 5.7 μg/m3, respectively.
(Rosero-Montalvo et al., 2022)	NOx and CO	NOx, CO, UV, HUM, TEMP	ANN	Acc: 96%
(Huang et al., 2022)	PM2.5 and PM10	influence indicators, including vehicle speed, temperature, relative humidity, and dew point temperature, were used as feature variables.	Shallow neural network	R2 = 0.94 for PM2.5 and R2 = 0.89 for PM10 for all training samples.
(Rakholia et al., 2023)	NO2, SO2, O3, and CO	NO2, CO, O3, SO2, Humidity, Temperature, Visibility, Pressure, uvIndex, Wind speed, Cloud cover, Dew point Hour	N-BEATS	
(Bai et al., 2022)	PM2.5 concentration	Daily PM2.5 concentration series of three major cities	Hybrid model based on ML	Acc: 71.5%
(Abu El-Magd et al., 2023)	PM10	LST, Landsat bands, SAVI, distance from the road and distance from the station	RF	Acc: 98.3%
(Gu et al., 2022)	PM 2.5	Automatic feature generation and feature selection procedures	Hybrid Interpretable Predictive ML model.	The accuracies of 1, 3 and 6-h-ahead prediction are 98.7%, 0.93.3% and 85.8%, respectively

(continued on next page)

Table A1 (*continued*)

Author, year	Contaminants	Features	Classifier	Findings/Results (%)
(Heydari et al., 2022)	NO2 and SO2	The datasets include wind speed, air temperature, NO2, and SO2 for five months.	LSTM and multi-verse optimization algorithm	RMSE: 0.0310; MAE: 0.0263; MAPE: 17.2091
(Ravindra et al., 2023)	Primary and secondary pollutants	Concentration of air pollutants	RF	Best performance with R2 = 0.872, 0.871 without lag and 1-day lag, respectively on total patients
(Kumar and Pande, 2023)	PM2.5, PM10 NO, NO2, NOX, NH3, CO, SO2, and O3 Benzene Toluene	PM2.5, PM10 NO, NO2, NOX, NH3, CO, SO2, and O3 concentrations	XBoost	Acc: 91%
(Sun et al., 2023)	Warning model of mixed exposure to air pollutants	30 variables including death date; age; sex; daily average temperature; average relative humidity; NO2; PM10, PM2.5 SO2 and CO and O3	XBoost	Sen: 94.5%; Spe: 89.6%; Acc: 98.0%; AUC: 94.4%
(Ku et al., 2022)	PM2.5, PM10, O3, NO2, CO, and SO2	Climatic factors and air-pollution indicators, including PM2.5, PM10, O3, NO2, CO, and SO2,	XGBoost and GPR-based ML models	R2 values of over 0.67 and RMSE values below 13.9 per 10,000 inhabitants.
(Lai et al., 2022)	CO, O3, and NO2	Concentration of air pollutants	Ensemble model	The R2 value of the CO gas model is 0.73; it reached 0.51 for the O3 gas model and 0.37 for the NO2 gas model.
(Zhang et al., 2023a)	PM2.5	PM2.5, PM10, CO, NO2, O3, and SO2	RF	–
(Fang et al., 2022)	PM2:5, CO, SO2, and NO2	PM2:5, CO, SO2, and NO2	Regional feature selection-based ML (RFSML)	Acc: 0.72
(Zhou et al., 2022)	PM2.5	Aerosol and gaseous species (Org, SO42, NO3, NH4+, Cl, nc-POA, COA, SOA, SO2, NO2, CO and O3)	RF	–
(Almalawi et al., 2022)	CO2, SO2, NO2, and atmospheric PM	CO2, SO2, NO2, and atmospheric PM2.5 and PM10	GBDT ensemble	RMSE MAE MSE (04.146 5.209 3.135)
(Khan et al., 2022)	PM2.5	PM2.5	RF	R2 MAE RMSE (0.94 11.33 22.77)
(Wan et al., 2022)	PM2.5, PM10, SO2, CO, NO2, and O3	PM2.5, PM10, SO2, CO, NO2, and O3	RF	AUROCs by 0.712 up to 0.771
(Zou et al., 2022)	SO2, NO2, and PM10	SO2, NO2, and PM10	GBDT	The optimal R2, MAE, and RMSE values obtained by the GBDT model are 10.320, 7.087, and 0.975, respectively.
(Liu et al., 2022)	PM2.5	PM2.5, SO2, NO2, and CO	RF	–
(Tao et al., 2023)	PM2.5	Meteorological and soil parameters	LSTM integrated with Bayesian optimizer	RMSE = 13.4%, Kling-Gupta efficiency = 0.89
(Wijnands et al., 2022)	NO2, PM10, PM2.5 and O3	Total precipitation, mean temperature and mean solar radiation over three days.	XBoost	
(Habeebullah et al., 2022)	O3, NO2, and PM10	O3, NO2, and PM10	Supervised ML	
(Ji et al., 2023)	NO2, CO, and PM (PM10 and PM2.5)	NO2, CO, PM10, and PM2.5	RF	NO2, CO, O3, PM2.5, and PM10 are significantly correlated with the numbers of clinic visits (PCC:0.35)
(Ren et al., 2020)	PM2.5	PM2.5	Rapidly exploring random tree star (RRT*)	R2 value equal to 0.9459
(Liang et al., 2023)	PM2.5	The traffic, LULC, and PA features are the most informative, followed by LiDAR tree heigh	Multi-scale, attention-enhanced CNN architecture	The R2 and RMSE values for this model were 0.94 and 1.34 μg/m3, respectively
(Xie et al., 2022)	PM2.5, PM10, SO2, CO, NO2, and O3	Concentration of air pollutants	Bayesian neural network	95.6% prediction accuracy
(Hashad et al., 2021)	No specific pollutant, in general	Vegetation dimension (width or height) and particle size were the top two selected features	RF, ANN, and XGB	RF, ANN, and XGB performed well with a normalized RMSE of 6–7% and an average test R2 value > 0.91
(Ali et al., 2021b)	CO and NO2 concentrations and an infrared sensor to measure PM levels	Raw air pollutant concentration, temperature, and humidity correction	ANN-based calibration method	MAPE of 38.89% and R2 of 0.78
(Lautenschlager et al., 2020)	NOx or PM	Industry usage, commercial usage, residential usage, light/heavy traffic	RFOstochastic	RMSE of 1.74 and R2 equal to 0.19
(Lovrić et al., 2021)	NO2, PM10, O3 (ozone), and Ox (total oxidant)	City average concentrations of PM10 and NO2 concentrations	RF	R2: 0.84
(Cole et al., 2020)	SO2, NO2, CO, and PM10	Temperature, wind direction, wind speed, and atmospheric pressure	RF	–
(Delavar et al., 2019)	PM10 and PM2.5	Meteorological parameters, topography, and pollutant concentrations.	Autoregressive nonlinear neural network	Acc: 94%
(Stirnberg et al., 2021b)	PM1	Concentration of air pollutants	Tree-based ML	R2 is 0.58
(Mele and Magazzino, 2020)	CO2, PM2.5	Concentration of air pollutants	LSTM	CO2 PM2.5:0.92111 0.88457
(Razavi-Termeh et al., 2021)	CO, PM10,PM2.5, NO2, SO2, and O3	Season information and concentrations of CO, PM10 and PM2.5, NO2, SO2, and O3	RF	Acc: 0.823, 0.821, 0.83, and 0.827 for three seasons
(Song et al., 2021a)	PM2.5, PM10, O3, and CO	Meteorological factors and concentration of air pollutants	RF	RMSE for the PM2.5, PM10, O3, and CO validation dataset were 9.027 g/m3, 20.312 g/m3, 10.436 g/m3, and 0.097 mg/m3, respectively
(Cazzolla Gatti et al., 2020)	PM2.5	PM10, O3, SO2, NO2, CO, and Benzene	RF	R2 is 0.95 and RMSE is 28.9

(continued on next page)

Table A1 (*continued*)

Author, year	Contaminants	Features	Classifier	Findings/Results (%)
(Zhao et al., 2021)	PM2.5	Land use, characteristics, traffic, and road network, as well as external factors such as meteorological information	XBoost	R2 is 0.80 and RMSE is 8.1
(Ren et al., 2020)	CO, SO2, PM10, PM2.5, NOx, VOC, NH3	Deciduous forest coverage, longitude, daily maximum relative humidity, wind speed, and wind direction	XBoost	R2 is: 0.96 RMSE is 2
(Adams et al., 2020)	PM2.5	Data source, attribute name, units, and tapered elemental oscillating microbalance measurement	Stacked Ensemble Model	R2 is from: 0.8–0.99
(Du et al., 2021)	PM2.5, PM10, CO, NO2, O3, and SO2	PM2.5, PM10, CO, NO2, O3, and SO2	Time-Varying Filter Empirical Mode Decomposition	MAPE is 2.03%
(Xiao et al., 2019)	General	Concentration of air pollutants	A parameterized non-intrusive reduced order model (P-NIROM)	RMSE is 8%
(Li et al., 2021)	PM2.5, PM1.0, and NO2	Concentration of air pollutants	Grid-LSTM	Acc: 82%
(Song et al., 2021b)	PM2.5	Wind speed, pressure, water vapor pressure, temperature, and humidity	ML framework (Deep-MAPS)	SMAPE< 15%
(Usmani et al., 2021)	PM10, CO, NOx, NO2, NO, and SO2	PM10, CO, NOx, NO2, NO, and SO2	ELSTM model	RMSE: 0.002

Acc: accuracy; Sen: sensitivity; Spe: specificity; RMSE: root mean square error; MAPE: mean absolute percentage error.

Table A2
Summary of studies that DL approaches for air pollution forecasting and monitoring.

Author, year	Contaminants	Features	Classifier	Findings/Results (%)
(Usmani et al., 2023)	CO, O3, NO, NO2, NOx, SO2, and PM10	AQM Station, Date, cardiorespiratory mortal. it count, and the air pollutants, i.e., CO, O3, NO, NO2, NOx, SO2, and PM10.	Enhanced LSTM	ELSTM: 0.004 (in terms of RMSE and MAE
(Shu et al., 2023)	PM2.5, PM10, SO2, NO2, CO, and O3	PM2.5, PM10, SO2, NO2, CO, and O3	Proposed DW-CAE model is more accurate than other baseline models	Acc: 93%
(Sun et al., 2022)	PM2:5 and O3	Ground observation data for different source stations	Hybrid WRF-CMAQ LSTM	RMS is 40% and 20% lower than the Weather Research and Forecasting–Community Multiscal. e Air Quality model
(Zhu et al., 2018)	Temperature, CH4, NO, CO, relative humidity, NMHC (non-methane hydrocarbons), PM10, NOx, O3, SO2, PM2.5, and NO2	Temperature, CH4, NO, CO, relative humidity, NMHC (non-methane hydrocarbons), PM10, NOx, O3, SO2, PM2.5, and NO2	Conv. LSTM	Acc: 91%
(Han et al., 2022b)	PM2:5 and PM10	PM2.5 and PM10	Bayesian deep-learning model	DL model can reduce the prediction errors by a maximum of 3.7% and 12.4%
(Xie et al., 2021)	Ammonia (NH3), CO2, and hydrogen sulphide (H2S)	NH3, CO2, and H2S	Driven sequential Concentration Transport Emission Model (DL-CTEM)	Mean errors: 0.1 ppm, 79.2 ppm, and 106.3
(Xie et al., 2022)	NH3, CO2, and H2S	Gas concentrations and air pollution from pig buildings	DL-CTEM	R2 for concentrations of NH3, CO2, and H2S at PLS and pit in both winter and summer were 0.8875, 0.7808, and 0.6335, respectively.
(Kabir et al., 2020)	PM2.5	Automatic feature extraction from outdoor images	BRB based CNN	Sen: 97.1%; Spe: 95.6%; AUC: 93.6
(Tao et al., 2019)	PM2.5	Meteorological data and air pollutant concentration	Convolutional-based bidirectional gated recurrent unit (CBGRU) model	RMSE as: 14.5 and MAE is: 10.48
(Dairi et al., 2021)	NO2, SO2, CO, and O3.	Concentration of air pollutants	Integrated multiple directed attention variational autoencoder (IMDA-VAE)	R2 is 0.87 and RMSE is 13.44
(Steininger et al., 2020)	NO2	Entity, area, and distance features	MapLUR	R2 is: 0.673 RMSE is 8.002
(Gao et al., 2021)	PM2.5, PM10, CO, NO2, O3, and SO2	Concentration of air pollutants	LSTM Forecast Model	Air quality evaluation score has improved from 0.3494 in 2015 to 0.4504 in 2019
(Hähnel et al., 2020)	NO2 and PM10	Pollution measurements, traffic data, weather data	Geometric DL	MAE of the DL computed values was 1.7μg/cm3
(Bekkar et al., 2021)	PM2.5	PM2.5, PM10, and CO	CNN-LSTM multivariate	MAE: 9.03; RMSE: 16.6; R2: 0.979
(Wang et al., 2022a)	PM2.5	PM2.5, PM10, SO2, CO, NO2 and O3 and meteorological data including wind speed, air temperature, and dew point	GWO-LSTM	Improves the MAE, RMSE and MAPE by 32.92%, 27.65% and 30.02%, respectively

Data availability

Data will be made available on request.

References

Abu El-Magd, S., Soliman, G., Morsy, M., Kharbish, S., 2023. Environmental hazard assessment and monitoring for air pollution using machine learning and remote sensing. Int. J. Environ. Sci. Technol. 20 (6), 6103–6116. https://doi.org/10.1007/s13762-022-04367-6.

Achilleos, S., Evans, J.S., Yiallouros, P.K., Kleanthous, S., Schwartz, J., Koutrakis, P., 2014. PM10 concentration levels at an urban and background site in Cyprus: the impact of urban sources and dust storms. J. Air Waste Manag. Assoc. 64 (12), 1352–1360. https://doi.org/10.1080/10962247.2014.923061.

Adams, M.D., Massey, F., Chastko, K., Cupini, C., 2020. Spatial modelling of particulate matter air pollution sensor measurements collected by community scientists while cycling, land use regression with spatial cross-validation, and applications of machine learning for data correction. Atmos. Environ. 230 (Jun). https://doi.org/10.1016/j.atmosenv.2020.117479.

Agbehadji, I.E., Obagbuwa, I.C., 2024. Systematic review of machine learning and deep learning techniques for spatiotemporal air quality prediction. Atmosphere 15, 1352. https://doi.org/10.3390/ATMOS15111352.

Ahmed, M., Seraj, R., Islam, S.M.S., 2020. The K-Means Algorithm: A Comprehensive Survey and Performance Evaluation. MDPI AG. https://doi.org/10.3390/electronics9081295.

Ali, S., Glass, T., Parr, B., Potgieter, J., Alam, F., 2021a. Low cost sensor with IoT LoRaWAN connectivity and machine learning-based calibration for air pollution monitoring. IEEE Trans. Instrum. Meas. 70. https://doi.org/10.1109/TIM.2020.3034109.

Ali, S., Glass, T., Parr, B., Potgieter, J., Alam, F., 2021b. Low cost sensor with IoT LoRaWAN connectivity and machine learning-based calibration for air pollution monitoring. IEEE Trans. Instrum. Meas. 70. https://doi.org/10.1109/TIM.2020.3034109.

Almalawi, A., et al., 2022. An IoT based system for magnify air pollution monitoring and prognosis using hybrid artificial intelligence technique. Environ. Res. 206 (Apr). https://doi.org/10.1016/j.envres.2021.112576.

Aminizadeh, S., et al., 2024. Opportunities and challenges of artificial intelligence and distributed systems to improve the quality of healthcare service. Artif. Intell. Med. 149, 102779. https://doi.org/10.1016/J.ARTMED.2024.102779.

P. Arden et al., "Lung Cancer, Cardiopulmonary Mortality, and Long-term Exposure to Fine Particulate Air Pollution." [Online]. Available: https://jamanetwork.com/.

Bai, L., Liu, Z., Wang, J., 2022. Novel hybrid extreme learning machine and multi-objective optimization algorithm for air pollution prediction. Appl. Math. Model. 106, 177–198. https://doi.org/10.1016/j.apm.2022.01.023.

Balogun, A.L., Tella, A., Baloo, L., Adebisi, N., 2021. A Review of the Inter-correlation of Climate Change, Air Pollution and Urban Sustainability Using Novel Machine Learning Algorithms and Spatial Information Science. Elsevier B.V. https://doi.org/10.1016/j.uclim.2021.100989

Bekkar, A., Hssina, B., Douzi, S., Douzi, K., 2021. Air-pollution prediction in smart city, deep learning approach. J Big Data 8 (1). https://doi.org/10.1186/s40537-021-00548-1.

Bell, M.L., Dominici, F., Samet, J.M., 2005. A Meta-Analysis of Time-Series Studies of Ozone and Mortality with Comparison to the National Morbidity, Mortality, and Air Pollution Study [Online]. Available: www.pubmed.com.

Bellinger, C., Mohomed Jabbar, M.S., Zaïane, O., Osornio-Vargas, A., 2017. A systematic review of data mining and machine learning for air pollution epidemiology. BioMed Central Ltd. https://doi.org/10.1186/s12889-017-4914-3.

Brook, R.D., et al., 2010. Particulate Matter Air Pollution and Cardiovascular Disease: an Update to the Scientific Statement from the American Heart Association. https://doi.org/10.1161/CIR.0b013e3181dbece1.

Bert Brunekreef and Stephen T Holgate, "Air pollution and health." [Online]. Available: http://europa.eu.int/comm/environment/docum/pos_paper.pdf.

Cazzolla Gatti, R., Velichevskaya, A., Tateo, A., Amoroso, N., Monaco, A., 2020. Machine learning reveals that prolonged exposure to air pollution is associated with SARS-CoV-2 mortality and infectivity in Italy. Environmental Pollution 267. https://doi.org/10.1016/j.envpol.2020.115471.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Association for Computing Machinery, pp. 785–794. https://doi.org/10.1145/2939672.2939785.

Chen, R., et al., 2017. Fine particulate air pollution and daily mortality. A nationwide analysis in 272 Chinese cities. Am. J. Respir. Crit. Care Med. 196 (1), 73–81. https://doi.org/10.1164/RCCM.201609-1862OC.

Chen, K.C., Tsai, S.W., Shie, R.H., Zeng, C., Yang, H.Y., 2022. Indoor air pollution increases the risk of lung cancer. Int J Environ Res Public Health 19 (3). https://doi.org/10.3390/ijerph19031164.

Choe, S.A., et al., 2018. Increased proportion of mature oocytes with sustained-release growth hormone treatment in poor responders: a prospective randomized controlled study. Arch. Gynecol. Obstet. 297 (3), 791–796. https://doi.org/10.1007/s00404-017-4613-4.

Cole, M.A., Elliott, R.J.R., Liu, B., 2020. The impact of the wuhan covid-19 lockdown on air pollution and health: a machine learning and augmented synthetic control approach. Environ. Resour. Econ. 76 (4), 553–580. https://doi.org/10.1007/s10640-020-00483-4.

Conibear, L., et al., 2021. Statistical emulation of winter ambient fine particulate matter concentrations from emission changes in China. Geohealth 5 (5). https://doi.org/10.1029/2021GH000391.

Cover, T.M., Hart, P.E., 1952. Approximate Formulas for the Information Transmitted Bv a Discrete Communication Channel.

Dairi, A., Harrou, F., Khadraoui, S., Sun, Y., 2021. Integrated multiple directed attention-based deep learning for improved air pollution forecasting. IEEE Trans. Instrum. Meas. 70. https://doi.org/10.1109/TIM.2021.3091511.

Darbandi, M., 2017. Proposing new intelligence algorithm for suggesting better services to cloud users based on kalman filtering. J. Comput. Sci. Appl. 5 (1), 11–16. https://doi.org/10.12691/JCSA-5-1-2.

Darbandi, M., Haghgoo, S., Hajiali, M., Khabir, A., 2018. Prediction and estimation of next demands of cloud users based on their comments in CRM and previous usages. Proceedings of the 2018 International Conference on Communication. Computing and Internet of Things, pp. 81–86. https://doi.org/10.1109/IC3IOT.2018.8668119.

Delavar, M.R., et al., 2019. A novel method for improving air pollution prediction based on machine learning approaches: a case study applied to the capital city of Tehran. ISPRS Int. J. Geo-Inf. 8 (2). https://doi.org/10.3390/ijgi8020099.

Du, Z., Heng, J., Niu, M., Sun, S., 2021. An innovative ensemble learning air pollution early-warning system for China based on incremental extreme learning machine. Atmos. Pollut. Res. 12 (9). https://doi.org/10.1016/j.apr.2021.101153.

E. C. C. and H. WHO TEAM Air quality and health, 2006. Air Quality Guidelines Global Update 2005.

Eftim, S.E., Samet, J.M., Janes, H., McDermott, A., Dominici, F., 2008. Fine particulate matter and mortality: a comparison of the Six Cities and American Cancer Society cohorts with a medicare cohort. Epidemiology 19 (2), 209–216. https://doi.org/10.1097/EDE.0b013e3181632c09.

Fang, L., et al., 2022. Development of a regional feature selection-based machine learning system (RFSML v1.0) for air pollution forecasting over China. Geosci. Model Dev. (GMD) 15 (20), 7791–7807. https://doi.org/10.5194/gmd-15-7791-2022.

Gal, Y., Ghahramani, Z., 2015. Bayesian convolutional neural networks with Bernoulli approximate variational inference. http://arxiv.org/abs/1506.02158.

Gao, H., Yang, W., Wang, J., Zheng, X., 2021. Analysis of the effectiveness of air pollution control policies based on historical evaluation and deep learning forecast: a case study of chengdu-chongqing region in China. Sustainability 13 (1), 1–28. https://doi.org/10.3390/su13010206.

Gaskins, A.J., et al., 2019. Exposure to fine particulate matter and ovarian reserve among women from a fertility clinic. Epidemiology 30 (4), 486–491. https://doi.org/10.1097/EDE.0000000000001029.

Gu, Y., Li, B., Meng, Q., 2022. Hybrid interpretable predictive machine learning model for air pollution prediction. Neurocomputing 468, 123–136. https://doi.org/10.1016/j.neucom.2021.09.051.

Gugnani, V., Singh, R.K., 2022. Analysis of deep learning approaches for air pollution prediction. Multimed Tools Appl 81 (4), 6031–6049. https://doi.org/10.1007/s11042-021-11734-x.

Habeebullah, T.M., Munir, S., Zeb, J., Morsy, E.A., 2022. Modelling the effect of COVID-19 lockdown on air pollution in makkah Saudi arabia with a supervised machine learning approach. Toxics 10 (5). https://doi.org/10.3390/toxics10050225.

Hähnel, P., Mareček, J., Monteil, J., O'Donncha, F., 2020. Using deep learning to extend the range of air pollution monitoring and forecasting. J. Comput. Phys. 408. https://doi.org/10.1016/j.jcp.2020.109278.

Han, K., Wang, Y., 2021. A Review of Artificial Neural Network Techniques for Environmental Issues Prediction. Springer Science and Business Media B.V. https://doi.org/10.1007/s10973-021-10748-9.

Han, Y., Lam, J.C.K., Li, V.O.K., Zhang, Q., 2022a. A domain-specific bayesian deep-learning approach for air pollution forecast. IEEE Trans Big Data 8 (4), 1034–1046. https://doi.org/10.1109/TBDATA.2020.3005368.

Han, Y., Lam, J.C.K., Li, V.O.K., Zhang, Q., 2022b. A domain-specific bayesian deep-learning approach for air pollution forecast. IEEE Trans Big Data 8 (4), 1034–1046. https://doi.org/10.1109/TBDATA.2020.3005368.

Hashad, K., et al., 2021. Designing roadside green infrastructure to mitigate traffic-related air pollution using machine learning. Sci. Total Environ. 773 (Jun). https://doi.org/10.1016/j.scitotenv.2020.144760.

Heidari, A., Shishehlou, H., Darbandi, M., Navimipour, N.J., Yalcin, S., 2024. A reliable method for data aggregation on the industrial internet of things using a hybrid optimization algorithm and density correlation degree. Cluster Comput. 27 (6), 7521–7539. https://doi.org/10.1007/S10586-024-04351-4/TABLES/3.

Heydari, A., Majidi Nezhad, M., Astiaso Garcia, D., Keynia, F., De Santoli, L., 2022. Air pollution forecasting application based on deep learning model and optimization algorithm. Clean Technol. Environ. Policy 24 (2), 607–621. https://doi.org/10.1007/s10098-021-02080-5.

Howard, A., et al., 2019. Searching for mobileNetV3. Proceedings of the IEEE International Conference on Computer Vision 1314–1324. https://doi.org/10.1109/ICCV.2019.00140.

Huang, C., et al., 2022. Effect of urban morphology on air pollution distribution in high-density urban blocks based on mobile monitoring and machine learning. Build. Environ. 219 (Jul). https://doi.org/10.1016/j.buildenv.2022.109173.

Jerrett, M., et al., 2001. A GIS - environmental justice analysis of particulate air pollution in Hamilton, Canada. Environ Plan A 33 (6), 955–973. https://doi.org/10.1068/a33137.

Ji, Y., et al., 2023. Regression analysis of air pollution and pediatric respiratory diseases based on interpretable machine learning. Front. Earth Sci. 11. https://doi.org/10.3389/feart.2023.1105140.

Joharestani, M.Z., Cao, C., Ni, X., Bashir, B., Talebiesfandarani, S., 2019. PM2.5 prediction based on random forest, XGBoost, and deep learning using multisource remote sensing data. Atmosphere 10 (7). https://doi.org/10.3390/atmos10070373.

Kabir, S., Ul Islam, R., Hossain, M.S., Andersson, K., 2020. An integrated approach of belief rule base and deep learning to predict air pollution. Sensors 20 (7), 1–25. https://doi.org/10.3390/s20071956.

Khan, A., Sharma, S., Chowdhury, K.R., Sharma, P., 2022. A novel seasonal index–based machine learning approach for air pollution forecasting. Environ. Monit. Assess. 194 (6). https://doi.org/10.1007/s10661-022-10092-x.

Kim, J.G., Lee, S.Y., Lee, I.B., 2023. The development of an LSTM model to predict time series missing data of air temperature inside fattening pig houses. Agriculture (Switzerland) 13 (4). https://doi.org/10.3390/agriculture13040795.

Ku, Y., Bin Kwon, S., Yoon, J.H., Mun, S.K., Chang, M., 2022. Machine learning models for predicting the occurrence of respiratory diseases using climatic and air-pollution factors. Clin Exp Otorhinolaryngol 15 (2), 168–176. https://doi.org/10.21053/ceo.2021.01536.

Kumar, K., Pande, B.P., 2023. Air pollution prediction with machine learning: a case study of Indian cities. Int. J. Environ. Sci. Technol. 20 (5), 5333–5348. https://doi.org/10.1007/s13762-022-04241-5.

Lai, W.I., Chen, Y.Y., Sun, J.H., 2022. Ensemble machine learning model for accurate air pollution detection using commercial gas sensors. Sensors 22 (12). https://doi.org/10.3390/s22124393.

Lautenschlager, F., et al., 2020. OpenLUR: off-the-shelf air pollution modeling with open features and machine learning. Atmos. Environ. 233 (Jul). https://doi.org/10.1016/j.atmosenv.2020.117535.

Li, V.O.K., Lam, J.C.K., Han, Y., Chow, K., 2021. A big data and artificial intelligence framework for smart and personalized air pollution monitoring and health management in Hong Kong. Environ Sci Policy 124, 441–450. https://doi.org/10.1016/j.envsci.2021.06.011.

Li, T., et al., 2023a. Contributions of various driving factors to air pollution events: interpretability analysis from Machine learning perspective. Environ. Int. 173 (Mar). https://doi.org/10.1016/j.envint.2023.107861.

Li, Y., Sha, Z., Tang, A., Goulding, K., Liu, X., 2023b. The application of machine learning to air pollution research: a bibliometric analysis. Ecotoxicol. Environ. Saf. 257 (Jun). https://doi.org/10.1016/j.ecoenv.2023.114911.

Li, T., et al., 2023c. Contributions of various driving factors to air pollution events: interpretability analysis from Machine learning perspective. Environ. Int. 173, 107861. https://doi.org/10.1016/J.ENVINT.2023.107861.

Liang, L., Daniels, J., Bailey, C., Hu, L., Phillips, R., South, J., 2023. Integrating low-cost sensor monitoring, satellite mapping, and geospatial artificial intelligence for intra-urban air pollution predictions. Environmental Pollution 331 (Aug). https://doi.org/10.1016/j.envpol.2023.121832.

Liu, H., Yue, F., Xie, Z., 2022. Quantify the role of anthropogenic emission and meteorology on air pollution using machine learning approach: a case study of PM2.5 during the COVID-19 outbreak in Hubei Province, China. Environmental Pollution 300. https://doi.org/10.1016/j.envpol.2022.118932.

Loh, H.W., Ooi, C.P., Seoni, S., Barua, P.D., Molinari, F., Acharya, U.R., 2022. Application of Explainable Artificial Intelligence for Healthcare: A Systematic Review of the Last Decade (2011–2022). Elsevier Ireland Ltd. https://doi.org/10.1016/j.cmpb.2022.107161.

Lovrić, M., Pavlović, K., Vuković, M., Grange, S.K., Haberl, M., Kern, R., 2021. Understanding the true effects of the COVID-19 lockdown on air pollution by means of machine learning. Environmental Pollution 274. https://doi.org/10.1016/j.envpol.2020.115900.

Masood, A., Ahmad, K., 2021. A Review on Emerging Artificial Intelligence (AI) Techniques for Air Pollution Forecasting: Fundamentals, Application and Performance. Elsevier Ltd. https://doi.org/10.1016/j.jclepro.2021.129072

Mele, M., Magazzino, C., 2020. A Machine Learning analysis of the relationship among iron and steel industries, air pollution, and economic growth in China. J. Clean. Prod. 277 (Dec). https://doi.org/10.1016/j.jclepro.2020.123293.

Meng, Q., et al., 2022. Prediction of COPD acute exacerbation in response to air pollution using exosomal circRNA profile and Machine learning. Environ. Int. 168. https://doi.org/10.1016/j.envint.2022.107469.

Nazmul Hoq, Md, Alam, Rakibul, Amin, Ashraful, 2019. Prediction of possible asthma attack from air pollutants: towards a high density air pollution map for smart cities to improve living. IEEE.

Oreshkin, B.N., Carpov, D., Chapados, N., Bengio, Y., 2019. N-BEATS: neural basis expansion analysis for interpretable time series forecasting. http://arxiv.org/abs/1905.10437.

Pielke, R.A., Adegoke, J.O., Chase, T.N., Marshall, C.H., Matsui, T., Niyogi, D., 2007. A new paradigm for assessing the role of agriculture in the climate system and in climate change. Agric. For. Meteorol. 142 (2–4), 234–254. https://doi.org/10.1016/j.agrformet.2006.06.012.

Probst, P., Wright, M.N., Boulesteix, A.L., 2019. Hyperparameters and Tuning Strategies for Random Forest. Wiley-Blackwell. https://doi.org/10.1002/widm.1301.

Rakholia, R., Le, Q., Quoc Ho, B., Vu, K., Simon Carbajo, R., 2023. Multi-output machine learning model for regional air pollution forecasting in Ho Chi Minh City, Vietnam. Environ. Int. 173 (Mar). https://doi.org/10.1016/j.envint.2023.107848.

Raub, J.A., Mathieu-Nolf, M., Hampson, N.B., Thom, S.R., 2000. Carbon Monoxide Poisoning-A Public Health Perspective [Online]. Available: www.elsevier.com/locate/toxicol.

Ravindra, K., et al., 2023. Application of machine learning approaches to predict the impact of ambient air pollution on outpatient visits for acute respiratory infections. Sci. Total Environ. 858 (Feb). https://doi.org/10.1016/j.scitotenv.2022.159509.

Razavi-Termeh, S.V., Sadeghi-Niaraki, A., Choi, S.M., 2021. Effects of air pollution in Spatio-temporal modeling of asthma-prone areas using a machine learning model. Environ. Res. 200 (Sep). https://doi.org/10.1016/j.envres.2021.111344.

Reid, C.E., Brauer, M., Johnston, F.H., Jerrett, M., Balmes, J.R., Elliott, C.T., 2016. Critical review of health impacts of wildfire smoke exposure. Environ. Health Perspect. 124 (9), 1334–1343. https://doi.org/10.1289/EHP.1409277/SUPPL_FILE/EHP.1409277.S001.ACCO.PDF.

Ren, X., Mi, Z., Georgopoulos, P.G., 2020. Comparison of Machine Learning and Land Use Regression for fine scale spatiotemporal estimation of ambient air pollution: modeling ozone concentrations across the contiguous United States. Environ. Int. 142 (Sep). https://doi.org/10.1016/j.envint.2020.105827.

Rosero-Montalvo, P.D., López-Batista, V.F., Arciniega-Rocha, R., Peluffo-Ordóñez, D.H., 2022. Air pollution monitoring using WSN nodes with machine learning techniques: a case study. Log. J. IGPL 30 (4), 599–610. https://doi.org/10.1093/jigpal/jzab005.

Seoni, S., et al., 2023. Application of uncertainty quantification to artificial intelligence in healthcare: a review of last decade (2013-2023). Comput. Biol. Med. https://doi.org/10.1016/j.compbiomed.2023.107441.

X. Shi et al., "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting.".

Shu, Y., Ding, C., Tao, L., Hu, C., Tie, Z., 2023. Air pollution prediction based on discrete wavelets and deep learning. Sustainability 15 (9). https://doi.org/10.3390/su15097367.

Smagulova, K., James, A.P., 2019. A Survey on LSTM Memristive Neural Network Architectures and Applications. Springer Verlag. https://doi.org/10.1140/epjst/e2019-900046-x.

Soleimani, Z., Darvishi Boloorani, A., Khalifeh, R., Griffin, D.W., Mesdaghinia, A., 2019. Short-term effects of ambient air pollution and cardiovascular events in Shiraz, Iran, 2009 to 2015. Environ. Sci. Pollut. Control Ser. 26 (7), 6359–6367. https://doi.org/10.1007/s11356-018-3952-4.

Song, Z., Bai, Y., Wang, D., Li, T., He, X., 2021a. Satellite retrieval of air pollution changes in central and eastern China during covid-19 lockdown based on a machine learning model. Remote Sens (Basel) 13 (13). https://doi.org/10.3390/rs13132525.

Song, J., Han, K., Stettler, M.E.J., 2021b. Deep-MAPS: machine-learning-based mobile air pollution sensing. IEEE Internet Things J. 8 (9), 7649–7660. https://doi.org/10.1109/JIOT.2020.3041047.

Steininger, M., Kobs, K., Zehe, A., Lautenschlager, F., Becker, M., Hotho, A., 2020. MapLUR: exploring a new paradigm for estimating air pollution using deep learning on map images. ACM Transactions on Spatial Algorithms and Systems 6 (3). https://doi.org/10.1145/3380973.

Stirnberg, R., et al., 2021a. Meteorology-driven variability of air pollution (PM1) revealed with explainable machine learning. Atmos. Chem. Phys. 21 (5), 3919–3948. https://doi.org/10.5194/ACP-21-3919-2021.

Stirnberg, R., et al., 2021b. Meteorology-driven variability of air pollution (PM1) revealed with explainable machine learning. Atmos. Chem. Phys. 21 (5), 3919–3948. https://doi.org/10.5194/acp-21-3919-2021.

Strubell, E., Ganesh, A., McCallum, A., 2020. Energy and policy considerations for modern deep learning research. Proc. AAAI Conf. Artif. Intell. 34 (9), 13693–13696. https://doi.org/10.1609/AAAI.V34I09.7123.

Subramaniam, S., et al., 2022. Artificial Intelligence Technologies for Forecasting Air Pollution and Human Health: A Narrative Review. MDPI. https://doi.org/10.3390/su14169951.

Sun, H., et al., 2022. Development of an LSTM broadcasting deep-learning framework for regional air pollution forecast improvement. Geosci. Model Dev. (GMD) 15 (22), 8439–8452. https://doi.org/10.5194/gmd-15-8439-2022.

Sun, H.Y., Chen, S.Y., Li, X.Y., Cheng, L.P., Luo, Y.P., Xie, L.L., 2023. Prediction and early warning model of mixed exposure to air pollution and meteorological factors on death of respiratory diseases based on machine learning. Environ. Sci. Pollut. Control Ser. 30 (18), 53754–53766. https://doi.org/10.1007/s11356-023-26017-1.

Tao, Q., Liu, F., Li, Y., Sidorov, D., 2019. Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU. IEEE Access 7, 76690–76698. https://doi.org/10.1109/ACCESS.2019.2921578.

Tao, H., et al., 2023. Machine learning algorithms for high-resolution prediction of spatiotemporal distribution of air pollution from meteorological and soil parameters. Environ. Int. 175. https://doi.org/10.1016/j.envint.2023.107931.

Usmani, R.S.A., Pillai, T.R., Hashem, I.A.T., Marjani, M., Shaharudin, R., Latif, M.T., 2021. Air pollution and cardiorespiratory hospitalization, predictive modeling, and analysis using artificial intelligence techniques. Environ. Sci. Pollut. Control Ser. 28 (40), 56759–56771. https://doi.org/10.1007/s11356-021-14305-7.

Usmani, R.S.A., Pillai, T.R., Hashem, I.A.T., Marjani, M., Shaharudin, R.B., Latif, M.T., 2023. Artificial intelligence techniques for predicting cardiorespiratory mortality caused by air pollution. Int. J. Environ. Sci. Technol. 20 (3), 2623–2634. https://doi.org/10.1007/s13762-022-04149-0.

Vakili, A., Al-Khafaji, H.M.R., Darbandi, M., Heidari, A., Jafari Navimipour, N., Unal, M., 2024. A new service composition method in the cloud-based Internet of things environment using a grey wolf optimization algorithm and MapReduce framework. Concurr. Comput. 36 (16), e8091. https://doi.org/10.1002/CPE.8091.

Wan, S., et al., 2022. Influence of ambient air pollution on successful pregnancy with frozen embryo transfer: a machine learning prediction model. Ecotoxicol. Environ. Saf. 236. https://doi.org/10.1016/j.ecoenv.2022.113444.

Wang, J., Xu, W., Dong, J., Zhang, Y., 2022a. Two-stage deep learning hybrid framework based on multi-factor multi-scale and intelligent optimization for air pollutant prediction and early warning. Stoch. Environ. Res. Risk Assess. 36 (10), 3417–3437. https://doi.org/10.1007/s00477-022-02202-5.

Wang, J., Xu, W., Dong, J., Zhang, Y., 2022b. Two-stage deep learning hybrid framework based on multi-factor multi-scale and intelligent optimization for air pollutant prediction and early warning. Stoch. Environ. Res. Risk Assess. 36 (10), 3417–3437. https://doi.org/10.1007/S00477-022-02202-5/TABLES/5.

Wei, J., Yang, F., Ren, X.C., Zou, S., 2021. A short-term prediction model of PM2.5 concentration based on deep learning and mode decomposition methods. Appl. Sci. 11 (15). https://doi.org/10.3390/app11156915.

WHO TEAM, 2016. Ambient Air Pollution: A Global Assessment of Exposure and Burden of Disease.

Wijnands, J.S., Nice, K.A., Seneviratne, S., Thompson, J., Stevenson, M., 2022. The impact of the COVID-19 pandemic on air pollution: a global assessment using machine learning techniques. Atmos. Pollut. Res. 13 (6). https://doi.org/10.1016/j.apr.2022.101438.

Wu, Y., Hoi, S.C.H., Mei, T., Yu, N., 2017. Large-scale online feature selection for ultra-high dimensional sparse data. ACM Trans. Knowl. Discov. Data 11 (4). https://doi.org/10.1145/3070646.

Xiao, D., Fang, F., Zheng, J., Pain, C.C., Navon, I.M., 2019. Machine learning-based rapid response tools for regional air pollution modelling. Atmos. Environ. 199, 463–473. https://doi.org/10.1016/j.atmosenv.2018.11.051.

Xie, X., Zuo, J., Xie, B., Dooling, T.A., Mohanarajah, S., 2021. Bayesian network reasoning and machine learning with multiple data features: air pollution risk monitoring and early warning. Nat. Hazards 107 (3), 2555–2572. https://doi.org/10.1007/s11069-021-04504-3.

Xie, Q., Ni, J.Q., Li, E., Bao, J., Zheng, P., 2022. Sequential air pollution emission estimation using a hybrid deep learning model and health-related ventilation control in a pig building. J. Clean. Prod. 371. https://doi.org/10.1016/j.jclepro.2022.133714.

Yamamoto, S.S., Phalkey, R., Malik, A.A., 2014. A Systematic Review of Air Pollution as a Risk Factor for Cardiovascular Disease in South Asia: Limited Evidence from India and Pakistan. https://doi.org/10.1016/j.ijheh.2013.08.003.

Zhang, W., Qian, C.N., Zeng, Y.X., 2014. Air Pollution: A Smoking Gun for Cancer. Landes Bioscience. https://doi.org/10.5732/cjc.014.10034.

Zhang, B., et al., 2023a. Machine learning assesses drivers of PM2.5 air pollution trend in the Tibetan Plateau from 2015 to 2022. Sci. Total Environ. 878 (Jun). https://doi.org/10.1016/j.scitotenv.2023.163189.

Zhang, B., et al., 2023b. Machine learning assesses drivers of PM2.5 air pollution trend in the Tibetan Plateau from 2015 to 2022. Sci. Total Environ. 878, 163189. https://doi.org/10.1016/J.SCITOTENV.2023.163189.

Zhang, Z., Zhang, S., Chen, C., Yuan, J., 2024. A systematic survey of air quality prediction based on deep learning. Alex. Eng. J. 93, 128–141. https://doi.org/10.1016/J.AEJ.2024.03.031.

Zhao, B., et al., 2021. Urban air pollution mapping using fleet vehicles as mobile monitors and machine learning. Environ. Sci. Technol. 55 (8), 5579–5588. https://doi.org/10.1021/acs.est.0c08034.

Zheng, H., Lin, F., Feng, X., Chen, Y., 2021. A hybrid deep learning model with attention-based conv-LSTM networks for short-term traffic flow prediction. IEEE Trans. Intell. Transport. Syst. 22 (11), 6910–6920. https://doi.org/10.1109/TITS.2020.2997352.

Zhou, W., et al., 2022. Machine learning elucidates the impact of short-term emission changes on air pollution in Beijing. Atmos. Environ. 283. https://doi.org/10.1016/j.atmosenv.2022.119192.

Zhu, J., Wu, P., Chen, H., Zhou, L., Tao, Z., 2018. A hybrid forecasting approach to air quality time series based on endpoint condition and combined forecasting model. Int J Environ Res Public Health 15 (9). https://doi.org/10.3390/ijerph15091941.

Zou, G., Lai, Z., Li, Y., Liu, X., Li, W., 2022. Exploring the nonlinear impact of air pollution on housing prices: a machine learning approach. Economics of Transportation 31 (Sep). https://doi.org/10.1016/j.ecotra.2022.100272.