

SURVEY

Open Access



# A systematic scrutiny of artificial intelligence-based air pollution prediction techniques, challenges, and viable solutions

Meenakshi Malhotra<sup>1</sup>, Savita Walia<sup>1</sup>, Chia-Chen Lin<sup>2\*</sup>, Inderdeep Kaur Aulakh<sup>3</sup> and Saurabh Agarwal<sup>4\*</sup>

\*Correspondence:

Chia-Chen Lin  
ally.cclin@ncut.edu.tw  
Saurabh Agarwal  
saurabh@yu.ac.kr;  
saurabhnsit2510@gmail.com  
<sup>1</sup>Chitkara University Institute of  
Engineering and Technology,  
Chitkara University, Rajpura, Punjab,  
India  
<sup>2</sup>Department of Information  
Computer Science and Information  
Engineering, National Chin-Yi  
University of Technology,  
Taichung 411, Taiwan  
<sup>3</sup>University Institute of Engineering  
and Technology, Panjab University,  
Chandigarh, India  
<sup>4</sup>Department of Information and  
Communication Engineering,  
Yeungnam University,  
Gyeongsan  
38541, Republic of Korea

## Abstract

Air is an essential human necessity, and inhaling filthy air poses a significant health risk. One of the most severe hazards to people's health is air pollution, and appropriate precautions should be taken to monitor and anticipate its quality in advance. Among all the countries, the air quality in India is decreasing daily, which is a matter of concern to the health department. Many studies use machine learning and Deep learning methods to predict atmospheric pollutant levels, prioritizing accuracy over interpretability. Many research studies confuse researchers and readers about how to proceed with further research. This paper aims to give every detail of the considered air pollutants and brief about the techniques used, their advantages, and challenges faced during pollutant prediction, which leads to a better understanding of the techniques before starting any research related to air pollutant prediction. This paper has given numerous prospective questions on air pollution that piqued the study's interest. This study discussed various machine and deep learning methods and optimization techniques. Despite all the discussed machine learning and deep learning techniques, the paper concluded that more datasets, better learning techniques, and a variety of suggestions would enhance interpretability while maintaining high accuracy for air pollution prediction. The purpose of this review is also to reveal how a family of neural network algorithms has helped researchers across the globe to predict air pollutant(s).

**Keywords** Air pollutants, AQI, Correlation, Meteorological factors, Module, Pollution, Prediction, Uncertainty

## Introduction

Pollution is a significant issue, affecting water, air, noise levels, land, or any other aspect of our environment. Every living human seeks a healthy life. They need clean water, fresh air, a peaceful day, and eco-friendly surroundings. But imagine - what if all these significant sources are contaminated? The drinking water gets infected with chemicals, breathing becomes difficult due to toxins in the air, and the surroundings are noisy and chaotic. How difficult would life become? Pollution, in a way, is a concern. However, water and air pollution are the most alarming of all.

**Table 1** Birth rate and death rate over the last eight years in India

| Population   | Year | Birth Rate     | Death Rate    | Infant mortality rate      | Death rate Due to Pollution |
|--------------|------|----------------|---------------|----------------------------|-----------------------------|
| 1.32 billion | 2016 | 18.3 births*   | 7.3 deaths*   | 33.1 deaths <sup>#</sup>   | 1.0 million [3]             |
| 1.33 billion | 2017 | 18 births*     | 7.3 deaths*   | 31.4 deaths <sup>#</sup>   | 1.2 million [4]             |
| 1.35 billion | 2018 | 17.8 births*   | 7.3 deaths*   | 29.7 deaths <sup>#</sup>   | 1.2 million [5]             |
| 1.36 billion | 2019 | 17.6 births*   | 7.2 deaths*   | 28.3 deaths <sup>#</sup>   | 1.6 million [6]             |
| 1.38 billion | 2020 | 17.4 births*   | 7.3 deaths*   | 29.07 deaths <sup>#</sup>  | 0.12 million [7]            |
| 1.39 billion | 2021 | 17.37 births*  | 7.3 deaths*   | 28.77 deaths <sup>#</sup>  | -                           |
| 1.4 billion  | 2022 | 17.163 births* | 7.380 deaths* | 27.693 deaths <sup>#</sup> | -                           |
| 1.43 billion | 2023 | 16.949 births* | 7.416 deaths* | 26.619 deaths <sup>#</sup> | -                           |

\*per 1000 population

<sup>#</sup>per 1000 live births**Table 2** Air pollutants details

| S. No. | Air Pollutant     | Name                                 | Detail   |
|--------|-------------------|--------------------------------------|--|
| 1      | PM <sub>2.5</sub> | Particulate Matter (2.5 micrometers) | This pollutant can enter the lower respiratory system, producing respiratory illnesses, and is linked to increased death rates in children with asthma and allergies.  |
| 2      | PM <sub>10</sub>  | Particulate Matter (10 micrometers)  | This pollutant penetrates the pulmonary alveoli after setting in the upper respiratory tract, affecting respiratory and cardiovascular diseases and possibly causing cancer.   |
| 3      | O <sub>3</sub>    | Ozone                                | It is an oxidizing contaminant that impacts living organisms. As a result, it is linked to several human health problems and is one of the most frequent pollutants in the atmosphere.   |
| 4      | CO                | Carbon Monoxide                      | This pollutant is a highly hazardous gas that is colorless, tasteless, and odorless. Prolonged exposure can be lethal. Furthermore, a carbon monoxide oxidation procedure produces carbon dioxide, the primary greenhouse gas precursor. |
| 5      | NO <sub>2</sub>   | Nitrogen Oxide                       | This pollutant causes the development of suspended particles and is a precursor to ozone and acid rain. High concentrations cause health concerns by cracking tissues and disrupting the respiratory tract.                              |
| 6      | SO <sub>2</sub>   | Sulfur Oxide                         | The pollutant is a precursor to acid rain-suspended particles affecting the respiratory tract.   |

With the rapid increase in the worldwide population, the demand for necessities like food, land, and water has also peaked. Countries like China and India are topping the list for being the most populated countries. Still, according to diverse research found on the internet, the population in China is moderately high compared to others [1]. The increasing rate of population explicitly affects the environment, and air pollution has emerged as a significant challenge in recent years. India has been ranked among the most populated countries, with 1.40 billion people [2] accounting for 17.71% of the global population. In the past few years, the birth and death ratios, and especially the death rate due to pollution, have drastically increased, as shown in Table 1.

### Air pollutants

Air pollution has reached crisis levels in most major cities, necessitating air quality predictions. The variety of temporal sequence information, combining one-dimensional and multidimensional panels, makes air quality prediction challenging. Before delving into the details of pollutant prediction, it is essential to identify the types of pollutants that exist and understand how they react to various situations (Table 2).

- a. **Primary Pollutants:** These pollutants are found directly in the air. Emissions from automobiles, power plants, biomass burning, forest fires, volcanoes, and other sources can all contribute to these problems. Gasoline and diesel are the key sources of air pollution in transportation. Industries produce primary pollutants such as NO<sub>2</sub>, CO, SO<sub>x</sub>, PM, Mercury, Organic Compounds (VOCs), and others in large quantities.
- b. **Secondary Pollutants:** Secondary pollutants form from primary pollutants reacting with each other or with molecules in the atmosphere rather than being directly generated from sources. O<sub>3</sub>, sulphuric acid and nitric acids are secondary pollutants, as are Peroxyacyl Nitrates (PANs), Smog, PM, and NO<sub>2</sub>.
- c. **Airborne air pollutants, such as persistent organic pollutants (POPs), CO, sulfur oxide, and nitrogen oxide, exist in the atmosphere.** Penetration can occur in various forms: as volatile gases like ozone and benzene, as droplets such as sulfuric acid and nitrogen dioxide, or as particulate matter including diesel exhaust and aromatic hydrocarbons.
- d. **Aerosol Pollutants:** Fog, dust, forest exudates, and geyser steam are naturally occurring aerosols. Aerosols are invisible. Microscopic particles comprise a few molecules; some are visible but still very small. Haze, delicate particulate matter, and smoke are examples of anthropogenic aerosols. Aerosols have an indirect or negligible influence on greenhouse gases, causing cooling. It can scatter light and alter the reflectance of the earth's surface. Frequent aerosols include sea salt, dust, and volcanic ash.
- e. **Particulate Matter:** It can be caused by both human and natural factors. Natural activities that create PM include road dust into the air, agriculture (stub burning, woodland burning, etc.), industrial processes, fossil fuel combustion, etc. The wind blows dust, and wildfires contribute to the production of PM. Particles with dimensions of 10 micrometers or 2.5 micrometers are produced by forest fires, volcanoes, dust storms, and human activities such as stubble burning and burning fossil fuels through vehicles, power stations, and road dust.
- f. **Meteorological Data:** It is known as a weather pattern. It helps to locate pollution sources and predict days with high pollutant concentrations. Computation models also contribute to the estimation of air quality. For example, weather data can include sun radiation, temperature, humidity, wind direction, and speed.
- g. **Air Quality Index:** It determines the cleanliness of the air and its impact on public health. To calculate AQI according to the Indian Government (CPCB), out of the available air pollutants, at least three air pollutants should be present for the calculation. Out of those three pollutants, one must be PM<sub>10</sub> or PM<sub>2.5</sub>, or both. It is used to keep the public's health safe. A higher number on the AQI scale, which ranges from 0 to 500, denotes more polluted air and a higher danger to health, and depending upon the AQI level, the impact on people's health can be identified (Table 3).

#### **Pollution affects three broad areas**

- a. **Pollution Affects Ecosystem:** Acid Rain is caused mainly by air pollution, and it has the most significant impact on plants that rely on rainwater. The most dangerous pollutants impacting plants include ammonia, nitrogen dioxide, sulfur dioxide, and ozone, which can kill the entire flora.

**Table 3** Air pollutants range as per AQI levels in  $\mu\text{g}/\text{m}^3$ 

| AQI Level | PM <sub>10</sub><br>24 h | PM <sub>2.5</sub><br>24 h | NO <sub>2</sub><br>24 h | O <sub>3</sub><br>8 h | CO<br>8 h | SO <sub>2</sub><br>24 h | NH <sub>3</sub><br>24 h | Pb<br>24 h |
|-----------|--------------------------|---------------------------|-------------------------|-----------------------|-----------|-------------------------|-------------------------|------------|
| Level 1   | 0–50                     | 0–30                      | 0–40                    | 0–50                  | 0–1.0     | 0–40                    | 0–200                   | 0–0.5      |
| Level 2   | 51–100                   | 31–60                     | 41–80                   | 51–100                | 1.1–2.0   | 41–80                   | 201–400                 | 0.6–1.0    |
| Level 3   | 101–250                  | 61–90                     | 81–180                  | 101–168               | 2.1–10    | 81–380                  | 401–800                 | 1.1–2.0    |
| Level 4   | 251–350                  | 91–120                    | 181–280                 | 169–208               | 10.1–17   | 381–800                 | 801–1200                | 2.1–3.0    |
| Level 5   | 351–430                  | 121–250                   | 281–400                 | 209–748*              | 17.1–34   | 801–1600                | 1201–1800               | 3.1–3.5    |
| Level 6   | 430+                     | 250+                      | 400+                    | 748+*                 | 34+       | 1600+                   | 1800+                   | 3.5+       |

\*Hourly Monitoring

- i. Sulphur: Sulphur levels in lakes and streams that are too high can harm plants and forest soil.
  - ii. Nitrogen: When deposited on surface water, it may affect aquatic vegetation and species.
  - iii. Ozone: It damages tree leaves in protected natural areas and hurts scenic.
  - iv. Mercury: It is a heavy metal complex released into the environment and deposited on land and water, accumulating in plants and animals, some of which people consume directly or indirectly.
  - v. Particulate Matter: When particle pollution levels are high in one place, it can seriously affect forests, wildlife, and coastal areas. Large areas of dead trees are common in PM-affected ecosystems because when groundwater becomes too acidic, critical minerals drain out of the soil, preventing plants from growing.
- b. **Pollution Affects Public Health:** Air pollution, whether indoor or outdoor, impacts people's lives. Particulate matter, sulfur dioxide, nitrogen dioxide, and ozone are the pollutants that have the most significant evidence to be harmful to public health. The health risks associated with particulate matter smaller than 10 and 2.5 millimeters in diameter are detailed in Table 4. PM can enter the lungs and circulatory system, causing cardiovascular, cerebrovascular, and respiratory consequences [8, 9]. Particulate matter (PM), methane, carbon monoxide, polyaromatic hydrocarbons (PAH), and volatile organic compounds (VOC) are produced when wood and coal are burned in inefficient stoves or open hearths. When kerosene burns in classic wick lamps, many fine particles and other contaminants are created [10]. Indoor air pollution, which significantly influences health issues, is caused by vapors from building materials, paints, furniture, cooking, heating, and tobacco smoking [11, 12]. A study of indoor contaminants discovered that 80% of particle mass concentrations were created [13]. An office building, school, house, cafeteria, or restaurant are examples of indoor locations. Most pollutants had substantially greater concentrations indoors than outside, and new regulations and recommendations should be created to cope with PM concentrations in kindergartens and schools [14]. Because there is always a link between indoor and outdoor quality, it is critical to keep track of the pollutants on any given day or at any given moment [15]. NO<sub>2</sub>, PM<sub>2.5</sub>, and PM<sub>10</sub> concentrations in classrooms and outdoors have distinct compositions, sources, and contributions [16]. Outdoor air quality, as well as other factors such as the kind of windows, have a significant impact on indoor PM concentrations [17]. Seasons affect indoor and outdoor pollutants; for example, during the rainy season, the indoor/outdoor ratio

**Table 4** AQI Range, Color, and their health impact

| AQI Category        | Level   | Range   | Color       | Health Impact   |
|---------------------|---------|---------|-------------|---|
| <b>Good</b>         | Level 1 | 0-50    | Green       | Minimal Impact  |
| <b>Satisfactory</b> | Level 2 | 51-100  | Light Green | Sensitive individuals may experience mild breathing difficulties  |
| <b>Moderate</b>     | Level 3 | 101-200 | Yellow      | Breathing difficulties can also be experienced by persons with heart problems, older people, children, and people with lung disorders such as asthma.   |
| <b>Poor</b>         | Level 4 | 201-300 | Orange      | People with heart problems may face breathing trouble with extended exposure.   |
| <b>Very Poor</b>    | Level 5 | 301-400 | Red         | Individuals with heart and lung conditions may be significantly impacted.   |
| <b>Severe</b>       | Level 6 | 401-500 | Maroon      | Even healthy people can experience respiratory problems, and people with heart or lung diseases may be at serious risk for serious health problems. Even a small amount of physical activity might harm one's health. |

of PM<sub>2.5</sub>, PM<sub>10</sub>, and TVOC increases [18]. People worldwide are now contributing to doing smaller to more prominent research to gain information about air pollutants [19–21].

- c. **Pollution Affects Sustainable Development:** Sustainable development satisfies current requirements without compromising future generations’ ability. It balances economic progress, environmental preservation, and human well-being. Many organizations are developing sustainable development goals to protect the environment and restore ecological balance since environmental contamination seriously threatens human growth and sustainable development.

**Research questions (RQs)**

When discussing air pollution, several questions commonly arise. Some of these questions include:

**RQ1:** How does air pollution affect different aspects?

**RQ2:** How does one city’s pollution level affect the neighboring city’s AQI index?

**RQ3:** What are the methods used to analyze the future air quality?

**RQ4:** What steps are the government or organizations taking to deal with the air pollution problem?

The first thing that springs to mind is how air pollution influences different aspects. Is it simply affecting a few points, or is it affecting the entire society? In the [Introduction](#) section, this topic was addressed. The second question is: How are other cities or places affected if one city is harmed by air pollution? Because pollutants have some particle weight (very little), do they travel? Correlation is the answer to this. The correlation between the two particles might be positive or negative depending on the climatic variable values. The relationship between meteorological data and air pollutants changes depending on the season, such as spring, summer, autumn, or winter. The temperature has a negative relationship with  $PM_{2.5}$ .

In contrast, precipitation has a positive relationship, and wind speed transports pollutants to neighboring border areas, which depends on how fast or slow the wind is [22]. The monsoon, pre-monsoon, and post-monsoon all impact the correlation. Different geographies can also produce different correlation results [23]. WS, AT, RH, and sun radiation were chosen in the winter to determine the connection between the two. WS and AT were the most prevalent meteorological components in the winter, while the highest absorptions of  $SO_2$ ,  $NO_2$ , NO, and  $NO_x$  were found in December, January, and February [24]. The two most challenging aspects of determining the correlation between the two are seasonal and geographical characteristics [25]. The relationship between the two is influenced by geographical location and seasonal change [26]. Temperature is inversely related to  $PM_{2.5}$  and  $PM_{10}$  levels and humidity all of the time, but a positive relationship during the monsoon [27]. The third research question focuses on the methodologies used to analyze air quality or anticipate air pollution levels and the obstacles researchers encountered during pollutant prediction.

Additionally, it examines whether the government or responsible authorities have taken action to address poor air quality. The literature review section explored these issues, and the [Need for Prediction](#) section sought potential answers.

The main contribution of the paper is as follows:

- a. The paper starts by conducting a systematic review, thoroughly examining existing literature to gather insights on causes, health impacts, and recent advancements in air pollution prediction. This approach ensures that the paper is grounded in existing knowledge and can provide a holistic view of the subject matter.
- b. The paper aims to illuminate the various factors crucial for accurately predicting air pollution levels. This inclusivity helps identify key variables and methodologies that contribute to effective forecasting. Furthermore, by detailing aspects such as datasets, data sources, and the artificial intelligence methods utilized for prediction, the paper equips future researchers with essential information to navigate this complex field.
- c. Furthermore, by discussing the initiatives undertaken by the Indian government to address air pollution, the paper contextualizes the research within real-world efforts to mitigate its adverse effects.
- d. This paper aids as a valued source for researchers and policymakers, offering perspectives on the complexities of air pollution prediction and advocating for continued efforts toward its effective management.

Many machine learning and deep learning methods used to predict air pollutant concentration struggled with managing data uncertainty and addressing the spatial-temporal aspect of the dataset, which negatively impacted the prediction models' performance.

The current ANN model has enhanced performance by utilizing hybrid methods and optimizing parameters. The improvement in the model results in more accurate predictions while maintaining high precision.

The article is structured as follows: Sect. 2 reviews recent studies on single and multiple air pollutants using machine and deep learning techniques. Section [Need for Prediction](#) addresses the current need for air pollutant prediction and outlines the measures being taken by the Indian Government to address this critical issue. Finally, Sect. 4 concludes with future perspectives.

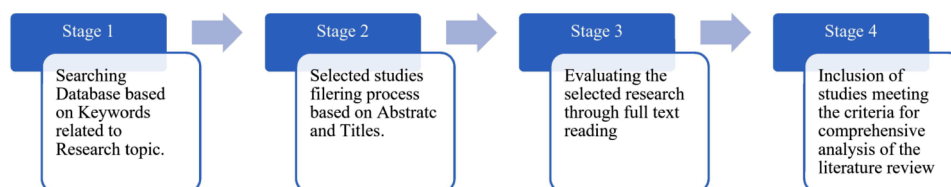
### Literature review on air quality prediction

Data collection is the first and the most essential part of the analysis and prediction. For the air pollution prediction, the data can be traffic data (fuel consumption, age of the vehicle, emission rate, and traffic volume), Environmental data (pollutant concentration, humidity, wind speed, wind direction, etc.), Geographical Data (longitude, latitude, land use, etc.), Socioeconomic Data (population density and rate of urbanization). The data collection part contains data that can include a combination of outliers, missing values, etc.

This step indicates the preparation of data before feeding it to the model. This step fines the raw data into information by eliminating the outliers, missing data, erroneous data, irrelevant data, and inconsistent data from the data collection step. This step is also known as the data cleaning step. Different techniques are available to process raw data, including normalization/scaling and interpolation methods for handling missing data. Feature selection can also be included in the data pre-processing step to enable the model to train faster. In feature selection, only those features are selected, which improves the model's accuracy, depending on the proper subset selection.

Choosing the proper technique is one of the tricky parts of the research field. There are various techniques to deal with problems related to the real world, depending on the research problems. Any method can be chosen, whether machine learning or deep learning. Artificial intelligence is the superset of machine learning, whereas deep learning is the subset of machine learning. Whether machine learning or deep learning is used, the primary focus is always on achieving the highest accuracy or the best possible results from the model. Any Machine Learning or Deep Learning algorithm can be used, depending on the size of the dataset and the problem requirements. Deep learning includes several varieties, including Neural Networks (Convolutional Neural Networks, Recurrent Neural Networks, Back Propagation Neural Networks [28], Long Short Term Memory Neural Networks, etc.), Stacked Auto Encoder, Deep Boltzmann Machines, Deep Belief Networks, and others [29–31].

In stage 1 (Fig. 1), we conducted keyword searches in Google Scholar, IEEE Xplore, and ScienceDirect, which initially resulted in results collection of a broad range of studies. In



**Fig. 1** Four-phase research design for developing a State-of-the-Art literature review methodology

stage 2, after searching based on keywords, the further selections were based on titles and abstracts. In addition to that, we filtered out the studies with insufficient focus on AI applications in air pollutant prediction. In stage 3, we thoroughly read the full text and found out the methodologies used and the challenges faced by the researchers. A comprehensive list of studies for in-depth analysis was done in the last stage. More details regarding the same can be seen in Table 5.

One can forecast a single air pollutant and multiple air pollutants using machine learning and deep learning approaches. The below-mentioned literature review on air pollutant(s) prediction is divided into three categories: Single air pollutant prediction, Multiple air pollutant prediction, and Single & Multiple air pollutant predictions using deep learning approaches. Each defined prediction method contains all the essential points about the research in Tables 6 and 7, and Table 8.

### Single air pollutant prediction

Single air pollutant prediction involves forecasting future concentrations of a specific pollutant using historical data and relevant environmental factors. Zhou et al. [32] used the Kendall tau coefficient to identify essential spatiotemporal features from regional meteorological and air quality data. Second, Multi-Task Learning (MTL) was used to train the Multi-Output Support Vector Machine (M-SVM) model to detect nonlinear interactions and disseminate correlation data across tasks. Finally, the M-SVM model was evaluated using  $PM_{2.5}$  concentration and meteorological and air quality information. A spatial and temporal correlation was also considered so that the correlated station could deal with one missing value in one station. The conclusion focused on addressing the robustness and uncertainty of the discussed models in the future. The Auto-Regressive Integrated Moving Average with Explanatory Variable-Multi-Layer Perception (ARIMAX-MLP) approach [33] was employed to train the model using expected values, achieving high accuracy. The number of hidden layers utilized here was one or two depending on the input, with up to twenty hidden neurons in each layer. Four different training algorithms were examined with 50 population sizes. The stopping criteria were 30 (number of generations). They concluded that due to the uncertainty, the result gets affected, and the model can predict the data 10 days in advance. Cabaneros et al. [34] collected data from two monitoring sites, and a nonlinear activation function was used

**Table 5** State-of-art literature review methodology

| Stage                 | Description   | Action Taken  | Criteria  |
|-----------------------|---|---|---|
| Initial Stage         | Keywords 'air pollution prediction,' 'air pollutants prediction,' and 'Prediction' were used to search massive Databases. | We conducted a keyword search using Google Scholar, IEEE Xplore, and ScienceDirect.   | Relevance, Presence of Keywords.                    |
| Preliminary Screening | Titles and abstracts of studies relevant to the research topic.   | <ul style="list-style-type: none"> <li>• Titles and abstracts were reviewed for relevance.</li> <li>• Filter out the studies with insufficient focus on AI applications in air pollutant prediction.</li> </ul> | Research topics relevant to ML, DL, and prediction. |
| Full-Text Evaluation  | Reading and evaluating the full text of the selected articles for reliability and relevance.                              | <ul style="list-style-type: none"> <li>• Full-text reading</li> <li>• Evaluating the soundness of the methodology used to predict air pollution.</li> <li>• Results and methods were documented.</li> </ul>     | Reliability, finding, methodology, Results.         |
| Final inclusion       | Inclusion of studies meeting the criteria for comprehensive analysis of the literature review                             | <ul style="list-style-type: none"> <li>• Completed a comprehensive list of studies for in-depth analysis</li> <li>• Created summaries of each selected study.</li> </ul>  | Comprehensive coverage, methodological soundness    |



**Table 6** Single air pollutant prediction techniques

| Ref  | Year | Country   | Forecasting Pollutants | Dataset  |  | Meteorological Data  | Dataset Source  |   | Evaluation               | Time Scale          | Technique |
|------|------|-----------|------------------------|--|--|--|---|---|--------------------------|---------------------|-----------|
|      |      |           |                        | Air Pollutant Data   | Air Pollutants Source  |  | Meteorological data Source  |   |                          |                     |           |
| [32] | 2019 | Taiwan    | PM <sub>2.5</sub>      | PM <sub>2.5</sub> , PM <sub>10</sub> , O <sub>3</sub> , NO <sub>x</sub> , NO <sub>2</sub> , CO                                   | Rainfall, AT, WS, WD, RH   | "Environmental Protection Administration, ROC"   | "Environmental Protection Ad- ministration, ROC"                          | RMSE, G <sub>bench</sub>                        | 2010 to 2016             | MM-SVM (M-SVM+ MTL) |           |
| [33] | 2018 | Iran      | PM <sub>2.5</sub>      | NO <sub>2</sub> , SO <sub>2</sub> , NO <sub>x</sub> , CO, O <sub>3</sub> , PM <sub>10</sub> , PM <sub>2.5</sub> , O <sub>3</sub> | AT, RH, Precipitation, Sunshine duration, Evaporation, WS, WD, Bulb temperature, Dew point, Pressure | "Environment Pollution Monitoring Centre of Mashhad"                                   | "Razavi Khorasan Meteorological Organization"                             | RMSE, MAE, MAPE                                 | 21/03/2014 to 20/03/2018 | ARIMAX- MLP         |           |
| [34] | 2017 | London    | NO <sub>2</sub>        | NO, NO <sub>x</sub> , CO, PM <sub>10</sub> , PM <sub>2.5</sub> , O <sub>3</sub>  | AT, WD, WS, RH, Solar Radiation  | "London Marylebone Road and Bloomsbury"  | "London Marylebone Road and Bloomsbury"                                   | RMSE, R <sup>2</sup> , FB                       | Jan 2008 to Dec 2009     | MLP-NN              |           |
| [35] | 2017 | China     | PM <sub>2.5</sub>      | SO <sub>2</sub> , NO <sub>2</sub> , CO, O <sub>3</sub> , PM <sub>2.5</sub>   | AT, RH, Pressure, Visibility, WS   | "Shenzhen air quality monitoring network (China Ministry of Environmental Protection)" | "Shenzhen Meteorological Monitoring Central Station"                      | MAE, RMSE, DA, IA                               | 01/09/2013 to 10/09/2013 | Hybrid-GARCH        |           |
| [36] | 2016 | London    | NO <sub>2</sub>        | NO <sub>2</sub>  | WS, WD, AT   | "Cabin housing precision air quality monitors"   | "Cabin housing precision air quality monitors"                            | ×   | 1998 to 2007             | ANN- ARIMAX         |           |
| [37] | 2015 | China     | PM <sub>2.5</sub>      | PM <sub>2.5</sub> , PM <sub>10</sub> , NO <sub>2</sub> , SO <sub>2</sub> , CO, O <sub>3</sub>                                    | AT, WS, WD, RH   | "Environmental Protection Administration of China"                                     | "China weather website platform"  | MAE, RMSE, IA                                   | Sep 2013 to Oct 2014     | MLP- BPNN           |           |
| [38] | 2015 | Mexico    | PM <sub>10</sub>       | PM <sub>10</sub>   | WD, WS, AT, RH, Atmospheric Pressure, Precipitation and Solar Radiation                              | "Automatic Environmental Monitoring Network in Salamanca"                              | "Automatic Environmental Monitoring Network in Salamanca"                 | RMSE, MAE, MSE, R <sup>2</sup> , d <sub>2</sub> | ×                        | MLP- Cluster- ing   |           |
| [39] | 2013 | Australia | O <sub>3</sub>         | NO <sub>x</sub> , O <sub>3</sub>   | AT   | "O <sub>3</sub> - Ecotech Ozone Monitor 9810, Nox- Ecotech 9841 instrument"            | "Commonwealth Scientific and Industrial Research Organization, Australia" | RMSE, MAE, R <sup>2</sup> , d <sub>2</sub>      | 2004                     | RBFFNN metamodel    |           |

**Table 6** (continued)

| Ref  | Year | Country | Forecasting Pollutants   | Dataset   |   | Dataset Source  |   | Evaluation                                       | Time Scale           | Technique               |
|------|------|---------|--|---|---|---|---|--|----------------------|-------------------------|
|      |      |         |  | Air Pollutant Data  | Meteorological Data                               | Air Pollutants Source   | Meteorological data Source                          |  |                      |                         |
| [40] | 2012 | Poland  | SO <sub>2</sub> /NO /CO /O <sub>3</sub> /PM <sub>2.5</sub> /PM <sub>10</sub> | PM <sub>10</sub> , PM <sub>2.5</sub> , SO <sub>2</sub> , NO, CO, O <sub>3</sub> | AT, WS, WD, RH, Rain, Cloud, Snow Cover, Pressure | ×   | ×   | α – standardization<br>αβ – standardization      | ×                    | Fuzzy Network Algorithm |
| [41] | 2012 | Poland  | PM <sub>10</sub>   | PM <sub>10</sub>  | AT, WS, WD, RH                                    | "Suburb Ursynow of Warsaw"  | ×   | ×  | 2006 to 2008         | Neural Network          |
| [42] | 2011 | China   | O <sub>3</sub>   | NO <sub>2</sub> , O <sub>3</sub>  | AT, RH, WS, Wind velocity,                        | UV radiation "Detector (Model 202, American 2B technology Company)" | "Ambient monitoring instrument and UV-A radiometer" | R <sup>2</sup> , d <sub>2</sub> , MAE, RMSE, MBE | Mar 2009 to Aug 2009 | SVM- GABPNN             |

**Table 7** Multiple air pollutant prediction techniques

| Ref  | Year | Country                    | Forecasting   |   | Dataset   |   | Dataset Source  |                                    | Evaluation               | Time Scale                  | Technique |
|------|------|----------------------------|---|---|---|---|---|------------------------------------|--------------------------|-----------------------------|-----------|
|      |      |                            | Pollutants  | Air Pollutant Data  | Air Pollutant Data  | Meteorological Data   | Air Pollutants Source   | Meteorological data Source         |                          |                             |           |
| [43] | 2019 | China                      | PM <sub>2.5</sub> , SO <sub>2</sub> , NO <sub>2</sub> , CO  | PM <sub>2.5</sub> , SO <sub>2</sub> , NO <sub>2</sub> , CO  | ×   | ×   | ×   | MAE, MAPE, RMSE                    | ×                        | EWT-MAEGA-NARX              |           |
| [44] | 2018 | India                      | C <sub>6</sub> H <sub>6</sub> , NO <sub>2</sub> , CO, O <sub>3</sub> , SO <sub>2</sub> , PM <sub>2.5</sub> , PM <sub>10</sub> | C <sub>6</sub> H <sub>6</sub> , NO <sub>2</sub> , CO, O <sub>3</sub> , SO <sub>2</sub> , PM <sub>2.5</sub> , PM <sub>10</sub> | RH, AT, WS, WD  | “Central Pollution Control Board, India”                      | “Central Pollution Control Board, India”                                  | AUC, Accuracy, Success Index       | Jan 2015 to 03-Aug 2017  | Random Forest               |           |
| [45] | 2018 | Greece                     | NO <sub>2</sub> , NO, O <sub>3</sub> , CO, SO <sub>2</sub>  | NO <sub>2</sub> , NO, O <sub>3</sub> , CO, SO <sub>2</sub>  | ×   | “Hellenic Ministry of Environment, Energy and Climate Change” | ×   | MAE, RMSE, R <sup>2</sup>          | 2001 to 2013             | FFNN                        |           |
| [46] | 2018 | China                      | Rd, Rs  | PM <sub>2.5</sub> , PM <sub>10</sub> , NO <sub>2</sub> , SO <sub>2</sub> , CO, O <sub>3</sub> and AQI                         | Precipitation, AT, RH, and location (longitude and latitude)  | “China National Environmental Monitoring Centre”              | “National Meteorological Information Centre of China”                     | R <sup>2</sup> , RMSE, MAE         | Jan 2014 to Dec 2016     | SVM                         |           |
| [47] | 2017 | China                      | PM <sub>2.5</sub> , SO <sub>2</sub> , NO <sub>2</sub> , O <sub>3</sub>  | PM <sub>2.5</sub> , SO <sub>2</sub> , NO <sub>2</sub> , O <sub>3</sub>  | Location, Precipitation, AT, WS, WD, RH                       | “Environmental Monitoring Centre” China                       | “Meteorological Data Service Centre”                                      | R <sup>2</sup> , RMSE, LOOCV, RMSE | 01/01/2013 to 31/12/2013 | LUR                         |           |
| [48] | 2016 | Shanghai, Bulgaria, Berlin | PM <sub>2.5</sub> , O <sub>3</sub>  | PM <sub>2.5</sub> , NO <sub>x</sub> , NO <sub>2</sub> , NO, O <sub>3</sub> , SO <sub>2</sub>                                  | AT, Pressure, RH, WS, WD, Solar Radiation                     | ×   | ×   | AUC, Accuracy, Success Index       | Feb 2013 to Feb 2014     | Multilabel Bayesian Network |           |
| [49] | 2016 | Mexico                     | O <sub>3</sub> , SO <sub>2</sub> , NO <sub>2</sub> , CO, PM <sub>10</sub> , PM <sub>2.5</sub>                                 | O <sub>3</sub> , SO <sub>2</sub> , NO <sub>2</sub> , CO, PM <sub>10</sub> , PM <sub>2.5</sub> , AQI                           | ×   | “Mexican Ministry of Environment”                             | ×   | AQI-AHP                            | 2015                     | FIS- AHP                    |           |
| [50] | 2016 | China                      | PM <sub>10</sub> , SO <sub>2</sub> , NO <sub>2</sub>  | PM <sub>10</sub> , SO <sub>2</sub> , NO <sub>2</sub>  | WS, AT, RH, Pressure, Visibility, Illumination, Precipitation | “Ministry of Environmental Protection of China”               | “ <a href="http://www.tianqihoubao.com">http://www.tianqihoubao.com</a> ” | MAPE, RMSE, R <sup>2</sup>         | 01/01/2011 to 26/12/2011 | W- BPNN                     |           |

**Table 8** Single and multiple air pollutants deep learning-based Prediction techniques

| Ref  | Year | Country        | Dataset                                      |  | Dataset Source   |   | Evaluation  | Time Scale               | Technique |
|------|------|----------------|--|--|--|---|---|--------------------------|-----------|
|      |      |                | Forecasting Pollutants                       | Air Pollutant Data   | Air Pollutants Source                                    | Meteorological data Source  |   |                          |           |
| [66] | 2019 | Korea          | Indoor PM <sub>2.5</sub>                     | NO, NO <sub>2</sub> , CO, CO <sub>2</sub> , PM <sub>2.5</sub> , PM <sub>10</sub> , AT, RH              | "Ministry of Environment"                                | "Ministry of Environment"   | RMSE, MAPE, R <sup>2</sup>  | Jan 2009 to Dec 2009     | GRU       |
| [67] | 2019 | Beijing, China | PM <sub>2.5</sub>                            | PM <sub>2.5</sub> , PM <sub>10</sub> , NO <sub>2</sub> , CO, O <sub>3</sub> , SO <sub>2</sub>          | "Urban Air project of Micro-soft Research"               | "Urban Air project of Micro-soft Research"  | MAE, RMSE   | 01/05/2014 to 30/04/2015 | LSTM-FC   |
| [68] | 2019 | Beijing, China | PM <sub>2.5</sub>                            | PM <sub>2.5</sub>  | "Ministry of Environmental Protection of China"          | "MERRA2 data"   | RMSE, MAE, MAPE   | 01/01/2016 to 31/12/2017 | C-LSTM    |
| [69] | 2019 | China          | PM <sub>2.5</sub>                            | PM <sub>2.5</sub> , PM <sub>10</sub> , NO <sub>2</sub> , SO <sub>2</sub> , O <sub>3</sub> , CO         | "Environmental Protection Administration (EPA) of China" | "China Meteorological Administration (CMA)"   | IA, MAE, RMSE, Recall rate, False alarm rate, Correlation coefficient | 01/01/2015 to 01/04/2016 | GC-LSTM   |
| [70] | 2018 | China          | PM <sub>10</sub>                             | PM <sub>2.5</sub> , PM <sub>10</sub> , NO <sub>2</sub> , CO, O <sub>3</sub> , SO <sub>2</sub> , AQI    | "China National Environmental Monitoring Center"         | "China National Environmental Monitoring Center"  | MSE, MAPE   | 01/04/2015 to 01/09/2017 | GRU       |
| [71] | 2018 | Beijing, China | Air Quality                                  | CO, NO <sub>2</sub> , SO <sub>2</sub> , O <sub>3</sub> , PM <sub>10</sub> and PM <sub>2.5</sub>        | X  | X   | MAE, RMSE, Accuracy   | 01/05/2013 to 30/04/2017 | Deep-LSTM |
| [72] | 2017 | Beijing, China | PM <sub>2.5</sub>                            | PM <sub>2.5</sub>  | "Ministry of Environmental Protection of China"          | "National Oceanic and Atmospheric Administration's (NOAA's) national climate data center" | RMSE, MAE, MAPE   | 01/01/2014 to 28/05/2016 | LSTM      |
| [73] | 2019 | Taiwan         | PM <sub>2.5</sub> , PM <sub>10</sub> and NOx | PM <sub>2.5</sub> , PM <sub>10</sub> , NOx, NO <sub>2</sub> , NO, SO <sub>2</sub> , CO, O <sub>3</sub> | "Environmental Protection Administration in Taiwan"      | "Central Weather Bureau in Taiwan"  | RMSE, G <sub>bench</sub>  | 2010 to 2016             | DM-LSTM   |

to process it. Stepwise Regression, Principal Component Analysis (PCA), and Classification and Regression Technique (CART) were used for feature selection. The network was trained using the Levenberg-Marquardt backpropagation technique, which changed the weights of interconnected neurons. Gradient descent was used as a nonlinear optimization method. The process was performed three times to account for the random initialization of weights across neurons.

To estimate the  $PM_{2.5}$  concentration of six stations in China, the hybrid-Garch (Generalized Auto-Regressive Conditional Heteroskedasticity) model was considered [35]. ARIMA model was also considered for capturing the linear part of the data. On the other hand, SVM was used to deal with non-linearity. The authors failed to choose the proper data set as the data set was small, which led to an accuracy problem. An ANN approach and ARIMAX model predicted [36] the concentration of  $NO_2$ . To process the expected results, the relationship between the  $NO_2$  concentration and various meteorological variables and traffic pollutants was also considered for the prediction. The correlation between  $PM_{2.5}$  and various meteorological data was identified [37] using the MLP method. The wavelet transform of the time series was used to forecast the average daily concentrations of  $PM_{2.5}$  based on the concentration from the day before and the expected values of the day's meteorological data. As the activation function, the sigmoid function was utilized. The model predicted  $PM_{2.5}$  values two days ahead. The novel aspect of the method was adding a trajectory-based geographic parameter to the ANN algorithm as an additional input predictor.

A scheme was proposed to estimate  $PM_{10}$  concentrations for the subsequent 24 h using a combination of a multilayer perceptron (MLP) and a clustering algorithm [38]. The clustering algorithm explored the relationship between meteorological data and  $PM_{10}$  pollutant data. An interpolation method was applied to handle missing values, and a log-sigmoid activation function was used. It was found that it would be challenging to add relevant information to the ANN's input patterns by identifying groups with similar data features and uncovering associations between them. A combination of three techniques was recommended [39]. The goal was to precisely forecast the geographical distribution of ozone as an air pollutant across a sizable area while handling noisy input using the hidden neuron output information from the preceding iteration. Scheme [40] utilized forecast, actual weather, and pollution data. The model incorporated all data, including those with inaccurate, chaotic, and unpredictable values. Weather forecasts were first sorted, and then certain meteorological conditions were processed into Fuzzy sets, characterized using Fuzzy numbers. A collection of pollutant concentrations was derived using Fuzzy grouping, followed by the projection of an aerosanitary scenario using standardized approaches. To achieve the best results [41], predictors of equivalent quality but autonomous operation were selected. Four neural predictors were considered: MLP, RBFN, ERN, and SVM. The MLP employed the sigmoid activation function, RBF networks used a local (Gaussian) function with a distinct learning technique, and SVR utilized the kernel concept and a robust statistical learning method. Another study [42] used meteorological data as input to anticipate ozone concentrations in the atmosphere. The SVM algorithm was used to classify the data into specific categories. Using sigmoid as the activation function, BPNN with a genetic algorithm was employed to achieve higher forecast stability. It was noted that in cities with less traffic and more factories, this pattern might not perform well.

#### Challenges:

- To deal with the uncertainty that affects the prediction performance.
- To feed the model for better prediction, identify the relationship between the variables.
- Lack of Spatio-Temporal consideration in the study.
- To deal with sudden changes in input data.
- Improve prediction accuracy with a time extension.
- To improve computational time concerning complexity.

Due to the involvement of natural climate changes and human activities, the data on air pollution are inherently ambiguous. For better prediction, an enhanced algorithm is needed to handle the uncertainty of data. To predict the quantity or concentration of air pollutants, it is crucial to understand how each pollutant behaves and whether there is a positive, negative, or neutral relationship between those pollutants. A more profound comprehension of the link between pollutants can help determine which features should be added to the model. There is a link between the forecast performance and the time and area. The atmospheric behavior of the data gathered from various locations, such as the sea, mountains, deserts, etc., would vary. Significant fluctuations in the data over time may be encountered when working with time series data, such as air pollution data. As a result, if the algorithm used in the training phase is not well-trained, the prediction system may perform poorly. A sufficient amount of data is required for a prediction model to produce accurate results, and if this is not considered from the beginning, the system's processing time may increase.

#### Multiple air pollutants prediction

Multiple Air Pollutants Prediction integrates advanced models and techniques to forecast contaminant concentrations accurately. Two models were incorporated, the Multi-Agent Evolutionary Genetic Algorithm (MAEGA) and the Nonlinear Auto Regressive Network with Exogenous Input (NARX) neural network, which achieved good performance with minimal error [43]. To estimate pollutant concentrations, Empirical Wavelet Transformation (EWT) was utilized to decompose the time series data, followed by MAEGA to optimize the weights in a multi-step process. The improved NARX neural network then forecasted the contaminants. A hybrid method combining differential evolution and random forest techniques was devised to forecast contaminants [44]. The differential evolution approach selected the candidate solution based on fitness value, and the random forest method created multiple trees to project pollutants using the candidate-chosen value. MLR and ANN, two approaches with linear and nonlinear behavior, were used to forecast pollutant concentrations [45]. A leave-one-out cross-validation process was employed to produce accurate results for prediction. The dataset spanned 13 years, with 60% used for training, 15 to 20% for validation, and the final two years for testing. Activation functions such as hyperbolic tangent were utilized for the input and hidden layers, while the output layer used a linear function. Seven air pollution variables, fifteen combinations of two parameters, and twenty combinations of three parameters were used to train the proposed algorithm and achieve reliable forecasting results [46]. It was claimed that selecting the right pollutant is crucial for the accuracy of the forecasting model. The SVM was utilized to forecast  $R_s$  and  $R_d$  ("Global and Diffuse

Solar Radiation”), starting with no pollutants and incrementally increasing to one, two, and three pollutants. Fifty-nine potential predictors were analyzed to forecast various air contaminants [47]. Comparisons were made between air pollution data and meteorological factors, industrial emissions data and the population density of the study region, and topographical data and meteorological data. The preference for a Land Use Regression (LUR) model was demonstrated in the prediction process. An attempt was made to increase the accuracy of the proposed prediction system with the help of an independent predictor [48].

PM<sub>2.5</sub> was first predicted in 8 stations for the same day and the next day, followed by the prediction of ozone in two stations with 4 class variables. A Bayesian network was used to justify the performance of the multilabel classifier over the independent approach, which improved the AUC and success index. O<sub>3</sub>, SO<sub>2</sub>, CO, PM<sub>10</sub>, and PM<sub>2.5</sub> were inputs, with responses indicating air quality levels from excellent to dangerous [49]. A Fuzzy Interference System (FIS) method was presented for assessing air quality, prioritizing parameters with a more significant damaging weight assignment. A membership function was implemented to cope with ambiguity, and the Analytic Hierarchy Process (AHP) was considered to analyze more problematic pollutants. Wavelet transformation with BPNN was performed on the input dataset to forecast air contaminants [50]. The relationship between air pollutants was considered since it could impact the accuracy of the suggested forecasting system. Stationary wavelet transformation was used to decompose the time series, and wavelet coefficients were then used to alter the weights in the BPNN model for future forecasting. Kumar and Pande [51] forecast air quality by analyzing six years of air pollution data from 23 Indian cities. The dataset is pre-processed, and feature selection is conducted using correlation analysis. Exploratory data analysis reveals hidden patterns and identifies key pollutants affecting the air quality index. A significant reduction in most pollutants was observed in 2020, attributed to the pandemic. Five machine learning models are used for prediction, with resampling techniques addressing data imbalance. Model outcomes, compared using standard metrics, show the Gaussian Naive Bayes model with the highest accuracy and the Support Vector Machine model with the lowest. Performance evaluation indicates that the XGBoost model has the strongest correlation between predicted and actual data.

Challenges:

- To examine the predictive ability of a more advanced ANN model.
- Consideration of knowledge-based systems with improved computational time.
- Deeper analysis to optimize the interference for better assessment.
- Better analysis of membership function.

The increasing amount of data necessitates using advanced artificial neural networks (ANNs), such as deep learning, two advanced machine learning model hybridization, etc., because the machine learning model performs poorly when data increases (particularly when considering meteorological and air pollution data). Knowledge-based systems are required to handle the processing time and performance of large volumes of data. The massive amount of data also necessitates parameter optimization for improved algorithm performance. By using the membership function, non-fuzzy inputs can be transformed into fuzzy ones. The membership function can represent a number between 0

and 1. An element can belong to a class if its value is closer to 1, and it is not to belong to a class if its value is closer to 0.

### Single and multiple air pollutant prediction using deep learning approaches

Several methods have been utilized in the literature, ranging from statistical approaches to more recent developments in machine learning. Deep feedforward and recurrent neural networks are the most prevalent network topologies in the literature [52–54]. Deep learning has proven effective in discovering hidden relationships within complicated issues, and the more specialized architecture recurrent neural network has proven to be a helpful tool for time series prediction. Additionally, ensemble learning is advantageous since it is susceptible to noise and variation. Researchers have opted for various deep-learning methods for the prediction [55–57]. Air quality data is continuous, so an effective and efficient prediction model is needed. Unlike machine learning techniques, deep learning comes with the idea of learning the data long-term. With the help of inclusions of gates in LSTM [58] and GRU, the techniques are currently being used worldwide for predicting single air pollutants, multiple air pollutants, or air quality indexes. Table 7 shows the deep learning techniques used by researchers for single or multiple air pollutant predictions.

An innovative hybrid model was proposed [59] by combining the strengths of two deep learning approaches. Furthermore, to forecast PM concentrations, hybrid deep-learning approaches CNN-LSTM and CNN-GRU were compared to several individual methods, including LSTM and GRU. These training models considered hourly air pollution and meteorological data. The experiment findings demonstrated that the model could forecast PM pollutant concentrations for seven days. In five randomly selected regions, for PM<sub>10</sub> and PM<sub>2.5</sub> prediction, hybrid models outperformed single models with the lowest RMSE and MAE values. It was found that the PM<sub>10</sub> prediction CNN-GRU model performed better, and the PM<sub>2.5</sub> CNN-LSTM model performed well. To increase the accuracy of PM<sub>2.5</sub> prediction, an early warning system was developed to predict PM<sub>2.5</sub> concentration based on extraction and optimization mechanisms [60]. To begin, a feedback VMD method was created to break down the sequence of PM<sub>2.5</sub> concentration, whereas fuzzy entropy was employed to rebuild comparable complexity patterns. Copula entropy was then utilized to identify the most significant influencing elements on PM<sub>2.5</sub>. Following that, the rebuilt elements and persuading factors were fed into three separate training forecast schemes, i.e., LSTM, GRU, and TCN. Individual prediction model outputs were nonlinearly merged and optimized using Gaussian process regression and multi-objective grey wolf optimization. Lastly, the predictions of the various rebuilt elements were nonlinearly combined to get the concluding PM<sub>2.5</sub> predictions.

A graph convolutional temporal sliding long short-term memory (GT-LSTM) scheme [61] was given to predict various air contaminants. Both models were merged using a temporal sliding method. To study the dynamic changes of air pollution over time, LSTM was applied with a temporal sliding technique. Experiments showed that the scheme could extract high-level spatiotemporal characteristics more accurately and efficiently than the existing benchmark. To anticipate and analyze air pollution from Combined Cycle Power Plants, develop an innovative hybrid intelligence model combining the Multi-verse Optimization Algorithm (MVO) and LSTM [62]. In this model, the MVO approach was used to improve the LSTM parameters, which decreased the forecasting



error. At the same time, the LSTM functioned as the forecasting engine to anticipate the amounts of  $\text{NO}_2$  and  $\text{SO}_2$  released by the Combined Cycle Power Plant. The model's effectiveness was then assessed using actual data gathered from a combined cycle power plant in Kerman, Iran. Over five months (May to September 2019), wind speed measurements, air temperature,  $\text{NO}_2$ , and  $\text{SO}_2$  levels were made every three hours.

A multi-point deep learning scheme was proposed [63] for highly dynamic air quality prediction based on convolutional LSTM (ConvLSTM). ConvLSTM designs integrated LSTM with CNN, allowing for the extraction of temporal and spatial data characteristics. Furthermore, uncertainty quantification schemes were constructed above the model's architecture, and its performance was investigated. The ConvLSTM outperformed cutting-edge technologies of machine learning and deep learning schemes. A method [64] was developed to forecast air pollution and create an early warning system. This approach uniquely integrates advanced optimization, feature selection, and extraction methods. The  $\text{PM}_{2.5}$  sequence was initially broken down into several smaller sequences using full ensemble empirical mode decomposition with adaptive noise. Subsequently, fuzzy entropy was employed to reconstruct the new components of the sub-sequences, resulting in varying levels of complexity. The influential elements of the reconstructed components were identified using the Max-Relevance and Min-Redundancy techniques. To forecast and nonlinearly integrate the rebuilt components, a dual-phase deep learning scheme was designed utilizing LSTM, which has been improved using the grey wolf optimization method. Ultimately, the proposed hybrid scheme successfully anticipated air pollution levels and provided an effective early alerting mechanism. This hybrid framework performed better than different approaches concerning accuracy, warning precision, and prediction resilience. The outcomes indicated that this hybrid framework could be valuable for forecasting air pollution and providing early alerts.

Beijing's hourly  $\text{PM}_{2.5}$  concentrations and meteorological data were utilized as input by researchers [65]. Using the GRU model, four distinct models were trained, each corresponding to one of the four seasons: spring, summer, autumn, and winter. The effectiveness of these models in predicting seasonal  $\text{PM}_{2.5}$  levels was then evaluated using appropriate test sets for each season. The model's prediction error and prediction accuracy were studied and compared after multiple trials. Continual modification of model parameters and the benefits of this technique were confirmed. The findings showed that the model's prediction accuracy was high. The concentration of contaminants was forecasted by considering three primary areas [66]. First, the dataset was characterized, processed, and examined to identify crucial properties, including the autocorrelation function and evidence of non-Gaussianity through a normality test. A Structural Recurrent Neural Network (SRNN) and various RNN structures linked with memory cells were evaluated to produce an accurate sequential prediction model of  $\text{PM}_{2.5}$  indoor health risk assessment.

In the end, many performance indicators were established to assess each prediction model's performance. The researchers identified various work constraints, such as the difficulty in implementing dynamic hyper parameterization for offline structure training and capturing instantaneous disturbed  $\text{PM}_{2.5}$  data using the Dynamic-Window size and sequence length. A subway station in Korea was used to forecast  $\text{PM}_{2.5}$  using a deep-learning approach. Telemonitoring systems were installed in the metro, waiting rooms, and subterranean regions. To forecast indoor  $\text{PM}_{2.5}$  concentrations, point-to-point

prediction was utilized to choose the top RNN models. After identifying the finest RNN for the research, the multi-sequence prediction was used, yielding the best prediction accuracy results. Statistical analysis revealed that  $PM_{2.5}$  concentrations were higher between February and December compared to other months.

A data-driven model was employed to measure air pollution concentration in the atmosphere [67]. “Historical data, meteorological data, weather prediction data, and day-of-the-week data” from 36 monitoring stations were used in the operation. An LSTM-Fully Connected neural network focused on the spatiotemporal relationship to better predict  $PM_{2.5}$  concentrations. A basic linear interpolation approach was used to fill in the missing data. The procedure was multidimensional and multi-step, with results validated using 12-fold cross-validation. While the model successfully predicted  $PM_{2.5}$  concentration levels, it failed to capture abrupt changes in the input data. Meteorological and aerosol data from 1233 monitoring sites were collected, and Convolutional Long Short-Term Memory Extended (C-LSTME) [68] was used to extract spatial-temporal characteristics. The results were then validated using a 5-fold cross-validation method. The  $k$ -nearest neighbor technique was used to determine the influence of the center station on the nearby stations. After a one-hour to six-day time lag, the proposed method was compared to existing methodologies and demonstrated superior prediction results. The study accounted for the spatiotemporal relationship with a time lag to effectively anticipate contaminants.

The objective [69] was to address the abovementioned constraints by proposing a hybrid model to enhance  $PM_{2.5}$  concentration forecasts. A Graph Convolutional Network (GCN) was utilized to identify spatial dependencies between stations, whereas an LSTM network was used to capture temporal dependencies within the data over various periods. The study utilized geographical and air quality data, including historical air quality records, meteorological factors, and geographical and temporal predictors, to visually represent spatio-temporal fluctuations. Airnet data was utilized to find the best-performing predictor in the RNN model [70]. The data was trained using several deep learning models, such as RNN, LSTM, and GRU, and the findings showed that GRU performed better than the other two predictors. Four layers were evaluated for each deep learning model, and RMSE and MAPE values for each prediction layer were considered. A novel deep learning-based system [71] was created to forecast future air quality by leveraging historical air quality and weather data. Granger Causality represented the spatial interdependence between two stations, with formulated relative stations and relative areas reflecting the spatial connection. The study divided air quality’s temporal features into short-term and long-term dependencies, both learned using LSTM. The method was tested using air quality data and one-year weather forecasts from Beijing.

A spatiotemporal correlation was used [72] to forecast air pollution concentrations with LSTM neural networks. A total of 20,196 records of historical data, meteorological data, and supplementary data were employed for the research. After standardizing the data, Missing data was addressed through a simple linear interpolation approach. Time-stamping was done using one-hot encoding. While the approach effectively improved multi-score predictions, it did not enhance prediction performance. To address issues like spatiotemporal instability and time-lag effects, the Deep multi-output LSTM (DM-LSTM) technique [73] was combined with three deep learning algorithms: mini-batch gradient descent, dropout neuron, and L2 regularization. The DM-LSTM model was

created with numerous future stages in mind. The correlation between the input and output variables was ascertained using the Kendall-tau coefficient.

The analysis of atmospheric contaminants was conducted [74] using time series regression forecasting at two major monitoring stations in India, Shadipur and Anandvihar, out of a total of 20 sites during the period from 2016 to 2017. The examination involved both descriptive and predictive analysis utilizing Rstudio and Tableau. Another study [75] focused on determining atmospheric contaminant concentrations with a Deep Learning system using datasets from Denmark and Romania collected between 2013 and 2015. The model was trained on 69.5% of the data from 17,568 samples, with the remainder used for testing. This prediction model consisted of four steps: sensor data input, computation by the deep learning model, data labeling, and final prediction by the decision unit. The k-means algorithm with the Mahout library was utilized to improve air quality prediction using the Citypulse dataset as a benchmark for Romania. A new dataset was generated by combining five air pollution data components with four meteorological data components. The fuzzy clustering technique was recommended for future research to achieve better results. A study [76] employed a two-phase decomposition approach combining CEEMD and VMD to manage high accuracy and dispersed signal information. Following this decomposition, the Extreme Learning Machine (ELM) with Differential Evolution (DE) was utilized to achieve accurate prediction performance. The training data comprised China's data from July 1, 2014, to May 31, 2016, with testing data from June 1, 2016, to June 30, 2016. Two models [77] for predicting air pollutant concentrations were introduced: the EMD-SVR and the EMD-IMFs model. These models were used to model AQI data, and S-ARIMA was utilized to forecast IMFs.

However, the specific pollutant concentrations impacting the environment were not defined. The Mean Absolute Percentage Error (MAPE) was used to identify the best value by comparing the two proposed hybrid models with six others, using China data between June 2014 and August 2015. In previous studies, only hourly data was used for forecasting, and yearly data was often overlooked. To address this, using the Grey model, a method was proposed to predict yearly data on air pollutants such as  $PM_{2.5}$ ,  $PM_{10}$ ,  $SO_2$ , and  $O_3$  [78]. Irulegi et al. [79] employed the NetZero energy-building technique to measure room temperature at various intervals, with sensors installed in different locations. The readings over a week showed significant temperature variations on only two days. A system [80] for predicting future levels of air pollutants ( $SO_2$ ,  $NO_2$ ,  $O_3$ ,  $PM_{2.5}$ ) was suggested based on synoptic forecasting, using data from 20 monitoring sites in Israel collected between 2002 and 2006. The processing utilized a numerical weather prediction categorization model, although the method's main flaw was constant pollutant coefficients. A time series model (ARIMA) was used to estimate  $SO_2$  and PM levels accurately, analyzing data from Hawaii obtained between January 1, 2014, and August 21, 2018, considering wind direction and speed [81].

The impact of noise on a two-DOF pitch-plunge aeroelastic system was investigated. Variations in the inlet flow velocity were modeled in [82] using the Ornstein-Uhlenbeck (O-U) stochastic process. The study utilized a 2D airfoil with linear spring-based aeroelastic characteristics to investigate Limit Cycle Oscillations (LCOs) in dynamic stall circumstances. According to the survey, adding stochastic noise causes notable qualitative changes in the Limit Cycle Oscillations (LCOs) regime and reveals intrinsic volatility absent from deterministic models. A more accurate representation of environmental

influences was obtained by introducing noise to fluid velocity, demonstrating that fluctuations in these parameters cause LCO limit values to oscillate around the mean.

The researchers in [83] used Artificial Neural Network (ANN), Gaussian process regression (GPR), Decision Tree (DT), Ensemble Learning (EL), Support Vector Machine (SVM), and Linear Regression (LR) algorithms, including optimized versions of GPR, EL, DT, and SVM, to assess the amount of CO<sub>2</sub> in an office space. It used 169 real-time data sets that included variables like air quality index, temperature, wind speed, relative humidity, occupancy, and area per person. 30% of the data was used for testing and 70% for training after it had been scaled and normalized between 0 and 1. A 5-fold cross-validation procedure was performed. With R, RMSE, MAE, NS, and 20-index values of 0.98874, 4.20068 ppm, 3.35098 ppm, 0.9817, and 1, respectively, the results demonstrated that optimized GPR achieves excellent accuracy. The model can help to estimate CO<sub>2</sub> levels and their effect on health.

Here are some suggestions and guidelines for extending the utility of these models to manage and effectively utilize large datasets, especially in scenarios involving rapid data influx:

- Assess each model's current scalability to determine how well it can handle increased data volumes without a significant loss in performance or speed. Outline the architectural limits of simpler models and identify whether enhancements or replacements are necessary for big data applications.
- Implement data preprocessing steps such as data normalization, dimensionality reduction, and feature selection to manage large datasets more efficiently before they are input into the model. Use data batching and mini-batch gradient descent for training models on large datasets to manage memory resources effectively and speed up the learning process.
- Upgrade hardware or migrate to cloud-based solutions to ensure that the infrastructure can handle the computational demands of big data analytics. Utilize distributed computing environments like Apache Hadoop or Spark to process large datasets effectively, mainly when dealing with real-time data streams.
- Incorporate streaming algorithms that can process data in real-time, which is crucial for scenarios involving sudden data rushes. Adapt models to use incremental learning techniques, where the model updates its parameters continuously as new data arrives without the need for retraining from scratch.
- Evaluate model performance under varying conditions of data volume, velocity, and variety to identify potential performance degradation or failures. Test models against scenarios of sudden data influx to gauge their responsiveness and accuracy in real-time predictions.
- Conduct stress tests to identify potential bottlenecks or weaknesses in data processing and model predictions. Discuss possible model enhancements such as hybrid architectures combining simpler models with more complex algorithms to improve handling and predictive accuracy with large datasets.
- Recommend integrating advanced machine learning techniques such as deep learning and ensemble methods, which are generally more adept at handling complex and large-scale data. Suggest employing adaptive algorithms that can adjust to changes in data patterns over time, enhancing model flexibility and durability against the variability in big data streams.

### Need for prediction

Air pollution is impacting countries all over the world. Delhi is India's capital city and one of the most polluted. Air pollution in Delhi and the NCR is often so terrible that even a slight improvement in air quality won't significantly impact it. A variety of sources causes smog and air pollution in Delhi and the NCR. Emissions from motor vehicles, industry, and building activities are among them. During the winter, the wind pattern across northern India leaves New Delhi and the NCR sensitive to air patterns. The reasons behind the increase in air pollution in Delhi, as well as the efforts taken by the government to address the issue, are detailed below.

- a. **Border Sharing:** Delhi has its border with Punjab, Haryana, and Uttar Pradesh, and rice stubble burning in any of these towns contributes to Delhi's pollution levels. The climatic elements play a vital role in exacerbating the pollution problem. For example, any unpleasant activity done by any city affects the pollution level of neighboring cities. Stubble burning is cleansing an agricultural field by burning the residue left on the soil after harvesting to prepare it for the next seeding cycle. While it is a significant factor in air pollution, it is not the sole contributor.

**Step(s) taken by government** The government is taking significant measures to address the problem of stubble burning. The authorities have created a public awareness campaign to educate people about the dangers of crop burning. New frameworks and action plans for preventing and managing stubble burning are also being established, like the Turbo Happy Seeder (THS) machine, Pusa bio decomposer, and Custom hiring centers (CHSs) are some solutions to the problem. The government has raised funds to purchase crop residue management equipment [84]. In addition, breaking the regulations can result in significant fines, prison time, or both [85].

- b. **Vehicular Emission:** One of society's significant difficulties today is pollution due to vehicle emissions. Any vehicle, such as a car, bus, truck, scooter, etc., can contribute to air pollution. The most crucial consideration is which sort of vehicle emits the least amount of pollution. Compared to other modes of transportation, the bus is a viable option since it consumes less fuel per person.

**Step(s) taken by government** The government banned the sale of diesel personal vehicles with engines larger than 2000 cc and levied a "green tax" of 4% on more prominent cars and 2.5% on compact cars. The authorities considered steps such as introducing Bharat Stage IV and Bharat Stage VI norms, electrification of vehicles, and the usage of LNG-powered buses to tackle the issue [86]. On January 1, 2016, the Delhi government welcomed the odd-even scheme, under which vehicles with odd plates were allowed to park, drive, and purchase gasoline, among other things [87–89], resulting in a reduction in traffic congestion, a slight decrease in pollution levels, and a boost in public transportation demand, including more buses and increased metro frequency. Essential actions include expanding green areas, implementing automated road sweeping, particularly close to building sites, and starting a public awareness campaign to encourage non-motorized mobility within a 5 to 6-km radius of the city.

- c. **Seasons:** In winter, dust particles and airborne contaminants become stationary. The lack of wind causes these pollutants to remain trapped in the atmosphere, impacting meteorological conditions and forming smog. In the winter, coal combustion raises

indoor and outdoor pollution levels [22–25, 27]. If meteorological factors have a negative relationship with air pollution, they also have a positive relationship.

**Step(s) taken by government** The government advises citizens to avoid going outside in the early morning and late evening when pollution levels are high. Several groups have developed products that can be purchased in stores or online to take better safeguards. Air purifiers can be used to combat interior pollution, or indoor plants can be installed within the house. For protection against outdoor pollution, masks are readily available on the market.

d. **Overpopulation:** In 2019, Delhi's population was 33 million, making it the world's fifth most populous metropolis. As the population grows, so does the need for food, land, and transportation, all of which impact pollution directly or indirectly.

**Step(s) taken by government** The government has proposed various strategies to address the growing population problem. Programs to raise awareness have been launched and are spreading across the country [90].

e. **Crackers burning on Festivals:** India is characterized by diverse ethnicities, religions, and languages. Where people joyfully celebrate each holiday. Every year during Diwali (the Festival of Lights), people ignite many crackers, contributing to pollution.

**Step(s) taken by government** The Government has introduced Green Cracker in association with CSIR (Council of Scientific and Industrial Research) [91]. Every year, the government establishes a particular time for cracker burning and awareness of the celebration, encouraging people to celebrate the festival of light rather than with fireworks. Challans/fines, a ban on cracker burning by cities, awareness against crackers, and practical hours are some steps the government considers for the betterment of the people [92].

#### Limitation of the work

- Despite many researchers on air pollutant prediction, it is still challenging to collect datasets; datasets like air pollutants, meteorological data, traffic data, and satellite data are still lacking.
- The lack of a relationship between air pollutants and weather data results in a low prediction rate. Researchers should focus more on the correlation factors as air pollution is affected by many factors.
- Predictions should be made regarding the geographical and temporal scope. All other regions should be considered to make the prediction more accurate.
- Uncertainty in data can result in the reliability of prediction.

#### Conclusion and future scope

Air pollution is a growing problem that significantly impacts current and future generations. Long-term solutions are necessary to combat this issue, not just during peak pollution months but throughout the year. Addressing major contributors to air pollution requires infrastructure for electric vehicles, enhanced public transit networks, and prioritization of renewable energy sources year-round.

This paper comprehensively reviews air quality, covering various pollutants, meteorological factors, and statistical data on death rates due to air pollution. It examines the impact of air pollution on three broad areas of society, offering readers a deeper understanding of its consequences. Additionally, the study explores several approaches to predicting air quality using machine learning and deep learning, highlighting the challenges in each methodology to guide future research.

While traditional artificial intelligence outperforms statistical methods, hybrid models show superior performance. The accuracy and robustness of chemical weather and air quality forecasts are improved via ensemble approaches and data decomposition. Effective weather prediction requires integrating spatiotemporal components, meteorological features, and geographical deliberations, which many researchers have neglected. Different techniques are needed to forecast air pollution based on contaminants and regions since no single method can guarantee accurate predictions. Large datasets can make training computationally expensive, while small datasets may reduce accuracy. Despite these challenges, AI technology is becoming a preferred alternative to traditional mathematical techniques due to its quick and reliable response.

In addition to the advantages mentioned above, this paper has some theoretical and practical implications. The study provides a deeper understanding of the mechanisms behind air pollution formation, dispersion, and transformation. Introducing new methodologies or data analysis techniques may push theoretical boundaries and set new standards in the field. It may enhance theories regarding the broader environmental impacts of air pollution. It may help regulatory agencies better monitor and enforce compliance with air quality standards, ensuring that industries and other sources adhere to regulations. The research may inform the design of public health campaigns and protective measures such as air quality alerts and recommendations for vulnerable populations. The research might drive innovation in air quality monitoring technologies, leading to more accurate and real-time data collection tools. The study can contribute to educational efforts to raise awareness about air pollution and its health impacts, encouraging behavioral changes and community action. The research might encourage collaborations between researchers, policymakers, industry leaders, and non-governmental organizations to address air pollution challenges more effectively.

Most studies reviewed used diverse performance metrics to evaluate model prediction accuracy, yet more research is needed to explore other aspects of model efficacy. Additionally, researchers often overlook other environmental contaminants impacting human health.

Despite significant progress in air quality prediction, many obstacles remain. Machine learning and deep learning advances yield good predictions, but uncertainty can undermine these results. For instance, during the COVID-19 pandemic, reduced industrial and transportation activities led to healthier air quality. Conversely, cities with poor air quality experienced higher health risks. Understanding data patterns is crucial for developing effective algorithms and raising public awareness in critical situations.

Future research should address the following gaps identified in this study:

- Insufficient research addressing uncertainty factors affecting prediction outcomes highlights the need for enhanced models to incorporate real-time data updates and feedback mechanisms.

- There is a need for a more inclusive investigation of meteorological and pollutant parameters in forecasting accuracy, which should include a thorough investigation of the formation and behavior of secondary pollutants to understand their complex dynamics.
- There are few analyses on the effectiveness of sophisticated Artificial Neural Network (ANN) models in improving computational efficiency in complex prediction systems.
- Satellite data, meteorological data, and spatial-temporal resolution of air pollution forecasts should be utilized to create a comprehensive and accurate prediction model.

#### Acknowledgements

We thank the anonymous reviewers for their insightful suggestions, which greatly enhanced the quality of this article.

#### Author contributions

Conceptualization, M.M., S.W., C.L., I.K.A., and S.A.; writing—original draft preparation, M.M., S.W., C.L., I.K.A., and S.A.; writing—review and editing, M.M., S.W., C.L., I.K.A., and S.A.; All authors have read and agreed to the published version of the manuscript.

#### Funding

This work was supported in part by the National Science and Technology Council under Grant NSC 111-2410-H-167-005-MY2 and NSC 112-2634-F-005-001-MBK.

#### Data availability

No datasets were generated or analysed during the current study.

#### Declarations

##### Institutional review board statement

Not applicable.

##### Informed consent

Not applicable.

##### Competing interests

The authors declare no competing interests.

Received: 10 June 2024 / Accepted: 22 September 2024

Published online: 09 October 2024

#### References

1. Total Population by Country. 2019 [internet]. [cited 2019 December 12]. <http://worldpopulationreview.com/countries/>
2. India Population. (2019) - Worldometers [Internet]. [cited 2019 December 12]. <https://www.worldometers.info/world-population/india-population/>
3. Air pollution claims over 10 lakh. lives every year and yet we tend to ignore it [internet]. [cited 2022 December 2]. <https://www.firstpost.com/health/air-pollution-claims-over-10-lakh-lives-every-year-and-yet-we-tend-to-ignore-it-10051431.html>
4. Air. pollution linked to 12.4L deaths in India in '17: Report | India News - Times of India [Internet]. [cited 2022 December 2]. <https://timesofindia.indiatimes.com/india/air-pollution-linked-to-12-4l-deaths-in-india-in-17-report/article-show/66978223.cms>
5. 12.4 lakh. deaths reported in India due to air pollution - India Today [Internet]. [cited 2022 December 2]. <https://www.indiatoday.in/education-today/gk-current-affairs/story/12-4-lakh-deaths-reported-in-india-due-to-air-pollution-1404411-2018-12-07>
6. Explained. India topped air pollution death toll in 2019, says report [internet]. <https://indianexpress.com/article/explained/explained-india-topped-air-pollution-death-toll-2019-7922560/>
7. 120,000 Deaths in India Due to Air Pollution. in 2020 [internet]. <https://smartairfilters.com/en/blog/120000-deaths-in-india-due-to-air-pollution-in-2020/>
8. WHO | Ambient air pollution: Health impacts [internet]. WHO, World O. 2018 [cited 2019 December 10]. <https://www.who.int/teams/environment-climate-change-and-health/air-quality-energy-and-health/sectoral-interventions/ambient-air-pollution/health-risks#:~:text=4.2%20million%20people%20die%20prematurely,and%206%25%20to%20lung%20cancer>
9. Kazemi Z, Jonidi Jafari A, Farzadkia M, Amini P, Kermani M. Evaluating the mortality and health rate caused by the PM2.5 pollutant in the air of several important Iranian cities and evaluating the effect of variables with a linear time series model. *Heliyon*. 2024;10:e27862.
10. WHO | Household air pollution. Health impacts [internet]. [cited 2019 December 10]. [https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health#:~:text=The%20combined%20effects%20of%20ambient,\(COPD\)%20and%20lung%20cancer](https://www.who.int/news-room/fact-sheets/detail/household-air-pollution-and-health#:~:text=The%20combined%20effects%20of%20ambient,(COPD)%20and%20lung%20cancer)



11. Bai L, He Z, Li C, Chen Z. Investigation of yearly indoor/outdoor PM<sub>2.5</sub> levels in the perspectives of health impacts and air pollution control: Case study in Changchun, in the northeast of China. *Sustain Cities Soc.* 2020;53:101871.
12. Cincinelli A, Martellini T. Indoor air quality and health. *Int J Environ Res Public Health.* MDPI AG; 2017.
13. Alves CA, Vicente ED, Evtugina M, Vicente AM, Nunes T, Lucarelli F, Calzolari G, Nava S, Calvo AI, del Blanco Alegre C, Oduber F. Indoor and outdoor air quality: A university cafeteria as a case study. *Atmospheric Pollution Res.* 2020;11(3):531–44.
14. Alves C, Nunes T, Silva J, Duarte M. Comfort parameters and Particulate Matter (PM<sub>10</sub> and PM<sub>2.5</sub>) in School classrooms and Outdoor Air. *Aerosol Air Qual Res.* 2013;13:1521–35.
15. Buczyńska AJ, Krata A, Van Grieken R, Brown A, Polezer G, De Wael K, et al. Composition of PM<sub>2.5</sub> and PM<sub>10</sub> on high and low pollution event days and its relation to indoor air quality in a home for the elderly. *Sci Total Environ.* 2014;490:134–43.
16. Bennett J, Davy P, Trompetter B, Wang Y, Piersie N, Boulic M, et al. Sources of indoor air pollution at a New Zealand urban primary school; a case study. *Atmos Pollut Res.* 2019;10:435–44.
17. Scibor M. Are we safe inside? Indoor air quality in relation to outdoor concentration of PM<sub>10</sub> and PM<sub>2.5</sub> and to characteristics of homes. *Sustain Cities Soc.* 2019;48.
18. Chamseddine A, Alameddine I, Hatzopoulou M, El-Fadel M. Seasonal variation of air quality in hospitals with indoor–outdoor correlations. *Build Environ.* 2019;148:689–700.
19. Malhotra M, Aulakh IK, Kaur N, Aulakh NS. Air Pollution Monitoring Through Arduino Uno. *Advances in Intelligent Systems and Computing.* 2020.
20. Kumar A, Malhotra S, Kaur DP, Gupta L. Weather Monitoring and Air Quality Prediction using Machine Learning. 2022 1st International Conference on Computational Science and Technology, ICCST 2022 - Proceedings. Institute of Electrical and Electronics Engineers Inc.; 2022. pp. 364–8.
21. Pandithurai O, Bharathiraja N, Pradeepa K, Meenakshi D, Kathiravan M, Vinoth Kumar M. Air Pollution Prediction using Supervised Machine Learning Technique. *Proceedings of the 3rd International Conference on Artificial Intelligence and Smart Energy, ICAIS 2023.* Institute of Electrical and Electronics Engineers Inc.; 2023. pp. 542–6.
22. Wang J, Ogawa S. Effects of meteorological conditions on PM<sub>2.5</sub> concentrations in Nagasaki, Japan. *Int J Environ Res Public Health.* 2015;12:9089–101.
23. Giri D, Krishna Murthy V, Adhikary PR. The influence of meteorological conditions on PM<sub>10</sub> concentrations in Kathmandu Valley. *Int J Environ Res.* 2008;2:49–60.
24. Zyromski A, Biniak-Pieróg M, Burszta-Adamiak E, Zamiar Z. Evaluation of relationship between air pollutant concentration and meteorological elements in winter months. *J Water Land Dev.* 2014;22:25–32.
25. Yang Q, Yuan Q, Li T, Shen H, Zhang L. The relationships between PM<sub>2.5</sub> and meteorological factors in China: seasonal and regional variations. *Int J Environ Res Public Health.* 2017;14(12):1510.
26. Malhotra M, Aulakh IK. Meteorological Factors Correlation with Air pollutants: a Case Study in Delhi. *Int J Environ Sci Dev.* 2023;14:91–105.
27. Kayes I, Shahrir SA, Hasan K, Akhter M, Kabir MM, Salam MA. The relationships between meteorological parameters and air pollutants in an urban environment. *Global J Environ Sci Manage.* 2019;5:265–78.
28. Lin X, Fu Y, Peng DZ, Liu C-H, Chu M, Chen Z, et al. CFD- and BPNN- based investigation and prediction of air pollutant dispersion in urban environment. *Sustain Cities Soc.* 2024;100:105029.
29. Yang F, Huang G. An optimized decomposition integration model for deterministic and probabilistic air pollutant concentration prediction considering influencing factors. *Atmos Pollut Res.* 2024;15:102144.
30. Shaban WM, Dongxi X, Daef KS, Elbaz K. Real-time early warning and the prediction of air pollutants for sustainable development in smart cities. *Atmos Pollut Res.* 2024;15:102162.
31. Monteiro TO, Alves PAA da S, de AN, Barradas Filho AO, Villa-Vélez HA, Cruz G. Estimation of the main air pollutants from different biomasses under combustion atmospheres by artificial neural networks. *Chemosphere.* 2024;352:141484.
32. Zhou Y, Chang FJ, Chang LC, Kao IF, Wang YS, Kang CC. Multi-output support vector machine for regional multi-step-ahead PM<sub>2.5</sub> forecasting. *Sci Total Environ.* 2019;651:230–40.
33. Ahani IK, Salari M, Shadman A. Statistical models for multi-step-ahead forecasting of fine particulate matter in urban areas. *Atmospheric Pollution Res.* 2019;10(3):689–700.
34. Cabaneros SMLS, Calautit JKS, Hughes BR. Hybrid Artificial Neural Network Models for Effective Prediction and Mitigation of Urban Roadside NO<sub>2</sub> Pollution. *Energy Procedia.* 2017;142:3524–30.
35. Wang P, Zhang H, Qin Z, Zhang G. A novel hybrid-Garch model based on ARIMA and SVM for PM<sub>2.5</sub> concentrations forecasting. *Atmospheric Pollution Research.* 2017;8(5):850–60.
36. Catalano M, Galatioto F, Bell M, Namdeo A, Bergantino AS. Environmental Science & Policy Improving the prediction of air pollution peak episodes generated by urban transport networks. *Environ Sci Policy.* 2016;60:69–83.
37. Feng X, Li Q, Zhu Y, Hou J, Jin L, Wang J. Artificial neural networks forecasting of PM<sub>2.5</sub> pollution using air mass trajectory based geographic model and wavelet transformation. *Atmos Environ.* 2015;107:118–28.
38. Cortina-Januchs MG, Quintanilla-Dominguez J, Vega-Corona A, Andina D. Development of a model for forecasting of PM<sub>10</sub> concentrations in Salamanca, Mexico. *Atmos Pollut Res.* 2015;6:626–34.
39. Wahid HA, Ha QP, Duc H, Azzi M. Neural network-based meta-modelling approach for estimating spatial distribution of air pollutant levels. *Applied Soft Comput.* 2013;13(10):4087–96.
40. Domańska D, Wojtylak M. Application of fuzzy time series models for forecasting pollution concentrations. *Expert Syst Appl.* 2012;39:7673–9.
41. Siwek K, Osowski S. Engineering Applications of Artificial Intelligence improving the accuracy of prediction of PM<sub>10</sub> pollution by the wavelet transformation and an ensemble of neural predictors. *Eng Appl Artif Intell.* 2012;25:1246–58.
42. Feng Y, Zhang W, Sun D, Zhang L. Ozone concentration forecast method based on genetic algorithm optimized back propagation neural networks and support vector machine data classification. *Atmos Environ.* 2011;45:1979–85.
43. Liu H, Wu H, Lv X, Ren Z, Liu M, Li Y, et al. An intelligent hybrid model for air pollutant concentrations forecasting: case of Beijing in China. *Sustain Cities Soc.* 2019;47:101471.
44. Kumar D. ScienceDirect ScienceDirect Evolving Differential evolution method with random forest for Evolving Differential evolution method with random forest for prediction of Air Pollution prediction of Air Pollution. *Procedia Comput Sci.* 2018;132:824–33.
45. Alimisis A, Philippopoulos K, Tzani CG, Deligiorgi D. Spatial estimation of urban air pollution with the use of artificial neural network models. *Atmos Environ.* 2018;191:205–13.

46. Fan J, Wu L, Zhang F, Cai H, Wang X, Lu X. Evaluating the effect of air pollution on global and diffuse solar radiation prediction using support vector machine modeling based on sunshine duration and air temperature. *Renew Sustain Energy Rev.* 2018;94:732–47.
47. Huang L, Zhang C, Bi J. Development of land use regression models for PM<sub>2.5</sub>, SO<sub>2</sub>, and O<sub>3</sub> in. *Environ Res.* 2017;158:542–52.
48. Corani G, Scanagatta M. Air pollution prediction via multilabel classification. *Environ Model Softw.* 2016;80:259–64.
49. Olvera-garcía MÁ, Carbajal-hernández JJ, Sánchez-fernández LP, Hernández-bautista I. Ecological Informatics Air quality assessment using a weighted fuzzy inference system. *Ecol Inf.* 2016;33:57–74.
50. Bai Y, Li Y, Wang X, Xie J, Li C. Air pollutants concentrations forecasting using back propagation neural network based on wavelet decomposition with meteorological conditions. *Atmospheric pollution res.* 2016;7(3):557–66.
51. Kumar K, Pande BP. Air pollution prediction with machine learning: a case study of Indian cities. *International Journal of Environmental Science and Technology* [Internet]. 2023;20:5333–48. <https://link.springer.com/https://doi.org/10.1007/s13762-022-04241-5>
52. Lin Y-C, Lin Y-T, Chen C-R, Lai C-Y. Meteorological and traffic effects on air pollutants using bayesian networks and deep learning. *J Environ Sci.* 2025;152:54–70.
53. Wu Z, Tian Y, Li M, Wang B, Quan Y, Liu J. Prediction of air pollutant concentrations based on the long short-term memory neural network. *J Hazard Mater.* 2024;465:133099.
54. Li D, Wang J, Tian D, Chen C, Xiao X, Wang L, et al. Residual neural network with spatiotemporal attention integrated with temporal self-attention based on long short-term memory network for air pollutant concentration prediction. *Atmos Environ.* 2024;329:120531.
55. Wu Z, Tian Y, Li M, Wang B, Quan Y, Liu J. Prediction of air pollutant concentrations based on the long short-term memory neural network. *J Hazard Mater.* 2023;133099.
56. Ma Z, Wang B, Luo W, Jiang J, Liu D, Wei H, et al. Air pollutant prediction model based on transfer learning two-stage attention mechanism. *Sci Rep.* 2024;14:7385.
57. Bekkar A, Hssina B, Douzi S, Douzi K. Air-pollution prediction in smart city, deep learning approach. *J Big Data* [Internet]. 2021;8:161. <https://journalofbigdata.springeropen.com/articles/https://doi.org/10.1186/s40537-021-00548-1>
58. Drevil GI, Al-Bahadili RJ. Air pollution prediction using LSTM deep learning and metaheuristics algorithms. *Measurement: Sensors* [Internet]. 2022;24:100546. <https://linkinghub.elsevier.com/retrieve/pii/S2665917422001805>
59. Yang G, Lee H, Lee G. A hybrid deep learning model to forecast particulate matter concentration levels in Seoul, South Korea. *Atmosphere.* 2020;11(4):348.
60. Wang J, Xu W, Zhang Y, Dong J. A novel air quality prediction and early warning system based on combined model of optimal feature extraction and intelligent optimization. *Chaos Solitons Fractals.* 2022;158:112098.
61. Mao W, Jiao L, Wang W, Wang J, Tong X, Zhao S. A hybrid integrated deep learning model for predicting various air pollutants. *Gisci Remote Sens.* 2021;58:1395–412.
62. Heydari A, Majidi Nezhad M, Astiaso Garcia D, Keynia F, De Santoli L. Air pollution forecasting application based on deep learning model and optimization algorithm. *Clean Technol Environ Policy.* 2022;24:607–21.
63. Mokhtari I, Bechkit W, Rivano H, Yaici MR. Uncertainty-aware deep learning architectures for highly dynamic air Quality Prediction. *IEEE Access.* 2021;9:14765–78.
64. Wang J, Xu W, Dong J, Zhang Y. Two-stage deep learning hybrid framework based on multi-factor multi-scale and intelligent optimization for air pollutant prediction and early warning. *Stoch Env Res Risk Assess.* 2022;36:3417–37.
65. Zhou X, Xu J, Zeng P, Meng X. Air Pollutant Concentration Prediction based on GRU Method. *J Phys Conf Ser.* 2019;1168.
66. Loy-Benitez J, Vilela P, Li Q, Yoo C. Sequential prediction of quantitative health risk assessment for the fine particulate matter in an underground facility using deep recurrent neural networks. *Ecotoxicol Environ Saf* [Internet]. 2019;169:316–24. <https://linkinghub.elsevier.com/retrieve/pii/S0147651318311606>
67. Zhao J, Deng F, Cai Y, Chen J. Chemosphere Long short-term memory - fully connected (LSTM-FC) neural network for PM<sub>2.5</sub> concentration prediction. *Chemosphere.* 2019;220:486–92.
68. Wen C, Liu S, Yao X, Peng L, Li X, Hu Y, et al. Science of the total environment a novel spatiotemporal convolutional long short-term neural network for air pollution prediction. *Sci Total Environ.* 2019;654:1091–9.
69. Qi Y, Li Q, Karimian H, Liu D. A hybrid model for spatiotemporal forecasting of PM<sub>2.5</sub> based on graph convolutional neural network and long short-term memory. *Sci Total Environ.* 2019;664:1–10.
70. Athira V, Geetha P, Vinayakumar R, Soman KP. ScienceDirect ScienceDirect DeepAirNet: applying recurrent networks for Air Quality Prediction. *Procedia Comput Sci.* 2018;132:1394–403.
71. Wang J, Song G. Neurocomputing A, Deep. Spatial-Temporal Ensemble Model for Air Quality Prediction. *Neurocomputing.* 2018;314:198–206.
72. Li X, Peng L, Yao X, Cui S, Hu Y, You C. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation \*. *Environ Pollut.* 2017;231:997–1004.
73. Zhou Y, Chang F, Chang L, Kao J, Wang Y. Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts. *J Clean Prod.* 2019;209:134–45.
74. Sharma N, Taneja S, Sagar V, Bhatt A. Forecasting air pollution load in Delhi using data analysis tools. *Procedia Comput Sci. Elsevier BV;* 2018. pp. 1077–85.
75. Kök I, Şimşek MU, Özdemir S. A deep learning model for air quality prediction in smart cities. *Proceedings – 2017 IEEE International Conference on Big Data, Big Data.* 2017. pp. 1983–90.
76. Wang D, Wei S, Luo H, Yue C, Grunder O. Science of the total environment a novel hybrid model for air quality index forecasting based on two-phase decomposition technique and modified extreme learning machine. *Sci Total Environ.* 2017;580:719–33.
77. Zhu S, Lian X, Liu H, Hu J, Wang Y, Che J. Daily air quality index forecasting with hybrid models: a case in. *Environ Pollut.* 2017;231:1232–44.
78. Wu L, Li N, Yang Y. Prediction of air quality indicators for the Beijing-Tianjin-Hebei region. 2018;196:682–7.
79. Irulegi O, Serra A, Hernández R. Data in brief data on records of indoor temperature and relative humidity in a University building. *Data Brief.* 2017;13:248–52.
80. Broday DM, Alpert P. Exploring the applicability of future air quality predictions based on synoptic system forecasts. *Environ Pollut.* 2012;166:65–74.

81. Reikard G. Atmospheric Environment: X volcanic emissions and air pollution : forecasts from time series models. *Atmos Environ X*. 2019;1:100001.
82. Ketseas D. Stochastic response of an Airfoil and its effects on Lco's Behavior under Stall Flutter Regime. *Int J Math Stat Comput Sci*. 2024;2:168–72.
83. Kapoor NR, Kumar A, Kumar A, Kumar A, Mohammed MA, Kumar K et al. Machine Learning-Based CO<sub>2</sub>Prediction for Office Room: A Pilot Study. *Wirel Commun Mob Comput*. 2022;2022.
84. Measures to Reduce Pollution Due to Stubble Burning. 2020. pp. 151–6.
85. All about stubble. burning, its alternatives and steps taken by Centre and state govts.
86. A look at. key govt initiatives to keep air pollution from vehicles under check.
87. Odd–even rationing. - Wikipedia [Internet]. [cited 2019 December 9]. [https://en.wikipedia.org/wiki/Odd–even\\_rationing](https://en.wikipedia.org/wiki/Odd–even_rationing)
88. What is Delhi's new odd-even vehicle rule all about? Where did it come from? *India Today*. 2015.
89. Pollution in Delhi dips 62%. in one day, thanks to high wind speed and odd-even [internet]. [cited 2019 December 10]. <https://www.indiatoday.in/diu/story/pollution-in-delhi-dips-62-in-one-day-thanks-to-high-wind-speed-and-odd-even-1615717-2019-11-04>
90. Population Control [Internet]. [cited 2022 June 7]. <https://pib.gov.in/newsite/PrintRelease.aspx?relid=194837>
91. India launches. Green Crackers in its bid to curb air pollution [internet]. [cited 2022 June 7]. <https://pib.gov.in/newsite/PrintRelease.aspx?relid=193646>
92. No Firecracker, Only Green Crackers This Diwali. : How These States Taking Measures to Control Air Pollution [Internet]. [cited 2022 June 7]. <https://www.india.com/news/india/no-firecracker-only-green-crackers-this-diwali-how-these-states-taking-measures-to-control-air-pollution-5072307/>

### **Publisher's note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.