

Systematic Review

# Supervised Machine Learning Approaches for Predicting Key Pollutants and for the Sustainable Enhancement of Urban Air Quality: A Systematic Review

Ismail Essamlali \*, Hasna Nhaila and Mohamed El Khaili 

Electrical Engineering and Intelligent Systems Laboratory, ENSET Mohammedia, Hassan 2nd University of Casablanca, Mail Box 159, Mohammedia 28810, Morocco; hasna.nhaila@enset-media.ac.ma (H.N.); elkhalil@enset-media.ac.ma (M.E.K.)

\* Correspondence: essamlali.ismail01@gmail.com

**Abstract:** Urban air pollution is a pressing global issue driven by factors such as swift urbanization, population expansion, and heightened industrial activities. To address this challenge, the integration of Machine Learning (ML) into smart cities presents a promising avenue. Our article offers comprehensive insights into recent advancements in air quality research, employing the PRISMA method as a cornerstone for the reviewing process, while simultaneously exploring the application of frequently employed ML methodologies. Focusing on supervised learning algorithms, the study meticulously analyzes air quality data, elucidating their unique benefits and challenges. These frequently employed ML techniques, including LSTM (Long Short-Term Memory), RF (Random Forest), ANN (Artificial Neural Networks), and SVR (Support Vector Regression), are instrumental in our quest for cleaner, healthier urban environments. By accurately predicting key pollutants such as particulate matter (PM), nitrogen oxides (NO<sub>x</sub>), carbon monoxide (CO), and ozone (O<sub>3</sub>), these methods offer tangible solutions for society. They enable informed decision-making for urban planners and policymakers, leading to proactive, sustainable strategies to combat urban air pollution. As a result, the well-being and health of urban populations are significantly improved. In this revised abstract, the importance of frequently employed ML methods in the context of air quality is explicitly emphasized, underlining their role in improving urban environments and enhancing the well-being of urban populations.

**Keywords:** pollution monitoring; air quality; supervised; machine learning



**Citation:** Essamlali, I.; Nhaila, H.; El Khaili, M. Supervised Machine Learning Approaches for Predicting Key Pollutants and for the Sustainable Enhancement of Urban Air Quality: A Systematic Review. *Sustainability* **2024**, *16*, 976. <https://doi.org/10.3390/su16030976>

Academic Editors: Norazian Mohamed Noor, Ahmad Zia Ul-Saufie Mohamad Japeri and Mohd Remy Rozainy Mohd Arif Zainol

Received: 7 October 2023  
Revised: 13 November 2023  
Accepted: 17 January 2024  
Published: 23 January 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In recent years, there has been a significant increase in public awareness of the effects of population growth and activity on the environment. The United Nations estimates that there are currently 7.6 billion people living in the world, with 4.2 billion (or around 55%) of them living in cities [1]. By 2050, it is estimated that the urban population will double compared to its current size. This rapid growth in urban areas has led to a significant increase in industrialization to meet the demands of the expanding population.

However, this rapid pace of industrial and economic development has led to harmful repercussions for the environment, particularly in terms of air quality, posing a significant challenge to both human health and the environment. The influx of people into cities and the subsequent increase in industrial activity have resulted in elevated levels of air pollutants. These pollutants are often released from vehicular emissions, power generation, and manufacturing processes, accumulating in the atmosphere and contributing to the formation of harmful smog and haze. As a consequence, urban residents are increasingly exposed to poor air quality, which can lead to a range of respiratory and cardiovascular ailments. Additionally, the long-range transport of these pollutants can adversely affect not only local communities but also distant regions, amplifying the global scale of the

issue. Urgent measures are required to mitigate air quality degradation, including the implementation of cleaner technologies, enhanced urban planning, and the promotion of sustainable transportation systems to ensure a healthier future for both the planet and its inhabitants.

The concept of a smart city offers an opportunity to effectively and efficiently address these challenges. With air pollution posing a threat to urban living, it is crucial to address this issue promptly and wisely. A smart city refers to an urban locality that leverages information and communication technologies (ICT) to enhance the quality of life for its residents. It achieves this by offering improved healthcare, transportation, and energy services while enabling the government to optimize resource utilization for the well-being of its citizens.

In the context of air quality research, it is essential to recognize the value of many qualitative tools. These tools include chemical analysis: Ambient Sampling, Source Profiling, and Receptor Modeling, which hold profound significance in the field of air quality management. These methodologies offer a systematic approach to comprehending the multifaceted nature of air pollution, facilitating the precise identification and quantification of pollution sources. By combining the insights derived from chemical analysis with rigorous statistical techniques, these models equip environmental scientists and policymakers with a robust foundation for devising targeted pollution control strategies, establishing regulatory benchmarks, and executing efficacious measures to ameliorate air quality. Their pivotal role in safeguarding public health, preserving ecological integrity, and promoting sustainable development through the foundational mitigation of air pollution renders them indispensable tools in contemporary environmental science and policy paradigms. As such, these qualitative models occupy a central place in the arsenal of methodologies available for addressing the complex challenge of air quality management [2].

Moreover, the convergence of ICT with ML presents substantial potential for fostering groundbreaking solutions. An area that particularly stands out in smart city applications is the enhancement of air quality monitoring. ML algorithms demonstrate the capability to effectively interpret and analyze extensive datasets concerning air quality, sourced from diverse channels such as sensors and weather stations [3]. In the field of ML, we encounter a trio of techniques: supervised, unsupervised, and semi-supervised methods. Supervised ML entails instructing a model using meticulously labeled data, where the expected outcome is already discerned. On the other hand, unsupervised ML embarks on the uncharted terrain of unlabeled data, excelling in tasks such as categorizing data groups into clusters founded on shared characteristics. The hybrid amalgam known as semi-supervised ML presents an appealing avenue, especially in situations where labeled data are sparse. As a result of this analysis, these algorithms can discern patterns, predict pollution levels, and offer real-time monitoring of the city's air quality status.

The motivation behind this study is multifaceted and driven by a confluence of factors. First and foremost, there is a growing global concern for environmental well-being, with a pressing need to address the adverse impacts of air pollution on ecosystems and wildlife. Additionally, public health is a paramount consideration, as air pollution poses significant health risks to communities, making it imperative to gain a deeper understanding of the issue. Policy and regulatory frameworks related to air quality are of great importance, and this study seeks to provide evidence-based insights to inform and shape effective policies. Furthermore, the study is motivated by the remarkable technological advancements in supervised learning and data analytics, which offer new avenues for air quality research. It also aligns with urban planning efforts, aiming to create cities that minimize air pollution and enhance residents' quality of life. In light of global sustainability goals and commitments, this research contributes to broader objectives. Moreover, there is a genuine aspiration for scientific advancement and the pursuit of interdisciplinary collaboration, uniting fields such as environmental science, data science, and public health to tackle the complex challenge of air quality comprehensively.

This study aims to conduct a thorough review of recent literature on air pollution, with a specific focus on the use of supervised learning methods in the field and the identification of key pollutants. We aspire to identify and highlight the most prevalent pollutants and the most employed techniques. We commence with an examination of related work (Section 2) and describe our study methodology within the context of prior research (Section 3). Section 4 delves into the application of supervised ML techniques in air quality monitoring, while Section 5 discusses our analysis results, limitations, and potential future research directions. Finally, Section 6 summarizes key findings and conclusions.

## 2. Related Works

### 2.1. Qualitative Approaches in the Air Quality Field

A wide range of qualitative approaches work together in the complex field of environmental science and air quality management to provide a deep comprehension and efficient reduction of air pollution. These essential instruments include the critical procedures of Chemical Analysis, Source Profiling, Receptor Modeling, and Ambient Sampling. Ambient Sampling is the first step in the process, which is a fundamental step that involves the systematic collection of air samples from various geographic locations. With the use of this methodical technique, it is possible to obtain discrete air quality snapshots and perform detailed quantification of a wide range of contaminants, including heavy metals, volatile organic compounds, and particle matter. The next step involves subjecting the gathered samples to a thorough chemical analysis, which reveals a wealth of information about the makeup of the air environment and the existence of various pollutants. After that, Source Profiling takes the lead in locating and describing the potential sources of these contaminants. A variety of methodological techniques are used by environmental scientists to track the sources of pollutants identified in the samples they have gathered. This involves a multifaceted approach that amalgamates data from chemical analysis, isotopic inquiry, and meteorological insights, enabling the discrimination between natural contributors, such as dust and sea salt, and anthropogenic sources, encompassing industrial emissions, vehicular traffic, and residential heating. Source Profiling stands as an instrumental undertaking in ascertaining the principal contributors to air pollution. Finally, Receptor Modeling assumes a pivotal role, capitalizing on the outcomes of chemical analysis to meticulously allot responsibility among the varied pollution sources. By employing a blend of statistical and mathematical techniques that meticulously consider the chemical composition of pollutants, Receptor Modeling meticulously unravels the intricate amalgam found within the samples. This necessitates the application of methodologies such as chemical mass balance, positive matrix factorization (PMF), and factor analysis, all directed toward delivering a comprehensive dissection of the contributions from disparate sources. These receptor models not only foster a more profound understanding but also serve as invaluable resources for both environmental agencies and policymakers, endowing them with the requisite insights to develop meticulously precise pollution control strategies and regulatory frameworks, all directed toward the augmentation of air quality [2].

### 2.2. Air Quality Analysis and Forecasting

Multiple reviews have illuminated various facets of air quality analysis, forecasting, and related domains. Notably, the role of data quality in constructing effective Air Quality Models (AQMs), highlights challenges stemming from limited data availability [4]. Some studies provide a contemporary overview of air pollution forecasting, spotlighting models like the Air Quality Index (AQI) and discussing predictor variables and emerging trends [5]. In the domain of air quality prediction, there has been a substantial body of research focused on ML algorithms, with particular attention given to Artificial Neural Networks (ANN). The primary objective of this research was to elucidate the connection between the Air Quality Index (AQI) and various ML approaches, including ANN and Multiple Linear Regression (MLR) [6]. Furthermore, the investigation extends to encompass ANN, Deep Neural Networks (DNN), Support Vector Machines (SVM), and Fuzzy Logic, highlighting

the need for further inquiry, particularly within the realm of DNN [7]. This research journey culminates in an assessment of data decomposition techniques in conjunction with ML models such as ANN, SVM, and the Emerging Learning Machine (ELM) for air pollution forecasting. This underscores the sustained endeavor to craft more precise and efficacious strategies to combat the pressing issue of air pollution [8].

Deep learning has gained significant prominence in the field of air pollution epidemiology, a trend that becomes readily apparent in the reviewed literature. One study [9] delves into the potential of deep learning for source apportionment and forecasting, with an emphasis on its spatial and temporal analysis capabilities. Additionally, another study [10] provides a comparative analysis of non-deep and deep learning approaches in modeling air pollutant correlations. Furthermore, the application of deep learning takes center stage in [11], focusing specifically on its utilization for time series air quality forecasting.

Shifting the focus to indoor environments, the versatility of ML comes to the forefront. In [12], an extensive review examines studies employing ML to predict occupancy behavior and patterns with a focus on its related applications to indoor air quality and thermal comfort, favoring algorithms grounded in neural networks. Furthermore, Ref. [13] establishes a connection between ML and household cooking practices, exploring its implications for carbon neutrality. In parallel, Ref. [14] conducts a comprehensive survey of ML applications in heating, ventilation, and air conditioning (HVAC) systems and their impact on building performance. Lastly, Ref. [15] employs ANN and Reinforcement Learning (RL) models to investigate indoor air quality and thermal comfort.

### *2.3. Machine Learning in Urban and Industrial Planning*

Other reviews have highlighted the substantial role of Machine Learning (ML) models in analyzing air quality data in city planning and urban sustainability. By identifying areas with poor air quality and comprehending the contributing factors, city planners can make informed decisions to optimize land use, transportation infrastructure, and green spaces, ultimately enhancing the quality of life for urban residents. These reviews emphasize the importance of ML in advancing urban sustainability and smart city development [16,17].

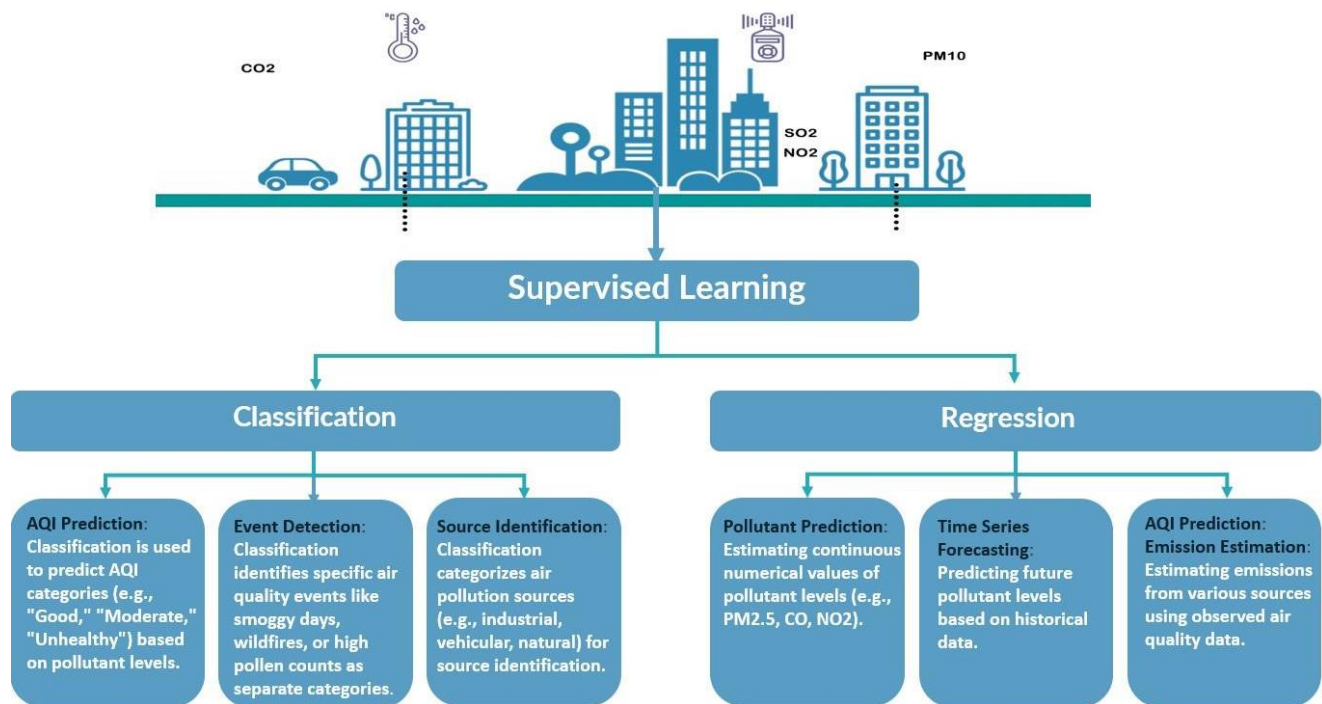
At the dynamic intersection of industrial and urban planning, Machine Learning (ML) has emerged as a transformative and versatile tool, as affirmed by numerous comprehensive reviews. These comprehensive analyses accentuate the pivotal role that Machine Learning plays in the context of Industry 4.0, where data-driven decision-making takes center stage [18]. Notably, some reviews shed light on Machine Learning's influential applications in optimizing energy efficiency within the industrial sector [19]. They also provide valuable insights into the practical challenges and opportunities encountered when implementing large-scale Machine Learning systems in real-world industrial settings [20]. Collectively, these reviews underscore the profound impact of Machine Learning on industrial planning, heralding greater efficiency and sustainability within the dynamic landscape of modern manufacturing.

### *2.4. Machine Learning in Climate Change Context*

This interconnected web of research showcases the diverse applications of ML within urban and industrial planning. However, it is essential to recognize that the influence of ML extends beyond these domains, finding relevance within the pressing global challenge of climate change. The transition to this broader context is imperative. As articulated in [21], ML lends its capabilities to the exploration of spatial techniques in climate change research, offering new avenues for understanding and mitigating climate-related challenges. Additionally, Ref. [22] demonstrates how ML can be harnessed to discern the multifaceted impacts of climate change on public health, particularly emphasizing disparities in evidence distribution. This seamless bridge from urban and industrial planning to the wider climate change landscape underscores the indispensable and interconnected role of Machine Learning in addressing contemporary global challenges.

These methods, including qualitative ones, play a crucial role in addressing climate change and air pollution mitigation while aligning with the Sustainable Development Goals (SDGs). They offer precision in identifying and characterizing pollution sources, enabling targeted mitigation efforts and emission reductions, supporting public health (SDG 3), fostering sustainable urban planning (SDG 11), and contributing to climate action by reducing greenhouse gas emissions (SDG 13). Moreover, they have broader positive impacts on terrestrial biodiversity (SDG 15) and exemplify the importance of partnerships (SDG 17) among governments, industries, and the scientific community for effective implementation and achieving the SDGs [23].

These reviews collectively deepen our understanding of various aspects encompassing air quality analysis, forecasting techniques, and the associated challenges, underscoring the pivotal role prediction plays in safeguarding human health, the environment, and energy efficiency. However, our approach diverges as we concentrate on investigating the implementation of Supervised ML methods in the context of air quality (Figure 1). This distinct perspective redirects our attention toward the application of Supervised ML techniques, yielding novel insights into comprehending air quality analysis and forecasting, accompanied by unique challenges that demand attention.

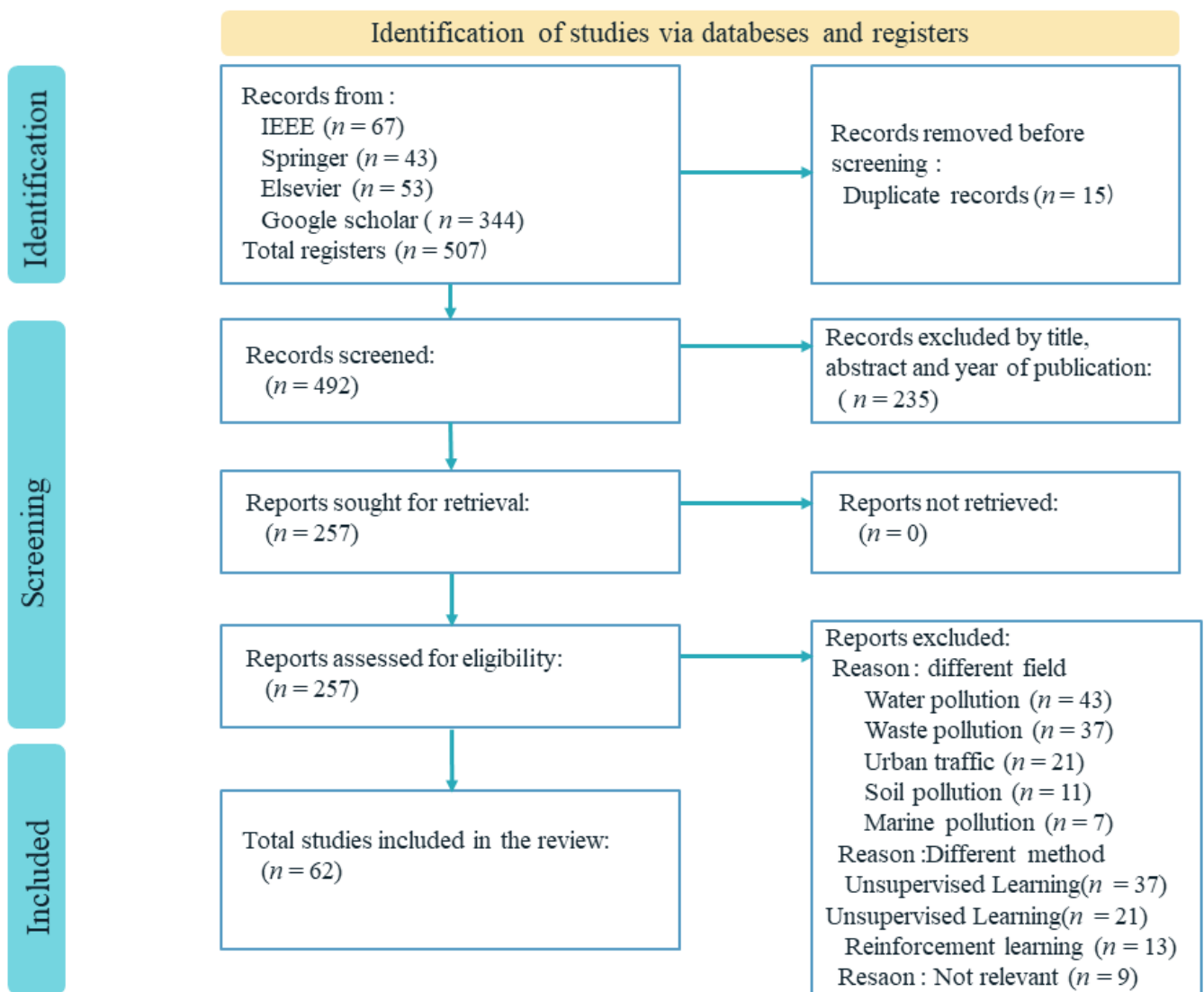


**Figure 1.** Supervised learning approaches in air quality monitoring.

### 3. Method

The study followed the PRISMA guidelines [24], and a detailed checklist is provided in Table S1 in the Supplementary Material. PRISMA ensures a structured and transparent approach to comprehensive literature reviews, enhancing rigor and reproducibility. It involves defining research questions, selecting studies, assessing quality, and synthesizing findings. By applying PRISMA, we collected recent articles on ML and urban air quality, focusing on environmental challenges in urban areas. We analyzed articles using supervised ML evaluating algorithm choice, preprocessing, and more. This yielded 59 relevant articles for our comprehensive analysis (Figure 2).





**Figure 2.** The PRISMA method used in the reviewing process.

### 3.1. Database Collection

The review process commenced with a comprehensive collection of academic articles related to ML and urban quality. Reputable databases such as Google Scholar, MDPI, IEEE Xplore, Springer, and Elsevier were utilized to access a diverse array of journal articles. To maintain the review's relevance and currency, the focus was on recent publications.

### 3.2. Initial Selection

The initial selection phase involved a meticulous process of narrowing down a pool of approximately 507 academic articles obtained from reputable databases such as Google Scholar, MDPI, IEEE Xplore, Springer, and Elsevier. This aimed to identify articles aligned with the research objective, focusing on titles and abstracts that explored the application of ML methods (supervised, unsupervised, reinforcement, and semi-supervised) in addressing environmental challenges in the context of urban air quality. To maintain relevance and currency, emphasis was placed on recent publications, and this phase served the purpose of refining the initial dataset to a more manageable set of articles by the research criteria.

### 3.3. Preliminary Screening

Within the selected articles, special emphasis was placed on the concept of urban air quality. Articles on urban air quality typically cover sources of pollution, health impacts,

environmental sustainability, monitoring methods, regulatory aspects, mitigation strategies, community engagement, technological solutions, case studies, and policy recommendations. The review delved deeper into the selected articles, conducting a comprehensive evaluation of the employed supervised ML methodologies. Key factors such as algorithm selection, data preprocessing techniques, feature engineering, and model evaluation, were closely examined to understand the efficacy and robustness of the approaches.

### 3.4. Assessment and Retrieval

In the selection of final articles for this review, a meticulous approach was adopted to ensure the inclusion of high-quality, indexed research with a diverse representation of supervised learning methods. The chosen articles spanned a wide range of methodologies, including classification, regression, deep learning, and hybrid models, providing a comprehensive examination of their applications in the context of urban air quality research. Additionally, special consideration was given to incorporating various pollutants, such as PM<sub>10</sub>, PM<sub>2.5</sub>, SO<sub>2</sub>, CO<sub>2</sub>, and AQI, thus offering insights into the multifaceted challenges of air pollution management. This stringent selection process resulted in the compilation of 62 articles that collectively form the foundation for a comprehensive and insightful analysis of the intersection between supervised learning and diverse pollutants in urban air quality research.

### 3.5. Synthesis and Presentation

The reviewed articles were synthesized to create a coherent and well-structured narrative. The review commenced with an introduction that outlined the importance of applying ML in the context of urban air quality research. It then proceeded to explore the various supervised ML approaches used to tackle urban challenges and their implications on urban air quality outcomes. Special attention was given to highlighting the most noteworthy studies and their potential impact on future research and urban planning strategies.

## 4. Air Quality Monitoring with Supervised Learning

### 4.1. Air Quality Field

#### 4.1.1. Air Quality Landscape

Economic expansion leads to urbanization and industrialization, contributing to environmental decline. The United Nations notes that 55% of the global population (4.2 billion) resides in cities, with an anticipated doubling by 2050 [1]. Such rapid urban growth heightens energy demands, driving resource consumption and carbon emissions [25,26]. Urbanization is consistently linked to environmental deterioration [27].

Industrialization, a pivotal contributor to pollution, exhibits a direct impact on emissions. A 1% industrialization rise is associated with an 11% per capita emissions increase [28]. This has immediate and long-term environmental effects [29]. These links intertwine urbanization, industrialization, climate, and urban residents' well-being [30]. Climate change, an outcome of this intricate interplay, profoundly affects urban environments, notably air quality. Addressing local air pollution and climate change necessitates joint exploration [31]. It is predicted that there will be an increase in ozone-related fatalities due to climate change induced by greenhouse gases [32]. They are showcasing the intimate connection between air pollution and climate change.

According to [33], the average person inhales approximately 11,000 L of air per day. This air is composed of 78.19% nitrogen, 21.94% oxygen, and 0.032% carbon dioxide, with trace amounts of other gasses [34]. However, when additional elements, known as pollutants, are introduced into the air, the composition can change, leading to air pollution [35]. It is concerning that 95% of the global population breathes this polluted air, as reported by the World Health Organization (WHO). The 2014 data report identifies air pollution as one of the top eight causes of death worldwide. Specifically, in the European Union, it is estimated to be responsible for 400,000 premature deaths [36]. The WHO reports a staggering 6.5 million annual deaths attributed to air pollution [37]. Regrettably, the

COVID-19 pandemic has exacerbated this already critical situation. Studies addressing the connection between air pollution and COVID-19 mortality have emerged, with a particular focus on the valuable contribution of AI-enabled imaging techniques [38]. In addition, investigations were carried out to analyze death data through ML experiments, as well as to assess the impact of the Wuhan lockdown on air quality and health. These assessments employed RF-based weather normalization methods for four pollutants in 30 Chinese cities [39]. Furthermore, research findings have indicated that a significant proportion, over 70%, of SARS-CoV-2 deaths in Italy may be attributed to air pollution and PM<sub>2.5</sub> [40].

The intricate interplay between air quality and health is just one facet of a much broader nexus that encompasses several critical disciplines, as depicted in Figure 3. This complex interconnection serves as a pivotal focal point for addressing contemporary urban challenges. The amalgamation of air quality considerations with urban planning, industrial planning, transportation, green infrastructure, public health, and climate change has garnered significant attention in recent research and policy discussions.



**Figure 3.** Branches of air quality: a graph of interdisciplinary fields (colors and dimensions).

#### 4.1.2. Pollutants and Air Quality Indices

Air pollution can be categorized into indoor and outdoor pollution [41]. Outdoor pollution can stem from natural events like fires and volcanic eruptions, as well as human activities such as manufacturing, burning, and transportation. On the other hand, indoor pollution can be influenced by activities within the home, such as cooking, smoking, and burning fuels, as well as proximity to outdoor pollution sources like highways and industrial areas [42]. It is estimated that indoor air pollution causes 4.3 million deaths, while outdoor air pollution causes 3 million deaths [43]. Studies reveal that the concentration of various pollutants, such as PM<sub>2.5</sub> and CO<sub>2</sub>, is sometimes 2–5 times higher indoors than outdoors, and occasionally even more than 100 times higher, far exceeding the allowable



limit established by the WHO [44]. Being exposed to pollution when staying in homes, offices, educational institutions, etc., can have a substantial negative influence on human health because the average person spends about 80–90% of their time indoors [42].

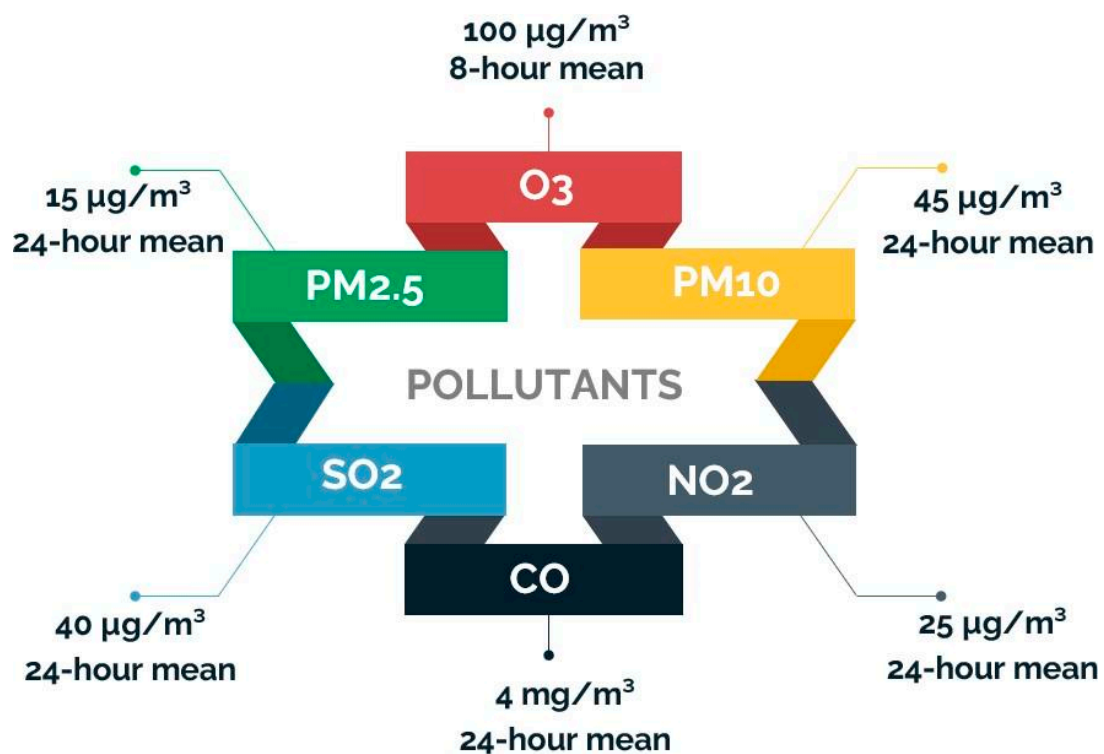
Because of their considerable effects on both air quality and human health, some essential pollutants stand out among the diverse range of contaminants. Among these significant contaminants are:

- PMs, including PM<sub>1</sub>, PM<sub>2.5</sub>, and PM<sub>10</sub>, which refers to particles with a diameter less than 1 µm, 2.5 µm, and 10 µm, respectively, are linked to illnesses and fatalities. Reducing PM<sub>2.5</sub> levels from 35 µg/m<sup>3</sup> to 10 µg/m<sup>3</sup> could potentially decrease air pollution-related deaths by 15% [43,45]. PM<sub>2.5</sub> was the fifth-ranked mortality risk factor in the world and was responsible for 7.6% of all fatalities [46].
- Ozone (O<sub>3</sub>) is produced through photochemical reactions and plays a dual role in greenhouse gas emissions and its impact on human health and the environment. High concentrations of ground-level ozone can be particularly harmful. O<sub>3</sub> exists as a gas both in the upper atmosphere (stratosphere) and at ground level. Stratospheric ozone is beneficial as it acts as a protective shield against ultraviolet rays. However, at the ground level and in the troposphere, ozone becomes a secondary air pollutant. It is formed through a series of intricate photochemical reactions involving solar radiation and ozone precursors [47].
- Nitrogen dioxide (NO<sub>2</sub>) and sulfur dioxide (SO<sub>2</sub>), produced from fuel burning [48], particularly in power plants and vehicles, are associated with respiratory issues and were responsible for 39% of NO<sub>x</sub> emissions in Europe's road transportation industry in 2017 [49]. These gasses are the primary acidic gases released by human activities. They not only contribute to the creation of acid rain and photochemical smog but also have detrimental effects on human health, vegetation, and materials [50].
- Carbon dioxide (CO<sub>2</sub>), produced by burning fossil fuels, respiration, and natural processes, is a greenhouse gas contributing to global warming and pollution concentration, accounting for a significant percentage of emissions [35,51]. CO<sub>2</sub>, as one of the greenhouse gasses (GHGs), plays a significant role in the global warming issue intertwined with industrial development in the globalized world. According to the current literature, the adoption of low-carbon practices is considered the most effective strategy for mitigating global warming. The combustion of fossil fuels by human activities is the primary source of CO<sub>2</sub> emissions, which greatly contributes to the creation of an environment conducive to global warming [52].
- Carbon monoxide (CO) is a hazardous gas emitted from various sources such as incineration, power plants, and urban road traffic. Inhalation of this gas can be fatal, as it converts to CO<sub>2</sub> in the atmosphere. CO poisoning is a prevalent form of toxicity in the modern world and is the leading cause of poisoning-related deaths in the United States. It is a highly toxic gas that lacks taste, odor, and irritants. Detecting CO is challenging due to these properties and the absence of a distinctive clinical signature, often mimicking other common disorders. CO is produced when hydrocarbons undergo incomplete combustion. Sources of CO include poorly ventilated garages with motor vehicle exhaust, as well as areas near garages. Combustion appliances can also generate CO when there is partial combustion of fuels like oils, coal, wood, kerosene, and others. A common scenario involves infrequently used and poorly maintained heating units [53].
- Methane (CH<sub>4</sub>), mainly from natural gas and human activities like landfills and livestock, is another potent greenhouse gas. Methane contributes to the enhanced greenhouse effect. Methane production is a microbiological process, which is predominantly controlled by the absence of oxygen and the amount of easily [54]. CH<sub>4</sub> plays a significant role in intensifying the greenhouse effect, as it is approximately 20 times more potent than CO<sub>2</sub> on a molar basis. It is the second most influential greenhouse gas, following CO<sub>2</sub>, and its overall impact, considering both direct and indirect effects

on tropospheric ozone and stratospheric water vapor, is equivalent to about half of CO<sub>2</sub> [55].

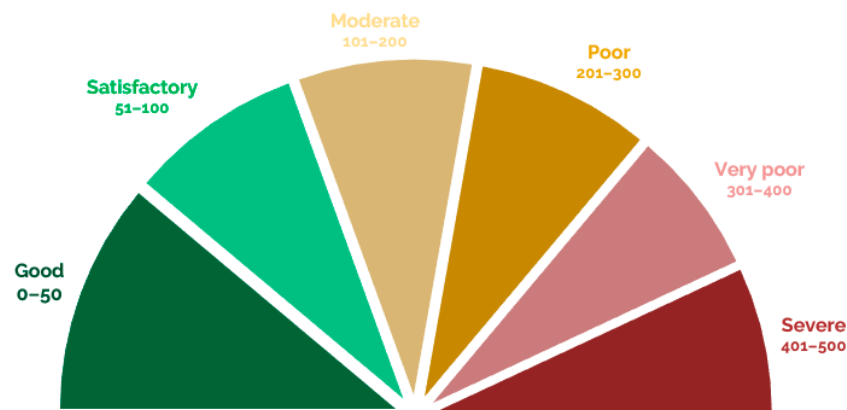
- Volatile organic compounds (VOC), are considered significant contributors to air pollution, affecting the environment through both indirect and direct means. Indirectly, they act as precursors to the formation of ozone and smog. Directly, they pose toxicity risks to the environment. The rise of industrialization and urbanization has resulted in an increase in VOC emissions from various sources, both indoors and outdoors. These sources include the chemical industry, paper manufacturing, food processing, transportation, petroleum refineries, vehicle manufacturing, textiles, electronics, solvents, and cleaning products [56].

The World Health Organization (WHO) has taken significant steps by establishing thresholds and limits for the most significant pollutants (Figure 4).



**Figure 4.** Recommended 2021 air quality guideline (AQG) levels and 2005 AQG [44].

The introduction of Air Quality Indices (AQIs) has simplified the process of assessing air quality. These indices offer a convenient way for individuals to understand the level of air pollution in their vicinity. AQIs are numerical values that correspond to different degrees of pollution, where higher numbers indicate poorer air quality. Although each country or region may have its own specific AQI standard, the Environmental Protection Agency (EPA) standard is widely adopted and encompasses six levels of health concern, ranging from "Good" to "Severe" [41] (Figure 5). The AQI is derived from measurements of five pollutants: SO<sub>2</sub>, CO, O<sub>3</sub>, NO<sub>2</sub>, and PMs [57]. The Indoor Air Quality Index (IAQ) is an important indicator used to assess indoor pollution. It takes into account various biological, chemical, and physical factors that contribute to indoor air quality, including CO<sub>2</sub>, CO, VOC, and O<sub>3</sub>, as well as physical factors such as temperature, humidity, and particulate matter. The IAQ provides a comprehensive measure of the overall quality of the air indoors by considering these different parameters [58]. This index serves as a measure of indoor pollution in enclosed spaces, where people spend approximately 90% of their time [59].



**Figure 5.** Air quality index (AQI).

#### 4.2. Supervised Learning Field

ML is a natural outgrowth of the intersection of Computer Science and Statistics [60]. A subfield of ML, supervised learning. It makes use of labeled datasets to train algorithms that can properly categorize data or predict outcomes. Its beginnings may be seen in the 1950s and 1960s when scientists created linear regression models to examine and forecast correlations between variables. In order to anticipate future results, supervised learning trains a model using input/output pairs or labeled data from the past. To accurately anticipate outputs for novel inputs, a function that can be roughly approximated must be developed. regression when we predict quantitative outputs, and classification when we predict qualitative outputs are two types of difficulties that might arise with supervised learning [61]. In the 1970s and 1980s, as computers became more accessible, researchers started to create increasingly complex supervised learning algorithms. As an illustration, decision trees (DT) were created in the 1970s as a technique to describe complicated decision-making processes [62]. An array of approaches exists, among which is the supervised ML algorithm developed by Vapnik, named SVM. This algorithm is versatile, catering to both classification and regression tasks. Another method involves the embedded technique called RF, which manifests as an assembly of unpruned classification or regression trees. In the context of the RF configuration, each separate tree produces a class prediction, and the ensemble makes its decision based on the class that receives the highest number of votes. ANN has made considerable progress in enhancing supervised learning by enabling the representation of complex non-linear relationships among variables. Various utilization techniques exist for ANN, including Recurrent Neural Network (RNN), LSTM networks, and Gated Recurrent Unit (GRU). These architectures, which fall under the umbrella of ANNs, are specifically tailored to handle continuous or sequential data. The resurgence of interest in supervised learning was ignited by the accessibility of large datasets and improved computational capabilities. Alongside this, DL algorithms were developed and applied across diverse fields by researchers, spanning domains like computer vision, and natural language processing. In supervised learning, a classification model uses known inputs to predict unknown outputs, specifically for categorical outputs. The dataset is divided into classes, and a classification algorithm learns from a training dataset to assign new data points to specific classes. Through a mapping function derived from the training data, the classification model can predict the class label for test data. The classification process involves data collection, preprocessing to eliminate noise and duplicates, and splitting the data into training and test sets using cross-validation. Once trained, the model can predict the class or label for new datasets, and its performance is evaluated using the test data. Classification can be binary, with two possible outcomes, or multi-label, dealing with multiple classes. In contrast, regression in supervised learning focuses on predicting continuous values based on variables, identifying correlations, and making predictions for continuous outputs. Regression can be categorized as simple linear regression, establishing

a relationship between two variables with a straight line, or multiple regression, involving multiple variables and further classified as linear or non-linear regression [63].

#### 4.3. Supervised Learning Approaches for Air Quality Analysis

The use of supervised ML in air quality analysis is widespread due to its effectiveness. There are two main approaches in this field: classification and regression models. Classification models categorize air quality based on pollutant readings, while regression models predict continuous values such as pollutant levels based on factors like time, location, and weather.

##### 4.3.1. PM and Beyond: Exploring Pollutant Prediction in Air Quality Analysis

Particulate matter (PM) stands as a significant concern in the realm of air pollution, and numerous studies are devoted to employing ML models for PM-level forecasting. In one instance, high-resolution spatial air pollution maps of Charlotte, North Carolina were crafted utilizing linear regression and ML techniques [64]. In Ankara, Turkey, Bozdağ et al. introduced a hybrid model aimed at forecasting PM<sub>10</sub> concentrations. Multiple models, such as K-Nearest Neighbor (KNN), eXtreme Gradient Boosting (xGBoost), and Artificial Neural Networks (ANN), were employed in parallel to forecast PM<sub>10</sub> concentrations [65]. In another study, a variety of techniques such as Adaptive Neuro-Fuzzy Inference System (ANFIS), Support Vector Regression (SVR), Classification and Regression Trees (CART), and Extreme Learning Machine (ELM) were employed to forecast PM<sub>10</sub> and PM<sub>2.5</sub> levels in an industrial area. The results of this investigation indicated that ANFIS proved to be the most effective model [37]. Furthermore, a study that compared four algorithms for detecting peak PM<sub>2.5</sub> values found that Random Forest (RF) emerged as the most effective choice [46]. Expanding beyond PM, several studies have directed their focus toward other pollutants. One such study employed 16 alternative methods to predict annual average NO<sub>2</sub> concentrations and discovered that SVR and ANN exhibited lower accuracy in comparison to Generalized Boosted Machine (GBM), RF, and bagging [66]. Additionally, a dynamic indoor CO<sub>2</sub> model was developed using ML techniques like Support Vector Machines (SVM), Adaptive Boosting (AdaBoost), RF, GBM, Logistic Regression (LR), and Multilayer Perceptron (MLP) to manage the campus classroom's HVAC system based on CO<sub>2</sub> levels [67]. Finally, a novel non-linear air quality regression model, GaussODM, was introduced, drawing upon the Gaussian dispersion model. This model demonstrated superior performance when compared to both a simple regression model and a benchmark interpolation method for estimating geographic ambient NO<sub>x</sub> concentrations [68].

##### 4.3.2. Regression Techniques for Air Pollution Prediction

Regression techniques have proven to be highly effective in the domain of air pollution prediction. For example, indoor PM<sub>2.5</sub> concentrations were predicted using both Multiple Linear Regression (MLR) and Random Forest Regression (RFR) [69]. In Colombia's Aburrá Valley, a method for predicting pollutant concentrations utilized temporal features as input variables for an advanced Artificial Neural Network (ANN) and Support Vector Regression with Particle Swarm Optimization (SVR-PSO), achieving the best results [70]. More precise spatiotemporal Land Use Regression (LUR) models for pollutants like PM<sub>2.5</sub>, PM<sub>10</sub>, O<sub>3</sub>, NO<sub>2</sub>, CO, and SO<sub>2</sub> were created using mixed effect models and Least Absolute Shrinkage and Selection Operator (LASSO). These models outperformed previous LUR models, particularly at time scales of a day or longer [71]. ANN models have gained popularity in the prediction of air pollution and estimation of hospital admissions due to exposure to pollution. In a study focused on two major Brazilian cities, five ANNs were tested to estimate hospital admissions caused by PM<sub>10</sub> and meteorological factors [72]. Another study demonstrated improved one-step-ahead prediction results when combining ANNs and Multilayer Perceptron (MLP) models for PM<sub>10</sub> and PM<sub>2.5</sub> prediction [73]. Beyond ANN models, various other ML models, including Least-Squares Support Vector Machines (L-SVM), Gaussian Process Regression (GPR), RFR, and the PROPHET time

series model, were evaluated in Bangladesh for monitoring PM and air quality. Among these models, GPR performed most favorably according to metrics such as R2, RMSE, and MAE [74]. Neural network models also showed lower error rates compared to linear regression models in forecasting CO concentrations [75].

#### 4.3.3. Enhancing Air Quality Classification Methods

This study explores the advanced classification techniques for air quality analysis, addressing class imbalance and improving precision in assessing air quality. The introduction of the Adjusting Kernel Scaling (AKS) method, alongside classification algorithms like Adaboost, Multi-Layer Perceptron, GaussianNB, and SVM, demonstrates remarkable performance with an accuracy rate of 99.66% when applied to data from the Indian Central Pollution Control Board (CPCB), surpassing other classification methods [76]. Another study presents a laboratory study focusing on challenging multi-class classification problems related to indoor air contaminants. The study evaluates the proposed hybrid SVM (HSVM) model. It compares it with five existing methods, such as Euclidean distance to centroids (EDC), simplified fuzzy ARTMAP network (SFAM), multilayer perceptron neural network (MLP) based on back-propagation learning, individual FLDA, and single SVM, showcasing the HSVM model's superior performance in addressing discrimination challenges in various electronic nose applications [77]. PCA was used to identify vehicular emissions and fuel combustion as significant pollution sources. Ensemble models, including Single Decision Tree (SDT), Decision Tree Forest (DTF), and Decision Tree Boost (DTB), are constructed and compared with Support Vector Machines (SVM). These models effectively discriminate seasonal air quality and predict air quality indices. The DT models show better performance in both classification and regression compared to SVM, credited to their use of bagging and boosting algorithms [78].

The authors of another study highlighted the use of machine learning algorithms, including Naïve Bayes, Random Forest, and K-Nearest Neighbor, for predicting PM<sub>10</sub> hotspots, a major air pollutant in Malaysia. These models offer effective tools for spatial PM<sub>10</sub> assessment, particularly in urbanized and industrialized areas with high PM<sub>10</sub> concentrations. The RF model's output highlights high PM<sub>10</sub> concentrations in urbanized and industrialized areas, emphasizing the detrimental impact of air pollutants in urban regions. These models have the potential to support the Sustainable Development Goal (SDG) for Sustainable Cities and Communities by facilitating spatial PM<sub>10</sub> assessment and management [79]. Regarding the relationship between air pollution and COVID-19 infections, a Reduced-Space Gaussian Process Regression was employed to develop a classification model. This analysis reveals a correlation between high COVID-19 infection areas and elevated levels of NO<sub>2</sub> and PM<sub>10</sub>, emphasizing the role of industrial factors and environmental conditions in the context of the pandemic [80].

#### 4.3.4. Deep Learning's Role in Reliable Air Pollution Forecasting

The significance of deep learning in the domain of air quality prediction has evolved into a transformative force, particularly within the specialized field of air pollution forecasting. The distinguishing hallmark of deep learning lies in its profound ability to uncover intricate patterns and interrelationships concealed within vast and complex datasets. This remarkable attribute has empowered deep learning (DL) models with an extraordinary capacity to not only predict air quality levels with precision but also to transcend conventional prediction methods. DL models bring a novel dimension to the table by extending their utility beyond mere accuracy; they contribute to the augmentation of forecast reliability. This is achieved through a unique capability—the assessment of uncertainty associated with each prediction, a feature that proves invaluable in air pollution forecasting.

To illustrate the practical application of DL in air quality prediction, consider the work of [81], which employed Long Short-Term Memory (LSTM), a modified RNN, to monitor PM<sub>2.5</sub> levels at various time stamps. Additionally, another study introduced the Convolutional Bidirectional Gated Recurrent Unit (CBGRU), a DL model tailored for short-



term PM<sub>2.5</sub> forecasting [82]. Furthermore, the study [83] harnessed a Bayesian LSTM DL model to assess the impact of air quality regulations in Beijing, China. These instances underscore the versatility and effectiveness of DL models in addressing intricate challenges related to air quality prediction, with potential implications for mitigating the impact of air pollution on public health and the environment.

#### 4.3.5. Enhancing Air Pollution Forecasting with Hybrid Models

The integration of hybrid models into air quality forecasting represents a promising approach, offering the potential to significantly enhance prediction accuracy and robustness in this critical field. These hybrid models ingeniously combine distinct methodologies, skillfully capitalizing on the complementary strengths of each component. A noteworthy illustration of this approach is demonstrated in [84], where researchers introduced the W-ANN hybrid model, skillfully merging the Wavelet transform method with traditional ANN models. The study's findings revealed that the W-ANN model surpassed traditional ANN models in terms of forecasting accuracy, underlining the effectiveness of hybridization. Another compelling example of a hybrid model applied in the domain of air pollution forecasting is the ICEEMDAN-BPNN-ICA model, as presented by the authors in [85]. This multifaceted model seamlessly integrates three techniques: Intrinsic Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN), serving as a preprocessing step for data decomposition into intrinsic mode functions; Back Propagation Neural Network (BPNN), a versatile feedforward neural network employed to predict future values of each function; and Independent Component Analysis (ICA), a sophisticated signal processing technique adept at extracting independent components from the output of the BPNN. This comprehensive model was successfully employed to predict concentrations of various pollutants, including PM<sub>2.5</sub>, SO<sub>2</sub>, NO<sub>2</sub>, CO, and O<sub>3</sub>, thereby showcasing the immense potential of hybrid models in advancing air quality forecasting methodologies.

The findings of this section are listed in Table 1. This table provides a comprehensive summary of the results obtained, emphasizing the key ML techniques utilized, and parameters predicted. Two graphical representations, namely Figures 6 and 7, are provided in the study. They display the predicted shares of pollutants as outlined and the proportion of methods employed in the review, respectively.

**Table 1.** Supervised machine learning for air quality monitoring: a table overview.

	Source	ML Method	Predicted Value
01	Ravindra et al. [86]	RF, K-NN, LASSO, Decision Tree(DT), SVR Xgboost, DNN	Hospital admissions related to Acute Respiratory Infections
02	Dutta and Pal [87]	stacked-bidirectional long short-term memory (stacked-BDLSTM)	PM <sub>2.5</sub> , PM <sub>10</sub>
03	Van et al. [88]	DT, RF, XGBoost	AQI
04	Eren et al. [89]	LSTM, RNN, GRU	PM <sub>2.5</sub>
05	Barthwal [90]	Markov chain (DTMC) models	AQI
06	Wang et al. [91]	EML	Ozone
07	Persis and Amar [92]	NNs, SVM, DT, RF, XGboost.	AQI
08	Koo et al. [93]	DNN, RNN, Convolutional Neural Network (CNN),	PM <sub>2.5</sub>
09	Natsagdorj et al. [94]	Bayesian optimized LSTM, CNN-LSTM	PM <sub>2.5</sub>
10	Falah et al. [95]	RF, XGboost	PM <sub>2.5</sub>
11	Xie et al. [96]	Deep Learning-based Complex Trait Estimation Model(DL-CTEM)	NH <sub>3</sub> , CO <sub>2</sub> , H <sub>2</sub> S
12	Muthukumar et al. [97]	Convolutional Long Short-Term Memory (ConvLSTM), Graph Convolutional Network (GCN)	PM <sub>2.5</sub>
13	Abu El-Magd et al. [98]	RF	PM <sub>10</sub>

Table 1. Cont.

	Source	ML Method	Predicted Value
14	Huang et al. [99]	LR, RF, SVM, GPR, NN, ensemble tree	PM <sub>2.5</sub> , PM <sub>10</sub>
15	Gilik et al. [100]	CNN, LSTM	PM, NO <sub>x</sub> , SO <sub>2</sub>
16	Kumar and Pande [101]	Gaussian naive bayes (GNB), SVM XGBoost	AQI
17	Sethi and Mittal [102]	weighted naive bayes(WNB)	AQI
18	Abirami et al. [103]	SVR, Decision Tree Regression (DTR) RFR, MLR	AQI
19	Chen et al. [104]	DNN, LSTM	Ozone
20	Cheng et al. [105]	ResNet-LSTM	PM <sub>2.5</sub>
21	Cho and Moon [106]	ANN	CO <sub>2</sub> , PM <sub>10</sub> , PM <sub>2.5</sub>
22	Geetha et al. [107]	LSTM, RNN	SO <sub>2</sub> , CO <sub>2</sub> , NO <sub>2</sub> , CO, CFCs
23	Liu et al. [108]	SVM	AQI
24	Martín-Baos et al. [109]	LR, GPR, RF	AQI
25	Asha et al. [110]	Edited Nearest Neighbor (ENN)	NH <sub>3</sub> , CO, NO <sub>2</sub> , CH <sub>4</sub> , CO <sub>2</sub> , PM <sub>2.5</sub>
26	Ferreira et al. [111]	fuzzy ARTMAP	PMs
27	Choudhury et al. [112]	KNN, SVR, Hidden Markov Model(HMM) CNN,LSTM	NO <sub>2</sub> , O <sub>3</sub>
28	Qader et al. [113]	NNs, GPR	CO <sub>2</sub>
29	Magazzino et al. [38]	NNs, DT	Deaths
30	Mokhtari et al. [81]	CNN, LSTM	Propylene
31	De Mattos Neto et al. [73]	ANN, MLP	PM <sub>10</sub> , PM <sub>2.5</sub>
32	Wei et al. [114]	EMD-FUSION	SO <sub>2</sub>
33	Cihan et al. [37]	ANFIS, SVR, CART, RF, KNN, ELM	PM <sub>10</sub> , PM <sub>2.5</sub>
34	Taheri and Razban [67]	SVM, AdaBoost, RF, GBM, LR, MLP	CO <sub>2</sub>
35	K et al. [115]	LR	PM <sub>2.5</sub>
36	Cole et al. [39]	RF	PM <sub>2.5</sub>
37	Shahriar et al. [74]	L-SVM, GPR, RFR	PM <sub>2.5</sub>
38	Chang et al. [116]	Gradient Boosting Trees (GBT), SVR LSTM, LSTM2	PM <sub>2.5</sub>
39	Bozdag et al. [65]	LASSO, SVR, RF, kNN, xGBoost, ANN	PM <sub>10</sub>
40	Chang et al. [117]	LSTM, SVR, GBT	PM <sub>2.5</sub>
41	Magazzino et al. [118]	ANNs	Deaths
42	AlOmar et al. [84]	W-ANN	Ozone
43	Cazzolla Gatti et al. [40]	RF	Deaths
44	Han et al. [83]	LSTM	PM <sub>2.5</sub>
45	Shams et al. [75]	MLR, ANN	CO
46	Zeinalnezhad et al. [119]	ANFIS	SO <sub>2</sub> , O <sub>3</sub> , NO <sub>2</sub> , CO
47	Alyousifi et al. [120]	Multi-Wave Fuzzy Time Series (MWFTS)	API
48	Lu et al. [121]	Density-Based Spatial Clustering of Applications with Noise (DBSCAN), DNN	PM <sub>2.5</sub>
49	Tao et al. [82]	CBGRU	PM <sub>2.5</sub>
50	Chen et al. [66]	16 methods	NO <sub>2</sub>
51	Araujo et al. [72]	ELM, MLR, Radial Basis Function(RBF) Echo State Network(ESN),ENN	Hospitalizations

Table 1. Cont.

	Source	ML Method	Predicted Value
52	Murillo-Escobar et al. [70]	SVR-PSO	NO, NO <sub>2</sub> , O <sub>3</sub> , PM <sub>10</sub> , PM <sub>2.5</sub>
53	Zhou et al. [122]	GPM	NO <sub>2</sub> , HC
54	Yadav et al. [123]	CTSPD Algorithm	CO, Ozone, NO <sub>2</sub> , PM <sub>2.5</sub> , PM <sub>10</sub>
55	Jiang et al. [85]	ICEEMDAN-BPNN-ICA	PM <sub>2.5</sub> , SO <sub>2</sub> , NO <sub>2</sub> , CO, O <sub>3</sub>
56	Yuchi et al. [69]	MLR, RFR	PM <sub>2.5</sub>
57	Son et al. [71]	LASSO	PM <sub>2.5</sub> , PM <sub>10</sub> , O <sub>3</sub> , NO <sub>2</sub> , CO, SO <sub>2</sub>
58	Ketu et al. [76]	Adjusting Kernel Scaling (AKS)Adaboost, Multi-Layer Perceptron, GaussianNB, and SVM	AQI
59	Zhang, Lei, et al. [77]	hybrid SVM (HSVM)Euclidean distance to centroids (EDC), simplified fuzzy ARTMAP network (SFAM), multilayer perceptron neural network (MLP), individual FLDA, and single SVM	SO <sub>2</sub> , NO <sub>2</sub> , CO, CO <sub>2</sub> , NH <sub>3</sub> , O <sub>3</sub> , formaldehyde, benzene, toluene, inhalable particle, and VOCs
60	Singh et al. [78]	PCA, Single Decision Tree (SDT), Decision Tree Forest (DTF), Decision Tree Boost (DTB)SVM	AQI
61	Tella, Abdulwaheed, et al. [79]	Naïve Bayes, Random Forest, and K-Nearest Neighbor	PM <sub>10</sub>
62	Velásquez et al. [80]	Reduced-Space Gaussian Process Regression	NO <sub>2</sub> , PM <sub>10</sub>

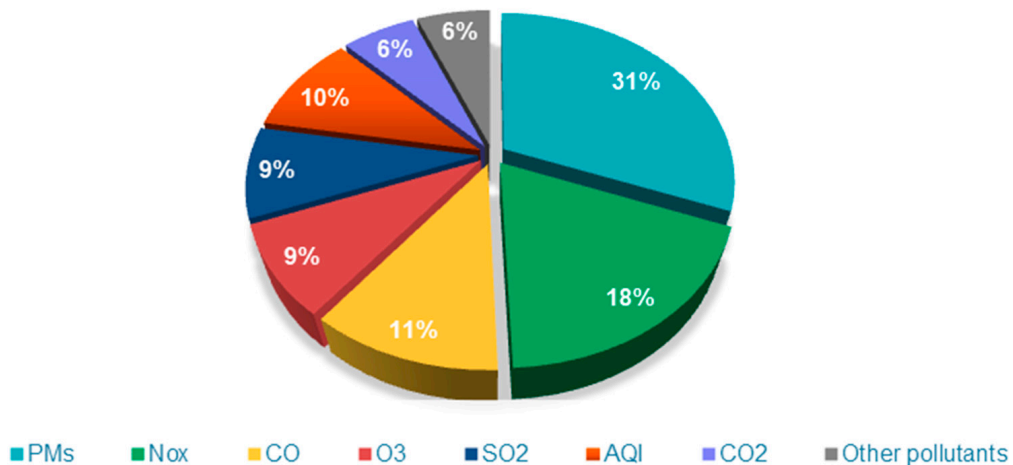


Figure 6. Predicted pollutant shares in the survey.

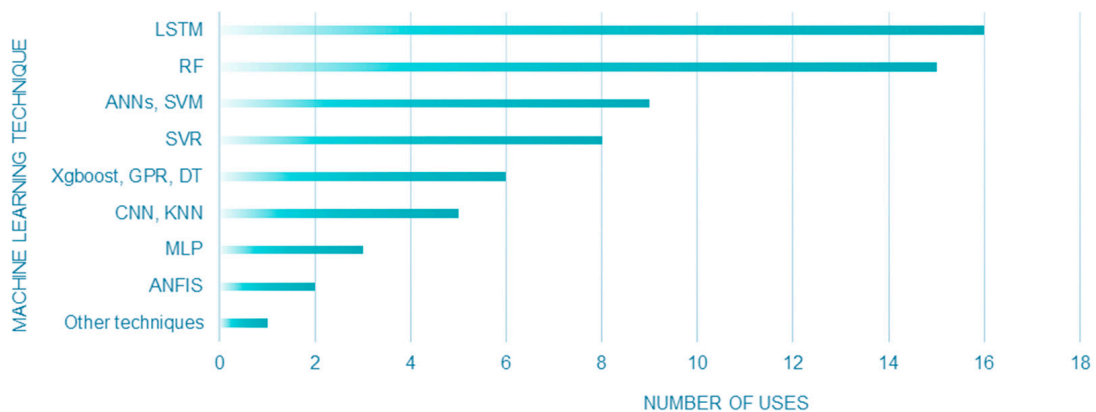


Figure 7. Proportion of methods employed in the review.

## 5. Challenges and Future Directions

### 5.1. Findings, Limitations, and Challenges in Air Quality Research

The analysis of the existing research has highlighted the significance of exploring the link between COVID-19 fatalities, economic expansion, and air pollution. Several studies indicate that air pollution, particularly PM<sub>10</sub> and PM<sub>2.5</sub>, may play a significant role in COVID-19-related deaths. The application of ML techniques was identified as an effective approach to forecasting air pollution and minimizing its negative impact on public health. However, it is worth noting that the studies included in the literature review were conducted in specific regions, and their conclusions may not apply to other locations. Further studies are necessary to gain a thorough understanding of the relationship between COVID-19 mortality, economic growth, and air pollution across various regions. The literature review also found that PM, NO<sub>x</sub>, SO<sub>2</sub>, O<sub>3</sub>, CO, and CO<sub>2</sub> are frequently identified as pollutants of concern in air quality assessments. It indicates that these pollutants are causing significant apprehension and are expected to attract more attention and stricter regulations. This knowledge holds great significance for policymakers and researchers as it helps them direct their endeavors toward decreasing the concentrations of these pollutants and enhancing air quality on the whole. This review reveals also that in the domain of air quality research, certain algorithmic approaches have emerged as particularly prominent and widely employed. Among these, algorithms such as LSTM, RF, ANN, SVR, XGBoost, DT, Convolutional Neural Networks (CNN), and Gaussian Process Regression (GPR) stand out.

Temperature and humidity also play a vital role in air quality research. They influence how pollutants behave and their concentrations in the air. Factors like chemical reactions, dispersal patterns, the respiratory system of humans, and the functionality of air quality monitoring devices are all affected by temperature and humidity levels. It is important to consider that the choice of pollutants studied and predicted depends on the location and specific sources of pollution under investigation. For example, in heavily industrialized areas, studies may focus on predicting levels of PM and NO<sub>x</sub>, while in rural regions, attention may be directed toward predicting levels of ozone and carbon monoxide. Therefore, it is crucial to account for the local context when designing and implementing IoT-based air quality monitoring systems to effectively address and mitigate pollution concerns.

A key challenge revolves around the necessity for more robust and reliable sensors. IoT sensor networks heavily rely on numerous low-cost devices, which can be susceptible to failures and inaccuracies. Consequently, this can lead to data gaps and inaccurate measurements, ultimately affecting the effectiveness and reliability of ML models. To overcome this obstacle, researchers should prioritize the development of more resilient and dependable sensors, alongside the formulation of approaches to handle missing data and sensor errors. Another significant challenge pertains to the concerns surrounding privacy issues and security [124]. IoT sensor networks collect and transmit vast amounts of sensitive data, making them vulnerable to hacking and various cyber-attacks. These breaches can lead to the compromise of sensitive information and even manipulation of data, ultimately impeding the effectiveness of ML models. To address this issue, researchers should concentrate on developing secure and privacy-preserving techniques for data collection, storage, and transmission. Researchers commonly rely on classification and regression methods when analyzing air quality data. This preference can be attributed to the fact that these approaches yield results that are more easily understood, in comparison to clustering and dimensionality reduction techniques. The ability to comprehend the contribution of various factors to air quality outcomes, as well as the impact of interventions, holds paramount importance in this field of study. One potential critique concerning the utilization of supervised ML techniques in air quality research is their potential inability to capture the intricacies and variability present in the data, potentially leading to inaccurate predictions. Moreover, these supervised techniques heavily rely on labeled data, which can be challenging to acquire, and they often require significant computational resources. Ensemble methods, including RF, XGBoost, AdaBoost, Gradient Boosting, GBT, and ensemble trees,

have proven to be effective in air quality prediction. However, their effectiveness varies depending on the specific dataset and problem at hand. On the other hand, techniques such as decision tree-based methods (DT, CART) and Naive Bayes (Gaussian Naive Bayes) have shown less robust performance. DL techniques, such as Deep Learning-CTEM, ConvLSTM, GCN, and CNN+LSTM, were also proposed. Nonetheless, it is worth noting that these approaches can be computationally expensive and may not be practical for real-time predictions. Furthermore, these advanced techniques, such as neural networks (NN, MLP, NNs, LSTM, ELM) and hybrid models (EMD-FUSION, ANFIS), exhibit commendable performance in air quality analysis. However, their implementation and interpretation can pose challenges, alongside the need for larger datasets and computational resources. It is important to acknowledge that these approaches may lack interpretability and offer limited insights into the underlying mechanisms that drive air quality.

### 5.2. The Role of ML Models in Mitigating Climate Change and Air Pollution: A Sustainable Development

According to the *Global Sustainable Development Report 2023*, addressing air quality is a critical component of the Sustainable Development Goals (SDGs), particularly those related to environmental conservation and sustainability. The report emphasizes that efforts to protect and restore terrestrial ecosystems, a cornerstone of SDGs, are intrinsically linked to air quality improvements. Preserving air quality not only supports goals related to life on land, such as SDG 15, which focuses on halting biodiversity loss, but it is also essential for achieving SDG 3, which centers on good health and well-being. In the context of data and ML, the *Global Sustainable Development Report* underscores the importance of data-driven strategies and innovative technologies in enhancing air quality control. The report advocates for leveraging ML to advance data analysis and predictive modeling for more effective air quality monitoring and mitigation efforts. By embracing ML and data technologies, as highlighted in the report, it is possible to make well-informed decisions and implement policies that effectively address air pollution, mitigating its adverse effects on human health and the environment. As we progress through the second half of Agenda 2030, these novel partnerships and data technologies are integral in ensuring clean air for all, aligning with the SDGs, and enhancing the quality of life and the environment, as emphasized in the *Global Sustainable Development Report 2023* [23].

Artificial intelligence and machine learning technologies play a pivotal role in addressing the challenges posed by climate change. Both climate change and the field of artificial intelligence are intricate subjects of academic exploration. The environmental consequences of ML, both in a positive and negative sense, are currently a subject of in-depth analysis. The United Nations has established a set of 17 Sustainable Development Goals (SDGs) that encompass environmental, social, and economic aspects. AI has the potential to reduce global emissions of greenhouse gasses by up to 4%, thereby assisting in alleviating the impact of climate change. AI methodologies have the capability to enhance the accuracy of forecasting tools utilized for predicting and evaluating extreme environmental occurrences and for analyzing long-term climate change data. Consequently, AI can serve as a crucial tool in the battle against climate change [125].

The reported potential effects of AI encompass both positive and negative ramifications on sustainable development. Nevertheless, as of now, there is no published research that systematically examines the extent to which AI could influence all facets of sustainable development, as defined in this investigation, encompassing the 17 Sustainable Development Goals (SDGs) and the 169 internationally agreed targets outlined in the 2030 Agenda for Sustainable Development. A substantial body of relevant evidence indicates that AI might serve as a facilitator for achieving 134 targets (approximately 79%) spanning all SDGs, primarily using technological enhancements that have the potential to overcome existing limitations. However, the development of AI may have an adverse impact on 59 targets (about 35%), also distributed across all SDGs [126].



Looking ahead, while ML offers substantial potential to mitigate pollution and carbon emissions through enhanced optimization of industrial processes and transportation systems, their implementation must be approached sustainably to prevent inadvertent increases in carbon footprints. ML, in particular, can have a negative impact on the environment through energy consumption in the training and use of large models, the operation of data centers, and the manufacturing of specialized hardware such as GPUs. These activities contribute to carbon emissions and other types of pollution [127]. The findings of various studies demonstrate that AI has a negative impact on carbon intensity. For example, a study in China presents the first evidence of a relationship between AI and carbon intensity, supporting the notion that AI has a positive effect on carbon emissions [128]. Another study reveals that CO<sub>2</sub> emissions increase with the number of Edge-AI G-IoT devices deployed and with the greenness of a country's energy production [129]. The use of electronic devices and the infrastructure required to support them can contribute to more than CO<sub>2</sub> emissions as mentioned [130]. Data centers and high-performance computing facilities are estimated to contribute 100 megatonnes of CO<sub>2</sub> emissions per year, similar to the amount emitted by commercial aviation in the US [130]. This significantly contributes to climate change. Furthermore, the McKinsey Global Institute estimates that AI could potentially increase global GDP growth by 1.24% per year by 2030. However, there is a lack of consensus in the literature on the relationship between ICT and CO<sub>2</sub> emissions, with a viewpoint falling into three categories: ICT can promote economic development in an environmentally friendly manner, ICT is an energy consumer, and the relationship between ICT and CO<sub>2</sub> emissions is uncertain [128]. Studies have shown that energy savings in data centers can range from 154% to 604% depending on the specific scenario, reducing CO<sub>2</sub> emissions [131]. Therefore, it is crucial to evaluate the environmental consequences of integrating AI and IoT and take steps to mitigate their negative influence. The energy consumption, materials, electronic waste, and transportation associated with IoT contribute significantly to pollution. To alleviate the environmental impact of AI, it is essential to emphasize renewable energy, sustainable design, recycling of electronic waste, and energy efficiency. By adopting these strategies, we can reduce the ecological impact of these technologies while retaining their advantages.

### *5.3. Future Directions and Open Perspectives in Urban Planning for Air Quality Research*

The future of air quality research holds great promise, with potential advancements in the field of machine learning (ML) and data-driven approaches. Beyond the current capabilities of supervised and unsupervised methods, one notable avenue for exploration is reinforcement learning, a concept receiving significant attention in various domains. This approach, known for its capacity to learn through interaction with the environment and optimize actions to maximize rewards, has the potential to revolutionize air quality research. By training agents to make real-time decisions based on air quality data and feedback loops, reinforcement learning can be employed to dynamically optimize interventions and policies, particularly in complex urban environments with variable pollution sources.

Semi-supervised learning offers another avenue for future research. Combining the strengths of labeled and unlabeled data, it is particularly well-suited for scenarios where acquiring large amounts of labeled data is challenging and cost-prohibitive. In the context of air quality research, where obtaining labeled data for various pollutants across diverse locations can be a formidable task, semi-supervised learning presents a viable solution. By effectively harnessing both labeled monitoring station data and unlabeled information from satellite imagery or sensor networks, semi-supervised learning models can potentially offer more precise and comprehensive insights into air quality patterns and the underlying drivers of pollution. These future research directions are poised to enhance our understanding and management of air quality, contributing to the well-being of both the environment and human health.

Regarding optimizing urban layouts, the potential of AI and data analytics to reshape urban landscapes is truly transformative. Urban planners now have the tools to conduct

nuanced analyses aimed at reconfiguring the spatial arrangement of urban features. This includes a strategic reevaluation of the placement of residential zones, industrial areas, and traffic arteries. The overarching objective is to tactically reduce the proximity of sensitive receptors, such as residential communities, to pollution sources, thereby fostering cleaner urban environments. The impact of AI-driven algorithms extends to the relocation of industrial facilities away from densely populated areas, a move that not only mitigates harmful emissions exposure but also contributes to the enhancement of air quality. This facet underscores the profound influence of AI on urban planning, with air quality improvement at its core.

The integration of AI into the domain of smart transportation systems presents an auspicious path toward revolutionizing urban mobility and, consequently, elevating air quality. AI-infused traffic management solutions hold the potential to alleviate congestion, a well-documented contributor to air pollution in urban settings. Furthermore, the development of electric and autonomous vehicles, under the guidance of AI technologies, ushers in an era of reduced emissions, constituting a significant boon for air quality. Implementing predictive modeling further accentuates the positive impact of AI by facilitating traffic flow optimization, ultimately mitigating idling time and the pollution associated with congestion. As urban areas continue to grapple with the multifaceted challenge of air quality, AI-driven solutions in transportation systems stand out as a promising avenue for meaningful change, ensuring cleaner and healthier urban environments for residents.

## 6. Conclusions

In conclusion, this comprehensive review provides a nuanced analysis of the current landscape about the utilization of supervised machine learning in the fields of air quality management and climate change mitigation, with a keen focus on its alignment with the Sustainable Development Goals (SDGs). Through a meticulous examination of a variety of scholarly works, it becomes evident that these technological advancements hold great promise in elevating the precision and efficiency of air pollution monitoring and predictive capabilities. By harnessing advanced algorithms in conjunction with sensor networks, supervised learning has distinctly showcased its potential to uncover critical insights into the sources and intricacies of air pollution. The models generated through data-driven methodologies offer a heightened degree of accuracy, thereby facilitating timely interventions aimed at ameliorating environmental and public health risks. It is imperative to acknowledge, however, that certain challenges persist, most notably those about data quality and the interpretability of models. Despite these challenges, the integration of supervised learning undeniably represents a proactive and forward-looking approach to environmental stewardship. The confluence of cutting-edge technology and environmental imperatives has led to a more informed and efficient approach to the management of air quality.

In summary, the intersection of pioneering technology and environmental priorities has paved the way for a more enlightened and effective approach to air quality management and climate change mitigation. This journey towards a cleaner and healthier future remains reliant on sustained interdisciplinary collaboration and the continuous refinement of methodological approaches. The path forward, in alignment with the principles of the SDGs, holds the promise of a sustainable and improved future.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/su16030976/s1>, Table S1: The PRISMA 2020 checklist [132].

**Author Contributions:** All the authors have considerably contributed to this paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We would like to express our gratitude to Omar Bouattane, Hassan Ouajji, and Mohamed Youssfi for their invaluable support in this research.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. United Nations. *World Population Is Projected to Reach 9.8 Billion in 2050, and 11.2 Billion in 2100*; United Nations: San Francisco, CA, USA, 2022. Available online: <https://www.un.org/en/desa/world-population-projected-reach-98-billion-2050-and-112-billion-2100> (accessed on 6 October 2023).
2. Johnson, T.M.; Guttikunda, S.; Wells, G.J.; Artaxo, P.; Bond, T.C.; Russell, A.G.; Watson, J.G.; West, J. *Tools for Improving Air Quality Management: A Review of Top-Down Source Apportionment Techniques and Their Application in Developing Countries*; World Bank: Washington, DC, USA, 2011.
3. Kaginalkar, A.; Kumar, S.; Gargava, P.; Niyogi, D. Review of urban computing in air quality management as smart city service: An integrated IoT, AI, and cloud technology perspective. *Urban Clim.* **2021**, *39*, 100972. [[CrossRef](#)]
4. Karroum, K.; Lin, Y.; Chiang, Y.-Y.; Ben Maissa, Y.; El Haziti, M.; Sokolov, A.; Delbarre, H. A Review of Air Quality Modeling. *MAPAN* **2020**, *35*, 287–300. [[CrossRef](#)]
5. Méndez, M.; Merayo, M.G.; Núñez, M. Machine learning algorithms to forecast air quality: A survey. *Artif. Intell. Rev.* **2023**, *56*, 10031–10066. [[CrossRef](#)] [[PubMed](#)]
6. Patil, R.M.; Dinde, H.T.; Powar, S.K. A Literature Review on Prediction of Air Quality Index and Forecasting Ambient Air Pollutants using Machine Learning Algorithms. *Int. J. Innov. Sci. Res. Technol.* **2020**, *5*, 1148–1152. [[CrossRef](#)]
7. Masood, A.; Ahmad, K. A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance. *J. Clean. Prod.* **2021**, *322*, 129072. [[CrossRef](#)]
8. Liu, H.; Yin, S.; Chen, C.; Duan, Z. Data multi-scale decomposition strategies for air pollution forecasting: A comprehensive review. *J. Clean. Prod.* **2020**, *277*, 124023. [[CrossRef](#)]
9. Bellinger, C.; Mohamed Jabbar, M.S.; Zaïane, O.; Osornio-Vargas, A. A systematic review of data mining and machine learning for air pollution epidemiology. *BMC Public Health* **2017**, *17*, 907. [[CrossRef](#)]
10. Zhang, B.; Rong, Y.; Yong, R.; Qin, D.; Li, M.; Zou, G.; Pan, J. Deep learning for air pollutant concentration prediction: A review. *Atmos. Environ.* **2022**, *290*, 119347. [[CrossRef](#)]
11. Zaini, N.; Ean, L.W.; Ahmed, A.N.; Malek, M.A. A systematic literature review of deep learning neural network for time series air quality forecasting. *Environ. Sci. Pollut. Res.* **2022**, *29*, 4958–4990. [[CrossRef](#)]
12. Zhang, W.; Wu, Y.; Calautit, J.K. A review on occupancy prediction through machine learning for enhancing energy efficiency, air quality and thermal comfort in the built environment. *Renew. Sustain. Energy Rev.* **2022**, *167*, 112704. [[CrossRef](#)]
13. Jia, J.-J.; Zhu, M.; Wei, C. Household cooking in the context of carbon neutrality: A machine-learning-based review. *Renew. Sustain. Energy Rev.* **2022**, *168*, 112856. [[CrossRef](#)]
14. Tien, P.W.; Wei, S.; Darkwa, J.; Wood, C.; Calautit, J.K. Machine Learning and Deep Learning Methods for Enhancing Building Energy Efficiency and Indoor Environmental Quality—A Review. *Energy AI* **2022**, *10*, 100198. [[CrossRef](#)]
15. Ma, N.; Aviv, D.; Guo, H.; Braham, W.W. Measuring the right factors: A review of variables and models for thermal comfort and indoor air quality. *Renew. Sustain. Energy Rev.* **2021**, *135*, 110436. [[CrossRef](#)]
16. Ben Atitallah, S.; Driss, M.; Boulila, W.; Ben Ghézala, H. Leveraging Deep Learning and IoT big data analytics to support the smart cities development: Review and future directions. *Comput. Sci. Rev.* **2020**, *38*, 100303. [[CrossRef](#)]
17. Li, F.; Yigitcanlar, T.; Nepal, M.; Nguyen, K.; Dur, F. Machine Learning and Remote Sensing Integration for Leveraging Urban Sustainability: A Review and Framework. *Sustain. Cities Soc.* **2023**, *96*, 104653. [[CrossRef](#)]
18. Usuga Cadavid, J.P.; Lamouri, S.; Grabot, B.; Pellerin, R.; Fortin, A. Machine learning applied in production planning and control: A state-of-the-art in the era of industry 4.0. *J. Intell. Manuf.* **2020**, *31*, 1531–1558. [[CrossRef](#)]
19. Narciso, D.A.; Martins, F. Application of machine learning tools for energy efficiency in industry: A review. *Energy Rep.* **2020**, *6*, 1181–1199. [[CrossRef](#)]
20. Lwakatara, L.E.; Raj, A.; Crnkovic, I.; Bosch, J.; Olsson, H.H. Large-scale machine learning systems in real-world industrial settings: A review of challenges and solutions. *Inf. Softw. Technol.* **2020**, *127*, 106368. [[CrossRef](#)]
21. Balogun, A.-L.; Tella, A.; Baloo, L.; Adebisi, N. A review of the inter-correlation of climate change, air pollution and urban sustainability using novel machine learning algorithms and spatial information science. *Urban Clim.* **2021**, *40*, 100989. [[CrossRef](#)]
22. Berrang-Ford, L.; Sietsma, A.J.; Callaghan, M.; Minx, J.C.; Scheelbeek, P.F.D.; Haddaway, N.R.; Haines, A.; Dangour, A.D. Systematic mapping of global research on climate and health: A machine learning review. *Lancet Planet. Health* **2021**, *5*, 514–525. [[CrossRef](#)]
23. Sachs, J.D.; Lafortune, G.; Fuller, G.; Drumm, E. Implementing the SDG Stimulus. In *Sustainable Development Report 2023*; Dublin University Press: Dublin, Ireland, 2023. [[CrossRef](#)]
24. Harie, Y.; Gautam, B.P.; Wasaki, K. Computer vision techniques for growth prediction: A prisma-based systematic literature review. *Appl. Sci.* **2023**, *13*, 5335. [[CrossRef](#)]

25. Madlener, R.; Sunak, Y. Impacts of urbanization on urban structures and energy demand: What can we learn for urban energy planning and urbanization management? *Sustain. Cities Soc.* **2011**, *1*, 45–53. [[CrossRef](#)]
26. Zhou, W.; Zhu, B.; Chen, D.; Griffy-Brown, C.; Ma, Y.; Fei, W. Energy consumption patterns in the process of China's urbanization. *Popul. Environ.* **2012**, *33*, 202–220. [[CrossRef](#)]
27. Mahmood, H.; Alkhateeb, T.T.Y.; Furqan, M. Industrialization, urbanization and CO<sub>2</sub> emissions in Saudi Arabia: Asymmetry analysis. *Energy Rep.* **2020**, *6*, 1553–1560. [[CrossRef](#)]
28. Cherniwchan, J. Economic growth, industrialization, and the environment. *Resour. Energy Econ.* **2012**, *34*, 442–467. [[CrossRef](#)]
29. Liu, X.; Bae, J. Urbanization and industrialization impact of CO<sub>2</sub> emissions in China. *J. Clean. Prod.* **2018**, *172*, 178–186. [[CrossRef](#)]
30. Pizzulli, V.A.; Telesca, V.; Covatariu, G. Analysis of Correlation between Climate Change and Human Health Based on a Machine Learning Approach. *Healthcare* **2021**, *9*, 86. [[CrossRef](#)]
31. Bollen, J.; van der Zwaan, B.; Brink, C.; Eerens, H. Local air pollution and global climate change: A combined cost-benefit analysis. *Resour. Energy Econ.* **2009**, *31*, 161–181. [[CrossRef](#)]
32. Kinney, P.L. Interactions of Climate Change, Air Pollution, and Human Health. *Curr. Environ. Health Rep.* **2018**, *5*, 179–186. [[CrossRef](#)]
33. Thu, M.Y.; Htun, W.; Aung, Y.L.; Shwe, P.E.E.; Tun, N.M. Smart Air Quality Monitoring System with LoRaWAN. In Proceedings of the 2018 IEEE International Conference on Internet of Things and Intelligence System (IoTIS), Bali, Indonesia, 1–3 November 2018; IEEE: Bali, India, 2018; pp. 10–15. [[CrossRef](#)]
34. Firdaus, R.; Murti, M.A.; Alinursafa, I. Air quality monitoring system based internet of things (IoT) using lora. In Proceedings of the 2019 IEEE International Conference on Internet of Things and Intelligence System (IoTIS), Bali, Indonesia, 5–7 November 2019; pp. 195–200.
35. Bougoudis, I.; Demertzis, K.; Iliadis, L.; Anezakis, V.-D.; Papaleonidas, A. Fuzzfra, a fuzzy semi-supervised forecasting framework: The case of the air pollution in athens. *Neural Comput. Appl.* **2018**, *29*, 375–388. [[CrossRef](#)]
36. Badicu, A.; Suci, G.; Balanescu, M.; Dobrea, M.; Birdici, A.; Orza, O.; Pasat, A. Pms concentration forecasting using arima algorithm. In Proceedings of the 2020 IEEE 91st Vehicular Technology Conference (VTC2020-spring), Antwerp, Belgium, 25–28 May 2020; pp. 1–5.
37. Cihan, P.; Ozel, H.; Ozcan, H.K. Modeling of atmospheric particulate matters via artificial intelligence methods. *Environ. Monit. Assess.* **2021**, *193*, 287. [[CrossRef](#)] [[PubMed](#)]
38. Magazzino, C.; Mele, M.; Sarkodie, S.A. The nexus between COVID-19 deaths, air pollution and economic growth in new york state: Evidence from deep machine learning. *J. Environ. Manag.* **2021**, *286*, 112241. [[CrossRef](#)]
39. Cole, M.A.; Elliott, R.J.R.; Liu, B. The Impact of the Wuhan COVID-19 Lockdown on Air Pollution and Health: A Machine Learning and Augmented Synthetic Control Approach. *Environ. Resour. Econ.* **2020**, *76*, 553–580. [[CrossRef](#)] [[PubMed](#)]
40. Cazzolla Gatti, R.; Velichevskaia, A.; Tateo, A.; Amoroso, N.; Monaco, A. Machine learning reveals that prolonged exposure to air pollution is associated with SARS-CoV-2 mortality and infectivity in italy. *Environ. Pollut.* **2020**, *267*, 115471. [[CrossRef](#)] [[PubMed](#)]
41. Senthilkumar, R.; Venkatakrishnan, P.; Balaji, N. Intelligent based novel embedded system based IoT enabled air pollution monitoring system. *Microprocess. Microsyst.* **2020**, *77*, 103172. [[CrossRef](#)]
42. Sharma, P.K.; Mondal, A.; Jaiswal, S.; Saha, M.; Nandi, S.; De, T.; Saha, S. Indoairsense: A framework for indoor air quality estimation and forecasting. *Atmos. Pollut. Res.* **2021**, *12*, 10–22. [[CrossRef](#)]
43. Kanabkaew, T.; Mekbungwan, P.; Raksakietisak, S.; Kanchanasut, K. Detection of PM<sub>2.5</sub> plume movement from IoT ground level monitoring data. *Environ. Pollut.* **2019**, *252*, 543–552. [[CrossRef](#)]
44. World Health Organization. *WHO Global Air Quality Guidelines: Particulate Matter (PM<sub>2.5</sub> and PM<sub>10</sub>), Ozone, Nitrogen Dioxide, Sulfur Dioxide and Carbon Monoxide: Executive Summary*; World Health Organization: Geneva, Switzerland, 2021.
45. Lin, L.; Di, L.; Yang, R.; Zhang, C.; Yu, E.; Rahman, M.S.; Sun, Z.; Tang, J. Using machine learning approach to evaluate the PM<sub>2.5</sub> concentrations in china from 1998 to 2016. In Proceedings of the 2018 7th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), Hangzhou, China, 6–9 August 2018; pp. 1–5.
46. Ameer, S.; Shah, M.A.; Khan, A.; Song, H.; Maple, C.; Islam, S.U.; Asghar, M.N. Comparative analysis of machine learning techniques for predicting air quality in smart cities. *IEEE Access* **2019**, *7*, 128325–128338. [[CrossRef](#)]
47. Zoran, M.A.; Savastru, R.S.; Savastru, D.M.; Tautan, M.N. Assessing the relationship between ground levels of ozone (O<sub>3</sub>) and nitrogen dioxide (NO<sub>2</sub>) with coronavirus (COVID-19) in milan, italy. *Sci. Total Environ.* **2020**, *740*, 140005. [[CrossRef](#)]
48. El Khaili, M.; Bakkoury, J.; Khat, A.; Alloubane, A. Crowdsourcing by IoT using labview for measuring the air quality. In Proceedings of the 3rd International Conference on Smart City Applications, Tetouan, Morocco, 10–11 October 2018; pp. 1–8.
49. Li, Z.; Yim, S.H.-L.; Ho, K.-F. High temporal resolution prediction of streetlevel PM<sub>2.5</sub> and NO<sub>x</sub> concentrations using machine learning approach. *J. Clean. Prod.* **2020**, *268*, 121975. [[CrossRef](#)]
50. Chang, M.B.; Lee, H.M.; Wu, F.; Lai, C.R. Simultaneous removal of nitrogen oxide/nitrogen dioxide/sulfur dioxide from gas streams by combined plasma scrubbing technology. *J. Air Waste Manag. Assoc.* **2004**, *54*, 941–949. [[CrossRef](#)] [[PubMed](#)]
51. Lara-Cueva, R.A.; Meneses, P.B.; Marquez, M.D.; Gordillo, R.X.; Benitez, D.S. Air quality monitoring system within campus by using wireless sensor networks. In Proceedings of the 2019 14th Iberian Conference on Information Systems and Technologies (CISTI), Coimbra, Portugal, 19–22 June 2019; IEEE: Piscataway, NJ, USA; pp. 1–4. [[CrossRef](#)]



52. Mardani, A.; Streimikiene, D.; Cavallaro, F.; Loganathan, N.; Khoshnoudi, M. Carbon dioxide (CO<sub>2</sub>) emissions and economic growth: A systematic review of two decades of research from 1995 to 2017. *Sci. Total Environ.* **2019**, *649*, 31–49. [[CrossRef](#)] [[PubMed](#)]
53. Prockop, L.D.; Chichkova, R.I. Carbon monoxide intoxication: An updated review. *J. Neurol. Sci.* **2007**, *262*, 122–130. [[CrossRef](#)]
54. Segers, R. Methane production and methane consumption: A review of processes underlying wetland methane fluxes. *Biogeochemistry* **1998**, *41*, 23–51. [[CrossRef](#)]
55. Beerling, D.; Berner, R.A.; Mackenzie, F.T.; Harfoot, M.B.; Pyle, J.A. Methane and the ch<sub>4</sub> related greenhouse effect over the past 400 million years. *Am. J. Sci.* **2009**, *309*, 97–113. [[CrossRef](#)]
56. Kamal, M.S.; Razzak, S.A.; Hossain, M.M. Catalytic oxidation of volatile organic compounds (VOCs)—A review. *Atmos. Environ.* **2016**, *140*, 117–134. [[CrossRef](#)]
57. Raghuvveera, E.; Kanakaraja, P.; Kishore, K.H.; Sriya, C.T.; Prasad, B. D.; Lalith, B.S.K.T. An IoT enabled air quality monitoring system using LoRa and LPWAN. In Proceedings of the 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 8–10 April 2021; IEEE: Piscataway, NJ, USA; pp. 453–459. [[CrossRef](#)]
58. Yang, C.-T.; Liao, C.-J.; Liu, J.-C.; Den, W.; Chou, Y.-C.; Tsai, J.-J. Construction and application of an intelligent air quality monitoring system for healthcare environment. *J. Med. Syst.* **2014**, *38*, 15. [[CrossRef](#)]
59. Tran, T.V.; Dang, N.T.; Chung, W.-Y. Battery-free smart-sensor system for real-time indoor air quality monitoring. *Sens. Actuators B Chem.* **2017**, *248*, 930–939. [[CrossRef](#)]
60. Mitchell, T.M. *The Discipline of Machine Learning*; Machine Learning, School of Computer Science, Carnegie Mellon University: Pittsburg, PA, USA, 2006; Volume 9.
61. Hastie, T.; Tibshirani, R.; Friedman, J. Overview of Supervised Learning. In *The Elements of Statistical Learning*; Springer Series in Statistics; Springer: New York, NY, USA, 2009; pp. 9–41. [[CrossRef](#)]
62. Quinlan, J.R. Induction of decision trees. *Mach. Learn.* **1986**, *1*, 81–106. [[CrossRef](#)]
63. Dridi, S. Supervised learning—A systematic literature review. *OSF* **2021**. [[CrossRef](#)]
64. Adams, M.D.; Massey, F.; Chastko, K.; Cupini, C. Spatial modelling of particulate matter air pollution sensor measurements collected by community scientists while cycling, land use regression with spatial cross-validation, and applications of machine learning for data correction. *Atmos. Environ.* **2020**, *230*, 117479. [[CrossRef](#)]
65. Bozdağ, A.; Dokuz, Y.; Gökçek, B. Spatial prediction of PM<sub>10</sub> concentration using machine learning algorithms in ankara, turkey. *Environ. Pollut.* **2020**, *263*, 114635. [[CrossRef](#)] [[PubMed](#)]
66. Chen, J.; de Hoogh, K.; Gulliver, J.; Hoffmann, B.; Hertel, O.; Ketzel, M.; Bauwelinck, M.; van Donkelaar, A.; Hvidtfeldt, U.A.; Katsouyanni, K.; et al. A comparison of linear regression, regularization, and machine learning algorithms to develop Europe-wide spatial models of fine particles and nitrogen dioxide. *Environ. Int.* **2019**, *130*, 104934. [[CrossRef](#)] [[PubMed](#)]
67. Taheri, S.; Razban, A. Learning-based CO<sub>2</sub> concentration prediction: Application to indoor air quality control using demand-controlled ventilation. *Build. Environ.* **2021**, *205*, 108164. [[CrossRef](#)]
68. Chen, S.; Yuval; Broday, D.M. Re-framing the gaussian dispersion model as a nonlinear regression scheme for retrospective air quality assessment at a high spatial and temporal resolution. *Environ. Model. Softw.* **2020**, *125*, 104620. [[CrossRef](#)]
69. Yuchi, W.; Gombojav, E.; Boldbaatar, B.; Galsuren, J.; Enkhmaa, S.; Beejin, B.; Naidan, G.; Ochir, C.; Legtseg, B.; Byambaa, T.; et al. Evaluation of random forest regression and multiple linear regression for predicting indoor fine particulate matter concentrations in a highly polluted city. *Environ. Pollut.* **2019**, *245*, 746–753. [[CrossRef](#)]
70. Murillo-Escobar, J.; Sepulveda-Suescun, J.P.; Correa, M.A.; Orrego-Metaute, D. Forecasting concentrations of air pollutants using support vector regression improved with particle swarm optimization: Case study in aburrá valley, colombia. *Urban Clim.* **2019**, *29*, 100473. [[CrossRef](#)]
71. Son, Y.; Osornio-Vargas, R.; O'Neill, M.S.; Hystad, P.; Texcalac-Sangrador, J.L.; Ohman-Strickland, P.; Meng, Q.; Schwander, S. Land use regression models to assess air pollution exposure in Mexico City using finer spatial and temporal input parameters. *Sci. Total Environ.* **2018**, *639*, 40–48. [[CrossRef](#)]
72. Araujo, L.N.; Belotti, J.T.; Alves, T.A.; de Souza Tadano, Y.; Siqueira, H. Ensemble method based on artificial neural networks to estimate air pollution health risks. *Environ. Model. Softw.* **2020**, *123*, 104567. [[CrossRef](#)]
73. De Mattos Neto, P.S.G.; Firmino, P.R.A.; Siqueira, H.; De Souza Tadano, Y.; Alves, T.A.; De Oliveira, J.F.L.; Da Nobrega Marinho, M.H.; Madeiro, F. Neural-based ensembles for particulate matter forecasting. *IEEE Access* **2021**, *9*, 14470–14490. [[CrossRef](#)]
74. Shahriar, S.A.; Kayes, I.; Hasan, K.; Salam, M.A.; Chowdhury, S. Applicability of machine learning in modeling of atmospheric particle pollution in Bangladesh. *Air Qual. Atmos. Health* **2020**, *13*, 1247–1256. [[CrossRef](#)] [[PubMed](#)]
75. Shams, S.R.; Jahani, A.; Moeinaddini, M.; Khorasani, N. Air carbon monoxide forecasting using an artificial neural network in comparison with multiple regression. *Model. Earth Syst. Environ.* **2020**, *6*, 1467–1475. [[CrossRef](#)]
76. Ketu, S.; Mishra, P.K. Scalable kernel-based SVM classification algorithm on imbalance air quality data for proficient healthcare. *Complex Intell. Syst.* **2021**, *7*, 2597–2615. [[CrossRef](#)]
77. Zhang, L.; Tian, F.; Nie, H.; Dang, L.; Li, G.; Ye, Q.; Kadri, C. Classification of multiple indoor air contaminants by an electronic nose and a hybrid support vector machine. *Sens. Actuators B Chem.* **2012**, *174*, 114–125. [[CrossRef](#)]



78. Singh, K.P.; Gupta, S.; Rai, P. Identifying pollution sources and predicting urban air quality using ensemble learning methods. *Atmos. Environ.* **2013**, *47*, 426–437. [[CrossRef](#)]
79. Tella, A.; Balogun, A.-L.; Adebisi, N.; Abdullah, S. Spatial assessment of PM<sub>10</sub> hotspots using random forest, K-nearest neighbour and Naïve Bayes. *Atmos. Pollut. Res.* **2021**, *12*, 101202. [[CrossRef](#)]
80. Velásquez, R.M.A.; Lara, J.V.M. Gaussian approach for probability and correlation between the number of COVID-19 cases and the air pollution in Lima. *Urban Clim.* **2020**, *33*, 100664. [[CrossRef](#)] [[PubMed](#)]
81. Mokhtari, I.; Bechkit, W.; Rivano, H.; Yaici, M.R. Uncertainty-aware deep learning architectures for highly dynamic air quality prediction. *IEEE Access* **2021**, *9*, 14765–14778. [[CrossRef](#)]
82. Tao, Q.; Liu, F.; Li, Y.; Sidorov, D. Air pollution forecasting using a deep learning model based on 1D convnets and bidirectional GRU. *IEEE Access* **2019**, *7*, 76690–76698. [[CrossRef](#)]
83. Han, Y.; Lam, J.C.; Li, V.O.; Reiner, D. A Bayesian LSTM model to evaluate the effects of air pollution control regulations in Beijing, China. *Environ. Sci. Policy* **2021**, *115*, 26–34. [[CrossRef](#)]
84. AlOmar, M.K.; Hameed, M.M.; AlSaadi, M.A. Multi hours ahead prediction of surface ozone gas concentration: Robust artificial intelligence approach. *Atmos. Pollut. Res.* **2020**, *11*, 1572–1587. [[CrossRef](#)]
85. Jiang, P.; Li, C.; Li, R.; Yang, H. An innovative hybrid air pollution early-warning system based on pollutants forecasting and extenics evaluation. *Knowl.-Based Syst.* **2019**, *164*, 174–192. [[CrossRef](#)]
86. Ravindra, K.; Bahadur, S.S.; Katoch, V.; Bhardwaj, S.; Kaur-Sidhu, M.; Gupta, M.; Mor, S. Application of machine learning approaches to predict the impact of ambient air pollution on outpatient visits for acute respiratory infections. *Sci. Total Environ.* **2023**, *858*, 159509. [[CrossRef](#)]
87. Dutta, D.; Pal, S.K. Prediction and assessment of the impact of COVID-19 lockdown on air quality over kolkata: A deep transfer learning approach. *Environ. Monit. Assess.* **2023**, *195*, 223. [[CrossRef](#)]
88. Van, N.; Van Thanh, P.; Tran, D.; Tran, D.-T. A new model of air quality prediction using lightweight machine learning. *Int. J. Environ. Sci. Technol.* **2023**, *20*, 2983–2994. [[CrossRef](#)]
89. Eren, B.; Aksangür, İ.; Erden, C. Predicting next hour fine particulate matter (PM<sub>2.5</sub>) in the istanbul metropolitan city using deep learning algorithms with time windowing strategy. *Urban Clim.* **2023**, *48*, 101418. [[CrossRef](#)]
90. Barthwal, A. A markov chain-based IoT system for monitoring and analysis of urban air quality. *Environ. Monit. Assess.* **2023**, *195*, 235. [[CrossRef](#)]
91. Wang, L.; Zhao, Y.; Shi, J.; Ma, J.; Liu, X.; Han, D.; Gao, H.; Huang, T. Predicting ozone formation in petrochemical industrialized lanzhou city by interpretable ensemble machine learning. *Environ. Pollut.* **2023**, *318*, 120798. [[CrossRef](#)]
92. Persis, J.; Amar, A.B. Predictive modeling and analysis of air quality—visualizing before and during COVID-19 scenarios. *J. Environ. Manag.* **2023**, *327*, 116911. [[CrossRef](#)]
93. Koo, Y.-S.; Kwon, H.-Y.; Bae, H.; Yun, H.-Y.; Choi, D.-R.; Yu, S.; Wang, K.-H.; Koo, J.-S.; Lee, J.-B.; Choi, M.-H.; et al. A development of PM<sub>2.5</sub> forecasting system in south korea using chemical transport modeling and machine learning. *Asia-Pac. J. Atmos. Sci.* **2023**, *59*, 577–595. [[CrossRef](#)]
94. Natsagdorj, N.; Zhou, H. Prediction of PM<sub>2.5</sub> concentration in Ulaanbaatar with deep learning models. *Urban Clim.* **2023**, *47*, 101357.
95. Falah, S.; Kizel, F.; Banerjee, T.; Broday, D.M. Accounting for the aerosol type and additional satellite-borne aerosol products improves the prediction of PM<sub>2.5</sub> concentrations. *Environ. Pollut.* **2023**, *320*, 121119. [[CrossRef](#)]
96. Xie, Q.; Ni, J.-Q.; Li, E.; Bao, J.; Zheng, P. Sequential air pollution emission estimation using a hybrid deep learning model and health-related ventilation control in a pig building. *J. Clean. Prod.* **2022**, *371*, 133714. [[CrossRef](#)]
97. Muthukumar, P.; Cocom, E.; Nagrecha, K.; Comer, D.; Burga, I.; Taub, J.; Calvert, C.F.; Holm, J.; Pourhomayoun, M. Predicting PM<sub>2.5</sub> atmospheric air pollution using deep learning with meteorological data and ground-based observations and remote-sensing satellite big data. *Air Qual. Atmos. Health* **2022**, *15*, 1221–1234. [[CrossRef](#)]
98. Abu El-Magd, S.; Soliman, G.; Morsy, M.; Kharbush, S. Environmental hazard assessment and monitoring for air pollution using machine learning and remote sensing. *Int. J. Environ. Sci. Technol.* **2022**, *20*, 6103–6116. [[CrossRef](#)]
99. Huang, C.; Hu, T.; Duan, Y.; Li, Q.; Chen, N.; Wang, Q.; Zhou, M.; Rao, P. Effect of urban morphology on air pollution distribution in high-density urban blocks based on mobile monitoring and machine learning. *Build. Environ.* **2022**, *219*, 109173. [[CrossRef](#)]
100. Gilik, A.; Ogrenici, A.S.; Ozmen, A. Air quality prediction using CNN+LSTM-based hybrid deep learning architecture. *Environ. Sci. Pollut. Res.* **2022**, *29*, 11920–11938. [[CrossRef](#)]
101. Kumar, K.; Pande, B.P. Air pollution prediction with machine learning: A case study of Indian cities. *Int. J. Environ. Sci. Technol.* **2022**, *20*, 5333–5348. [[CrossRef](#)]
102. Sethi, J.K.; Mittal, M. Efficient weighted naive bayes classifiers to predict air quality index. *Earth Sci. Inform.* **2022**, *15*, 541–552. [[CrossRef](#)]
103. Abirami, G.; Girija, R.; Das, A.; Sreenivasan, N. Predicting air quality index with machine learning models. In *Machine Learning and Deep Learning in Efficacy Improvement of Healthcare Systems*; CRC Press: Boca Raton, FL, USA, 2022; pp. 353–371.
104. Chen, Y.-W.; Medya, S.; Chen, Y.-C. Investigating variable importance in ground-level ozone formation with supervised learning. *Atmos. Environ.* **2022**, *282*, 119148. [[CrossRef](#)]
105. Cheng, X.; Zhang, W.; Wenzel, A.; Chen, J. Stacked ResNet-LSTM and CORAL model for multi-site air quality prediction. *Neural Comput. Appl.* **2022**, *34*, 13849–13866. [[CrossRef](#)]

106. Cho, J.H.; Moon, J.W. Integrated artificial neural network prediction model of indoor environmental quality in a school building. *J. Clean. Prod.* **2022**, *344*, 131083. [[CrossRef](#)]
107. Yadav B, V.; Geetha, D. Prediction of concentration of air pollution using deep and machine learning. In Proceedings of the 2022 8th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 25–26 March 2022; IEEE: Piscataway, NJ, USA, 2022; Volume 1, pp. 1369–1375.
108. Liu, C.-C.; Lin, T.-C.; Yuan, K.-Y.; Chiueh, P.-T. Spatio-temporal prediction and factor identification of urban air quality using support vector machine. *Urban Clim.* **2022**, *41*, 101055. [[CrossRef](#)]
109. Martín-Baos, J.Á.; Rodríguez-Benitez, L.; García-Ródenas, R.; Liu, J. IoT based monitoring of air quality and traffic using regression analysis. *Appl. Soft Comput.* **2022**, *115*, 108282. [[CrossRef](#)]
110. Asha, P.; Natrayan, L.; Geetha, B.; Beulah, J.R.; Sumathy, R.; Varalakshmi, G.; Neelakandan, S. IoT enabled environmental toxicology for air pollution monitoring using ai techniques. *Environ. Res.* **2022**, *205*, 112574. [[CrossRef](#)]
111. Ferreira, W.d.A.P.; Grout, I.; da Silva, A.C.R. Application of a fuzzy artmap neural network for indoor air quality prediction. In Proceedings of the 2022 International Electrical Engineering Congress (IEECON), Khon Kaen, Thailand, 9–11 March 2022; IEEE: Piscataway, NJ, USA, 2022; pp. 1–4.
112. Choudhury, A.; Middya, A.I.; Roy, S. A comparative study of machine learning and deep learning techniques in forecasting air pollution levels. In Proceedings of the International Conference on Data Science and Applications, Kolkata, India, 26–27 March 2022; Springer: Berlin/Heidelberg, Germany, 2022; pp. 607–619.
113. Qader, M.R.; Khan, S.; Kamal, M.; Usman, M.; Haseeb, M. Forecasting carbon emissions due to electricity power generation in Bahrain. *Environ. Sci. Pollut. Res.* **2022**, *29*, 17346–17357. [[CrossRef](#)]
114. Wei, X.; Wang, X.; Zhu, T.; Gong, Z. Fusion prediction model of atmospheric pollutant based on self-organized feature. *IEEE Access* **2021**, *9*, 8110–8120. [[CrossRef](#)]
115. Meena, K.; Raja Sekar, R.; Mayuri, A.V.R.; Preetha, V.; Krishna Veni, N.N. 5G narrow band-IoT based air contamination prediction using recurrent neural network. *Sustain. Comput. Inform. Syst.* **2022**, *33*, 100619. [[CrossRef](#)]
116. Chang, Y.-S.; Abimannan, S.; Chiao, H.-T.; Lin, C.-Y.; Huang, Y.-P. An ensemble learning based hybrid model and framework for air pollution forecasting. *Environ. Sci. Pollut. Res.* **2020**, *27*, 38155–38168. [[CrossRef](#)]
117. Chang, Y.-S.; Chiao, H.-T.; Abimannan, S.; Huang, Y.-P.; Tsai, Y.-T.; Lin, K.-M. An LSTM-based aggregated model for air pollution forecasting. *Atmos. Pollut. Res.* **2020**, *11*, 1451–1463. [[CrossRef](#)]
118. Magazzino, C.; Mele, M.; Schneider, N. The relationship between air pollution and COVID-19-related deaths: An application to three french cities. *Appl. Energy* **2020**, *279*, 115835. [[CrossRef](#)]
119. Zeinalnezhad, M.; Chofreh, A.G.; Goni, F.A.; Klemeš, J.J. Air pollution prediction using semi-experimental regression model and adaptive neuro-fuzzy inference system. *J. Clean. Prod.* **2020**, *261*, 121218. [[CrossRef](#)]
120. Alyousifi, Y.; Othman, M.; Faye, I.; Sockalingam, R.; Silva, P.C.L. Markov Weighted Fuzzy Time-Series Model Based on an Optimum Partition Method for Forecasting Air Pollution. *Int. J. Fuzzy Syst.* **2020**, *22*, 1468–1486. [[CrossRef](#)]
121. Lu, X.; Wang, J.; Yan, Y.; Zhou, L.; Ma, W. Estimating hourly PM<sub>2.5</sub> concentrations using Himawari-8 AOD and a DBSCAN-modified deep learning model over the YRDUA, China. *Atmos. Pollut. Res.* **2021**, *12*, 183–192. [[CrossRef](#)]
122. Zhou, Y.; Zhao, X.; Lin, K.-P.; Wang, C.-H.; Li, L. A gaussian process mixture model-based hard-cut iterative learning algorithm for air quality prediction. *Appl. Soft Comput.* **2019**, *85*, 105789. [[CrossRef](#)]
123. Yadav, M.; Jain, S.; Seeja, K.R. Prediction of air quality using time series data mining. In Proceedings of the International Conference on Innovative Computing and Communications, Ostrava, Czech Republic, 21–22 March 2019; Bhattacharyya, S., Hassanien, A.E., Gupta, D., Khanna, A., Pan, I., Eds.; Springer: Singapore, 2019; pp. 13–20.
124. Khiat, A.; Bahnasse, A.; Bakkoury, J.; El Khaili, M.; Louhab, F.E. New approach based internet of things for a clean atmosphere. *Int. J. Inf. Technol.* **2019**, *11*, 89–95. [[CrossRef](#)]
125. Sahil, K.; Mehta, P.; Bhardwaj, S.K.; Dhaliwal, L.K. Development of mitigation strategies for the climate change using artificial intelligence to attain sustainability. In *Visualization Techniques for Climate Change with Machine Learning and Artificial Intelligence*; Elsevier: Amsterdam, The Netherlands, 2023; pp. 421–448.
126. Vinuesa, R.; Azizpour, H.; Leite, I.; Balaam, M.; Dignum, V.; Domisch, S.; Felländer, A.; Langhans, S.D.; Tegmark, M.; Nerini, F.F. The role of artificial intelligence in achieving the Sustainable Development Goals. *Nat. Commun.* **2020**, *11*, 233. [[CrossRef](#)]
127. Sharma, N.; Panwar, D. Green IoT: Advancements and Sustainability with Environment by 2050. In Proceedings of the 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 4–5 June 2020; IEEE: Noida, India, 2020; pp. 1127–1132. [[CrossRef](#)]
128. Liu, J.; Liu, L.; Qian, Y.; Song, S. The effect of artificial intelligence on carbon intensity: Evidence from China’s industrial sector. *Socio-Econ. Plan. Sci.* **2022**, *83*, 101002. [[CrossRef](#)]
129. Fraga-Lamas, P.; Lopes, S.I.; Fern´andez-Caram´es, T.M. Green IoT and Edge AI as Key Technological Enablers for a Sustainable Digital Transition towards a Smart Circular Economy: An Industry 5.0 Use Case. *Sensors* **2021**, *21*, 5745. [[CrossRef](#)]
130. Lannelongue, L.; Grealey, J.; Inouye, M. Green Algorithms: Quantifying the Carbon Footprint of Computation. *Adv. Sci.* **2021**, *8*, 2100707. [[CrossRef](#)]

131. Fernandez-Cerero, D.; Fernandez-Montes, A.; Jakobik, A. Limiting Global Warming by Improving Data-Centre Software. *IEEE Access* **2020**, *8*, 44048–44062. [[CrossRef](#)]
132. Page, M.J.; McKenzie, J.E.; Bossuyt, P.M.; Boutron, I.; Hoffmann, T.C.; Mulrow, C.D.; Shamseer, L.; Tetzlaff, J.M.; Akl, E.A.; Brennan, S.E.; et al. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* **2021**, *372*, n71. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.