# Air quality monitoring

Ganesh college of engineering

Submitted by ; C.Nandhini

T.Swathi

S.Sevvanthi

M.Arivumathi

## Introduction

fuels, deforestation, and transportation emissions, not only create air pollution but also contribute to global climate change.

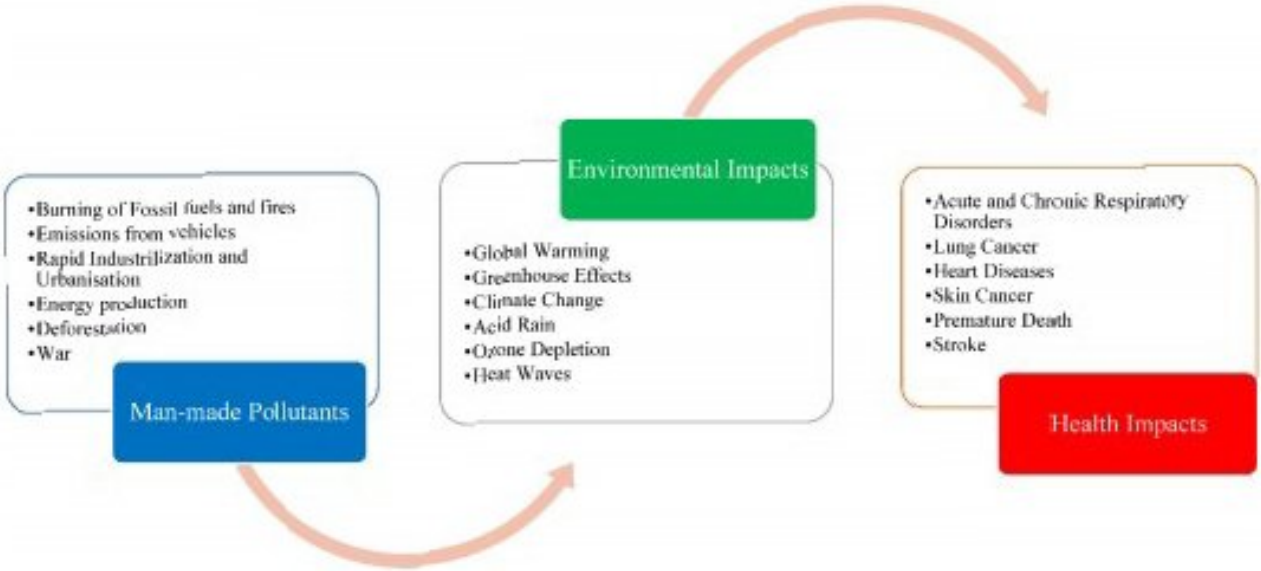


**Figure 1.** Potential Relationship between Man-made Pollutants, Environmental Impacts and Health Impacts.

Human health and life quality, which are determined by environmental psychosocial, physical, biological, and social characteristics, are factors in environmental health [5]. Climate changes lead to alterations in weather and temperature pattern. Although these changes are natural, humans have been the dominant contributor to climate change since the 19th century, primarily to the burning of fossil fuels which generates $CO_2$. Ahmed et al. determined the regional and periodic change of $CO_2$ and Particulate matter in the important Libyan city of Misurata and the measurements were made during the months of November and February once every three days [6]. The results were very useful to understand the regional and periodical characteristics of air pollution. Cetin et al. evaluated the regional and temporal changes in air pollution and measured the $CO_2$ concentration and particulate matter concentration in various regions of Bursa city. The results of the study revealed that the $CO_2$ was statistically not significant with respect to the season but the Particulate matter shows, statistically, a 99.9% confidence level by season [7]. In the Turkish case study, the author examined the variation in the indoor $CO_2$ concentration in the examination hall and the findings indicate that indoor $CO_2$ levels are higher than 1500 ppm while the start of exams of threshold value within 10 min [8]. In another interesting study, the authors examined the Pb and Cr pollution in the capital city of Turkiye by collecting topsoil samples from 50 regions; thus, it is clear that air pollution is very dangerous to the environment and ecosystem. The earth is currently warming faster than at any other time in recorded history. Warmer temperatures are shifting weather patterns and disturbing nature's normal equilibrium. As a result of climate change, storms, floods, cold spells, and heat waves are expected to have a greater socioeconomic impact. Heat Waves (HWs) are expected to grow increasingly strong and common as a result of man-made climate change. HWs are clearly major events that can induce fast changes in biodiversity patterns as well as ecosystem structure and functioning as a result of human climate change [9]. To overcome the above-mentioned issues, it is very important to forecast air pollution, weather conditions, and climate changes to implement an early-warning system.

Artificial Intelligence is used to imitate the human mind's problem-solving and decision-making abilities. Figure 2 shows the relationship between Artificial Intelligence methods, Machine Learning algorithms, and Deep Learning algorithms. To limit public exposure risks due to pollution, AI should be used as an important method for environ-

mental pollution forecasting and which helps policymakers to develop better policies in the case of environmental protection [10]. AI can manage complicated and non-linear interactions between spatial-temporal factors, which makes it a better method for air pollution predicting and forecasting [11]; it offers an outstanding capacity for tracking the current pollution condition and demonstrates a precise and quick method for identifying pollution hotspot areas. Deep learning is a subset of the AI method which is also used for forecasting and predicting extreme weather conditions. The important ability of AI is the easy management of a lot of data and real-world complex problems; further, it increases the accuracy in the case of weather forecasting and modelling which helps policymakers and decision-makers [12].

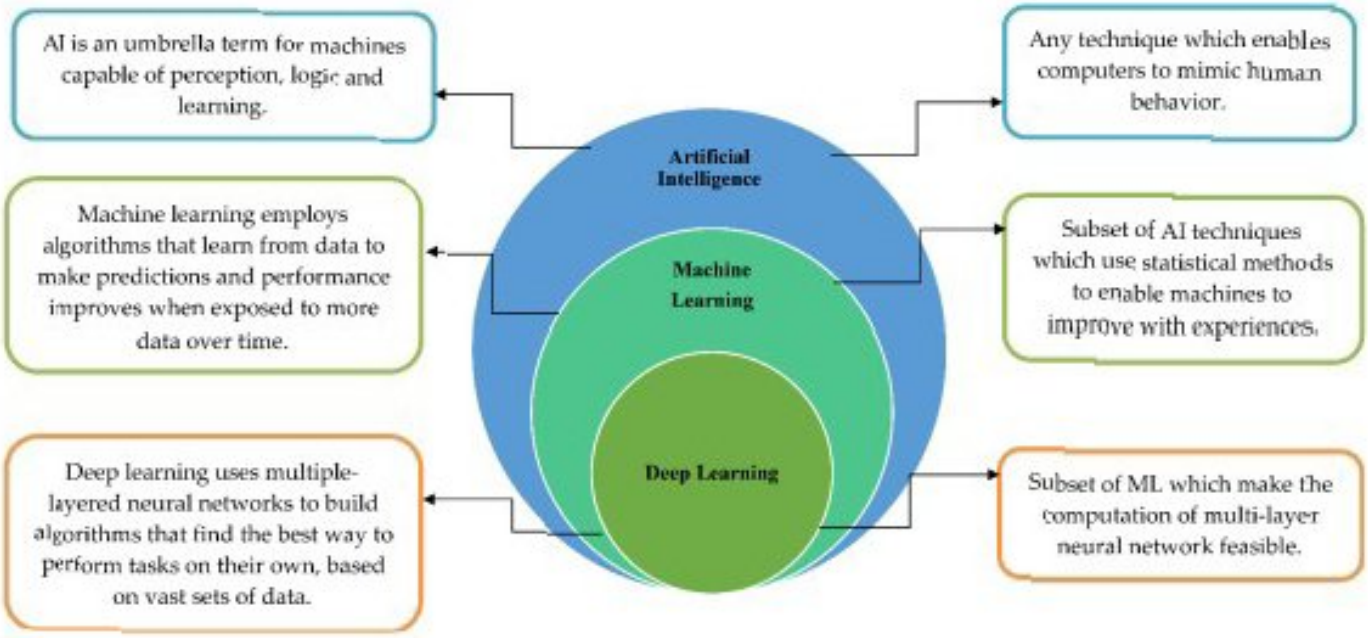


**Figure 2.** Relationship between Artificial Intelligence, Machine Learning and Deep Learning.

### *1.1. Sources and Impacts of Pollutants on Environmental and Human Health*

In many parts of the world, air pollution is a bigger crisis due to its direct impact not only on environmental health but also on human health, which leads to various environmental crises and notably health crises such as early death, respiratory illness, stroke, cardiovascular disease and etc. [13]. Natural and anthropogenic activity generates a harmful and toxic mixture of particles and gases that are emitted into the atmosphere which leads to several crises [14]. Natural calamities such as sudden occurrences of forest fires and volcanic eruptions emitted toxic pollutants such as $SO_2$, $CO_2$, $NO_2$, Carbon monoxide, and Particulate emission [15]. Nowadays, due to increased industry and automotive gas emissions, atmospheric air has become more poisonous, contaminating the air we breathe. The main air pollutants include $CO_2$, $NO_2$, $SO_2$, CO, and particulate matter, all of which have adverse effects on the ecosystem and individuals. Figure 3 represents the major air pollutants and their causes on the environment and humans. Various major air pollutants lead to the environmental atmosphere causing the greenhouse effect, ozone reduction, and photochemical smog, aggravating the environment and producing significant environmental disasters [16], which are responsible for several severe health issues (i.e., chronic disease) [17]. Volcanic eruptions and industrial growth are the major causes of sulphur dioxide ($SO_2$). The presence of $SO_2$ in the atmosphere may also contribute to acid rain, which harms the ecosystem, and also leads to a worsening of the human respiratory system [18].

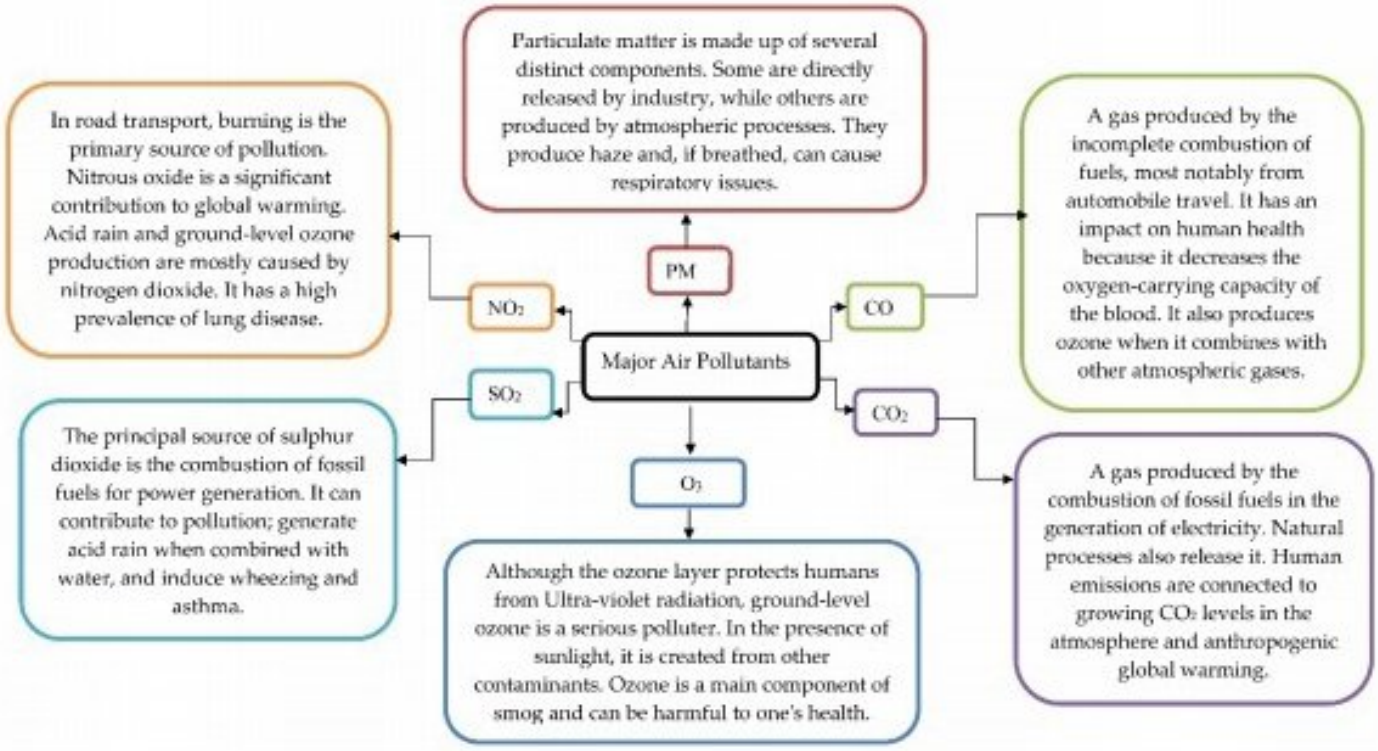


**Figure 3.** Major Air Pollutants and their Sources and Effects.

The burning of fuels in industry and transportation is the major cause of Nitrogen dioxide emissions [3,17]. $NO_2$ exposure results in adverse effects on human individuals and ecosystems, resulting in a rise in lung disease prevalence. The ozone gas at ground level causes severe health impacts to humans due to its high oxidizing capabilities [3]. When we are exposed to air pollution for an extended period, our pulmonary function suffers significantly. Lung cancer, asthma and chronic obstructive pulmonary disease can all be caused by a decline in pulmonary function [4,17]. According to studies, particles with a diameter of 10 μm can pass through the respiratory tract, whereas particles with a diameter of less than 5 μm can penetrate the deep section of the bronchioles [18]. Furthermore, particles smaller than 1 μm can enter the alveoli [17]. Poor air quality may impair cognition and contribute to mental illnesses [17,18]. As a consequence, reducing air pollution can be an effective way to combat the diseases indicated above while also benefiting society [3,19]. Air quality forecasting, monitoring, and early warning systems, as preventive measures, form the foundation for successful pollution control measures and the development of environmental laws to improve air quality, and it is very helpful in designing sustainable smart cities, environmental sustainability, and pollution control management. As preventive measures, air quality forecasting, early warning systems and monitoring are the foundation for successful pollution control measures and the development of environmental laws to improve air quality and it is very helpful in designing sustainable smart cities, environment sustainability and pollution control management.

### *1.2. Current Status of Research on Environmental Pollution Forecasting Techniques*

Air pollution is an inescapable fact of life; it is well known that the effects of environmental pollution are more severe than those of soil and water pollution [19]. Governments all across the world are attempting to improve air quality by enacting various environmental laws. Without a question, most developed countries want to use AI technologies and approaches in their environmental policy today. The forecasting models may be classified into three categories: physical models, statistical models, and AI models; it primarily studies weather phenomena and meteorological conditions for the physical forecasting

mental pollution forecasting and which helps policymakers to develop better policies in the case of environmental protection [10]. AI can manage complicated and non-linear interactions between spatial-temporal factors, which makes it a better method for air pollution predicting and forecasting [11]; it offers an outstanding capacity for tracking the current pollution condition and demonstrates a precise and quick method for identifying pollution hotspot areas. Deep learning is a subset of the AI method which is also used for forecasting and predicting extreme weather conditions. The important ability of AI is the easy management of a lot of data and real-world complex problems; further, it increases the accuracy in the case of weather forecasting and modelling which helps policymakers and decision-makers [12].

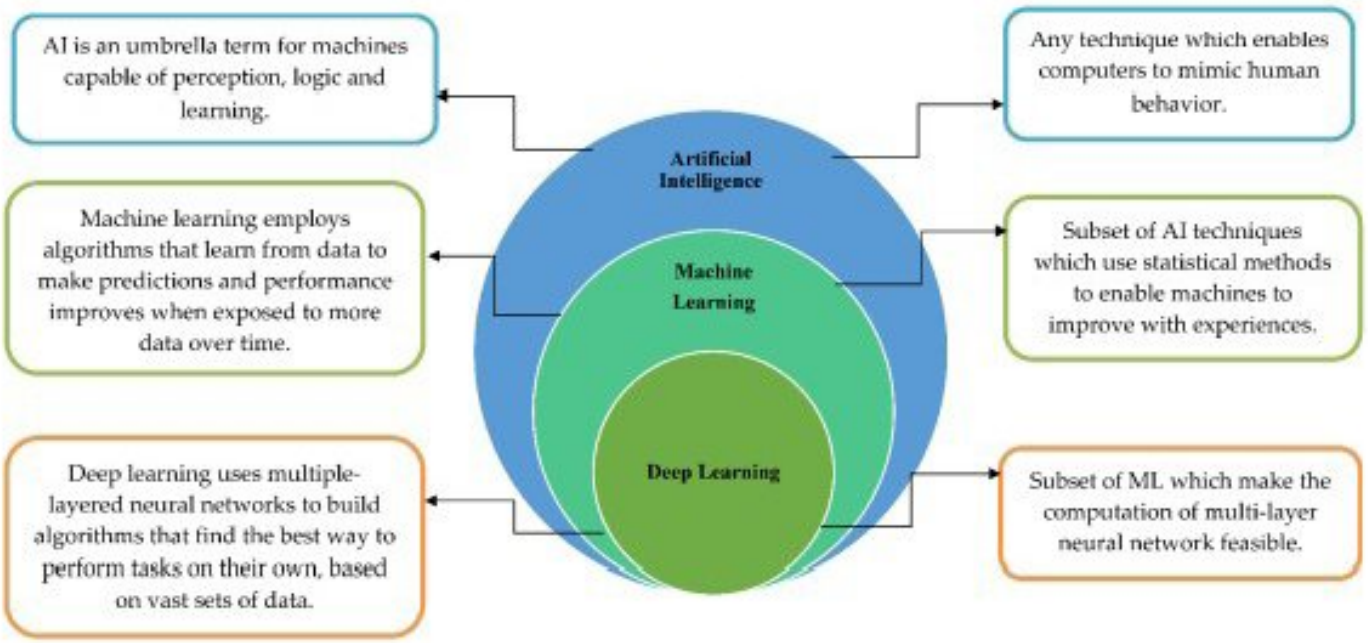


**Figure 2.** Relationship between Artificial Intelligence, Machine Learning and Deep Learning.

### *1.1. Sources and Impacts of Pollutants on Environmental and Human Health*

In many parts of the world, air pollution is a bigger crisis due to its direct impact not only on environmental health but also on human health, which leads to various environmental crises and notably health crises such as early death, respiratory illness, stroke, cardiovascular disease and etc. [13]. Natural and anthropogenic activity generates a harmful and toxic mixture of particles and gases that are emitted into the atmosphere which leads to several crises [14]. Natural calamities such as sudden occurrences of forest fires and volcanic eruptions emitted toxic pollutants such as $SO_2$, $CO_2$, $NO_2$, Carbon monoxide, and Particulate emission [15]. Nowadays, due to increased industry and automotive gas emissions, atmospheric air has become more poisonous, contaminating the air we breathe. The main air pollutants include $CO_2$, $NO_2$, $SO_2$, CO, and particulate matter, all of which have adverse effects on the ecosystem and individuals. Figure 3 represents the major air pollutants and their causes on the environment and humans. Various major air pollutants lead to the environmental atmosphere causing the greenhouse effect, ozone reduction, and photochemical smog, aggravating the environment and producing significant environmental disasters [16], which are responsible for several severe health issues (i.e., chronic disease) [17]. Volcanic eruptions and industrial growth are the major causes of sulphur dioxide ($SO_2$). The presence of $SO_2$ in the atmosphere may also contribute to acid rain, which harms the ecosystem, and also leads toa worsening of the human respiratory system [18].

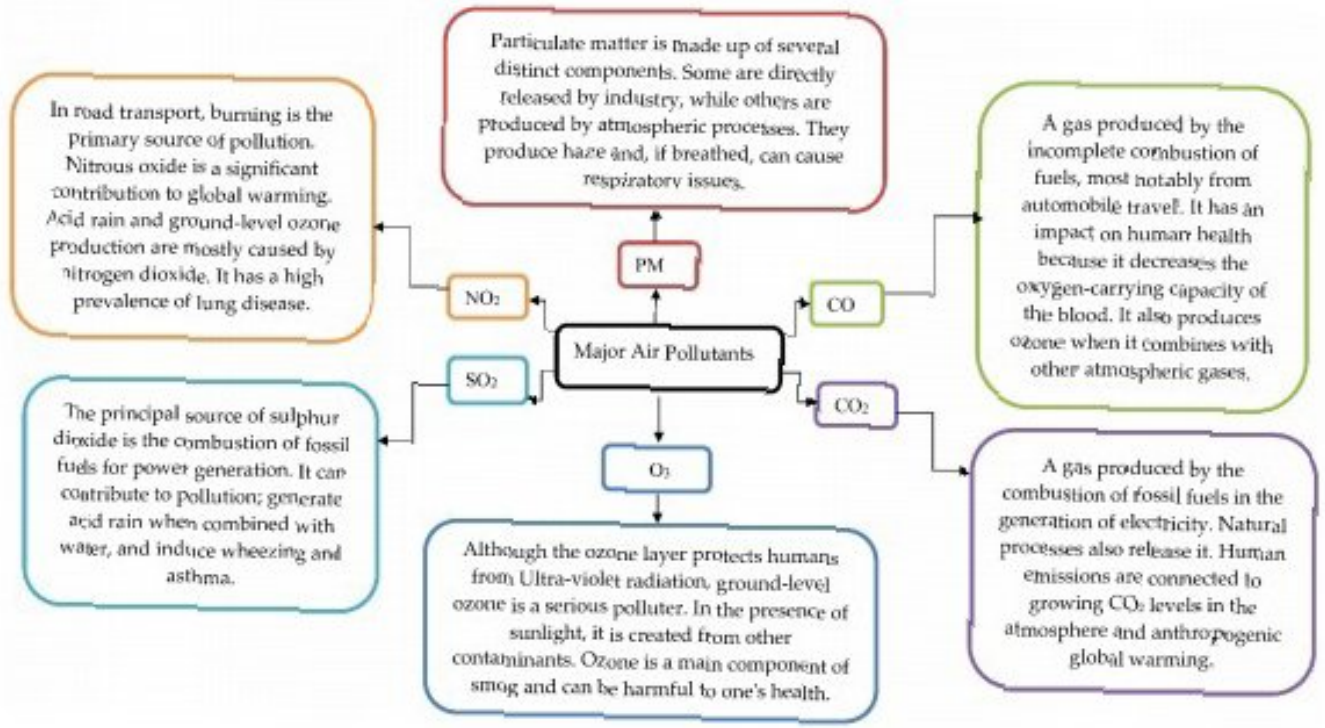


**Figure 3.** Major Air Pollutants and their Sources and Effects.

The burning of fuels in industry and transportation is the major cause of Nitrogen dioxide emissions [3,17]. $NO_2$ exposure results in adverse effects on human individuals and ecosystems, resulting in a rise in lung disease prevalence. The ozone gas at ground level causes severe health impacts to humans due to its high oxidizing capabilities [3]. When we are exposed to air pollution for an extended period, our pulmonary function suffers significantly. Lung cancer, asthma and chronic obstructive pulmonary disease can all be caused by a decline in pulmonary function [4,17]. According to studies, particles with a diameter of 10 µm can pass through the respiratory tract, whereas particles with a diameter of less than 5 µm can penetrate the deep section of the bronchioles [18]. Furthermore, particles smaller than 1 µm can enter the alveoli [17]. Poor air quality may impair cognition and contribute to mental illnesses [17,18]. As a consequence, reducing air pollution can be an effective way to combat the diseases indicated above while also benefiting society [3,19]. Air quality forecasting, monitoring, and early warning systems, as preventive measures, form the foundation for successful pollution control measures and the development of environmental laws to improve air quality, and it is very helpful in designing sustainable smart cities, environmental sustainability, and pollution control management. As preventive measures, air quality forecasting, early warning systems and monitoring are the foundation for successful pollution control measures and the development of environmental laws to improve air quality and it is very helpful in designing sustainable smart cities, environment sustainability and pollution control management.

### 1.2. Current Status of Research on Environmental Pollution Forecasting Techniques

Air pollution is an inescapable fact of life; it is well known that the effects of environmental pollution are more severe than those of soil and water pollution [19]. Governments all across the world are attempting to improve air quality by enacting various environmental laws. Without a question, most developed countries want to use AI technologies and approaches in their environmental policy today. The forecasting models may be classified into three categories: physical models, statistical models, and AI models; it primarily studies weather phenomena and meteorological conditions for the physical forecasting

model. Chemical transport models (CTM) aim to construct a mathematical model that describes chemical and meteorological processes in the atmosphere, focusing on the emission, transport, and mixing of air pollutant concentrations [20]. The enormous number of computations required by physical models is one of their key restrictions, and the quality of these deterministic techniques is dependent on a vast quantity of information and data from pollution sources [21]. Compared with the statistical model, the physical model is easy to use and simple to calculate [22].

Figure 4 represents the classifications of different air pollution forecasting models. Traditional statistical models for air pollution forecasting include the autoregressive integrated moving average (ARIMA) [20], and the grey model (GM) [23]. The statistical models work by describing the link between variables based on feasibility and statistical average, and forecasting accuracy is tough to achieve. To perform effective $PM_{2.5}$ forecasting, Auto-Regressive Integrated Moving Average (ARIMA) model was previously utilized [22,24]. Regional characteristics in statistical models can be complicated, chaotic, and extremely nonlinear [25] and nonlinear fitting is one of the drawbacks of these models [22]. According to several academic studies, AI models outperform physical forecasting methods and statistical models [26]. To tackle complicated issues, AI replicates human vision, learning, and reasoning. Artificial Intelligence includes machine learning as a subset. Machine learning performs well in classification and regression series, and it is widely regarded as one of the best methods in pollutant forecasting because of its great resilience and accuracy. An artificial neural network is a regularly used predictor that models nonlinear series by simulating the human brain and nervous system [3]. For environmental pollution prediction, generally used AI techniques such as Back Propagation Neural Network (BPNN), Extreme Learning Machine (ELM), Long Short-Term Memory (LSTM), the generalized regression neural network (GRNN), Gated Recurrent Unit (GRU), the wavelet neural network (WNN) and other models are Fuzzy logic model and Support Vector Machine (SVM). Furthermore, deep learning models based on forecasting algorithms are capable of performing functions with numerous layers. Recently, considerable work has been expended in reviewing and researching the properties of various types of intelligent prediction models in the field of air quality. Because of these constraints, it was commonly assumed that no one predictor could be competent for all aspects of modelling and that there was no single intelligent technique suitable for all individual difficulties.

In general, hybrid models relate to the combination of not just different algorithms or techniques, but also the advantages of each component, resulting in increased performance [27]. The different Hybrid AI models for pollutant forecasting are represented in Table 1.

**Table 1.** Various Hybrid AI models for Environmental pollution forecasting.

| S. No | Hybrid AI Models | Pollutant Forecasting | Ref. |
|---|---|---|---|
| 1 | Ensemble empirical mode decomposition—Least squares support vector machine (EEMD-LSSVM) | $PM_{2.5}$ | [28] |
| 2 | Principal component analysis- -Cuckoo search—Least squares support vector machine (PCA-CS-LSSVM) | $PM_{2.5}$ | [29] |
| 3 | Wavelet packet decomposition—Particle swarm optimization—Backpropagation neural network—Adaptive Boosting (WPD-PSO-BNN-Adaboost) | $PM_{2.5}$ | [30] |
| 4 | Particle swarm optimization- -Extreme learning machine (PSO-ELM) | $CO_2$ | [31] |
| 5 | Genetic algorithm—Random forest- Backpropagation neural network (GA-RF-BPNN) | $PM_{10}$ | [32] |
| 6 | Complementary empirical mode decomposition—Particle swarm optimization and gravitational search algorithm—Support vector regression—Generalized regression neural network (CEMD-PSOGSA-SVR-GRNN) | $PM_{2.5}$ | [33] |

Table 1. *Cont.*

| S. No | Hybrid AI Models | Pollutant Forecasting | Ref. |
|---|---|---|---|
| 7 | Wavelet packet decomposition—complete ensemble empirical mode decomposition with adaptive noise—Least squares support vector regression—chaotic particle swarm optimization method and gravitation search algorithm (WPD-CEEMD-LSSVR-CPSWOM-GSA) | $PM_{25}$ | [34] |
| 9 | Variational mode decomposition—Sample entropy—Least squares support vector machine (VMD-SE-LSSVM) | AQI | [35] |
| 10 | Complementary empirical ensemble mode decomposition—Cuckoo search—Grey wolf optimizer- support vector machine (CEEMD-CS-GWO-SVM) | $NO_2$& $SO_2$ | [36] |
| 11 | Wavelet packet decomposition (WPD)—Bidirectional Long Short-Term Memory (Bi-LSTM)—Stacked auto encoder Non-dominated Sorting Genetic Algorithm II (NSGA-II). | $PM_{25}$ | [37] |

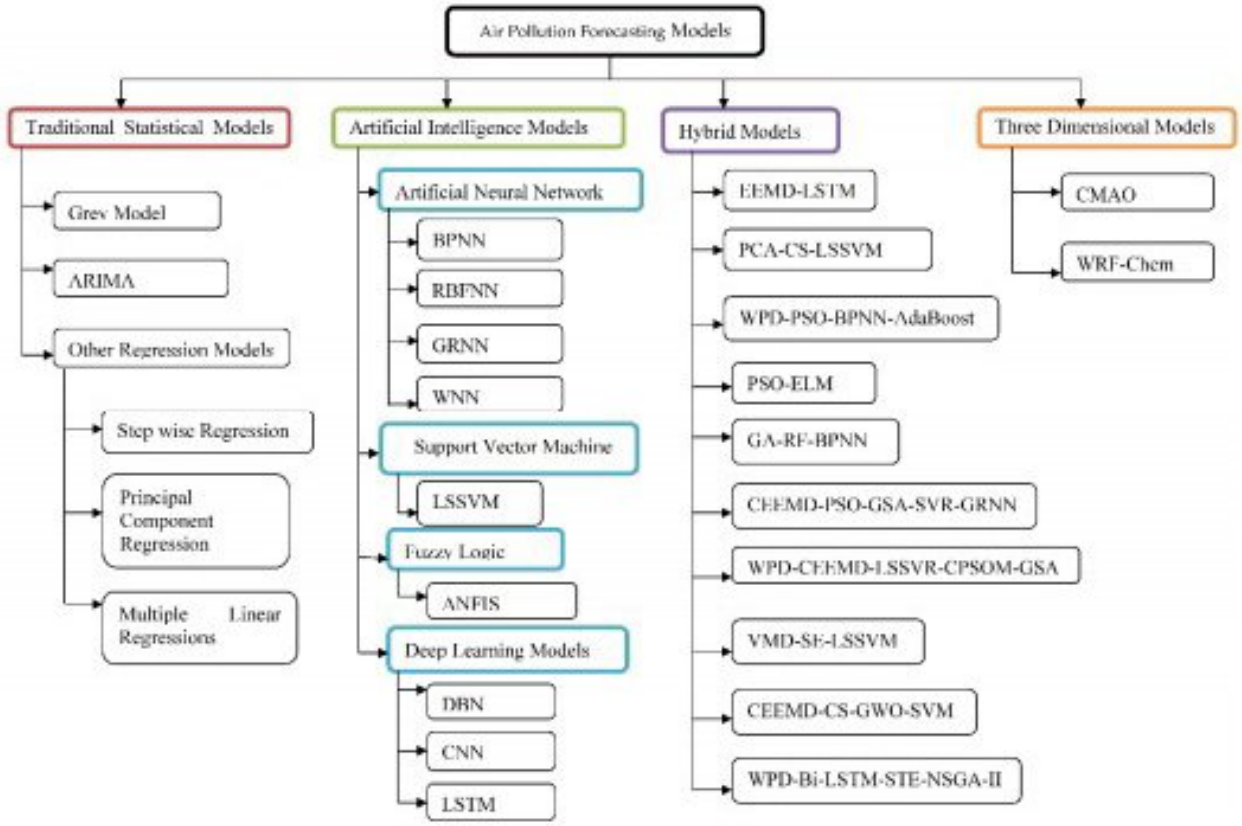


**Figure 4.** Different Methods of Air Pollution Forecasting.

Artificial intelligence (AI) has recently emerged as the most widely used technology tool for regulating and mitigating the detrimental effects of various air pollutants, generating considerable interest in the disciplines of atmospheric and medical science [38]. Many researchers have utilized AI approaches as a better healthcare decision support device to detect, manage, and cure illnesses caused by air pollution [19]. AI is one such instrument, with tremendous potential to expedite climate change mitigation and adaptation methods in sectors such as energy, land use, and disaster response; however, several bottlenecks and hurdles now exist that prevent AI from reaching its full potential in this area [39].

### 1.3. The Main Contributions of the Study

This narrative review study gives insights into the present status of Artificial Intelligence use in air quality monitoring systems and air quality forecasts that have been produced to date (January 2015–June 2022). The article's major goal is to look into various artificial intelligence technologies that are utilized for air quality monitoring, early warning systems and air quality prediction; this study focuses in particular on the most extensively utilized AI and ML methods in air pollution forecasting. Furthermore, numerous hybrid models are being emphasized because of their superior performance over single predictors. Artificial intelligence may also be employed in digital healthcare and chronic airway disease diagnosis. Overall, the goal of this project is to examine research on air pollution that use artificial intelligence and machine learning methods and get a general knowledge of relevant techniques based on performance criteria. The present study also emphasized various pollutants and its forecasting techniques. In addition, the role of machine learning and artificial intelligence in diagnosing pulmonary illnesses and COVID-19 induced by air pollution is discussed. The present literature is examined in light of several research concerns and AI-based technologies. In addition, the use of AI approaches to anticipate climate change and heat waves is also discussed.

## 2. Methods

The paper follows the format of a classic narrative review. Although this is not a full systematic review, the aspects of identifying literature sources and search methodologies, as well as study selection, are taken into account; however, the search results and references were disclosed, so that the reader may make an informed decision. Because this topic is still in its early stages, the authors included abstracts as part of the discussion in this narrative review.

For identifying papers from well-known research journals and publications, a literature search was undertaken using academic and science databases (e.g., Google Scholar, PubMed, Science Direct, Springer, IEEE explore, Scopus, and Web of Science). The keywords that were chosen for these literary analyses were: "Artificial Intelligence", "Machine learning", "Fuzzy Logic", "Air Pollution", "Artificial Intelligence in air pollution forecasting", Artificial Intelligence in Air quality monitoring", Artificial Intelligence and Human health", Machine Learning in digital health care", "Environment", "Climate change", Global warming", "Air quality monitoring", Air pollution Prediction", "Support vector machine", "Environmental health", "Artificial neural network", "Deep learning", "random forest", "generalized additive models", "gradient boosting machines", "xgboost", "decision trees", "Heat waves forecasting", "causes of climate change and global warming", "Hybrid AI models for air quality forecasting", "Application of artificial intelligence and machine learning", "Gaseous pollutants", "Ozone concentration forecasting using AI techniques", "$NO_2$ and $SO_2$ forecasting using AI techniques, "AI-based Particulate matters forecasting", "Early warning systems", "Diagnosis of Chronic air way diseases by AI and ML methods", Greenhouse Gas emission", "Causes of air pollutants on environment and human "and "Hybrid and Novel AI models for air pollution forecasting.

The following peer-reviewed international journals were used to choose the papers for this review paper: Environmental Science and Pollution Research, Atmospheric Environment, Atmospheric Research, Environmental Monitoring and Assessment, Journal of Hydrology, Environmental Geochemistry and Health, Atmospheric Pollution Research, Buildings and Environment, Buildings, Journal of forecasting, Health and technology, Journal of Cleaner Production, Engineering Applications of Artificial Intelligence, Environmental Modelling and Assessment, Environmental pollution, International Journal of Environmental Research and Public health, Journal of Environmental Engineering and Science, Journal of Environmental Management, Journal of Environmental Protection and Ecology, Computers in Industry, Applied Soft Computing, Computational Intelligence and Neuroscience, Diagnostics, Environmental Science and Policy, Lung Cancer, International Journal of Medical Sciences, Earth Science Informatics, Journal of Big data, Journal of

Ambient Intelligence and Humanized Computing, Journal of Medical Internet Research, Journal of Marine Systems, Sustainability, Aerosol and Air Quality Research, Environmental Modeling and Software, Remote Sensing of Environment, Science of Total Environment and Atmospheric Environment.

The discovery of relevant publications' techniques was repeated until the citation trails were no longer apparent. In addition, the reference lists of the selected research articles were searched in order to find additional references. Each relevant article's title, abstract, keywords, and conclusion were carefully examined to verify that they aligned with the literature review's goal. Second, the selected articles' content was analyzed to assess their applicability. The review period was set to run from January 2015 through June 2022. The review includes only papers written in English; this review covered only research articles dealing with Artificial Intelligence and Air Pollution Forecasting, Artificial Intelligence and Human Health, Climate Change, and Artificial Intelligence; this review research only included papers published in peer-reviewed journals; all other publications, such as conference papers and reports, were eliminated.

## 3. Potential Application of AI for Air Pollution Control

Artificial intelligence had a tough ride during the twentieth century; however, the growth of the internet in the twenty-first century is allowing AI to reach its full potential; this potential can only be restricted by the unknown limit of AI itself. The rise of artificial intelligence is one of the most spectacular advances in the scientific world, attracting the interest of practically all disciplines of study (AI). Artificial intelligence technology might be utilized to monitor air quality and regulate pollution. A great number of studies have found that AI technology can help with environmental pollution control. The use of AI technology to model complicated environmental problems is becoming more prevalent in the environmental industry, particularly in air quality regulation. Artificial intelligence is one of the most sophisticated technologies, and it is employed in practically every sector. As a result, using artificial intelligence in environmental monitoring is the best option, and its anticipated value is the most accurate [40]. Instead of simple machine learning and artificial intelligence approaches, most researchers and writers now use advanced and complex algorithms for early warning systems, air quality prediction and environmental quality monitoring. Alimissis et al. [41] investigated two ways for modelling urban air pollution using interpolation; it gives a great capacity for tracking the present pollution scenario and demonstrates a precise and quick method for identifying pollution hotspot areas [19].

### *Various AI Techniques for Air Quality Forecasting and Monitoring System*

An air quality monitoring system lays the framework for air quality forecasting and traceability, and more accurate air quality forecasting is made possible by improved monitoring data. Air quality forecasting is an effective approach to protecting human health by providing advanced warnings of harmful air pollution [42]. Accurate forecasting of air quality is extremely important for the environment and persons dealing with real-world air pollution. Cities and people can respond to air pollution warnings in before by decreasing traffic, restricting outdoor activities and shutting down industries [43]. Many countries have pollution early warning systems. As a result, all governments should emphasize air pollution forecasting as a foundation for pollution warning and control systems [15].

Numerous forecasting models, mostly for pollution concentrations, have been proposed [38]. Based on their concepts, forecasting models may be divided into three types: statistical forecasting models, numerical forecasting models, and machine learning models [38]. Because of their simplicity, statistical models might be employed in air quality forecasting; they can anticipate pollutant concentrations in the future only by evaluating the link between pollutant concentration and climatic parameters in previous data, without knowing the sources of pollution [38]. The Markov model, grey model (GM), autoregressive integrated moving average model (ARIMA) [44], and multiple linear regression model

(MLR) are examples of general statistical models [38]. Artificial intelligence includes machine learning. Because of its high resilience and error tolerance, machine learning works well in regression and classification problems, and it is commonly recognized as one of the most effective approaches to pollutant forecasting. A combination of AI algorithms was allegedly used to anticipate air pollutants and develop air quality monitoring and early-warning systems [45]. Many AI algorithms have been proposed and used to forecast air quality. To demonstrate AI techniques for predicting various levels of air pollution, vast amounts of historical pollutant data and temporal factorsare required [15]. AI-based estimates of air quality and pollution are gaining momentum as a result of climate change and urbanization. According to a recent study, AI-augmented models are superior for univariate time series data forecasting.

Mauro et al. utilized the ARIMA statistical model for air quality forecasting. ARIMA was compared to ML algorithms such as SVR and ANN. The results showed a significant improvement in prediction quality [46]. ANN has been demonstrated to be especially successful for increasingly complicated jobs. ANN models also use a complex algorithm that has been successfully used to predict air pollution [47]. Elangasinghe et al. [48], established anartificial neural network air pollutant prediction model capable of fully capturing the time fluctuation of pollutant concentrations under defined conditions by extracting key information from daily accessible meteorological parameters. Pardo and Malpica [49] utilized a double-layered LSTM neural network to predict the Madrid air quality model. Deeper LSTM networks enhance prediction accuracy while increasing computational cost and time [50]. Song et al. [51] proposed a hybrid model based on LSTM and Kalman filtering to estimate the concentration of various air quality components [52]. Qi et al. [53], propose a hybrid model for $PM_{2.5}$ air quality forecasting that incorporates graph convolutional networks and LSTM (GC-LSTM). The time-space mapping information includes historical meteorological parameters, geographical characteristics, and other time-series properties [54].

A spatiotemporal convolution-long short-term memory neural network extension (C-LSTME) model was introduced by Wen et al. [19,55]. Furthermore, climatic and particulate data are used to enhance forecast outcomes [56]; however, both weather and temporal features model frameworks need a massive series of data records, considerably increasing the suggested technique's time complexity. Zeng et al. [57] developed a novel predicting model for the concentration of $PM_{2.5}$ that combines the nested LSTM (NLSTM) neural network and the extended stationary wavelet transform (ESWT) [20,50]. To improve prediction performance, the suggested technique blends deep learning technology, namely layered LSTM neural networks, with a modified SWT based on the artificial intelligence method SWT [50]. In Turkey, Güler Dincer et al. developed a unique Fuzzy K-Medoid clustering approach to predict $SO_2$ concentration [15,58]. Random forest regressor (RFR), decision tree regressor (DTR), and Linear regression (LR) approaches are utilised in another investigation to predict air pollution [59]. The results indicated that the RFR outperformed the LR and DTR on the given data set. Althuwaynee et al. assessed the air pollution hazard by using decision tree algorithms and evaluated the correlation clusters of $PM_{10}$ and other pollutants [60]. Shaziayani et al. [61], predicted the concentration of $PM_{10}$ for the next day by using a tree-based machine learning approach (the models have boosted regression trees, random forest and decision tree algorithms). The study of Wang and Kong developed air quality predictive modelling based on improved decision tree algorithms to enhance the prediction accuracy and time performance [62]. By using air mass trajectory analysis, the daily approximate $PM_{2.5}$ concentration prediction accuracy was improved by introducing ANN techniques by Feng et al. [21]. Yan et al. [63] employed GRNN to forecast $PM_{2.5}$ concentration values in three Chinese urban clusters. For precise forecasting of the $PM_{2.5}$ level in the urbanized area, the GRNN technique could be very useful and reliable.

Mo et al. created a revolutionary air quality early-warning system based on artificial intelligence and superior data preprocessing technologies [38]. The prediction model was built using the potent swarm intelligence method: The improved Complete

Ensemble Empirical Mode Decomposition with Adaptive Noise (ICEEMDAN) [20,57], Whale Optimization Algorithm (WOA) and the efficient ANN Extreme Learning Machine (ELM) [20,38,57]. The prediction findings were then further examined using the fuzzy comprehensive assessment approach, which provided intuitive air quality information and appropriate measures [38]. The simulation results demonstrated that the proposed ICEEMDAN-WOA-ELM model outperformed previous models and that the ICEEMDAN decomposition technique, in conjunction with the WOA optimization approach, played major roles in increasing neural network prediction accuracy [38]. Because of its accuracy and efficacy, this planned air quality early-warning system is expected to play an important role in the future [38]. To examine the present pollutant levels, Amuthadevi et al., employed Deep Learning Long-Short-Term Memory (DL-LSTM), Neuro-Fuzzy and Statistical Multilevel Regression such as Non-Linear Artificial Neural Network (ANN). The findings demonstrated that DL-LSTM outperforms ANN, Neuro-fuzzy, and regression methods [46]. Qi et al. [53], developed a hybrid model that employs a graph convolutional neural network to discover geographical correlations from surrounding air quality and meteorological data, as well as an RNN model to account for temporal dependency. Similarly, Du et al. [64], created a hybrid deep learning model for air pollution forecasting that includes both a convolution neural network and an RNN model for spatial-temporal modelling. Neither model, however, has generated high-resolution estimates for locations with limited air quality measurements. The data-driven Granger Causality technique is a more advanced data-driven approach for pollution assessment [65]; it achieves 82% accuracy by predicting air quality at high spatial-temporal resolution using data from four domains, including transportation, meteorology, pollutant exposure, and urban morphology [66].

Bekker et al. [67] used CNN-LSTM to predict the hourly prediction of $PM_{2.5}$ concentration in Beijing, China, using a spatial-temporal feature by incorporating meteorological data, historical pollutant data, and $PM_{2.5}$ concentration in neighbouring stations [68]; they investigated the performance differences across Deep Learning algorithms such as CNN, GRU, Bi-LSTM, Bi-GRU, LSTM and a hybrid CNN-LSTM model. Experiment findings show that their "hybrid CNN-LSTM multivariate" technique makes more accurate predictions than any of the classic models stated and outperforms them in predictive performance [67]. The hybrid AI techniques are more reliable and accurate in terms of developing early-warning systems and environmental pollution monitoring. Table 2 depicts several Artificial Intelligence models and Machine Learning technologies used in environmental pollution forecasting, as well as performance evaluation criteria. The authors determined the linear and nonlinear relationships between the air pollution index (API) and meteorological factors in the two Chinese cities of Xi'an and Lanzhou using correlation analysis and artificial neural networks (ANNs; incorporating wavelet ANNs [WANNs]). Both the WANN and ANN models successfully replicated the APIs in Xi'an and Lanzhou, although the WANN model outperformed the ANN during the forecasting phase ($R = 0.8037$ for Xi'an and $R = 0.7742$ for Lanzhou; $R = 0.8846$ and $R = 0.8906$ respectively). Wavelet-ANN, Wavelet-ARIMA, and Wavelet-SVM are three hybrid models that were created to predict the 2016 $PM_{2.5}$ trends in five Chinese cities (Beijing, Chengdu, Guangzhou, Shanghai and Taiyuan). The authors investigated the meteorological impacts and restriction during COVID-19 lockdown, and reduction in $NO_2$ and $PM_{2.5}$ using generalized additive models including meteorological parameters and multi-temporal variations was used to quantify the impact from each factor on $NO_2$ and $PM_{2.5}$ [68]. GAM helpful tool for estimating the air quality responses to human activities because they can balance fitting performance and have good interpretability [68]. Karl Ropkins and James Tete investigated the impacts of COVID-19 lockdown on air quality trends across UK by using break point-segment methods and the results showed that the NO and $NO_2$ decreased up to 32% to 50% near roadsides during lockdown whereas the $O_3$ concentration increased 20% during lock-down across UK; moreover, the $PM_{10}$ (5.9 to 6.3 $\mu g/m^3$) and $PM_{2.5}$(3.9 to 5.0 $\mu g/m^3$) concentration Increased at both urban and rural areas station [69]. Jasper et al. [70], explored global assessment of non-linear correlations between atmospheric processes and daily ground-level $NO_2$, $PM_{10}$,

$PM_{2.5}$, and $O_3$ observations were captured in city- and pollutant-specific XGBoost models for over 706 cities, while weather, seasonality, and trends were taken into account, this study specifically employed XGBoost, an eXtreme Gradient Boosting technique based on decision trees [70]. XGBoost is a gradient boosting machine adaption that does additive optimisation in functional space. The author mentioned limitation as smaller nations lacked the assessment criteria of XGBoost models for a multi-model ensemble forecast, resulting in bigger observed variances [71]. Pardo et al. conducted a case study at urban traffic sites in Spain to estimating changes in air pollution levels as a result of the COVID-19 lockdown by using Multiple Linear Regression (MLR) models with seasonal and meteorological predictors and the Q-Q Mapping post-correction enhances model performance and predicts severe condition [69]. Meaningful climate science requires collecting huge amounts of data on many different variables such as temperature and humidity, but working with such massive data sets is challenging across different geographical areas due to unavailability of datasets, climatic condition and seasonal variations. Due to its accuracy and predictive capability, the hybrid model, which is based on data decomposition, the ensemble approach, and spatial-temporal approaches, surpasses the others. In terms of pollution forecasts, the Hybrid models outperformed other models. Environmental epidemiology can benefit significantly from geoAI technologies' ability to incorporate large amounts of big spatial and temporal data in a variety of formats, as well as their computational effectiveness, flexibility in algorithms and workflows to account for important aspects of spatial (environmental) processes, such as spatial nonstationarity, and scalability to model other environmental exposures across different geographic areas. GeoAI techniques can be used to overcome issues with environmental exposure modelling, such as computational processing and time inefficiencies (especially when massive data are combined with extensive geographic research areas), as well as data-related limitations that affect spatial and temporal resolution.

Numerous research have demonstrated that traditional statistical methods are less accurate than artificial intelligence technology in terms of prediction accuracy. Artificial neural networks (ANN) can manage massive volumes of data sets, implicitly identify complicated nonlinear relationships between dependent and independent factors, and detect all possible interactions between predictor variables, among other major advantages. Traditional AI performance outperforms statistical approaches but falls short of the hybrid model. According to reports, many researchers employ an individual ANN technique or a hybrid method that includes ANN (e.g., adaptive network-based fuzzy inference system (ANFIS) and fuzzy neural network (FNN) since it is simple to design and have a high level of accuracy [57]. There are, however, many different machine learning techniques, and determining which one is optimal for the task at hand can be difficult. In other words, if additional factors are examined or a hybrid model is developed and used, the forecasting model will perform better [15]. Finally, the hybrid framework provides benefits in prediction stability, prediction accuracy, and air pollution warning accuracy. Although there is no perfect way to predict air quality, hybrid models beat single models; however, the intricacy of model building raises the computational cost. Because of the structure, hybrid models with complicated frameworks frequently require extra computational time [20].

**Table 2** Statistics of various pollutants and AQI in the *CPCB* dataset

| Pollutants → Statistics ↓ | $PM_{2.5}$ | $PM_{10}$ | NO | $NO_2$ | $NO_x$ | $NH_3$ | CO | $SO_2$ | $O_3$ | Benzene | Toluene |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Count | 24,933 | 18,391 | 25,949 | 25,946 | 25,346 | 19,203 | 27,472 | 25,677 | 25,509 | 23,908 | 21,490 |
| Mean | 67.450 | 118.127 | 17.574 | 28.560 | 32.309 | 23.483 | 2.248 | 14.531 | 34.491 | 3.280 | 8.700 |
| Std | 64.661 | 90.605 | 22.785 | 24.474 | 31.646 | 25.684 | 6.962 | 18.133 | 21.694 | 15.811 | 19.969 |
| Min | 0.040 | 0.010 | 0.020 | 0.010 | 0.078 | 0.010 | 0.253 | 0.010 | 0.010 | 0.063 | 0.238 |
| 25% | 28.820 | 56.255 | 5.630 | 11.750 | 12.820 | 8.580 | 0.510 | 5.670 | 18.860 | 0.120 | 0.600 |
| 50% | 48.570 | 95.680 | 9.890 | 21.690 | 23.520 | 15.850 | 0.890 | 9.160 | 30.840 | 1.070 | 2.970 |
| 75% | 80.590 | 149.745 | 19.950 | 37.620 | 40.127 | 30.020 | 1.450 | 15.220 | 45.570 | 3.080 | 9.150 |
| Max | 949.990 | 1000.102 | 390.680 | 362.210 | 467.630 | 352.890 | 175.818 | 193.860 | 257.730 | 455.03 | 454.85 |

| Pollutants → Statistics ↓ | Xylene | AQI |
|---|---|---|
| Count | 11,422 | 24,850 |
| Mean | 3.070 | 166.463 |
| Std | 6.323 | 140.696 |
| Min | 0.134 | 13.000 |
| 25% | 0.140 | 81.000 |
| 50% | 0.980 | 118.00 |
| 75% | 3.350 | 208.00 |
| Max | 170.370 | 2049.00 |

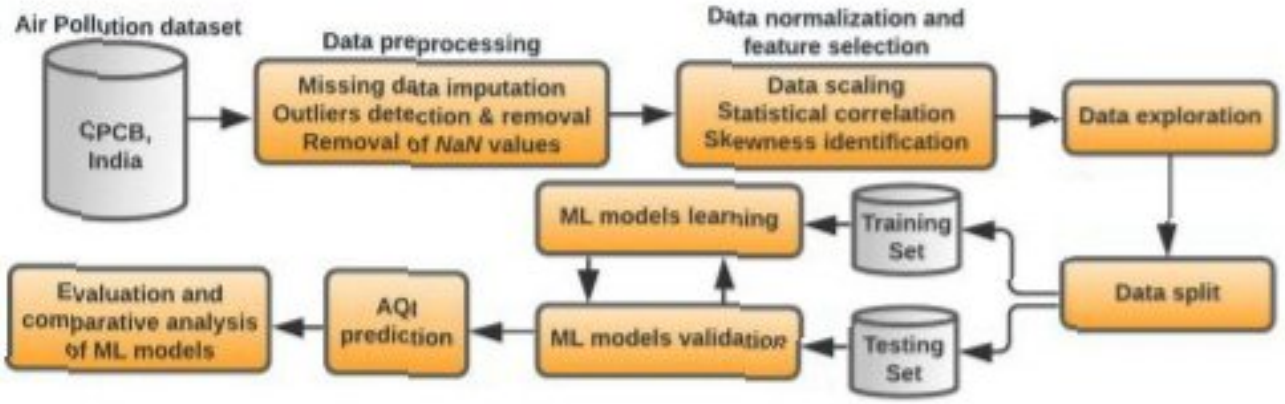


**Fig. 1** Flowchart of the proposed model

speed and generalization capability of ML algorithms. Outliers and missing data are the two most common errors in data extraction and monitoring applications. The data preprocessing step performs various operations on data such as filling out *not-a-number (NAN)* data, removing or changing outlier data, etc. Figure 2 shown below presents a view of the missing values in each feature of the dataset. Observe that among all other features, *Xylene* has the most missing values and *CO* has the least missing values. A large number of missing values may be existing due to a variety of factors, such as a station that can sense data but does not possess a device to record it.

All the missing values are filled with the median values against each feature to solve the missing data problem. Next, a normalisation process has been applied to standardize the data, ensuring that the significance of variables is unaffected by their ranges or units. The data normalisation process helps to bring different data attributes into a similar scale of measurement. This process plays a vital role in the stable training of ML models and boosts performance. The datatypes of all the variables are also examined during normalisation. For example, the dataset is collected from different monitoring stations which deal with different representations of dates. Thus, the date 'Monday, May 17, 2021'

| | Missing Values | % of Total Values |
|---|---|---|
| Xylene | 18109 | 61.300000 |
| PM10 | 11140 | 37.700000 |
| NH3 | 10328 | 35.000000 |
| Toluene | 8041 | 27.200000 |
| Benzene | 5623 | 19.000000 |
| AQI | 4681 | 15.900000 |
| AQI_Bucket | 4681 | 15.900000 |
| PM2.5 | 4598 | 15.600000 |
| NOx | 4185 | 14.200000 |
| O3 | 4022 | 13.600000 |
| SO2 | 3854 | 13.100000 |
| NO2 | 3585 | 12.100000 |
| NO | 3582 | 12.100000 |
| CO | 2059 | 7.000000 |

**Fig. 2** Missing values of the features and their percentages

may be represented as '17/5/2021' or as '17–05-2021' etc. Such date feature has been normalised through the *datetime* Python library.

## Feature selection

The *CPCB* dataset under study involves a specific parameter viz, AQI and government agencies use this parameter to alert people about the quality of the air and also practice forecasting it. According to the *National Ambient Air Quality Standards*, there are six AQI categories: good (0–50), satisfactory (51–100), moderate (101–200), poor (201–300), very poor (301–400), and severe (401–500). Scholars in the realm suggest that reducing input variables lowers the computational cost of modeling and enhances prediction performance. A correlation-based feature selection method has been exploited in the present work to determine the optimal number of input variables (pollutants) when developing a predictive model. Statistical correlation-based feature selection algorithms compute correlations between every pair of the input variable and the target variable. The variables possessing the strongest correlation with the target variable are then filtered for further study. Since many ML algorithms are sensitive to outliers, any feature in the input dataset which does not follow the general trend of that data must be found. For the present dataset, a correlation-based statistical outliers detection method has been applied to identify the outliers. To select significant features, the correlation analysis of the AQI feature has been exercised with features of other pollutants. Figure 3, shown below clearly reveals that pollutants $PM_{10}$, $PM_{2.5}$, CO, $NO_2$, $SO_2$, $NO_X$, and NO are generally responsible for the AQI to attain higher values.


**Fig. 3** Correlation heatmap of AQI with other pollutants (Threshold: 0.4)

**Table 3** Correlation between AQI and pollutants

| S. No | Features | Correlation value | S. No | Features | Correlation value |
|---|---|---|---|---|---|
| 1 | $PM_{10}$ | 0.80331 | 7 | NO | 0.452191 |
| 2 | CO | 0.68334 | 8 | Toluene | 0.279992 |
| 3 | $PM_{2.5}$ | 0.65918 | 9 | $NH_3$ | 0.252019 |
| 4 | $NO_2$ | 0.53707 | 10 | $O_3$ | 0.198991 |
| 5 | $SO_2$ | 0.52586 | 11 | Xylene | 0.165532 |
| 6 | NOx | 0.486450 | 12 | Benzene | 0.044407 |

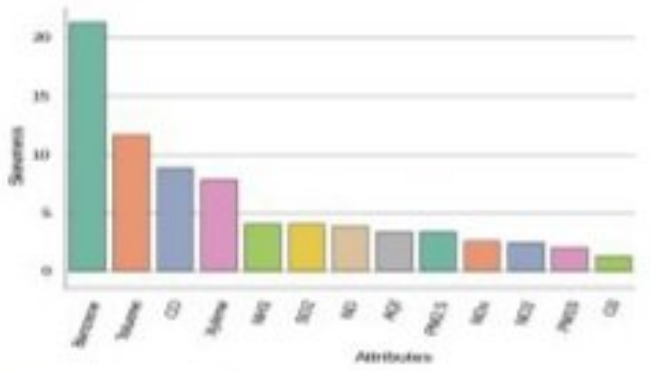


**Fig. 4** Skewness present in dataset features

These pollutants are correlated with AQI based on the correlation values above the threshold of 0.4.

Table 3 given below shows the exact correlation values of each pollutant of the dataset with AQI.

Many ML models function better when data have a normal distribution and underperform when data have a skewed distribution. Therefore, it is necessary to identify the skewness being present in the features and to perform

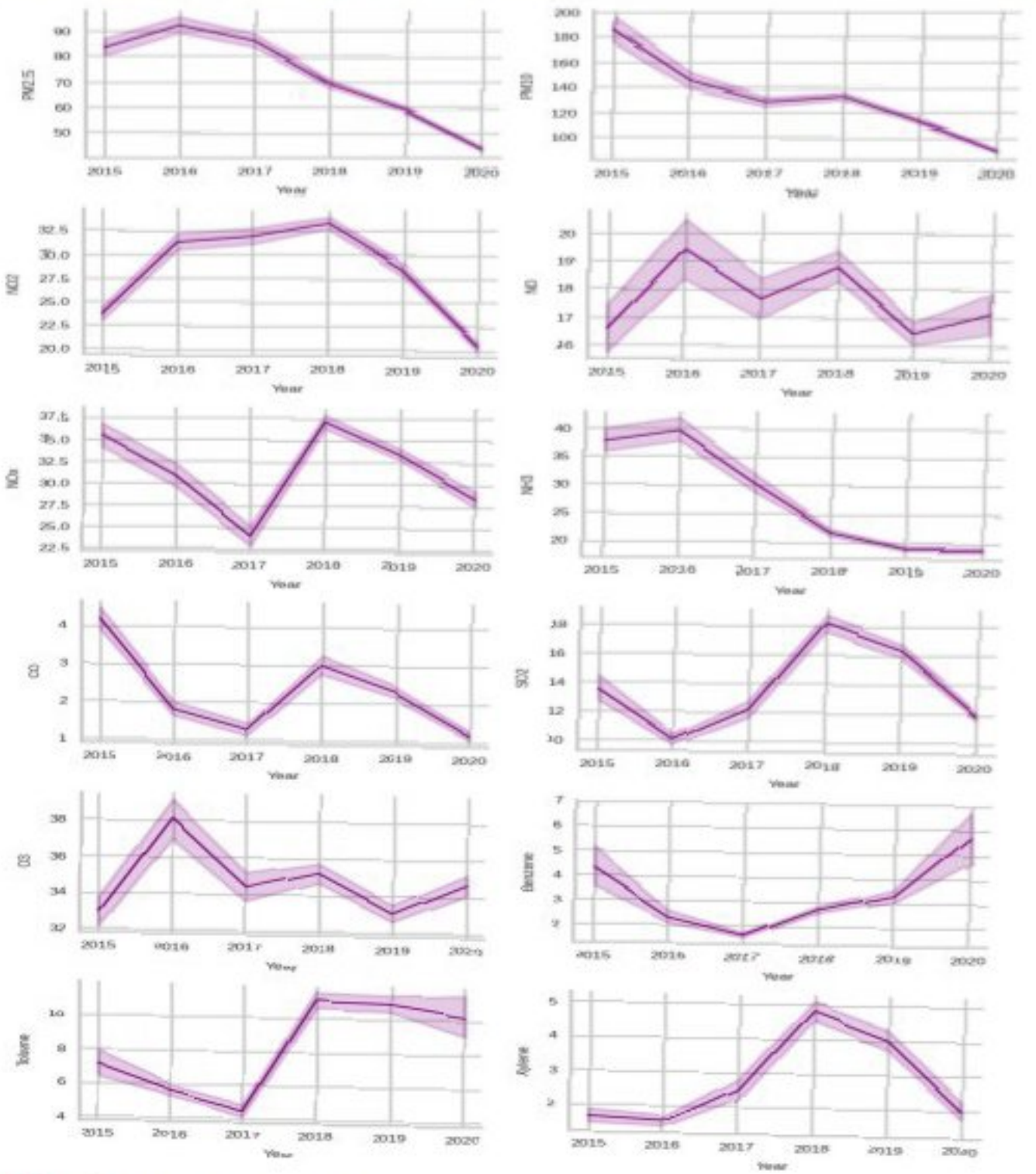


**Fig. 5** Intensities of various pollutants from 2015 to 2020

attribute, *AQI_Bucket* has some missing values which result in the unequal splitting of the classes. Many ML models ignore this imbalanced datasets problem which may lead to poor classification and prediction performances. To overcome this data imbalance problem, the *SMOTE (Synthetic Minority Oversampling Technique)*

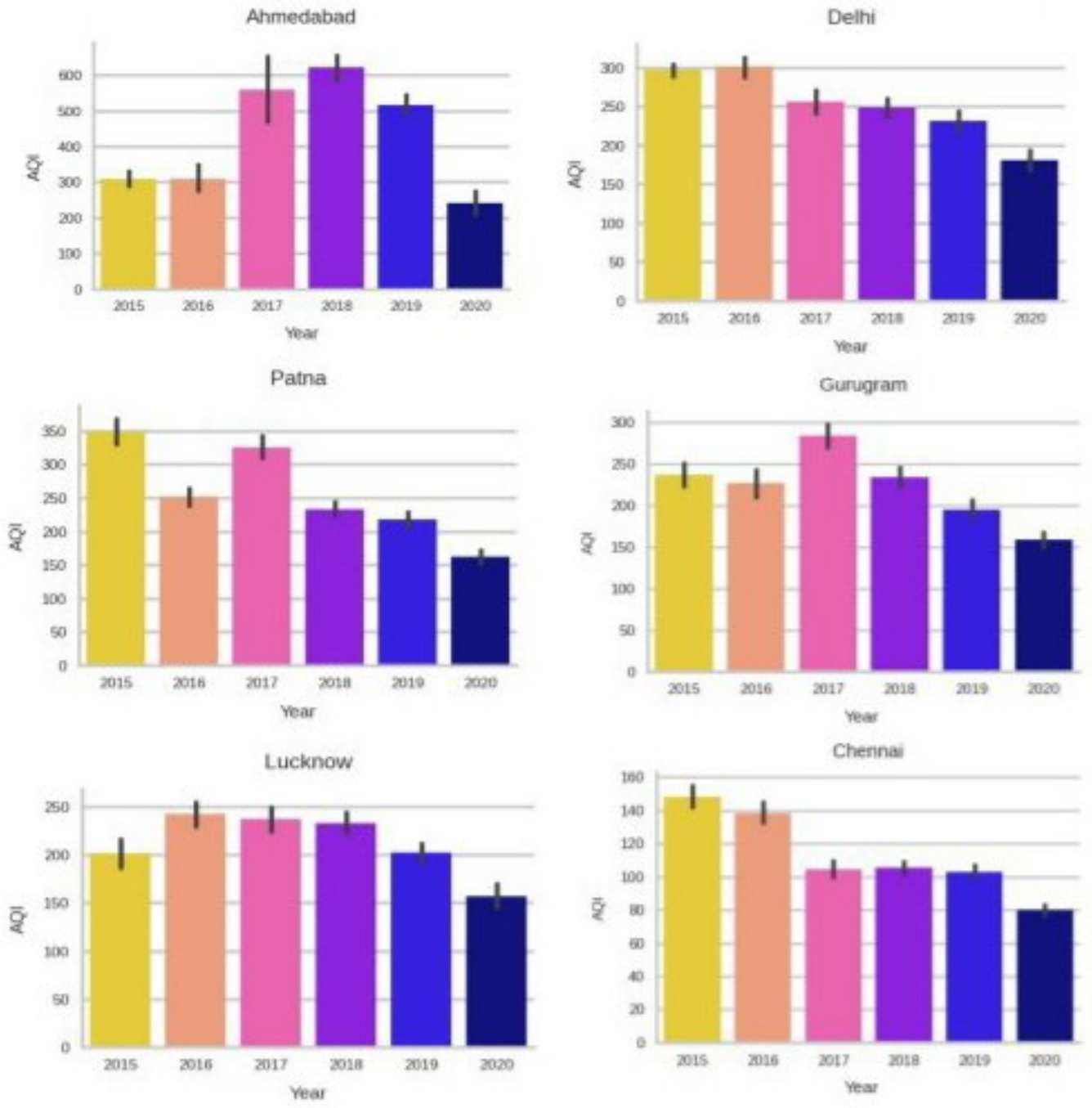


**Fig. 6** The six most polluted Indian cities with their average AQI values from 2015 to 2020

has been applied. In this technique, the algorithm synthesizes new elements for minority classes rather than creating copies of already existing elements. It functions by randomly choosing a point from the minority class and computing the *k-nearest neighbor* distances for the selected point. The newly created synthetic points are added between the chosen point and its neighbors. To implement *SMOTE* for class imbalance, we have used an imbalanced-learn Python library in the *SMOTE* class. Now, five popular ML models, *KNN*, *Gaussian*

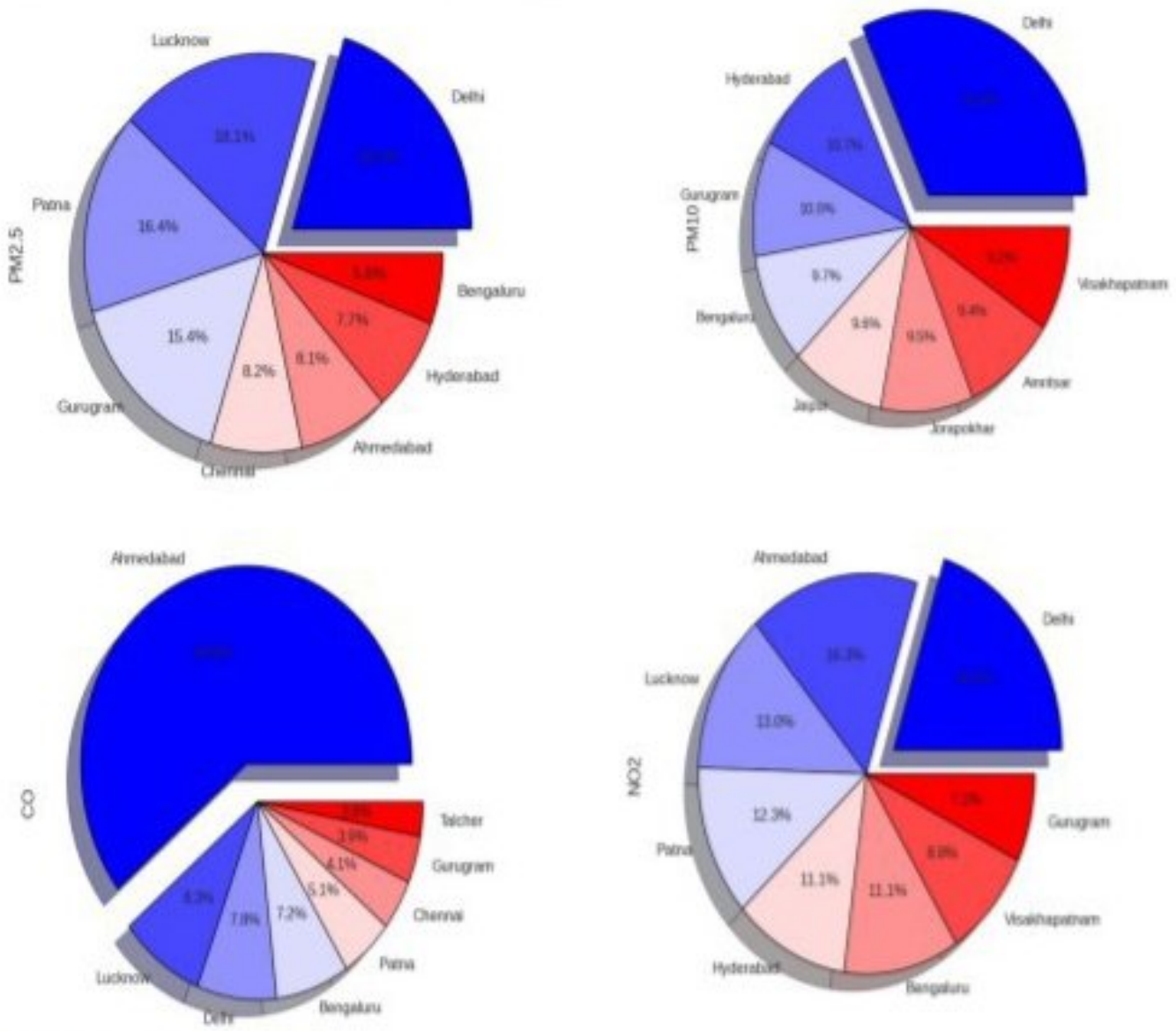


**Fig. 7** Pollutants governing AQI directly

*Naive Bayes (GNB)*, *SVM*, *RF*, and *XGBoost* have been employed to predict the AQI level with *SMOTE* and without *SMOTE* resampling technique. Table 4 shown below presents the results of used ML models in terms of accuracy, precision, recall, and F1-score during the training phase. Precision tells the fraction of relevant instances present in the retrieved instances, while recall is the fraction of relevant instances that have been retrieved. Accuracy is the ratio of the correctly labeled attributes to the whole pool of variables. F1-score is a weighted average of precision and recall. Note that the *XGBoost* model achieved the highest accuracy, while the *SVM* model exhibited the lowest accuracy.

The performances of the ML models for the training set are evaluated against the standard performance parameters, viz *MAE*, *RMSE*, *Root Mean Squared Logarithmic Error (RMSLE)*, and coefficient of determination, i.e. $R^2$ (Table 5). These performance measures have been exploited

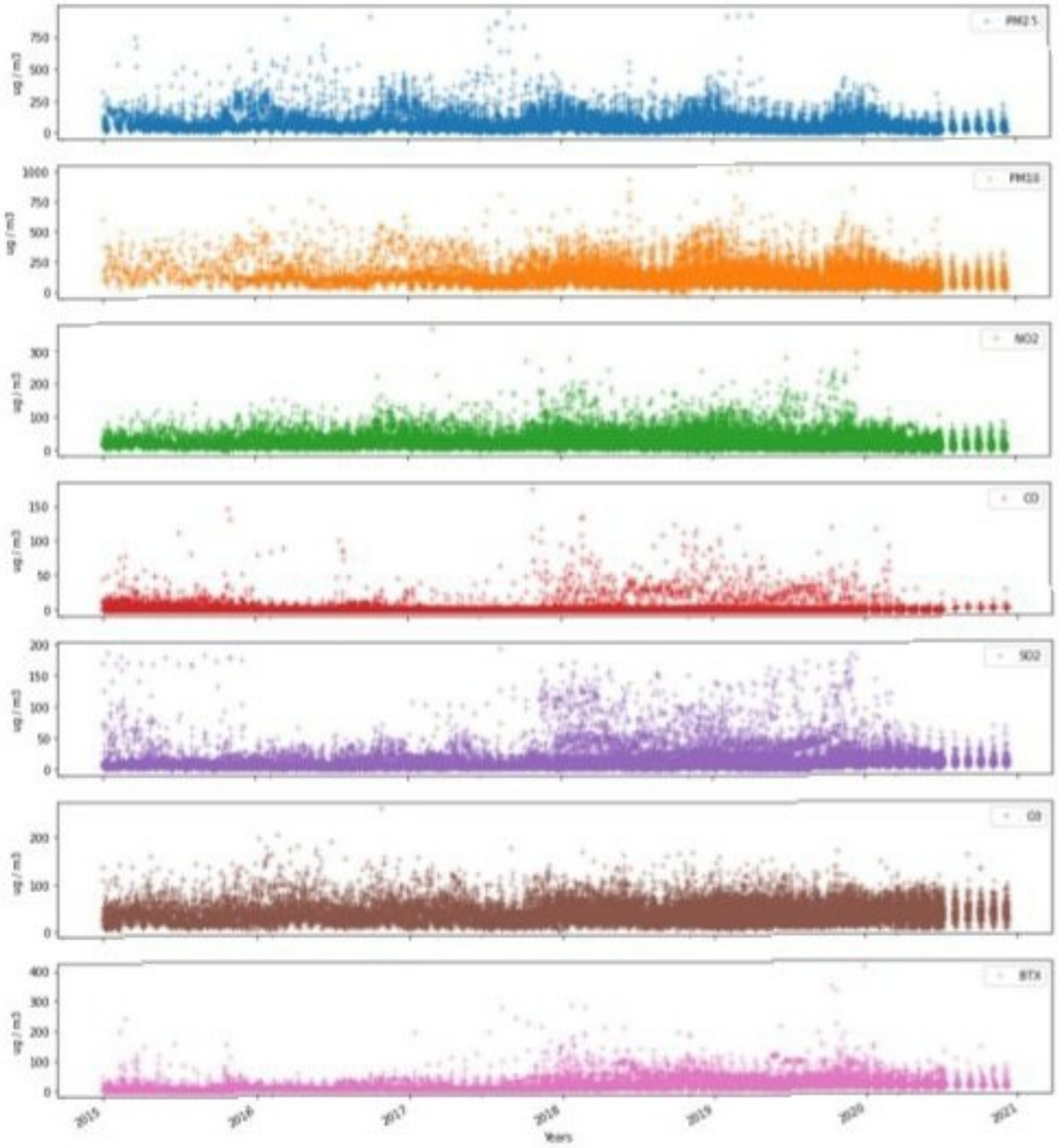


**Fig. 8** Timeline graph of AQI with respect to specific pollutants

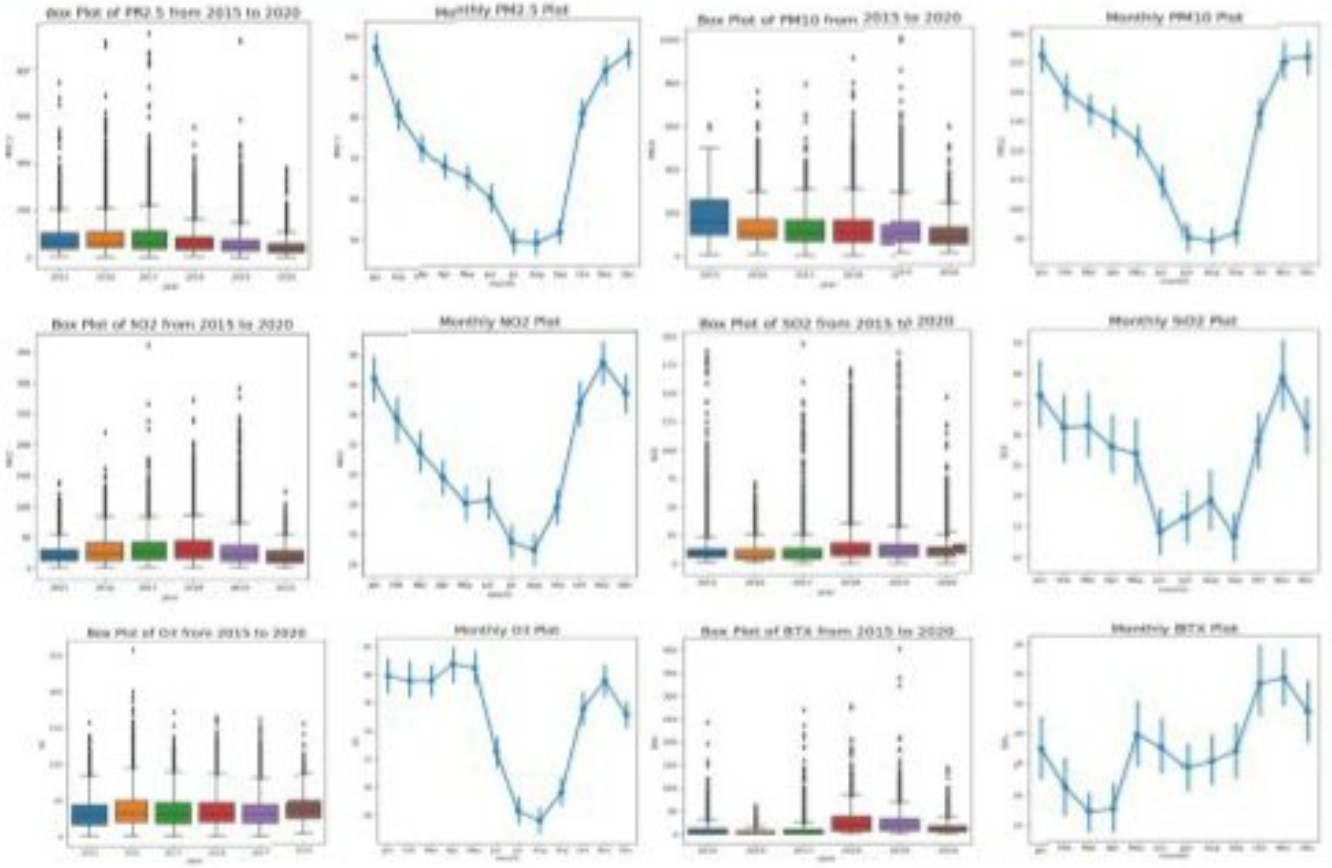


**Fig. 9** Variation analysis of pollutants through *Box plots*

**Table 4** Comparison of model results in the training set

| Model | Accuracy | Precision | Recall | F1-score | Training time (in seconds) |
|---|---|---|---|---|---|
| KNN | 89 | 94 | 90 | 96 | 0.104 |
| GNB | 85 | 91 | 94 | 88 | 0.110 |
| SVM | **81** | 90 | 93 | 88 | 0.258 |
| RF | 88 | 93 | 88 | 92 | 0.102 |
| XGBoost | **91** | 95 | 95 | 91 | 0.532 |

**Table 6** Comparison of model results in the testing set

| Model | Accuracy | Precision | Recall | F1-Score | Prediction time (in seconds) |
|---|---|---|---|---|---|
| KNN | 85 | 92 | 85 | 94 | 0.018 |
| GNB | 83 | 88 | 89 | 92 | 0.016 |
| SVM | **78** | 91 | 90 | 83 | 0.027 |
| RF | 86 | 92 | 91 | 90 | 0.023 |
| XGBoost | 90 | 96 | 95 | 91 | 0.041 |

**Table 5** Results of ML algorithms for AQI Prediction with and without SMOTE (training set)

| Models | Without SMOTE | | | | With SMOTE | | | |
|---|---|---|---|---|---|---|---|---|
| | MAE | RMSE | RMSLE | $R^2$ | MAE | RMSE | RMSLE | $R^2$ |
| KNN | 0.627 | 3.834 | 0.153 | 0.913 | 0.023 | 1.003 | 0.063 | 0.864 |
| GNB | 0.622 | 2.454 | 0.164 | 0.856 | 0.027 | 1.212 | 0.045 | 0.801 |
| SVM | 0.537 | 2.238 | 0.078 | 0.820 | 0.026 | 1.003 | 0.043 | 0.772 |
| RF | 0.331 | 1.973 | 0.082 | 0.643 | 0.022 | **0.863** | **0.023** | 0.583 |
| XGBoost | **0.298** | **1.465** | **0.045** | **0.612** | **0.012** | 0.963 | 0.062 | **0.545** |

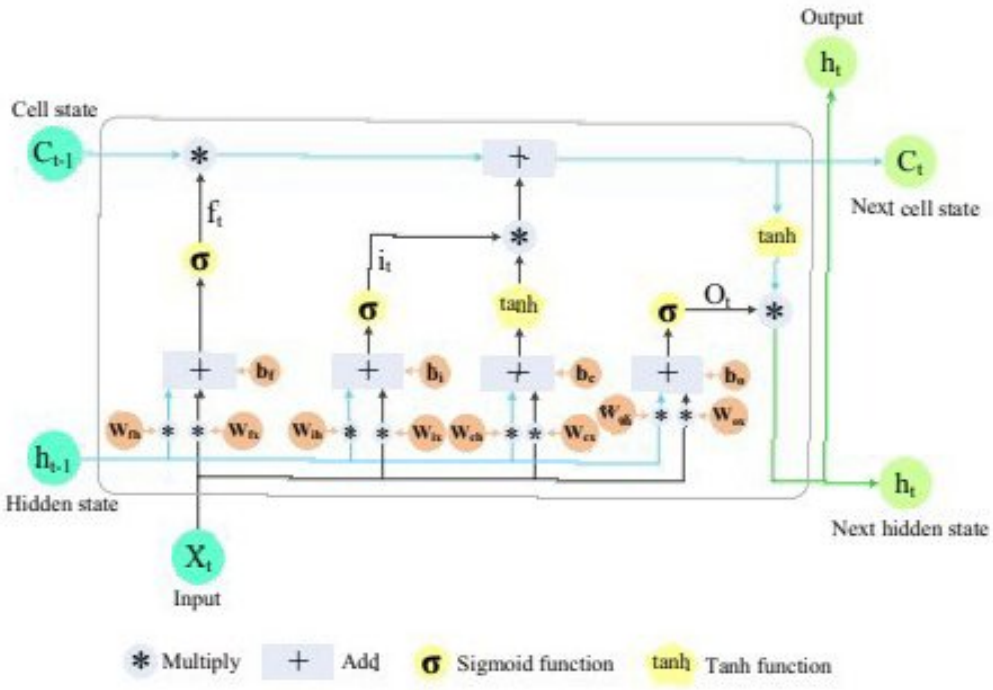


**Figure 2.** LSTM structure.

feature extraction of time series data[41–43]. *Therefore, CNN* can effectively extract features from non-linear and unstable air quality data.

**ILSTM.** *Model structure.* LSTM has excellent advantages in mining long-term dependence relationships of sequence data. Figure 2 shows three types of gates: forget gate, input gate, and output gate, respectively. A gate can be regarded as a full connection layer, and LSTM stores and updates information by these gates[44,45]. Gated Recurrent Unit (GRU) has only two gates. GRU combines the input gate and forget gate in LSTM into one, which is called the update gate[46], as shown in Fig. 3. Based on the gated technology, the ILSTM model proposed in this paper consists of input gate and forget gate, as shown in Fig. 4.

Compared with LSTM, ILSTM deletes the output gate. Compared with GRU, ILSTM structure is simpler. The parameters of LSTM, GRU, and ILSTM are shown in Table 1. Compared with LSTM, ILSTM reduces weight parameters from 8 to 4 and bias parameters from 4 to 2. Compared with GRU, ILSTM reduces weight parameters from 6 to 4 and bias parameters from 3 to 2.

In terms of algorithm, ILSTM adds the cell state $c_{t-1}$ of the previous moment to the algorithm of the forget gate to generate the mainline forgetting $k_t$, which affects the data retention degree at the current time. In addition, when updating the cell state $c_t$ of the current moment, the CIM is introduced to prevent supersaturation in the learning process.

The forget gate of the ILSTM $f_t$ is a crucial component of ILSTM unit, which can control what information should be retained and what information should be forgotten. $\sigma(x)$ is a Sigmoid function, as shown in formula (1). $x_t$ is the input data of the t-th time step. $h_{t-1}$ is the hidden layer of the previous time step $t-1$. $W_{fh}$ is the weight of $h_{t-1}$ of forget gate, and $W_{fx}$ is the weight of $x_t$. $b_f$ is the bias of forget gate, as shown in formula (2).

$$\sigma(x) = \frac{1}{1+e^{-x}}, \tag{1}$$

$$f_t = \sigma(W_{fh} \cdot h_{t-1} + W_{fx} \cdot x_t + b_f). \tag{2}$$

Mainline forgetting $k_t$ is calculated by cell state $c_{t-1}$ and $f_t$. Mainline forgetting represents the influence of information on current cell state $c_t$, where $c_{t-1}$ is cell state information from the beginning to the previous moment, as shown in formula (3).

$$k_t = f_t \times c_{t-1}. \tag{3}$$

The input gate $i_t$ controls how much of the current input data $x_t$ flows into the memory cell, that is, how much can be saved to $c_t$. Compared with the input gate of LSTM, ILSTM adds $c_{t-1}$ to the input gate algorithm, that is, the cell state information up to the previous moment. The introduction of $c_{t-1}$ makes the input gate of the model have a memory effect on the retention of data at the current time, as shown in formula (4). $W_{ih}$ and $W_{ix}$ are the weights of the input gate's $h_{t-1}$ and $x_t$, respectively, and $b_i$ is the bias of the input gate.

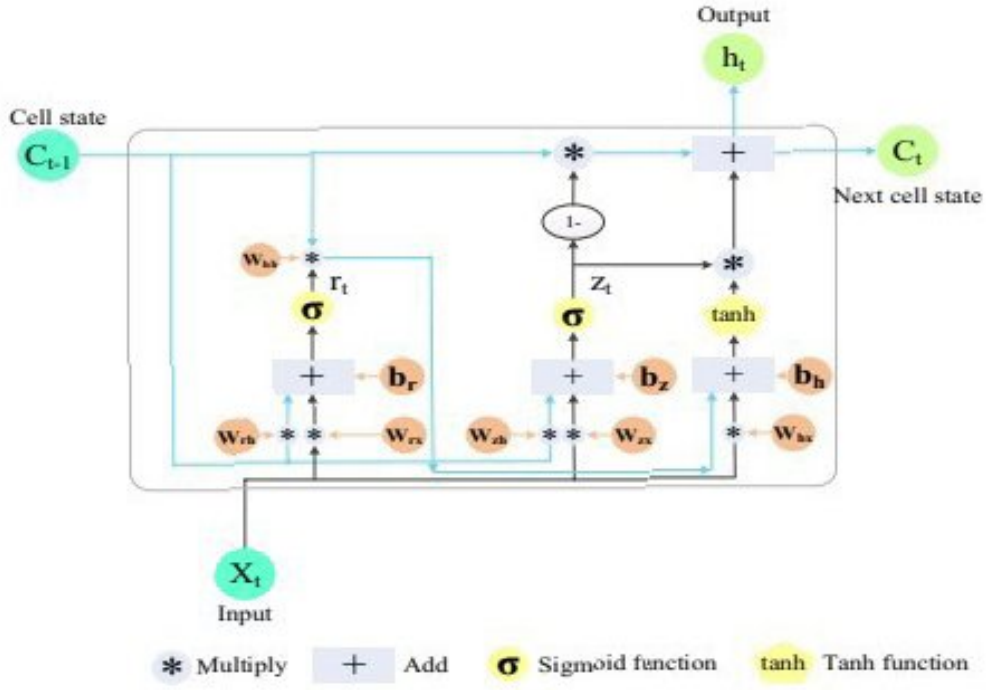


**Figure 3.** GRU structure.

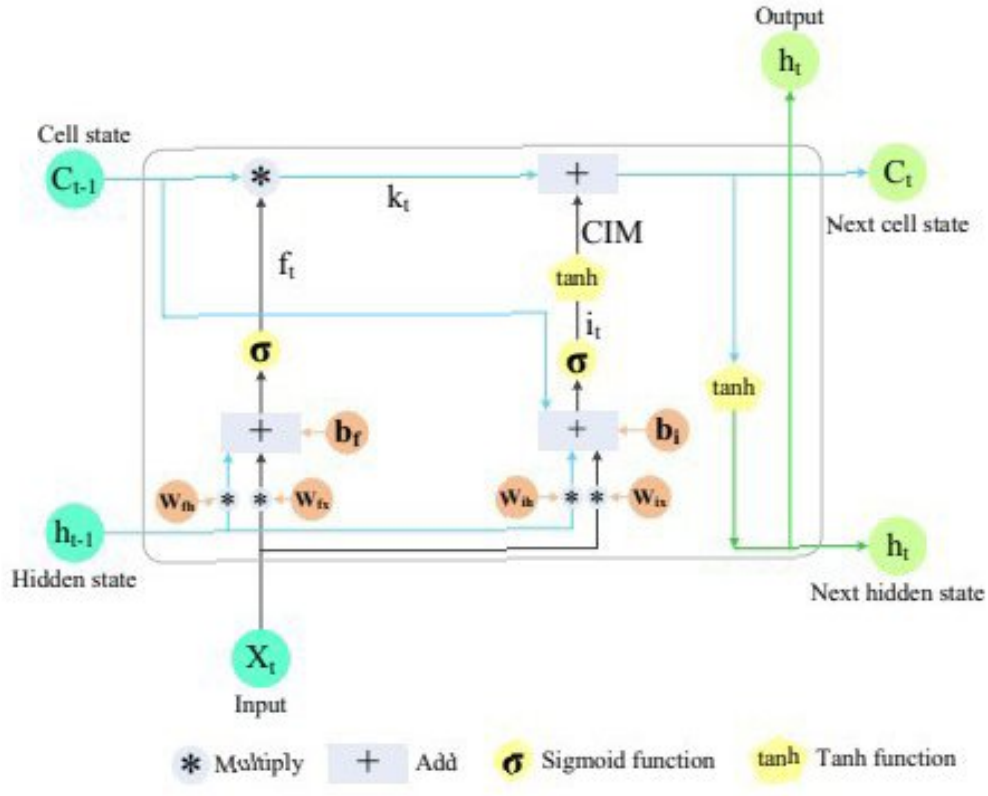


**Figure 4.** ILSTM structure.

| LSTM | GRU | ILSTM |
|---|---|---|
| 3 gates | 2 gates | 2 gates |
| 8 weight parameters | 6 weight parameters | 4 weight parameters |
| 4 bias parameters | 3 bias parameters | 2 bias parameters |

**Table 1.** The parameters of LSTM and ILSTM model.

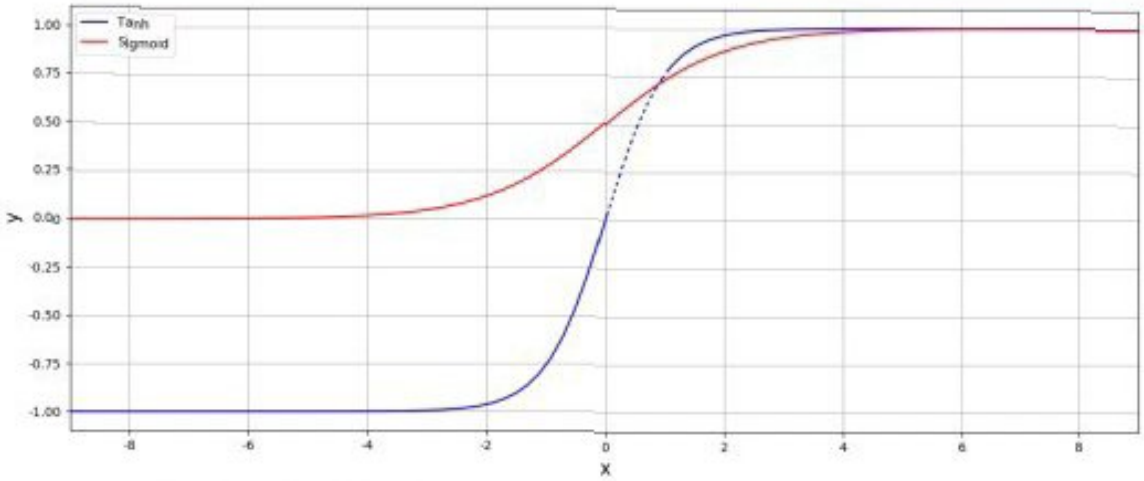


**Figure 5.** Tanh and Sigmoid function.

$$i_t = \sigma(W_{ih} \cdot h_{t-1} + W_{ix} \cdot x_t + c_{t-1} + b_i). \tag{4}$$

Due to the characteristics of the Sigmoid activation function, when the value of x is outside − 3 and 3, the value of the Sigmoid activation function will fall into a supersaturation interval. Therefore, in formula (4), when the input data enters the input gate's supersaturation, the value does not change significantly, decreasing learning sensitivity. A CIM is introduced into ILSTM model to prevent this phenomenon, As shown in formula (5). The Sigmoid function value (ranging from 0 to 1) calculated by the above formula is taken as the input of Tanh. Tanh and Sigmoid function as shown in Fig. 5.

The value of $\tanh(i_t)$ will be between $[0, 0.762)$, as shown in the dotted line part by the tanh function in Fig. 5, so the obtained value will be more uniform and significant. Therefore, the value output by the CIM will greatly reduce the supersaturation degree, and the significant difference makes the model calculation more recognizable, thereby making the model learning more sensitive.

$$\mathrm{CIM} = \tanh(i_t). \tag{5}$$

Formula (6) shows that $c_t$ is the information kept from the beginning to the present.

$$c_t = k_t + \mathrm{CIM}. \tag{6}$$

$h_t$ indicates the information preserved at the current time. $c_t$ controls how much information can be kept through tanh function, as shown in formula (7).

$$h_t = \tanh(c_t). \tag{7}$$

*Formula derivation of ILSTM.* ILSTM is proposed to improve the model's prediction accuracy and reduce the model's training time on the premise that the model can alleviate the issues of "gradient explosion" and "gradient disappearance" of the RNN. Input gate and forget gate use two parameter matrices $[W_{fh}, W_{fx}]$ and $[W_{ih}, W_{ix}]$. Record $W_f = [W_{fh}, W_{fx}]$, $W_i = [W_{ih}, W_{ix}]$,and $W = [W_f, W_i]$. The $L_t$ function of W is the loss corresponding to $h_t$. L is the total loss. As for the derivative of W of L, as shown in formula (8):

$$\frac{\partial L}{\partial W} = \sum_{t=1}^{T} \frac{\partial L_t}{\partial W}. \tag{8}$$

The RNN updates the W parameter by formula (9):

$$W = W - \frac{\partial L_t}{\partial W}, \tag{9}$$

where $\frac{\partial L_t}{\partial W}$ can be written as formula (10):

$$\frac{\partial L_k}{\partial W} = \frac{\partial L_k}{\partial h_k}\frac{\partial h_k}{\partial c_k}\frac{\partial c_k}{\partial c_{k-1}}\cdots\frac{\partial c_2}{\partial c_1}\frac{\partial c_1}{\partial W}. \tag{10}$$

Formula (10) can be simplified to formula (11):

$$\frac{\partial L_k}{\partial W} = \frac{\partial L_k}{\partial h_k}\frac{\partial h_k}{\partial c_k}\cdots\frac{\partial c_2}{\partial c_1}\frac{\partial c_1}{\partial W} = \frac{\partial L_k}{\partial h_k}\frac{\partial h_k}{\partial c_k}\left(\prod_{t=2}^{k}\frac{\partial c_t}{\partial c_{t-1}}\right)\frac{\partial c_1}{\partial W}, \tag{11}$$

where $c_t$ is shown in formula (12):

$$c_t = f_t \times c_{t-1} + CIM. \tag{12}$$

The $CIM = \tanh(i_t)$, so formula (12) can be written as formula (13):

$$c_t = f_t \times c_{t-1} + \tanh(i_t). \tag{13}$$

The derivative of $c_t$ can be obtained by formula (14):

$$\frac{\partial c_t}{\partial c_{t-1}} = f_t + (1-\tanh(i_t)^2)(i_t)'. \tag{14}$$

Then the total loss can be written as formula (15):

$$\frac{\partial L_k}{\partial W} = \frac{\partial L_k}{\partial h_k}\frac{\partial h_k}{\partial c_k}\left(\prod_{t=2}^{k}\left(f_t + \left(1-\tanh\left(i_t\right)^2\right)\left(i_t\right)'\right)\right)\frac{\partial c_1}{\partial W}. \tag{15}$$

Then record that x is equal to formula (16):

$$x = W_{fh}\cdot h_{t-1} + W_{fx}\cdot x_t + b_f. \tag{16}$$

Then that $f_t$ can be written as formula (17):

$$f_t = \sigma(W_{fh}\cdot h_{t-1} + W_{fx}\cdot x_t + b_f) = \sigma(x). \tag{17}$$

Then record that y is equal to formula (18):

$$y = W_{ih}\cdot h_{t-1} + W_{ix}\cdot x_t + c_{t-1} + b_i. \tag{18}$$

Then that $i_t$ can be written as formula (19):

$$i_t = \sigma(W_{ih}\cdot h_{t-1} + W_{ix}\cdot x_t + c_{t-1} + b_i) = \sigma(y). \tag{19}$$

Then the formula (15) can be written as formula (20):

$$\frac{\partial L_k}{\partial W} = \frac{\partial L_k}{\partial h_k}\frac{\partial h_k}{\partial c_k}\left(\prod_{t=2}^{k}\left(\sigma(x) + \left(1-\tanh\left(\sigma(y)\right)^2\right)\sigma'(y)\right)\right)\frac{\partial c_1}{\partial W}. \tag{20}$$

Then record that z(x, y) is equal to formula (21):

$$z(x,y) = \sigma(x) + (1-\tanh(\sigma(y))^2)\sigma'(y). \tag{21}$$

Then the formula (20) can be written as formula (22):

$$\frac{\partial L_k}{\partial W} = \frac{\partial L_k}{\partial h_k}\frac{\partial h_k}{\partial c_k}\left(\prod_{t=2}^{k}\left(z(x,y)\right)\right)\frac{\partial c_1}{\partial W}. \tag{22}$$

As shown in formula (22), the gradient of the function is $\frac{\partial L_k}{\partial h_k}\frac{\partial h_k}{\partial c_k}\left(\prod_{t=2}^{k}\left(z(x,y)\right)\right)\frac{\partial c_1}{\partial W}$. When z(x, y) is greater than 1, the gradient may be too large with the increase of data amount. When z(x, y) is too small, the gradient disappears easily.

In this model, the $\sigma(x)$ function is shown in Fig. 6, and the $(1-\tanh(\sigma(y))^2)\sigma'(y)$ function is shown in Fig. 7. $\sigma(x)$'s range is [0,1] and $\sigma(y)$'s range is [0, 1], so the function range of $(1-\tanh(\sigma(y))^2)\sigma'(y)$ is (0.1720, 0.1880). Function gradient z(x, y) is shown in Fig. 8. It can be seen from the figure that the value range of function gradient z(x, y) will be more reasonable. Therefore, this model can alleviate the problems of "gradient disappearance" and "gradient explosion" to a great extent.

**CNN-ILSTM.** The structure of CNN-ILSTM is shown in Fig. 9. The CNN-ILSTM model is generally divided into four parts. The first layer is the data input layer. This paper takes AQI as the research object and air quality data as the model input. The second layer is the data preprocessing layer. To ensure the reliability of the prediction results and improve the accuracy of the prediction results, it is necessary to conduct standardized processing of the original data, three-dimensional time series construction and other pre-processing operations. The third layer is the feature extraction layer, which realizes feature extraction of air quality data by taking advantage of

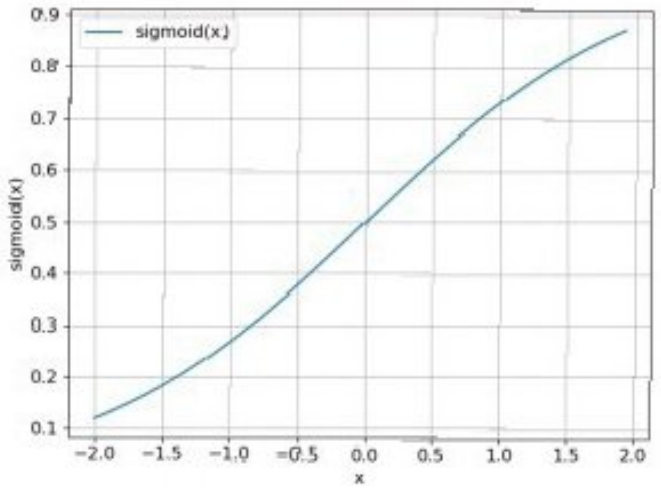


**Figure 6.** $\sigma(x)$ Function.

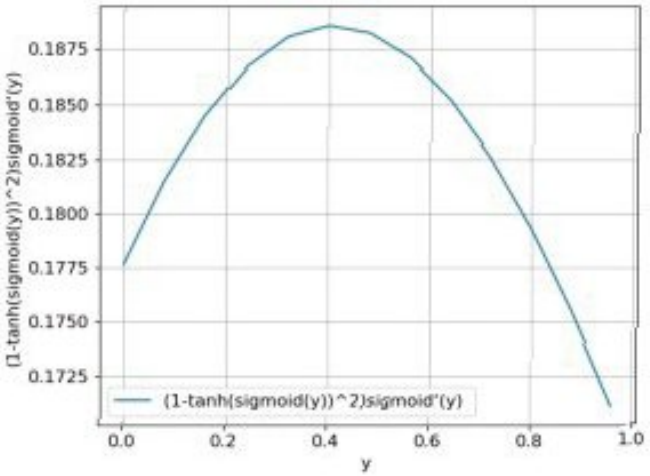


**Figure 7.** $(1-\tanh(\sigma(y))^2)\sigma'(y)$ Function.

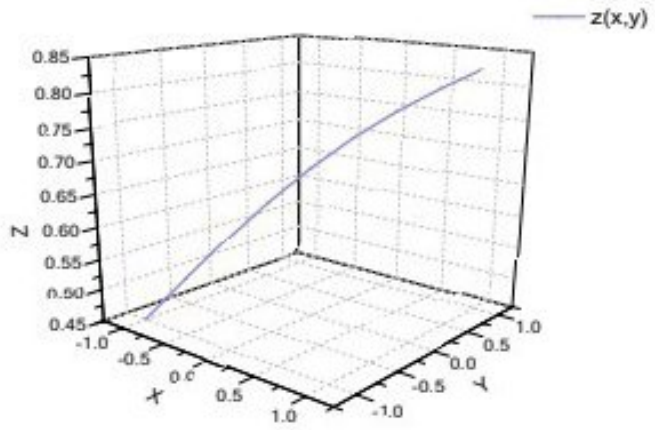


**Figure 8.** z (x, y) function.

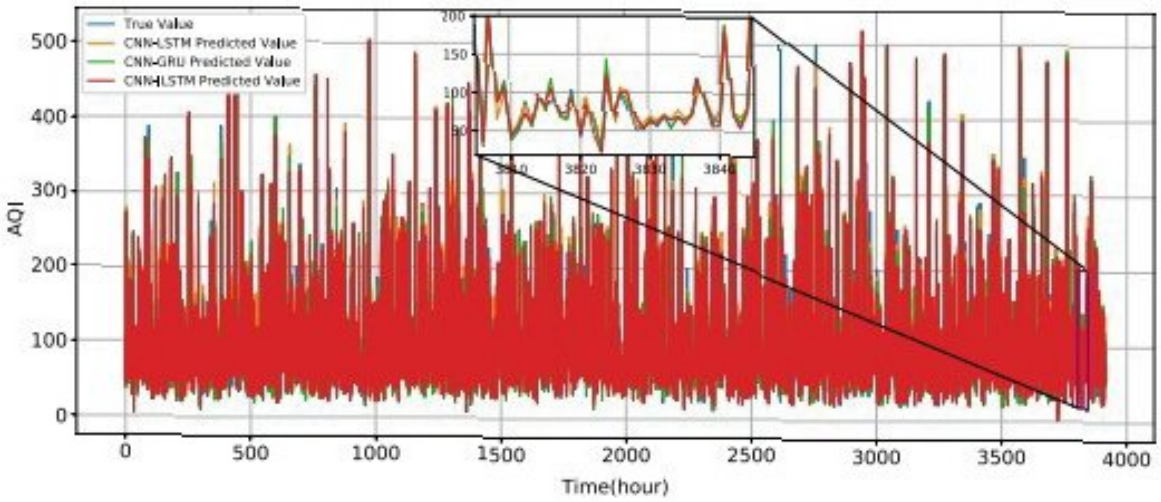


**Figure 16.** True value and the predict value of CNN-LSTM, CNN-GRU, and CNN-ILSTM.

1. The introduction of the model makes up for the shortcomings of the single model in some aspects. For example, CNN can realize the advantages of data feature extraction, which makes up for the RNN's deficiencies in eigenvalues screening and learning before data input.
2. ILSTM deletes the output gate, improves the input gate and forget gate, and introduces a CIM to learn historical data more thoroughly.

ILSTM and CNN-ILSTM have been greatly improved in model training time because:

1. The structure of the ILSTM model is simpler. ILSTM consists of input gate and forget gate. Compared with LSTM, ILSTM has no output gate.
2. ILSTM has fewer parameters. Compared with LSTM, ILSTM reduces weight parameters from 8 to 4 and bias parameters from 4 to 2. Compared with GRU, ILSTM reduces weight parameters from 6 to 4 and bias parameters from 3 to 2.

## Conclusions

This paper presents an AQI prediction model based on CNN-ILSTM. Compared with the traditional regression models of SVR, RFR, and MLP, and the deep learning models of LSTM, GRU, ILSTM, CNN-LSTM, and CNN-GRU, the overall evaluation of prediction results of CNN-ILSTM is best. ILSTM is proposed for the first time. ILSTM is improved and optimized in model design and parameter ratio on the premise of high prediction accuracy and alleviating the issues of "gradient explosion" and "gradient disappearance" in the RNN caused by long-term data dependence. Compared with LSTM and GRU, the training time of ILSTM is reduced by 48.6% and 10.34%, and ILSTM has the best AQI prediction results. In addition, the introduction of CNN makes up for the deficiency of ILSTM feature extraction and learning. The experiment results show that the MAE of CNN-ILSTM decreases by 0.284798, and the $R^2$ increases by 0.013951 compared with ILSTM AQI prediction. The conclusions of this paper are as follows:

1. ILSTM has performed better than LSTM in my tests. ILSTM is an improvement of LSTM, which deletes the output gate in LSTM, improves its input gate and forget gate, and introduces a CIM to prevent supersaturation in the learning process. On the premise of ensuring that the model can alleviate the issues of "gradient explosion" and "gradient disappearance" of RNN and has high prediction accuracy. Compared with LSTM and GRU, ILSTM significantly reduces the training time.
2. The AQI prediction model of CNN-ILSTM makes up for the shortcomings of the single prediction model, such as insufficient feature data extraction and insufficient historical data learning. In this experiment, the AQI prediction model of CNN-ILSTM is the best.
3. The model design and parameter tuning are improved and optimized, so the convergence rate of the AQI prediction model based on CNN-ILSTM is improved.

However, the AQI prediction model of CNN-ILSTM does not perform well in extreme value prediction. Therefore, the following research will carry out the high-precision prediction of extreme values.

Received: 13 February 2022; Accepted: 10 May 2022
Published online: 19 May 2022

| Model | MAE | MSE | $R^2$ | Training time (s) |
|---|---|---|---|---|
| SVR | 19.2644 | 483.6269 | 0.8750 | 411.2 |
| RFR | 15.364 | 402.6292 | 0.8905 | 341.3 |
| MLP | 13.6186 | 386.3648 | 0.9106 | 50.3 |
| LSTM | 9.4974 | 266.8523 | 0.9447 | 124.9 |
| GRU | 11.0032 | 270.8179 | 0.9411 | 71.6 |
| ILSTM | 9.2420 | 248.9234 | 0.9508 | 64.2 |
| CNN-LSTM | 9.2314 | 258.5143 | 0.9466 | 149.5 |
| CNN-GRU | 9.0615 | 248.6389 | 0.9478 | 92.5 |
| CNN-ILSTM | 8.4134 | 202.1923 | 0.9601 | 85.3 |

**Table 6.** Experimental results.

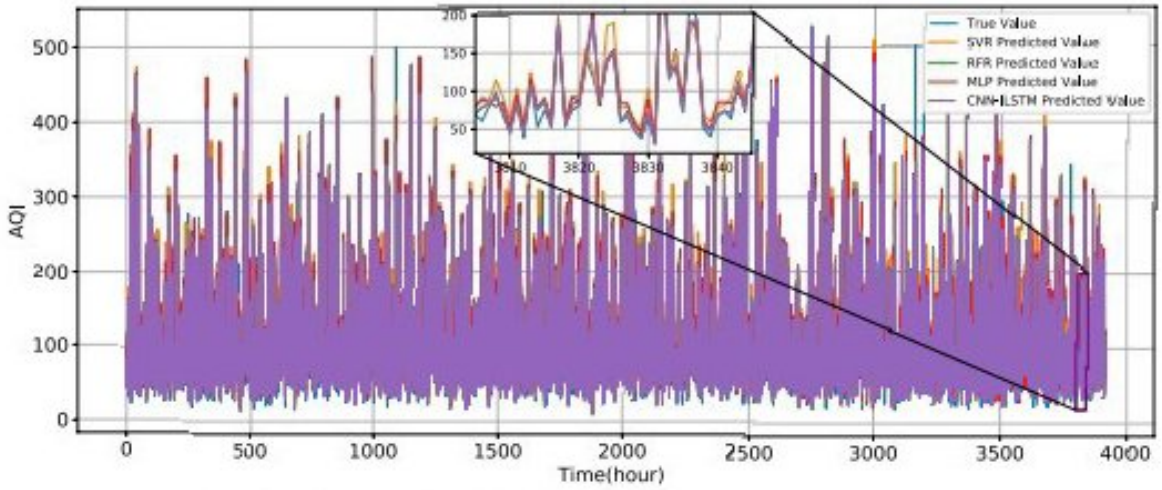


**Figure 13.** True value and the predicted value of SVR, RFR, MLP, and CNN-ILSTM.

*Evaluation index.* To scientifically evaluate the prediction accuracy of the model, this experiment uses MAE, MSE, $R^2$, and model training time as the overall evaluation index of the model. MAE describes how different the predicted value is from the true value. MSE measures the average modulus length of the predicted value error, regardless of direction. $R^2$ describes how similar the predicted value is to the true value. The model training time describes the calculation time of the model. The calculation methods of MAE, MSE and $R^2$ are shown in formula (25–27).

$$\mathrm{MAE} = \frac{1}{m}\sum_{i=1}^{m}\left|\left(y_i - \hat{y}_i\right)\right|, \tag{25}$$

$$\mathrm{MSE} = \frac{1}{m}\sum_{i=1}^{m}\left(y_i - \hat{y}_i\right)^2, \tag{26}$$

$$R^2 = 1 - \frac{\sum_i \left(\hat{y}_i - y_i\right)^2}{\sum_i \left(\overline{y}_i - y_i\right)^2}, \tag{27}$$

where m is the number of data in the test set; $\hat{y}_i$ is the predicted value; $y_i$ is the true value; $\overline{y}_i$ is the average value of the true values.

*Experimental results.* To verify the accuracy of CNN-ILSTM in predicting AQI, traditional regression models (SVR, RFR, and MLP), recurrent neural network models based on gated technology (LSTM, GRU, ILSTM), and hybrid prediction models (CNN-LSTM, CNN-GRU) are introduced as comparison models.

Experimental results are shown in Table 6. The test set prediction evaluation results show that the traditional regression models SVR, RFR, and MLP have a lower prediction fitting degree than the recurrent neural network model based on gated technology. The $R^2$ of LSTM is 0.0697 higher than that of SVR, the $R^2$ of LSTM is 0.0542 higher than that of RFR, and the $R^2$ of LSTM is 0.0341 higher than that of MLP. The predicted and true values of SVR, RFR, MLP, and CNN-ILSTM are shown in Fig. 13.

fuels, deforestation, and transportation emissions, not only create air pollution but also contribute to global climate change.

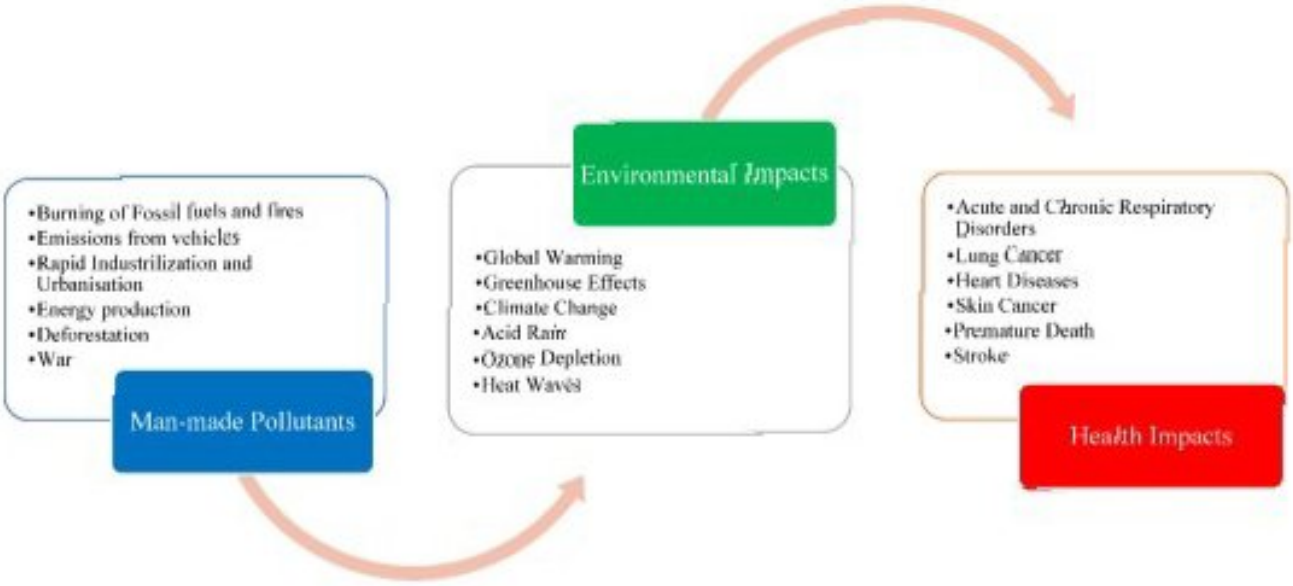


**Figure 1.** Potential Relationship between Man-made Pollutants, Environmental Impacts and Health Impacts.

Human health and life quality, which are determined by environmental psychosocial, physical, biological, and social characteristics, are factors in environmental health [5]. Climate changes lead to alterations in weather and temperature pattern. Although these changes are natural, humans have been the dominant contributor to climate change since the 19th century, primarily to the burning of fossil fuels which generates $CO_2$. Ahmed et al. determined the regional and periodic change of $CO_2$ and Particulate matter in the important Libyan city of Misurata and the measurements were made during the months of November and February once every three days [6]. The results were very useful to understand the regional and periodical characteristics of air pollution. Cetin et al. evaluated the regional and temporal changes in air pollution and measured the $CO_2$ concentration and particulate matter concentration in various regions of Bursa city. The results of the study revealed that the $CO_2$ was statistically not significant with respect to the season but the Particulate matter shows, statistically, a 99.9% confidence level by season [7]. In the Turkish case study, the author examined the variation in the indoor $CO_2$ concentration in the examination hall and the findings indicate that indoor $CO_2$ levels are higher than 1500 ppm while the start of exams of threshold value within 10 min [8]. In another interesting study, the authors examined the Pb and Cr pollution in the capital city of Turkiye by collecting topsoil samples from 50 regions; thus, it is clear that air pollution is very dangerous to the environment and ecosystem. The earth is currently warming faster than at any other time in recorded history. Warmer temperatures are shifting weather patterns and disturbing nature's normal equilibrium. As a result of climate change, storms, floods, cold spells, and heat waves are expected to have a greater socioeconomic impact. Heat Waves (HWs) are expected to grow increasingly strong and common as a result of man-made climate change. HWs are clearly major events that can induce fast changes in biodiversity patterns as well as ecosystem structure and functioning as a result of human climate change [9]. To overcome the above-mentioned issues, it is very important to forecast air pollution, weather conditions, and climate changes to implement an early-warning system.

Artificial Intelligence is used to imitate the human mind's problem-solving and decision-making abilities. Figure 2 shows the relationship between Artificial Intelligence methods, Machine Learning algorithms, and Deep Learning algorithms. To limit public exposure risks due to pollution, AI should be used as an important method for environ-

mental pollution forecasting and which helps policymakers to develop better policies in the case of environmental protection [10]. AI can manage complicated and non-linear interactions between spatial-temporal factors, which makes it a better method for air pollution predicting and forecasting [11]; it offers an outstanding capacity for tracking the current pollution condition and demonstrates a precise and quick method for identifying pollution hotspot areas. Deep learning is a subset of the AI method which is also used for forecasting and predicting extreme weather conditions. The important ability of AI is the easy management of a lot of data and real-world complex problems; further, it increases the accuracy in the case of weather forecasting and modelling which helps policymakers and decision-makers [12].

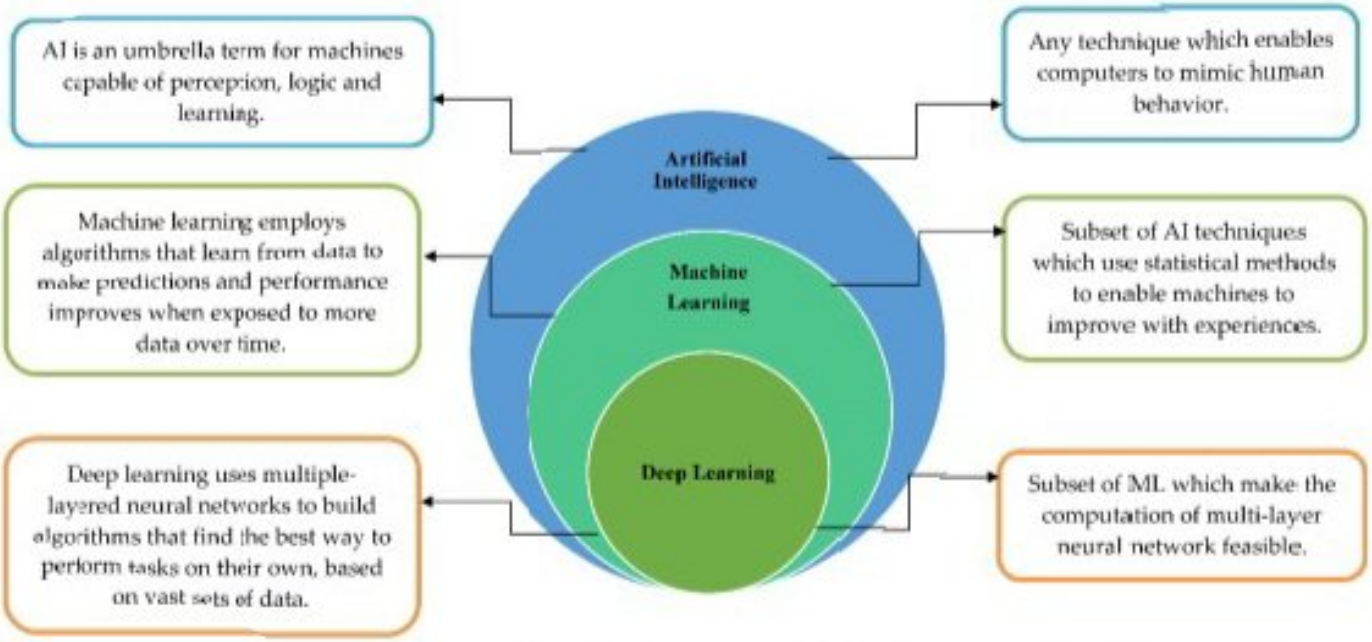


**Figure 2.** Relationship between Artificial Intelligence, Machine Learning and Deep Learning.

### *1.1. Sources and Impacts of Pollutants on Environmental and Human Health*

In many parts of the world, air pollution is a bigger crisis due to its direct impact not only on environmental health but also on human health, which leads to various environmental crises and notably health crises such as early death, respiratory illness, stroke, cardiovascular disease and etc. [13]. Natural and anthropogenic activity generates a harmful and toxic mixture of particles and gases that are emitted into the atmosphere which leads to several crises [14]. Natural calamities such as sudden occurrences of forest fires and volcanic eruptions emitted toxic pollutants such as $SO_2$, $CO_2$, $NO_2$, Carbon monoxide, and Particulate emission [15]. Nowadays, due to increased industry and automotive gas emissions, atmospheric air has become more poisonous, contaminating the air we breathe. The main air pollutants include $CO_2$, $NO_2$, $SO_2$, CO, and particulate matter, all of which have adverse effects on the ecosystem and individuals. Figure 3 represents the major air pollutants and their causes on the environment and humans. Various major air pollutants lead to the environmental atmosphere causing the greenhouse effect, ozone reduction, and photochemical smog, aggravating the environment and producing significant environmental disasters [16], which are responsible for several severe health issues (i.e., chronic disease) [17]. Volcanic eruptions and industrial growth are the major causes of sulphur dioxide ($SO_2$). The presence of $SO_2$ in the atmosphere may also contribute to acid rain, which harms the ecosystem, and also leads toa worsening of the human respiratory system [18].

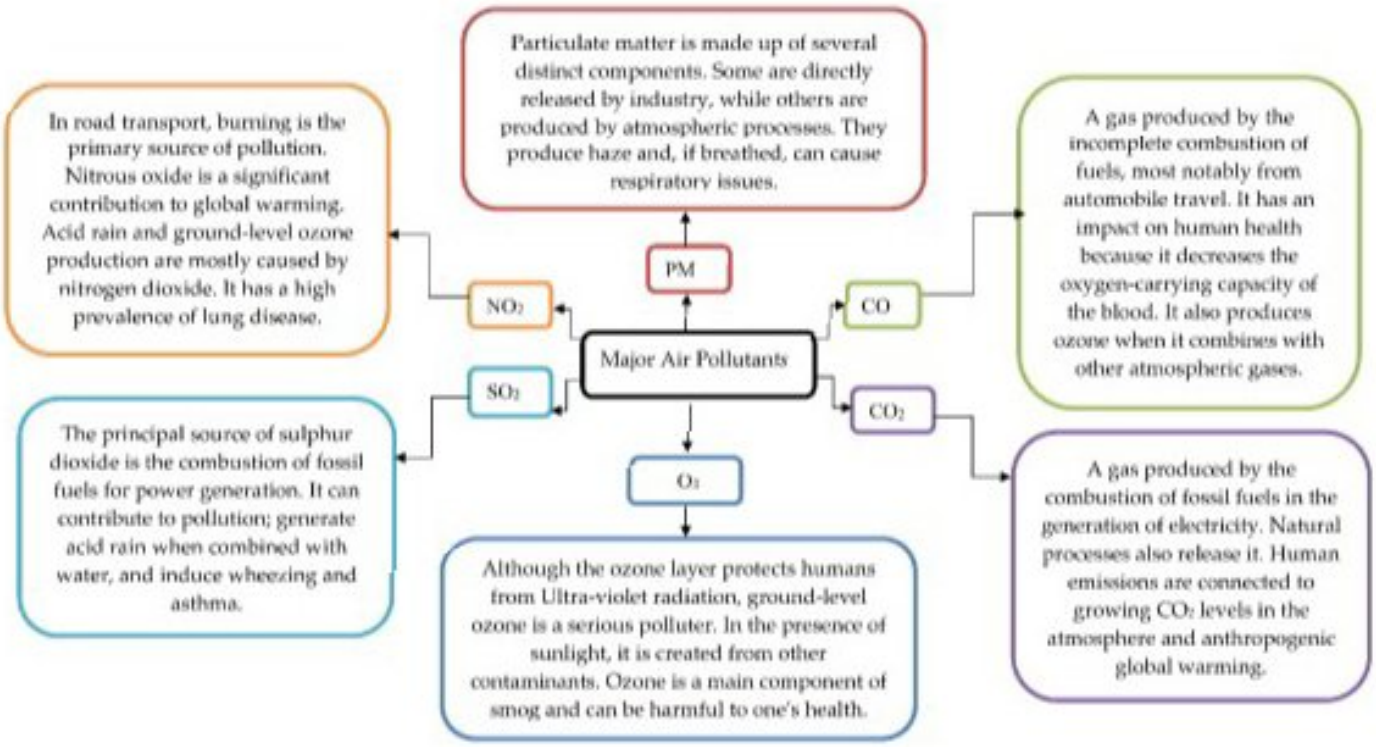


**Figure 3.** Major Air Pollutants and their Sources and Effects.

The burning of fuels in industry and transportation is the major cause of Nitrogen dioxide emissions [3,17]. $NO_2$ exposure results in adverse effects on human individuals and ecosystems, resulting in a rise in lung disease prevalence. The ozone gas at ground level causes severe health impacts to humans due to its high oxidizing capabilities [3]. When we are exposed to air pollution for an extended period, our pulmonary function suffers significantly. Lung cancer, asthma and chronic obstructive pulmonary disease can all be caused by a decline in pulmonary function [4,17]. According to studies, particles with a diameter of 10 μm can pass through the respiratory tract, whereas particles with a diameter of less than 5 μm can penetrate the deep section of the bronchioles [18]. Furthermore, particles smaller than 1μm can enter the alveoli [17]. Poor air quality may impair cognition and contribute to mental illnesses [17,18]. As a consequence, reducing air pollution can be an effective way to combat the diseases indicated above while also benefiting society [3,19]. Air quality forecasting, monitoring, and early warning systems, as preventive measures, form the foundation for successful pollution control measures and the development of environmental laws to improve air quality, and it is very helpful in designing sustainable smart cities, environmental sustainability, and pollution control management. As preventive measures, air quality forecasting, early warning systems and monitoring are the foundation for successful pollution control measures and the development of environmental laws to improve air quality and it is very helpful in designing sustainable smart cities, environment sustainability and pollution control management.

### *1.2. Current Status of Research on Environmental Pollution Forecasting Techniques*

Air pollution is an inescapable fact of life; it is well known that the effects of environmental pollution are more severe than those of soil and water pollution [19]. Governments all across the world are attempting to improve air quality by enacting various environmental laws. Without a question, most developed countries want to use AI technologies and approaches in their environmental policy today. The forecasting models may be classified into three categories: physical models, statistical models, and AI models; it primarily studies weather phenomena and meteorological conditions for the physical forecasting

**Table 1.** *Cont.*

| S. No | Hybrid AI Models | Pollutant Forecasting | Ref. |
|---|---|---|---|
| 7 | Wavelet packet decomposition—complete ensemble empirical mode decomposition with adaptive noise—Least squares support vector regression—chaotic particle swarm optimization method and gravitation search algorithm (WPD-CEEMD-LSSVR-CPSWOM-GSA) | $PM_{2.5}$ | [34] |
| 9 | Variational mode decomposition—Sample entropy—Least squares support vector machine (VMD-SE-LSSVM) | AQI | [35] |
| 16 | Complementary empirical ensemble mode decomposition—Cuckoo search—Grey wolf optimizer- support vector machine (CEEMD-CS-GWO-SVM) | $NO_2$ & $SO_2$ | [36] |
| 11 | Wavelet packet decomposition (WPD)—Bidirectional Long Short-Term Memory (Bi-LSTM)—Stacked auto encoder Non-dominated Sorting Genetic Algorithm II (NSGA-II). | $PM_{2.5}$ | [37] |

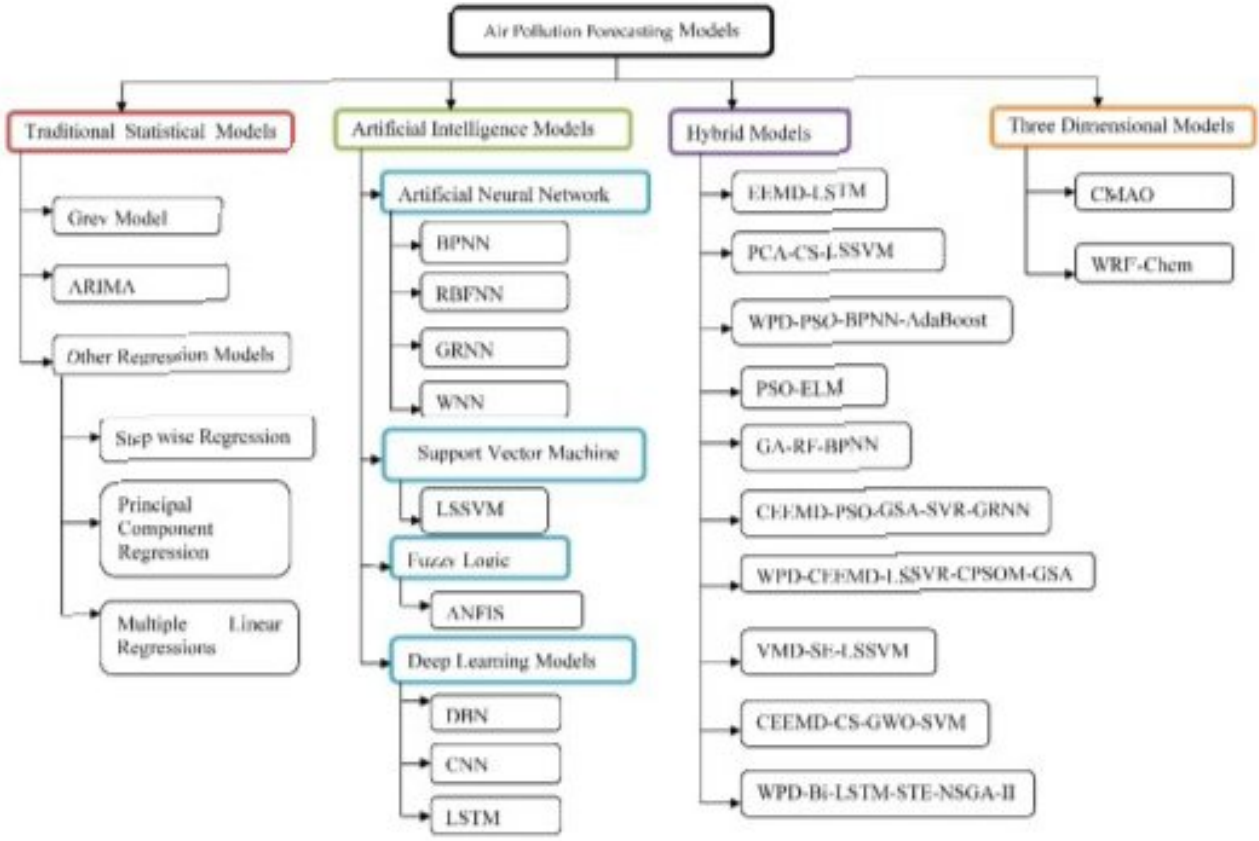


**Figure 4.** Different Methods of Air Pollution Forecasting.

Artificial intelligence (AI) has recently emerged as the most widely used technology tool for regulating and mitigating the detrimental effects of various air pollutants, generating considerable interest in the disciplines of atmospheric and medical science [38]. Many researchers have utilized AI approaches as a better healthcare decision support device to detect, manage, and cure illnesses caused by air pollution [19]. AI is one such instrument, with tremendous potential to expedite climate change mitigation and adaptation methods in sectors such as energy, land use, and disaster response; however, several bottlenecks and hurdles now exist that prevent AI from reaching its full potential in this area [39].

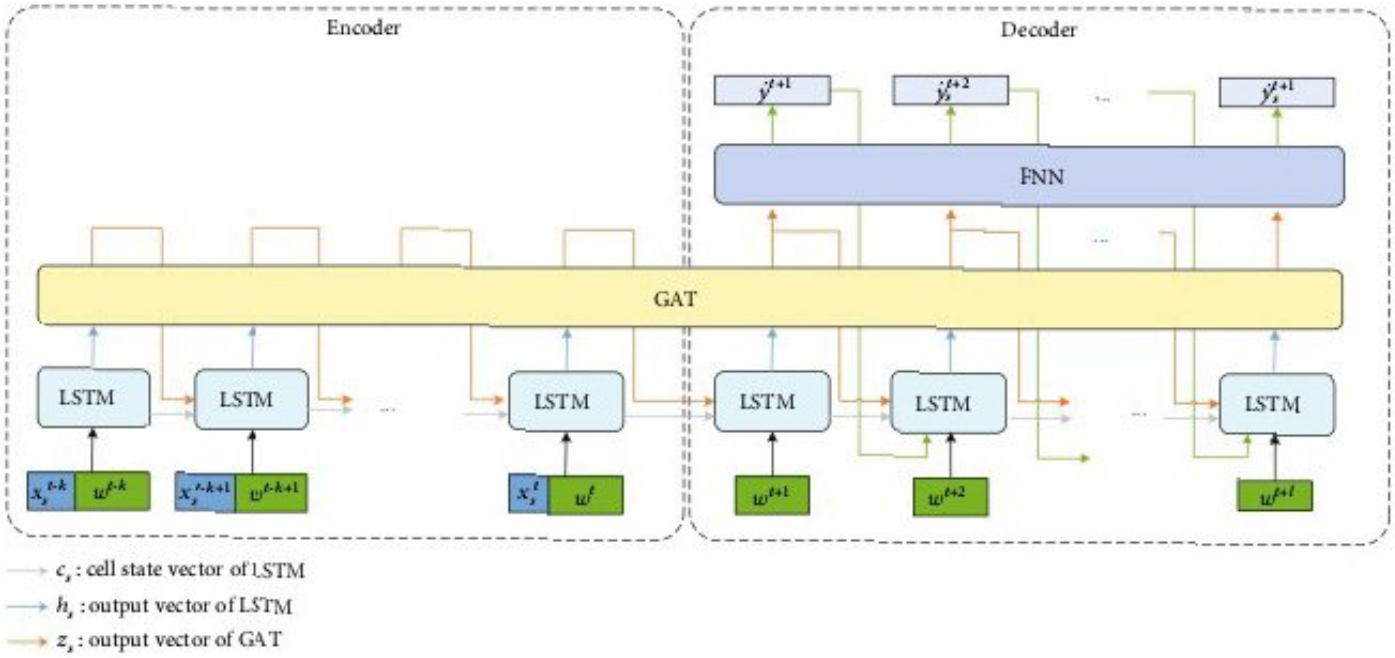


Figure 2: The air quality prediction model GAT-LSTM.

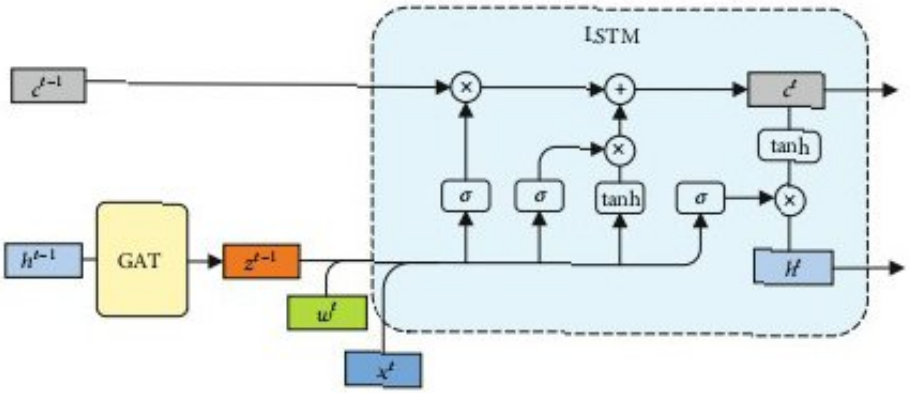


Figure 3: The LSTM unit in GAT-LSTM.

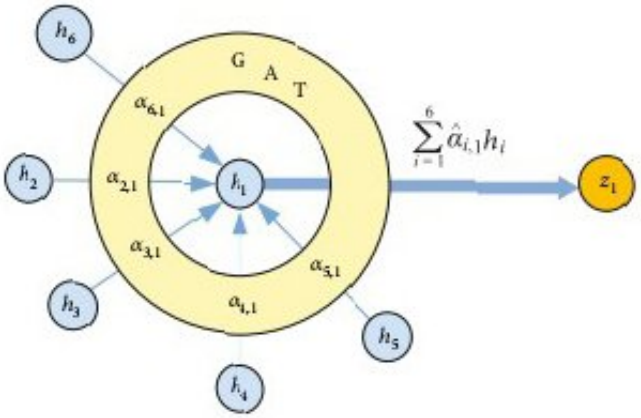


Figure 4: The GAT mechanism for node 1 (suppose that node 2 ~ 6 are 1's neighbors in $G_s$).

where $\mathscr{L}_{B_i}(f_\theta)$ is the loss of the original model $f_\theta$ on training batch $B_i$, $\nabla\mathscr{L}_{B_i}(f_\theta)$ is gradient of $\mathscr{L}_{B_i}(f_\theta)$ and $\beta$ is the learning rate in the first step. In the second step (Line 11 in Algorithm 1), we get the second-step adapted parameter by

$$\theta = \theta - \gamma \sum_{i=1}^{k} \nabla\mathscr{L}_{B_0}\left(f_{\theta_i}\right), \tag{8}$$

where $\mathscr{L}_{B_0}(f_{\theta_i})$ is the loss of first-step adapted model $f_{\theta_i}$ on training batch $B_0$ and $\gamma$ is the learning rate in the second step.

## 5. Experimental Results and Analysis

*5.1. Dataset.* We use real datasets collected from four cities in China (Beijing, Tianjin, Shenzhen, Guangzhou) to verify the effectiveness and efficiency of the proposed model and meta-learning algorithm. These cities are very different in

Table 2: Data details.

| Attribute | | Number of different items | | | |
|---|---|---|---|---|---|
| | | Beijing | Tianjin | Shenzhen | Guangzhou |
| Time span | | 2014/5/1-2015/4/30 | 2014/5/1-2015/4/30 | 2014/5/1-2015/4/30 | 2014/5/1-2015/4/30 |
| AQIs | | 278023 | 189604 | 88139 | 281436 |
| Station ID | | 36 | 27 | 11 | 42 |
| Major pollutants | $PM_{10}$<br>$NO_2$<br>CO<br>$O_3$<br>$SO_2$ | 278023 | 189604 | 88139 | 281436 |
| Historical weather | District ID | 17 | 20 | 7 | 5 |
| | Basic weather<br>Temperature<br>Pressure<br>Humidity<br>Wind speed<br>Wind direction | 116867 | 106614 | 30305 | 55632 |
| Weather forecast | District ID | 17 | 20 | 6 | 5 |
| | Basic weather<br>Wind strength<br>Wind direction | 390702 | 361624 | 106380 | 51870 |

$$\mathrm{ACC} = 1 - \frac{1}{|\mathcal{T}|} \sum_{(X,y)\in\mathcal{T}} \frac{\|f(X) - y\|_1}{\|y\|_1}. \tag{11}$$

Here, $\mathcal{T}$ is the test set. $y$ is sample's label (true monitoring data in the future) and $f(X)$ is the predicted value of $y$. $\|\cdot\|_1$ and $\|\cdot\|_2$ are L1 and L2 norm, respectively.

In GAT-LSTM, the dimension of the GAT's output vector, the LSTM's output vector, and cell state vector are all set to 128. While training, we use dropout [34] and batch normalization [35] to strengthen the training effect. The batch size is set to 64. The number of epochs is set to 3.

*5.3. Experiment Results.* At first, we need to find an appropriate influence radius $r$ for building the directed graph $G_r$ in GAT-LSTM, so we compare the performance of GAT-LSTM with different $r$. Table 3 and Figure 5 give the comparison results on the dataset from Beijing. They show that when $r < 20$km, the accuracy of GAT-LSTM increases as $r$ increases. The reason for this phenomenon is that when $r$ is within a reasonable range, a larger $r$ allows the model to consider more spatial correlation, thereby providing a more accurate prediction. The accuracy reaches its peak at $r = 20$ km. When $r > 20$km, the accuracy decreases slightly as $r$ increases. This phenomenon is because too large $r$ will cause the model to incorrectly estimate the correlation among some remote monitoring sites based on the data similarity. Thus, we set $r$ =20 km in the following experiments.

In the first group of experiments, we use the dataset from Beijing to evaluate all the prediction models. We divide the dataset into training set (70%), validation set (20%), and test set (10%). Each model takes data from the past 48 hours as input ($k + 1 = 48$), then outputs prediction values for the next $l$ hours. Table 4 shows the experiment results with different $l$. The best results are marked in bold. It can be seen that the traditional linear model ARIMA does not perform well under the influence of multiple complex factors. LSTM's performance is acceptable for short-term prediction and drops quickly with the increase of $l$. Spatial correlation plays an important role in air quality prediction. By using CNN to extract spatial correlation among monitoring stations, the ST-DNN performs much better than ARIMA and LSTM. However, the spatial correlation built by ST-DNN cannot change dynamically with the change of weather, which reduces its predictive effects. By using GAT to dynamically model spatial correlation, GAT-LSTM gives the best performance in all cases. The performance of all models declines with the increase of $l$, but the decline rate of GAT-LSTM is lower than the other three, which s that it is suitable for long-term prediction.

Table 3: The performance of GAT-LSTM with different influence radius $r$ on Beijing dataset ($l = 6$).

| $r$ (km) | RMSE | MAE | ACC |
|---|---|---|---|
| 5 | 39.1 | 24.8 | 0.767 |
| 10 | 33.3 | 21.9 | 0.778 |
| 20 | 28.5 | 20.1 | 0.799 |
| 25 | 30.3 | 20.5 | 0.797 |
| 40 | 29.7 | 20.2 | 0.796 |
| 60 | 28.8 | 20.5 | 0.798 |
| 100 | 29.3 | 20.2 | 0.796 |

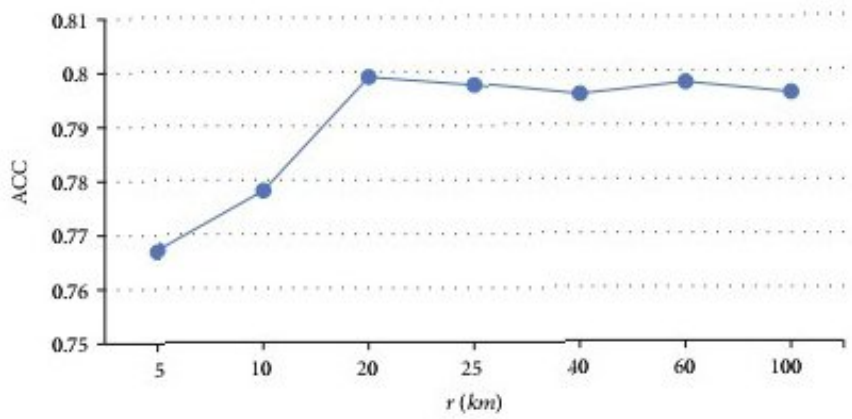


Figure 5: The relationship between influence radius $r$ and prediction accuracy ($l = 6$).

Table 4: Comparison of mean prediction results by different methods among 36 monitoring stations in Beijing (RMSE, MAE, ACC).

| Methods | +6 h | | | +12 h | | | +24 h | | | +48 h | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | RMSE | MAE | ACC | RMSE | MAE | ACC | RMSE | MAE | ACC | RMSE | MAE | ACC |
| ARIMA | 53.3 | 40.7 | 0.622 | 70.7 | 59.2 | 0.432 | 94.1 | 68.2 | 0.321 | 104.3 | 78.2 | 0.201 |
| LSTM | 48.2 | 34.5 | 0.723 | 57.3 | 47.3 | 0.582 | 71.2 | 55.1 | 0.473 | 85.7 | 66.3 | 0.367 |
| ST-DNN | 35.6 | 23.8 | 0.761 | 52.2 | 37.1 | 0.67 | 63.2 | 49.3 | 0.546 | 70.4 | 57.7 | 0.474 |
| GAT-LSTM | 28.5 | 20.1 | 0.799 | 47.9 | 34.8 | 0.698 | 55.7 | 45.2 | 0.621 | 65.8 | 48.9 | 0.501 |

Table 5: RMSE of transfer learning methods over different sizes of training dataset in the target city. (taking Beijing as the target city and other cities as the source cities).

| Methods | | 24 h | 72 h | 240 h | 720 h | 2400 h |
|---|---|---|---|---|---|---|
| Single-FT | Tianjin | 62.7 | 62.1 | 53.5 | 50.5 | 45.6 |
| | Shenzhen | 71.3 | 70.7 | 67.6 | 61.0 | 55.2 |
| | Guangzhou | 70.9 | 69.0 | 66.2 | 59.8 | 54.5 |
| Multi-FT | | 63.1 | 61.2 | 55.3 | 52.5 | 43.1 |
| MAML | | 59.4 | 57.8 | 49.4 | 45.1 | 36.2 |
| MetaGAT-LSTM | | 55.8 | 53.7 | 45.3 | 40.2 | 31.4 |

Table 6: RMSE of transfer learning methods over different sizes of training dataset in the target city. (taking Shenzhen as the target city and other cities as the source cities).

| Methods | | 24 h | 72 h | 240 h | 720 h | 2400 h |
|---|---|---|---|---|---|---|
| Single-FT | Tianjin | 77.3 | 76.2 | 72.0 | 66.3 | 58.1 |
| | Beijing | 79.2 | 77.2 | 73.1 | 68.9 | 59.2 |
| | Guangzhou | 65.6 | 60.3 | 51.9 | 49.3 | 43.7 |
| Multi-FT | | 66.3 | 65.1 | 58.9 | 56.7 | 45.4 |
| MAML | | 61.5 | 60.5 | 51.6 | 48.2 | 40.3 |
| MetaGAT-LSTM | | 58.2 | 57.5 | 52.3 | 47.7 | 35.6 |

In the second group of experiments, we execute two experiments by taking Beijing and Shenzhen as the target cities, respectively. We delete most of the data from the target city and only keep a small part for training. With different sizes of the training dataset (in target city), the results of comparing MetaGAT-LSTM with other transfer learning methods are given by Tables 5 and 6. It can be seen all the methods performs better with larger training dataset. Table 5 shows that Single-FT from the Tianjin is better than that from the other two cities. Table 6 shows that Single-FT from Guangzhou is better than that from the other two cities. The climate and geographical location cause similarity

# THANK YOU