



A Review of Machine Learning Models for Predicting Air Quality in Urban Areas

El-Sayed M. El-kenawy

School of ICT, Faculty of Engineering, Design and Information & Communications Technology (EDICT),
Bahrain Polytechnic, PO Box 33349, Isa Town, Bahrain

Emails: sayed.elkenawy@polytechnic.bh

Abstract

Air pollution is a critical environmental issue that threatens almost the world, and public health, ecosystems, and the sustainability of cities are affected by the severe impacts of air pollution. Urbanization and industrialization have been on the run, with escalating pollution levels. Hence, air monitoring and air quality prediction are necessary for such challenges. This review discusses advanced machine learning (ML), deep learning (DL) techniques, and IoT-based study hybrid frameworks for air-quality prediction in urban settings. Integration of different data sets such as meteorological parameters, concentrations of pollutants, and data from satellite imagery, these technologies provide strong and scalable solutions for real-time monitoring and forecasting. Some of the advancements include the use of IoT-enabled sensors, the use of convolutional and recurrent neural networks, and the development of location-specific predictive models. Despite significant evolution, several challenges of data sparsity, computational requirements, and model adaptability remain. This paper casts the technologies as transforming cities into smart and green cities and advancing the cause for continuous innovation and interdisciplinary collaboration to strengthen their effectiveness. These findings add to the advancement of knowledge on air quality prediction methodologies and their crucial role in sustainable urban development.

Keywords: Air pollution; machine learning; deep learning; IoT; air quality prediction; urban sustainability.

1. Introduction

Pollution of the environment through emissions of greenhouse gases is one of the biggest concerns in the world today. It has far-reaching implications for people's health, the environment and even the future sustainability of urban centers. As urbanization and industrialization intensify, cities are experiencing significant difficulties in regulating air quality, which is crucial for supporting populace health. Consequently, ML and DL applications have been proposed as practical approaches for accurately solving the problem in various urban settings. Dominant studies concerning innovations of ML and DL models

combined with IoT data collection, hybrid approach, and location-based prediction algorithms are reviewed in this paper. These methodologies enhance the accuracy of the air quality forecast and offer beneficial information on the design of actual cities, thus helping to design better or more intelligent, sustainable cities. Therefore, using these sophisticated technologies, researchers and policymakers want to solve air quality issues sustainably and support the SDG goals in resilient urban environments.

1. *The Growing Burden of Air Pollution in Urban Centers*

Significant attention should be paid to the problem of air pollution as one of the key environmental problems of the present day, which affects large cities disproportionately. This, coupled with increased emissions from vehicles, industries and other human activities, has made births unsafe and highly charged with health hazards. Habitat areas, primarily associated with densely populated urban regions with significant industrial and vehicle traffic density, are currently ranked highest in the emission of dangerous pollutants, including PM_{2.5}, PM₁₀, NO₂, and SO₂. This has drastic impacts on public health and can lead to respiratory diseases, cardiovascular diseases, and even premature deaths. Besides, secondary effects of these conditions, such as acid rain and climate change, exacerbate the problem even more regarding the environmental cascading effects, which ultimately influence ecosystems and biodiversity. Still, regarding the economic implications, there are also short-magnitude economic costs of air pollution, such as government funds for health-related issues or pollution-fighting strategies. Besides negative impacts on human health and the environment, impairment of air quality in urban areas affects people's conditions and their ability to see, the value of their property, and, worst of all, their quality of life. Meeting this growing burden calls for sound, effective, and efficient solutions to monitor, forecast, and make appropriate policies that help the cities grow sustainably and protect the population's health [1].

2. *Machine Learning Innovations in Air Quality Prediction*

Machine learning (ML) advancements have significantly facilitated modern air quality prediction, giving the industry the tools to approach these complex and constantly changing pollution issues within cities. Conventional statistical paradigms do not come out well while handling big and multiple parameter data sets, a strength of the ML algorithms, as they can capture the complex relationship between different environmental variables. They can all consume data from various sources, such as satellite images, IoT sensors, and historical air quality records, and make accurate and timely forecasts. This has widely incorporated methods like the Random Forest, the Support Vector Machine and the Gradient Boosting, which have proven responsive in determining the most influential pollutants and even the AQI.

More innovations have recently featured longitudinal learning with LSTM systems and CNN to model temporal and spatial dependency in pollutant data accurately. More so, machine learning – deep learning frameworks integrate two or more algorithms to improve predictability as they adopt the strength of two or more frameworks. These innovations also allow for incorporating additional or accessory data, including meteorological factors and urban dynamics, to capture the complex factors affecting air quality.

Furthermore, the ML models provide flexibility for the implementation in different urban settings, including both large megapolises and small cities, which makes them effective instruments for governing air quality. This suggests that these models help produce relevant recommendations, such as the high-risk areas and the predicted trend in pollution levels in a city, which are helpful for city planners/policymakers/other stakeholders. Therefore, this paper concludes that as machine learning advances, air quality prediction will be more effective in delivering sustainable cities [2].

3. *IoT and Hybrid Models in Air Quality Monitoring*

IoT has revolutionized air quality tracking using internet-connected gadgets to manage and analyze data in near real-time. The IoT-based system gives a rich, real-time and efficient control and monitoring over the pollutants having particulate matter such as PM_{2.5}, PM₁₀, nitrogen dioxide (NO₂), Sulfur dioxide (SO₂), Carbon monoxide (CO) and Ozone (O₃). These systems also used meteorological parameters, including temperature and humidity, to enrich data and add depth to primary air quality data. The new generation of vital IoT sensors placed frequently at various spots in urban environments can fill the coverage deficits of monitoring stations and provide contextual data on air pollution.

Similarly, the call for blending different approaches of ML and DL to improve IoT data analysis has brought forth influential hybrid models. For example, models that incorporate Random Forest and Gradient Boosting algorithms build on the solidity of ensemble methodologies; conversely, frameworks that blend CNNs and LSTMs are best suited to seize spatial and temporal features in air quality data models. These kinetic-kinematic models make sense of the big data collected by IoT sensors; they turn data into knowledge.

Moreover, an IoT-powered system is characterized by honest life feedback that can influence deterministic policies such as traffic control measures, industrial emission standards, and public health warnings. Compared to other approaches, hybrid models can predict upcoming pollution levels and suggest prevention measures to the stakeholders. Such innovations increase the accuracy of measuring air quality and stimulate people's participation with the help of application programs and website panels. In general, such synergy of IoT and hybrid models can be considered a significant advance toward effective data-driven MAQS and building more innovative and sustainable cities [3].

4. *The Role of Location-Specific Approaches in Forecasting*

Promised, location-oriented procedures for air quality forecasting garner significant attention as each urbanized area faces unique environmental challenges defined by its natural geography, climate, and socio-economic conditions. While general models do not circulate engines, specific-location methods draw directly from localized variables like emission sources, meteorological conditions, population density, and urban topographical features. These site-relevant approaches refine the understanding of pollution mechanisms and enable an enhanced air quality prediction in any region.

For example, cities that are densely populated with buildings are usually visited by what is called by some as the "street canyon effect," where concentrated polluted air gets trapped between tall buildings, making the local air even more polluted. Unlike inland cities, coastal cities would be governed differently because sea breezes introduce mechanisms that disperse or concentrate pollutants. Location-specific models include these environmental peculiarities in their prediction processes, usually needing unique databases like localized traffic records, industrial emissions, and weather factors.

Machine learning (ML) and deep learning (DL) techniques are best-suited approaches for location-specific forecasting because they can handle different, complicated datasets for region-specific discovery. These are those presented by Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, both well-known for capturing the spatial and temporal dependencies in pollution data. With IoT sensors attached to such models for data input into real-time events, they are more capable of in-time environmental adaption, which forecasts timely and without error.

More than that, these approaches are specifically suited to address location issues that are practical for propagation purposes in important instances. For instance, housing-specialized forecasting models for industrialized regions can give results [4].

5. Towards Sustainable Cities Through Predictive Modeling

Modern predictive modeling is one of the pillars that underpin the dynamic city with sustainable urban development, besides being an avenue that cities use to address complex challenges such as air pollution at the urban level. These models build on vast machine learning (ML) and deep learning (DL) techniques applied over massive datasets to predict air quality with exceptional accuracy. This includes predicting pollutants and high-risk areas, allowing city planners, decision-makers, and others to take preemptive steps to reduce pollution and improve sustainability.

Predictive models are also capable of assimilating different sources of data, including but not limited to IoT-enabled air quality sensors, satellite imagery, and meteorological observations. This tip-of-the-detector approach can help identify pollution patterns, sources, and contributory factors and generate recommendations to reach the SDGs, such as SDG 3: Good Health and Well-being and SDG 11: Sustainable Cities and Communities. For instance, the predictive models can draw in optimized traffic management systems, control industrial emissions, and plan for green infrastructure projects, directly reducing pollution levels for better city air quality.

Forecasting models also involve their users in real-time data updates about air quality and communication through more accessible and popular platforms like mobile applications and web dashboards. Aiding the public's ability to decide about activities during that day, this technology includes route selection when commuting to particular destinations or delaying outdoor plans when pollution levels are high. Some platforms, in addition, reward individuals for using greener means of transport by offering them incentives for reducing emissions so that there is mutualization for being able to commit towards sustainability. It offers cost-effective urban planning for the economy by identifying the critical intervention points that waste resources and maximizing the return of pollution control measures. It also helps long-term planning by simulating the expected outcome of various policies; thus, a city could adopt those strategies, balancing environmental health against economic growth.

In conclusion, predictive modeling is a paradigm tool in the modern mystical involvement toward urban sustainability. Data-driven, evidenced decision-making in the public and resource movement is made possible through these models, eventually making these cities more intelligent, greener, and resilient. As urbanization accelerates, integration between urban planning processes and predictive modeling becomes important."

Current user scores vary with the area: Cont'd - the higher the resolution, the better will be integrity; therefore, once integrated, this point calculates a given area, the higher the resolution, the better as more integration best opted for Times Square-level planning. Predation models also link their users to real-time updates on air quality and communication data on more accessible and popular platforms, such as mobile applications or web dashboards. These empower citizens to decide on their daily activities, from which routes to take when commuting to a particular destination to whether or not outdoor plans should be postponed when pollution levels are at their peak. Besides, some platforms reward incentives for emissions reductions to promote the use of greener modes of transport among people. From the budget point of view, predictive modeling would provide an economical urban planning context by identifying the significant intervention points that cut across the waste of resources and maximize the return of pollution control measures. "It helps in long-term planning, simulating the expected outcome of various policies; therefore, a city could adopt those strategies, balancing environmental health against economic growth. To sum up, predictive modeling is a tool-driven paradigm of modern mystical involvement in urban sustainability. Data-driven evidenced decision-making

in public and resource movements is possible through these models, making all these cities more intelligent, greener, and resilient [5].

In summary, predictive modeling is a breakthrough way forward for the betterment of cities and their sustainability. Such models would assist in encouraging decisions based on solid data to be developed and would help governments, businesses, and citizens work toward a common goal. They would be critical to the bright, green, and resilient cities of the future. Integration of the forecasting model in urban planning will become progressive, i.e., fast times, for strategic development soon for waves of continuity for generations to come.

2. Literature Review

Continued advancements in industrialization and urbanization of most modern cities have put the environment's health in a difficult position, and air pollution has become one big problem today. Currently, the air quality in cities has become one of the biggest challenges to human, ecological, and overall well-being; hence, there is a need for the best methods for estimating air quality. In return, advanced tools like machine learning and deep learning have emerged with the capability to study massive environmental-based data and predict air quality with a higher accuracy rate. The works discussed outline various sources, such as the mentioned deep learning structures built in parallel or combined, the usage of IoT devices for data gathering and assessment, and learning models tailored for particular cities. These improvements not only help improve the accuracy of air quality predictions but can also offer meaningful data for policymakers and urban developers about the increasing issue of air pollution sustainably.

Ambient air pollution has been ranked among the leading environmental risk factors for health since it claims about six and a half million lives annually, most of which occur in urban settings. As noted in the study labeled [6], these megacities are defined as those with populations over 10 million. They are found to be among the most common areas of the highly contaminated environment, which may be enduring several weeks of hazardous AQI levels attributable to concentrated emissions. In light of these factors, this paper will develop an approach for the air quality prediction in such regions, using an attention-enhanced deep multitask spatiotemporal machine learning model based on Long-Short Term Memory Networks. This innovative approach accurately predicts levels of major pollutants such as sulfur dioxide and carbon monoxide while detailing the inherently complicated pattern and variations and guiding policymakers to ameliorate air quality in urban cities.

The general supervision of air pollution is essential because it directly affects the quality of life of all inhabitants of the environment. An analysis in [7] provided an approach to predicting air pollutant levels by applying machine learning methods where the K-Means algorithm was used for clustering. Multinomial Logistic Regression and Decision Tree classifiers were used in the study to classify the results depending on the data available and elaborated in the R programming language. The comparison of these classifiers indicated that the proposed Multinomial Logistic Regression model was better than the assigned Decision Tree model for air pollution analysis.

Noise pollution, in particular, has been identified as impacting human health most in urban centers. As described in [8], a set of Machine and Deep Learning models was built to predict the future concentrations of five key pollutants using data from the Sentinel-5P satellite, climate, and topography. Experiments were undertaken in the Milan metropolitan area to show the accuracy of the models developed initially, where the estimated values of the overall percentage error would be approximately 30% of the next-day pollutant levels.

These models are beneficial as they do not rely on the monitoring stations and, therefore, can be applied where none are present. Furthermore, the dataset made available for public use will trigger similar studies to develop more knowledge about urban air quality changes.

One of the most significant environmental issues that prevail in bright city surroundings is how to address the issue of air pollution. In the research described in [9], the use of pollution data collected in real-time by IoT-based devices for anticipating air quality and improving the performance of traffic control systems was noted as being revolutionary. The study used performance measures, for example, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE), to assess the efficiency of four superior regression methods for forecasting pollution. Using multiple datasets approaches in experiments based on Apache Spark processing unveiled the efficiency of the best-fit model and standalone learning alongside hyperparameter tuning in minimizing error rate in predicting air quality.

External air quality has been one of the most compelling environmental health problems for many years, especially in large, populated areas where the rate of the development of tall buildings and effects from the street canyon lead to increased concentrations of pollutants. In the analysis conducted in [10], a novel hybrid deep learning framework, Deep-AIR, has been proposed to model the complicated inter-localization relationship between air pollution and urban dynamics by integrating a convolutional neural network with an extended short-term memory network. Using 1x1 convolution layers, the framework can learn spatial characteristics of air pollutants and other features such as road density, building density and the effect of the street canyon. Finally, applying Deep-AIR to the cities of Hong Kong and Beijing, we further showed higher performance of Deep-AIR as compared to baseline models, providing comprehensive estimations of air pollutants and their hourly, daily average, and the following 24 hours changes with interpretability of saliency analysis focusing on street canyon effect and meteorological factors.

Consequently, the Internet of Things (IoT) is becoming a popular research area in industry, commerce, education, and others. As we explain in the paper [11], smart cities utilize IoT-based services and applications to encourage sustainable urban lifestyles based on ICT to improve awareness, interactivity and efficiency among city's stakeholders. IoT applications have grown significantly in smart cities; consequently, the generation of a vast amount of data calls for the efficient management of data and the application of measures to allow sustainable development. Utilizing the findings of previous studies in handling forecasting issues in big data, this study develops a new deep learning model using LSTM. It applies it to forecast further air quality values in smart cities. The evaluation results using the lifted glass model establish that the model has utility in various smart city prediction scenarios.

Globalization, extra-urban expansion, and present-day industrial development have affected air quality so severely that airborne pollution is a significant menace to the environment and life. It is argued that air pollution is a 'hidden pandemic' of public health and that forecasting the trend is relevant and crucial, as presented in the paper [12]. The study conducts a comparative analysis of various deep learning-based single-step forecasting models, including Long Short-Term Memory (LSTM) networks, Gated Recurrent Unit (GRU) networks, and a statistical model to predict five major air pollutants. Some of these pollutants include Nitrogen Dioxide (NO₂), Ozone (O₃), Sulphur Dioxide (SO₂), and Particulate Matter (PM_{2.5} and PM₁₀). Using a publicly available dataset from an air quality monitoring station in Belfast city center, the models' predictive performances were evaluated using three metrics: This is a list of four metrics which are Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and R-squared. The conclusions strengthen the impact of deep learning models for air quality forecasting, which can be a solution to the problems concerning urban pollution.

In the past few decades, urban and industrial growth in the newly industrialized and developed countries has deteriorated the quality of air and its effects on life, thus shifting attention to sustainable development. As mentioned in [13], modern industries release liquid droplets, solid particles, and gaseous molecules into the atmosphere, which cause a high density of damaging particulate matter that affects people's health. To this end, new forecasting methods have become important because they can predict the degree of pollution at specific locations and suggest what should be done to prevent it. The paper offers a rich analysis of various aspects of machine learning and deep learning methodologies in air pollution and air quality prediction. It underlines their applicability and usefulness in advancing the existing forecasting models. From this analysis, researchers can quickly obtain architectural ideas for developing air pollution prediction models and discover new trends for improvement with the potential for future issues.

Modernization, increased industrialization rate, and urbanization have led to a high density of airborne pollutants that are hazardous to human life and the environment. As described in the paper [14], the present research focuses on air quality prediction and assessment based on mathematical models that forecast pollutant concentrations. The study combines high levels of artificial intelligence with real-time environmental data to produce sound air quality forecasts. One of the research concern's chief variables is the Indoor Air Quality index, where CO₂ concentration is a proxy for poor IAQ. Therefore, the continuous value of temperature and relative humidity has been used to ascertain how these factors affect air quality. The research has established high dependency, direct air exchange intensity and, at the same time, low heat transfer rates to maintain a high standard of indoor air quality in living spaces with certain geographical constraints.

Environment is an important parameter for evaluating pollution and the efficiency for regulating the greatness of a city and its industries with its population. This paper[,] as pointed out in paper [15] shows that air quality monitoring in urban areas has continued to be problematic, especially with emerging industrialization. Regardless of the constant advances in technology, the problem of polluted air remains among the most considerable threats to humanity: it leads to problems like acid rain, global warming, heart diseases, and skin cancer. The precursor of this work involves predicting the Air Quality Index (AQI) to combat pollution before the conditions worsen. Specifically, the research employs the Central Pollution Control Board (CPCB) of India data to determine AQI by employing the Neural Networks and Support Vector Machines to make an accurate AQI forecast at the Delhi location. The findings underscore how the applied and proposed Machine Learning models enhance prediction accuracy, making the models ready for adoption in other smart cities to optimize air quality control.

Internalizing the degradation of the air quality within India increases the health consequences in that extreme pollution levels cause preventable deaths. An architecture based on deep learning is proposed in the publication introduced as [16] to predict the outdoor air conditions in Delhi based on the meteorological and air quality data obtained from the monitoring stations installed at fourteen sites. About this revised aim, the study employs carbon monoxide (CO) and particulate matter (PM_{2.5} and PM₁₀) concentrations to compute an Air Quality Index (AQI) time series for the 1765 days beginning September 2014 to July 2020. The data set was split into 80% training and 20% testing, while AQI values employed in the model were grouped into 5 classes from satisfactory to severe. The novel DCNN-LSTM system combines the ability of deep convolutional neural networks (DCNN) for feature learning and LSTM for modeling time-series effects. On the level of assessment indicators, with a sensitivity of 0.97, specificity of 0.97, accuracy of 97.48%, and F1 score of 97.48%, the AUC-ROC of 0.97 also indicates the effectiveness of the proposed DCNN-LSTM model compared to other models. Based on these results, employing more sophisticated DL approaches would enhance the quality of AQ forecasts.

Urbanization remains a central sociopolitical process, with cities being the predominant agents of social interaction, with many opportunities and challenges, such as air pollution. As mentioned in [17], filtering urban air quality is critical to achieving enhanced and more sustainable living conditions. The paper presents a disruptive innovation that links smart, sustainable mobility and air quality alternatives. The platform relies on real-time data analytics and information that will benefit not only the commuters but also the planners of the city. It uses innovative computations that help to forecast poor air quality and recommend environment-friendly commuting options using easy-to-use mobile applications and web portals. One of its significant utilities is its ability to predict air quality, allowing users to make appropriate, health-wise commuting decisions. Further, there is an aspect of community engagement since sustainable behaviors are now prompted and motivated by incentives/ reimbursement on any emission reduction. Thus, by collecting and analyzing information about the distribution of people's commute and air quality, the service provides valuable information for city design and pollution control. This approach is a revolution since it shows that densely populated urban centers can be transformed today to become more innovative, greener and sustainable by applying technology, data and people's togetherness.

The AQI is an essential marker in determining the health conditions of urban ecosystems with the help of prediction regarding the traffic system and a listing of pollutants most dangerous to human life. Based on the study identified in [18], different machine learning models have been used to design an air quality prediction system. The research focused on measuring concentrations of seven atmospheric pollutants: PM_{2.5}, PM₁₀, CO, NO₂, NH₃, SO₂, and O₃ for densely populated areas of Varanasi city from January 2019 to October 2023. Data pre-processing was performed to enhance the model's feasibility on the data set. Further, the performance of machine learning models, such as Random Forest, Decision Tree, SVM, Multilayer Perceptron Model KNN, and Linear Regression models, were also tested for AQI prediction. Analyses showed that Random Forest and Decision Tree had the best accuracy of roughly 100%, which gave them a better performance than SVM (98%), Multilayer Perceptron (95%), KNN (93%), and Linear Regression (79%), revealing the models' high capability of air quality forecasting.

Breathing is one of the most significant challenges humanity will face in the future, and it is one of the most acute problems of the world's environment. As discussed in [19], urban air quality worsens upstream, affecting water and land quality. The present work aims to analyze existing machine-learning solutions for estimating and predicting air quality in smart cities: Ahmedabad, Delhi, Lucknow, Gurugram and Mumbai. This paper aims to conduct a comparative study of different machine learning algorithms using random forest regression, decision tree regression, linear regression, xgboost, and a newly developed model that comprises random forest and xgboost. By applying such algorithms, research has established the major pollutants that cause the decline in air quality in different cities and their prevention. The hybrid model showed realistic possibilities for improving the accuracy of air quality predictions within innovative city environments.

Air quality improvement is a critical component of urban environmental health; targets important for sustainable development, such as SDG 3, good health and well-being, and SDG 11-sustainable cities and communities. According to the study in [20], Dehradun has significant challenges concerning air quality and upholding geographical and climatic factors. To overcome these issues, in the current study, the ML technique is applied for predicting AQI in Dehradun city using the historical dataset for two years, namely NO₂, CO, SO₂, PM 2.5, PM 10, and O₃. The models are assessed mainly by the R-squared (R²) statistic, Root Mean Square Error (RMSE) and Mean Absolute Error (MAE). Comparing the tested models' results, the best one was identified as a Lasso Regressor model, characterized by MAPE = 0.0269, MAE = 0.0185, RMSE = 0.0272, R² = 0.9999. All these confirm the capability of the proposed model to predict AQI levels and its ability to help implement early interventions to address air quality deterioration. The study is relevant

for policymakers or urban planners who want to create plans for eradicating air pollution, which is not always a major issue discussed in Himalayan towns such as Dehradun.

Table 1 presents a thorough consolidation of the reviewed literature on models for air quality prediction regarding various methods and innovations employed in different studies and shows significant contributions, including the use of machine learning (ML) and deep learning (DL) techniques such as LSTM networks, Random Forest, and hybrid CNN-LSTM frameworks that are geared at improving prediction accuracy. Data sources include everything that ranges from IoT-enabled sensors and satellite observations to localized datasets from urban monitoring stations, ensuring the diverse applicability and scalability of the models. It shows how real-time data analytics and hybrid modeling approaches can reap great dividends in addressing urban air pollution issues. Also, Table 1 shows that advanced predictive tools are more utilized than just to forecast potential deterioration in air quality through actionable "smart" urban planning and policymaking.

Table 1: Summary of Literature Review

STUDY	KEY FOCUS	METHODOLOGY	DATA SOURCE	FINDINGS
[6]	Predicting AQI in megacities using deep learning	Attention-enhanced deep multitask spatiotemporal model based on LSTM networks	Emissions data from urban megacities	Achieved high accuracy in predicting pollutants like SO ₂ and CO; guided policymaking for urban air quality improvement
[7]	Urban air quality analysis and prediction	K-means clustering, Multinomial Logistic Regression, Decision Tree classifiers	R programming; air pollution data	Multinomial Logistic Regression outperformed Decision Tree inaccuracy
[8]	Urban air pollution forecasting	ML models leveraging satellite data, meteorology, and topography	Sentinel-5P satellite data	Predicted pollutant levels with ~30% error; applicable in areas lacking monitoring stations
[9]	IoT-based air quality prediction in smart cities	Real-time IoT data; regression techniques like MAE and RMSE	IoT devices and Apache Spark datasets	Hybrid models showed superior error minimization and prediction accuracy
[10]	Deep learning framework for urban air quality modeling	Deep-AIR: Hybrid CNN-LSTM framework	Urban data on road density, building density, and meteorological factors	Achieved better performance than baseline models in cities like Hong Kong and Beijing
[11]	Deep learning for air quality forecasting in smart cities	LSTM-based forecasting model	IoT-enabled sensors in urban environments	Effective in various smart city prediction scenarios

[12]	Comparative analysis of air pollutant prediction models	LSTM, GRU, and statistical methods	Dataset from air quality monitoring station in Belfast	LSTM models demonstrated strong predictive performance
[13]	Machine learning in air pollution forecasting	ML/DL approaches for pollution data analysis	Pollutant concentration and meteorological data	Highlighted effectiveness and potential future improvements in forecasting methodologies
[14]	Air quality analysis using mathematical models	Regression models, real-time data integration	CO ₂ concentration, temperature, humidity	Identified the need for balancing heat transfer and ventilation for indoor air quality
[15]	Smart City AQI prediction	Neural Networks, SVM	Central Pollution Control Board data (India)	Random Forest and Decision Tree achieved ~100% accuracy in AQI prediction
[16]	Deep learning for Delhi air conditions	DCNN-LSTM model	Meteorological and AQI data from 14 stations	DCNN-LSTM achieved 97.48% accuracy in classification
[17]	Smart mobility and air quality alternatives	Real-time data analytics for air quality	Urban commuter and pollutant data	Encouraged eco-friendly commuting through predictive tools and incentives
[18]	Machine learning in Varanasi AQI prediction	Random Forest, Decision Tree, SVM	Local pollutant datasets (2019–2023)	Random Forest and Decision Tree outperformed other models in accuracy
[19]	Hybrid ML models in Indian cities	Random Forest, XGBoost, and hybrid models	Data from cities like Ahmedabad, Delhi, and Mumbai	Hybrid models improved prediction accuracy and identified major pollutants
[20]	AQI forecasting in Dehradun city	Lasso Regressor model	Historical AQI data and meteorological parameters	Lasso Regressor achieved a 0.9999 R ² score, demonstrating superior predictive capability.

The literature highlighted in this paper emphasizes that future work in air pollution prediction and management will likely benefit from complex machine learning and deep learning procedures. As these models couple various environmental parameters and use data analysis, they show the potential of enhancing the efficiency of monitoring and predicting air quality in urban environments. Moreover, such innovations as hybrid models and IoT integration give solutions that can be easily applied to various geographical and urban conditions. That is, every analysis has also revealed a challenge to continually improve and develop the SD models due to growth in the environmental conditions and betterment of prediction. Altogether, this

line of research makes the basis for wiser, more environmentally friendly municipalities that contribute to local, national, and global attempts at providing sustainable urban livelihood and enhancing the population's well-being.

3. Discussion

Air pollution has been a challenge everywhere in recent years, but it has become more pronounced in cities concerning their population, industrial activities, and vehicular emissions. If pollution sources are known, high-precision, advanced tools can be used for prediction and mitigation. Machine learning (ML) and deep learning (DL) have changed the game for air quality forecasting by drawing upon complex datasets and giving excellent results regarding actionable insights. These applications integrate various datasets, including IoT-enabled sensors, satellite imagery data, and meteorological parameters, to define city air quality comprehensively.

The problem persists in their faces; the prevailing challenge is air pollution; however, in metropolitan cities, where population density, industrial processes, and vehicle emissions significantly lower air quality, this is way beyond handling it. This part of the problem involves bringing knowledge of the sources of pollution with advanced tools for predicting and managing its impacts. New ML and DL technologies promise a revolution in air quality forecasting by analyzing and interpreting complex datasets into actionable insights at a highly accurate level. Developed models integrate various data sources, such as IoT (Internet of Things) sensors, satellite images, and meteorological parameters, to robustly assess urban air quality [21].

Incorporating IoT technology has further advanced air quality monitoring, real-time data capture, and dynamic analysis. Hybrid models based on ML and IoT solutions will improve prediction accuracy and scalability to suit very complicated problems of city diversity. They will also facilitate intelligent, localized solutions, from traffic optimizations to industry emission controls, thus informing SDGs and building healthier and more sustainable cities [22].

Community engagement plays a vital role in addressing air quality issues. Thus, ML-powered interfacing platforms provide real-time access to air quality updates and personalized recommendations to the general citizenry on reducing pollutant exposure. Making sustainable behavior like using eco-friendly transport available promotes collective responsibility towards pollutants that the community actively participates in releasing into the environment [23].

It should be known that much has progressed, but a lot remains, like sparsity of data, non-adaptiveness of the model, and lack of adequate resources for developing countries. Thus, these stakeholders must put collaborative efforts in place to address those gaps. This brings the air quality management system to a new level of efficiency in making cities sustainable and conducive to healthy living [24], [25].

Quality improvement concerning air management is made possible and progressive by employing machine learning and IoT technologies, which come with integrated innovations and impacts. World Center Real-time monitoring and accurate forecasting coupled with public participation go a long way in minimizing the effects of air pollution and promoting sustainability inefficient urban practices. Despite data sparsity and adaptability to the model, the collective effort of the stakeholders can further improve these systems, making them more reachable and efficient. Their future lies in adopting improved systems to create innovative and green cities. Advancements would continue; hopefully, this can be the reality for future generations living in a sustainable urban setting.

4. Conclusion

It is wrong to assume that air pollution is not a threat to the globe; even more, people in most cities within industrialized or vehicular environments have increasing evidence of rapid urban expansion. This review has shown how applying innovative, advanced machine learning and deep learning methods in air quality prediction and monitoring provides rigorous tools for predicting and monitoring air quality. Employing massive data collected from IoT sensors, satellite images under exposure, and all-weather parameters provides some insights into the effective pollution dynamics while doing dynamic analysis. Such technological advancement would pave the way for more innovation, leading to intelligently responsive urban air quality management systems that can be sidelined on public health protection and how urban living is made sustainable.

Combined with hybrid ML models, IoT technology has become an important development in air quality prediction. It is an enabled IoT system that allows the user to collect real-time data. It is enhanced through a hybrid framework with techniques like CNNs and LSTMs, thus guaranteeing excellent spatial and temporal forecasting abilities. These models can quickly point out the exact spot that predicts high pollution concentrations and forecasts pollution levels in the future. They also offer excellent interventions and make valuable tools for urban planners and policymakers. These systems quickly adapt to almost all geographical and socio-economic aspects, bringing potential scalability and universal applicability.

Although technological intervention has put many advantages in this area, challenges persist. These models are still not widely applicable in resource-scarce settings due to data sparseness, high variability of the environment, and computational resource costs. Further, their dynamism in urban environments necessitates consistent updates and improvements to keep predictive models relevant and accurate. This would require collaboration between researchers, governments, and private stakeholders to get better quality data, more efficient computation, and cheaper solutions.

Thus, this would set the tone for the future- air management has to be transformed into advanced technologies for sustainable cities. They could be expected to better manage air pollution and its associated health risks through strengthening collaboration across disciplines and embedding these requirements into urban planning and policy frameworks. With time, air quality research and advancements will continue to play significant roles in the evolution of these systems that would match the demands of an increasingly dynamic world. If this effort rallies, it can put forward a tangible vision of a sustainable, livable, and healthier city for future generations.

References

- [1] M. Méndez, M. G. Merayo, and M. Núñez, "Machine learning algorithms to forecast air quality: a survey," *Artificial Intelligence Review 2023* 56:9, vol. 56, no. 9, pp. 10031–10066, Feb. 2023, doi: 10.1007/S10462-023-10424-4.
- [2] S. Havemann, P. Kishcha, I. E. Agbehadji, and I. C. Obagbuwa, "Systematic Review of Machine Learning and Deep Learning Techniques for Spatiotemporal Air Quality Prediction," *Atmosphere 2024*, Vol. 15, Page 1352, vol. 15, no. 11, p. 1352, Nov. 2024, doi: 10.3390/ATMOS15111352.

- [3] I. Gryech, C. Assad, M. Ghogho, and A. Kobbane, "Applications of machine learning and IoT for Outdoor Air Pollution Monitoring and Prediction: A Systematic Literature Review," Jan. 2024, Accessed: Jan. 01, 2025. [Online]. Available: <https://arxiv.org/abs/2401.01788v1>
- [4] S. Sellamuthu, Niranjana, Ayush, V. Manori, Y. Chaurasia, and P. Mohan, "A Comprehensive Review of Air Quality Prediction Techniques: Methods, Performance and Challenges," *8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), I-SMAC 2024 - Proceedings*, pp. 1147–1156, 2024, doi: 10.1109/I-SMAC61858.2024.10714587.
- [5] J. Han, W. Zhang, H. Liu, and H. Xiong, "Machine Learning for Urban Air Quality Analytics: A Survey," *J. ACM*, vol. 37, no. 111, p. 35, Oct. 2023, doi: 10.1145/1122445.1122456.
- [6] H. Khan *et al.*, "Novel Approach for Predicting the Air Quality Index of Megacities through Attention-Enhanced Deep Multitask Spatiotemporal Learning," Jul. 2024, Accessed: Dec. 31, 2024. [Online]. Available: <https://arxiv.org/abs/2407.11283v1>
- [7] K. Nandini and G. Fathima, "Urban Air Quality Analysis and Prediction Using Machine Learning," *1st International Conference on Advanced Technologies in Intelligent Control, Environment, Computing and Communication Engineering, ICATIECE 2019*, pp. 98–102, Mar. 2019, doi: 10.1109/ICATIECE45860.2019.9063845.
- [8] G. Blanco, L. Barco, L. Innocenti, and C. Rossi, "Urban Air Pollution Forecasting: a Machine Learning Approach leveraging Satellite Observations and Meteorological Forecasts," *2024 IEEE International Workshop on Metrology for Living Environment, MetroLivEnv 2024 - Proceedings*, pp. 421–426, 2024, doi: 10.1109/METROLIVENV60384.2024.10615605.
- [9] S. Ameer *et al.*, "Comparative Analysis of Machine Learning Techniques for Predicting Air Quality in Smart Cities," *IEEE Access*, vol. 7, pp. 128325–128338, 2019, doi: 10.1109/ACCESS.2019.2925082.
- [10] Y. Han, Q. Zhang, V. O. K. Li, and J. C. K. Lam, "Deep-AIR: A Hybrid CNN-LSTM Framework for Air Quality Modeling in Metropolitan Cities," Mar. 2021, Accessed: Dec. 31, 2024. [Online]. Available: <https://arxiv.org/abs/2103.14587v1>
- [11] I. Kk, M. U. ŐimŐek, and S. zdemir, "A deep learning model for air quality prediction in smart cities," *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017*, vol. 2018-January, pp. 1983–1990, Jul. 2017, doi: 10.1109/BIGDATA.2017.8258144.
- [12] F. Naz *et al.*, "Comparative Analysis of Deep Learning and Statistical Models for Air Pollutants Prediction in Urban Areas," *IEEE Access*, vol. 11, pp. 64016–64025, 2023, doi: 10.1109/ACCESS.2023.3289153.
- [13] M. Bhushan, I. Dawar, S. Sharma, T. K. Bawaniya, U. Anand, and A. Negi, "Air Quality Prediction Using Machine Learning and Deep Learning: An Exploratory Study," *2023 7th International Conference On Computing, Communication, Control And Automation, ICCUBEA 2023*, 2023, doi: 10.1109/ICCUBEA58933.2023.10392048.
- [14] A. Gururaj, V. Nagaraj, A. S. Rajesh, K. N. Achyuth, M. U. Somesh, and B. Ashish Dubay, "Air Quality Prediction and Analysis using Machine Learning," *Proceedings - 2024 5th International Conference on Mobile Computing and Sustainable Informatics, ICMCSI 2024*, pp. 543–550, 2024, doi: 10.1109/ICMCSI61536.2024.00086.

- [15] U. Mahalingam, K. Elangovan, H. Dobhal, C. Valliappa, S. Shrestha, and G. Kedam, "A machine learning model for air quality prediction for smart cities," *2019 International Conference on Wireless Communications, Signal Processing and Networking, WiSPNET 2019*, pp. 452–457, Mar. 2019, doi: 10.1109/WISPNET45539.2019.9032734.
- [16] A. Barthwal and A. K. Goel, "Advancing air quality prediction models in urban India: a deep learning approach integrating DCNN and LSTM architectures for AQI time-series classification," *Model Earth Syst Environ*, vol. 10, no. 2, pp. 2935–2955, Apr. 2024, doi: 10.1007/S40808-023-01934-9/METRICS.
- [17] Z. Shahbazi, Z. Shahbazi, and S. Nowaczyk, "Enhancing Air Quality Forecasting Using Machine Learning Techniques," *IEEE Access*, 2024, doi: 10.1109/ACCESS.2024.3516883.
- [18] P. Chaturvedi, "Air Quality Prediction System Using Machine Learning Models," *Water Air Soil Pollut*, vol. 235, no. 9, pp. 1–16, Sep. 2024, doi: 10.1007/S11270-024-07390-0/METRICS.
- [19] G. Sharma, S. Khurana, N. Saina, Shivansh, and G. Gupta, "Comparative Analysis of Machine Learning Techniques in Air Quality Index (AQI) prediction in smart cities," *International Journal of System Assurance Engineering and Management*, vol. 15, no. 7, pp. 3060–3075, Jul. 2024, doi: 10.1007/S13198-024-02315-W/METRICS.
- [20] I. Dawar, M. Singal, V. Singh, S. Lamba, and S. Jain, "Predicting air quality index using machine learning: a case study of the Himalayan city of Dehradun," *Natural Hazards 2024*, pp. 1–27, Nov. 2024, doi: 10.1007/S11069-024-07027-9.
- [21] A. Gangwar, S. Singh, R. Mishra, and S. Prakash, "The State-of-the-Art in Air Pollution Monitoring and Forecasting Systems using IoT, Big Data, and Machine Learning," *Wirel Pers Commun*, vol. 130, no. 3, pp. 1699–1729, Apr. 2023, doi: 10.1007/s11277-023-10351-1.
- [22] I. Gryech, C. Assad, M. Ghogho, and A. Kobbane, "Applications of machine learning and IoT for Outdoor Air Pollution Monitoring and Prediction: A Systematic Literature Review," Jan. 2024, Accessed: Jan. 01, 2025. [Online]. Available: <https://arxiv.org/abs/2401.01788v1>
- [23] N. Bandara, S. Hettiarachchi, and P. Athukorala, "AirSPEC: An IoT-empowered Air Quality Monitoring System integrated with a Machine Learning Framework to Detect and Predict defined Air Quality parameters," Nov. 2021, Accessed: Jan. 01, 2025. [Online]. Available: <https://arxiv.org/abs/2111.14125v1>
- [24] D. M and R. V, "Discretized Linear Regression and Multiclass Support Vector Based Air Pollution Forecasting Technique," *International Journal of Engineering Trends and Technology*, vol. 70, no. 11, pp. 315–323, Nov. 2022, doi: 10.14445/22315381/IJETT-V70I11P234.
- [25] I. Gryech, C. Assad, M. Ghogho, and A. Kobbane, "Applications of machine learning and IoT for Outdoor Air Pollution Monitoring and Prediction: A Systematic Literature Review," Jan. 2024, Accessed: Jan. 01, 2025. [Online]. Available: <https://arxiv.org/abs/2401.01788v1>