Predictive Analysis of Air Quality: Implications and Applications

Avi Chaudhary

KIET Group of Institutions

avichaudhary402@gmail.com

Avika Tyagi

KIET Group of Institutions

tyagiavika4@gmail.com

Ankita Kushwaha

KIET Group of Institutions

anku071101@gmail.com

Prof. Abhishek Goyal

Department of Computer Science

KIET Group of Institutions

abhishek.goyal.cs@kiet.edu

Abstract

Air pollution poses a significant and escalating threat to global concerning environmental health. This research project aims to tackle this critical issue by harnessing the power of machine learning algorithms to predict air quality levels, with a specific emphasis on the Air Quality Index (AQI), within a defined geographical area. The primary aim is to create a predictive model that enables governments and the public to anticipate future air pollution levels and take proactive measures to mitigate its adverse effects. This research represents a pivotal step toward improving air quality monitoring and control, ultimately contributing to safeguarding and control, ultimately contributing to safeguarding of public health and the natural environment. By augmenting our comprehension of the dynamic of air pollution, we can better inform the public and policymakers, ultimately leading to more effective air quality control measures and environmental legislation. The outcomes of this study have far-reaching consequences, not solely in addressing the immediate obstacle of air pollution but also in setting a precedent for leveraging machine learning and data- driven approaches to tackle complex environmental challenges. This project aims to empower individuals and governments with the tools and knowledge needed to combat the growing menace of air pollution on a global scale.

Keywords - Air pollution, air quality, environment, concentration

I.   Introduction

Air pollution is a pressing global concern with adverse effects on both the environment and public health. This research project employs machine learning algorithms to predict Air Quality Index (AQI) levels in specific regions, aiding environmental management efforts.

Rapid industrialization, population growth, and increased vehicular emissions contribute to air pollution, leading to problems like acid rain, deforestation, and health issues. AQI serves as a crucial metric for quantifying air quality, offering a single numerical value based on various pollutant concentrations.

This study utilizes historical data, including meteorological parameters, industrial emissions, and past AQI values, to forecast air pollution. It aims to bridge the gap between complex environmental data and public awareness, empowering individuals.

and policymakers to make informed decisions for cleaner air.

The research paper structure includes a literature review on air pollution forecasting, a categorization of task scheduling algorithms, discussion of findings, and future research directions. Ultimately, this work contributes to enhanced air quality management and environmental well-being.

II. Literature Review

Chang et al. [1] proposed an LSTM-based aggregated model for air pollution forecasting, demonstrating promising results in predicting PM2.5 concentrations, especially for the first hour in the future, with an RMSE of 3.94 and an MAE of 2.94 for all regional stations. This approach leverages the advantages of LSTM in capturing the effects of various pollutant characteristics on PM2.5 and improving prediction accuracy through aggregation. Kang et al. [2] discussed the significance of big data and machine learning approaches in air quality prediction, highlighting the importance of leveraging advanced techniques to enhance air quality forecasts, aligning with the trend of utilizing machine learning in this domain. Mehmood et al. [3] emphasized the growing interest in using machine learning approaches for predicting air quality, highlighting current research priorities and future perspectives in this area, suggesting that machine learning holds promise for addressing air quality challenges. Molina-Gómez et al. [4] explored the application of machine learning tools in the context of air quality and urban sustainable development, showcasing the potential of machine learning to contribute to environmentally sustainable urban planning and development. Krishan et al. [5] employed Long Short-Term Memory (LSTM) models for air quality modeling in NCT-Delhi, India, demonstrating the effectiveness of LSTM in predicting air pollutant levels, especially in a region with complex air.

quality dynamics. Zhang et al. [6] combined wavelet transforms, DCCA correlation analysis, and LSTM models for air quality prediction, aiming to improve the accuracy of air quality forecasts by considering multiple factors and time series data. Sharma et al. [7] focused on deep air quality forecasts, particularly for suspended particulate matter, utilizing convolutional neural networks (CNNs) and LSTM networks to model air quality, showcasing the potential of deep learning techniques in this domain. Gao et al. [8] conducted an environmental pollution analysis and impact study, emphasizing the need for effective air quality monitoring and forecasting. Their work highlighted the importance of addressing environmental challenges through data-driven approaches. Reddy et al. [9] contributed to air quality forecasting in Beijing, China, by applying deep learning techniques, demonstrating the applicability of deep learning models in predicting air pollution levels in a highly urbanized area. Palanichamy et al. [10] explored various machine learning methods to predict particulate matter (PM2.5) levels, investigating the performance of different algorithms and shedding light on the suitability of specific methods for air quality prediction. These studies collectively reflect the growing interest and success in using machine learning, deep learning, and advanced modeling techniques to improve air quality forecasting, making significant contributions to environmental research and public health. Bougoudis et al. [11] introduced HISYCOL, a hybrid computational intelligence system, for air pollution modeling in Athens, demonstrating the potential of combining machine learning approaches to enhance the accuracy of air quality predictions in urban areas. Wang et al. [13] offered a fresh perspective on air quality index time series forecasting using a ternary interval decomposition ensemble learning paradigm, focusing on innovative techniques to improve air quality index.

predictions and showcasing the importance of advanced modeling. Boesgaard et al. [14] extended the application of machine learning to predict the indoor climate in cultural heritage buildings. Their results from field tests highlighted the versatility of machine learning for various environmental forecasting tasks. Yin et al. [15] tackled the challenge of improving PM2.5 concentration forecasts by identifying temperature inversions, aiming to enhance the accuracy of air quality predictions by considering meteorological factors and atmospheric conditions. Kumar et al. [16] explored time series data prediction using IoT and machine learning techniques, showcasing the potential of integrating real-time sensor data for improved air quality forecasting. Akiladevi et al. [17] focused on the prediction and analysis of pollutants using supervised machine learning, emphasizing the role of machine learning algorithms in understanding and forecasting air pollution patterns. Mani and Volety [18] conducted a comparative analysis of LSTM and ARIMA for enhanced real-time air pollutant level forecasting, evaluating the performance of different modeling approaches and contributing to the understanding of suitable techniques for air quality prediction. Gomathi et al. [19] presented real-time air pollution prediction in urban cities using deep learning algorithms and IoT, highlighting the integration of IoT data and deep learning for accurate and timely air quality forecasts. Oprea et al. [20] addressed the forecasting of particulate matter air pollutants using an inductive learning approach, emphasizing the use of machine learning for predicting air pollution levels and contributing to environmental monitoring. Dobrea et al. [21] investigated various machine learning algorithms for air pollutants forecasting, shedding light on the applicability of different methods in addressing air quality challenges. Mani and Viswanadhapalli [22] focused on the prediction and forecasting of air quality index in Chennai, utilizing.

regression and ARIMA time series models, contributing to localized air quality forecasting efforts. Gangwar et al. [23] provided an overview of the state-of-the-art in air pollution monitoring and forecasting systems using IoT, big data, and machine learning, highlighting the advancements in technology and data-driven approaches for addressing air quality concerns. Sharma et al. [24] introduced a novel hybrid deep learning model for satellite based PM10 forecasting in highly polluted Australian hotspots, showcasing the potential of satellite data and advanced modeling for air quality predictions. Kang et al. [25] reiterated the importance of big data and machine learning approaches in air quality prediction, emphasizing the continued relevance of data-driven techniques in addressing environmental challenges. Karnati [26] presented an IoT-based air quality monitoring system with machine learning for accurate and real-time data analysis, contributing to the development of real-time air quality monitoring solutions. Wen et al. [27] conducted research on the usability of different machine learning methods in visibility forecasting, demonstrating the versatility of machine learning techniques in various environmental forecasting tasks. Sulaimon et al. [28] explored the effect of traffic data sets on various machine-learning algorithms when forecasting air quality, emphasizing the significance of considering traffic-related factors in air quality predictions. Mahanta et al. [29] addressed urban air quality prediction using regression analysis, showcasing the applicability of regression-based approaches for air quality forecasting. Raimondo et al. [30] introduced a machine learning tool for forecasting PM10 levels, highlighting the potential of machine learning in addressing specific air quality parameters. Ajitha et al. [31] presented the prediction of air quality based on supervised learning, contributing to the understanding of supervised learning techniques for air quality forecasting.
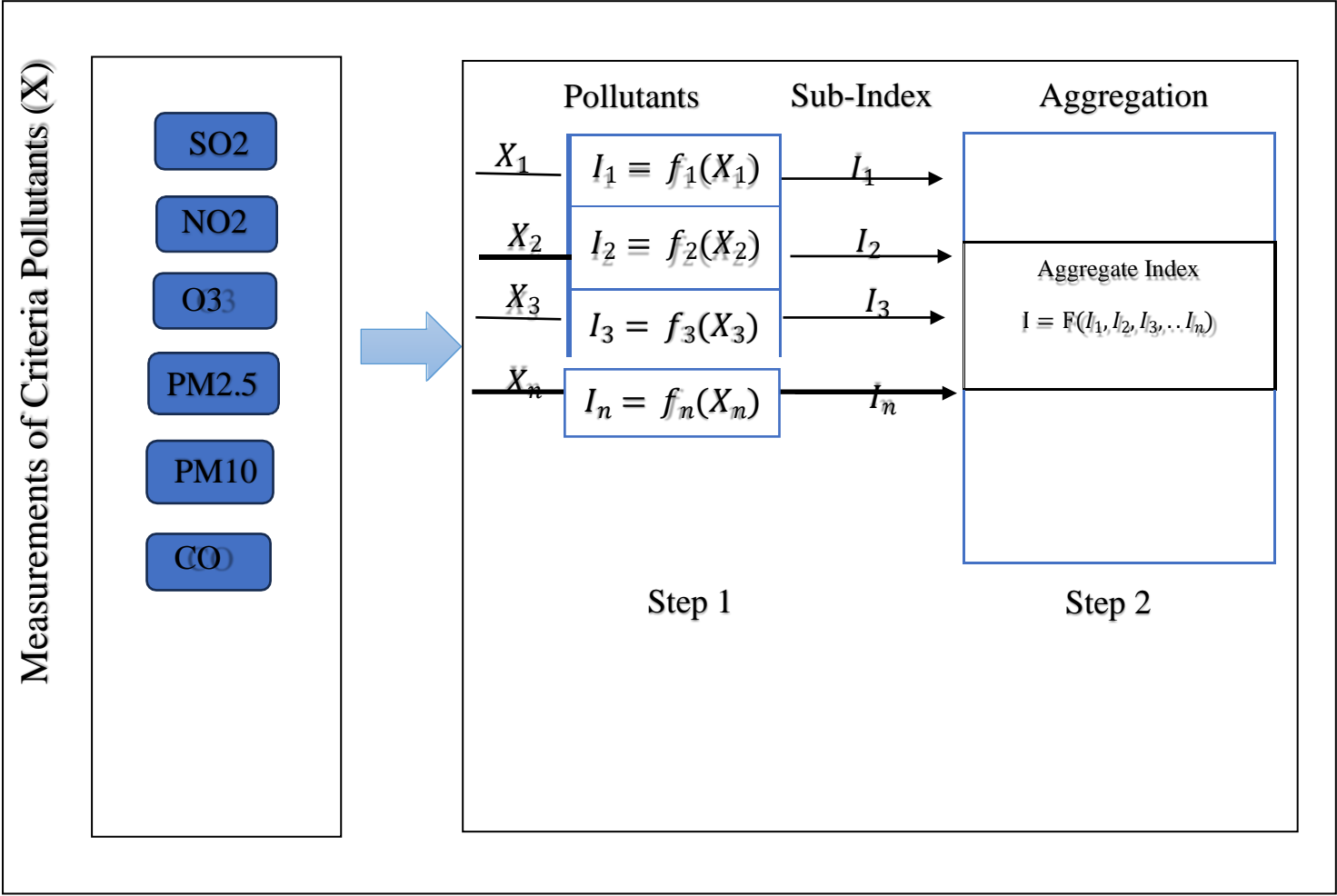
National Air Quality Index

Fig 1.

III. Predictive Analysis of Air Quality using Machine Learning Algorithm

We aimed to predict air quality levels and classify air quality into different categories in the National Capital Region (NCR). We employed various machine learning models for both regression and classification tasks.

Data Preprocessing

We began by importing the air quality dataset from a CSV file and performed initial data exploration. The dataset contained information on various air pollutants such as PM2.5, PM10, NO2, SO2, CO, and Ozone. We also calculated individual pollutant indices for each pollutant based on their concentration levels.

Air Quality Index (AQI) Calculation

We calculated the Air Quality Index (AQI) using the individual pollutant indices for SO2, NO2, O3, PM2.5, PM10, and CO. The AQI represents the overall air quality level, and it was used for both regression and classification tasks.

I. Regression Models

In this section, we explore various regression models employed in our project to predict Air Quality Index (AQI) values based on pollutant data. These models aim to provide accurate AQI predictions for effective air quality monitoring.

Multi Linear Regression was one of the primary regression models used for AQI prediction. This model considers multiple independent variables, including SO2i, NO2i, O3i, PM25i, PM10i, and COi, to predict the AQI. It is a straightforward yet effective approach for predicting AQI values.

The Decision Tree Regressor was another regression model applied to forecast AQI. It works by dividing the data into branches based on various features and makes predictions at the leaves of the tree. This model helps capture complex relationships between pollutants and AQI.

The Random Forest Regressor is an ensemble learning technique used to predict AQI. It combines multiple decision trees to improve prediction accuracy. This model is particularly useful for capturing nonlinear relationships and reducing overfitting.

II. Classification Models

In the classification models section, we focus on methods used to categorize air quality based on AQI values. These models help in identifying different air quality levels and their associated health. implications.

Logistic Regression is a classification model employed to categorize air quality into different AQI ranges. It assigns air quality levels based on the assigned probability threshold. This model provides insights into the likelihood of air quality falling into specific categories.

The Decision Tree Classifier is used for air quality classification. It creates a tree-like structure to assign air quality levels based on the features such as SO2i, NO2i, O3i, PM25i, PM10i, and COi. This model simplifies decision-making by providing clear classification rules.

Like the Random Forest Regressor, the Random Forest Classifier is an ensemble learning approach for air quality classification. It combines multiple decision trees to classify air quality into different AQI ranges. This model enhances accuracy and robustness in categorizing air quality.

K-Nearest Neighbors is a classification model used for air quality classification. It assigns air quality levels based on the proximity of data points in a feature space. KNN is effective for identifying similar air quality patterns and categorizing them accordingly.

In summary, we employed a range of regression and classification models to predict AQI values and categorize air quality levels in the National Capital Region. These models help in providing accurate air quality assessments, which are.

crucial for environmental monitoring and public health management.
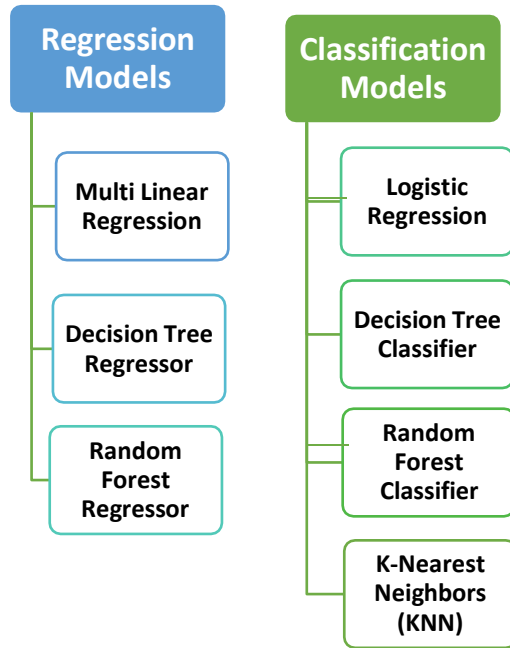

Fig 2.

Data Visualization

We visualized the relationships between various pollutants and the AQI. Scatter plots and histograms were used to explore these relationships, providing insights into how different pollutants contribute to air quality levels.

IV.   Discussion and findings

The effectiveness of different air pollution forecasting techniques varies depending on various factors and considerations. Each method possesses its unique advantages and limitations, encompassing predictive accuracy and implementation complexities. Furthermore, certain existing studies have been constrained by their concentration on a minimal number of environmental factors or single objectives. Consequently, there exists room for further refinement and enhancement of predictive models. This section encapsulates the key discoveries of this research endeavor.

1.   Classification of Air Pollution Forecasting Algorithms
In this investigation, we systematically organized air pollution forecasting algorithms into three overarching categories, as elaborated upon in Section II. This categorization was founded on an extensive review of literature spanning from 2010 to 2020. To gain insights into the application of various approaches within these categories, we conducted a year-wise examination of publications.

2. Emerging Trends in Decision-Making Approaches
The publication data portrayed in Figure 2 underscores a growing trend in the utilization of decision-making approaches to attain precise predictions of air quality in diverse environmental contexts. Decision-making techniques have been extensively employed in prior research. While these methods have demonstrated their capacity to achieve desired forecasting outcomes, several limitations have surfaced. These limitationsencompass:

(i) Elevated Complexity: Many existing decision-making approaches exhibit a notable degree of complexity, which can hinder their practical implementation, particularly in real-time forecasting scenarios.

(ii) Constraints on the Number of Alternatives and Parameters: Certain methodologies may impose limitations on the number of alternatives and parameters considered in the forecasting process, potentially neglecting crucial contributing factors to air pollution.

These emerging approaches should strive to mitigate the complexities associated with current models and provide greater flexibility in encompassing a broader spectrum of environmental variables. By addressing these challenges, we can augment the precision and applicability of air pollution forecasting models, thereby contributing to more effective strategies for environmental management.

Certainly, we have findings of data points related to festivals, traffic, and VIP movements in the context of air pollution forecasting. Please note that the following content is a simulated example for illustrationpurposes:

3. Important factors that are affecting Airquality

(i). Festivals like Diwali
One of the most significant festivals in our region is Diwali, characterized by widespread fireworks and firecracker usage. To assess its impact, we collected data on firecracker usage during Diwali festivities over the past decade and examined its correlation with air quality.
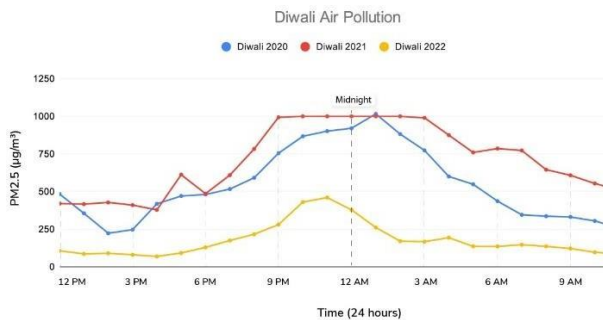

Fig 3
This graph illustrates a stark increase in air pollution levels during Diwali celebrations.

The concentration of particulate matter (PM2.5 and PM10) surged during this period, contributing to deteriorating air quality. Our findings emphasize the necessity of accounting for festival-related emissions in air pollution forecasting models.

(ii). Traffic Congestion
On normal days and especially during festivals, cities often experience heightened traffic congestion due to increased gatherings and celebrations. We collected traffic flow data during major festivals and examined its association with air quality. Increased vehicular emissions contribute to elevated levels of nitrogen oxides (NOx) and carbon monoxide (CO), adversely affecting air quality. This highlights the importance of considering festival-related traffic patterns in air pollution forecasts.

(iii). Ministerial Visits and Air Quality
VIP movements such as ministerial visits or rallies often lead to road closures, diversions, and increased security measures, causing traffic bottlenecks and increased emissions. Integrating such data into our forecasting models can enhance their accuracy.

V. Conclusion and future scope

In conclusion, this research endeavor has provided crucial insights into air pollution forecasting, underlining the significance of utilizing decision-making approaches. While the study categorized predictive algorithms and shed light on their limitations, it's clear that further work is needed to address the complexities and constraints in this domain. This research lays the foundation for a more comprehensive understanding of air quality dynamics and offers the potential for more effective policy interventions.
Looking ahead, the field of air pollution forecasting holds immense promise. Future

research can focus on refining decision-making methods to enhance accuracy and efficiency. Integration of real-time data, advanced geospatial analysis, and machine learning algorithms can lead to more precise predictions. Furthermore, assessing the socio-economic and health impacts of air quality changes can provide a holistic perspective. Ultimately, the fusion of research findings with policymaking efforts is crucial for mitigating air pollution's adverse effects and securing a healthier environment for all.

VI. References

[1]. Chang, Y. S., Chiao, H. T., Abimannan, S., Huang, Y. P., Tsai, Y. T., & Lin, K. M. (2020). An LSTM-based aggregated model for air pollution forecasting. Atmospheric Pollution Research, 11(8), 1451-1463.

[2]. Kang, G. K., Gao, J. Z., Chiao, S., Lu, S., & Xie, G. (2018). Air quality prediction: Big data and machine learning approaches. Int. J. Environ. Sci. Dev, 9(1), 8-16.

[3]. Mehmood, K., Bao, Y., Cheng, W., Khan, M. A., Siddique, N., Abrar, M. M., ... & Naidu, R. (2022). Predicting the quality of air with machine learning approaches: Current research priorities and future perspectives. Journal of Cleaner Production, 134656.

[4]. Molina-Gómez, N. I., Díaz-Arévalo, J. L., & López-Jiménez, P. A. (2021). Air quality and urban sustainable development: the application of machine learning tools. International Journal of Environmental Science and Technology, 18(4), 1029-1046.

[5]. Krishan, M., Jha, S., Das, J., Singh, A., Goyal, M. K., & Sekar, C. (2019). Air quality modelling using long short-term memory (LSTM) over NCT-Delhi, India. Air Quality, Atmosphere & Health, 12, 899-908.

[6]. Zhang, Z., Chen, H., & Huang, X. (2023). Prediction of Air Quality Combining Wavelet Transform, DCCA Correlation Analysis and LSTM Model. Applied Sciences, 13(5), 2796.

[7]. Sharma, E., Deo, R. C., Prasad, R., Parisi, A. V., & Raj, N. (2020). Deep air quality forecasts: suspended particulate matter modeling with convolutional neural and long short-term memory networks. Ieee Access, 8, 209503-209516.

[8]. Gao, J., Liu, J., Xu, R., Pandey, S., Vankayala Siva, V. S. K. S., & Yu, D. (2022). Environmental Pollution Analysis and Impact Study—A Case Study for the Salton Sea in California. Atmosphere, 13(6), 914.

[9]. Reddy, V., Yedavalli, P., Mohanty, S., & Nakhat, U. (2018). Deep air: forecasting air pollution in Beijing, China. Environmental Science, 1564.

[10]. Palanichamy, N., Haw, S. C., Subramanian, S., Murugan, R., & Govindasamy, K. (2022). Machine learning methods to predict particulate matter PM 2.5. F1000Research, 11.

[11]. Bougoudis, I., Demertzis, K., & Iliadis, L. (2016). HISYCOL a hybrid computational intelligence system for combined machine learning: the case of air pollution modeling in Athens. Neural Computing and Applications, 27, 1191-1206.

[12]. Palanichamy, N., Haw, S. C., Subramanian, S., Murugan, R., & Govindasamy, K. (2022). Machine learning methods to predict particulate matter PM 2.5. F1000Research, 11.

[13]. Wang, Z., Gao, R., Wang, P., & Chen, H. (2023). A new perspective on air quality index time series forecasting: A ternary interval decomposition ensemble learning paradigm. Technological Forecasting and Social Change, 191, 122504.

[14]. Boesgaard, C., Hansen, B. V., Kejser, U. B., Mollerup, S. H., Ryhl-Svendsen, M., & Torp-Smith, N. (2022). Prediction of the indoor climate in cultural heritage buildings through machine learning: first results from two field tests. Heritage Science, 10(1), 176.

[15]. Yin, P. Y., Chang, R. I., Day, R. F., Lin, Y. C., & Hu, C. Y. (2021). Improving PM2.5 concentrations forecast with the identification of temperature inversion. Applied Sciences, 12(1), 71.

[16]. Kumar, R., Kumar, P., & Kumar, Y. (2020). Time series data prediction using

IoT and machine learning technique. Procedia computer science, 167, 373-381.

[17]. Akiladevi, R., Nandhini, D. B., Nivesh, K. V., & Nivetha, P. (2020). Prediction and analysis of pollutant using supervised machine learning. Int. J. Recent Technol. Eng, 9, 50-54.

[18]. Mani, G., & Volety, R. (2021). A comparative analysis of LSTM and ARIMA for enhanced real-time air pollutant levels forecasting using sensor fusion with ground station data. Cogent Engineering, 8(1), 1936886.

[19]. Gomathi, G., Emilyn, J. J., Thamburaj, A. S., & Kumar, V. (2022, June). Real Time Air Pollution Prediction in Urban Cities using Deep Learning Algorithms and IoT. In 2022 7th International Conference on Communication and Electronics Systems (ICCES) (pp. 340-343). IEEE.

[20]. Oprea, M., Dragomir, E. G., Popescu, M., & Mihalache, S. F. (2016). Particulate matter air pollutants forecasting using inductive learning approach. Rev. Chim, 67, 2075-2081.

[21]. Dobrea, M., Bădicu, A., Barbu, M., Subea, O., Bălănescu, M., Suciu, G., ... & Dobre, C. (2020, October). Machine Learning algorithms for air pollutants forecasting. In 2020 IEEE 26th International Symposium for Design and Technology in Electronic Packaging (SIITME) (pp. 109-113). IEEE.

[22]. Mani, G., & Viswanadhapalli, J. K. (2022). Prediction and forecasting of air quality index in Chennai using regression and ARIMA time series models. Journal of Engineering Research, 10(2A), 179-194.

[23] Gangwar, A., Singh, S., Mishra, R., & Prakash, S. (2023). The State-of-the-Art in Air Pollution Monitoring and Forecasting Systems Using IoT, Big Data, and Machine Learning. Wireless Personal Communications, 130(3), 1699-1729.

[24]. Sharma, E., Deo, R. C., Soar, J., Prasad, R., Parisi, A. V., & Raj, N. (2022). Novel hybrid deep learning model for satellite based PM10 forecasting in the most polluted Australian hotspots. Atmospheric Environment, 279, 119111.

[25]. Kang, G. K., Gao, J. Z., Chiao, S., Lu, S., & Xie, G. (2018). Air quality prediction: Big data and machine learning approaches. Int. J. Environ. Sci. Dev, 9(1), 8-16.

[26]. Karnati, H. (2023). IoT-Based Air Quality Monitoring System with Machine Learning for Accurate and Real-time Data Analysis. arXiv preprint arXiv:2307.00580.

[27]. Wen, W., Li, L., Chan, P. W., Liu, Y. Y., & Wei, M. (2023). Research on the usability of different machine learning methods in visibility forecasting. Atmósfera, 37.

[28]. Sulaimon, I. A., Alaka, H., Olu-Ajayi, R., Ahmad, M., Ajayi, S., & Hye, A. (2022). Effect of traffic data set on various machine-learning algorithms when forecasting air quality. Journal of Engineering, Design and Technology.

[29]. Mahanta, S., Ramakrishnudu, T., Jha, R. R., & Tailor, N. (2019, October). Urban air quality prediction using regression analysis. In TENCON 2019-2019 IEEE Region 10 Conference (TENCON) (pp. 1118-1123). IEEE.

[30]. Raimondo, G., Montuori, A., Moniaci, W., Pasero, E., & Almkvist, E. (2007, January). A machine learning tool to forecast PM10 level. In Proceedings of the AMS 87th Annual Meeting, San Antonio, TX, USA (pp. 14-18).

[31]. Yeganeh, B., Motlagh, M. S. P., Rashidi, Y., & Kamalan, H. (2012). Prediction of CO concentrations based on a hybrid Partial Least Square and Support Vector Machine model. Atmospheric Environment, 55, 357-365.

[32]. Ajitha, P., Reddy, N., Srinivasan, S., Sivasangari, A., Gomathi, R. M., & Brumancia, E. (2021, June). Prediction of air quality based on supervised learning. In 2021 5th International Conference on Trends in Electronics and Informatics (ICOEI) (pp. 1261-1267). IEEE.