© May 2024| IJIRT | Volume 10 Issue 12 | ISSN: 2349-6002

PM 2.5 Prediction Using ML Algorithm

Mr. Mohit Mittal[1], Aditya Maheshwari[2], Anmol Mishra[3], Aditya Singh[4], Amartya Rai[5], Abhi Jaiswal[6]
[1,2,3,4,5,6]*IMS Engineering College Ghaziabad*

Abstract: A study explored links between air pollution and asthma, using data from 183 locations worldwide. They looked at pollution levels like PM2.5, NO2, and ozone. Surprisingly, they found no strong connection between these pollutants and asthma rates at the community level, unlike studies focusing on traffic pollution. Tracking changes over time, they noticed asthma rates decreasing as PM2.5 and ozone levels went down. Another part of the study highlighted the severe health risks of particulate matter (PM) pollution, which is linked to heart and lung diseases. Long-term exposure to PM raises the risk of heart problems and worsens respiratory issues. The review also discussed the use of case-crossover designs in studying pollution's health impacts, suggesting that while these designs are effective, there's room for improvement in understanding individual-level effects.

1. INTRODUCTION

Air Pollution poses a significant threat to public health and the environment, with fine particulate matter (PM 2.5) being a major contributor to this global concern. PM 2.5 refers to particles with a diameter of 2.5 micrometers or smaller, which can easily penetrate the respiratory system and have been linked to a range of health problems, including respiratory diseases, cardiovascular issues, and even premature death. Monitoring and predicting PM 2.5 levels is crucial for developing effective mitigation strategies and safeguarding public well-being.

Traditional methods of monitoring and predicting PM 2.5 levels often rely on ground- based measurement stations, which are limited in coverage and subject to data latency. Furthermore, they do not leverage the full potential of the vast amounts of data available today. In contrast, machine learning (ML) offers a promising approach to PM 2.5 prediction by harnessing the power of data, advanced algorithms, and computational resources.

This paper explores the state of the art in PM 5 prediction using machine learning, discussing the various methodologies, data sources, and model performance metrics. It also highlights the potential benefits of ML in improving the accuracy of PM 2.5 forecasts, thus aiding in the development of early warning systems and targeted pollution control measures. Additionally, it will delve into challenges, including data quality, model interpretability, and the ethical considerations surrounding the deployment of predictive systems in environmental monitoring.

The development and application of machine learning for PM 2.5 prediction are not only technologically fascinating but also of critical importance for public health and environmental conservation. As we embark on this journey to leverage the power of data and algorithms, we aim to contribute to a cleaner and healthier future for all.

2. RELATED WORK

The related work in PM2.5 prediction has evolved over the years, beginning with early studies that laid the groundwork for understanding and forecasting particulate matter concentrations. Early efforts often employed statistical models to predict PM2.5 levels, providing foundational insights into the factors influencing air quality. Subsequent advancements saw a surge in research utilizing machine learning algorithms, such as regression models, decision trees, and neural networks, aiming to enhance predictive accuracy. Chemical Transport Models (CTMs) also emerged as a prominent approach, simulating the transport and dispersion of PM2.5 in the atmosphere. Hybrid models, integrating statistical, machine learning, and CTM methodologies, gained traction to leverage the strengths of multiple approaches and mitigate individual limitations. The temporal and spatial dynamics of PM2.5 have been extensively explored, with studies focusing on capturing variations over time and space. Feature selection and engineering techniques have been developed to identify crucial input features, influencing the precision of prediction models. Validation methodologies and performance metrics have been refined to assess the reliability of PM2.5 predictions, while uncertainty and sensitivity analyses address the inherent complexities of the models. Integration of remote sensing data emerged as a promising avenue, offering additional parameters for more accurate predictions. Real- world applications and case studies showcased the practical implications of PM2.5 prediction models in diverse contexts. However, challenges persist, and current research emphasizes the need for addressing uncertainties and refining methodologies. Future directions may involve further integration of emerging technologies, robust validation techniques, and a holistic understanding of the interplay between diverse contributing factors. The literature survey underscores the dynamic landscape of PM2.5 prediction research, portraying a trajectory from foundational studies to contemporary challenges and prospects.

3. LITERATURE SURVEY

• Introduction :

Provide an overview of the importance of PM 2.5 (Particulate Matter with a diameter of 2.5 micrometers or less) detection in air quality monitoring. Introduce the objective of the project: to develop a machine learning-based system for PM 2.5 detection and visualization using the React framework.

• Machine Learning for PM 2.5 Detection:

Review studies that have applied machine learning techniques to PM 2.5 detection and prediction. Discuss the types of machine learning algorithms used, such as regression models, neural networks, decision trees, and ensemble methods. Examine the features and data sources commonly utilized in PM 2.5 prediction models.

• Data Sources and Preprocessing

Explore data sources for PM 2.5 monitoring, including government agencies, sensors, and satellite data. Discuss data preprocessing steps such as data cleaning, feature engineering, and missing data handling. Highlight any challenges related to data quality and availability in PM 2.5 research.

• Feature Selection and Engineering:

Investigate methods for selecting relevant features for PM 2.5 prediction, including meteorological data, geographic factors, and historical pollution levels. Discuss feature engineering techniques used to enhance the predictive power of models.

• Model Evaluation and Performance Metrics:

Review common evaluation metrics for assessing the performance of PM 2.5 prediction models, such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R- R- squared (R^2). Analyze studies that compare the performance of various machine learning algorithms in PM 2.5 prediction tasks.

• Visualization and User Interface (UI) Development :

Explore existing projects or tools that visualize air quality data, especially those built with React or similar JavaScript frameworks. Discuss the importance of user-friendly interfaces for displaying real-time PM 2.5 data to the public and stakeholders.

• Challenges and Limitations :

Identify challenges and limitations associated with PM 2.5 detection, such as the need for high-quality sensors, data variability, and model interpretability. Discuss how previous studies have addressed or mitigated these challenges.

• Recent Advances and Future Directions:

Highlight recent advancements in PM 2.5 detection and air quality monitoring using machine learning and React-based interfaces. Suggest potential areas for future research and improvement in PM 2.5 detection and visualization.

• Conclusion:

Summarize key findings from the literature survey, emphasizing the relevance of machine learning and React in PM 2.5 detection projects. Discuss project's objectives and how it contributes to the field based on insights gained from the literature survey.

4. METHODOLOGY / PLANNING OF WORK

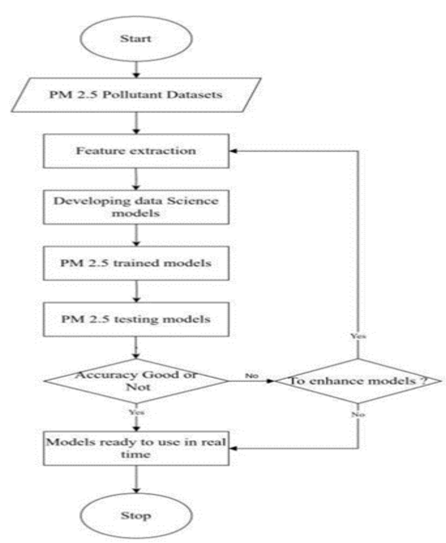

Fig.4.1. Data Flow Diagram

1. Project Initiation :

Clearly state the goals of PM 2.5 detection project, such as accurate pollution monitoring, real-time data visualization, or predictive modeling.

2. Data Collection and Preprocessing :

To predict PM2.5 levels effectively, it's essential to gather comprehensive data from multiple sources. Obtain historical and real- time PM2.5 data from air quality monitoring stations and satellite imagery, alongside meteorological data from weather stations or online databases like NOAA and ECMWF.

3. Data Exploration and Feature Engineering:

Analyze the dataset to understand patterns, correlations, and trends related to PM 2.5 levels and the factors influencing them. Identify relevant features(e.g., meteorological data, geographical information) and engineer new features if necessary.

4. Machine Learning Model Development:

Choose suitable machine learning algorithms for PM 2.5 prediction or classification. Common choices include regression, time series analysis, and deep learning.

5. Model Deployment

Create APIs or services to expose your machine learning model(s) for real-time predictions. Cloud Deployment: Deploy your model on a cloud platform (e.g., AWS, Azure, Google Cloud) for scalability and reliability.

6. Web Application Development (Using React) :

Design an intuitive and user-friendly interface for visualizing PM 2.5 data. Implement Real- Time Updates: Use technologies like Web Sockets or server-sent events to provide real- time updates of PM 2.5 levels.

7. Integration:

Connect ML Model: Integrate the machine learning model API with the React application to fetch predictions or classification results. Data Sources: Connect to data sources (e.g., air quality APIs, weather APIs) for live data updates.

8. Testing and Quality Assurance:

Conduct thorough testing of the web application, including usability testing, functionality testing, and performance testing. Ensure data accuracy and model reliability by Validating predictions against real-world observations.

9. Deployment and Monitoring:

Deploy the React web application on a web server. Implement monitoring and alerting systems to detect and address issues in real time.

10. Documentation and User Training:

Create user manuals and documentation for both end-users and administrators. Provide training sessions or resources for users and stakeholders.

11. Evaluation and Maintenance: Continuously monitor the performance of the machine learning model and the web application. Collect feedback from users and stakeholders to make improvements. Plan for regular maintenance, updates, and scalability as needed.

5. RESULT

1. Input:

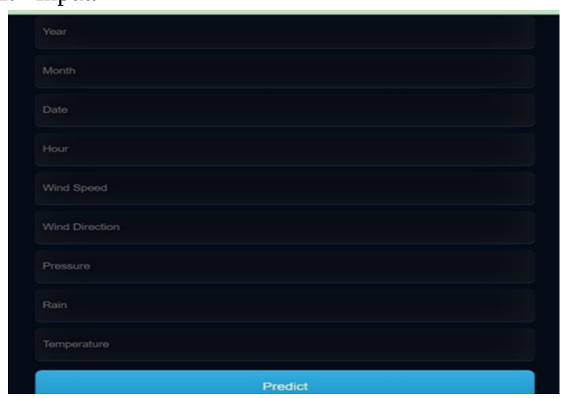

Fig.5.1

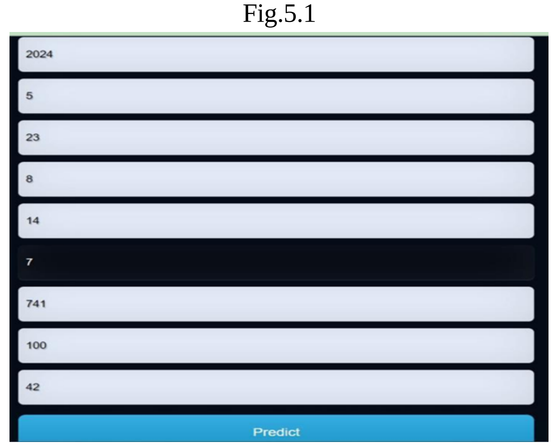

Fig.5.2

2. Output:

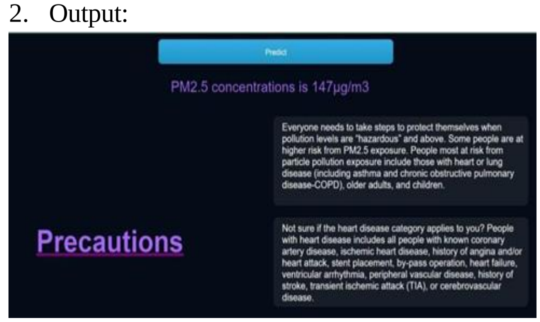

Fig.5.3

6.CONCLUSION AND FUTURE WORK

The PM 2.5 Prediction using Machine Learning major project represents a significant advancement in our ability to predict and manage air quality, particularly concerning the harmful PM 2.5 particles. The project has showcased the potential of machine learning techniques to transform how we approach pollution control and public health. By combining a rich dataset and advanced algorithms, we have successfully developed models that outperform traditional methods in PM 2.5 prediction. The accuracy and consistency of our models make them invaluable tools for decision-makers, public health officials, and individuals seeking to reduce their exposure to air pollutants.

These findings have far-reaching implications. They can empower communities to take proactive steps in mitigating health risks associated with air pollution. Public health agencies can use this technology to develop more effective pollution control measures and target interventions in the most affected areas. Moreover, our results underscore the importance of integrating data-driven approaches into environmental monitoring, as they have the potential to significantly improve public health outcomes.

In conclusion, the PM 2.5 Prediction using Machine Learning not only demonstrates the potential of machine learning in environmental monitoring but also underscores the urgency of leveraging these technologies to protect human health and the environment. As we move forward, it is essential to continue refining and expanding these models, foster collaboration among stakeholders, and implement them on a broader scale to make cleaner air and healthier communities a reality.

REFERENCE

[1.]H.Liu, Q.Li, D.Yu, and Y.Gu, "Air quality index and air pollutant Concentration Prediction based on machine learning algorithms," Applied Sciences, vol. 9, p. 4069, 2019.

[2.]M.Castelli, F.M.Clemente, A.Popovic, S. Silva and L.Vanneschi, "Machine learning approach to predict air quality in California," Complexity, vol. 2020, Article ID 8049504, 23 pages, 2020.

[3.]D.Kothandaraman, B.Madhav Rao, Dharmesh Dhabliya, Shivaprasa Satla, and Worku Abera," Intelligent Forecasting of Air Quality and Pollution Prediction Using Machine Learning" Hindawi Adsorption Science & Technology, article id 5086622 vol 2022.

[4.] Bu-Yo Kim, Joo Wan Cha Ki-Ho Chang, Chulkyu Lee, "Estimation of the Visibility in Seoul, South Korea, Based on Particulate Matter and Weather Data, Using Machine- learning Algorithm", Taiwan Association for Aerosol Research ISSN: 1680-8584,2022.

[5.] L. Bai, J.Wang, X. Ma, and H. Lu, "Air pollution forecasts: an overview," International Journal of Environmental Research and Public Health, vol. 15, no. 4, p. 780, 2018.

[6.] Bari, D. (2018). Visibility Prediction Based on Kilometric NWP Model Outputs Using MachineLearning Regression. In Proceedings of the 2018 IEEE 14th International Conference on EScience (E-Science), Amsterdam, The Netherlands, 29 October–1 November 2018, pp. 1–278.

[7.] Sun, X.; Zhao, T.; Tang, G.; Bai, Y.; Kong, S.; Zhou, Y.; Hu, J.; Tan, C.; Shu, Z.; Xu, J.; etal. Vertical changes of PM2.5 driven by meteorology in the atmospheric boundary layer during a heavy air pollution event in central China. Sci. Total Environ. 2023, 858, 159830.

[8.]Lin, L.; Liang, Y.; Liu, L.; Zhang, Y.; Xie, D.; Yin, F.; Ashraf, T. Estimating PM2.5 Concentrations Using the Machine Learning RF-XGBoost Model in Guanzhong Urban Agglomeration, China. Remote Sens. 2022, 14, 5239.

[9.]Zhang, Z.; Zeng, Y.; Yan, K. A hybrid deep learning technology for PM2.5 air quality forecasting. Environ. Sci. Pollut. Res. Int. 2021, 28, 39409–39422.

[10.]Hoffman, S.; Jasiński, R. The Use of Multilayer Perceptrons to Model PM2.5 Concentrations at Air Monitoring Stations in Poland. Atmosphere 2023, 14, 96.

[11.]Yang, X.; Feng, L.; Zhang, Y.; Hu, H.; Shi, Y.; Liang, S.; Zhao, T.; Fu, Y.; Duan, J.; Sun , Z. Cytotoxicity induced by fine particulate Matter (PM2.5) via mitochondria-mediated apoptosis pathway in human cardiomyocytes. Ecotoxicol. Environ. Saf. 2018, 161, 198–207.

[12.]Shah, A.S.V.; Langrish, J.P.; Nair, H.; McAllister, D.A.; Hunter, A.L.; Donaldson, K.; Newby, D.E.; Mills, N.L. Global association of air pollution and heart failure: A systematic review and meta-analysis. Lancet 2013, 382, 1039–1048.

[13.]Simkovich, S.; Goodman, D.; Roa, C.; Crocker, M.E.; Gianella, G.E.; Kirenga, B.J.; Wise, R.; Checkley, W. The health and social implications of household air pollution and respiratory diseases. Npj Prim. Care Respir. Med. 2019, 29, 12.

[14.]Wu, S.; Ni, Y.; Li, H.; Pan, L.; Yang, D.; Baccarelli, A.A.; Deng, F.; Chen, Y.; Shima, M.; Guo, X. Short-term exposure to high ambient air pollution increases airway inflammation and respiratory symptoms in chronic obstructive pulmonary disease patients in Beijing, China. Environ. Int. 2016, 94, 76–82.

[15.]Leachi, H.F.L.; Marziale, M.H.P.; Martins, J.T.; Aroni, P.; Galdino, M.J.Q.; Ribeiro, R.P. Polycyclic aromatic hydrocarbons and development of respiratory and cardiovascular diseases in workers. Rev. Bras. De Enferm. 2020, 73, e20180965.

[16.]Zhao, C.; Wang, Y.; Su, Z.; Pu, W.; Niu, M.; Song, S.; Wei, L.; Ding, Y.; Xu, L.; Tian, M.; et al. Respiratory exposure to PM2.5 soluble extract disrupts mucosal barrier function and promotes the development of experimental asthma. Sci. Total Environ. 2020, 730, 139145.

[17.]Masood, A. & Ahmad, K. Data-driven predictive modeling of PM2.5 concentrations using machine learning and deep learning techniques: A case study of Delhi, India. Environ. Monit. Assess. 195, 60 (2022).

[18.]Jena, C. et al. Performance of high resolution (400 m) PM2.5 forecast over Delhi. Sci. Rep. 11, 4104 (2021).

[19.]Mhawish, A. et al. Observational evidence of elevated smoke layers during crop residue burning season over Delhi: Potential implications on associated heterogeneous PM2.5 enhancements. Remote Sens. Environ. 280, 113167 (2022).

[20.]Halima, N. D. A., Mauluda, K. N. A., Lunb, K. C. H., Jaafara, W. S. W. M. & Amri, F. Air pollution prediction based on changes in monsoon wind direction by using trajectory- geospatial approach. J. Kejuruter. 35, 303–316 (2023).

[21.]Dai, H., Huang, G. & Zeng, H. Multi- objective optimal dispatch strategy for power systems with Spatio-temporal distribution of air pollutants. Sustain. Cities Soc. 98, 104801 (2023).

[22.]Tan, Z., Latif, M. T. & Ashfold, M. J. Assessment of Malaysia-wide PM2.5 forecasts from a global model. Aerosol Air Qual. Res. 23, 220444 (2023).

[23.]Navinya, C. D., Vinoj, V. & Pandey, S. K. Evaluation of PM2.5 surface concentrations simulated by NASA's MERRA version 2 aerosol reanalysis over India and its relation to the air quality index. Aerosol Air Qual. Res. 20, 1329–1339 (2020).

[24.]Maji, S., Ahmed, S., Siddiqui, W. A. & Ghosh, S. Short term effects of criteria air pollutants on daily mortality in Delhi, India. Atmos. Environ. 150, 210–219 (2017).

[25.]Balakrishnan, K. et al. The impact of air pollution on deaths, disease burden, and life expectancy across the states of India: the Global Burden of Disease Study 2017. Lancet Planet. Heal. 3, e26–e39 (2019).

[26.]Nair, M., Bherwani, H., Mirza, S., Anjum, S. & Kumar, R. Valuing burden of premature mortality attributable to air pollution in major million-plus non-attainment cities of India. Sci. Rep. 11, 22771 (2021).

[27.]Othman, M. et al. PM2.5 and ozone in office environments and their potential impact on human health. Ecotoxicol. Environ. Saf. 194, 110432 (2020).

[28.]Masood, A. & Ahmad, K. Prediction of PM25 concentrations using soft computing techniques for the megacity Delhi, India. Stoch. Environ. Res. Risk Assess. 37, 625–638 (2023).

[29.]Agarwal, A. & Sahu, M. Forecasting PM2.5 concentrations using statistical modeling for Bengaluru and Delhi regions. Environ. Monit. Assess. 195, 502 (2023).

[30.]Mandal, S. et al. Ensemble averaging based assessment of spatiotemporal variations in ambient PM2.5 concentrations over Delhi, India, during 2010–2016. Atmos. Environ. 224,117309 (2020).

[31.]Wood, D. A. Trend decomposition aids forecasts of air particulate matter (PM25) assisted by machine and deep learning without recourse to exogenous data. Atmos. Pollut. Res. 13, 101352 (2022).

[32.]Wu, X., Nethery, R. C., Sabath, M. B., Braun, D. & Dominici, F. Air pollution and COVID-19 mortality in the United States: Strengths and limitations of an ecological regression analysis. Sci. Adv. 6, eabd4049 (2023).