A Survey on Data-Driven Air Quality Prediction Using Machine Learning

M.Thanigavel[1], G Gayathri[2], M Saivardhan[3], P Naveen[4], V Leelavathi[5], N Guru Charan[6], B Venkateswarlu[7]

[1]*M.Tech ., Asst. Professor in CSE Department., Gokula Krishna College of Engineering*

[2,3,4,5,6,7] *IV B.Tech (Student) Dept of CSE, Gokula Krishna College of Engineering,*

***Abstract*—The project aims to ensure optimal air quality in targeted urban areas by employing a sophisticated air quality monitoring system that collects data on contaminants from various locations. Pollutants like particulate matter (PM), nitrogen dioxide (NO2), sulfur dioxide (SO2), ozone (O3), carbon monoxide (CO), and others, accumulate in the atmosphere, causing a deterioration in air quality and posing serious risks to both human health and the environment. Severe pollution detection and traffic routing may benefit from the prediction of the air pollution index. Using sophisticated algorithms to represent the complex connections between these factors is a promising area in machine learning. Comparing different machine learning techniques, including SARIMA, SVM, and LSTM, in order to forecast the Ahmedabad air quality index for Gujarat, India, is the aim of this piece of work. The project leverages advanced machine learning algorithms to analyze historical data on air quality and predict air quality index. By accurately predicting air quality levels, the project can help individuals and authorities take preventive measures to reduce exposure to pollutants and improve public health. These models measure local air contamination and collect data on pollutant concentrations. The proposed research uses various machine learning models to predict air quality, including Random Forest (100%), Logistic Regression (79%), Decision Tree (100%), Support Vector Machine (93%), Linear SVC (98%), K-Nearest Neighbor (99%), and Multinomial Naïve Bayes (52%). A user-friendly Django-based web interface offers an accessible platform for users to monitor air quality in real-time, based on the two best-performing models: Random Forest and Decision Tree techniques**

***Index Terms*—Air Quality Index, Linear Regression, Artificial Neural Network, Decision tree regression, Air pollution Detection, Machine learning, Air quality, statistical analysis, Air pollutant feature extraction, Real-time air quality prediction**

INTRODUCTION

Air pollution remains a critical global environmental issue, with both short-term and long-term consequences for public health, ecosystems, and the climate. It greatly contributes to the occurrence of breathing problems, heart issues, stroke, and cancer, and it is also a major driver of climate change. Efforts to combat air pollution continue through regulatory measures, technological innovations, and public awareness campaigns, with the goal of achieving cleaner and healthier air for all. Reducing air pollution not only saves lives but also ensures future generations will live in a healthier and more sustainable environment. Serious population growth and an increase in vehicles will result in a number of serious environmental issues, including acid rain, deforestation, air and water pollution, toxic material emissions, and more. In order to meet the demands of an expanding population, there has been a sharp growth in industry. This might lead to a range of enterprises releasing dangerous substances into the environment, which would significantly exacerbate the worldwide issue of pollutants in the air in urban areas. This indicates that the air that humans breathe is contaminated with several dangerous gases and particles that have a negative impact on human health. To overcome the constraints of conventional primary air quality forecasting models and the shortcomings in meteorological data exploitation, this study conducts an analysis of the hierarchical impact of various meteorological factors on air quality utilizing a random forest (RF) model. Subsequently, sophisticated data mining techniques, encompassing machine learning algorithms, neural networks, and various regression based prediction models, are deployed to delineate the interrelations among primary weather forecast data, actual meteorological measurements, and air pollutant concentrations. In the concluding phase, leveraging the predictive performance and evaluative metrics of the established models, a comparative analysis is undertaken to highlight the merits, limitations, and contextual applicability of each model, thereby providing a

nuanced understanding of their operational efficacy in air quality prediction.

LITERATURE SURVEY

Mangayarkarasi, R., Vanmathi, C., Khan, M. Z., Noorwali, A., Jain, R., & Agarwal, P. (2021).COVID19: Forecasting Air Quality Index and Particulate Matter (PM2. 5). Computers, Materials & Continua, 67(3). The concentrations of different air pollutants are measured to establish the level of air quality. To raise people's consciousness, It is necessary to have a system of automation that predicts the quality. In several Indian towns, air pollution has decreased as a result of the COVID-19 pandemic and the limitations it has placed on anthropogenic activity. The maximum band for each pollutant at any given moment is provided by the overall air quality index (AQI). PM2.5, or fine particulate matter, is smaller than 2.5 micrometers and can have harmful effects on those with asthma and other cardiovascular disorders when inhaled. Inventors Weimin Zhang and Mengying Zhang, Patent No: US20190113445A1, have introduced a unique system for continuous monitoring of air pollution in a specific area. The system comprises multiple monitoring devices strategically placed at various heights and intervals. Halsana, S. (2020). Air quality prediction model using supervised machine learning algorithms. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 8, 190-201. The "largest environmental health threat in the world" is air pollution, which is responsible for 7 million fatalities globally each year. The only practical way to address this global problem is to use algorithms based on machine learning to forecast the AQI (Air The quality Index), which can inform the public about the state of the air in a particular area and enable the government to appropriate action to improve the quality of the air in the future. This project's main goal is to forecast the AQI depending on the level of concentration. Machine learning regression strategies for PM2.5 prediction are surveyed in the work ”Forecasting air pollution particulate matter (PM2.5) using machine learning regression models” by Harish Kumar, Yogesh, Gad, et al.’s (2020). It evaluates the performance of models such as gradient boosting, SVR, random forest, and linear regression by comparing them and using measures like RMSE and MAE. Lela, D. B., Jogu, V., & Balu, H. (2022). Analysing Air Pollution Through Maps & Tools and Predicting AQI Using Python. Int. Res. J. Mod. Eng. This paper predicts the atmospheric concentration of SO2 using machine learning techniques. Sulfur dioxide causes irritation to the skin and ocular mucous membranes.

METHODOLOGY

A. Deepak, Dr. Amrapali S. Chavan , Bodhankar. Aniruddha,
Dr. L. Sherly Pushpa Annabel, Nimmala. harathi,
A. Vanathi
They discussed: Advancing Air Quality Prediction in Specific Cities Using Machine Learning

Architecture or Data flow diagram:

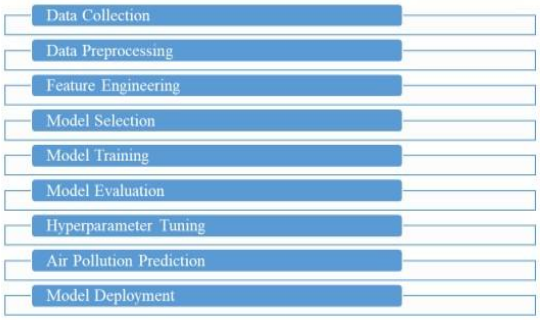


Fig 1. Proposed Methodology

Algorithms: Support Vector Machine (SVM), Random Forest, and XG Boost.

B. Aravind, S. Arumugam.
They Discussed: Air Quality Index Prediction Using Machine Learning and Deep Learning
Architecture or Data flow diagram

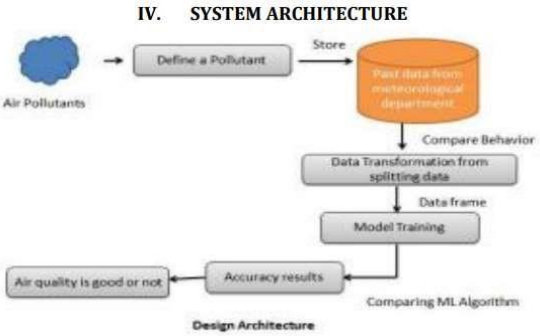
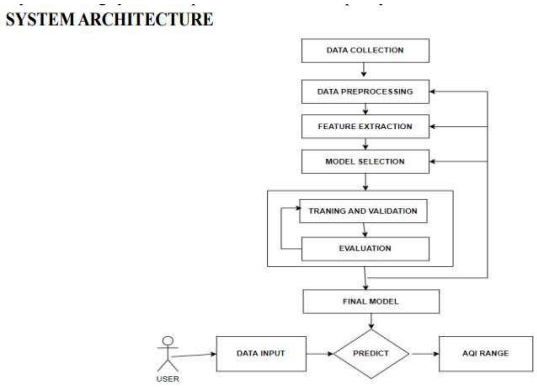


Fig 1: System architecture

Algorithms: Support Vector Machine (SVM), Random Forest, and confusion matrix.
B. Raviteja, P. Tejaswini,U. Swetha Reddy. Discussed: Air Quality Prediction Using Machine Learning

Architecture or Data flow diagram



Algorithms: Support Vector Machine (SVM), Random Forest, and neural networks.

Sheela S Maharajpet, Likhitha. S, Kiran .T

They Discussed: Air Quality Prediction Using Machine Learning

Architecture or Data flow diagram

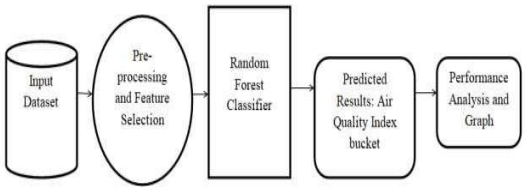


Figure 1. System Architecture

Algorithms: Support Vector Machine (SVM), Random Forest, and Decision tree.

Md. Mahbub Rahman, Md. Emran Hussain Nayeem, Md. Shorup Ahmed, Khadiza Akhtar Tanha, Md. Shahriar Alam Shakib, Khandaker Mohammad Mohi Uddin, Hafz Md. Hasan Babu.

They Discussed: AirNet: predictive machine learning model for air quality forecasting using web interface.

Architecture or Data flow diagram

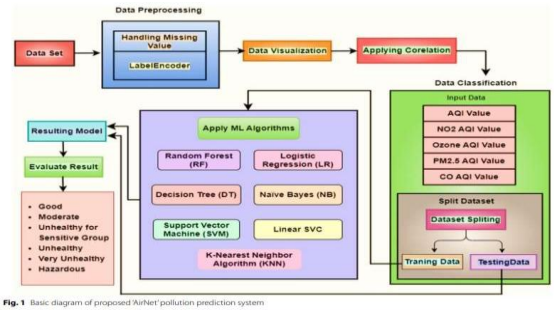


Fig. 1 Basic diagram of proposed 'AirNet' pollution prediction system

Algorithms: Support Vector Machine (SVM), Random Forest, confusion matrix and Decision tree.

Safaa Berkani, Ihsane Gryech, Mounir Ghogho (Fellow, IEEE),Bassma Guermah1, Abdellatif Kobbane3, (Senior Member, IEEE)

They Discussed: Data Driven Forecasting Models for Urban Air Pollution: More Air Case Study

Architecture or Data flow diagram

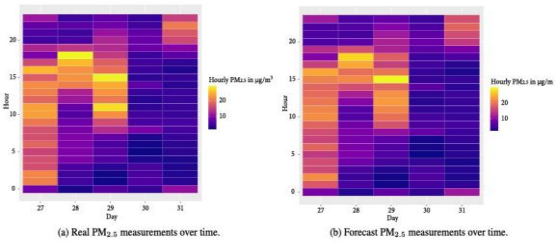


FIGURE 5. Comparison between the forecast and real values for $PM_{2.5}$.

Algorithms: XGBoost, Support Vector Machine (SVM), LSTM, and Decision tree

RESULTS & DECISIONS

It investigated the utilization of machine learning techniques to advance air quality prediction in specific cities.

It includes. The machine learning methods Support Vector Machine (SVM), Random Forest, Decision Tree, confusion matrix, linear regression and XG Boost were evaluated in order to determine which one would be the most accurate at forecasting air quality. These models use a range of input variables, such as meteorological data, traffic data, and emission data, to predict pollutant concentrations in the atmosphere. Studies have shown that machine learning models can accurately predict air quality, with some models achieving up to 90% accuracy. By combining machine learning with traditional air quality monitoring techniques, it may be possible to better understand the sources and impacts of air pollution and develop effective strategies for reducing its effects on human health and the environment

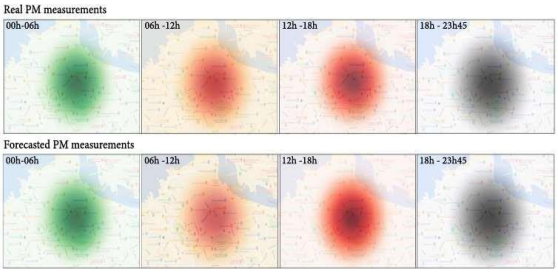


FIGURE 6. Geographical heatmap.

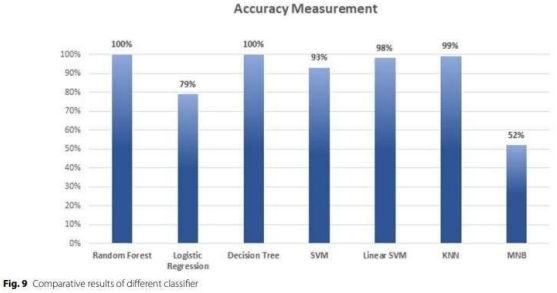


Fig. 9 Comparative results of different classifier

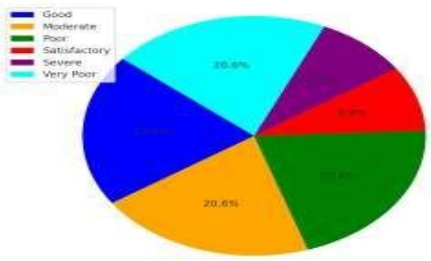


Figure 3. Chart

CONCLUSION

The application of machine learning techniques to improve air quality prediction in specific cities. It compiled an extensive dataset incorporating data from air quality monitoring stations, weather stations, traffic patterns, and industrial activities in the targeted city. It is a suitable option because of its ability to manage a large number of input variables and reduce overfitting. In this work, we focus on the early detection of air quality, crucial for safeguarding public health and the environment, with the potential to save millions of lives globally. Our research introduces an advanced air quality prediction system that integrates various machine learning techniques, including K-Nearest Neighbor, Random Forest, Support Vector Machine, Logistic Regression, Decision Tree, and Linear SVC, achieving a remarkable 100% classification accuracy with both Random Forest and Decision Tree models. It is work contributes uniquely by combining conventional and advanced methods and offering a user-friendly web interface for real time monitoring and alerts.

REFERENCES

[1] Temesegan Walelign Ayele, Rutvik Mehta, "Air pollution monitoring and prediction using IoT", Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018.

[2] Saba Ameer, Munam Ali Shah, Abid Khan, Houbing Song, Carsten Maple, Saif Ul Islam, Muhammad Nabeel Asghar, "Comparative Analysis of Machine Learning Techniques for Predicting Air Quality in Smart Cities", IEEE Access (Volume: 7), 2019.

[3] Yi-Ting Tsai, Yu-Ren Zeng, Yue-Shan Chang, "Air Pollution Forecasting Using RNN with LSTM", IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress, 2018.

[4] Cifuentes, J., Marulanda, G., Bello, A., & Reneses, J. (2020). Air temperature forecasting using machine learning techniques: A review. Energies, 13(6). https://doi.org/10.3390/en13164215 .

[5] Fu, L., Li, J., & Chen, Y. (2023). An innovative decision making method for air quality monitoring based on big data assisted artificial intelligence technique. Journal of Innovation and Knowledge,8(2)https://doi.org/10.1016/j.jik.2022.100294.

[6] Gokhale, S., & Raokhande, N. (2008). Performance evaluation of air quality models for predicting PM10 and PM2.5 concentrations at urban traffic intersection during winter period. Science of the Total Environment, 394(1), 9–24. https://doi.org/10.1016/j.scitotenv.2008.01.020.

[7] Abirami S, Chitra P (2023) Probabilistic air quality forecasting using deep learning spatial–temporal neural network. Geoinformatics 27(2):199–235

[8] Akhtar J (2020) Non-small cell lung cancer classification from histopathological images using feature fusion and deep CNN. Int J Eng Adv Technol. 9:2249–8958

[9] Alahmad B, Khraishah H, Althalji K, Borchert W, Al-Mulla F, Koutrakis P (2023) Connections between air pollution, climate change, and cardiovascular health. Canad J Cardiol. 39:1182–1190.

[10] Mangayarkarasi, R., Vanmathi, C., Khan, M. Z., Noorwali, A., Jain, R., & Agarwal, P. (2021). COVID19: Forecasting Air Quality Index and Particulate Matter (PM2. 5).

[11] Hossain, E., Shariff, M. A. U., Hossain, M. S., & Andersson, K. (2020, December). A novel deep learning approach to predict air quality index. In Proceedings of International Conference on Trends in Computational and Cognitive Engineering: Proceedings of TCCE 2020 (pp. 367-381). Singapore: Springer Singapore.

[12] Nahar, K. M., Ottom, M. A., Alshibli, F., & Shquier, M. M. A. (2020). Air quality index using machine learning–a Jordan case study. Compusoft, 9(9), 3831-3840.

[13] Madan T, Sagar S, Virmani D (2020) Air quality prediction using machine learning algorithms–a review. In: 2nd international conference on advances in computing, communication control and networking (ICACCCN) pp 140–145.

[14] Anurag Shrivastava, S. J. Suji Prasad, Ajay Reddy Yeruva, P. Mani, Pooja Nagpal & Abhay Chaturvedi (2023): IoT Based RFID Attendance Monitoring System of Students using Arduino ESP8266 & Adafruit.io on Defined Area, Cybernetics and Systems, DOI: 10.1080/01969722.2023.2166243.

[15] P. William, A. Shrivastava, H. Chauhan, P. Nagpal, V. K. T. N and P. Singh, "Framework for Intelligent Smart City Deployment via Artificial Intelligence Software Networking," 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM), 2022, pp. 455-460, Doi: 10.1109/ICIEM54221.2022.9853119.

[16] World Health Organization. (Sep. 22, 2021). Air Pollution is One of the Biggest Environmental Threats to Human Health, Alongside Climate Change. [Online]. Available: https://www.who.int/news/item/22-09-2021 new-who-global-air-quality guidelines-aim-to-save-millions-of-lives from-air-pollution.

[17] State of Global Air 2020. Special Report, Health Effects Institute, Boston, MA, USA, 2020.

[18] R. Fuller et al., ‘‘Pollution and health: A progress update,’’ Lancet Planet. Health, vol. 6, no. 6, pp. e535e547, 2022