Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.13, May : 2024

SMART AIR POLLUTION HOTSPOT DETECTION AND MONITORING USING IOT AND ML

Dr. A.S. Chavan, Associate Professor, Department of Computer Engineering, AISSMS Institute of InformationTechnology, Pune
Anurag Dharmadhikari, Student, Department of Computer Engineering, AISSMS Institute of Information Technology, Pune.
Rohan Kumbharkar, Student, Department of Computer Engineering, AISSMS Institute of Information Technology, Pune.
Aniket Patil, Student, Department of Computer Engineering, AISSMS Institute of Information Technology, Pune.
Sejal Zende, Student, Department of Computer Engineering, AISSMS Institute of Information Technology, Pune.

ABSTRACT
The alarming rise in pollution levels in both developed and developing countries has made air pollution a major global concern. The deployment of web and affordable monitoring systems is necessary to address this problem. This essay suggests a thorough approach to tracking air pollution and educating the public. The system is installed in certain areas with elevated air pollution, and it is designed to periodically check the concentration of dangerous pollutants. It measures the amounts of pollutants detected using Air Quality Sensors and uses a specialized mechanism to communicate this information to the public. This method improves public awareness of local air quality conditions by providing both numerical and graphical displays of pollutant concentrations and data from air quality sensors. Furthermore, there are plans to incorporate user registration using a web application as part of the system's evolving architecture. Users of this program will effortlessly receive weekly or monthly notifications regarding the quality of the air. The suggested system uses machine learning methods for hotspot prediction in addition to an Arduino controller and an Air Quality Sensor for pollution detection. By enabling proactive steps to reduce air pollution, this integration gives communities the power to make educated decisions about how to enhance air quality

Keywords:
Air Pollution, Monitoring System, Web-Based Monitoring, Air Quality Sensor, Arduino Controller.

I. Introduction
As reported by the World Health Organization , air pollution is the term used to describe the variations in environmental features that occur when chemical and biological agents contaminate indoor and outdoor habitats. In addition to contributing to noise pollution, common causes of air pollution include combustion appliances in homes, automobile emissions, and forest fires. The main pollutants that pose a threat to human health are nitrogen dioxide(NO_2), sulphur dioxide(SO_2), ozone(O_3),carbon monoxide(CO) and particulate matter. Air pollution has serious adverse effects on human health, including respiratory and other illnesses that can be lethal.

The World Health Organization has carried out comprehensive evaluations of air quality in around 1500 cities across the globe, exposing concerning pollution levels. Interestingly, the capital of India has continuously been listed as one of the most heavily polluted cities globally. With the greatest amounts of particulate matter smaller than 2.5 micrometers, Pune in particular confronts major issues. However, major environmental and technological challenges face cities worldwide due to pervasive air pollution and inadequate infrastructure for monitoring air quality.

Pollution in all forms has increased dramatically as society has progressed. Air pollution levels have significantly grown as a result of population growth, more industrial activity, increased transportation, and greater use of fossil fuels. As per the 2016 WHO report, the International Energy

Agency and the World Health Organization estimate that in 2012, air pollution caused over 7 million deaths globally. India, the second-most populous country in the world and the largest democracy, suffers from extreme air pollution, which results in 1.2 million premature deaths annually.

In India, vehicle emissions account for 27% of air pollution, followed by industrial emissions (51%), agricultural and waste burning practices (22%), and other sources. Based on WHO assessments, a number of Indian cities are regularly ranked in the top 20 most polluted cities. To tackle this issue, dependable systems that can identify and track pollution hotspots must be developed. Vehicle emissions are a primary source of pollution in major cities such as Delhi, Mumbai, Chennai, and Kolkata. Effective pollution control therefore requires the deployment of devices that can locate and reduce pollution hotspots, especially in densely populated urban areas.

II. Literature Survey

Numerous studies have been conducted on the subject of air pollution detection. Using Decision Trees in conjunction with IoT and ML technologies is a critical step in the pursuit of more intelligent air pollution hotspot detection and monitoring. While Internet of Things devices provide real-time data collecting, decision trees provide interpretability and efficiency in the analysis of complicated datasets. Our goal is to create a precise and scalable system for locating pollution hotspots and tracking air quality by utilizing these technologies. This study supports the broader trend of addressing environmental problems with sophisticated computational techniques, opening the door to more proactive pollution mitigation initiatives and better informed decision-making. Sharma et al. [21] proposed that solar-powered Internet of Things sensor nodes should be used to monitor and run agricultural areas. In the field of agriculture, operations like as crop management, crop harvesting, control of water supply, control of animals, distribution of pesticide, and temperature monitoring technologies will also be monitored and managed.

K B Gurumoorthy et al. [1] forecasts the amount of precipitation and wind (velocity and direction) that will contribute to PM2.5 buildup. It incorporates an algorithm for machine learning (ML) that is based on conclusions from six years of earth science and pollution data.

Soumyadeep Sur et al. [2] In this study, pollution levels in a selected area of Delhi are estimated, and hotspots for air pollution are identified using various techniques and algorithms. Time series AQI data from CPCB sensors in Delhi are collected. SVM is employed for hotspot classification, while LSTM and PROPHET are used for time series analysis based on pollutants such as PM2.5, PM10, CO, and NO. These models are utilized to forecast pollution levels for the day ahead.

Rohit Adke et al.[3] They focused on creating and predicting air pollution using linear regression based prediction system.

Aditya C R et al.[4] In this paper, Logistic Regression is used to classify whether the data sample is polluted or not. Auto Regression is employed to forecast future values of PM2.5 based on previous PM2.5 readings.

X. Xi, et al. [5] This paper focuses on predicting AQI levels by using machine learning methods and incorporates data from the WRF-Chem model, which includes pollution, weather, and chemical component forecasts. The paper aims to improve AQI prediction by addressing the issue of outdated source lists.

Zhang, et al.[6] This paper says that Environmental problems, especially air pollution, have a profound impact on human health and ecosystem leading governments to make significant efforts in pollution control.

Yawen Zhang et al.[7] This study utilizes mobile sensing data, specifically air quality sensors installed on vehicles, to detect pollution hotspots. To address the challenge of uneven sampling, they propose a two-step approach that involves local spike detection and sample-weighted clustering.

Saide Pablo E., et al. [8] The study introduces a forecasting system for carbon monoxide (CO) using the WRF-Chem platform. The system aims to predict CO levels, which can then be used to estimate

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.13, May : 2024
PM10 and PM2.5 concentrations. The accuracy of the forecasts is evaluated and compared with official data.
Balachandran S, et al.[9] This research explores the effects of airborne particles on respiratory health, highlighting the significance of particle size and the presence of potentially toxic substances.

III. Proposed System

Our proposed system with the help of hardware integration which includes an Arduino board and a MQ-135 sensor our suggested solution can deliver real-time air quality data to customers via an intuitive web interface. This system's main goal is to provide users with an easy-to-use interface via which they can obtain and interpret the Air Quality Index (AQI) and parts per million (PPM) concentrations of gases that the MQ135 sensor has identified.

The system has a number of important features that are designed to improve user experience. First off, customers are guided effortlessly through the process of acquiring and comprehending air quality data via the website's extremely easy interface. This guarantees a seamless and effective user experience, making it easy for users to explore the website.

Second, the system precisely monitors PPM concentrations of gases including ammonia, benzene, and carbon dioxide in the surrounding environment by combining the MQ-135 sensor with Arduino and displaying the results on a website. Convenience and accessibility are guaranteed because users can simply access this data from any internet-connected device.

Thirdly, the system provides users with a thorough picture of overall air quality by calculating the AQI based on the PPM concentrations of the identified gases. For those who are worried about the effects of air pollution on their health, this information is essential because it empowers them to make decisions that will safeguard their health.

Additionally, the website presents the PPM concentrations of each gas along with the AQI in an understandable and straightforward manner. The data is efficiently presented using graphs, charts, and visualizations, enabling users to understand and make use of the air quality information supplied by the system for well-informed decision-making.

In summary, our system integrates hardware (an Arduino board and an MQ-135 sensor) to provide a user-friendly method for real-time air quality monitoring, displaying the data through a simple web interface. The technology gives users the ability to make educated decisions about their exposure to the environment by giving accurate AQI and PPM data in an easily readable manner.

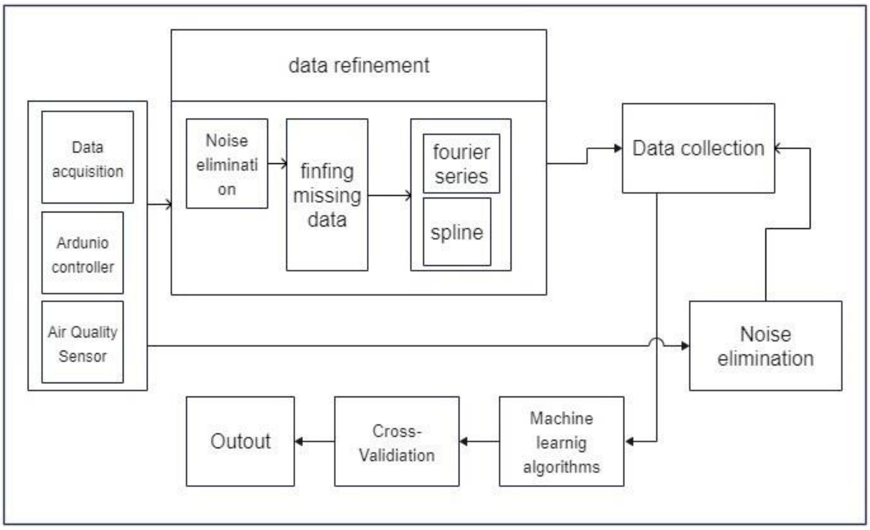

Fig. 1. System Architecture

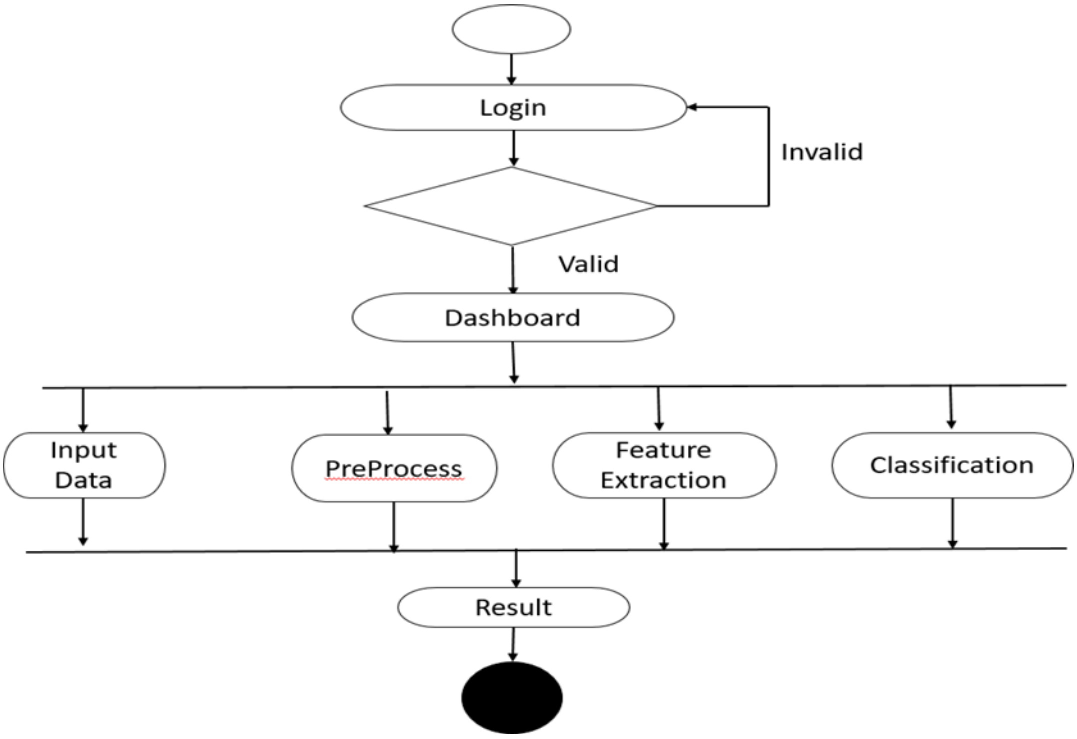

Fig. 2. Activity Diagram

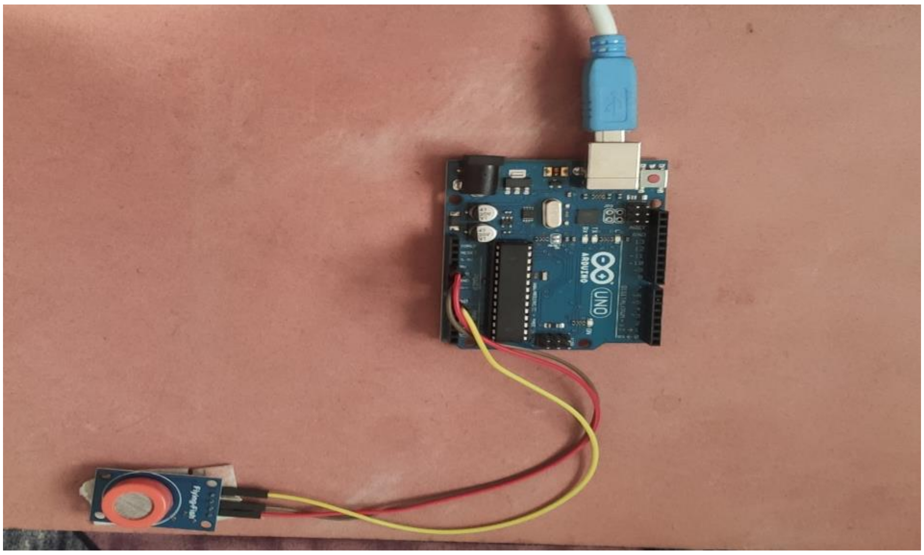

Fig. 3. Air Pollution Detection Unit

IV. Methodology

1. Data Collection

Data collection is a crucial and foundational stage in the training of any machine learning model. To ensure accurate model training, it is imperative to collect a substantial amount of high-quality data. This will help the model effectively learn the different features found in the dataset. The first step in the proposed system is gathering sensor data from several sources, such as satellite imaging, mobile sensors, and fixed monitoring stations. Our dataset consists of two features sensor value and air Quality

2. Data Preprocessing

An essential component of our suggested solution for monitoring air quality with an Arduino and MQ-135 sensor is data preprocessing. Preparing the unprocessed sensor data for analysis and model training entails a number of procedures.

Importing the required Python packages—such as numpy, pandas, and matplotlib—to enable data manipulation and visualization is the first stage in the preprocessing process. To ensure precise tracking of air quality over time, the datetime module is also used to maintain timestamps for real-time data collecting.

Importing the required Python packages—such as numpy, pandas, and matplotlib—to enable data manipulation and visualization is the first stage in the preprocessing process. To ensure precise tracking of air quality over time, the datetime module is also used to maintain timestamps for real-time data collecting.

The raw sensor data is loaded into a pandas Data Frame for additional processing when the packages are imported. It is possible to manipulate and analyze the data effectively thanks to its structured representation. The predictor and target variables must be extracted after the data has been loaded. This instance uses the MQ-135 sensor readings in the 'sensor Value' column as the predictor variable and the matching air quality classification in the 'air Quality' column as the target variable.

The train-test split function in the sklearn model selection module divides the dataset into training and testing sets after the predictor and target variables have been extracted. In order to evaluate the model's generalization ability, this stage makes sure that it is trained on a subset of the data and evaluated on unseen data.

Predictions are made using the training data after the model has been trained to assess its accuracy. Metrics are used to compute the accuracy score to measure how well the model performs in identifying the quality of the air based on sensor measurements.

A develop a predict Quality function that loads the learned DTC model from disk and predicts the air quality classification for fresh sensor data in real-time for air quality prediction. The function returns a qualitative assessment of the air quality, classifying it as Good Quality Air, Moderate Air Pollution, or Severe Air Pollution based on the projected class label.

3. Model Selection and Model Training

Given the nature of the data and the problem statement that is currently being addressed, the appropriate model is carefully selected based on a thorough evaluation of various factors. We chose to use the Decision Tree technique in our research because it is easy to understand, suitable for our dataset, and allows us to select and train models. Decision trees provide obvious insights into the decision-making process and are excellent at managing both numerical and categorical data. Our goal is to efficiently divide the feature space in order to identify air quality levels according to sensor readings from the MQ-135 sensor using the Decision Tree Classifier. The algorithm is perfect for our real-time air quality monitoring system because of its transparency and resilience to noisy data. Our preprocessed information is sent into the Decision Tree Classifier during model training, enabling it to identify underlying trends and adjust its parameters for precise air quality classification. Our goal is to create a dependable and comprehensible model that may support efficient environmental management and public health initiatives by utilizing the Decision Tree algorithm.

V. Results

We conducted a thorough data gathering campaign as part of our study project, utilizing real-time air pollution data from various geographical locations by integrating MQ135 sensors with Arduino microcontrollers. A rich dataset illustrating the various concentrations of air contaminants common in these places was produced as a result of this coordinated effort. By means of a thorough investigation, we were able to discover differences in air pollution levels between the investigated regions. Surprisingly, our research showed that about half of these areas had noticeably higher air pollution levels, indicat- ing that they could be hotspots for environmental concerns. The Air Quality Index (AQI) readings at these places were noticeably high, indicating a severe incident of air pollution. This

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 5, No.13, May : 2024

conclusion emphasizes how crucial our research is to identifying areas with elevated pollution levels and enabling focused interventions and mitigation plans. These kinds of revelations from our data gathering activities are a priceless addition to the continuing conversation about environmental surveillance and public health programs.

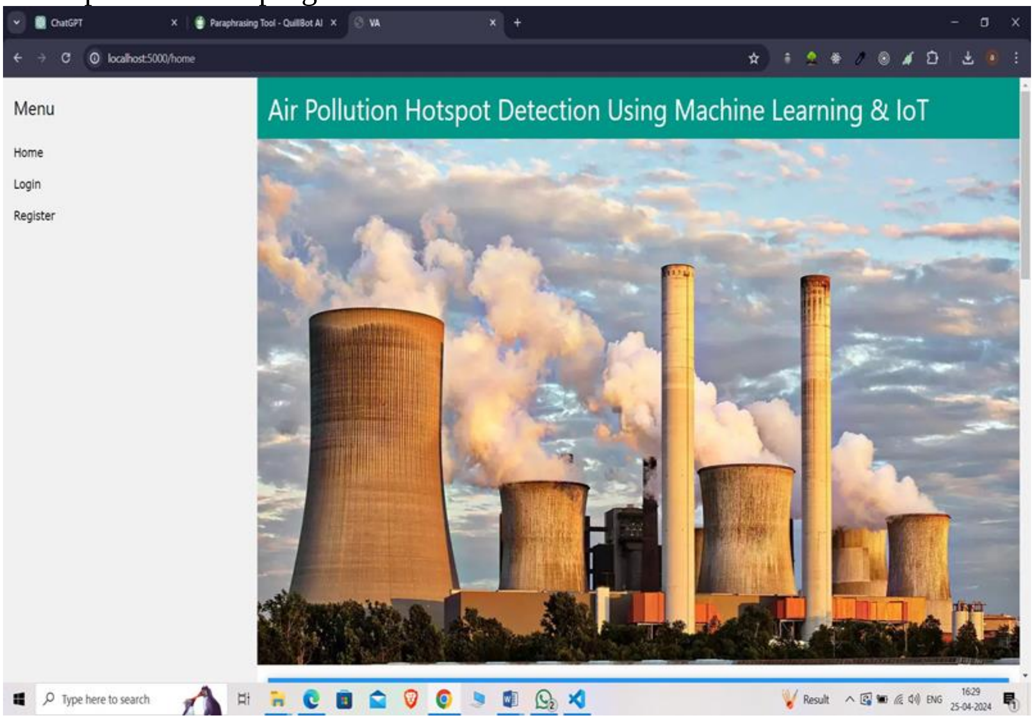

Fig. 4. Home Page

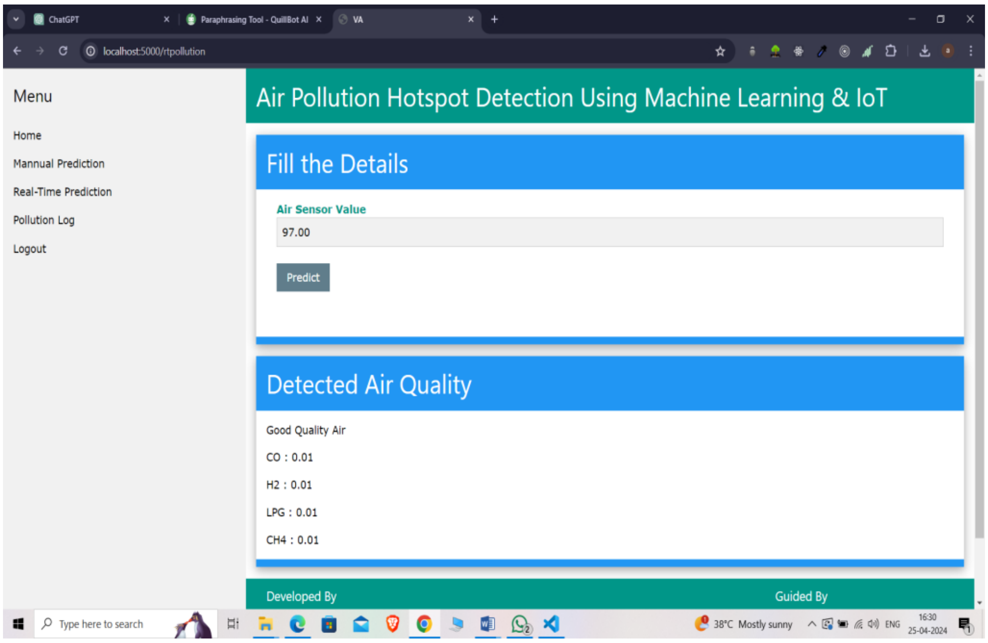

Fig. 5. Output/Result Page

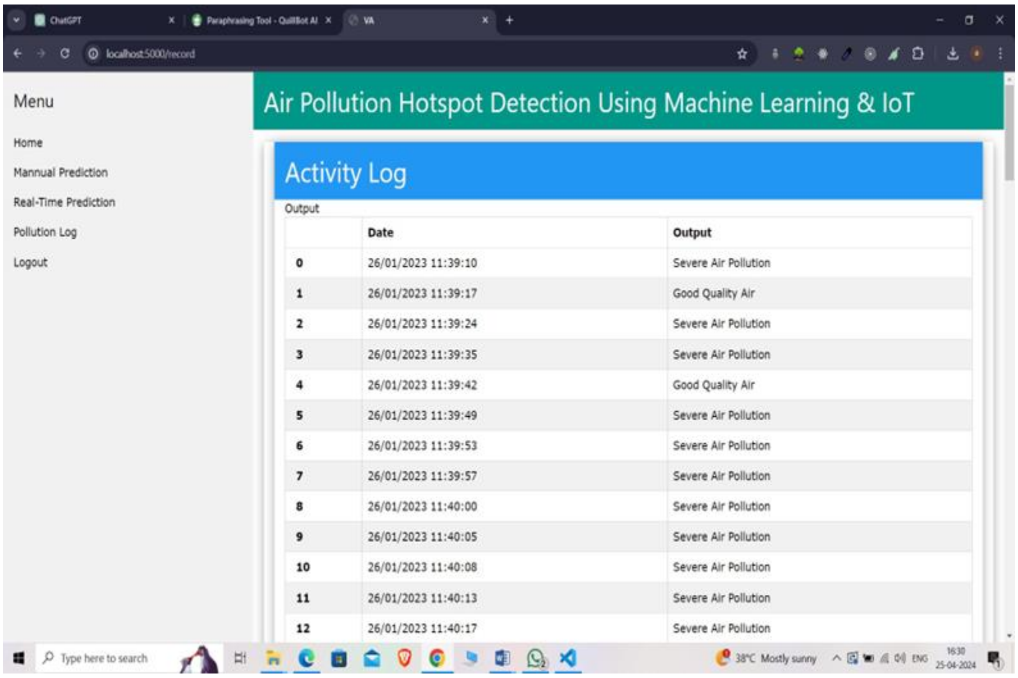

Fig. 6. Activity Log

VI. Applications

1) Urban Planning and Development

To help with decisions about urban planning, such as where to put roads, green spaces, and residential areas, identify locations with high pollution levels.

2) Public Health Management

When pollution events occur, monitor the state of the air in real time to safeguard public health and notify or advise susceptible groups.

3) Environmental Compliance

By locating pollution hotspots and sources, you can help regulatory bodies enforce environmental regulations.

4) Industrial Monitoring

To maintain compliance with environmental rules and limit the impact on nearby communities, keep an eye on emissions and air quality in the vicinity of industrial operations

5) Healthcare and Telemedicine

Utilizing data on air quality, telemedicine services that offer health recommendations based on regional pollution levels can be developed

VII. Conclusion

Regulating air pollution levels is increasingly becoming one of the most important responsibilities. People need to be aware of the extent of pollution in their environment and take action to reduce it. The results show how machine learning models—Decision Trees in particular—work well for determining air quality and pinpointing high-pollution locations. This study can be extended even further by using a simulation system capable of reproducing traffic and pollution conditions in a city. If multiple critical routes can be replicated, we can obtain a general understanding of real-world scenarios and infer the relationship between the city's traffic and pollution levels.

References

[1] Gurumoorthy, K. B., Vimal, S. P., Kumar, N. S., & Kasiselvanathan, M. (2021, June). Air pollution hotspot detection and identification of their source trajectory. In Journal of Physics: Conference Series (Vol. 1917, No. 1, p. 012029). IOP Publishing.Kamienski, C.; Soininen, J.-P.;

Taumberger, M.; Dantas, R.; Toscano, A.; Salmon Cinotti, T.; Filev Maia, R.; Torre Neto, A. Smart Water Management Platform: IoT-Based Precision Irrigation for Agriculture. Sensors 2019, 19, 276.
[2] Sur, S., Ghosal, R., & Mondal, R. (2020, September). Air Pollution Hotspot Identification and Pollution Level Prediction in the City of Delhi. In 2020 IEEE 1st International Conference for Convergence in Engineering (ICCE) (pp. 290-294). IEEE. Vijayan, T.; Sangeetha, M.; Kumaravel, A.; Karthik, B. Feature selection for Simple Color Histogram Filter based on Retinal Fundus Images for Diabetic Retinopathy recognition. IETE J. Res. 2020, 1–8.
[3] Kumar, V., & Ahmad, N. (2022, July). Deep Learning for Air Quality Prediction after COVID-19 Pandemic Based on Pollutant and Meteorological Data. In 4th International Conference on Communication & Information Processing (ICCIP).
[4] Aditya, C. R., Deshmukh, C. R., Nayana, D. K., & Vidyavastu, P. G. (2018). Detection and prediction of air pollution using machine learning models. International journal of engineering trends and technology (IJETT), 59(4), 204-207.Katarya, R.; Raturi, A.; Mehndiratta, A.; Thapper, A. Impact of Machine Learning Techniques in Precision Agriculture. In Proceedings of the 2020, 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things, ICETCE, Jaipur, India, 7–8 February 2020; pp. 1–6.
[5] Zhang, Y., Hannigan, M., & Lv, Q. (2021, November). Air Pollution Hotspot Detection and Source Feature Analysis using Cross-Domain Urban Data. In Proceedings of the 29th International Conference on Advances in Geographic Information Systems (pp. 592-595).Anand, R.; Karthiga, R.D.; Jeevitha, T.; Mithra, J.L.; Yuvaraj, S. Blockchain-Based Agriculture Assistance. Lect. Notes Electr. Eng. 2021, 700, 477.
[6] Kleine Deters, J., Zalakeviciute, R., Gonzalez, M., & Rybarczyk, Y. (2017). Modeling PM 2.5 urban pollution using machine learning and selected meteorological parameters. Journal of Electrical and Computer Engineering, 2017. Srisruthi, S.; Swarna, N.; Ros, G.M.S.; Elizabeth, E. Sustainable agriculture using eco-friendly and energy efficient sensor technology. In Proceedings of the 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, India, 20–21 May 2016; IEEE: Bangalore, India, 2016; pp. 1442–1446.
[7] Xi, X., Wei, Z., Xiaoguang, R., Yijie, W., Xinxin, B., Wenjun, Y., & Jin, D. (2015, November). A comprehensive evaluation of air pollution prediction improvement by a machine learning method. In 2015 IEEE international conference on service operations and logistics, and informatics (SOLI) (pp. 176-181). IEEE.
[8] Li, Y., Chen, Q., Zhao, H., Wang, L., & Tao, R. (2015). Variations in PM10, PM2. 5 and PM1. 0 in an urban area of the Sichuan Basin and their relation to meteorological factors. Atmosphere, 6(1), 150-163.
[9] Mead, M. I., Popoola, O. A. M., Stewart, G. B., Landshoff, P., Calleja, M., Hayes, M., ... & Jones, R. L. (2013). The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. Atmospheric Environment, 70, 186-203.
[10] Djebbri, N., & Rouainia, M. (2017, August). Artificial neural networks based air pollution monitoring in industrial sites. In 2017 International Conference on Engineering and Technology (ICET) (pp. 1-5). IEEE.
[11] Singh, K. P., Gupta, S., & Rai, P. (2013). Identifying pollution sources and predicting urban air quality using ensemble learning methods. Atmospheric Environment, 80, 426-437.
[12] Zhang, J., & Ding, W. (2017). Prediction of air pollutants concentration based on an extreme learning machine: the case of Hong Kong. International journal of environmental research and public health, 14(2), 114.
[13] Sun, W., & Sun, J. (2017). Daily PM2. 5 concentration prediction based on principal component analysis and LSSVM optimized by cuckoo search algorithm. Journal of environmental management, 188, 144-152.
[14] Saide, P. E., Carmichael, G. R., Spak, S. N., Gallardo, L., Osses, A. E., Mena-Carrasco, M. A., & Pagowski, M. (2011). Forecasting urban PM10 and PM2. 5 pollution episodes in very stable

nocturnal conditions and complex terrain using WRF–Chem CO tracer model. Atmospheric Environment, 45(16), 2769-2780.

[15] Chattopadhyay, S., Gupta, S., & Saha, R. N. (2010). Spatial and temporal variation of urban air quality: a GIS approach. Journal of Environmental Protection, 1(03), 264.

[16] Bhaskar, B. V., Rajasekhar, R. J., Muthusubramanian, P., & Kesarkar, A. P. (2008). Measurement and modeling of respirable particulate (PM10) and lead pollution over Madurai, India. Air Qual. Atmos. Health, 1(1), 45-55.

[17] Balachandran, S., Meena, B. R., & Khillare, P. S. (2000). Particle size distribution and its elemental composition in the ambient air of Delhi. Environment international, 26(1-2), 49-54.

[18] Gamble, J. F., & Lewis, R. J. (1996). Health and respirable particulate (PM10) air pollution: a causal or statistical association?. Environmental health perspectives, 104(8), 838-850.

[19] Finzi, G., & Tebaldi, G. (1982). A mathematical model for air pollution forecast and alarm in an urban area. Atmospheric Environment (1967), 16(9), 2055-2059.

[20] Sousa, S. I. V., Martins, F. G., Alvim-Ferraz, M. C., & Pereira, M. C. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. Environmental Modelling & Software, 22(1), 97-103.