Air Pollution Hotspot Detection using Machine Learning and IoT

Himanshu Singh #1, Shahid Tamboli*2, Aishwarya Desai#3, Mandar Ladkat*4

[1,2,3,4] Students, Department of Computer Engineering,
JSPM's Jayawantrao Sawant College of Engineering
Hadapsar, Pune-28
Affiliated to Savitribai Phule Pune University, Pune

***Abstract*— Air pollution rates nowadays are drastically increasing in all the developed and developing countries which require a more portable and cost effective solution. The proposed system includes the design for monitoring air pollution and creating awareness among the public .The proposed system is installed in a particular locality where there is acute air pollution. The level of each hazardous pollutant is monitored at periodic intervals. The Air Quality Sensor for the observed pollutants is determined and awareness is created among the public through a proposed system that displays the level of each observed pollutant and also the air quality sensor in that particular location. Thus, the quality of air in that area can be understood numerical and graphical formats. Further, this system is to be themselves in an app that pushers weekly or monthly air quality notification that is more comfortable to access. In this proposed Quality Sensor and Arduino controller to detect air pollution and use machine learning algorithm to predict whether it is an Air Pollution Hotspot or not.**

Keywords*—Air Pollution; Air Quality Sensor; Air Pollution Hotspot Prediction; Machine Learning & Classification; IoT; Decision Tree Algorithm; Arduino Controller; and Dataset; etc.

I. Introduction

As per World Health Organization [WHO], air pollution is the infectivity of the indoor or outdoor environment by any chemical and biological agent which changes the characteristics of the environment. Household combustion devices, vehicles, and forest fires are the common origins of air pollution and noise pollution. Pollutants that are responsible for health concerns include particulate matter, carbon monoxide, ozone, nitrogen dioxide, and sulfur dioxide. Air pollution causes respiratory and other diseases, which can be deadly. WHO has measured the quality of air in approximately 1500 cities and the Indian capital city was one of the most polluted cities in the world. Pune is having the highest concentration of particulate matter which is smaller than 2.5 micrometers[1].

Air pollution and lack of air quality monitoring points represent environmental and technological challenges for cities and environments around the world. To face this issue, the industry has focused its efforts on finding a versatile technological alternative that allows the improvement of the air quality measuring process and provides reference values in network sites where conventional monitoring fails to cover appropriately. Unfortunately, existing products and the generated results do not represent low-cost solutions.

As civilization proceeds, the amount of pollution in all mediums also increases exponentially. Whilst the population increases, there are more factories, transport, and fossil fuel consumptions, and the amount of air pollution is at large. According to the 2016 World Health Organization report [1], air pollution in 2012 caused the deaths of around 7 million people worldwide, an estimate roughly echoed by the International Energy Agency. India, which has the largest democracy and the second largest population in the whole world, is suffering very much from contamination in the air. Air pollution is the cause of 1.2 million premature deaths in India [6].

The main agents of chaos in air pollution in India are industrial pollution (51%), vehicular pollution (27%), and the rest caused by crops and waste burning, fireworks, etc. (22%). Already 13 cities in India place in the top 20s in WHO's list of most polluted cities. But to reduce pollution, there must be a reliable system that can detect pollution. In the case of metro cities like Delhi, Mumbai, Chennai, and Kolkata, the main source of pollution is vehicular pollution [4]. So, to reduce vehicular pollution, a system has to be made that can detect the main polluted areas or routes in a city[7].

II. Literature Survey

K B Gurumoorthy, S P Vimal, N Sathish Kumar and M Kasiselvanathan, "Air Pollution Hotspot Detection and Identification of Their Source Trajectory."[1] This paper predicts the accumulation of PM2.5 from wind (velocity and direction) and precipitation levels. It imbibes a machine learning (ML) algorithm supported by six years of earth science and pollution information inferences. At present, pollution may be a world downside. The Republic of India is additionally an enormous sufferer of this downside. Thus, it's necessary to spot the recent spots of pollutants and their transport specifically carbon monoxide gas (CO), sulfur dioxide (SO2), and oxides of element (NO+NO2) victimization advanced information analysis techniques. Challenges concerned during this current statement are mining the datasets from completely different parameters and providing the ultimate output with moderate abstraction resolution on pollution info. Therefore, the study illustrates that the employment of applied mathematics models supported by the ML algorithm is most relevant to predict PM2.5 accumulation from earth science information.

Soumyadeep Sur, Rohit Ghosal, Rittik Mondal, "Air Pollution Hotspot Identification and Pollution Level Prediction in the City of Delhi."[2] In this paper, we use various methods and algorithms to detect air pollution hotspots and predict pollution levels in a selected area in the city of Delhi. Time series AQI data is collected through the CPCB sensors in Delhi. Classification of hotspots is done using SVM, and the time series analysis based on pollutants like PM2.5, PM10, CO, NO data samples is done using LSTM and PROPHET. Pollution levels of a day in the future are predicted using the said models.

Rohit Adke, Suyog Bachhav, Akash Bambale, Bhushan Wawre, "Air Pollution Prediction using Machine Learning."[3] The urban air pollution rate has grown at an alarming state across India. Most of the cities are facing the issue of poor air quality which fails to meet standards of air for good health. It is indeed necessary to develop an air pollution measurement and prediction system for the smart city. Nowadays almost every city has its prediction system which uses linear regression. Exploitation browser to the visualized accurate result of prediction. For improving the linear regression algorithm with maximum accuracy, we are using a neural network. Which uses the Multilayer Perceptron algorithm.

Aditya C R, Chandana R Deshmukh, Nayana D K, Praveen Gandhi Vidyavastu, “Detection and Prediction of Air Pollution using Machine Learning Models.”[4] In populated and developing countries, governments consider the regulation of air as a major task. The meteorological and traffic factors, burning of fossil fuels, and industrial parameters such as power plant emissions play significant roles in air pollution. Among all the particulate matter that determines the quality of the air, Particulate matter (PM 2.5) needs more attention. When its level is high in the air, it causes serious issues to people’s health. Hence, controlling it by constantly keeping a check on its level in the air is important. In this paper, Logistic regression is employed to detect whether a data sample is either polluted or not polluted. Autoregression is employed to predict future values of PM2.5 based on the previous PM2.5 readings. Knowledge of the level of PM2.5 in nearing years, month or week, enables us to reduce its level to a lesser than the harmful range. This system attempts to predict PM2.5 levels and detect air quality based on a data set consisting of daily atmospheric conditions in a specific city.

Yawen Zhang, Michael Hannigan, Qin Lv, “Air Pollution Hotspot Detection and Source Feature Analysis using Cross-domain Urban Data.”[5] Air pollution is a major global environmental health threat, in particular for people who live or work near pollution sources. Areas adjacent to pollution sources often have high ambient pollution concentrations, and those areas are commonly referred to as air pollution hotspots. Detecting and characterizing pollution hotspots are of great importance for air quality management, but are challenging due to the high spatial and temporal variability of air pollutants. In this work, we explore the use of mobile sensing data (i.e., air quality sensors installed on vehicles) to detect pollution hotspots. One major challenge with mobile sensing data is uneven sampling, i.e., data collection can vary by both space and time. To address this challenge, we propose a two-step approach to detect hotspots from mobile sensing data, which includes local spike detection and sample-weighted clustering. Essentially, this approach tackles the uneven sampling issue by weighting samples based on their spatial frequency and temporal hit rate, to identify robust and persistent hotspots. To contextualize the hotspots and discover potential pollution source characteristics, we explore a variety of cross-domain urban data and extract features from them. As a soft validation of the extracted features, we build hotspot inference models for cities with and without mobile sensing data. Evaluation results using real-world mobile sensing air quality data as well as cross-domain urban data demonstrate the electiveness of our approach in detecting and inferring pollution hotspots. Furthermore, the empirical analysis of hotspots and source features yields useful insights regarding neighborhood pollution sources.

III. PROPOSED SYSTEM

A. *Problem Statement:*

Air pollution rates nowadays are drastically increasing in all the developed and developing countries which require a more portable and cost-effective solution. In this proposed system we design and develop a system with the help of an Air Quality Sensor and Arduino controller to detect air pollution and use a Machine Learning algorithm to predict whether it is an Air Pollution Hotspot or not.

B. *Architecture Design:*

We propose a two-step approach to automatically detect air pollution hotspots from mobile sensing data, which addresses the uneven sampling issue with mobile sensing. The discovery of hotspots can help in narrowing down the regions of interest, which are further investigated by domain experts and regulatory agencies.

Our analysis of hotspots and source features yields useful insights regarding neighborhood pollution sources. These insights may be directly used to complement conventional pollution inventories. As a supplement to empirical analysis, we build hotspot inference models for multiple cities to validate the effectiveness and generalizability of the extracted features in representing local pollution sources.

We evaluate our approaches using real-world mobile sensing air quality data collected by GSV vehicles in California and cross-domain urban data from various platforms[8].

Our methodology consists of three key components: hotspot detection, source-related feature extraction, and hotspot inference. Firstly, with mobile sensing air quality data, we detect pollution hotspots with a two-step approach[9].

Step 1 involves detecting local spikes from raw observations to pick out locations with elevated pollution concentrations. Step 2 aggregates the results of Step 1 by applying sample-weighted clustering to identify robust and persistent hotspots.

The sample-weighted clustering technique takes into account the spatial and temporal coverage of each sample. Secondly, with cross-domain urban data, such as POIs, trac, land use, and elevation, we extract source-related features and create a feature vector representation for each hotspot. Finally, with feature vector representation, we build machine learning (ML) models to infer hotspots in cities with and without mobile sensing data. For cities with mobile sensing data, the hotspot detection results are used as labels for model training. For cities without mobile sensing data, we perform domain adaption between cities for hotspot inference[10].

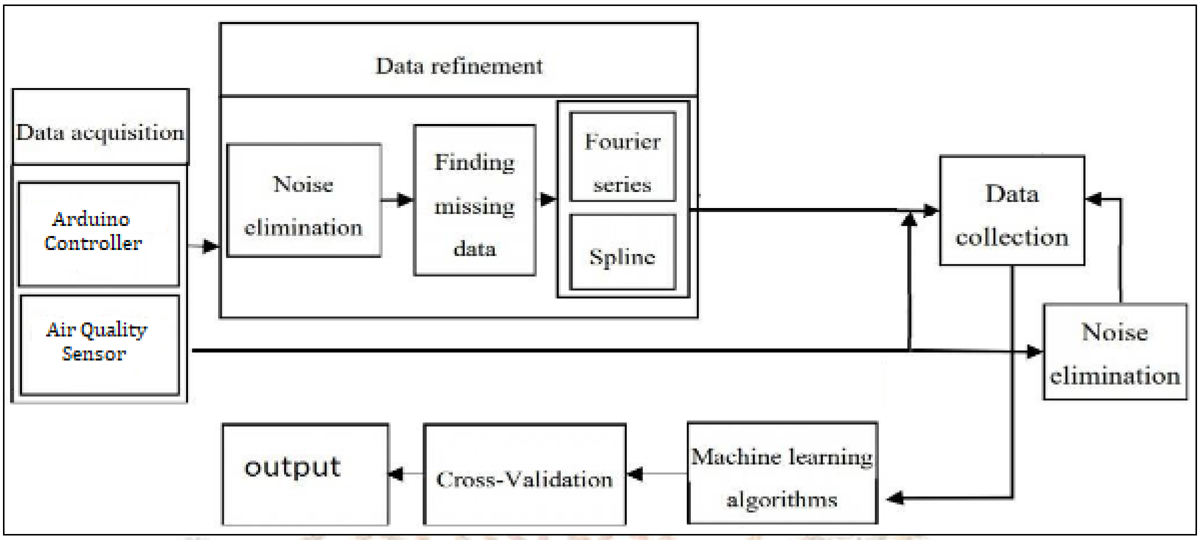

Fig.1: System Architecture

IV. RESULTS AND DISCUSSIONS

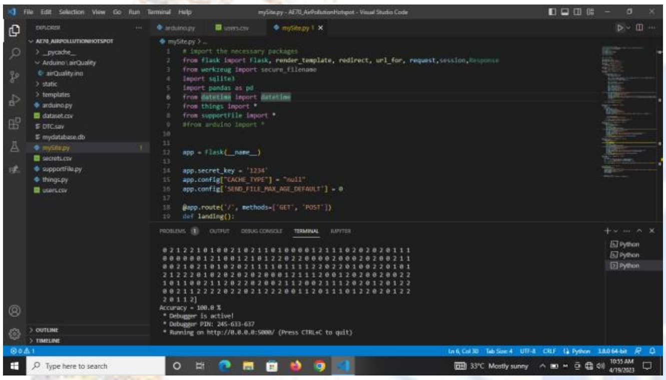

Fig-2 Python code

This is our source code for this project in Python. We have used several libraries such as aurdino, metplotlib, etc.As shown in the screenshots

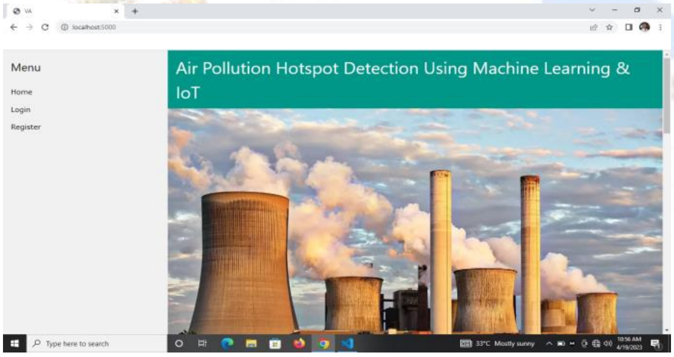

Fig-3 Pollution from Industries

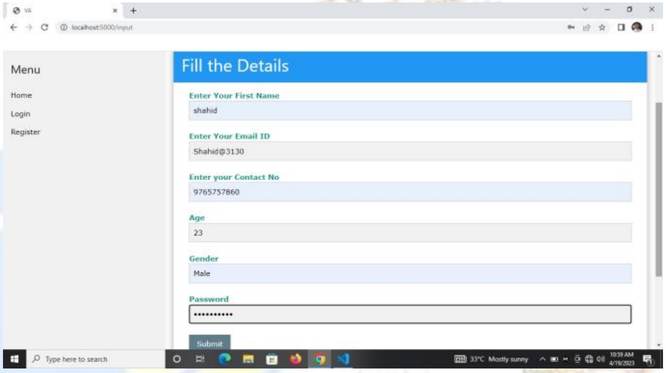

Fig-4 Menu Details

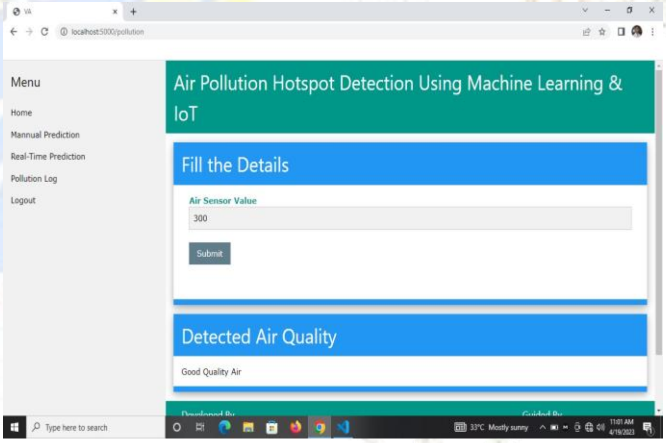

Fig-4 Menu Details with detected Air Quality

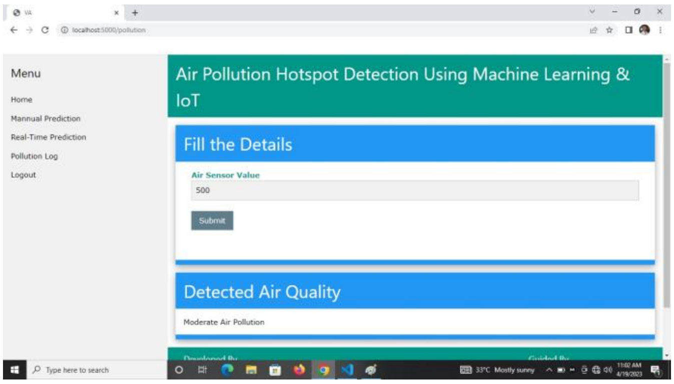

Fig-5 Menu Details with Detected Air Quality

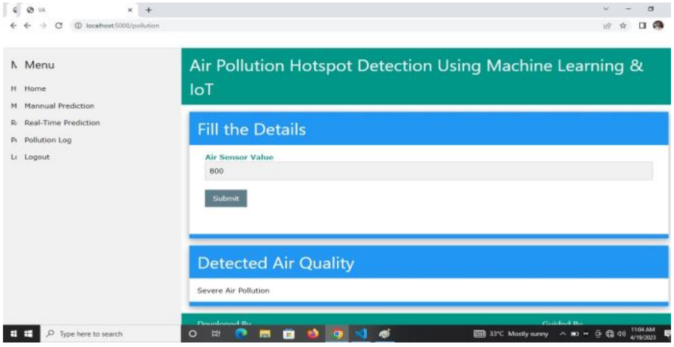

Fig-6 Menu Details with detected Air Quality

This figure shows the output for the input given manually to the system that will classify the quality of Air in Good, Moderate and Severe Quality manually. This will help us to get the Air quality of any location just by putting their PM value in it.

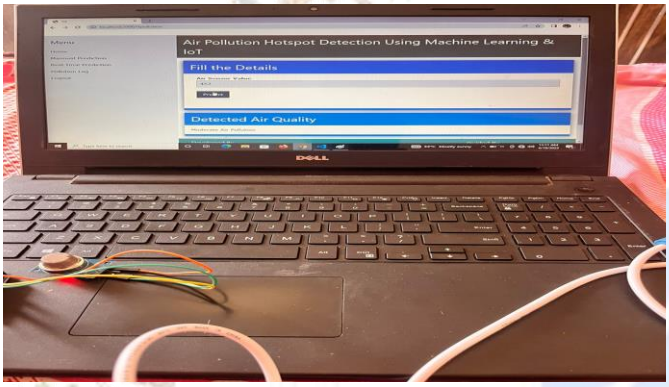

Fig-4 Hardware used in Laptop for detection of Air Quality

V. Conclusion

The regulation of air pollutant levels is rapidly becoming one of the most important tasks. It is important that people know what the level of pollution in their surroundings is and takes a step towards fighting against it. The results show that machine learning models (Decision Trees) can be efficiently used to detect the quality of air and predict Air Pollution hotspots. This work can be further extended by adding a simulation mechanism that can simulate congestion and pollution levels in a city. If various important routes can be simulated, we will get an approximate idea about a real-life scenario & can infer a relationship between the traffic pollution levels in the city. The project on Air Pollution Hotspot Detection using Machine Learning has successfully demonstrated the potential of machine learning techniques to identify and predict the quality of air in specified area. By analyzing data from various sources, including air quality sensors and satellite imagery, the project has developed a model that can accurately detect and classify air pollution hotspots in real-time. The result of this study can be used inform policy decisions aimed at reducing air pollution and improving public health.

Furthermore, there are several avenues for future research in the area of air pollution hotspot detection using machine learning. One possible direction is to explore the use of more advanced machine learning algorithms, such as deep learning,, to improve the accuracy and robustness of the models. Another possible direction is to incorporate more diverse data sources, such as weather patterns and traffic data, to further refine the models. Additionally, future studies could focus on developing predictive models that can forecast air pollution hotspots, enabling proactive interventions and preventive measures. Overall, the potential applications of machine learning in air pollution hotspot detection are vast and offer more promising opportunities for improving public health and environmental sustainability.

Acknowledgments

We would prefer to give thanks to the researchers likewise publishers for creating their resources available. We are conjointly grateful to the guide, and reviewer for their valuable suggestions and also thank the college authorities for providing the required infrastructure and support.

References

[1] K B Gurumoorthy, S P Vimal, N Sathish Kumar and M Kasiselvanathan, "Air Pollution Hotspot Detection and Identification of Their Source Trajectory". International Journal of Advanced Research in Science, Communication and Technology (IJARSCT) Volume 2, Issue 6, May 2022

[2] Soumyadeep Sur, Rohit Ghosal, Rittik Mondal, "Air Pollution Hotspot Identification and Pollution Level Prediction in the City of Delhi". Journal of Physics: Conference Series, 2019.

[3] Rohit Adke, Suyog Bachhav, Akash Bambale, Bhushan Wawre, "Air Pollution Prediction using Machine Learning". 2020 IEEE International Conference for Convergence in Engineering.

[4] Aditya C R, Chandana R Deshmukh, Nayana D K, Praveen Gandhi Vidyavastu, "Detection and Prediction of Air Pollution using Machine Learning Models". International Journal of Engineering Trends and Technology (IJETT) – volume 59 Issue 4 – May 2018.

[5] Yawen Zhang, Michael Hannigan, Qin Lv, "Air Pollution Hotspot Detection and Source Feature Analysis using Cross-domain Urban Data". IEEE, 2022

[6] Jan Kleine Deters, Rasa Zalakeviciute, Mario Gonzalez and Yves Rybarczyk 2017 Modeling PM2.5 Urban Pollution Using Machine Learning and Selected Meteorological Parameters Journal of Electrical and Computer Engineering, 5, pp.1-14.

[7] X. Xi, Z. Wei, R. Xiaoguang et al 2015 A comprehensive evaluation of air pollution prediction improvement by a machine learning method in the 10th IEEE International Conference on Service Operations and Logistics, and Informatics, pp. 176–181,Hammamet, Tunisia.

[8] Y. Li, Q. Chen, H. Zhao, L. Wang, and R. Tao 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, no. 1, pp. 150–163.

[9] M. Mead, O. Popoola, G. Stewart, P. Landshoff, M. Calleja, M. Hayesb, J. Baldovi, M. McLeod, T. Hodgson, J. Dicks, A. Lewis, J. Cohen, R. Baron, J. Saffell, and R. Jones 2013 The Use of Electrochemical Sensors for Monitoring Urban Air Quality in Low-Cost, High-Density Networks Atmospheric Environment, 70, pp. 186–203.

[10] Djebbri, N and Rouainia, M. 2017 Artificial neural networks based air pollution monitoring in industrial sites in the International Conference on Engineering and Technology (ICET), Antalya, Turkey, 21– 23.