Air Pollution Prediction Technique by Semi-supervisedLearning Approach using Machine Learning

Syeda Kaynath, Sthuthi Joanna, Thupakula Bhuvneshwari, Priyanka karki

Students of Acharya Institute of Technology Department of Information Science and Engineering,
Acharya Institute of Technology, Bangalore, Karnataka

Abstract - Air pollution detection using machine learning is a rapidly growing and developing field which involves the use of machine learning algorithms to detect and monitor the contamination of air pollution in various environments. The main goal of this approach is to create accurate, reliable, and automated systems that can continuously monitor and analyze the quality of the air we breathe.

There are several techniques that can be used for air pollution detection using machine learning. One of these popular techniques is to use a sensor data, which can be collected from a variety of sources, such as stationary air quality monitors or mobile devices like smartphones. This data can then be analyzed using machine learning algorithms to identify patterns and trends in the levels of pollutants in the air.

Another approach is to use satellite imagery to monitor air pollution levels. This technique involves using machine learning algorithms to analyze satellite data and identify areas with high levels of pollution. This can be useful for monitoring large areas and identifying pollution hotspots.

There are several benefits to using machine learning for air pollution detection. Firstly, machine learning algorithms can analyze vast amounts of data quickly and accurately, making it easier to detect changes in air quality over time. Additionally, these systems can be automated, meaning that they can continuously monitor air quality without the need for human intervention.

Overall, air pollution detection using machine learning has the potential to significantly improve our understanding of the impact of pollution on human health and the environment, and to help us take steps to reduce pollution levels and improve air quality.

Index Terms - Air pollution, Machine learning, Sensor data, Satellite imagery, Air quality monitoring, Pollution hotspots, automated systems, Human health, Environmental impact, Data analysis, Pattern recognition, Trend identification, Algorithm development, Pollution reduction.

I. INTRODUCTION (HEADING 1)

Pollution is calculated through particulate matter can be either human-made or naturally occur. Some examples include dust, ash and sea-spray. Particulate matter (including soot) is emitted during the combustion of solid and liquid fuels, such as for power generation, domestic heating and in vehicle engines. Particulate matter varies in size (i.e. the diameter or width of the particle). PM2.5 refers to the mass per cubic meter of air of particles with a size (diameter) generally less than 2.5 micrometres (μm) [13]. PM2.5 is also known as fine particulate matter (2.5 micrometres is one 400th of a millimetre). Fine particulate matter (PM2.5) is significant A m among the pollutant index because it is a big concern to people's health when its level in the air is relatively high [1]. PM2.5 refers to tiny particles in the air that reduce visibility and cause the air to appear hazy when levels are elevated. Different machine learning models have been applied to detect air pollution and predict PM2.5 levels based on a data set consisting of daily atmospheric conditions. Naive Bayes classification and support vector machine algorithms the air pollution detection using machine learning refers to the use of machine learning algorithms to detect and monitor the levels of air pollution in various environments. The goal of this approach is to Acharya Institute of Technology (AIT) May 2023 create accurate, reliable, and automated systems that can continuously monitor and analyze the quality of the air we breathe. Machine learning algorithms can be trained to analyze large amounts of data from various sources, such as sensor data, satellite imagery, and weather models. By analyzing this data, machine learning algorithms can identify patterns and trends in air pollution levels and predict future levels. This information can be used to develop targeted interventions to reduce pollution levels and protect public health. The use of machine learning for air pollution detection has the potential to significantly improve our

understanding of the impact of pollution on human health and the environment, and to help us take steps to reduce pollution levels and improve air quality.

II. LITERATURE SURVEY

A literature survey is a systematic review and analysis of published literature on a particular topic. It involves conducting a thorough search of databases, journals, books, and other sources to identify relevant articles, books, and other publications on a specific subject. The aim of a literature survey is to provide a comprehensive overview of the existing literature on a topic, identify gaps in knowledge, and suggest areas for further research.

SL.NO	Title of thePaper	Authors	Authors Approach / Method	Problem Addressed	Results
1	A novel methodfor Indoor Air Quality Control of Smart Homesusing a MachineLearning Model	Ali Majdia, Ali Jawad, AlrubaiebAlia HaiderAlWard yJamelBailic, HiteshPanchal d	Source control, treatment, andthe use of ventilation to reduce pollutants.	Indoor air quality check related to gasses any stressful mass or energy and can affect induction and health condition.	expected to predict air quality inside and around buildings and structures.
2	Air Cognizer: Predicting Air Quality with Tensor Flow Lite	Prerna Khanna, Tanmay Srivastava and Kanishk Jeet	Provides and collects images from the camera on the mobile phone, and processes them.	Predicting Air Quality with Tensor Flow Liteto improvethe operations.	Air quality is detected and AQI estimation modelon cloud.
3	Analysis of the Effectiveness ofAir Pollution Control PoliciesBased on Historical Evaluation and Deep Learning Forecast	Savitha Kamalapurkar	Effectiveness isin-depth and making scientific forecast of the air quality in one year.	Enrich existing literature on air pollution control by selecting the typical region	Provide a novel scientific tool for design and formulation of air pollution control policies
4	IoT based Air Pollution Monitoring System usingArduino	MuhammadAqib	trigger a alarm when the air quality goes down beyond acertain level	monitoring theAir Quality over a webserver using internet	It will show the air quality in PPM on the LCD and aswellas onwebpage.

SL.NO	Title of thePaper	Authors	Authors Approach / Method	Problem Addressed	Results
5	Real-time space occupancy sensing and human motion analysis using deep learning for indoor air quality control	Ivan MutisAbhijeet Ambekar, ViratJoshi	A foundationofa method to improve the current HVAC System with minimumcost	Sensing system to obviate energy wastageand enable costminimization	Builds new facility management techniques for effective management of existing building ventilation type.
6	AIR QUALITY INDEX USING MACHINE LEARNING – A JORDAN CASE STUDY	Khalid M.O.Nahar1 Mohammad Ashraf Ottom2 , Fayha Alshibli3 andMohammed M. Abu Shquier4	prioritizationair quality problems by establishing detection and monitoring stations in 12positions.	The model can predict themost pollutantfactors from real readings.	These estimateswill indicate the most influenced pollutants and theirbehavior inthe pollution.
7	Predict and Measure Air Quality Monitoring System Using Machine Learning	D.Saravanan	This allows us to choosea suitable model with a variety of regularizatio n strategies.	advanced analysis to simulate the hourly environment change	regularizations outperform open product relapse models and regularizations in terms of execution.
8	INDIAN AIR QUALITY PREDICTION AND ANALYSIS	Mrs. A. Gnana Soundari MTech, (PhD) Associate	Model wie capable fo successfully predicting th	Forecast theair quality ofIndia by usingML to predict the air	Better performance thanthe standard regression models and 96% on
	USING MACHINE LEARNIG	Professor Jeppiaar Engineering College	air qualit index of a tal county	qualityindex of a given area.	predicting the current air quality.
9	Understanding the true effects of the COVID-19on air pollution by means of Machine Learning	Mario Lovrića Kristina,Pavlovi Matej Vuković StuartK,Grange cdMichael Haberle RomanKerna	All pollutantswhich were analyzed decreased, with the exception of Ozone	Quantifying the effects was achieved with machine learning techniques.	The results have implications for air quality management.
10	Identification of high impact factors of air quality on a national scale using BD and ML	Jun Maab, Yuexiong, DingbJack, C.P.ChengaFeifeng, J.L Gana, Vincent	The second n objective is fact verification of proposed framework	Extremegradient Boosting Decesion tree (XGBoost) to model non linear relationships	Identify the most important factorson AQI on a national scale.

III. METHODOLOGY

Our goal is to predict the concentration of air pollutants of the next day on the basis of the historical meteorological and air pollutant data. In this work, we have focused on using the former day's data to predict the next day's hourly pollutants. The light model formulation had a clear advantage over the heavy model formulation and the baseline model formulation, which implied that controlling the number of parameters is important for improving generalization performance.

The regularizations, which verified that considering the similarities between models of consecutive hours is helpful. GPS location data to retrieve the assessment of the quality of the air from nearby air quality stations and they have applied dictionary learning and convolution neural network on the photos uploaded by the user to predict the air quality. The Air pollutants information is retrieved from the sensors which are processed in a unified schema and stored as a dataset. The disadvantage of this approach is they have not considered geographical and social environment characteristics, so the results may vary based on these factors. The network method consists of links and nodes which are used to calculate various analysis indicators for the network, such as centrality and the clustering coefficient. The station–station correlation coefficient was expressed as an inter-node link and calculated using the correlation coefficient. Centrality was used to analyze the network. Hyper parameters include the learning rate, hidden layer, hidden nodes, drop-out, and time period, increasing complexity so information about air pollutants is obtained from the sensors, analyzed, and then save it as a dataset.

This dataset has been pre-processed with a variety of features, which includes attribute selection and normalization. The prediction models built using machine learning have been shown to be more reliable and consistent. Data collecting is now simple and precise due to advanced technology and sensors. The required libraries are imported and then the dataset is read in the python code.

IV. DESIGN

Data collection: The first step is to collect data on air pollution levels using sensors or other monitoring devices. This data can include measurements of various pollutants such as particulate matter, ozone, and nitrogen dioxide, as well as meteorological data such as temperature, humidity, and wind speed.

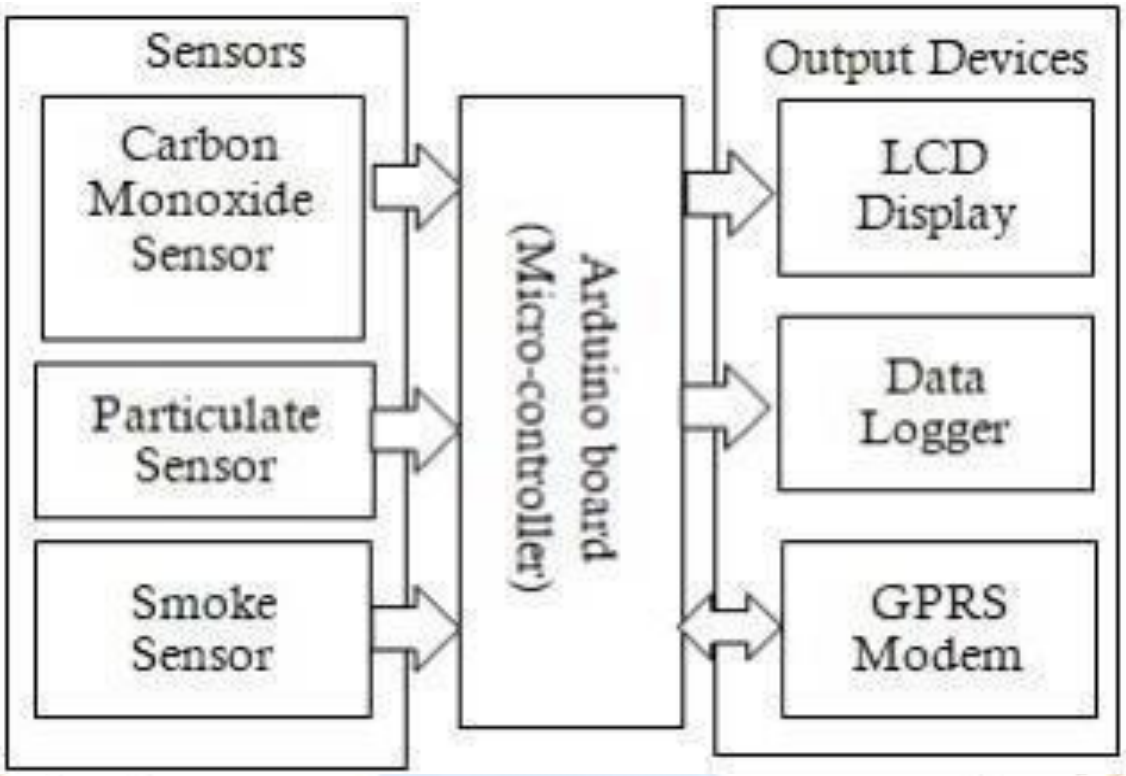
Data pre-processing: The collected data needs to be pre-processed to remove any noise or Model training: Once a model has been selected, it needs to be trained on the pre- processed data. This involves using a subset of errors, and to prepare it for analysis. This may involve data cleaning, normalization, and feature extraction.

Model selection: The next step is to select a suitable ML model for the task of air pollution detection. This may involve choosing between different types of models, such as decision trees, support vector machines, or neural networks.

The data to teach the model how to make accurate predictions of air pollution levels based on the available features.

Model evaluation: After training, the model needs to be evaluated on a separate set of data to determine its accuracy and generalization performance. This may involve using metrics such as mean squared error, root mean squared error, or coefficient of determination.

Deployment: Finally, the trained model can be deployed in a real-world setting to continuously monitor air pollution levels and provide real-time predictions. This may involve integrating the model into a mobile or web application, or into a larger environmental monitoring system.



Overall, the design of an air pollution detection system using ML requires expertise in data collection, data pre- processing, machine learning, and software engineering, as well as knowledge of environmental science and policy. Overall, air pollution detection using machine learning with PM2.5 and other molecules have the potential to significantly improve our understanding of the impact of this pollutant on human health and the environment, and to help us take steps to reduce these levels also improve air quality and develop predictive models that can forecast future levels, and to develop targeted interventions to reduce pollution levels and protect public health.

V. CONCLUSIONS

Air pollution is a serious threat to public health and the environment, and detecting and monitoring levels of pollutants in the air is essential to mitigating this threat. Air pollution detection using machine learning has emerged as a promising tool to improve our ability to monitor air quality and reduce pollution levels. Machine learning algorithms can analyze large amounts of data from various sources, including sensor data, satellite imagery, and weather models, to detect and predict air pollution levels. This information can be used to develop targeted interventions to reduce pollution levels, protect public health, and minimize environmental impact.

The use of machine learning for air pollution detection has shown promising results, including improved accuracy, early warning systems, real-time monitoring, identification of pollution hotspots, and better understanding of the impact of pollution on human health and the environment.

With the advancement of IoT infrastructures, big data technologies, and machine learning techniques, real-time air quality monitor and evaluation is desirable for future smart cities. This paper reports our recent literature study, reviews and compares current research work on air quality evaluation based on big data analytics, machine learning models and techniques.

Finally, it highlights some observations on future research issues, challenges, and needs.

In conclusion, air pollution detection using machine learning has the potential to significantly improve our ability to detect, monitor, and reduce levels of pollutants in the air, leading to improved public health and environmental outcomes. Continued research and development in this field is essential to maximize the potential of machine learning for air pollution detection and mitigation.

Finally, it highlights some observations on future research issues, challenges, and needs.

VI. REFERENCES

1. Zhang, S., Zhang, H., Li, X., & Yang, Q. (2020). Machine learning for air pollution forecasting: Model development and validation. Environmental Pollution, 257, 113558. https://doi.org/10.1016/j.envpol.2019.1135 58
2. Zhang, W., Zhang, C., Fu, C., & Liu, J. (2021). Deep Learning for Air Pollution Detection using Mobile Sensors and Social Media Data. Sensors, 21(1), 47. https://doi.org/10.3390/s21010047
3. Kumar, S., Mishra, S., Gupta, V., & Bhatia,

R. (2021). Air pollution prediction using machine learning algorithms: a review. Journal of Cleaner Production, 316, 128290. https://doi.org/10.1016/j.jclepro.2021.1282 90

4. Wei, Y., Hu, X., & Zhang, J. (2019). Air pollution forecasting with machine learning methods: A review. Science of The Total Environment, 657, 1243-1255. https://doi.org/10.1016/j.scitotenv.2018.12.439
5. Song, Y., & Kim, T. (2020). A review of machine learning methods for air pollution analysis. Journal of Cleaner Production, 267, 122153. M. Niharika and P. S. Rao, "A survey on air quality forecasting techniques," International Journal of Computer Science and Information Technologies, vol. 5, no. 1, pp.103-107,2014.
6. E. Kalapanidas and N. Avouris, "Applying machine learning techniques in air quality prediction," in Proc. ACAI, vol. 99, September 1999.
7. Questioning smart urbanism: Is data- driven governance a panacea? (November 2, 2015). [Online]. Available: http://chicagopolicyreview.org/2015/11/02

/questioningsmart-urbanis m-is-data- driven-governance-a-panacea/

8. D. J. Nowak, D. E. Crane, and J. C. Stevens, "Air pollution removal by urban trees and shrubs in the United States," Urban Forestry & Urban Greening, vol. 4, no. 3, pp. 115123, 2016. [7] T. Chiwewe and J. Ditsela,
9. "Machine learning based estimation of Ozone using spatio-temporal data from air quality monitoring stations," presented at 2016 IEEE 14th International Conference on Industrial Informatics (INDIN), IEEE, 2016.
10. Y. Zheng, X. Yi, M. Li, R. Li, Z. Shan, E. Chang, and T. Li, "Forecasting fine- grained air quality based on big data," in Proc. the 21th ACM SIGKDD International Conference on Knowledge Discovery and

Data Mining, pp. 2267-2276, August 10,2015.

11. "Qualitative and quantitative evaluation of MODIS satellite sensor data for regional and urban scale air quality," Atmospheric Environment, vol. 38, issue 16, pp. 2495–2509, May 2014.