A SURVEY ON PREDICTING AND CONTROLLING AIR POLLUTION USING MACHINE LEARNING AND DEEP LEARNING TECHNIQUES

*Ninsha V Unnikrishnan** [1], *S Mythili* [2]

ABSTRACT

Since air pollution has a significant negative influence on public health, early warning systems must be able to make precise long-term predictions about air quality. Predicting air quality has garnered a lot of interest, combining computer science, statistics, and environmental science. The Paper reviews a number of studies on air pollution, its effects on mental health, and the use of machine learning methods for air quality monitoring and prediction. Some of the studies investigate the connection between mental health outcomes and air pollution. They draw attention to the substantial influence that air pollutants, especially PM2.5, have on mental health conditions like depression and psychotic episodes. Machine learning techniques are used in many research to forecast pollution levels and air quality. To predict air quality indicators and particular pollutant concentrations, this research uses a variety of models, including Support Vector Machines (SVM), Random Forests, Neural Networks, and ensemble techniques like XGBoost. Innovative methods of environmental monitoring, such as the application of AI, IoT, and computer vision technologies, are the subject of certain studies. These technologies are used in urban settings to identify and categorize different types of pollution, including visual pollution. The combined findings of these studies highlight the substantial harm that air pollution causes to the general public, especially to mental health, and show how cutting-edge technologies can be used to monitor, forecast, and control air quality in metropolitan settings.

Department of Computer Science[1]
Karpagam Academy of Higher Education, Coimbatore, India[1]
ninshavu@gmail.com[1]
Department of Computer Science[2],
Karpagam Academy of Higher Education, Coimbatore, India[2]
smythili78@gmail.com[2]

* Corresponding Author

Keywords: Air pollution; Machine learning; Neural Network; Random Forest; XGBoost

I. INTRODUCTION

Air pollution is a serious environmental problem that has an effect on public health since it causes heart disease, respiratory illnesses, and even mental health issues. Effective monitoring and forecasting of air quality has become increasingly important as urbanization and industrial activity continue to increase. Researchers are looking into cutting-edge technology because traditional approaches to air pollution monitoring frequently have issues with accuracy and scalability. In this area, machine learning has become a potent instrument, providing data-driven methods for predicting air quality, analyzing pollution trends, and creating early warning systems. Machine learning helps to reduce the detrimental impacts of air pollution on public health by combining computer science, statistics, and environmental science to enable more accurate and timely responses.

Accurate long-term air quality forecasts are crucial given the pressing need for efficient monitoring and mitigating measures. Machine learning approaches are being used to predict and monitor air quality as a result of recent developments in computer science, statistics, and environmental science. This study examines a number of research that investigate the connection between mental health and air pollution, as well as the use of cutting-edge technologies like artificial intelligence (AI), the Internet of Things (IoT), and computer vision in the evaluation of air quality.

II. LITERATURE SURVEY

(i) Prediction of air pollution using Machine Learning.

In [1], the study focuses on employing a multi-task learning (MTL) framework to forecast hourly pollutant concentrations (such as SO_2, PM 2.5, and ozone) employing, PM2.5, and ozone) employing cutting-edge regularization techniques (such as the 2,1-norm, nuclear norm, and Frobenius norm) to improve forecast stability and accuracy. By predicting several time steps at once, it highlights temporal dependencies. On the other hand, in case of [2] the authors uses Support Vector Regression (SVR) to forecast pollutant levels and the Air Quality Index (AQI), with an emphasis on classification accuracy. Because of its capacity to represent nonlinear environmental data, the Radial Basis Function (RBF) kernel was used. [1] performs better than traditional regression models, with a focus on higher generalization and parameter reduction whereas [2] uses unseen validation to show 94.1% classification accuracy for AQI prediction

In order to support public health and urban planning, [3] is mainly focusing on predicting Tehran's PM2.5 and PM10 levels. It highlights the necessity of precise air pollution prediction models. Whereas [9] Focuses mostly on PM2.5 pollution and its effects on health, especially in emerging and heavily populated areas, emphasizing emissions from traffic and industry.

In [3] the authors make use of a variety of machine learning models, such as artificial neural networks (ANN), geo-weighted regression (GWR), support vector machines (SVM), and autoregressive nonlinear neural networks (ARNN). Also presents ARNN, a new improved prediction model that achieves a 94% error reduction in forecasting. To determine the important environmental factors influencing air pollution, [3] make use of genetic algorithms. In case of [9] authors use logistic regression and autoregression (AR) to anticipate and monitor air quality. Autoregression forecasts future PM2.5 levels based on past trends, whereas logistic regression categorizes pollution status. Thus, by using localized, real-world atmospheric data the paper increases the accuracy of the model.

An 11-year dataset from Taiwan's Environmental Protection Administration (EPA) is used in [10] to allow for a long-term examination of trends in air quality. The large dataset improves forecast accuracy by accounting for seasonal fluctuations and other important factor, whereas [12] employs bibliometric analysis, examining more than 900 peer-reviewed publications from 1990 to 2022. This study charts thedevelopment of research trends, important contributors, and international cooperation in machine learning applications for air pollution studies rather than concentrating on direct air pollution predictions. Since [6] uses data on air pollution from 23 Indian cities over a six-year period, it is extremely pertinent to emerging countries where air pollution is a major public health issue.

Support Vector Machines (SVM), Random Forests, Stackin g Ensembles, Adaptive Boosting (AdaBoost), and Artificial Ne ural Networks (ANN) are among the machine learning method s that are assessed in [10]which comes to the conclusion that AdaBoost is best at reducing mean absolute error (MAE), whi le the Stacking Ensemble model is the most successful, obtaini ng the highest R2 and lowest RMSE. But in case of methodology [12] offers a meta-analysis of research trends in air pollution prediction rather than implementing machine learning models. The report underlines the increasing importance of particulate matter prediction and lists important approaches utilized in the field. Gaussian Naive Bayes, Support Vector Machine, XGBoost, and five other machine learning models are compared in [6]. According to the study, XGBoost outperforms Gaussian Naive Bayes, which is efficient but lacks predictive precision, in terms of air pollution level prediction.

Though it could be enhanced with real-time data integration and wider geographic coverage, [10] provides a useful guidance for choosing machine learning models based on performance measures. Although[12] is a useful tool for scholars and decision-makers, the lack of forecasting models makes it inapplicable in the actual world. In the case of [6], the authors successfully apply machine learning to actual data, it would profit from more in-depth deep learning research as well as the inclusion of outside variables like the weather and legislative changes.

(ii) Analysing Health consequences from Ai pollution.

In order to observe the long-term, real-world consequences of air pollution on mental health, this [4] uses a natural experiment. Researchers can investigate the effects of pollution over time in an unaltered, natural environment by conducting a natural experiment. Participants are tracked for ten years. According to the study, there is a direct link between air pollution and clinical depression, with even slight increases in pollution being associated with a nearly 1% higher risk of getting depression. Long-term mental health results are highlighted here.

Whereas, Panel data regression is used in [5], and data from the China Family Panel Surveys (2010 and 2014) are used. The study [5] uses a sizable sample size (52,568 observations) to investigate the connection between mental health and air pollution. According to the [5], which measures mental health using the CES-D scale, there is a modest decline in CES-D scores for every 1 μg/m3 increase in PM2.5, which may indicate a detrimental effect on mental health. This study also shows that smokers and those with lower incomes are more susceptible to the negative effects of air pollution on mental health.

Because the study's [4] primary focus is clinical depression, it might not adequately account for the variety of mental health consequences associated with air pollution, such as anxiety, stress, and cognitive deterioration. In contrast to study [5], Study [4] doesn't address how pollutants and personal habits like smoking and exercise may combine to impact mental health. The study [5] examines smoking, but it skips over other health behaviors (such as diet and exercise) that may alter the relationship between pollution and mental health. Potential confounding factors, such as genetic predisposition or other environmental stresses that can influence the relationship between pollution and mental health outcomes are disregarded in this study.

Panel data regression is used in [5] on a large dataset that includes 52,568 observations from the China Family Panel Surveys (CFPS) conducted between 2010 and 2014. The study focuses on PM2.5 concentrations as a measure of air pollution and the CES-D scale to assess mental wellbeing. According to key findings, CES-D scale scores dropped by 0.012 for every 1 μg/m3 increase in PM2.5 concentration, indicating deteriorating mental health. The study also finds that smoking makes the detrimental impacts of air pollution on mental health worse, with low-income populations being more at risk.

Whereas, [11] employs a prospective longitudinal methodology, tracking 1,698 people in South East London over a five-year span (2008–2013). The impact of several air pollutants, such as NO2, NOx, O3, PM10, and PM2.5, on mental health outcomes is assessed in this study. The authors adjust for socioeconomic and environmental variables, such as exposure to traffic noise, and employ high-resolution air pollution data that is connected to the individuals' residential locations. The main conclusions of [11] demonstrate that while PM10 has the largest correlation with psychotic experiences (33% increase), elevated levels of PM2.5, NOx, and NO2 are linked to an 18–39% higher chance of mental health problems. The study emphasizes in particular that there were higher correlations between continuing pollution exposure and declining mental health among those who did not move during the study period.

[5] has certain limitations even if it successfully illustrates the long-term effects of PM2.5 exposure on mental health. Because it is observational in nature, it does not demonstrate causality, it does not include information on health behaviors other than smoking, and it leaves out potential confounding variables that might influence the relationship between pollution and mental health. While [11] suggests that their results have implications for public health and urban planning policy, indicating that enhancing the quality of the air in busy places may significantly improve mental health.

(iii) Analysing methods used for prediction of air pollution using deeplearning.

The main focus of [7] was Predicting air pollution levels (PM2.5) in Beijing but in [8] the authors concentrated on Identifying and classifying visual pollution in urban Dhaka, and in [13] authors contributed an innovative way of detecting air pollution by detecting and categorizing visual pollution on public highways. Different sorts of methodology was implemented in the three papers: [7] was used Hybrid CNN- LSTM deep learning model using meteorological and historical pollutant data, [8] used Deep learning-based object detection (YOLOv5, YOLOv7, Faster SegFormer) with Google Street View and [13] used Deep Active Learning (DAL) with YOLO- based VPP system for real-time detection.

While all three research employ deep learning to monitor the environment, Study [7] uses hybrid models to forecast air pollution, while research [8] and [13] use high-precision classification models to identify visual pollution. Study [13] incorporates deep active learning to increase accuracy and efficiency, whereas Study [8] offers a novel, inexpensive real- time solution. While Studies [8] and [13] support infrastructure management and urban aesthetics, In Study [7] author's forecasting methodology has a significant impact on public health. Future studies should look into combining thesestrategies and using hybrid AI models for real-time interventions and multi-dimensional pollution analysis.

III. CONCLUSION

Air pollution continues to pose a serious risk to public health, impacting both mental and physical health. The way we evaluate and react to pollution has been completely transformed by the use of machine learning into air quality monitoring and prediction. Researchers and policymakers can create more precise forecasting systems that allow for prompt responses and improved urban planning by utilizing sophisticated models and data-driven methodologies.

All of the papers that were evaluated highlight how seriously air pollution affects public health, especially when it comes to its association with mental health issues. Predicting air quality and monitoring the environment have been greatly improved by the application of machine learning models and new technologies, which present encouraging ways to reduce pollution in urban areas. Policymakers and researchers may create better early warning systems and intervention plans to safeguard public health and enhance urban air quality management by utilizing these cutting-edge techniques.

REFERENCES

[1] Dixian Zhu, Changjie Cai, Tianbao Yang and Xun Zhou, “A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization,” Big data and cognitive computing, 2018.

[2] Mauro Castelli ,Fabiana Martins Clemente, Aleˇ sPopoviˇ, Sara Silva, and Leonardo Vanneschi, A Machine Learning Approach to Predict Air Quality in California, Hindawi Complexity Volume 2020, 2020.

[3] Mahmoud Reza Delavar, Amin Gholami, Gholam RezaShiran,Yousef Rashidi Gholam Reza Nakhaeizadeh, Kurt Fedra and Smaeil Hatefi Afshar, “A Novel Method for Improving Air Pollution Prediction Based on Machine Learning Approaches: A Case Study Applied to the Capital City of Tehran”, Internation journal of geo information, 2019

[4] Younoh Kim, James Manley, and Vlad Radoias, “Air Pollution and Long Term Mental Health”, Antony.D, MDPI, 2020.

[5] Zhiming Yang, Qianhao Song, Jing Li, Yunquan Zhang, Xiao-Chen Yuan, Weiqing Wang and Qi Yu, “Air pollution and mental health: the moderator effect of health behaviors”, Environmental Health Letters, 2021.

[6] K. Kumar, B. P. Pande, “Air pollution prediction with machine learning: a case study of Indian cities”, International Journal of Environmental Science and Technology, 2022

[7] Abdellatif Bekkar, Badr Hssina, Samira Douzi and Khadija Douzi“, Air pollution prediction in smart city, deep learning approach”, Journal of Big Data, 2021.

[8] Md Fahim ShahoriarTitu , Abdul Aziz Chowdhury , S. M. Rezwanul Haque and Riasat Khan, “Deep-Learning-Based Real-Time Visual Pollution Detection in Urban and Textile Environments”, MDPI, 2024.

[9] 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, 2018

[10] Yun-Chia Liang, Yona Maimury, Angela Hsiang-Ling Chen and Josue Rodolfo Cuevas Juarez, "Machine Learning-Based Prediction of Air Quality", MDPI, 2020.

[11] Ioannis Bakolis, Ryan Hammoud, Robert Stewart, Sean Beevers, David Dajnak, Shirlee MacCrimmon, Matthew Broadbent, Megan Pritchard, NarushigeShiode, Daniela Fecht, John Gulliver, Matthew Hotopf, Stephani L. Hatch, Ian S. Mudway, "Mental health consequences of urban air pollution: prospective population based longitudinal survey", Social Psychiatry and Psychiatric Epidemiology, 2020.

[12] Shikha Jain, Navneet Kaur, Sahil Verma, Kavita, A.S. M. Sanwar Hosen, and Satbir S Sehgal, "Use of Machine Learning in Air Pollution Research: A Bibliographic Perspective", MDPI, 2022

[13] Mohammad AlElaiwi, Mugahed A. Al-antari, Hafiz Farooq Ahmad, Areeba Azhar, Badar Almarri and Jamil Hussain, "VPP: Visual Pollution Prediction Framework Based on a Deep Active Learning Approach Using Public Road Images", MDPI, 2022