A Personalized Healthcare Model for Air Pollution Monitoring and Prediction Using Machine Learning

Gade Navaneeth, Medi Praneeth, Mohammed Khasim and Y. Mohana Roopa
Department of CSE, Institute of Aeronautical Engineering, Gandimaisamma, Hyderabad, 500043, Telangana, India

Keywords: Personalized Healthcare, Air Pollution Prediction, Web Scraping, Linear Regression, AI-driven Health Recommendations, Air Quality Forecasting, Machine Learning.

Abstract: Air pollution, as a major contributor to respiratory, cardiovascular, and other chronic diseases, is a significant environmental and public health problem. This paper developed a personalized healthcare model based on air quality prediction and individual health recommendations-finally trained on machine learning skills and based on time-series data. Unlike traditional air quality monitoring systems that measure particulate matter and other gases, our approach is based on a Linear Regression algorithm that predicts PM2 levels and other pollutants. 5 by merging web-scraped infectious disease information with meteorological data. The model operates as a web-based application that allows users to enter their location to receive personalized health precautions and real-time air quality forecasts. The guidelines, which address the estimated risk of air pollution, tell consumers how to act preventatively, and what they should do to protect themselves, like staying indoors or using air filtration devices, they added. By combining real-time forecasts with individual participation, the system enables proactive health management and increases public awareness. The proposed methodology reports on the use of AI-driven insights to facilitate an evidence-based and straightforward approach for mitigating the negative health effects of air pollution.

1 INTRODUCTION

Air pollution, a leading cause of severe respiratory and cardiovascular ailments, is increasingly recognized as an important global health threat (Wang, X., Li, L., and Chen, Z., 2020). The intensification of human activity in urban areas and industrial areas has already started corroding the environment, even more so particularly in recent two decades, not limited to air pollutants such PM2. 5, PM10, NO_2, and SO_2 have become more serious, aggravated the baseline risk in the population and promote the risk of disease (P. Huang, et al., 2022). Accurate air quality prediction is essential for mitigating these hazards, and enabling people to take preventive action for their health (S. S. Kumar., et al., 2024).

This work presents a personalized healthcare platform that integrates health support with machine learning-based air quality prediction (S. Kumari., et al., 2025). This approach focuses on exposure at the individual level, rather than general environmental monitoring, by allowing users to input their location and receive real-time predictions of pollutants (I. Gryech., et al., 2024). It predicts pollution levels via Linear Regression using past meteorological data and also scrapes the web for air quality data. To enhance usability, the model is implemented as a web-based application that provides users with personalized health advice based on expected air quality (Kekulanadara., et al., 2021). These recommendations advise individuals to take precautionary actions such as wearing masks, staying indoors, or minimizing outdoor activitie during times of high pollution levels (R. Buvana., et al., 2022). This use of AI-driven data creates the base for a more proactive approach to air pollution reduction, by empowering the people to take educated health decisions based on the information (A. Mittal., et al., 2024).

2 RELATED WORKS

Air quality is an important health global health problem (1) and pollutants like PM2. 5, PM10, NO_2, and SO_2 relevant to respiratory and cardiovascular diseases. Thanks to machine learning and artificial

Navaneeth, G., Praneeth, M., Khasim, M. and Roopa, Y. M.
A Personalized Healthcare Model for Air Pollution Monitoring and Prediction Using Machine Learning.
DOI: 10.5220/0013935800004919

In *Proceedings of the 1st International Conference on Research and Development in Information, Communication, and Computing Technologies (ICRDICCT'25 2025) - Volume 5*, pages 575-579
ISBN: 978-989-758-777-1

intelligence technology, long-term air quality monitoring and personalized health advice by predictive models and data-driven methods have emerged as new avenues for research (S. B. Kasetty and S. Nagini., 2022). We divide this literature review into three sections that relate directly to the work in our study: machine learning in predicting air pollution, web scraping for data collection and AI-powered health recommendations.

2.1 Air Pollution Prediction Using Machine Learning

Historical weather data is often invaluable for improving the accuracy of predictions, as machine learning techniques have shown great promise in air pollution forecasting A personalized healthcare model (Behal and Singh, 2020) which integrates various machine learning techniques to predict the level of air pollution and its impact on parameters of health of individuals is proposed. By utilizing both environmental sensing data and ML algorithms, their study improved the spatial and temporal resolution of air quality predictions with personalized health recommendations. The ultimate goal of the machine learning research task is to create more accurate predictions of air pollution and develop AI-based health systems.

2.2 Web Scraping for Environmental Data Collection

Web scraping is behind accurate air quality prediction that relies on extensive datasets. Once published these methods let alone the huge-scale data they demand for the accurate time forecasting. It provided an overview of the significance of web scraping in automating data extraction processes and converting unstructured data available on the web into meaningful, stored, and analysable data (Sirisuriya 2023). Their research showed how web scraping can facilitate real-time updates to environmental data, which helps to retrain machine learning models using the most up-to-date pollution metrics. Such an approach can greatly enhance the responsiveness and accuracy of AI-based air quality monitoring solutions.

2.3 AI-Driven Health Recommendations Based on Pollution Exposure

AI-driven systems are increasingly being used to assess environmental risks and provide personalized health interventions. Olawade et al. (2024) highlighted how AI technologies enhance environmental monitoring by enabling pollution source detection, disaster forecasting, and air quality monitoring. Their research emphasized the role of predictive analytics in mitigating health risks, allowing individuals to take precautionary actions based on real-time pollution levels. Despite challenges such as data accessibility and privacy concerns, their study underscored the potential of AI to revolutionize public health strategies by integrating pollution forecasting with personalized healthcare recommendations (Y Mohana Roopa., et al., 2023, 2022).

3 METHODOLOGY

Thus, the personalized healthcare model developed for air pollution monitoring and prediction consists of a methodological framework entailing data collection, preprocessing, AQI prediction through machine learning approaches, correlation analysis between AQI levels and health conditions, and finally, health recommendation generation. This approach enables accurate prediction of air quality while offering personalized healthcare recommendations relative to the air pollution levels.

3.1 System Architecture

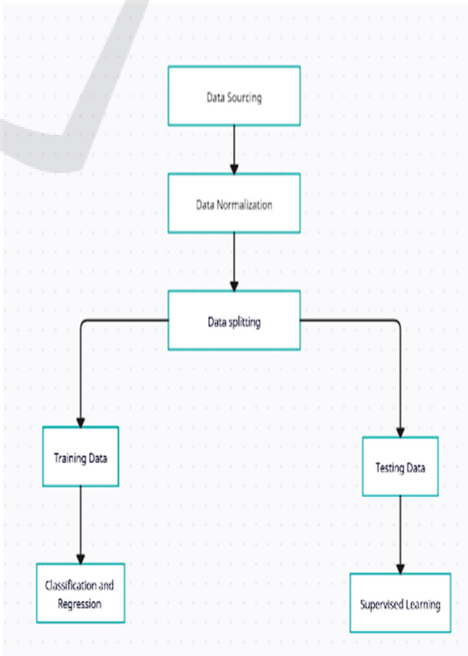

Figure 1: System Architecture.

This approach leverages the latest technologies to create an automated and intelligent air quality monitoring system that encompasses data collection, analysis, and predictive modelling. It includes different modules like data acquisition, pre-processing, machine learning model training, user interface integration, etc. The overall system architecture is shown in Figure 1.

3.2 Data Collection and Preprocessing

Historical and real-time air pollution data were collected from online sources such as Open Weather API using web scraping techniques. The dataset includes various meteorological and pollution-related parameters such as PM2.5, NO2, CO, temperature, humidity, precipitation, and wind speed.

The raw dataset undergoes preprocessing, including data cleaning, handling missing values, and normalization. The data is then aggregated into daily average values for the years 2013 to 2017. The cleaned dataset is structured to ensure consistency and readiness for machine learning analysis. Table 1 shows the Dataset Attributes Used for AQI Prediction.

Table 1: Dataset Attributes Used for AQI Prediction.

Feature	Description	Unit
PM2.5	Fine particulate matter concentration	µg/m³
NO2	Nitrogen dioxide concentration	ppb
CO	Carbon monoxide concentration	ppm
Temperature	Average daily temperature	°C
Humidity	Relative humidity	%
Wind speed	Wind speed at ground level	m/s
Precipitation	Rain or snow precipitation	mm

3.3 Machine Learning Model for AQI Prediction

The system employs supervised learning techniques, specifically Linear Regression, to predict future AQI levels. The dataset is divided into training (70%) and testing (30%) sets, ensuring model generalizability. Feature selection techniques are applied to retain the most significant parameters influencing air pollution trends.

The trained model evaluates its performance using error metrics such as Root Mean Squared Error (RMSE) and R^2 Score, ensuring the reliability of AQI predictions. The results are visualized to compare actual vs. predicted AQI values over time.

3.4 Personalized Healthcare Recommendations

Using predicted AQI readings, it generates customized healthcare recommendations. It then offers users precautionary advice depending on how severe the level of pollution exposure is. When pollution levels peak, users are notified as well as receive recommendations to reduce outdoor activities, wear protective masks and turn on air filtration systems.

3.5 Web-Based Deployment and User Interaction

I mean integrated model into the platform defined in web user models inputs the location and gets real time model prediction with health sector prediction and real time prediction. It securely authenticates the users, enabling them to access environmental conditions, pollution advisories, and health information. This dynamic updating helps it get a sense of how pollutants and dust flow around, so that it can be aware of pollution in real time.

3.6 Implementation

The personalized healthcare model for air pollution monitoring and prediction consists of three main parts, which are data collection, ML model development, and web deployment. The system ingests real-time pollution data, forecasts air quality levels, and suggests health care recommendations

It is built with Python for the data handling and machine learning aspect of running the text analysis, and leverages Scikit-learn for building and testing the models. Beautiful Soup is used to scrape the web for pollution and Meteorology data from datasets and open Weather API. The data from pollution will be stored in the structured form by using the SQL which is efficient to retrieve and management of the database. For the web-based implementation, I use the Flask framework for the backend API and React. as for the front-end allows the user to interact with the system without any disturbance.

It consists of a well-defined workflow that processes environmental data and generates custom insights. Data Collection: Obtain pollution and weather data from datasets and open Weather API.27 Data Pre-processing: Handle missing values if any, and normalize the data for consistency. The model training phase is executed using Linear Regression. After training, the system generates AQI forecasts along with healthcare recommendations, thus making users aware of pollution risks and how to take precautionary measures. The system receives input from users via a web-based interface, in which they share their location details and get updates on real-time pollution in their area.

Using Pickle, the trained model is exported to be used just fine with the Flask-based backend API. The API enables real-time predictions by processing user inputs to give AQI forecasts as well as personalized health suggestions. React -- the frontend. js in this project settles for easy-to-use interface for users to insert their location and visualize pollution trends.

4 RESULTS AND EVALUATION

We developed and implemented a smart and personalized healthcare model for air pollution monitoring and prediction with the following contributions: (1) reliable prediction of the PM2.5 AQI level and (2) personalized health advice. In the evaluated time period the air pollution we have trained a Linear Regression model on, are able to determine the best combination of parameters with air quality values. The red below the AQI prediction graph (Figure 2) indicates smooth variance between actual and prediction, demonstrating this system's predictive capability and therefore reliability in real-world scenarios.

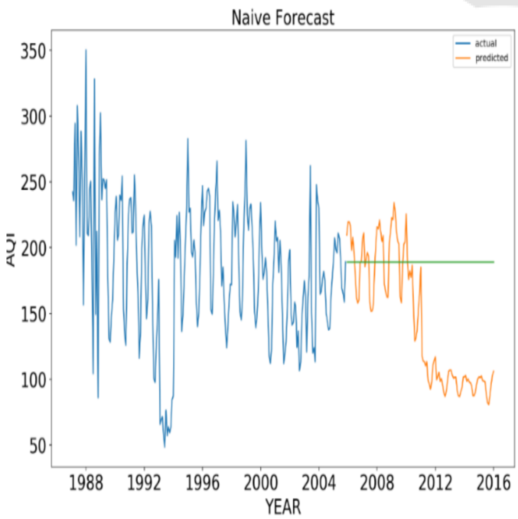

Figure 2: AQI Prediction Graph.

Moreover, correlation analysis between industrial waste emissions and AQI levels, indicates that there is a strong positive correlation between the two as well, implying that industrial emissions contribute significantly towards worsening air quality. This relationship is depicted in the AQI vs Industrial Waste Emissions Graph (Figure 3), which demonstrates how industrial emissions contaminate the air. The system also features a web-based user interface to input their location, view AQI prediction results, and get real-time health recommendations.

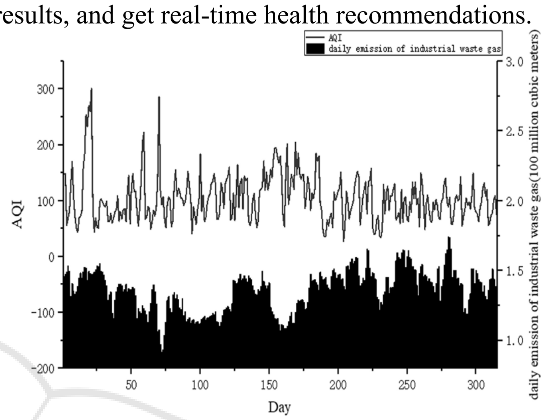

Figure 3: AQI and Industrial Waste Gas Emissions.

5 DISCUSSION

The system suggest which healthcare precautions users can take, such as adjusting their mask use, reducing their exposure to the outdoors or increasing filtration to improve their indoor air quality; the advice changes dynamically based on the level of pollutants. This interactive dashboard provides a user-friendly visualization of pollution trends, along with any matching health advisories. While the model can make accurate predictions, there are other areas to improve in the future by setting up additional machine learning models to increase prediction accuracy, obtaining real-time pollutant sensor data along with recommendations based on users' health risks. This will augment the system's capacity to combat pollution-induced health hazards and enhance environmental awareness in general.

6 CONCLUSIONS

A personalized healthcare model for air pollution monitoring and prediction using machine learning is presented through this study. By investigating historical and contemporary environmental data, the system accurately forecasts AQI with health advisories. The visually strong correlations between real return AQI values and Linear Regression predicted AQI values indicate a reliable predictive performance, so we can go ahead and move on to air pollution datasets training. Moreover, correlation analysis validates the influence of industrial discharge to the air, emphasizing the need for pollution stormers.

It moves here closer to users by building a web-based interface that lets users obtain health recommendations based on the AQI in their area. Depending on the pollution, the system decides real-time pollution precaution measures according to real-time pollution severity for users. The interactive dashboard additionally increases the interaction of users as the trends of pollution and their relevant health risk have been displayed in a user-friendly way.

While this model has a lot of positive impact in terms of forecasting air pollution and providing health advice, further enhancements can be made in terms of exploring additional machine learning techniques to improve model accuracy, integrating live sensors data, and giving offer health advice according to user-specific medical history.

REFERENCES

A. Mittal, S. Arora, S. Sharma and A. Garg, "Advancements in Air Pollution Prediction and Classification Models: Exploring Emerging Frontiers," 2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS), Coimbatore, India, 2024.

D. B. Olawade, O. Z. Wada, A. O. Ige, B. I. Egbewole, A. Olojo, and B. I. Oladapo, "Artificial intelligence in environmental monitoring: Advancements, challenges, and future directions," Hygiene and Environmental Health Advances, vol.12,2024

I. Gryech, C. Asaad, M. Ghogho, and A. Kobbane, "Applications of Machine Learning & Internet of Things for Outdoor Air Pollution Monitoring and Prediction: A Systematic Literature Review," Engineering Applications of Artificial Intelligence, vol. 137, Part B, 2024

J. J. Bosco and V. Kowsalya, "A Novel Approach for Air Pollution Prediction Using Machine Learning Techniques," 2024 Third International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT), Trichirappalli, India, 2024.

K. M. O. V. K. Kekulanadara, B. T. G. S. Kumara and B. Kuhaneswaran, "Machine Learning Approach for Predicting Air Quality Index," 2021 International Conference on Decision Aid Sciences and Application (DASA), Sakheer, Bahrain, 2021

P. Huang, K. Kim, and M. Schermer, "Ethical Issues of Digital Twins for Personalized Health Care Service: Preliminary Mapping Study," Journal of Medical Internet Research, vol. 24, no. 1, 2022

R. Buvana, A. Sathya, R. Chathana, C. Jaysri and D. Jashna, "Air Pollution Prediction with Machine Learning," 2022 1st International Conference on Computational Science and Technology (ICCST), CHENNAI, India, 2022.

S. B. Kasetty and S. Nagini, "A Survey Paper on an IoT-based Machine Learning Model to Predict Air Pollution Levels," 2022 4th International Conference on Advances in Computing, Communication Control and Networking (ICAC3N), Greater Noida, India, 2022.

S. D. S. Sirisuriya, "Importance of Web Scraping as a Data Source for Machine Learning Algorithms - Review," 2023 IEEE 17th International Conference on Industrial and Information Systems (ICIIS), Peradeniya, Sri Lanka, 2023.

S. S. Kumar, R. Chandra, and S. Agarwal, "Rule based complex event processing for an air quality monitoring system in smart city," Sustainable Cities and Society, vol. 112, 2024

S. Kumari, A. K. Tyagi, S. Tiwari, and G. Kulshreshtha, "Sensors and Digital Twin for Healthcare Monitoring Frameworks," in Digital Twin, Blockchain, and Sensor Networks in the Healthy and Mobile City, T. A. Nguyen, Ed. Elsevier, 2025

Wang, X., Li, L., and Chen, Z., "Prediction of air quality based on machine learning algorithms," Environmental Research, Vol. 188, pp. 109-118, 2020.

Y Mohana Roopa et, al," Prediction Evaluation of Gene Ontology Using Support Vector Machine", International Journal on Recent and Innovation Trends in Computing and Communication, June 2023, 11, pp. 522–526

Y. Mohana Roopa et.al," Deep Learning-Based Hybrid Framework Utilizing OpenCV and CNN for Automated Brain Tumor Detection and MRI Image Classification", 2024 4th International Conference on Sustainable Expert Systems (ICSES) on 15-17 October 2024 at Kaski, Nepal

Y. Mohana Roopa,et.al," A secured IoT-based model for human health through sensor data", Measurement: Sensors, Elsevier publisher, ISSN: 2665-9174 (Online), Volume 24, December 2022Page No 50-58