IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Climate-Aware Respiratory Risk Forecasting System

[1]R. Sri Divya, [2]P. Jaya Chaitanya, [3]Ch. Lakshmi Keertana, [4]Ch. Surya Prakash, [5]M. Vidya Sri,

[1]Assistant Professor, [2]Final Year B.Tech Student, [3]Final Year B.Tech Student, [4]Final Year B.Tech Student, [5]Final Year B.Tech Student

Department of CSE – ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Aditya College of Engineering and Technology (A), Surampalem, Andhra Pradesh, India.

Abstract: Air pollution has become a significant public health concern, especially for individuals suffering from respiratory diseases such as asthma and bronchitis. Existing air quality monitoring systems mainly provide real-time data and lack predictive capabilities for assessing future health risks. This paper presents a climate-aware respiratory risk forecasting system that utilizes environmental parameters including Air Quality Index (AQI), PM2.5, PM10, temperature, and humidity to predict respiratory risk levels using machine learning techniques such as Random Forest and XGBoost. The system classifies risk levels and provides timely alerts along with health recommendations to support preventive measures. The proposed approach improves decision-making by enabling early prediction of respiratory risks and enhancing public health awareness.

Index Terms **-** Air Quality Index (AQI), Respiratory Risk Prediction, Machine Learning, Environmental Data, Random Forest, XGBoost.

Introduction

Air pollution has become one of the most critical environmental challenges affecting public health worldwide. Rapid urbanization, industrialization, and increasing vehicular emissions have significantly contributed to the deterioration of air quality. Pollutants such as particulate matter (PM2.5 and PM10), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), and carbon monoxide (CO) have a direct impact on human respiratory health, leading to diseases such as asthma, bronchitis, and chronic obstructive pulmonary disease (COPD) [7]. Continuous exposure to these pollutants increases health risks and reduces overall quality of life.

The Air Quality Index (AQI) is widely used as a standard indicator to measure and represent air quality based on pollutant concentrations [6]. It helps in understanding pollution levels and their potential health effects. However, most existing air quality monitoring systems focus on providing real-time or historical AQI data. While this information is useful, these systems lack predictive capabilities and do not provide insights into future air quality conditions or associated health risks [2].

With the advancement of machine learning techniques, predictive models have been increasingly used for analyzing environmental data and forecasting AQI levels. Algorithms such as Random Forest and Extreme Gradient Boosting (XGBoost) have shown high accuracy in handling complex and non-linear environmental datasets [1], [8]. These techniques enable better prediction of air quality trends compared to traditional statistical methods.

To address the limitations of existing systems, this study proposes a climate-aware respiratory risk forecasting system that integrates environmental data with machine learning techniques to predict respiratory risk levels in advance. The system utilizes parameters such as AQI, particulate matter, temperature, and humidity to analyze patterns and forecast potential health risks. In addition to prediction, the system provides health advisories and precautionary measures to help users reduce exposure to harmful pollutants.

Furthermore, the system includes a route optimization feature that analyzes pollution levels across different travel paths and recommends safer routes with lower AQI exposure. By incorporating predictive analytics into air quality monitoring, the proposed approach supports proactive healthcare decision-making, enhances environmental awareness, and contributes to reducing the impact of air pollution on public health.

II. RELATED WORK

Air quality monitoring has been extensively studied to analyze environmental pollution and its impact on human health. Traditional systems rely on data collected from environmental monitoring stations, which measure pollutant concentrations such as PM2.5, PM10, NO_2, SO_2, and CO. These values are used to compute the Air Quality Index (AQI), which provides a standardized representation of air quality levels. Such systems provide real-time and historical information, enabling users to understand current pollution conditions; however, they are primarily descriptive and do not include predictive analysis for future air quality trends [6].

With the increasing availability of environmental datasets, recent research has focused on the use of data-driven approaches for air quality prediction. These datasets typically include both pollutant concentration data and meteorological parameters such as temperature, humidity, wind speed, and atmospheric pressure. By analyzing historical patterns in these datasets, researchers have developed predictive models to estimate AQI values and identify pollution trends over time [2].

Among various machine learning approaches, ensemble learning techniques have shown significant improvements in prediction accuracy. In particular, Extreme Gradient Boosting (XGBoost) has emerged as a highly effective method for air quality prediction due to its ability to handle large datasets, manage missing values, and capture complex non-linear relationships between environmental variables [1],[8]. XGBoost utilizes gradient boosting principles to iteratively improve model performance and reduce prediction errors, making it suitable for real-time forecasting applications.

Several studies have also explored the application of advanced data preprocessing techniques such as feature selection, normalization, and handling of missing data to improve the performance of prediction models. These preprocessing steps play a crucial role in enhancing the quality of input data and ensuring more accurate AQI predictions.

In addition to prediction models, research has also emphasized the importance of real-time forecasting systems that combine environmental and meteorological data for improved accuracy. Such systems aim to provide timely insights into air quality conditions and support decision-making processes related to environmental management and public safety.

Furthermore, studies have investigated the relationship between air pollution and respiratory health outcomes. These works highlight the strong correlation between exposure to pollutants and the occurrence of respiratory diseases such as asthma and bronchitis, emphasizing the need for accurate prediction and early warning systems [7].

Overall, existing research primarily focuses on air quality monitoring and prediction using environmental datasets and machine learning techniques, with XGBoost emerging as one of the most reliable and efficient approaches for AQI forecasting. However, most approaches are limited to prediction tasks and do not fully incorporate comprehensive health-related analysis within a single framework.

III. SYSTEM ARCHITECTURE

The system architecture describes the overall structure and interaction of different components involved in the climate-aware respiratory risk forecasting system. The system is designed as a web-based application that integrates environmental data collection, machine learning-based prediction, risk assessment, and user interaction.

The architecture consists of multiple interconnected components that work together to process environmental data and generate meaningful outputs. Initially, the system collects environmental parameters such as PM2.5, PM10, NO_2, SO_2, CO, temperature, and humidity from external web services and APIs, which provide real-time information based on the user's location.

The collected data is then processed to ensure it is suitable for prediction. This includes operations such as cleaning, normalization, and preparation of input features. The processed data is then used by the machine learning component, where the XGBoost model predicts the Air Quality Index (AQI). The predicted AQI values are further classified into standard categories such as Good, Moderate, Poor, and Severe.

Based on the predicted AQI values and environmental conditions, the system evaluates respiratory risk levels and provides health-related insights and precautionary measures. Additionally, the system analyzes pollution levels across different travel paths and identifies routes with lower AQI exposure to help users reduce their exposure to polluted environments.

Finally, the results, including AQI predictions, risk levels, and recommendations, are presented through a user-friendly interface, ensuring effective communication of information to the user. Overall, the system architecture integrates data collection, processing, prediction, and output generation into a unified framework for efficient respiratory risk forecasting.

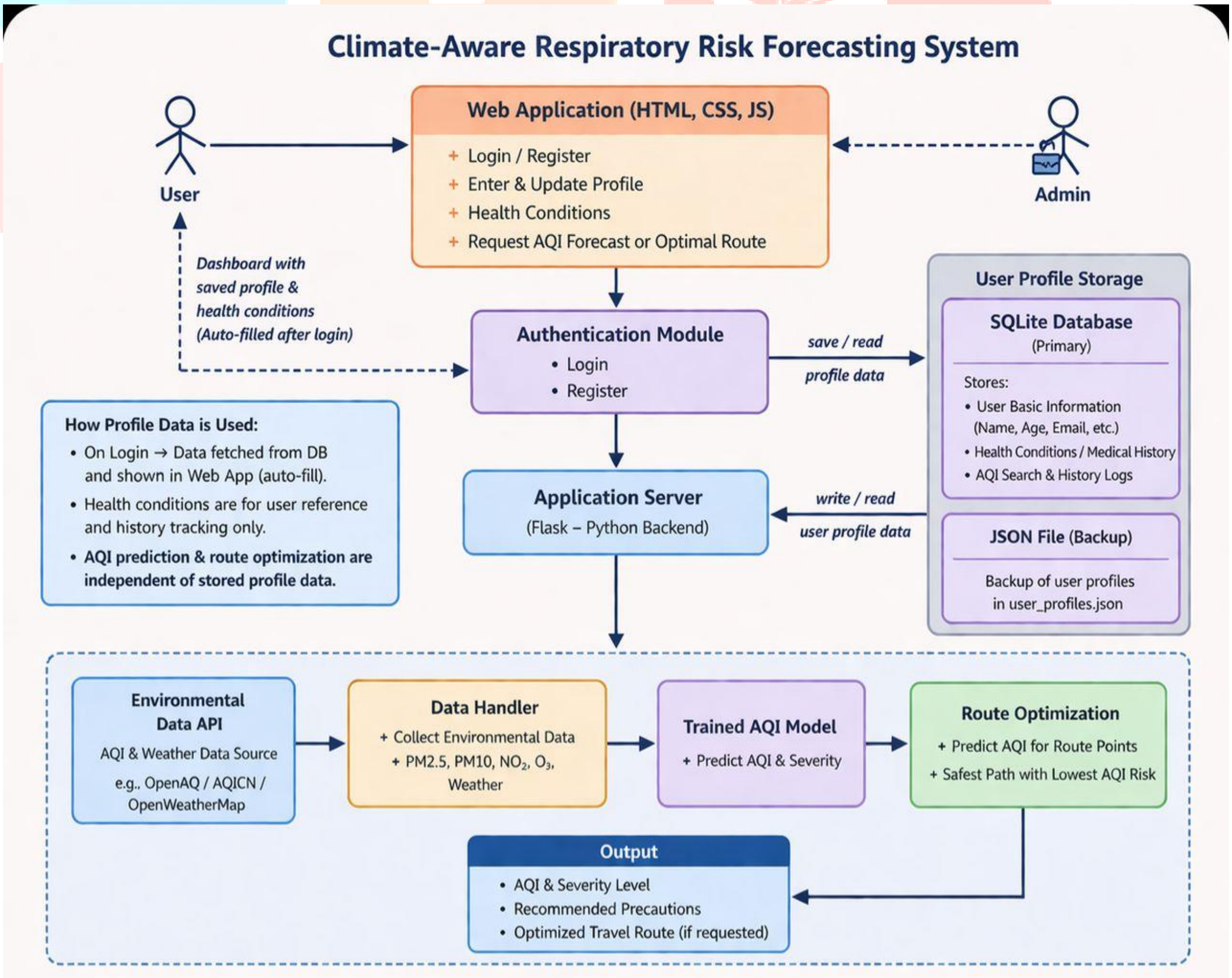


Figure 3.1: System Architecture

IV. IMPLEMENTATION

The system is implemented as a web-based application that integrates environmental data collection, machine learning-based prediction, and user interaction. The implementation focuses on processing environmental parameters and generating accurate AQI predictions along with respiratory risk insights.

A. Tech Stack

The system is developed using a combination of programming languages, frameworks, and tools. Python is used as the primary programming language for backend development and machine learning implementation. Flask is used as the backend framework to handle server-side operations. The frontend is developed using HTML, CSS, and JavaScript for user interaction and visualization. Libraries such as Pandas and NumPy are used for data processing, and XGBoost is used for AQI prediction. SQLite and JSON are used for data storage, while external APIs are used for retrieving environmental data.

Table I summarizes the complete technology stack used in Climate Aware Respiratory Risk Forecasting System

Table I: Tech Stack and Tools of Climate Aware Respiratory Risk Forecasting System

Component	Technology Used	Purpose
Programming Language	Python	Used for backend development and ML model
Backend Framework	Flask	Handles server-side logic and API requests
Frontend	HTML, CSS, JavaScript	Provides user interface and interaction
Machine Learning	XGBoost	Predicts AQI based on environmental data
Data Processing	Pandas, NumPy	Performs data cleaning and preprocessing
Database	SQLite / JSON	Stores user data and system information

APIs	Environmental Data APIs	Fetches real-time environmental parameters

B. Backend

The backend is implemented using Python and Flask. It acts as the core component of the system, handling user requests, retrieving environmental data, and performing data processing. The backend also integrates the trained XGBoost model, which takes environmental parameters as input and predicts AQI values. It ensures proper communication between different components and returns results to the frontend.

C. Frontend

The frontend is developed using HTML, CSS, and JavaScript, providing a user-friendly interface. It allows users to enter location details, view AQI predictions, and access health recommendations. The interface is designed to present information clearly and enable smooth interaction with the system.

D. Data Processing

Environmental data collected from APIs is preprocessed before being used for prediction. This includes handling missing values, organizing input features, and converting data into suitable formats. Proper preprocessing ensures that the machine learning model receives accurate and consistent input.

E. Machine Learning Model

The system uses the XGBoost algorithm for AQI prediction. The model is trained using environmental datasets, where pollutant parameters are used as input features and AQI values are used as output. The trained model predicts AQI values, which are further classified into categories such as Good, Moderate, Poor, and Severe.

F. Risk Evaluation

Based on predicted AQI values, the system evaluates respiratory risk levels and generates health recommendations. This helps users understand the potential impact of air pollution and take preventive measures.

G. Route Analysis

The system analyzes pollution levels across different travel routes and identifies paths with lower AQI exposure. This feature helps users minimize exposure to polluted environments during travel.

V. RESULTS AND EVALUATION

A. Dataset Description

The dataset used for this study consists of environmental parameters such as PM2.5, PM10, NO_2, SO_2, CO, temperature, and humidity. The data is collected from external environmental APIs and includes both real-time and historical values.

The dataset is limited and primarily focused on the Gurugram region, which is known for high pollution levels. Due to this regional limitation, the model is trained and evaluated mainly on data specific to Gurugram, which may affect its performance when applied to other locations.

B. Model Performance

The XGBoost model is used for predicting the Air Quality Index (AQI) based on environmental parameters. The model demonstrates good performance in capturing the relationship between pollutant levels and AQI values.

The predictions are consistent with expected air quality patterns, especially under varying pollution conditions. The model effectively handles non-linear relationships in the data, resulting in reliable AQI predictions.

C. AQI Prediction Results

The predicted AQI values are classified into standard categories such as Good, Moderate, Poor, and Severe. Based on these predictions, the system generates corresponding respiratory risk levels and health recommendations.

Table II: Sample AQI Prediction Results

Location	PM2.5	PM10	Temperature	Humidity	Predicted AQI	Category
Gurugram	120	180	32°C	65%	210	Poor
Gurugram	250	320	34°C	60%	350	Severe
Gurugram	60	90	28°C	70%	120	Moderate

D. Risk Assessment and Route Analysis

Based on the predicted AQI values, the system evaluates respiratory risk levels and provides appropriate health recommendations. The system also analyzes pollution levels across different travel routes and identifies paths with lower AQI exposure, helping users minimize their exposure to polluted environments.

E. Limitations

The system is trained on a dataset primarily focused on the Gurugram region. As a result, the model may not generalize effectively to other locations with different environmental conditions. Future improvements can include expanding the dataset to multiple regions for better scalability and accuracy.

F. System Results

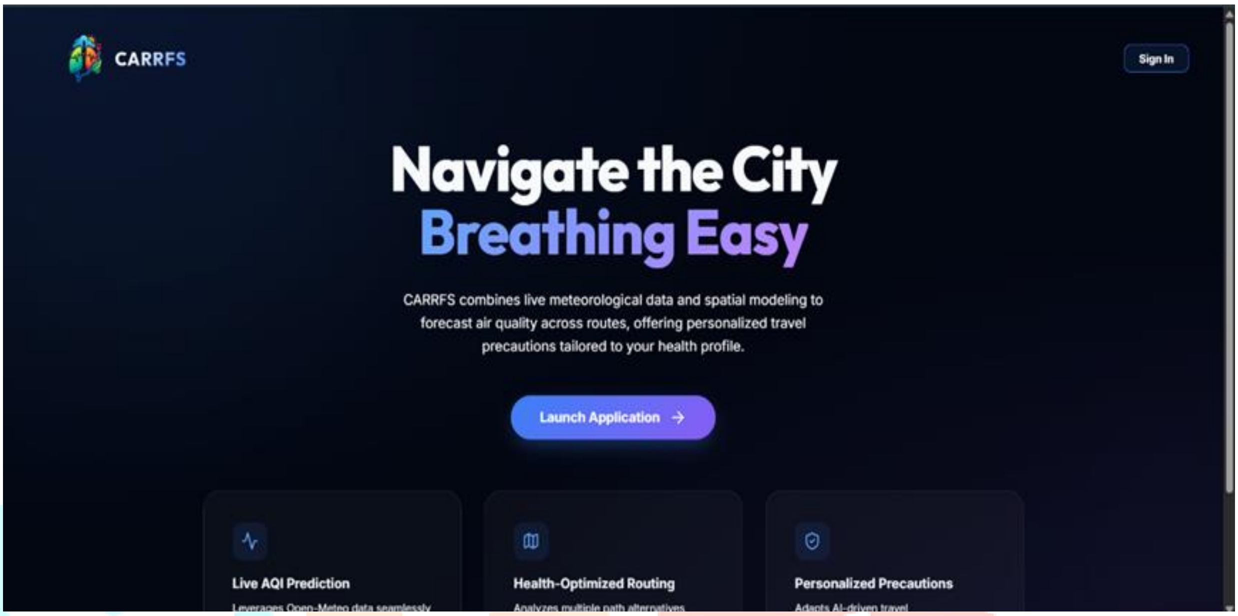


Figure 1.1: Landing Page

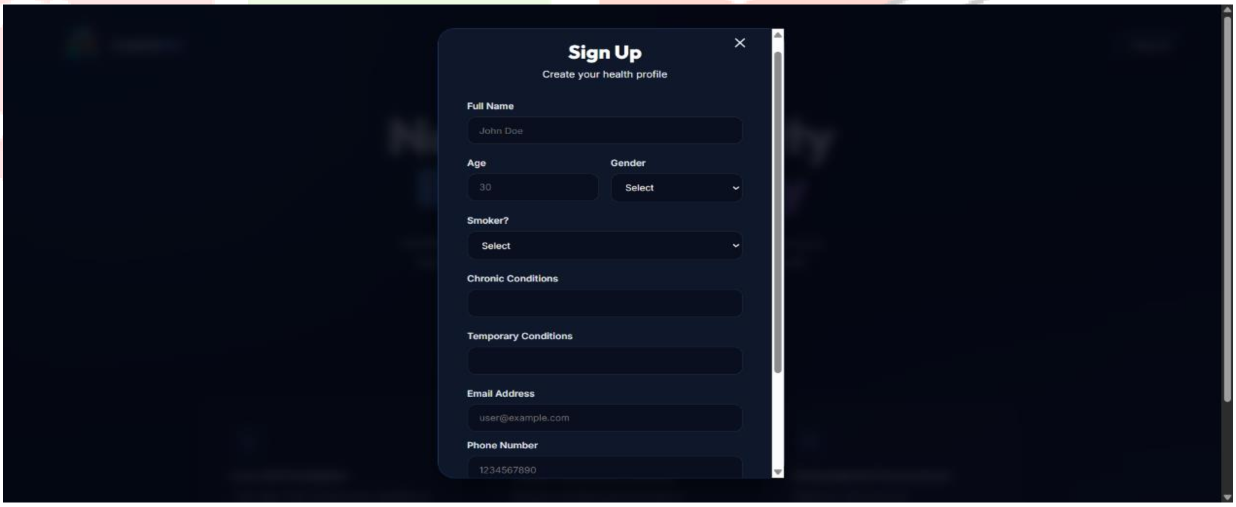


Figure 1.2: Sign Up Page

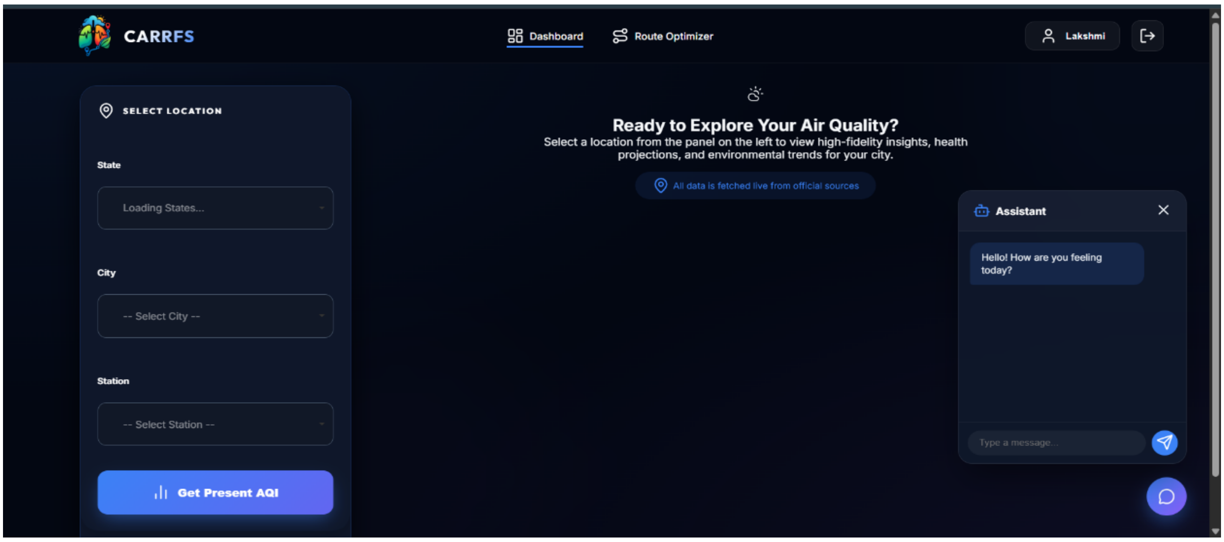


Figure 1.3: Dashboard

Figure 1.4: AQI Prediction

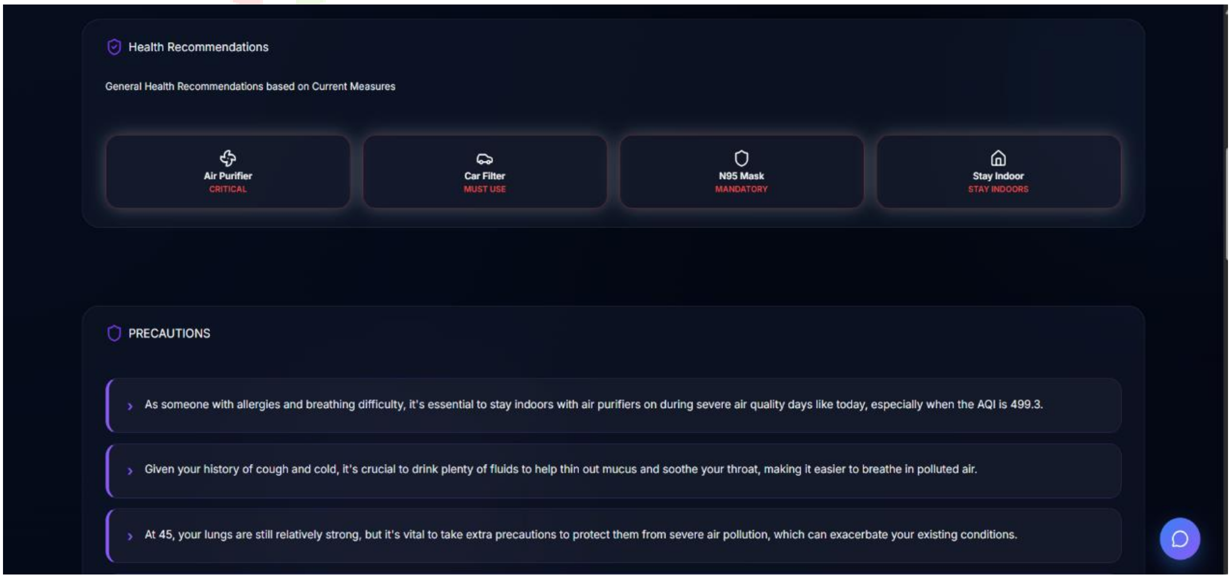


Figure 1.5: Health Recommendations

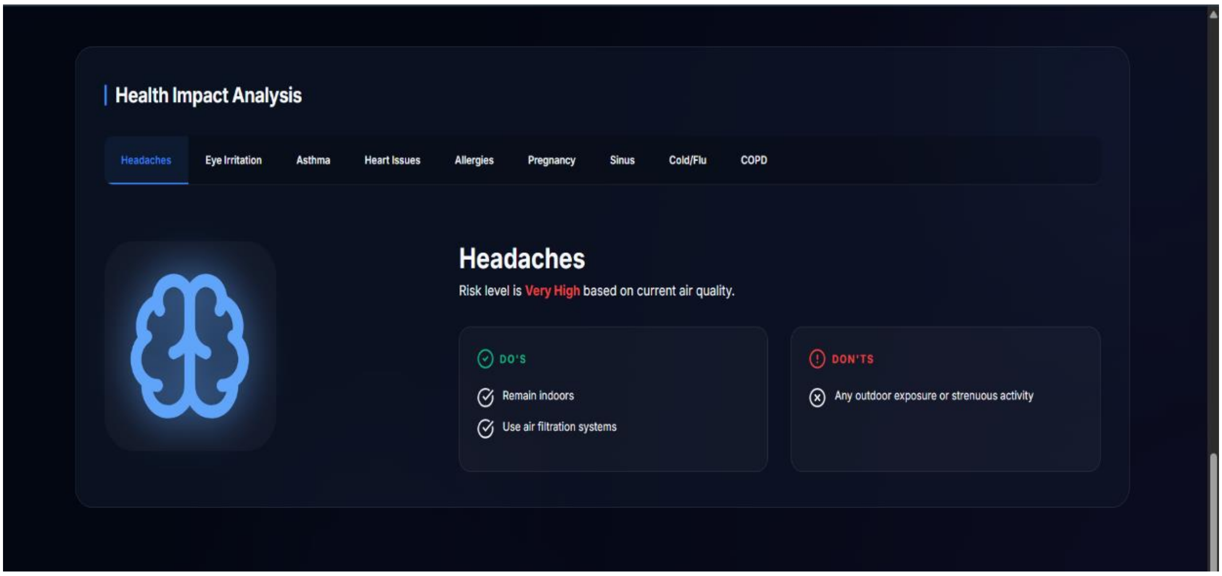


Figure 1.6: Health Impact Analysis

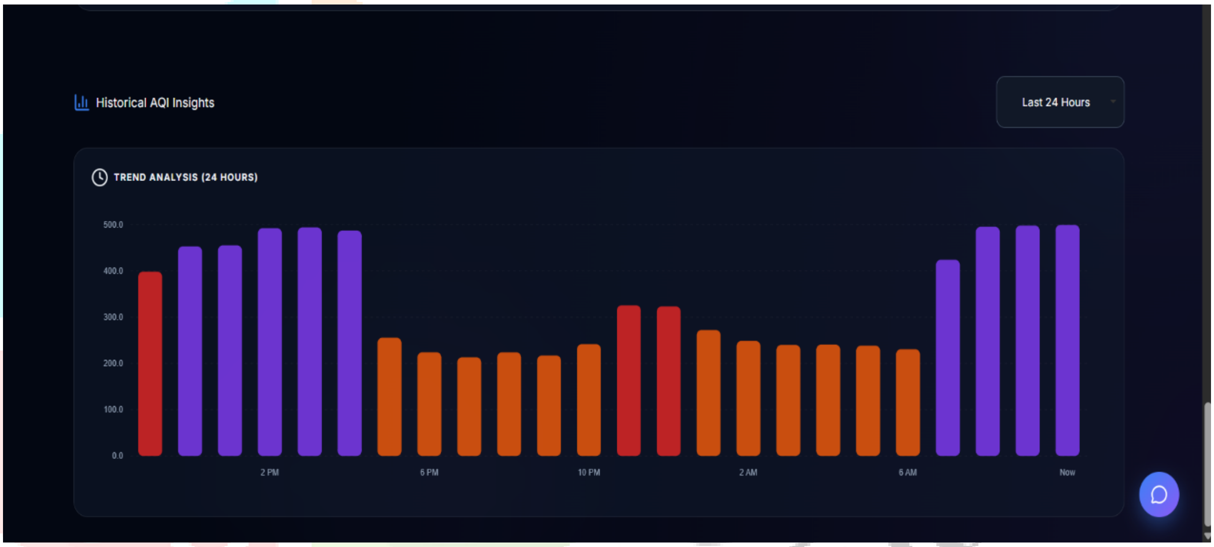


Figure 1.7: AQI Historical Trend Analysis (24hrs)

Fig 1.8: AQI Historical Trend Analysis (7 days)

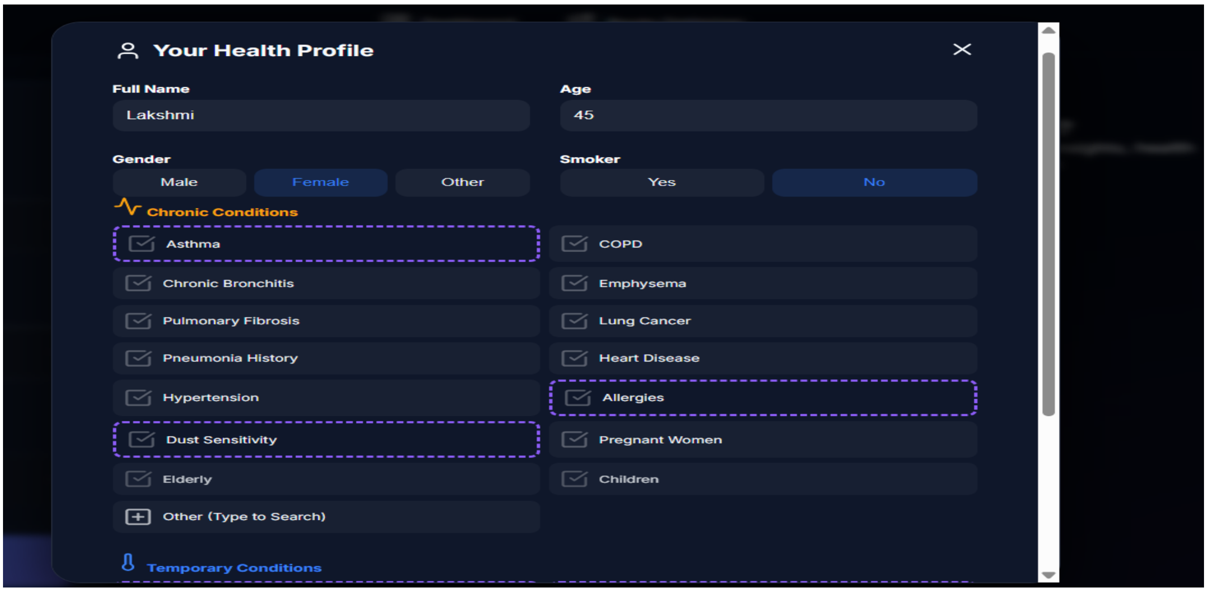


Figure 1.9: User Profile

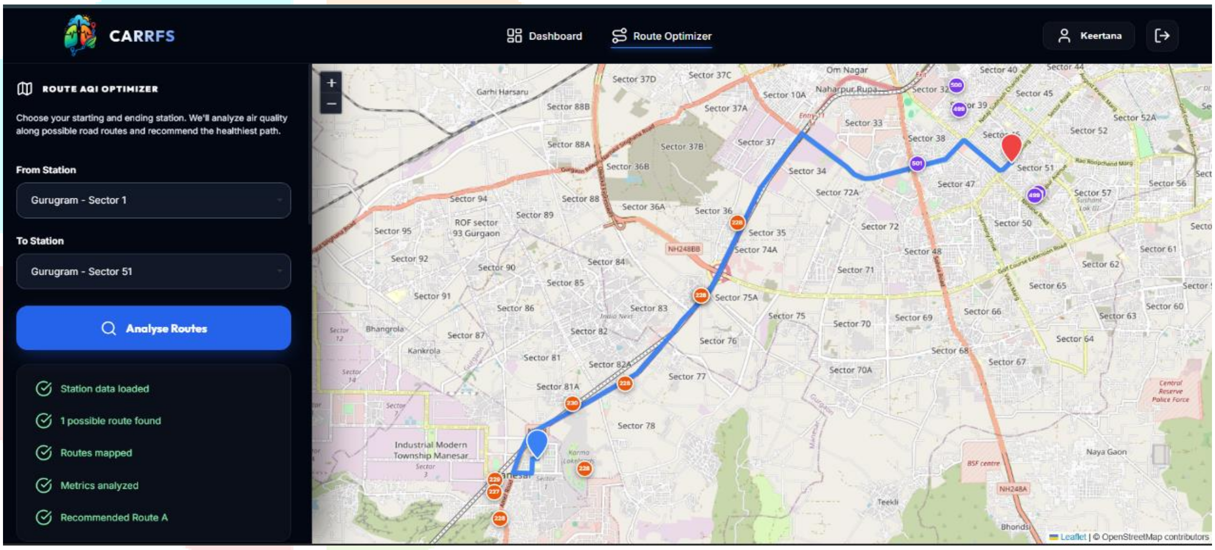
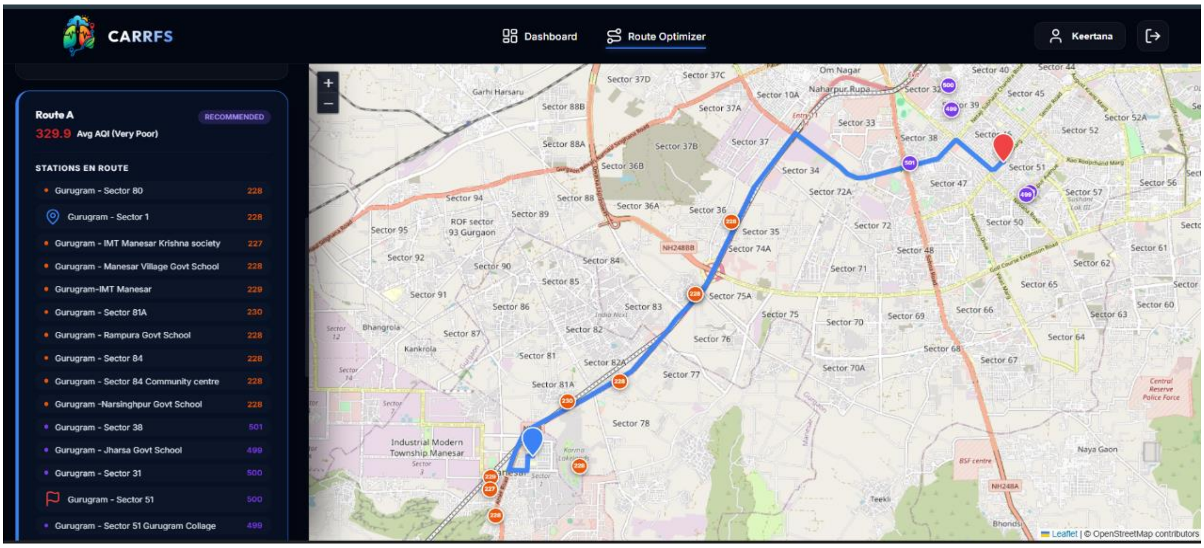


Figure 1.10: Route Analysis

Figure 1.11: Route Optimization



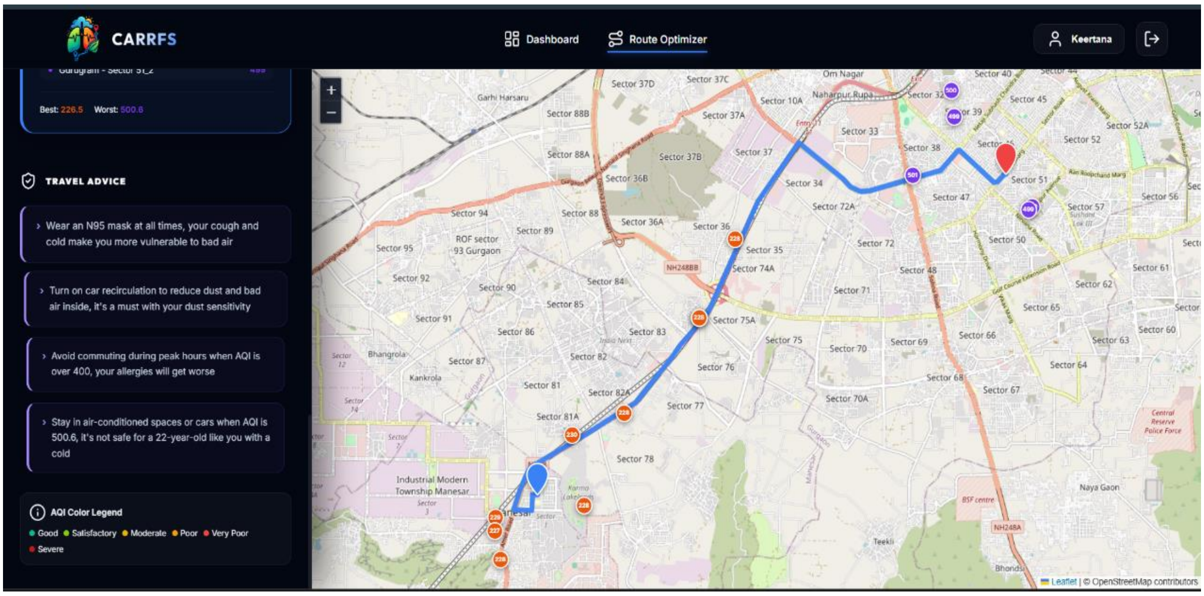


Figure 1.12: Travel Advice

VI. CONCLUSION

The proposed climate-aware respiratory risk forecasting system demonstrates that it is possible to build an intelligent solution that combines environmental data analysis with machine learning to predict air quality and assess respiratory risks. By utilizing the XGBoost model, the system achieves reliable prediction of Air Quality Index (AQI) based on multiple environmental parameters such as PM2.5, PM10, NO_2, SO_2, CO, temperature, and humidity.

The system not only predicts AQI values but also provides meaningful health recommendations and respiratory risk assessments, enabling users to take preventive measures. The inclusion of route optimization further enhances the system by helping users minimize their exposure to polluted environments during travel. The web-based implementation ensures accessibility and ease of use without requiring complex setup or technical expertise.

Overall, the system improves traditional air quality monitoring by incorporating predictive analytics and user-centric features. It serves as a supportive tool for increasing awareness about environmental health risks and assisting individuals in making informed decisions.

Future enhancements can include expanding the dataset to multiple regions beyond Gurugram, integrating real-time IoT sensor data for higher accuracy, improving model performance with advanced techniques, and developing a mobile application for wider accessibility.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Ms. R. SRI DIVYA, Assistant Professor, Department of Computer Science and Engineering (AI & ML), Aditya College of Engineering and Technology (ACET), for her continuous guidance, support, and encouragement throughout this project. The authors also thank the Department of Computer Science and Engineering (AI & ML) and the management of ACET for providing the necessary resources and academic environment to successfully complete this work. This work is submitted in partial fulfillment of the B.Tech degree requirements under Jawaharlal Nehru Technological University Kakinada (JNTUK).

REFERENCES

[1] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794, 2016.

[2] Y. Zhang, M. Bocquet, V. Mallet, C. Seigneur, and A. Baklanov, "Real-Time Air Quality Forecasting, Part I: History, Techniques, and Current Status," Atmospheric Environment, vol. 60, pp. 632–655, 2012.

[3] X. Li, L. Peng, Y. Hu, J. Shao, and T. Chi, "Deep Learning Architecture for Air Quality Predictions," Environmental Science and Pollution Research, vol. 23, no. 22, 2016.

[4] D. Jiang, Y. Zhang, X. Hu, Y. Zeng, J. Tan, and D. Shao, "Progress in Developing an ANN Model for Air Pollution Index Forecasting," Atmospheric Environment, vol. 38, no. 40, pp. 7055–7064, 2004.

[5] S. K. Grange and D. C. Carslaw, "Using Meteorological Normalisation to Detect Interventions in Air Quality Time Series," Science of the Total Environment, vol. 653, pp. 578–588, 2019.

[6] P. Kumar, M. Khare, R. M. Harrison, W. J. Bloss, A. C. Lewis, H. Coe, and L. Morawska, "New Directions: Air Pollution Challenges for Developing Megacities like Delhi," Atmospheric Environment, vol. 122, pp. 657–661, 2015.

[7] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.

[8] Y. LeCun, Y. Bengio, and G. Hinton, "Deep Learning," Nature, vol. 521, pp. 436–444, 2015.