ISSN 2319-5991 www.ijerst.com
© 2026 IJERST. All Rights Reserved
Peer Reviewed, Referred & Indexed Journal

Research Paper

AIR QUALITY PREDICTION USING MACHINE LEARNING ALGORITHMS FOR ENVIRONMENTAL MONITORING AND PUBLIC SAFETY

[1] S Nagaraju, [2] R Sowmya sree, [3] T Siddik Basha, [4] M Zakeer Hussain,[5]S Yasmeen Afroz, [6]K Sai
[1]Assistant Professor, [2,3,4,5,6] Students
Department of Computer Science and Engineering (Artificial Intelligence & Machine Learning),
Gates Institute of Technology, Gooty
Ananthapuramu, India

ABSTRACT- Air quality has become a major concern due to rapid urbanization, industrialization, and increasing vehicular emissions, affecting human health and the environment. This paper presents an air quality prediction system using a Hybrid Air Quality Prediction Algorithm (HAQP) that combines multiple machine learning techniques. The system utilize environmental parameters such as PM2.5, PM10, NO_2, SO_2, CO, O_3, temperature, and humidity. Data preprocessing methods including missing value handling, noise removal, and feature selection are applied to improve performance. The hybrid model integrates algorithms such as Linear Regression, Random Forest, Support Vector Machine (SVM), Decision Tree, and Gradient Boosting. The model is evaluated using MAE, RMSE, and R^2score. Experimental results show that the system provides accurate predictions and supports real-time environmental monitoring and public safety.

Keywords: Air Quality Prediction, Machine Learning, A QI, Environmental Monitoring, Public Safety, Random Forest, SVM, Gradient Boosting, Data Preprocessing.

I. INTRODUCTION

Air pollution is one of the most serious environmental challenges affecting human health and ecosystems worldwide. Rapid industrialization, urbanization, and increasing vehicle emissions have significantly contributed to the degradation of air quality. Poor air quality can cause severe health problems such as respiratory diseases, cardiovascular disorders, and reduced life expectancy. Therefore, accurate monitoring and prediction of air quality are essential for public safety and environmental protection.

Traditional air quality monitoring systems rely on physical sensors and statistical methods, which are often expensive and limited in coverage. Moreover, these systems fail to capture complex relationships among environmental pollutants. Traditional air quality monitoring systems rely on physical sensors and statistical methods, which are often expensive and limited in coverage.

Moreover, these systems fail to capture complex relationships among environmental pollutants.

In this work, a Hybrid Air Quality Prediction Algorithm (HAQP) is proposed, which combines multiple machine learning models to enhance prediction performance. The system processes environmental data, applies preprocessing techniques, and predicts air quality conditions efficiently. The proposed model aims to support real-time monitoring and help authorities take preventive measures.

II PROBLEM STSTEMENT

Air pollution has become a critical issue affecting public health and environmental sustainability. Traditional monitoring systems are limited by high cost, in sufficient coverage, and inability to provide real-time predictions. Additionally, existing machine learning models often struggle with handling missing data and capturing complex relationships among pollutants. Therefore, there is a need for an efficient and accurate air quality prediction system that can analyze environmental data and provide reliable predictions. The proposed system addresses these challenges by using a hybrid machine learning approach to improve prediction accuracy and support environmental monitoring

III. RELATED WORK

Several researchers have explored the use of machine learning and deep learning techniques for air quality prediction using environmental data. These automated systems help in analyzing pollution levels more accurately and support effective environmental monitoring.

Shreya et al. [1] presented a study on air pollution prediction using machine learning techniques and found that algorithms like Random Forest provide high accuracy. The study also high lighted the importance of combining multiple data sources, although challenges iike real-time implementation still exist.

A gbehadji and Obagbuwa [2] reviewed machine learning and deep learning methods for air quality prediction and emphasized the need for hybrid models to handle complex environmental data. They also discussed issues such as missing data and scalability.

Essamlali et al. [3] proposed machine learning models such as Support Vector Machine(SVM),Random Forest, and Gradient Boosting for predicting air pollutants. Their study showed that combining multiple models improves prediction accuracy.

Kumar Gupta [4] analyzed different machine learning techniques and highlighted that ensemble methods like Random Forest provide better performance due to their stability and accuracy.

Özüpak et al. [5] conducted a comparative study of machine learning models and found that optimized models provide higher accuracy and lower prediction errors.

Mottahedin [6] proposed a hybrid model combining ARIMA, CNN, and LSTM, which improved prediction accuracy by capturing both linear and nonlinear patterns in air pollution data.

Rahman [7] introduced a hybrid deep learning model integrating multiple algorithms such as ARIMA, CNN, LSTM, and XGBoost, achieving high accuracy in

air quality prediction.

Although many studies have shown promising results, challenges such as missing data, real-time prediction, and model complexity still exist. Therefore, more efficient and reliable models are required for accurate air quality prediction.

IV. EXISTING SYSTEM

Air quality prediction plays a crucial role in environmental monitoring as air pollution directly impacts human health and climate conditions. Existing systems collect environmental data from various sources such as monitoring stations, satellite data, and IoT sensors. These systemsmeasureparametersincludingPM2.5,PM10,CO, NO_2, SO_2, O_3, temperature, humidity, and wind speed. Traditional prediction methods such as Linear Regression, ARIMA, and time-series models were initially used, but they are limited in handling complex relationships among environmental factors.

With the advancement of machine learning, algorithms such as Random Forest, Support Vector Regression, Decision Trees, Gradient Boosting, and Artificial Neural Networks have been applied to improve prediction accuracy. These models can capture nonlinear relationships and provide better forecasting performance. However, most existing systems rely heavily on historical data, making them less effective in handling sudden environmental changes. Additionally, issues such as missing data, noise, limited scalability, and lack of real-time adaptability still exist. Although machine learning-based approaches have improved prediction accuracy, challenges related to performance, scalability, and interpretability remain

V. PROPOSED SYSTEM

The proposed system presents a machine learning- based approach for air quality prediction by integrating multiple algorithms to improve accuracy and reliability. The system collects environmental data from various sources such as air quality monitoring sensors, meteorological stations, and public datasets. Key parametersincludingPM2.5,PM10,NO_2,SO_2,CO,O_3, temperature, humidity, and wind speed are used as inputs. Data preprocessing techniques such as cleaning, normalization, and feature selection are applied to enhance data quality. Machine learning algorithms such as Random Forest, Support Vector Machine (SVM), Gradient Boosting, and Neural Networks are used for model training and prediction. The system also supports real-time data processing, visualization, and early warning alerts to ensure effective environmental monitoring and public safety

System Work flow

Data Input: Environmental data is collected from sensors, weather stations, and datasets, including pollutant and meteorological parameters.

Data Preprocessing: The data is cleaned, missing values are handled, noise is removed, and normalization is applied.

Feature Extraction: Important features are identified to understand patterns in air pollution levels.

Model Training: Machine learning algorithms such as Random Forest, SVM, Gradient Boosting, and Neural Networks are used to train the model.

Prediction: The trained model predicts the Air Quality Index (AQI) based on input data.

Visualization: Results are displayed using graphs and charts for easy understanding of pollution trends.

Early Warning System: Alerts are generated when AQI exceeds safe limits.
Output Result: The system shows AQI values along with categories like Good, Moderate, and Poor.
Continuous Improvement: The model is updated using new data and evaluated using MAE, RMSE, and R^2 score.

VI. MODELS USED

The proposed air quality prediction system utilizes multiple machine learning algorithms to analyze environmental data and improve prediction accuracy. These models help in identifying complex patterns in air pollution and provide reliable predictions. The system integrates several algorithms to form a hybrid approach for better performance.

Random Forest

Random Forest is an ensemble algorithm that combines multiple decision trees to improve accuracy and reduce over fitting. In this system, it analyzes environmental parameters and captures nonlinear relationships for reliable air quality prediction.

Support Vector Machine(SVM)

SVM is a supervised learning algorithm used for classification and regression. It models complex relationships between environmental factors and air quality, providing effective performance for high-dimensional data.

Decision Tree

Decision Tree is an interpretable algorithm that splits data based on feature values. It helps identify important environmental factors and provides simple and understandable prediction rules.

Gradient Boosting

Gradient Boosting is an ensemble technique that builds models sequentially to reduce errors. It improves prediction accuracy and enhances overall performance in air quality forecasting.

Linear Regression

Linear Regression models the relationship between input variables and output. It is used as a base line model to understand basic trends and compare performance with advanced algorithms.

In the proposed system, Linear Regression is used as a baseline model to compare performance with other advanced algorithms and to understand basic trends in air quality data.

VI. RESULTS AND DISCUSSION:

The proposed air quality prediction system was implemented using machine learning algorithms and tested on environmental datasets. The data was preprocessed by handling missing values, removing noise, and applying normalization. Relevant features were selected to improve prediction performance.

Multiple models such as Random Forest, Support Vector Machine (SVM), Decision Tree, Gradient Boosting, and Linear Regression were trained and evaluated using MAE, RMSE, and R^2 score. Among these, ensemble models like Random Forest and Gradient Boosting showed better accuracy

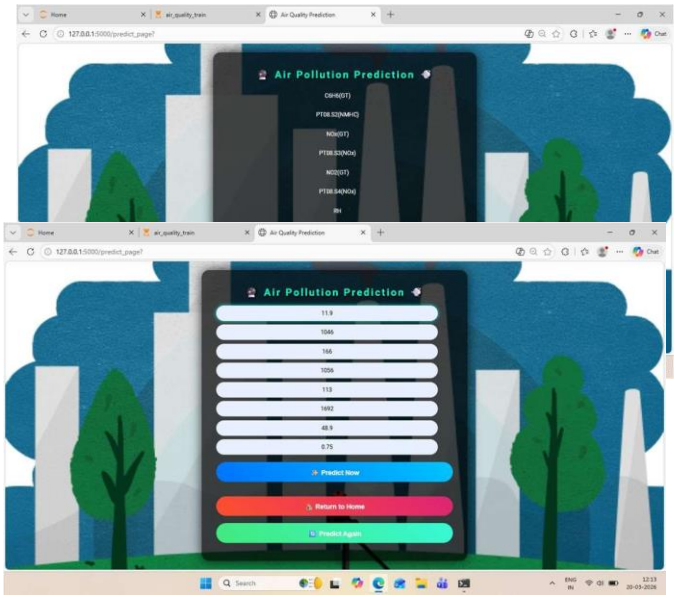
The system successfully predicts air quality levels and classifies them into categories such as Good, Moderate, and Poor. A user-friendly interface is developed to input data and display results. Graphs and charts are used to

visualize pollution trends and prediction outputs.

Home Page

The above screen shows the home page of the Air Quality Prediction System. It acts as the main interface of the application, where users can navigate to different functionalities. The page provides options such as “Predict Air Quality” and “About Air Quality.” This interface is developed using a web framework and allows users to easily interact with the system

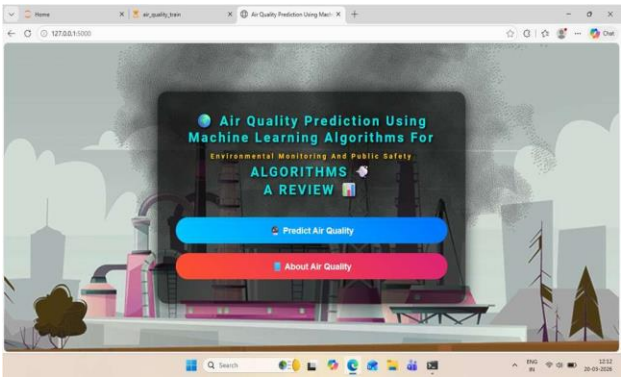
Prediction Page



The above screen shows the prediction page of the Air Quality Prediction System. In this page, the user is required to enter various environmental parameters such as C6H6(GT), NOx(GT), NO2(GT), humidity, and other related values. These inputs are used by the machine learning model to predict air quality. After entering the required data, the user can click on the “Predict Now” button to generate the prediction result.

Input Data Entry

The above screen shows the user entering input values for different environmental



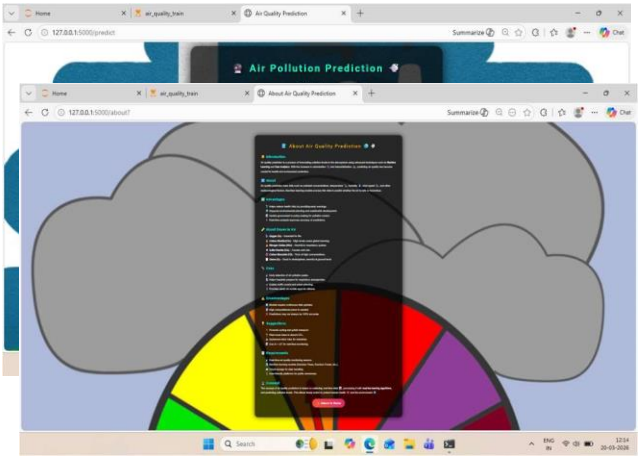
parameters in the Air Quality Prediction System. The values include pollutant concentrations and atmospheric conditions such as C6H6(GT), NOx(GT), NO2(GT), humidity, and others. These inputs are provided by the user and are used by the machine learning model to predict air quality.

Prediction Output

The above screen shows the prediction output generated by the Air Quality Prediction System. After entering the required environmental parameters and clicking on the “Predict Now” button, the system processes the input data using the trained machine learning model

Based on the analysis, the system displays the result indicating whether the air quality is affected by pollution or not. In this case, the output shows that the air is “Not affected by air pollution.” This result is presented in a clear and user-friendly interface.

About



The above screen shows the about page of the Air Quality Prediction System. This page provides detailed information about air quality, including its importance, factors affecting it, advantages of prediction, and its impact on environmental monitoring and public safety. It helps users understand the purpose and working of the system in a clear and informative manner

VII CONCLUSION AND FUTURE ENHANCEMENT:

The proposed system presents an efficient machine learning-based approach for air quality prediction using a hybrid model. By integrating multiple algorithms such as Random Forest, SVM, Decision Tree, Gradient Boosting, and Linear Regression, the system improves prediction accuracy and reliability. The model effectively analyzes environmental parameters and predicts air quality levels.

The system is implemented with a user-friendly interface that allows real-time prediction and visualization of results. It helps in understanding pollution patterns and supports decision-making for environmental monitoring and public safety. Overall, the proposed approach demonstrates that combining multiple machine learning techniques enhances performance and provides are liable solution for air quality prediction.

In future, the proposed system can be enhanced by integrating real-time data from IoT-based air quality sensors to improve prediction accuracy. Advanced deep learning models such as LSTM can be used for better time-series forecasting of air quality. The system can also be extended with Geographic Information System (GIS) integration for location-based visualization of pollution levels.

Additionally, the model can be deployed as a mobile or web application for wider accessibility. Using larger and more diverse datasets can further improve model performance and generalization. These improvements will make the system more scalable, accurate, and suitable for real-world environmental monitoring applications.

References

1.C. Zhang,X. Pan, H. Li, A. Gardiner, I. Sargent, and J.S. Hare, “A hybrid MLP-CNN classifier for very high resolution remotely sensed image classification, ”ISPRS Journal of Photo grammetry and Remote Sensing, vol. 140, pp.133–144,2017.

2.L. Bi, J. Kim, E. Ahn, and D. Feng, “Automatic skin lesion analysis using large-scale dermo scopy images and deepresidualnetworks,”arXivpreprintarXiv :1703.04197, 2017.

3.N. Hameed, A. M. Shabut, and M. A.

Hossain, "Multi- class skin diseases classification using deep convolutional neural network and support vector machine," in Proc.12th Int. Conf. Software, Knowledge, Information Management &Applications (SKIMA), 2018, pp. 1–7.
4.T.H.H.Aldhyani, A.Verma,M.H.Al-Adhaileh, and D. Koundal, "Multi-class skin lesion classification using a lightweight dynamic kernel deep-learning-based convolutional neural network, "Diagnostics, vol. 12,no. 9,p.2048,2022.
5.N. Vakili, W. Krathu, and N. Laoma neat tanaporn, "Multi-class primary morphology lesions classification using deep convolutional neural network,"inProc.12thInt. Conf. Advances in Information Technology (IAIT), 2021, pp. 1–7.
6.M. Iqbal, K. Younus, M. U. Kakar, J. Ma, "Automated multi-class classification of skin lesions through deep convolutional neural network with dermo scopic images,"Computerized Medical Imaging and Graphics, vol. 88, p. 101843, 2021.
7.S.S.Chaturvedi, J.V.Tembhurne, and T.Diwan,"A multi-class skin cancer classification using deep convolutional neural networks,"Multimedia Tools and Applications, vol. 79, pp. 28477–28498, 2020.
8.M.A.Anjum, J.Amin, M.Sharif, H.U.Khan, M.S.A. Malik, and S. Kadry, "Deep semantic segmentation and multi-class skin lesion classification based on convolutionalneuralnetwork,"IEEEAccess ,vol.8,pp. 129668–129678, 2020.
9.V. Anand, S. Gupta, D. Koundal, S. R. Nayak, J. Nayak, and S. Vimal, "Multi-class skin disease classification using transfer learning model," International Journal on Artificial Intelligence Tools, vol. 31, no. 2, p. 2250029, 2022.
10.P.N.Srinivasu, J.G.SivaSai, M.F.Ijaz, A.K. Bhoi, W. Kim, and J. J. Kang, "Classification of skin disease using deep learning neural networks with Mobile NetV2 and LSTM," Sensors, vol. 21, no. 8, p. 2852, 2021.
11.T. Shanthi, R. S. Sabeenian, and R. Anand, "Automatic diagnosis of skin diseases using convolution neural network, "Microprocessors and Microsystems, vol. 76, p. 103074, 2020.
12.V. R. Allugunti, "A machine learning model for skin disease classification using convolution neural network," International Journal of Computing, Programming and DatabaseManagement,vol.3,no.1,pp.141–147,2022.
[1]