International Journal of Scientific Research in Computer Science, Engineering and Information Technology

ISSN : 2456-3307 OPEN ACCESS

Available Online at : www.ijsrcseit.com

doi : https://doi.org/10.32628/IJSRCSEIT

# A Smart Air Pollution Detector Using Machine Learning

**K. Madhusudhan Reddy[1], Kanala Prakash[2]**

[1]Assistant Professor, Department of MCA, Annamacharya Institute of Technology and Sciences, Karakambadi, Tirupati, Andhra Pradesh, India

[2]Student, Department of MCA, Annamacharya Institute of Technology and Sciences, Karakambadi, Tirupati, Andhra Pradesh, India

ARTICLEINFO

**Article History:**

Accepted : 01 May 2025

Published: 05 May 2025

**Publication Issue**

Volume 11, Issue 3

May-June-2025

**Page Number**

119-129

ABSTRACT

The rapid growth of urbanization and industrial activities in cities has resulted in a significant deterioration in air quality, which poses an increasing threat to both public health and the environment. This study focuses on predicting air quality using ML algorithms, aiming to classify air quality into three distinct categories: Good, Satisfactory, and Poor. The dataset utilized for this research comprises key environmental factors such as PM2.5, PM10, nitrogen oxides, and carbon monoxide, which are considered critical indicators of air pollution. To enhance the accuracy of predictions, several ML models were employed, including Logistic Regression, MLP, Random Forest, Decision Tree, The data preprocessing phase involved several essential steps to prepare the dataset for model training. These steps included the handling of missing values, selection of relevant features, and addressing class imbalance through the use of the SMOTE, which was employed to balance the distribution of target labels. The models were then trained and evaluated based on their performance in predicting air quality categories, with accuracy being the primary evaluation metric. Moreover, it can help inform public health decisions by identifying regions with poor air quality and ensuring better management of air pollution levels.

**Keywords—**Air Quality Prediction, Machine Learning (ML), Classification, Synthetic Minority Over-sampling Technique (SMOTE), Air Quality Index (AQI).

## Introduction

Rapid urbanization and industrialization have exacerbated the decline in air quality, particularly in metropolitan areas, leading to a rise in a variety of health issues such as chronic respiratory diseases, heart-related conditions, and even premature deaths. With air pollution levels reaching hazardous thresholds, the need for effective monitoring and management of air quality has become more critical than ever.The AQI quantifies the severity of pollution, helping to categorize air quality into levels such as Good, Satisfactory, and Poor, enabling the timely implementation of health and environmental interventions.In urban environments, where pollution levels tend to be higher, real-time and accurate monitoring of air quality is essential to inform decision-making and prevent adverse health impacts. With the ability to predict and classify air quality, cities can proactively address pollution sources, improve urban planning, and safeguard the well-being of residents.This project seeks to develop a ML-based model to predict air quality levels using historical data. By examining pollutant concentrations, the model will classify air quality into three distinct categories: Good, Satisfactory, and Poor. The study explores a range of machine learning algorithms, includes logistic regression, MLP, random forest and decision tree, to identify the most effective method for air quality classification.The outcomes of this research will provide valuable insights for policy-makers, environmental experts, and the public, helping them to take preventive actions in response to poor air quality conditions. Additionally, the use of machine learning allows for a data- driven approach that predicts air quality.

**Objective**

The goal is to classify air quality into three distinct categories—Good, Satisfactory, and Poor. These pollutants play a significant role in determining the overall air quality in urban and industrialized regions, and accurately predicting their concentrations is essential for safeguarding public health and the environment.To achieve this objective, several machine learning algorithms will be explored and implemented. These algorithms will be evaluated based on their performance in accurately predicting air quality categories, and the best-performing model will be selected for further refinement.The project also emphasizes addressing challenges related to class imbalance within the dataset. Air quality categories often exhibit unequal distribution, with certain classes underrepresented. To mitigate this issueIn the present study gave, among other techniques, SMOTE, the synthetic minority over-sampling technique such that synthetic data samples will be generated for minority classes in order to facilitate learning of the model. predict all categories effectively, without bias toward the majority class.Moreover, the project will focus on optimizing the models to improve their accuracy and robustness. Hyperparameter tuning will be performed to fine-tune the settings of each algorithm, and feature selection will be carried out to identify and retain the most influential variables for predicting air quality. By optimizing both the model's structure and the features it relies upon, the system's performance will be enhanced, ensuring more precise and reliable predictions.

**Motivation**

The accelerating pace of industrial growth, urbanization, and increased vehicular emissions has resulted in a marked decline in global air quality. it has become a significant public health challenge. It has been linked to a rise in various health issues, particularly respiratory diseases, cardiovascular disorders, and premature deaths, making it one of the most pressing environmental health risks worldwide. The World Health Organization estimates that exposure to ambient air pollution causes millions of premature deaths annually, underscoring the urgency of addressing this issue.The impact of deteriorating air quality extends far beyond immediate health effects; it also places a substantial strain on healthcare systems, reduces life expectancy, and affects overall quality of life. Timely and accurate air quality monitoring is essential for identifying pollution hotspots, issuing health warnings, and implementing effective policies to protect communities and the environment.Historically, air quality has been monitored using manual sampling techniques and fixed monitoring stations. While these methods have been useful, they are often expensive, labor-intensive, and geographically constrained. Consequently, the need for more efficient, real-time solutions to assess air quality has become increasingly apparent. With the rapid development of ML technologies, we now have the opportunity to predict air quality more accurately and in real-time. By using historical environmental data, ML models can effectively classify and forecast air quality levels.With accurate and timely air quality forecasts, public health authorities can take swift action, such as issuing health advisories, enforcing pollution control measures, and educating the public about the risks of poor air quality. This project aims to make a meaningful contribution to environmental monitoring by developing an ML-based solution for air quality prediction. Beyond just environmental monitoring, the impact of this research extends to urban planning and public health policy, where data-driven insights can lead to better decision-making.

**scope**

This project is primarily focused on predicting and categorizing air quality levels based on various key environmental pollutants. The scope of this study is constrained to utilizing historical air quality data to develop and implement ML models that classify air quality into three primary levels: Good, Satisfactory, and Poor.The project also encompasses addressing key data preprocessing challenges that often arise when working with environmental datasets. These challenges include handling missing data, overcoming class imbalance, and selecting the most relevant features for training the models. To mitigate Balance classes, we will resort to using techniques like SMOTE so that the classifier operates fairly well for all categories of the air quality index. It is important to note that while the project focuses on building predictive models, it does not aim to develop a real-time air quality monitoring system. Instead, the primary goal is to train and evaluate models using historical air quality data. However, the insights gained from this study could lay the groundwork for future advancements in real-time air quality forecasting, should further research and data integration efforts be pursued.The findings of this project can have far-reaching applications across several domains. In the context of public health, the model could provide valuable information for anticipating health risks associated with poor air quality. In environmental monitoring, the predictions could be used to track pollution trends and inform pollution control strategies. Moreover, urban planning could benefit from the results, as city planners and policy-makers could use the insights to design better air quality management strategies and take preventive actions to reduce the adverse effects of pollution on communities. Ultimately, this project aims to provide a data-driven foundation that could contribute to improving air quality management practices and public health outcomes in urban areas.

**LITERATURE SURVEY**

With the rapid advancements in ML technologies, there is an increasing potential for leveraging these techniques to provide accurate and real-time air quality predictions. The following sections review various aspects of air quality prediction and monitoring[1], focusing on traditional methods, ML approaches, data preprocessing techniques, evaluation metrics, and challenges in the field.[2]

1. Air Quality Monitoring Techniques

These techniques have been fundamental in providing accurate measurements of various air contaminants. [3]However, they come with significant limitations, including high costs, long processing times, and restricted spatial coverage, making it difficult to monitor large urban or industrial regions comprehensively. [4]To overcome these challenges, researchers have increasingly turned to machine learning models, which can process large volumes of data from sensor networks, satellite imagery, and other remote sensing tools to offer real-time monitoring. [5]By utilizing these advanced technologies, air quality assessment can be scaled across large geographical regions, enabling more dynamic and efficient monitoring systems. [6]Furthermore, these ML-based systems provide the flexibility to integrate data from multiple sources, improving the precision and accuracy of air quality predictions.

2. ML Models for Air Quality Prediction

[7]Machine learning models have shown promise in predicting air quality levels, with various techniques applied across different studies. [8]some models have been used to predict the concentration of pollutants like PM2.5. These models have been valuable for their simplicity and interpretability but are often limited in handling more complex relationships within the data. [9]On the other hand, classification algorithms have gained popularity for categorizing air quality into specific levels, such as good, satisfactory, or poor. [10]Additionally, deep learning models, have been explored to tackle more intricate and non-linear patterns in air quality data, providing deeper insights into air pollution and its potential impact on health.

3. Data Preprocessing Techniques

Data preprocessing plays a crucial role in ensuring the effectiveness of machine learning models, especially when working with environmental data. [11]The raw data collected from monitoring stations often contains missing values, outliers, and imbalanced distributions, all of which can adversely affect model performance. More advanced strategies, such as interpolation and forward or backward filling, are also employed to preserve data integrity. The challenge of data imbalance, particularly when certain air quality categories are underrepresented, is another issue that requires attention. Techniques like SMOTE are commonly used to generate synthetic samples for underrepresented classes, ensuring the model learns to classify all categories effectively and accurately.

4. Evaluation Metrics and Model Comparison

Evaluating the performance of ML models is critical to determining their effectiveness in air quality prediction tasks. A range of evaluation metrics is commonly used which provide insight into how well the models classify different air quality levels. Research comparing the performance of various algorithms has consistently shown that decision trees and Random Forest models tend to outperform simpler models. Furthermore, ensemble methods, which combine multiple models to enhance performance, have also demonstrated high predictive power in air quality classification tasks. These ensemble methods are particularly beneficial in reducing variance and bias, leading to more reliable and stable predictions. Researchers have emphasized the importance of evaluating models under different environmental conditions, considering factors like regional pollution patterns and seasonal variations that can significantly affect air quality levels.

5. Challenges and Future Directions

While machine learning models have shown considerable success in predicting air quality, several challenges remain that hinder their full potential. Additionally, obtaining labeled datasets, particularly for real-time predictions, is a limitation that many researchers face. This scarcity of labeled data makes it difficult to train models effectively, particularly when aiming to predict future air quality levels. Future research is expected to focus on improving model generalization, enabling algorithms to better adapt to new, unseen data. Furthermore, there is significant potential for the development of real-time prediction systems using streaming data from IoT sensors and environmental monitoring stations, which could provide continuous, up-to-date air quality assessments.

## PROPOSED SYSTEM

The proposed system aims to predict air quality by analyzing various environmental factors using ML algorithms. It is designed to allow users to interact with the application, input relevant air quality indicators, and receive accurate predictions regarding the air quality in their respective locations. The system consists of two primary components: the User Module and the System Module.

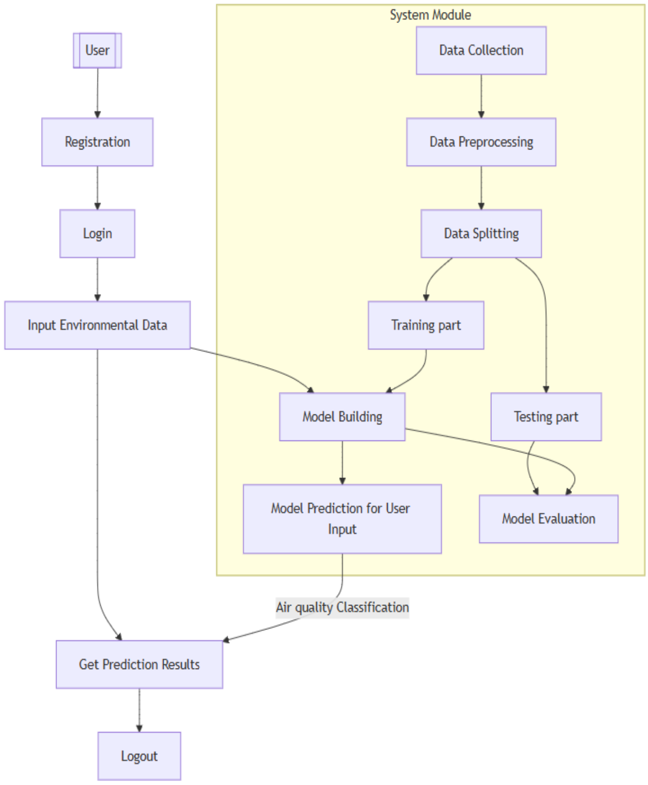


**Fig:** Block diagram for proposed system

User Module

The User Module is responsible for providing the necessary features for users to engage with the system. Upon accessing the system, users can first register. Next, users can log in to the system. After logging in, users can enter environmental data related to pollutant levels.

This input data will then be processed by the system to generate air quality predictions. The system will classify the air quality into one of three categories: Good, Satisfactory, or Poor, and provide the corresponding Air Quality Index value. The AQI value, which quantifies the level of pollution in the air, will serve as an additional indicator to inform users of the air quality conditions in their region. Once users have received the predictions, they can log out of the system to ensure the security of their session and personal information.

System Module

The System Module handles the backend processing necessary for generating accurate air quality predictions. The system will begin by gathering historical air quality data from various sources, including publicly available datasets such as those from Kaggle or government databases. The collected data will include a range of pollutants, along with AQI values that serve as indicators of air quality.

The data will undergo several preprocessing steps to prepare it for use in machine learning models. Additionally, the data will be normalized or scaled to ensure consistency across features, allowing the model to learn more effectively. The air quality categories will also be encoded into numerical labels, facilitating the classification process for the machine learning algorithms.

Once the preprocessing steps are completed, the dataset will be divided into train and test sets, with a common 70-30 split ratio. The training set will be used to train the models, while the testing set will be used to evaluate their performance. The system will employ various machine learning algorithms, such as Logistic Regression, Random Forest, XGBoost, and MLP Classifier, to build classification models for air quality prediction. These algorithms will be trained on the preprocessed data to classify air quality levels based on the input pollutant concentrations.

The models will be evaluated using standard performance metrics. Once these process are done, the best-performing model will be selected for deployment. This model will be used to predict air quality for user-provided data, categorizing it into one of the three air quality levels and displaying the corresponding AQI value and classification result.

Integrated Approach

By integrating both the User Module and the System Module, the proposed system will ensure that users receive accurate and reliable air quality predictions. This approach provides a comprehensive and user-friendly solution for air quality monitoring, offering valuable insights into the environmental conditions of different regions. By predicting air quality levels, the system will assist public health authorities, environmental agencies.

## METHODOLOGY

The methodology of this project is structured to predict and classify air quality based on various environmental factors. The process involves several key stages: data collection, preprocessing, model development, evaluation, and optimization. Below is a step-by-step description of the methodology applied in this study:

1) Data Collection

The Dataset contains historical air quality information, including measurements of different pollutants such as PM10, PM2.5, CO, NO2, NO, SO2, O3, Benzene, Toluene, and Xylene. Additionally, the dataset includes the Air Quality Index and corresponding air quality categories, labeled as AQI_Bucket. The data is collected from multiple cities and is available on a daily basis, providing an extensive overview of air quality levels over time.

2) Data Pre-processing

- Handling Missing Values: The dataset had missing values in some of the pollutant columns. These missing values were imputed by replacing them with the mean value of each respective column
- Feature Selection: Non-essential and redundant columns, such as 'City' and 'Date,' were removed from the dataset to minimize noise and reduce computational overhead. The focus was shifted to the relevant features related to air quality prediction, such as pollutant levels and AQI.
- Class Label Encoding: The air quality categories (e.g., Poor, Satisfactory, Good) were encoded into numerical values. The 'Moderate' and 'Satisfactory' categories were combined into one class (1), while 'Poor' and 'Very Poor' were merged into another class (0). The final classes were represented as 0 (Poor), 1 (Satisfactory), and 2 (Good).
- Handling Class Imbalance: The dataset exhibited class imbalance, with some air quality categories being underrepresented. To mitigate this, the SMOTE was applied to generate synthetic samples for the less frequent classes, ensuring a balanced distribution of target labels.

3) Model Development

Various ML algorithms were applied to predict air quality based on the pre-processed dataset. Thos ML algorithms are Logistic Regression, MLP, Random Forest, Decision Tree, AdaBoost, XGBoost,

4) Model Evaluation

- Accuracy: Accuracy is a fundamental metric that measures the overall effectiveness of a model by calculating the proportion of correct predictions out of all predictions made.
- Precision, Recall, and F1-Score: These metrics provide deeper insights into the model's performance, especially when dealing with class imbalance. Precision focuses on the accuracy of the positive predictions, assessing how many of the predicted positive instances were actually correct. Recall, on the other hand, measures the ability of the model to identify all relevant positive instances. The F1-Score is the harmonic mean of precision and recall, offering a balanced evaluation of the model's performance, particularly in scenarios where there is an uneven distribution of class labels.
- Confusion Matrix: The confusion matrix is a tool used to visually represent whether it is performing or not for the model classification.

5) Model Optimization

To improve model performance, hyperparameter tuning was carried out, particularly for more complex models like Random Forest, This involved experimenting with various hyperparameters to identify the optimal settings for each algorithm.

6) Results Visualization

The results of the models were visualized using bar charts to compare the accuracy of each model. Additionally, confusion matrices were used to illustrate which air quality categories were most accurately predicted. These visualizations were instrumental in interpreting the performance of each algorithm and selecting the most effective model for predicting air quality.

## ALGORITHM IMPLEMENDATION

This section outlines the implementation of various ML algorithms employed to predict and classify air quality categories, such as **Good**, **Satisfactory**, and **Poor**. Below is a detailed description of each algorithm used in this study:

### 1. Random Forest Classifier

I created the **Random Forest** model using the RandomForestClassifier() class from the sklearn.ensemble module. To train the model, I used the resampled training data (x_train and y_train) and applied the fit() method. This allowed the model to build multiple decision trees, which were then combined to form a forest, improving the accuracy of the predictions. Once the model was trained, I made predictions on both the training and test datasets using the predict() method. To evaluate how well the model performed, I compared the predicted labels (y_pred_rf) with the actual labels using the accuracy_score() function, which gave me a clear measure of the model's classification accuracy.

### 2. Decision Tree Classifier

I used the **Decision Tree** model, which I initialized using the DecisionTreeClassifier()classes sklearn.tree module. During, I applied fit() method to the resampled dataset (x_train and y_train). The model then split the data into different categories based on the feature values. Once the model was trained, I used the predict() method to make predictions on both the datasets. Measure model's performing or not, I compared the predicted labels with the actual labels (y_train and y_test) using the accuracy_score() function, which helped assess how well the model classified the air quality categories.

### 3. Cat Boost Classifier

I used the **Cat Boost** model, which I initialized using the CatBoostClassifier()class sklearn.tree modules. During, I applied fit() method to the resampled dataset (x_train and y_train). The model then split the data into different categories based on the feature values. Once the model was trained, I used the predict() method to make predictions on both the datasets. Will measure model's performing or not, I compared the predicted labels with the actual labels (y_train and y_test) using the accuracy_score() function, which helped assess how well the model classified the air quality categories

## RESULTS AND DISCUSSION

| Model | Accuracy |
|---|---|
| **Random forest** | 96.85 |
| **Decision Tree** | 96.85 |
| **Cat boost** | 99.79 |

**Fig:** Model accuracy comparison table

Performance Comparison

- Best Performing Models: Random Forest and demonstrated the highest performance, achieving accuracies of 96.47% and 96.42%, respectively. These ensemble methods are highly effective for classification tasks that involve complex datasets, as they combine the predictions of multiple decision trees. By aggregating the results from individual trees, both models enhance the generalizability and robustness of the predictions. Their ability to manage large amounts of data with numerous features contributed significantly to their success in this task.
- Moderate Performance: The MLP Classifier exhibited a balanced performance, achieving relatively high accuracy while also capturing the non-linear relationships inherent in the data. Although it did not surpass the ensemble methods in terms of accuracy, However, the

results suggest that neural networks, despite their power, require additional tuning and adjustments to handle the specific complexities of air quality prediction tasks effectively.

- Lower Performance: Logistic Regression performed the weakest among all models, with an accuracy of 85.70%. The simplicity of this model may explain its inability to accurately capture the non-linear relationships present in air quality data. While Logistic Regression is suitable for problems with fewer complexities, it struggles with datasets that contain intricate patterns or interactions, such as those typically seen in environmental data like air quality.

## Discussion

According to the result Random Forest are highly suited for predicting air quality, as they produced the highest accuracy rates. The capability of ensemble methods to aggregate the results from multiple decision trees allows them to capture more complex patterns within the data and handle diverse feature sets effectively. This is in contrast to simpler models, like Logistic Regression, which tend to perform well only with linearly separable data and are less capable of managing the complexities inherent in environmental datasets.

Although the ensemble models performed exceptionally well, challenges related to class imbalance remained despite the application of SMOTE. SMOTE was used to balance the dataset by generating synthetic samples for underrepresented classes. While this technique helped in improving model performance, the issue of class imbalance, particularly in predicting less frequent air quality levels, may still affect the models in real-world applications. This is especially the case in environments where pollution levels fluctuate significantly across different seasons or geographic regions. Despite these challenges, the models demonstrated good generalization capabilities.

Another observation was the potential overfitting of models like Decision Trees, which achieved high accuracy on the training set but showed a slight decrease in testing accuracy. This suggests that while Decision Trees can model complex relationships, they are more susceptible to overfitting when there is insufficient regularization.

Neural networks like MLPs are particularly useful for tasks where relationships between inputs and outputs are complex and non-linear, but they also require substantial computational resources and careful tuning of hyperparameters. In this study, although the MLP demonstrated good performance, it did not exceed the ensemble methods' accuracy.

In conclusion, the results of this study indicate that while advanced ensemble models like Random Forest provide the best performance in predicting air quality, challenges such as class imbalance and overfitting must still be addressed. Future work could explore techniques such as hyperparameter optimization, regularization, and the integration of additional real-time data sources.

## CONCLUSION

To conclude, the proposed air quality prediction system successfully applies ML algorithms to deliver accurate and reliable predictions based on various environmental factors. By utilizing a combination of ML models, the system is able to classify air quality into three categories: Good, Satisfactory, and Poor. These classifications are based on critical environmental indicators. This ability to accurately categorize air quality allows the system to offer valuable insights that can aid in decision-making processes related to public health and environmental management.

The project has also focused on ensuring that the models are trained on high-quality data through comprehensive data preprocessing techniques. Methods such as missing value imputation, feature scaling, and class balancing have been applied to address common challenges in environmental datasets. These preprocessing steps not only improve the quality of the data but also enhance the predictive accuracy of the machine learning models. By addressing issues like class imbalance through the use of SMOTE, the system ensures that predictions are more reliable, especially for underrepresented air quality categories.

This functionality makes the system highly applicable for raising awareness about air pollution and empowering citizens to take proactive steps to protect their health. Furthermore, it provides valuable information for policymakers, public health officials, and environmental agencies, allowing them to make informed decisions about pollution control measures and urban planning.

Looking ahead, future work on this system could focus on further optimizing the models by fine-tuning hyperparameters, exploring additional machine learning techniques, and integrating real-time data sources such as IoT sensors or environmental monitoring stations. This could improve the system's ability to offer up-to-date air quality predictions and expand its applicability to dynamic and fluctuating urban environments. Moreover, enhancing the system to support urban planning initiatives and pollution control strategies would provide significant value in the long term.

Overall, the proposed air quality prediction system demonstrates great potential in contributing to better air quality management. It can provide essential data-driven insights that support decision-making and public awareness efforts.

## FUTURE ENHANCEMENTS

Looking ahead, several key enhancements could be implemented to further improve the functionality, accessibility, and impact of the air quality prediction system. These enhancements would not only enhance the user experience but also expand the system's utility and reach.

Geographical Integration is one of the most promising future developments. By incorporating geographical data, the system could adjust predictions based on regional variations in air quality. For instance, users could input their specific city or location, and the system would use geolocation to access relevant local air quality data or forecasts. This would allow for more accurate and context-specific air quality predictions, especially in areas where pollution levels fluctuate widely based on weather patterns, industrial activities, or seasonal changes.

To further increase the accessibility and usability of the system, the development of a Mobile Application would be a crucial step forward. A dedicated mobile app would enable users to receive real-time air quality predictions and timely alerts directly on their smartphones. This feature would be particularly beneficial for individuals residing in high-pollution areas or those who need to monitor air quality on the go. The app could also push notifications to alert users of sudden air quality changes, ensuring that they are always aware of the conditions that may affect their health.

Health Recommendations based on predicted air quality and AQI values would add significant value to the system. Upon receiving air quality predictions, users could be provided with personalized health recommendations tailored to the current conditions. This feature would empower users to take proactive steps to protect their health, making the system not just a source of information but also a guide for healthier living.

Finally, implementing Multi-Language Support would greatly enhance the inclusivity and global applicability of the system. By offering the system in multiple languages, it could cater to a broader and more diverse user base, particularly in regions where multiple languages are spoken. This enhancement would make the system more accessible to people from different linguistic backgrounds, ensuring that individuals worldwide can benefit from accurate air quality predictions and health advice, regardless of the language they speak.

These future enhancements have the potential to transform the air quality prediction system into a more dynamic, accessible, and user-centric tool. By incorporating localized, real-time information, personalized health recommendations, and multilingual support, the system would not only contribute to improving public health but also create a more inclusive and responsive approach to environmental management.

## References

[1]. Aamer, H., Ba-Alawi, A. H., Kang, S., Lee, T., & Jo, Y. M. (2025). Prediction of school PM2.5 by an attention-based deep learning approach informed with data from nearby air quality monitoring stations. Chemosphere, 375, 144241. https://doi.org/10.1016/J.CHEMOSPHERE.2025.144241

[2]. Kant, S. (2024). From data to decision-making: utilizing decision tree for air quality monitoring in smart urban areas. International Journal of Information Technology (Singapore), 17(1), 665–672. https://doi.org/10.1007/S41870-024-02208-Y/METRICS

[3]. Lee, M. J., & Zhang, R. (2024). Multimodal Data Fusion and Deep Learning for Occupant-Centric Indoor Environmental Quality Classification. Journal of Computing in Civil Engineering, 39(2), 04024061. https://doi.org/10.1061/JCCEE5.CPENG-6249

[4]. Njaime, M., Abdallah, F., Snoussi, H., Akl, J., Chaaban, K., & Omrani, H. (2024). Transfer learning based solution for air quality prediction in smart cities using multimodal data. International Journal of Environmental Science and Technology, 22(3), 1297–1312. https://doi.org/10.1007/S13762-024-05722-5/METRICS

[5]. Pabitha, C., Pal Pandian, P., Saravanan, S., Rajendiran, M., Sneha Chaturya, A., & Natrayan, L. (2025). Development of robotic sensor nodes in wireless sensor networks using embedded systems and machine learning. Hybrid and Advanced Technologies, 513–518. https://doi.org/10.1201/9781003559139-76

[6]. Potharaju, S., Tirandasu, R. K., Tambe, S. N., Jadhav, D. B., Kumar, D. A., & Amiripalli, S. S. (2025). A two-step machine learning approach for predictive maintenance and anomaly detection in environmental sensor systems. MethodsX, 14, 103181. https://doi.org/10.1016/J.MEX.2025.103181

[7]. Rajendran, R. K., Mohana Priya, T., Musa, A. I. A., Mahalakshmi, S. B., & Anand, T. R. (1 C.E.). Smart Solutions for Climate Resilience Harnessing Machine Learning and Sustainable WSNs. Https://Services.Igi-Global.Com/Resolvedoi/Resolve.Aspx?Doi=10.4018/979-8-3693-3940-4.Ch010, 213–232. https://doi.org/10.4018/979-8-3693-3940-4.CH010

[8]. Rathnayaka, R. M. N., & Sujah, A. M. A. (2025). IoT-Driven Environmental Intelligence for Sustainable Tomorrow Through Advanced Machine Learning: A Systematic Literature Review. Journal of Information and Communication Technology (JICT), 02.

[9]. Sankaran, K. S. (2025). Design and development of sensor based crop monitoring system using deep learning approach for smart agriculture. Multimedia Tools and Applications, 1–22. https://doi.org/10.1007/S11042-024-20525-Z/METRICS

[10]. Varun, E., Rajesh, L., Lokeshwari, M., Ashwini, S., Bhat, D., & S., C. K. (1 C.E.). Machine Learning-Enhanced Water and Air Quality Monitoring Technologies and Applications. Https://Services.Igi-Global.Com/Resolvedoi/Resolve.Aspx?Doi=10.4018/979-8-3693-4759-1.Ch004, 79–110. https://doi.org/10.4018/979-8-3693-4759-1.CH004

[11]. Zeng, T., Ma, X., Luo, Y., Yin, J., Ji, Y., & Lu, S. (2025). Improving outdoor thermal environmental quality through kinetic canopy empowered by machine learning and control algorithms. Building Simulation 2025, 1–22. https://doi.org/10.1007/S12273-025-1246-6