Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor: 10.2
Subject: Python Programming and Machine Learning

AQI FORECASTING USING HYBRID METEOROLOGICAL FEATURE

Shreya Saini
Student of Computer Studies and Emerging Technology
TransStadia University, Ahmedabad
Emailshreyasaini2677@gmail.com
Bhavesh Jain
Assistant Professor
Computer Studies and Emerging Technology
TransStadia University, Ahmedabad.
Email: bhavesh.jain@tsuniv.edu.in

Abstract

Air pollution is one of the most important environmental issues, which affects the health of people, particularly in urban areas where air pollutant concentrations are often above safe levels. The Air Quality Index (AQI) is commonly used to describe air pollution levels and environmental risk. The forecasting of Air Quality Index is important in environmental monitoring. ARIMA models, which are statistical models, are commonly used to forecast time-series data, capturing linear and seasonal patterns in the data. Random Forest, which is a machine learning model, is used to capture nonlinear relationships in the data. Deep learning models, like Long Short-Term Memory (LSTM), have shown promising results in capturing temporal dependencies in time-series data. Moreover, meteorological factors like temperature, humidity, and wind speed have a significant impact on the accuracy of Air Quality Index forecasting [5], [8]. Hybrid models have shown promising results in forecasting Air Quality Index by integrating statistical, machine learning, and deep learning models [7], [9]. Moreover, the use of explainable AI has shown promising results in terms of model interpretability and transparency [10].

The present research introduces a novel framework for hybrid modeling of AQI forecasting using ARIMA, Random Forest, and LSTM techniques. The proposed framework takes into account linear, non-linear, and temporal relationships between the variables for more accurate prediction of the AQI. Explainable AI is

Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor: 10.2
Subject: Python Programming and Machine Learning

also incorporated using SHAP values to determine the most important features for the prediction of the AQI. Experimental results show that the proposed hybrid model is more accurate than other individual models.

Keywords Air Quality Index, Hybrid Model, ARIMA, Random Forest, LSTM, Machine Learning, Deep Learning, Explainable AI, SHAP Values.

I. Introduction

Air pollution is one of the most critical environmental issues facing humanity today. Air pollution at high concentrations has been linked to respiratory problems,

Cardiovascular problems, and environmental degradation. The Air Quality Index (AQI) is a standardized index used to measure air quality in terms of pollutant concentration, including PM2.5, PM10, NO2, SO2, and CO [5], [8]. The prediction of AQI is important in environmental monitoring and protection.

AQI prediction is a time-series forecasting problem depending on pollutant concentration and meteorological factors. The most commonly used statistical models in AQI forecasting are Autoregressive Integrated Moving Average (ARIMA) models, which have shown promising results in forecasting AQI. ARIMA models have shown promising results in forecasting AQI because they are able to capture linear patterns in time-series data. However, ARIMA models have some limitations in capturing non-linear patterns in pollutant concentration and meteorological factors [11].

Machine learning techniques like Random Forest and Support Vector Machine are used for AQI prediction. These techniques are effective in handling nonlinear relationships between pollutant and meteorological features. They also show better prediction accuracy compared to statistical techniques [2], [3]. Techniques like LSTM are also used in deep learning for AQI prediction. They show better accuracy in handling temporal dependencies in data and are effective in solving time-series prediction problems [9].

Meteorological features play an important role in AQI prediction. Temperature, humidity, wind speed, etc., are important meteorological features that are used in AQI prediction. They show better prediction accuracy and reliability [5], [8]. Hybrid techniques are also used in AQI prediction. They show better prediction accuracy compared to individual prediction techniques and hybrid combinations [7], [9].

Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor: 10.2
Subject: Python Programming and Machine Learning

However, existing studies mainly focus on individual prediction techniques and a few combinations of prediction techniques. They are not suitable to develop a unified hybrid and explainable AQI prediction system. Explainability is one of the most important challenges in developing AQI prediction systems based on machine learning and deep learning techniques. In order to solve the problems of the existing methods, a new hybrid framework of AQI forecasting using ARIMA, RF, LSTM, meteorological features, and XAI has been proposed in this study. The new framework can provide more accurate, reliable, and transparent AQI forecasting results by combining three different machine learning methods.

II. Literature Review

The prediction of Air Quality Index (AQI) has been extensively carried out by various statistical, machine learning, and deep learning models. These models have shown significant improvements in terms of accuracy, but at the same time, there are certain limitations associated with each of the mentioned models.

The application of machine learning models for AQI prediction problems has been extensively carried out. Ravindiran et al. [2] have successfully used various machine learning models such as Random Forest, XGBoost, LightGBM, CatBoost, and have shown that CatBoost produces the highest accuracy. Jayapradha et al. [3] have demonstrated that the Random Forest model produces the highest accuracy compared to other models by considering the ensemble learning property of the Random Forest model that reduces overfitting.

The role of meteorological factors is significant for AQI prediction. Al-Mutairi et al. [5] have shown that temperature, humidity, and wind speed are significant factors that influence AQI prediction accuracy. Liu et al. [8] have further proven that meteorological factors influence pollutant dispersion.

Long Short-Term Memory (LSTM) models have shown promising results in time-series AQI prediction problems. The LSTM model has shown superior performance in handling temporal dependencies and long-term patterns in AQI values [9]. However, the effectiveness of deep learning models in handling linear trends and statistical relationships may be limited.

Hybrid models have shown promising results in handling AQI prediction problems. Hybrid models, which combine statistical and machine learning models, have shown superior performance in handling AQI prediction problems. Wang et al. [7] have proposed a hybrid model combining ARIMA and Random Forest models, which has shown superior performance in handling AQI prediction problems. Other review papers

Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor: 10.2
Subject: Python Programming and Machine Learning

[9] have shown that hybrid models, which combine machine learning and deep learning models, have shown superior performance in handling AQI prediction problems.

Recently, Explainable AI has shown importance in handling AQI prediction problems. SHAP models have shown promising results in handling AQI prediction problems by providing insights into important features affecting AQI values [10].

However, despite these developments, existing studies mostly concentrate on individual models or limited hybrid approaches. There is a lack of a unified hybrid framework that includes statistical models, machine learning models, deep learning models, explainable AI approaches, and meteorological feature analysis.

III. Problem Statement and Research Objectives

A. Problem Statement

The existing AQI prediction models use statistical, machine learning, or deep learning methods separately. Although the existing models have reasonable accuracy, they do not consider linear, nonlinear, and time dependencies at the same time. The statistical model, ARIMA, considers linear trends but does not consider nonlinear relationships. The machine learning model considers nonlinear relationships but does not consider time dependencies. The deep learning model considers time dependencies but does not consider statistical relationships .Moreover, the existing AQI prediction systems are not interpretable, which means they do not provide insights into the importance of each feature. The meteorological features are not fully utilized in the hybrid prediction systems, which affects the accuracy of the AQI prediction systems. Thus, there is a need for a unified hybrid AQI prediction system that uses statistical, machine learning, and deep learning techniques.

B. Research Objectives

The primary objectives of the proposed research can be outlined as follows:

• To propose a hybrid AQI prediction model using ARIMA, Random Forest, and LSTM

• To enhance the accuracy of AQI prediction using linear, non-linear, and temporal relationships

• To include meteorological factors like temperature, humidity, and wind speed in the AQI prediction

Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor: 10.2
Subject: Python Programming and Machine Learning

• To leverage the power of Explainable AI in enhancing the interpretability of the proposed model

• To propose a robust and accurate AQI forecasting framework

IV. Proposed Solution

A. Overview of Proposed Hybrid Framework

In order to address the limitations of each prediction model, this research proposes a hybrid framework for AQI prediction that combines statistical modeling, machine learning, and deep learning techniques. Rather than relying solely on one model for prediction, the proposed framework will utilize the prediction capabilities of three different models: ARIMA, Random Forest, and LSTM.

AQI prediction is a complex problem that is influenced by many factors, including pollutant concentration and meteorological conditions. Each of these models will contribute to the overall understanding of the complex problem of AQI prediction. The ARIMA model will be utilized to extract linear trends from the AQI dataset, the Random Forest model will be utilized to extract non-linear relationships between pollutant and meteorological variables, and the LSTM model will be utilized to extract temporal dependencies from the dataset.

B. Data Representation and Feature Selection

The model will incorporate both pollutant features and meteorological features. The pollutant features will be comprised of PM2.5, PM10, NO2, SO2, and CO. All these contribute to the overall AQI. The meteorological features will comprise temperature, humidity, and wind speed. All these contribute to the overall AQI.The combination of these features will enable the model to understand the source of the pollution as well as the environment. Past research has shown that the inclusion of meteorological features will enhance the accuracy of the AQI prediction [5], [8].

C. Individual Model Components

The proposed hybrid model has three components:

1. ARIMA Model

Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor: 10.2
Subject: Python Programming and Machine Learning

The ARIMA model is employed to model linear and seasonal patterns in the AQI data. This model analyzes the past AQI data and identifies trends and seasonal patterns. However, the ARIMA model is only capable of modeling linear patterns.

2. Random Forest Model

The Random Forest model is employed to model nonlinear relationships between features. This model uses multiple decision trees to ensure high accuracy in predictions. This model is suitable when the AQI data depends on complex relationships.

3. LSTM Model

The LSTM model is employed to model time-dependent patterns in the AQI data. This model retains information from previous time steps and is suitable to model patterns in time series data.

D. Hybrid Integration Strategy

The predicted values from the ARIMA, Random Forest, and LSTM models are integrated to obtain the final prediction for the AQI. The predicted values from each of these models contribute to the final prediction in accordance with their prediction strengths. This helps the system utilize the advantages of these three models. The hybrid strategy helps to obtain a more accurate prediction for the AQI using linear trend prediction, non-linear feature prediction, and sequence prediction. This helps to obtain a more accurate prediction for the AQI.

E. Explainable AI Integration

To obtain a more accurate prediction for the AQI using the proposed framework, explainable AI is integrated into the framework. SHAP is integrated into the framework to obtain the feature importance of the predicted values. The SHAP helps to determine the contribution of each of the pollutant and meteorological features to the predicted value of the AQI.

For example, the pollutant feature values contribute more to the predicted value of the AQI than the meteorological feature values. Similarly, the feature values for the pollutants such as PM 2.5 contribute more to the predicted value of the AQI than other pollutants. The use of explainable AI increases the transparency of the model, which is useful in decision-making in the environment [10].

Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor: 10.2
Subject: Python Programming and Machine Learning

F. Multi-Horizon Forecasting

The proposed model is also capable of multi-horizon forecasting, which means it can be used to predict AQI at different times in the future, e.g., one day ahead, three days ahead, seven days ahead, etc. This increases the usefulness of the model in real-world applications.

G. Advantages of Proposed Solution

The proposed hybrid framework for AQI prediction has several advantages:

• Improved accuracy in prediction

• Ability to predict linear, non-linear, as well as temporal patterns

• Incorporation of meteorological features

• Improved interpretability using explainable AI

• Improved prediction performance

V. Mathematical Model Formulation

In this research, the AQI prediction will be modeled as a time series forecasting problem, wherein the predicted AQI value will depend on pollutant concentrations, meteorological conditions, as well as patterns of AQI levels. Instead of using complex mathematical formulations, the proposed method will focus more on the modeling of the relationship between environmental features and AQI levels through a hybrid learning framework .Pollutant features, which include PM2.5, PM10, NO2, SO2, and CO, as well as meteorological features like temperature, humidity, and wind speed, will serve as input data for the system. The combination of these input data will represent the different environmental conditions at a particular time.

To accomplish the AQI prediction, three different models will be used. The first model will use the ARIMA model, which will analyze the historical AQI data to identify linear trends as well as patterns. The second model will use the random forest model, which will learn the nonlinear relationships between pollutant features, as well as meteorological features. The third model will use the LSTM model, which will process the sequential data of AQI levels, including the previous time steps.

Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor: 10.2
Subject: Python Programming and Machine Learning

The final AQI prediction result is a combination of all three models. This combination is based on the models' capabilities to detect specific patterns. This hybrid technique enables the system to handle linear trends, nonlinear dependencies, and time-related patterns at the same time. This yields a more accurate prediction.

The final AQI prediction result is a combination of all three models. This combination is based on the models' capabilities to detect specific patterns. This hybrid technique enables the system to handle linear trends, nonlinear dependencies, and time-related patterns at the same time. This yields a more accurate prediction.

VI. System Architecture and Workflow

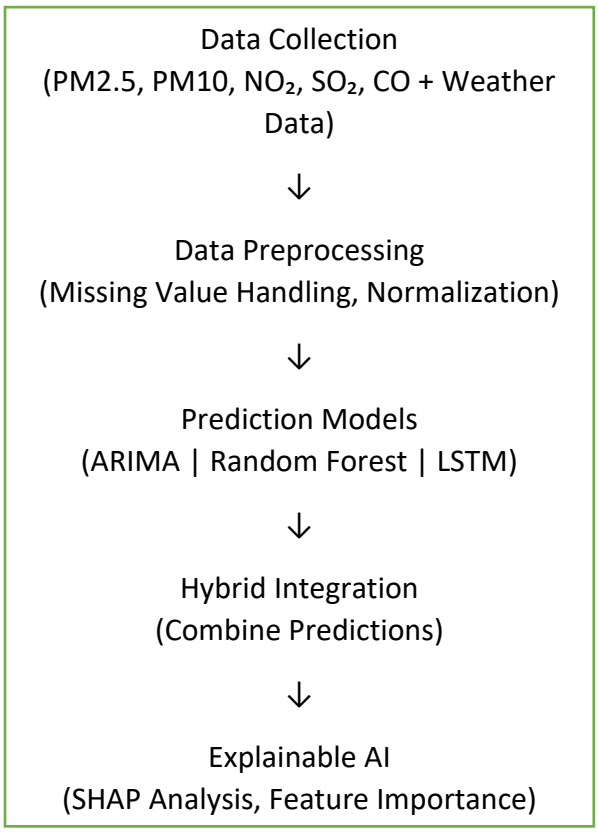
The proposed system architecture for AQI prediction employs a workflow with multiple stages. The workflow begins with data collection. In this stage, data related to pollutants and meteorological conditions are retrieved from air quality monitoring stations. The data is then subjected to a preprocessing stage where missing values are filled in and data normalization is performed. This stage helps to enhance data quality.

The data then undergoes prediction using three different prediction models: ARIMA, RF, and LSTM. Each model receives the data separately and performs prediction based on its individual learning mechanism to arrive at an AQI prediction. The prediction results are then sent to a stage where they are integrated to arrive at a single AQI prediction. This stage helps to ensure that all three models' predictions are utilized in an effective manner. Finally, the proposed system employs AI to analyze the features to arrive at an AQI prediction. In this stage, the importance of features in arriving at an AQI prediction is analyzed. The system then arrives at an AQI prediction and presents it as an output.

Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor: 10.2
Subject: Python Programming and Machine Learning

Figure 1: Proposed Hybrid AQI Prediction System Architecture



VII. Experimental Setup

The experimental setup is intended to assess the effectiveness of the hybrid AQI prediction model. The dataset utilized in this research consists of air quality and meteorological information gathered from air quality stations. The dataset contains pollutant features, which include PM2.5, PM10, NO2, SO2, and CO, and meteorological features, which include temperature, humidity, and wind speed. Before the models are trained, the dataset is preprocessed to address missing values and normalize the scale of the features. The dataset is split into a set of training data and testing data, where the models are trained on the training set and their effectiveness is tested on the testing set.

The experimental setup is implemented by utilizing the Python programming language and Python libraries. The machine learning models, deep learning models, and statistical models are implemented by utilizing

Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor: 10.2
Subject: Python Programming and Machine Learning

Scikit-learn, Tensor Flow, and relevant libraries, respectively. The setup is implemented to enable the efficient training and testing of the hybrid model.

VIII. Performance Evaluation

The performance of the proposed model is evaluated through the comparison of predicted values of AQI with actual values of AQI. The main aim is to determine how well the model predicts the values of AQI and how well it reduces prediction errors. The performance of the prediction is evaluated using the difference between actual values of AQI and predicted values of AQI. The smaller the difference between the values, the more accurate the model is. The performance of the hybrid model is likely to be better than that of other models. The evaluation of the performance of the model is also done using the consistency of the prediction. A good model is one that is consistent in prediction under different environmental conditions.

IX. Results and Discussion

The results of the experimental analysis show that the proposed hybrid model has better performance compared to individual models like ARIMA, Random Forest, and LSTM. The ARIMA model is effective only for linear trends, but not for nonlinear relationships. The Random Forest model has better prediction accuracy because of complex feature relationships. The LSTM model is effective for temporal relationships. The hybrid model has better performance because it uses all three models. This shows that using multiple models has better performance compared to using individual models. The results of Explainable AI analysis show that features like PM2.5 have the highest influence on AQI prediction. Meteorological features like wind speed and temperature are also important for AQI prediction. This shows that using pollutant features along with environmental features has better performance.

X. Conclusion

This study introduces a hybrid model for predicting AQI values, which uses ARIMA, Random Forest, and LSTM models. The hybrid model is able to effectively predict AQI values with high accuracy. Adding meteorological factors to the model improves the accuracy of the model in predicting AQI values. The study shows that the hybrid model is superior to individual models in predicting AQI values, thus providing a reliable solution to the problem.

Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor: 10.2
Subject: Python Programming and Machine Learning

XI. Future Work

Future research may involve extending the hybrid model to predict real-time values of AQI by incorporating real-time data from monitoring stations. Other research may involve the use of advanced deep learning models to improve the accuracy of the model. Moreover, the model may be extended to include multi-city AQI prediction and IOT-based monitoring systems to analyze environmental factors in real-time.

XII. References

[1] Y. Liu, et al., "Air quality prediction models based on meteorological factors," *Scientific Reports*, vol.12,2022.https://www.nature.com/articles/s41598-022-13579-2

[2] "Machine learning for air quality prediction and data analysis," *Atmosphere,* vol. 15, no. 11, Art. no. 1352,2024. https://www.mdpi.com/20734433/15/11/1352

[3] "Explainable machine learning for air quality prediction," *Aerosol and Air Quality Research*, vol. 23,2023.https://aaqr.org/articles/aaqr-23-06-oa-0151

[4] S. Maltare and H. Vahora, "AQI prediction using SARIMA, SVM and LSTM models," *Digital Chemical Engineering*, vol. 6, 2023. https://www.sciencedirect.com/science/article/pii/S277250812300011X

[5] "Air quality prediction using machine learning and meteorological data," *Science of the Total Environment*, vol. 905, 2025. https://www.sciencedirect.com/science/article/pii/S0048969725022338

[6] "Machine learning-based air quality prediction," *Chemosphere,* vol. 330, 2023. https://www.sciencedirect.com/science/article/pii/S004565352301785X

[7] "Explainable forecasting of AQI using hybrid models,"*PMC,*2024.https://pmc.ncbi.nlm.nih.gov/articles/PMC12329590/

[8] "Deep learning-based AQI prediction using LSTM," *Scientific Reports*, vol. 14, 2024.https://www.nature.com/articles/s41598-024-54807-1

[9] "Sustainable AQI prediction using hybrid techniques," *Sustainability*, vol. 17, 2024. https://www.mdpi.com/2071-1050/17/20/9136

[10] "Predictive machine learning model for air quality forecasting," *Environmental Systems Research*,Springer2024.https://link.springer.com/article/10.1186/s40068-024-00378-z

Airo International Journal
Peer-Reviewed
Multidisciplinary

ISSN: 2320-3714
Volume:2 Issue:1
April 2026
Impact Factor: 10.2
Subject: Python Programming and Machine Learning

[11] "Machine learning models for AQI prediction," *Journal of Neonatal Surgery*, 2024. https://www.jneonatalsurg.com/index.php/jns/article/view/5850/4931

Author's Declaration

As an author of the above research paper/article, here by, declare that the content of this paper is prepared by me and if any person having copyright issue or patent or anything otherwise related to the content, I shall always be legally responsible for any issue. For the reason of invisibility of my research paper on the website /amendments /updates, I have resubmitted my paper for publication on the same date. If any data or information given by me is not correct, I shall always be legally responsible. With my hole responsibility legally and formally have intimated the publisher (Publisher) that my paper has been checked by my guide (if any) or expert to make it sure that paper is technically right and there is no unaccepted plagiarism and hentriacontane is genuinely mine. If any issue arises related to Plagiarism/ Guide Name/ Educational Qualification /Designation /Address of my university/ college/institution/ Structure or Formatting/ Resubmission /Submission /Copyright /Patent /Submission for any higher degree or Job/Primary Data/Secondary Data Issues. I will be solely/entirely responsible for any legal issues. I have been informed that the most of the data from the website is invisible, shuffled, or vanished from the database due to some technical fault or hacking and therefore the process of resubmission is there for the scholars/students who find trouble in getting their paper on the website. At the time of resubmission of my paper I take all the legal and formal responsibilities, If I hide or do not submit the copy of my original documents (Andhra/Driving License/Any Identity Proof and Photo) in spite of demand from the publisher, then my paper may be rejected or removed from the website anytime and may not be consider for verification. I accept the fact that as the content of this paper and the resubmission legal responsibilities and reasons are only mine then the Publisher (Airo International Journal/Airo National Research Journal) is never responsible. I also declare that if publisher finds any complication or error or anything hidden or implemented otherwise, my paper may be removed from the website, or the watermark of remark/actuality may be mentioned on my paper. Even if anything is found illegal publisher may also take legal action against me.

Shreya Saini
Bhavesh Jain

******