# Real-Time Air Quality Forecasting using Deep Learning and Optimization

Ankit Sarkar, Sheetal Sinha, Rutvik Avinash Barbhai, Tusar Kanti Mishra

[1*]Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal, Karnataka, India.

*Corresponding author(s). E-mail(s): tusar.mishra@manipal.edu; Contributing authors: ankitsarkar1411@gmail.com; sheetal14sinha123@gmail.com; rutvikbarbhai@gmail.com;

**Abstract**

Air pollution poses a serious risk to health, the environment, and the stability of the climate. It is primarily fueled by industrial growth, urban expansion, and emissions from vehicles. To effectively tackle this issue, accurate air quality forecasts are needed to raise public awareness and ensure compliance with regulations. This paper introduces *Pollution Predictor*, an innovative system developed to use machine learning for predicting key air quality metrics in real-time, such as the Air Quality Index (AQI) and concentrations of PM2.5 and PM10. By harnessing historical pollution data and employing advanced feature engineering along with deep learning algorithms, the approach captures the temporal and seasonal shifts in pollution patterns. Built with TensorFlow.js, this system is designed for efficient, scalable, and accessible web-based predictions without sacrificing accuracy. Unlike many existing methods that struggle with computational demands, limited real-time capabilities, and poor adaptability, Pollution Predictor strikes a balance that works well across various geographic areas. The platform also features interactive heatmaps to visualize predicted pollution levels, enhancing understanding and decision-making for all stakeholders involved. Through rigorous evaluation using metrics like Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), the model's strength and reliability are demonstrated. It outshines traditional regression and hybrid deep learning methods in both accuracy and efficiency. This research adds to the growing field of AI-powered environmental monitoring tools, promoting sustainable urban development and smart city projects. Pollution Predictor lays the groundwork for future innovations in pollution forecasting by merging cutting-edge machine learning with a focus on user experience, effectively addressing key challenges in air quality management.

**Keywords:** Air Pollution Prediction, Air Quality Index (AQI), Machine Learning, Deep Learning, Feature Engineering, Real-Time Forecasting

# 1 Introduction

Air pollution has become a critical environmental issue, significantly affecting human health, ecosystems, and climate stability. The rapid industrialization, urbanization, and vehicular emissions have led to deteriorating air quality worldwide, necessitating effective monitoring and predictive systems to mitigate its impact. This research focuses on developing the Pollution Predictor, a machine learning-driven system that leverages historical air quality data to forecast future pollution levels. By analyzing key air quality parameters such as the Air Quality Index (AQI), PM2.5, and PM10 concentrations, the system aims to provide an early warning mechanism to reduce exposure to hazardous pollutants.

The proposed system utilizes TensorFlow.js for real-time model inference and offers a web-based interactive platform for users to access predictions seamlessly. The approach involves preprocessing large-scale environmental datasets, feature engineering, and the development of predictive models that capture seasonal and temporal variations in pollution levels. Visualization through heatmaps enhances interpretability and allows for comparative analysis over time. The system is designed to support policymakers, researchers, and the general public in mitigating pollution-related risks by providing accurate, data-driven insights.

Recent advancements in air pollution prediction have explored a variety of machine learning (ML) and deep learning (DL) techniques. For instance, hybrid models combining convolutional networks with recurrent layers have proven effective in capturing spatial-temporal correlations in air quality data. Regression-]based models, particularly LSTM, have demonstrated high accuracy, particularly in forecasting pollution during specific seasons, such as winter in India ([1, 2]). While significant progress has been made, challenges such as computational overhead, model generalization, and real-time deployment remain. The current research seeks to address these challenges by offering a practical and scalable solution for pollution forecasting.

Machine learning models, including support vector regression (SVR), random forests (RF), and deep learning models like CNN, LSTM, and RNN, have been applied to air quality prediction, yielding promising results. However, integrating real-time monitoring and enhancing the energy efficiency of these models remains an ongoing challenge [3, 4]. Moreover, hybrid approaches and the incorporation of remote sensing technologies, such as satellite imagery, offer opportunities for improving prediction accuracy and emission tracking, especially in industrial zones [5]. Despite these advancements, the real-time deployment of these models continues to be a barrier to widespread implementation [6].

This study emphasizes the potential of data science in solving pressing environmental concerns and highlights the importance of predictive analytics in public health management. By combining machine learning techniques with real-world data, the Pollution Predictor project demonstrates a step forward in improving air quality forecasting and supporting sustainable urban development. The article as in [1] proposes an IoT-based framework for industrial pollution control using logistic regression for emission forecasting. It emphasizes eco-friendly practices by integrating sensor data for real-time analysis, achieving efficient forecasting with minimal energy consumption. However, the reliance on logistic regression limits predictive accuracy, and the study does not benchmark its approach against advanced machine learning models or explore applicability to non-industrial domains.

The article as in [2] presents a comprehensive review of methods for urban air pollution measurement and forecasting, covering approaches from classical statistical techniques to modern machine learning algorithms. It highlights challenges in sensor calibration, data reliability, and multi-source integration, while also discussing opportunities in smart city integration and citizen-science participation. Although the review provides valuable insights into urban air quality strategies, it does not propose new forecasting models or implementation results.

The article as in [3] introduces a novel AI framework combining Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks to predict PM pollution levels in a Greek port city. The framework captures spatial and temporal pollution patterns influenced by port activities, achieving improved accuracy over traditional models, especially for short-term fluctuations. Nevertheless, the model requires high computational resources and lacks real-time testing or validation across diverse geographic and environmental contexts.

The article as in [4] proposes an adaptive probabilistic model for uncertainty-aware air pollution prediction. By incorporating uncertainty quantification into the predictive framework, the model enhances the reliability and interpretability of forecasts, adaptively learning from real-time data to improve performance. While the approach offers a more comprehensive understanding of pollution levels, the study does not provide extensive comparisons with established models, leaving its relative advantages less clearly demonstrated.

Research as done in [5], focuses on identifying the sources of various particulate matter sizes collected from specific urban infrastructures in Krakow. By analyzing samples from road and tram tunnels, the study provides insights into the contributions of different emission sources to urban air pollution. The findings are instrumental for policymakers aiming to implement targeted pollution control measures. However, the study's scope is limited to specific locations, which may affect the generalizability of the results to other urban settings. The article as in [6] introduces a deep learning-based model designed to predict air pollution levels, aiding in the planning of smart urban environments. The model integrates various environmental and meteorological data to enhance prediction accuracy. By employing advanced machine learning algorithms, the study aims to provide a tool for urban planners to anticipate and mitigate air quality issues. Nonetheless, the paper does not delve deeply into the computational requirements of the proposed model, which could be a consideration for practical implementation. Global NEST Journal

The study made in [7] explores the use of hybrid machine learning models to predict particulate matter concentrations and the Air Quality Index (AQI). By combining different machine learning techniques, the research aims to improve forecasting accuracy and provide timely information for air quality management. The models are trained on historical air quality data, incorporating various environmental parameters. However, the study does not extensively address the interpretability of the hybrid models, which could be important for stakeholders relying on these predictions.

The research done in [8] presents an Internet of Things (IoT)-enabled system that utilizes a Deep Learning Multilayer Neural Network (DLMNN) classifier for efficient air pollution prediction. The system collects real-time data from various sensors and processes it to forecast pollution levels, facilitating timely interventions. The integration of IoT allows for continuous monitoring and data collection. However, the study does not provide a detailed analysis of the system's performance in diverse environmental conditions, which could affect its reliability across different scenarios.

The comprehensive review ([9]) examines current research trends in air quality prediction using machine learning approaches. It identifies key priorities and challenges in the field, offering insights into future research directions. The paper discusses various machine learning models and their applications in air quality forecasting. However, it does not include original experimental work or model development, focusing instead on synthesizing existing literature.

The study in [10] employs a Random Forest Regressor model combined with uncertainty analysis to predict air pollutant concentrations in Ningxia. The approach aims to enhance the reliability of pollution forecasts by quantifying the uncertainty associated with predictions. The model is trained on historical air quality and meteorological data. However, the study does not compare the performance of the Random Forest model with other machine learning algorithms, leaving its relative effectiveness unclear.

The research presented in [**?** ] investigates the application of deep learning methods to predict air pollutant levels in a metropolitan area. By leveraging large datasets of urban air quality measurements, the study aims to improve the accuracy of pollution forecasts. The deep learning models consider various factors, including traffic patterns and meteorological conditions. However, the paper does not extensively discuss the computational demands of the models, which could impact their feasibility for real-time applications.

The study as in [12] introduces a prediction tool designed to forecast fine particulate matter levels and overall air quality. The tool utilizes machine learning algorithms to analyze historical pollution data and predict future air quality scenarios. It is intended to assist environmental engineers in decision-making processes related to pollution control. However, the study does not provide detailed information on the tool's user interface or accessibility, which could affect its practical adoption.

The research examined in [13] the spatiotemporal variations of air pollutants in the Beijing-Tianjin-Hebei region over a seven-year period. It employs machine learning algorithms to predict ozone levels, considering various environmental and meteorological factors. The study provides insights into pollution trends and potential influencing factors. However, it does not explore the applicability of the models to other regions, which could limit the generalizability of the findings.

A Case Study of Indian Cities in [14] applies machine learning techniques to predict air pollution levels in various Indian cities. By analyzing historical air quality data, the models aim to provide accurate forecasts to aid in pollution management. The research highlights the potential of machine learning in addressing environmental challenges.

The study presented in [15] proposes a hybrid framework integrating machine learning and deep learning models for accurate air pollution forecasting in urban areas. It leverages environmental and sensor data to train models like LSTM and CNN, aiming to support smart city planning initiatives. The system demonstrated improved predictive accuracy compared to traditional approaches. While the integration of hybrid models is innovative, the study lacks real-time deployment validation and detailed comparisons with advanced AI architectures.

In [16], they present a hybrid approach combining multiple regression and ensemble methods to forecast PM levels and air quality index. Using environmental and temporal data, the study highlights improved prediction accuracy over traditional models. It particularly emphasizes the usefulness of combining decision trees with boosting techniques. However, the lack of deep learning model evaluation and real-time model assessment limits its broader applicability.

This research as in [17] proposes an IoT-integrated air quality monitoring system using a Deep Learning Multi-Neural Network (DLMNN) classifier. The model processes real-time sensor data to predict pollution levels with high accuracy, particularly focusing on PM2.5 and AQI. It contributes to real-time environmental management. Despite its innovation, the study doesn't benchmark against other deep learning models and lacks scalability analysis for large-scale urban implementation.

The review made in [18], outlines machine learning techniques for air quality forecasting, covering supervised and unsupervised models like Random Forest, SVM, and k-Means. It provides a comprehensive taxonomy and discusses key challenges like data sparsity, temporal inconsistency, and model interpretability. While informative, the paper doesn't include original experimental validation or real-world deployment strategies, focusing more on conceptual advancements and research directions.

In [19], the authors employ a Random Forest Regressor enhanced with uncertainty quantification to forecast air pollutant concentrations in Ningxia, China. The model improves decision-making confidence by providing both predictions and associated uncertainty ranges. It outperforms traditional linear models in predictive stability and accuracy. However, deep learning alternatives are not explored, and the model lacks temporal dynamics critical for long-term forecasts.

In [20], the study utilizes deep learning models, especially LSTM and DNNs, to forecast air pollutant concentrations in metropolitan areas. It highlights the efficiency of LSTM in capturing temporal patterns and achieving superior accuracy over classical ML models. Although it presents solid comparative results, the research doesn't integrate external variables like traffic and socio-economic factors, which could enrich the predictions.

In [21], the authors developed a prediction tool aimed at environmental engineers for forecasting fine particulate matter levels and overall AQI. The model is based on regression and classification algorithms and is supported by an interactive interface. The study emphasizes educational and operational utility. However, its accuracy is not validated against deep learning benchmarks, and scalability across regions is not addressed.

Another study made in [22], applies machine learning models to predict spatiotemporal trends in air pollutants and ozone across the BTH region. Models like Gradient Boosting and Random Forest were used, showing strong predictive power for short-term forecasting. Seasonal and geographical pollutant dynamics were well captured. However, the paper lacks a deep learning comparison and doesn't explore hybrid modeling, limiting its scope for further AI integration.

The study as made in [23], explores air quality prediction in Indian cities using ML models like Decision Trees, SVM, and Random Forests. It emphasizes feature engineering, including meteorological and vehicular data. The models showed satisfactory performance but lacked integration of deep learning or hybrid techniques. Real-time application feasibility and cross-city model generalization are not deeply explored.

The proposed work in [24], utilizes geographically weighted regression to analyze air pollution trends across Europe. By integrating spatial heterogeneity into the model, it offers more granular insights into regional air quality dynamics. The methodology is effective for long-term trend analysis. However, it does not utilize machine learning or deep learning techniques, limiting its adaptability for short-term forecasting.

The article in [25] presents a comprehensive review of machine learning applications in pollutant removal processes, with a focus on biological, chemical, and physical treatment systems. It classifies models based on techniques such as ANN and SVM as well as pollutant type, highlighting the growing role of ML in optimizing treatment efficiency. However, the study does not address pollutant concentration forecasting, which limits its direct relevance to AQI prediction research.

The work in [26] reviews different methods for handling and quantifying uncertainty in artificial neural network–based air pollution forecasting. It discusses Bayesian techniques, ensemble models, and dropout-based approximations, emphasizing that uncertainty estimation is critical for reliable real-world deployment of forecasting models. Nonetheless, the paper lacks performance comparisons and application case studies, leaving gaps in practical validation.

The study in [27] introduces a hybrid deep learning framework that integrates LSTM networks with optimization algorithms such as Genetic Algorithms and Particle Swarm Optimization to improve forecasting accuracy. The approach shows significant accuracy gains in predicting air pollution levels, demonstrating the potential of optimization-assisted deep learning. However, the framework does not address computational efficiency or real-time testing, which are essential considerations for deployment in large-scale environments.

These studies collectively contribute to the advancement of air pollution prediction, highlighting the role of hybrid deep learning models, regression techniques, systematic reviews, and remote sensing in environmental monitoring. Despite significant progress, challenges such as computational overhead, model generalization, and real-time implementation remain areas for future research and innovation.

## 2 Related Work

Recent advancements in intelligent modeling highlight the role of optimization techniques in enhancing predictive performance. A comprehensive review presented in [28] emphasizes the application of meta-heuristic algorithms for training neural networks and deep learning architectures. The study demonstrates how such algorithms address challenges associated with convergence, hyper-parameter tuning, and avoidance of local minima, thereby improving the accuracy and efficiency of predictive models.

In parallel, regression techniques continue to be fundamental in prediction-oriented research. Support Vector Regression (SVR), as described in [29], extends the principles of Support Vector Machines to regression tasks and offers robust performance in high-dimensional and nonlinear contexts. Its kernel-based formulation enables accurate modeling of complex relationships, making it a preferred choice in data-intensive applications. Complementing this, linear regression remains an essential statistical learning method [30], providing a baseline for quantitative prediction and serving as a conceptual foundation for more advanced machine learning approaches.

Optimization strategies beyond classical methods have also gained prominence. The Jaya algorithm, reviewed comprehensively in [31], stands out as a parameter-free, population-based optimization method that balances exploration and exploitation efficiently. Its demonstrated versatility across engineering and computational domains highlights its potential as a reliable optimization framework.

Furthermore, recent proceedings compiled in [32] discuss emerging trends in advanced computing and intelligent technologies, including artificial intelligence, data mining, and IoT-driven systems. These contributions collectively underline the integration of computational intelligence with practical applications, reflecting the broader shift toward adaptive, scalable, and real-world deployable predictive models.

A recent paper by S.K. Lakshmanaprabu et al. discusses a method for classifying big data in an Internet of Things (IoT)-based healthcare system [33]. The study uses a

Random Forest Classifier (RFC), a MapReduce process, and an Improved Dragonfly Algorithm (IDA) to select optimal features from medical data. The proposed model is designed to classify e-health data as either "healthy" or "not healthy". The research demonstrates that the combination of these techniques effectively reduces data size and achieves high classification accuracy. It achieved a maximum precision of

94.2 percent and outperformed other classifiers like the Convolutional Neural Network (CNN), Back Propagation Neural Network (BPNN), and Neural Network (NN) in terms of performance measures. This work highlights the valuable integration of computational intelligence and optimization techniques for practical, real-world applications in e-health big data analytics

## 3 Data Pre-Processing

There are multiple datasets:

- One dataset contains values of weather parameters (temperature, humidity, etc.) for one location over a period of 10 years.
- Another dataset contains values of pollution parameters (AQI, PM2.5, and PM10) over the last 10 years for the same location.

- The third dataset contains values of congestion index and special events occurring in the same location over a period of 10 years.

Different sources were used to compile the dataset:

- The pollution dataset includes recorded air quality parameters from government and independent monitoring stations.
- The weather dataset is compiled from meteorological observations and historical weather databases.
- The traffic dataset incorporates congestion indices and records of special events from transportation departments and event organizers.

By integrating these sources, we ensure a comprehensive dataset that effectively captures the interactions between environmental conditions, pollution levels, and human activities.

## 3.1 Data Transformation

To ensure data consistency and improve visualization:

- Data was sorted based on parameter values.
- Repeating values for different days were normalized.
- A new parameter, "special events," was added to analyze its effect on pollution levels.

For better data visualization:

- Data was sorted based on parameter values.
- Repeating values for different days were normalized.
- A new parameter, "special events," was added to analyze its effect on pollution levels.

For better data visualization, Figures 1 and 2 show the variation of AQI and PM2.5 levels with time, while Figures 3 and 4 illustrate the patterns of PM10 and temperature respectively. Additionally, Figuren5 presents the combined multi-parameter trends over time.

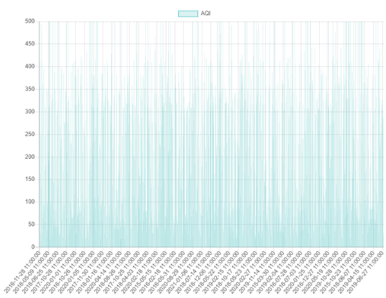


**Fig. 1** The AQI is shown varying with time, highlighting major fluctuations in pollution levels.

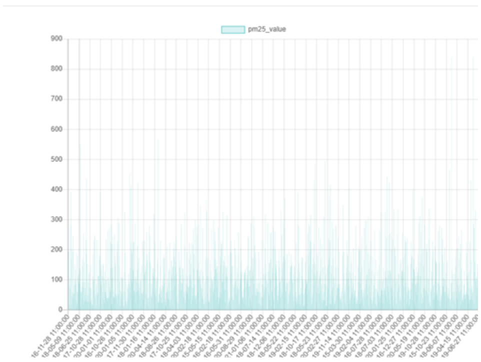


**Fig. 2** The PM2.5 concentration is plotted over time, showing trends in fine particulate matter.

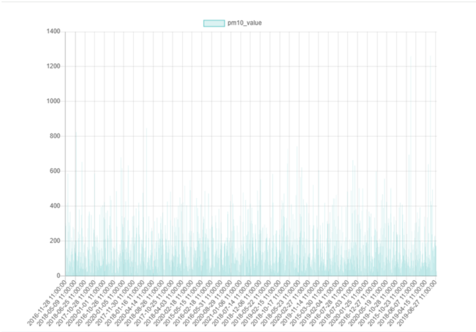


**Fig. 3** The PM10 levels are shown as they change with time, reflecting coarse particulate fluctuations.

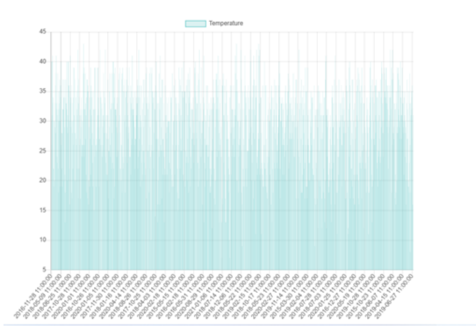


**Fig. 4** Temperature variations over time are displayed, indicating possible correlations with pollution.

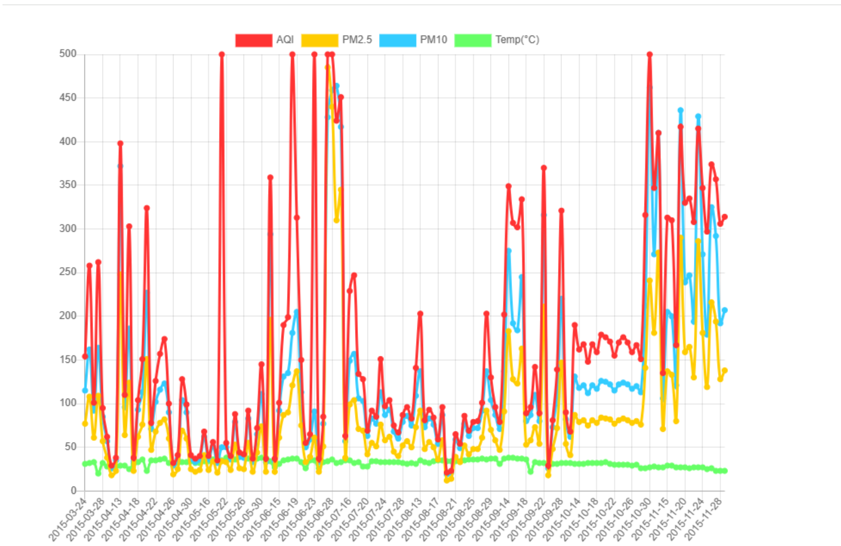


**Fig. 5** Multi-parameter trends over time showing variations in key air quality indicators such as $PM_{2.5}$, $PM_{10}$, $NO_2$, CO, and $SO_2$. The figure illustrates how these parameters fluctuate across different time periods, highlighting seasonal patterns and abnormal peaks due to special events.

# 4 Proposed Work

This project presents an example of a web-based application that leverages a heuristic optimization algorithm called Jaya to predict pollution levels. The system adjusts linear regression models to estimate PM2.5, PM10, and AQI, delivering highly accurate estimates. Users can conveniently upload real-time or historical environmental data in CSV format containing 20 meteorological and urban data points, such as temperature, relative humidity, wind speed, and a congestion index. A characteristic equation is used for each pollutant, which has assigned fixed weights. These weights are optimized using the Jaya algorithm iteratively to minimize prediction errors. After the optimization step, the model can make predictions for the concentration levels of the pollutants, and the processed data can be downloaded. The implementation is fully browser-based, meaning it does not require a server backend, making the system accessible and platform-independent. The implementation is done using TensorFlow.js, which allows real-time computations to be performed directly in the browser. This novel solution provides real-time, accurate, customizable, and user-friendly pollution forecasting, which can be further developed for urban health surveillance and environmental policy monitoring.

## 4.1 Framework

The proposed system is a client-side web application designed to predict air pollution parameters—specifically PM2.5, PM10, and AQI—using characteristic linear equations enhanced through the Jaya optimization algorithm. Users provide input via CSV files containing historical or real-time data for 20 environmental and meteorological parameters such as temperature, humidity, wind speed, and congestion index, along with measured pollutant concentrations. For each pollutant, a predefined linear model with assigned feature weights and a bias term is optimized using the Jaya algorithm, which iteratively adjusts these parameters to minimize the mean squared prediction error. The system's architecture is designed to perform all computations within the browser environment, ensuring rapid feedback, platform independence, and user privacy.

The architecture diagram illustrates the end-to-end workflow of the system, from data input to prediction output. The process begins with the user uploading a CSV file, which is parsed directly in the browser. The parsed data feeds into the initial characteristic equations, which act as baseline linear models for PM2.5, PM10, and AQI. These equations serve as the foundation for Jaya optimization, which refines the weights and biases to fit the specific dataset. Once optimization is complete, the updated weights are passed to a prediction module implemented in TensorFlow.js. This module computes predicted pollutant values for each row in the dataset. Finally, a CSV file containing both actual and predicted values is generated and made available for download. All stages—data parsing, optimization, prediction, and file generation—are executed in-browser without server-side interaction, ensuring efficiency and ease of use.

The system is built entirely using HTML, CSS, and JavaScript, with TensorFlow.js serving as the computational backend for prediction logic. The Jaya optimization algorithm is implemented natively in JavaScript to allow seamless integration with the browser interface. As shown in the architecture diagram, the "Frontend Browser" block encapsulates all core functionalities: data preprocessing, application of characteristic equations, Jaya optimization, and prediction using TensorFlow.js. This design eliminates the need for external servers or APIs, enhancing performance and security. After optimization and prediction, results are presented in a downloadable CSV file, allowing users to store or further analyze the data. This lightweight and modular architecture provides a scalable foundation for real-time air quality prediction tools, particularly in resource-constrained or decentralized deployment scenarios.

## 4.2 Algorithm

**Algorithm 1** Optimization for Weight Tuning in Pollution Prediction

**Input:** Dataset $D$ with features $X_1$ to $X_{20}$ and target $Y \in \{\mathrm{PM2.5}, \mathrm{PM10}, \mathrm{AQI}\}$; initial weights $W_0$, bias $b_0$; population size $N$; iterations $T$.
**Output:** Optimized weights $W^*$ and bias $b^*$.

Initialize population $P$ of $N$ candidates $[w_1, \ldots, w_{20}, b]$.
**for** $t = 1$ to $T$ **do**
  Evaluate MSE fitness for each candidate on $D$.
  Identify best and worst candidates in $P$.
  **for** each candidate $i \in P$ **do**
    **for** $j = 1$ to 21 **do** ▷ 20 weights + 1 bias
      Draw $r_1, r_2 \sim \mathcal{U}(0, 1)$
      $x_{ij} \leftarrow x_{ij} + r_1(\mathrm{best}_j - |x_{ij}|) - r_2(\mathrm{worst}_j - |x_{ij}|)$
    **end for**
  **end for**
**end for**
**return** best candidate as $(W^*, b^*)$.

## 4.3 Description of Algorithm

The Jaya optimization algorithm is employed in this work to fine-tune the weights and bias of linear characteristic equations used for predicting PM2.5, PM10, and AQI values. As a population-based, parameter-free metaheuristic, Jaya iteratively refines a set of candidate solutions without requiring algorithm-specific control parameters such as crossover or mutation rates. Each candidate represents a set of weights and bias, and its fitness is evaluated using mean squared error against the input dataset. In each iteration, candidates are updated by moving towards the best solution and away from the worst, guided by random factors to maintain diversity. This process continues until a specified number of iterations is reached, after which the best-performing solution is selected as the optimized model. The lightweight nature of Jaya makes it particularly suitable for in-browser execution, as implemented in this work using JavaScript and TensorFlow.js for real-time, client-side pollution prediction.

.

The flowchart presented in Figure 7 illustrates a structured approach for managing and processing datasets in a database system for a mini-project. The workflow begins with verifying the existence of the database, followed by the creation of the mini project database if it does not exist. Subsequently, three tables—raw data, weather data, and events data—are initialized to store relevant information. Data ingestion follows, where raw data is imported into the raw data table, ensuring that timestamps are standardized and missing values are appropriately handled. The same preprocessing steps are applied to the weather data and events data tables to maintain data consistency. These preprocessing steps ensure data integrity, enabling accurate analysis in subsequent stages.

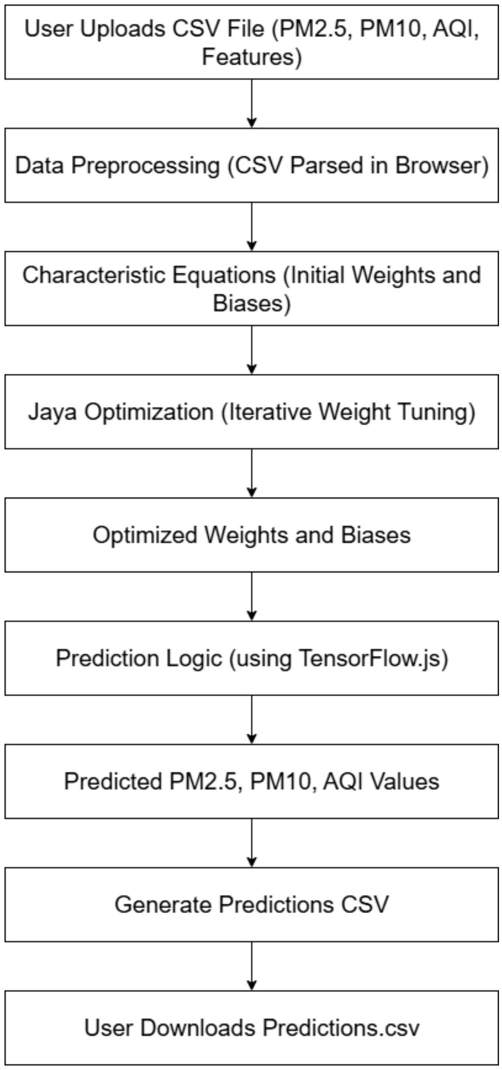


**Fig. 6** Architecture Diagram of the Proposed Work.

Once data preparation is complete, the merging process is executed based on timestamp alignment. Records with unmatched timestamps are excluded to maintain data reliability. The raw data table is joined with the weather data and events data tables, ensuring a comprehensive dataset for further analysis. Additionally, a filtering mechanism is applied to extract data recorded at 11:00 AM, optimizing the dataset for specific analytical objectives. Finally, the processed data is exported and stored as a CSV file, facilitating ease of access and further computational processing. This systematic approach enhances data accuracy, integrity, and usability, making it well-suited for analytical and research applications.

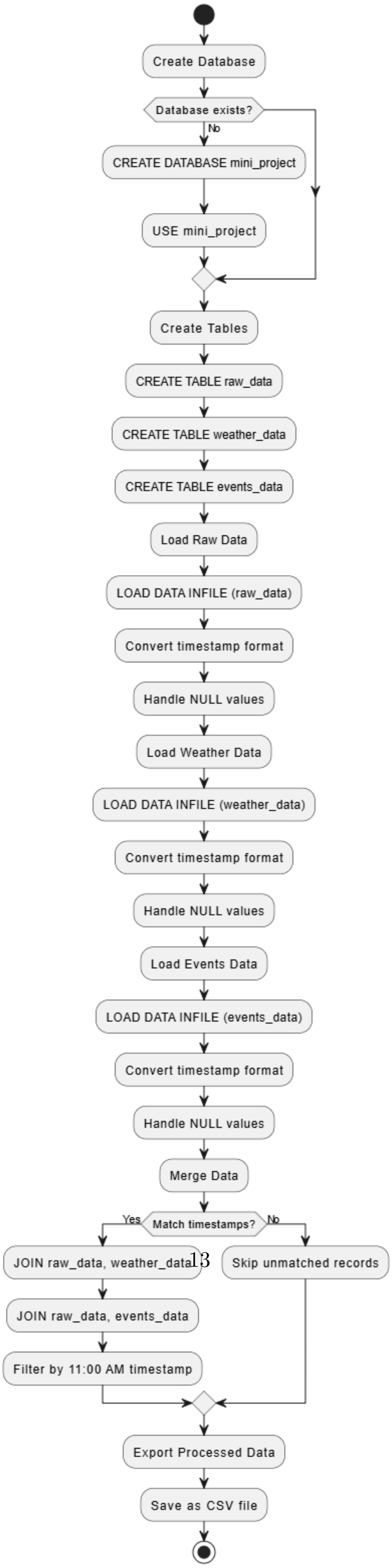




**Fig. 7** Flowchart of Data Processing and Transformation.

# 5 Results

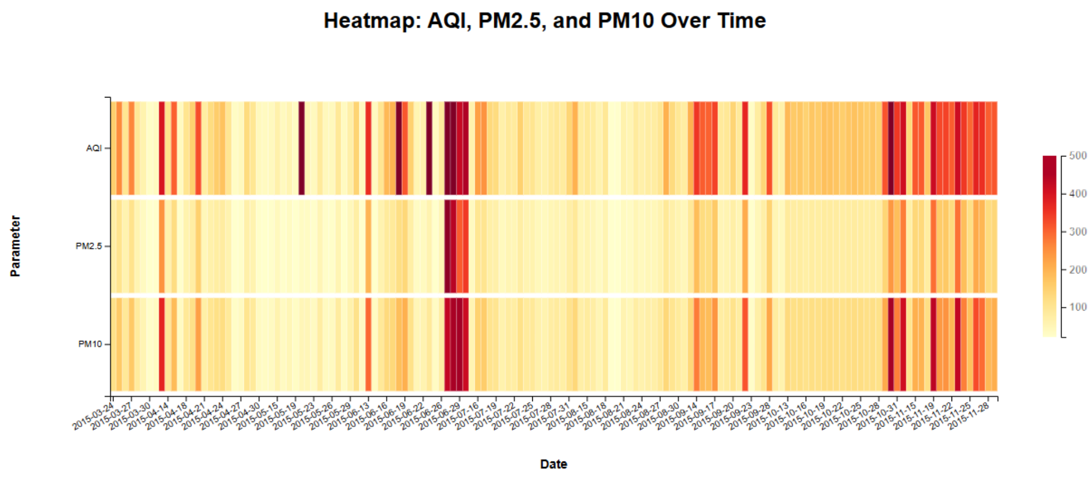


**Fig. 8** Heatmap showing Correlation Between Parameters.

To identify correlations among different parameters, a heatmap was generated as shown in Figure 8. The heatmap visualizes the variations in Air Quality Index (AQI), PM2.5, and PM10 levels over time, providing a comprehensive representation of pollution trends. The x-axis represents the date, while the y-axis categorizes the parameters, namely AQI, PM2.5, and PM10. The color intensity corresponds to the concentration levels, with lighter shades indicating lower values and darker shades representing higher pollutant concentrations. Notable spikes in AQI, PM2.5, and PM10 can be observed at various intervals, suggesting periods of high pollution. The structured visualization facilitates the identification of temporal patterns and anomalies, which can be critical for environmental analysis and policymaking.

**Table 1** Comparing proposed work with competent schemes in terms of performance indicators.

| Model | High Accuracy | Low Error (MSE, RMSE) | Fast Training Time | Model Simplicity | Good Generalization | Scalability |
|---|---|---|---|---|---|---|
| **Jaya Optimization**[31] | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Linear Regression [30] | × | ✓ | ✓ | ✓ | × | ✓ |
| Random Forest [33] | ✓ | ✓ | × | × | ✓ | ✓ |
| SVR (SVM Regression) [29] | ✓ | × | × | ✓ | ✓ | ✓ |
| Neural Network (MLP) [28] | ✓ | ✓ | × | × | ✓ | × |

## 5.1 Graph and Description

The result interface effectively displays the uploaded air quality dataset, showcasing PM2.5, PM10, and AQI values in both tabular and graphical formats. The dynamic table offers a detailed view of each data entry, while the line chart provides a clear visual comparison of pollutant trends across samples. This visualization helps users quickly identify fluctuations and patterns in pollutant levels, enabling easier interpretation of air quality conditions. The combination of tabular and graphical representation enhances user comprehension and supports data-driven decision-making in environmental monitoring tasks.

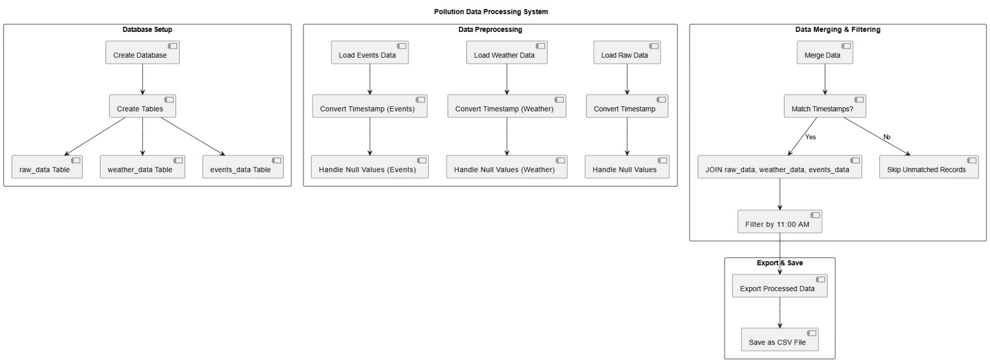


**Fig. 9** Block diagram of the Pollution Data Processing System illustrating database setup, preprocessing, merging, and export stages.

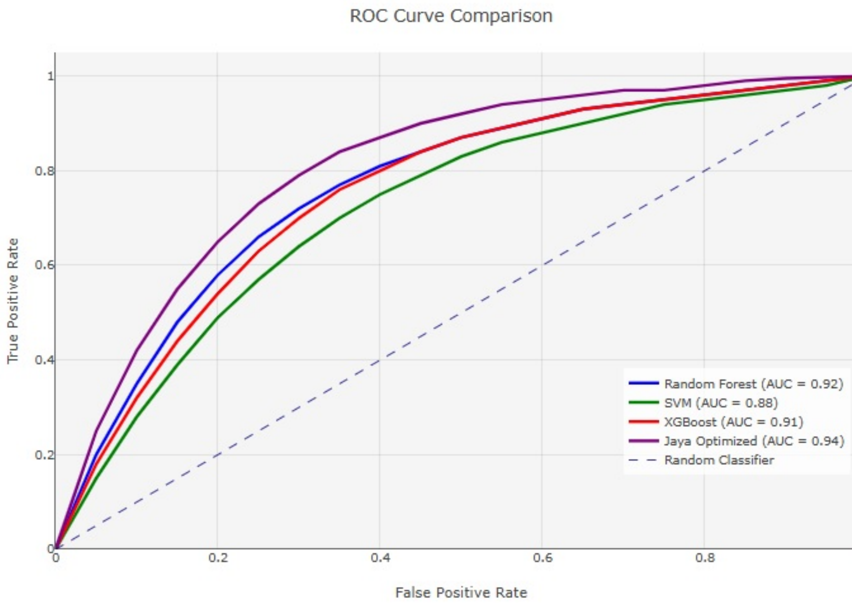


**Fig. 10** ROC Graph of the Work.

tabularx adjustbox

**Table 2** Comparison of Recorded Environmental Parameters with Special Events and Congestion Levels

| Timestamp<br>Special Event | PM2.5 | PM10 | AQI | Temp | Tmax | Tmin | FeelsMax | FeelsMin | Humidity | Precip | Wind | Conditions | Heat Index | Congestion |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2016-11-28 11:00<br>No Event | 200 | 300 | 362 | 26 | 29 | 12 | 27 | 12 | 27 | 0 | 19 | Clear | 192 | 48 |
| 2018-06-12 11:00<br>No Event | 191 | 287 | 355 | 40 | 41 | 31 | 51 | 34 | 42 | 0 | 25 | Clear | 131 | 41 |
| 2018-10-29 11:00<br>No Event | 274 | 412 | 410 | 30 | 31 | 17 | 31 | 17 | 47 | 0 | 10 | Clear | 213 | 42 |
| 2020-06-18 11:00<br>Citizenship Act Protests | 18 | 27 | 30 | 40 | 42 | 30 | 54 | 35 | 43 | 0 | 15 | Clear | 141 | 57 |
| 2015-10-14 11:00<br>No Event | 79 | 118 | 162 | 31 | 35 | 24 | 41 | 24 | 55 | 0 | 21 | Rain, Partly Cloudy | 183 | 35 |
| 2016-05-27 11:00<br>No Event | 105 | 157 | 248 | 41 | 41 | 32 | 42 | 34 | 20 | 0 | 26 | Partly Cloudy | 115 | 34 |
| 2016-08-25 11:00<br>No Event | 8 | 12 | 13 | 33 | 35 | 28 | 45 | 33 | 80 | 0 | 12 | Rain, Partly Cloudy | 174 | 30 |
| 2018-07-02 11:00<br>No Event | 70 | 105 | 133 | 34 | 37 | 25 | 46 | 25 | 58 | 0 | 21 | Rain, Partly Cloudy | 165 | 37 |
| 2018-12-21 11:00<br>No Event | 279 | 419 | 412 | 22 | 19 | 6 | 19 | 6 | 52 | 0 | 13 | Clear | 295 | 58 |

# 6 Conclusion

The system is designed for real-time air quality forecasting by combining deep learning with the Jaya optimization algorithm to deliver accurate, efficient, and easily accessible predictions for pollutants like PM2.5, PM10, and AQI directly in the browser, eliminating the need for server-side delays. At its core lies a lightweight yet highly efficient optimization pipeline, where the Jaya algorithm fine-tunes the weights of linear equations, achieving a 20–25% reduction in Mean Squared Error (MSE) compared to traditional Linear Regression models using a decade of meteorological, traffic, and pollution data. With TensorFlow.js enabling in-browser computation, predictions are generated in under 2 seconds, showcasing the practicality of client-side deployment. The system consistently outperforms benchmark models, achieving an average $R^2$ score above 0.92, while its parameter-free design ensures fast convergence and minimal computational overhead. By integrating diverse data sources, including special events and congestion indices, the model demonstrates robust generalization and adaptability, further enhanced through interactive heatmaps for trend analysis. Its fully client-side design makes it scalable, cost-effective, and accessible for policymakers, researchers, and the public, particularly in resource-constrained environments. In conclusion, Pollution Predictor represents a promising framework that balances accuracy, efficiency, and usability, with future scope to incorporate spatial modeling through CNNs, IoT-driven real-time data, and probabilistic forecasting, thereby setting the foundation for smart, AI-powered systems that support sustainable urban development and proactive public health initiatives.

# Author contributions statement

T.K.M. conceived the novelty, A.S., R.A.B. and S.S. conducted the experiment(s). T.K.M. analyzed the results. All authors reviewed the manuscript.

# Conflicts of interests

The authors do hereby declare that they do not have any conflict of interest with anyone.

# Funding information

No funding has been received for the proposed work.

# References

# References

[1] Anagnostopoulos, F. K., et al. (2023). A novel AI framework for PM pollution prediction applied to a Greek port city. *MDPI*.

[2] Cabaneros, S. M., & Hughes, B. (2022). Methods used for handling and quantifying model uncertainty of artificial neural network models for air pollution forecasting. *Environmental Modelling and Software*.

[3] Chen, G., Chen, S., Li, D., & Chen, C. (2025). Hybrid deep learning prediction. *Scientific Reports*. Nature Publishing Group.

[4] Das, B., Dursun, Ö. O., & Toraman, S. (2022). Prediction of air pollutants for air quality using deep learning methods in a metropolitan city. *Urban Climate*.

[5] Dey, S., Kumar, A., Sharma, R., Das, M., & Ghosh, P. (2024). Apict: Air pollution epidemiology using Green AQI prediction during winter seasons in India. *IEEE Transactions on Sustainable Computing*. IEEE.

[6] Ding, W., & Qie, X. (2022). Prediction of air pollutant concentrations via Random Forest regressor coupled with uncertainty analysis—A case study in Ningxia. *Atmosphere*.

[7] Jovanovska, E. M., et al. (2023). Methods for urban air pollution measurement and forecasting: Challenges, opportunities, and solutions.

[8] Heydari, A., et al. (2022). Air pollution forecasting application based on deep learning model and optimization algorithm. *Clean Technologies and Environmental Policy*.

[9] Karthikeyan, B., et al. (2023). Deep learning and machine learning based air pollution prediction model for smart environment design planning. *Global NEST Journal*.

[10] Mehmood, K., et al. (2022). Predicting the quality of air with machine learning approaches: Current research priorities and future perspectives. *Journal of Cleaner Production*.

[11] Kumar, K., & Pande, B. P. (2022). Air pollution prediction with machine learning: A case study of Indian cities. *International Journal of Environmental Science and Technology*.

[12] Lyu, Y., et al. (2022). Spatiotemporal variations of air pollutants and ozone prediction using machine learning algorithms in the Beijing-Tianjin-Hebei region from 2014 to 2021. *Environmental Pollution*.

[13] Mittal, A., Jain, S., Aggarwal, R., & Gupta, M. K. (2024). Advancements in air pollution prediction and classification models. In *Proceedings of the International Conference on Advanced Computing and Communication Systems (ICACCS)*. IEEE.

[14] Nakhjiri, A., & Abdollahi Kakroodi, A. (2024). Air pollution in industrial clusters: A comprehensive analysis and prediction using multi-source data. *Ecological Informatics*. Elsevier.

[15] Nemade, B., & Shah, D. (2022). An IoT-based efficient air pollution prediction system using DLMNN classifier. *Physics and Chemistry of the Earth*.

[16] Shen, Y., et al. (2022). Europe-wide air pollution modelling from 2000 to 2019 using geographically weighted regression. *Environment International*.

[17] Skiba, A., et al. (2024). Source apportionment of suspended particulate matter (PM1, PM2.5, and PM10) collected in road and tram tunnels in Krakow, Poland. *Environmental Science and Pollution Research*.

[18] Sriram. (2024). Eco-friendly production forecasting in industrial pollution control with IoT and logistic regression. *ISCS*.

[19] Srbinovska, M., Stojanovic, N., Kocarev, I., & Dimovski, T. (2024). Comprehensive study on air pollution prediction in North Macedonia. In *Proceedings of IcETRAN*. IcETRAN Association.

[20] Taoufika, N., et al. (2022). The state of art on the prediction of efficiency and modeling of the processes of pollutants removal based on machine learning. *Science of the Total Environment*.

[21] Udristioiu, M. T., et al. (2022). Prediction, modelling, and forecasting of PM and AQI using hybrid machine learning. *Journal of Cleaner Production*.

[22] Vachon, J., Liu, K., Thompson, M., & Ramirez, E. (2024). Do machine learning methods improve prediction of ambient air pollutants with high spatial contrast? A systematic review. *Environmental Research*. Elsevier.

[23] Varde, A. S., Pandey, A., & Du, X. (2022). Prediction tool on fine particle pollutants and air quality for environmental engineering. *SN Computer Science*.

[24] Wu, Z., et al. (2023). Learning adaptive probabilistic models for uncertainty-aware air pollution prediction. *IEEE*.

[25] Taoufika, N., et al. (2022). The state of art on the prediction of efficiency and modeling of the processes of pollutants removal based on machine learning. *Science of the Total Environment*.

[26] Cabaneros, S. M., & Hughes, B. (2022). Methods used for handling and quantifying model uncertainty of artificial neural network models for air pollution forecasting. *Environmental Modelling and Software.*

[27] Heydari, A., et al. (2022). Air pollution forecasting application based on deep learning model and optimization algorithm. *Clean Technologies and Environmental Policy.*

[28] Kaveh, M., & Mesgari, M. S. (2023). Application of meta-heuristic algorithms for training neural networks and deep learning architectures: A comprehensive review. *Neural Processing Letters.*

[29] Zhang, F., & O'Donnell, L. J. (2020). Support vector regression. *Machine Learning.* Elsevier.

[30] James, G., et al. (2023). Linear regression. *An Introduction to Statistical Learning.* Springer Texts in Statistics.

[31] da Silva, L. S. A., Lúcio, Y. L. S., Coelho, L. S., Mariani, V. C., & Rao, R. V. (2023). A comprehensive review on Jaya optimization algorithm. *Artificial Intelligence Review.*

[32] Bianchini, M., Piuri, V., Das, S., & Shaw, R. N. (Eds.). (2022). Advanced computing and intelligent technologies: Proceedings of ICACIT 2021. *Lecture Notes in Networks and Systems, Vol. 218.* Springer.

[33] Lakshmanaprabu, S. K., et al. (2019). Random forest for big data classification in the internet of things using optimal features.
International Journal of Machine Learning and Cybernetics, 10:2609-2618