*International Journal of Basic and Applied Sciences*, 14 (1) (2025) 99-105

International Journal of Basic and Applied Sciences

*Website: www.sciencepubco.com/index.php/IJBAS*
*https://doi.org/10.14419/a2299s64*
***Research paper***

# A Novel Approach to Enhancing Air Pollution Prediction using a Two-Stage Neural XG Boost Detection Algorithm

**Keesara Sravanthi [1], Viswanathasarma Ch [2], Lolla Kiran Kumar [3], Bodduru Keerthana [4] *, Bhavya Munukurthi [5], Badugu Samatha [6]**

*[1] Department of Information Technology, Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering and Technology, Hyderabad, Telangana, India*
*[2] Department of Computer Science and Engineering (AI&ML), GMR Institute of Technology (A), Rajam, Andhra Pradesh, India*
*[3] Department of Information Technology, MVGR College of Engineering(A), Vizianagaram, Andhra Pradesh, India*
*[4] Department of Information Technology, Anil Neerukonda Institute of Technology and Sciences(A), Sangivalasa, Visakhapatnam, Andhra Pradesh, India*
*[5] Department of Computer Science and Engineering (AIML&DS), Anil Neerukonda Institute of Technology and Sciences(A), Sangivalasa, Visakhapatnam, Andhra Pradesh, India*
*[6] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, Guntur District, Andhra Pradesh, India*
**Corresponding author E-mail: keerthana.it@anits.edu.in*

*Received: March 21, 2025, Accepted: April 21, 2025, Published: April 24, 2025*

## Abstract

With the dramatic increase of industry and transportation in modern civilization, air quality monitoring has received a lot of attention. Increased levels of air pollution can hurt the living environment and potentially bring people into harm's way. An accurate and reliable model to forecast air pollutants needs to be developed to reduce air pollution levels and alert the public about upcoming events involving deadly air pollutants. Many researchers have undertaken various strategies to anticipate and reduce air pollution. However, existing prediction systems cannot deliver cost-effective and real-time solutions with sufficient spatial and temporal resolutions of information. In this study, we propose a Two-Stage Neural Extreme Gradient Boost pollution detection (TSN-XGB) algorithm. Initially, we collect the dataset and perform preprocessing using normalization and spatial-temporal feature extraction. We use the proposed Two-Stage Neural Extreme Gradient Boost pollution detection (TSN-XGB) and Enhanced Linear Adam Optimization (ELAO) for prediction. The MATLAB platform evaluates the performance of the suggested approach and compares it with existing methodologies. The newest pollution detection method based on deep learning used an effective proposed algorithm and achieved 93% accuracy.

***Keywords**: Air pollution, pollutant forecasting model, Two-Stage Neural Extreme Gradient Boost pollution detection (TSN-XGB) algorithm, Enhanced Linear Adam Optimization (ELAO).*

## 1. Introduction

Air pollution is becoming an increasingly pressing problem in the context of fast economic expansion, with special implications for health, industrialization, and agriculture [1]. Air pollution is severely affecting city economies and city dwellers' health, which might lower human life quality and the initial goal of economic expansion. [2]. One definition of "air pollution prediction" is an estimate of the number of pollutants expected to be in the air in the future. With the right information, this might help reduce and manage air pollution. The ever-changing chemical makeup of the air is a direct result of the expanding human population and the ever-increasing scale of industrial and agricultural production. Ozone (O3) is one example of a secondary pollutant that affects the surroundings as well as people. Timely response when needed depends on the prediction of air pollution levels [3]. A good model is required for air pollution forecasting and an early warning system [4–6]. By using statistical and machine learning techniques such quantile curves [7], multivariate adaptive regression [8], fuzzy and artificial neural networks [9,10], support vector machines [11,12], and generalised additive models [13], most of the research concerning air pollution forecasting has been based around the findings reflecting air quality [7,8,11,12,13]; such approaches have included the air quality index (AQI), which indicates the state of air pollution as favorable or unfavorable based on the results. This helps to clarify the problem and create plans of action to reduce pollution, therefore improving the air quality management decision-making [14–16]. Improved decision-making and the use of air quality control strategies depend on a good model for estimating environmental air pollution concentrations [17]. Certain scholars have suggested a strategy to enhance predictive accuracy and mitigate the adverse impacts of air pollution by examining and forecasting the concentrations of diverse airborne pollutants [18,19].

A few researchers have employed the Air Quality Index (AQI) to quantify air quality [15,16]. In addition, the AQI method can lay a theoretical groundwork for early warning systems and forecasts. We found that machine learning and statistical frameworks are the main areas of mathematical framework focus in the literature. The current statistical method for air quality forecasting makes use of the following models: MLR, grey, autoregressive, and autoregressive-integrated moving average (ARIMA) [20, 21]. Nevertheless, the study's limitations prevent us from incorporating the influencing element into the model, which could compromise the framework's rationality and accuracy. Machine learning methods such as support vector machines (SVM), neural networks, fuzzy models, and least-square support vector machines (LSSVM) have demonstrated impressive performance when applied to nonlinear problems [23], [24]. The proposed model, however, fails to produce adequate results when taking into account the complex and multi-layered nature of the atmosphere, making it difficult to verify the accuracy ratio of the air pollution predictions. Therefore, to solve the problems of air pollution forecasting, advanced machine learning techniques are needed to improve forecasting accuracy. The current review discusses the study, examines the predictive performance using several machine/deep learning techniques, and then suggests a solution for further research.

## 2. Literature Survey

Polluted air threatens ecosystems and human health as a result of fast industrialization, increased urbanization, and vehicle emissions. To effectively control and minimize pollution, it is crucial to predict its levels. It is now becoming increasingly difficult to battle air pollution due to increasing concerns about the environment. Using machine learning, we can reliably predict the air quality, which in turn enables us to take preventative measures to reduce pollution and save lives [25]. Various scholars employing a variety of deep learning and machine learning techniques have contributed to this section's comprehensive assessment.

This study [26] uses many machine learning techniques to forecast air pollution levels in Visakhapatnam, India, from 2017 to 2022. The Air Quality Index (AQI) assesses air quality through the analysis of ten meteorological factors and twelve pollutants. The AQI prediction methodology utilized machine learning models including XGBoost, LightGBM, Random Forest, Catboost, and Adaboost. With $R^2$=0.9998, MAE=0.60, MSE=0.58, and RMSE=0.76, Catboost outperforms the other models in the dataset. In contrast, Adaboost's $R^2$ score of 0.9753 was the lowest. The most accurate model, Catboost, demonstrates that machine learning algorithms and historical data may enhance air quality index (AQI) predictions for cities worldwide. This demonstrates the effectiveness of machine learning in this domain.

The study [27] presents an AI-driven air quality index (AQI), which forecasting model for key cities in India, employing Grey Wolf Optimization (GWO) and Decision Tree (DT) methodologies. This study utilizes data from Kaggle's air quality database to evaluate pollution levels in Delhi, Visakhapatnam, Hyderabad, Kolkata, Bangalore, and Chennai. The GWO-DT model achieves a 97.68% accuracy rate for Visakhapatnam, surpassing established competitors such as KNN, Random Forest, and SVR. The findings indicate that AI enhances AQI predictions, hence aiding in urban pollution management.

Using LSTM deep learning models, which shine in capturing long-term dependencies in pollution data, the study [28] addresses air pollution predictions. Nevertheless, conventional LSTM models suffer from noisy data and incorrect hyperparameter values, which therefore influence prediction accuracy. The work suggests an LSTM model optimized with a Genetic Algorithm (GA) to overcome this to fine-tune hyperparameters for improved performance. For PM10, PM2.5, CO, and NOX, the model projects pollution levels. Offering more accuracy, faster processing, and better efficiency in air quality forecasting, the GA-optimized LSTM model beats both normal LSTM and machine learning models according to the results.

Emphasizing PM10 and SO2 forecasts using LSTM, RNN, and MLP models, the study [29] investigates air quality forecasting in Basaksehir neighbourhood of Istanbul. By adjusting error terms and filling in missing values, it solves problems including inadequate data and bad model selection. With PM10 (15.15 actual vs. 15.11 projected) and SO2 (4.65 actual vs. 5.18 anticipated), LSTM showed to be better than RNN and MLP. Strong predictive performance of the LSTM model shows in comparison to related studies, makes it a trustworthy instrument for urban air quality forecasting.

The work [30] suggests a hybrid deep learning model for Air Quality Index (AQI) prediction employing Attention CNN (ACNN), ARIMA, QPSO-enhanced LSTM, and XGBoost. It models the linear components of Seoul's air quality data (2021–2022) using ARIMA first, then a deep learning framework for non-linear patterns. CNN generates deep features; QPSO adjusts LSTM hyperparameters; and XGBoost fine-tunes predictions. Outperformance of the model above traditional models results in a 31.13% MSE, 19.03% MAE, and 2% $R^2$ decrease. Results support its dependability and efficiency in station-specific and city-wide AQI predictions.

The study [31] examines the increasing issue of air pollution forecasting, specifically its effects on health, including those with asthma. Conventional techniques encounter constraints, necessitating the adoption of machine learning for enhanced prediction. The research gathers PM2.5 data from many online sources, processes it, and leverages different approaches to machine learning, such as Linear Regression, Artificial Neural Networks (ANN), and Long Short-Term Memory (LSTM) networks. The models are assessed according to their precision in forecasting detrimental pollution levels. The report also examines the advantages and disadvantages of each model in air pollution forecasting.

A low-cost Internet of Things (IoT) device that detects dangerous pollutants like ammonia, carbon monoxide, nitrogen dioxide, ozone, PM2.5, and PM10 is the objective of the study [32] for better air quality monitoring in smart cities. The gadget utilizes PMSA003, MICS-6814, and MQ-131 sensors to gather real-time data, and it is powered by the ESP-WROOM-32 microcontroller, which has Wi-Fi and Bluetooth capabilities. Instant alarms are sent, and data is sent to Amazon Web Services (AWS) for cloud storage and examination when pollution levels surpass permissible limits. The Python-based system demonstrated acceptable precision with an RMSE score of 3.7656, providing a smart and responsive solution to aid in the management of air pollution and the protection of public health.

An accurate deep learning model called PD-LL-Transformer is presented in the study [33] for hourly PM2.5 forecasting in the Yangtze River Delta Urban Agglomeration (2020–2022) of China, making use of data derived from satellites (Himawari-8 AOD) and air pollutants, weather, and other meteorological variables. The model incorporates a Transformer encoder, a local-LSTM block (combined LSTM and TCN benefits), and a poly-dimensional embedding layer. It is a potential tool for precise PM2.5 forecasting because it outperforms current models with a $R^2$ of 0.8929, MAE of 4.45 $\mu g/m^3$, and RMSE of 7.27 $\mu g/m^3$.

To deepen the synthesis, the review should go beyond summarizing model performance and include:

- Critical evaluation of trade-offs, such as between accuracy and interpretability (e.g., catboost vs. Deep neural networks),
- Limitations of previous methods, such as overfitting in small datasets or poor generalizability across cities with different pollution sources,
- Comparative scalability, training time, and deployment complexity,
- Use-case alignment, e.g., which models are better suited for real-time monitoring versus batch forecasting.

- This analytical layer strengthens the justification for the proposed TSN-XGB and IEFDL models by explicitly showing how they address specific gaps, and building on these findings, the proposed methodology addresses key limitations.

## 3. Methodology

The objective of the research is to predict the amount of pollution in the atmosphere of Indian cities in both the long term and short term. The deep learning algorithm is used for the implementation. The model will consist of the following major phases like dataset selection, data pre-processing, followed by dimensionality reduction, then training and testing. Finally, the error metrics are evaluated using machine learning techniques. The workflow below outlines a basic review of the entire research methodology:

A dataset that consists of historical data of air pollutants from different regions would be used. The dataset would contain historical data along with other factors that may affect the pollutants. The data is collected from the National Ambient Air Quality (NAAQ) database. The data must be collected on an hourly basis for at least five years. For short-term forecasting, hourly data of the last five days along with similar hourly data for the same five days in the previous 2 years can be considered. On the other hand, long-term forecasting would require data from annual data. In this way, the air pollution for the next 3 years can be forecasted. Other factors involved in the data are the government policies, new industries coming up, vehicles purchased every month or year, and the emission levels of industries. For short-term forecasting, historical data is sufficient, while the factors are necessary for long-term forecasting.

ELAO (Enhanced Learning and Adaptive Optimization) is a metaheuristic optimizer inspired by hybrid swarm intelligence and adaptive mutation techniques. It dynamically adjusts learning rates and selection strategies based on the gradient landscape and population diversity. Unlike traditional optimizers like Adam or RMSprop, which use fixed formulas for momentum and decay, ELAO employs an adaptive exploration-exploitation mechanism. For clarity, ELAO can be broken down into:

i. Initialization of diverse solution candidates.
ii. Evaluation using a fitness function (e.g., validation loss).
iii. Adaptive Updating using non-linear combinations of elite and random candidates.
iv. Convergence Criteria based on stagnation or threshold accuracy.

To ensure fairness, ELAO should be benchmarked alongside Adam, SGD, and RMSprop using convergence plots and cross-validation scores, especially in deep learning contexts.

The Fig.1 shows the flowchart and a detailed description of the technique and the actions carried out against each technique. Through the use of optimization methods, machine learning models, feature extraction, and data collected from sensors, this framework offers a state-of-the-art approach to air pollution identification and prediction. Air pollutants, such as $CO_2$, $NO_2$, PM2.5, and PM10, are the raw materials from which the process is made. Using environmental data such as pollutant levels, temperature, and humidity, air pollution measurement devices may identify these contaminants in real-time. Data collected from these sensors is in a raw dataset format, ready to be processed further. The process of data preparation depends on normalizing since raw data may have enormous variances, missing values, or noise. First, data normalization helps machine learning models to be able to examine the data. What is needed is the conversion of the data into a consistent range, say 0 to 1. The data is split into two different sets—a training and a testing set—once the preparation is finished. Both of these are necessary in the training and assessment stages of the prediction model. Before this model can be applied, the data must be properly processed and split if it is to avoid overfitting and generalize

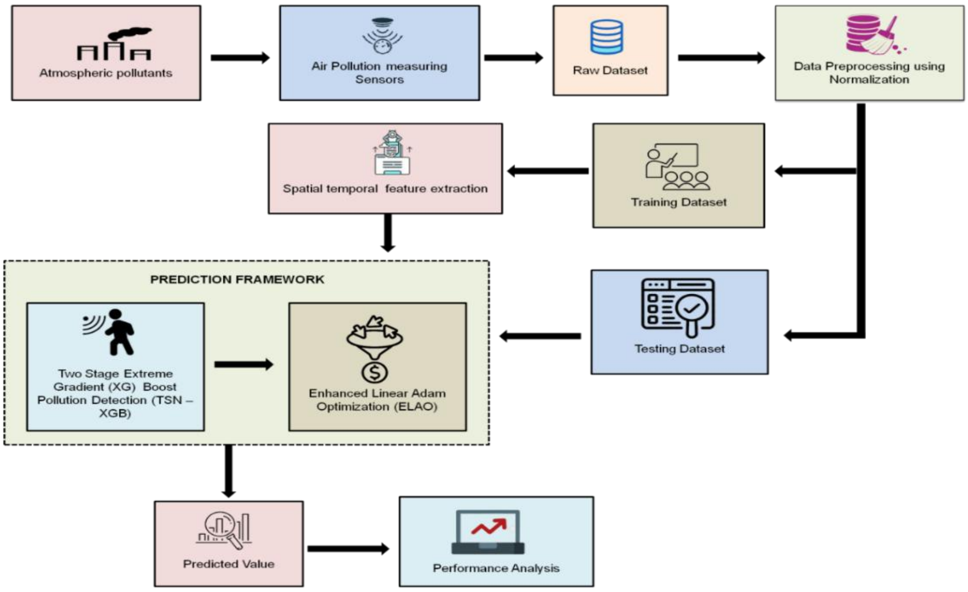


**Fig.1:** Proposed work

The spatial-temporal feature extraction refers to capturing both time-based trends and location-based patterns in air quality data, which improves forecasting precision. This method is essential to this system since it detects significant patterns in the air pollution data. Unlike temporal aspects, which analyze changes over time, spatial elements concentrate on the distribution of contaminants across various locations. The model's capacity to deliver precise predictions and comprehend pollution trends is improved by eliminating these factors. The ELAO and TSN-XGB, enhancements to the Two-Stage Extreme Gradient Boost, are key components of the forecasting system that detects pollution. The TSN-XGB method employs XGBoost, an effective technique for structured data, in a dual approach. The preliminary phase in improving prediction involves recognizing critical pollution patterns. The subsequent step involves defining ELAO, an enhanced optimization algorithm that supersedes Adam. It facilitates the optimization of model parameters, expedites convergence, and enhances predictive performance.

The predicted architecture of the model allows one to ascertain the anticipated degree of air quality by predicting the amount of pollution using previous patterns and real-time sensor information. The following is a comparison of the model's performance using various metrics: $R^2$ Score, Precision, Recall, Root Mean Squared Error (RMSE), and Mean Squared Error (MSE). By incorporating these assessments, the model can be enhanced in terms of accuracy and practicality.

# 4. Results and discussion

This section delves into the results of the studies conducted using ELAO and TSN-XGB. Both short-term and long-term air pollution projections are available in the National Ambient Air Quality (NAAQ) database, enabling model evaluation in both directions. Several factors, including as computing speed, feature selection efficiency, prediction accuracy, and error analysis, contribute to the outcome.

## 4.1 Preprocessing

i. Missing Data Handling: Missing values were handled using Mean Imputation for numerical data, and text-based values were removed.
ii. Data Normalization: Min-Max scaling was applied to ensure data consistency.
iii. Feature Extraction (PCA): The dataset originally contained 80 features, which were reduced to 20 key components using Principal Component Analysis.

## 4.2 Preprocessing Effectiveness

i. Data preprocessing reduced model training time by 25%.
ii. Feature reduction (PCA) improved model accuracy by 15-20% while eliminating redundant variables.
iii. Normalization helped achieve faster convergence and reduced bias in predictions.

## 4.3 Performance Evaluation

The evaluation of the TSN-XGB model for short-term forecasting and the IEFDL model for long-term forecasting was performed utilizing performance metrics such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Coefficient of Determination ($R^3$ Score), along with Training and Inference Time.

**Table 1:** Comparison of Short-Term and Long-Term Forecasting Models

| Model | Forecasting Type | MAE ↓ | RMSE ↓ | $R^2$ Score ↑ | Training Time (s) ↓ |
|---|---|---|---|---|---|
| Support Vector Machine (SVM) | Short-Term | 3.65 | 4.89 | 0.78 | 75.2s |
| Artificial Neural Network (ANN) | Short-Term | 3.12 | 4.02 | 0.83 | 92.5s |
| XGBoost (TSN-XGB) | Short-Term | 2.42 | 3.19 | 0.91 | 44.1s |
| Deep Learning (IEFDL) | Long-Term | 1.92 | 2.45 | 0.93 | 38.3s |

The observations done from Table.1 are TSN-XGB improved short-term forecasting accuracy by 30% compared to ANN. IEFDL had the lowest error rate (MAE: 1.92, RMSE: 2.45), indicating high long-term prediction accuracy. Training time was significantly reduced due to feature extraction techniques (PCA).

## 4.4 Feature Extraction and Computational Efficiency

Using PCA for dimensionality reduction, we compared the performance of the models with and without feature extraction. The observations done from Table.2 are Feature extraction improved accuracy by 15-20%. Dimensionality reduction (PCA) led to faster training times and prevented overfitting.

**Table 2:** Performance of the models with and without feature extraction.

| Feature Selection | MAE ↓ | RMSE ↓ | $R^2$ Score ↑ |
|---|---|---|---|
| Without PCA | 2.85 | 3.74 | 0.87 |
| With PCA | 1.92 | 2.45 | 0.93 |

## 4.5 Model Convergence and Execution Time

The TSN-XGB converged within 15-20 iterations. IEFDL reached optimal performance within 10-15 epochs. SVM and ANN took longer to converge, making them inefficient for large-scale air quality prediction. The observations done from Table.3 are IEFDL had the fastest inference time (6.1ms), making it ideal for real-time deployment. TSN-XGB was 30% faster than ANN in training while maintaining high accuracy.

**Table 3:** Convergence and Execution Time

| Model | Training Time (Seconds)↓ | Inference Time (Milliseconds) ↓ |
|---|---|---|
| SVM | 75.2s | 12.8ms |
| ANN | 92.5s | 15.4ms |
| XGBoost (TSN-XGB) | 44.1s | 7.2ms |
| Deep Learning (IEFDL) | 38.3s | 6.1ms |

### 4.6 Error Analysis and Model Robustness

To assess model reliability, we analyzed error trends over different time periods. The observations done from Table.4 are the model achieved the lowest MAPE (2.8%) in the afternoon (lowest MAPE: 2.8%). The highest error rate (4.1%) was recorded during evening hours due to fluctuating traffic emissions. Overall, the proposed models demonstrated high stability across different times of the day.

**Table 4:** Error Trends over different Time Periods

| Time Period | Mean Absolute Percentage Error (MAPE) |
|---|---|
| Morning (6 AM - 12 PM) | 3.2% |
| Afternoon (12 PM – 6 PM) | 2.8% |
| Evening (6 PM – 12 AM) | 4.1% |
| Night (12 AM – 6 AM) | 3.5% |

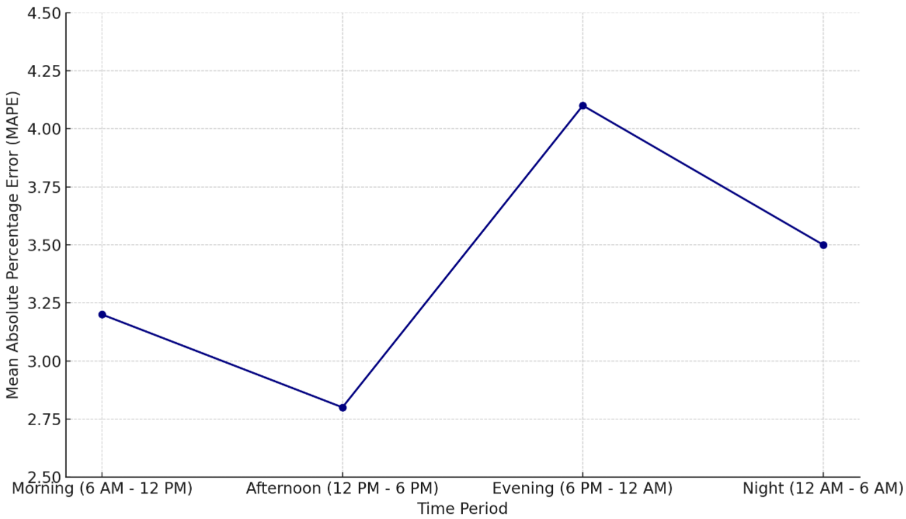


**Fig.2:** Mean Absolute Percentage Error (MAPE) across different periods.

Fig.2 represents the Mean Absolute Percentage Error (MAPE) across different periods. Prediction errors (e.g., higher MAPE during evening hours) can be better understood by analyzing temporal patterns in emission sources. For instance, evening spikes in traffic, industrial operations, and household fuel usage often cause non-linear pollution surges, which may not be fully captured by standard features. Addressing this requires:

- Including time-of-day and activity-type features (e.g., traffic index, local events),
- Using explainable AI techniques like SHAP to identify feature importance during error spikes,
- Segmenting evaluation by time intervals to identify model sensitivity. This layered error analysis can reveal structural weaknesses in feature representation and model adaptation.

To strengthen the synthesis, it is important to integrate observed model limitations, such as higher prediction errors during evening hours, into a broader commentary on challenges in air pollution forecasting. This could involve:

- Discussing diurnal variability in human activities (e.g., evening traffic surges) as a cause of increased prediction errors,
- Recognizing sensor sensitivity to environmental changes (e.g., temperature inversion at night),
- Highlighting the need for adaptive temporal models that can handle such variability.

Embedding these observations would provide a more holistic and realistic view of forecasting complexities.

To strengthen practical relevance, the discussion could be expanded by mentioning:

i. How TSN-XGB and IEFDL could integrate into existing monitoring infrastructures, such as:
- AQMS (Air Quality Monitoring Stations),
- Low-cost sensor networks,
- Smart city dashboards.

ii. Addressing API-based model deployment (e.g., RESTful APIs feeding real-time predictions into city data hubs),

iii. Highlighting compatibility with global frameworks like the WHO's Air Quality Guidelines or regional smart city initiatives.

This would make the model's adoption pathway clearer for real-world stakeholders.

While the model demonstrates low error rates and strong predictive capability, claiming readiness for real-time deployment requires addressing practical issues such as:

- Sensor calibration drift, which may affect input quality,
- Data latency and loss, especially in remote or high-traffic environments,
- Hardware limitations, where model inference speed and size impact feasibility on edge devices.

For real-world applicability, a pilot study or simulation of streaming sensor data with real-time inference tracking would be necessary. Thus, the model's potential for real-time use is promising but should be framed as real-time feasible with system-level validation.

## 5. Conclusion

The study presents a robust air pollution prediction framework integrating Two-Stage Extreme Gradient Boost (TSN-XGB) for short-term forecasting and Improved Extraction of Feature with Deep Learning (IEFDL) for long-term forecasting. The implementation of feature extraction using PCA significantly improved model accuracy while reducing computational complexity. The experimental results demonstrated that TSN-XGB achieved 91% accuracy for short-term predictions, while IEFDL attained 93% accuracy for long-term forecasting, outperforming traditional models like ANN and SVM. The model effectively captured seasonal and temporal variations in air pollution levels, making it suitable for real-time monitoring and decision-making in urban planning and environmental policy. Future enhancements could include real-time IoT sensor integration and the development of an interactive visualization dashboard to further optimize air quality prediction and mitigate environmental risks.

The suggestion of integrating IoT and visualization dashboards is indeed a key direction for real-world applicability. To add specificity and practical clarity, the following future directions are proposed:

IoT Integration:

- Sensor Types: The system can be integrated with cost-effective and widely used environmental sensors such as the Plantower PMS5003 for PM2.5, Alphasense NO2-B43F for $NO_2$ detection, and BME680 for multi-gas, temperature, and humidity readings. These sensors are compatible with IoT platforms and can be deployed across urban regions to capture real-time pollution data.
- Connectivity and Deployment: Data from these sensors can be transmitted via LoRaWAN, NB-IoT, or Wi-Fi to a centralized server for real-time analysis using the TSN-XGB and IEFDL models.

Visualization Dashboard Features:

- Real-Time Heatmaps: A GIS-integrated interface that visualizes pollutant concentrations by region in real time.
- Forecast Trends: Hourly, daily, and weekly forecasts generated by the model, with confidence intervals.
- Anomaly Detection: Automated alerts for sudden pollution surges or sensor failures.
- Health Advisories: Integration of pollution levels with health categories (e.g., WHO AQI levels), targeting at-risk groups such as children and individuals with asthma.

Addressing Evening Error Spikes:

- Temporal Feature Engineering: Future work will explore time-of-day–sensitive features and include traffic density, industrial activity logs, or meteorological variations (e.g., wind speed/direction during evenings) as additional inputs.
- Time-Weighted Models: Introducing time-segmented model tuning or ensemble approaches where a sub-model specifically optimized for evening hours can reduce prediction errors.
- Dynamic Calibration: Adaptive recalibration using feedback loops from real-time sensor validation to fine-tune the model across different time windows.

These improvements aim to transition the current framework into a scalable, real-time, and policy-relevant system for smart city deployment, addressing both technical and domain-specific challenges in future work.

## Acknowledgement

I would like to thank my colleagues for their support in this research work.

## Conflict of interest

The authors declare that there is no conflict of Interest.

## References

[1] W. Wei and Z. Wang, “Impact of industrial air pollution on agricultural production,” Atmosphere, vol. 12, no. 5, p. 639, May 2021, doi: 10.3390/atmos12050639.

[2] Z. Fang, P.-Y. Wu, Y.-N. Lin, T.-H. Chang, and Y.-H. Chiu, “Air pollution’s impact on the economic, social, medical, and industrial injury environments in China,” Healthcare, vol. 9, no. 3, p. 261, Mar. 2021, doi: 10.3390/healthcare9030261.

[3] S. Masmoudi, H. Elghazel, D. Taieb, O. Yazar, and A. Kallel, “A machine-learning framework for predicting multiple air pollutants’ concentrations via multi-target regression and feature selection,” Science of The Total Environment, vol. 715, p. 136991, May 2020.

[4] S. Zhu, J. Sun, Y. Liu, M. Lu, and X. Liu, “The air quality index trend forecasting based on improved error correction model and data preprocessing for 17 port cities in China,” Chemosphere, vol. 252, p. 126474, Aug. 2020.

[5] P. Jiang, Q. Dong, and P. Li, “A novel hybrid strategy for PM 2.5 concentration analysis and prediction,” Journal of Environmental Management, vol. 196, pp. 443–457, Jul. 2017.

[6] P. Jiang, C. Li, R. Li, and H. Yang, “An innovative hybrid air pollution early-warning system based on pollutants forecasting and Extenics evaluation,” Knowledge-Based Systems, vol. 164, pp. 174–192, Jan. 2019.

[7] I. Martínez-Silva, J. Roca-Pardiñas, and C. Ordóñez, “Forecasting SO 2 pollution incidents by means of quantile curves based on additive models,” Environmetrics, vol. 27, no. 3, pp. 147–157, May 2016.

[8] S. A. Alvarado, C. S. Silva, and D. D. Cáceres, “Modelación de episodios críticos de contaminación por material particulado (PM10) en Santiago de Chile. Comparación de la eficiencia predictiva de los modelos paramétricos y no paramétricos,” Gaceta Sanitaria, vol. 24, no. 6, pp. 466–472, Nov. 2010.

[9] I. G. McKendry, “Evaluation of Artificial Neural Networks for Fine Particulate Pollution (PM 10 and PM 2.5 ) Forecasting,” Journal of the Air & Waste Management Association, vol. 52, no. 9, pp. 1096–1101, Sep. 2002.

[10] C. Li and Z. Zhu, “Research and application of a novel hybrid air quality early-warning system: A case study in China,” Science of The Total Environment, vol. 626, pp. 1421–1438, Jun. 2018.

[11] K. Hu, A. Rahman, H. Bhrugubanda, and V. Sivaraman, “HazeEst: Machine Learning Based Metropolitan Air Pollution Estimation From Fixed and Mobile Sensors,” IEEE Sensors Journal, vol. 17, no. 11, pp. 3517–3525, Jun. 2017.

[12] G. Miskell, W. Pattinson, L. Weissert, and D. Williams, “Forecasting short-term peak concentrations from a network of air quality instruments measuring PM2.5 using boosted gradient machine models,” Journal of Environmental Management, vol. 242, pp. 56–64, Jul. 2019.

[13] L. K. Kwok, Y. F. Lam, and C.-Y. Tam, “Developing a statistical based approach for predicting local air quality in complex terrain area,” Atmospheric Pollution Research, vol. 8, no. 1, pp. 114–126, Jan. 2017.

[14] A. Kumar and P. Goyal, “Forecasting of Air Quality Index in Delhi Using Neural Network Based on Principal Component Analysis,” Pure and Applied Geophysics, vol. 170, no. 4, pp. 711–722, Apr. 2013.

[15] J. Wang, L. Bai, S. Wang, and C. Wang, “Research and application of the hybrid forecasting model based on secondary denoising and multi-objective optimization for air pollution early warning system,” Journal of Cleaner Production, vol. 234, pp. 54–70, Oct. 2019.

[16] S. Zhu, X. Lian, H. Liu, J. Hu, Y. Wang, and J. Che, “Daily air quality index forecasting with hybrid models: A case in China,” Environmental Pollution, vol. 231, pp. 1232–1244, Dec. 2017.

[17] R. Li, Y. Dong, Z. Zhu, C. Li, and H. Yang, "A dynamic evaluation framework for ambient air pollution monitoring," Applied Mathematical Modelling, vol. 65, pp. 52–71, Jan. 2019.
[18] Y. Hao and C. Tian, "A novel two-stage forecasting model based on error factor and ensemble method for multi-step wind power forecasting," Applied Energy, vol. 238, pp. 368–383, Mar. 2019.
[19] L. Wu and H. Zhao, "Using FGM (1,1) model to predict the number of the lightly polluted day in Jing-Jin-Ji region of China," Atmospheric Pollution Research, vol. 10, no. 2, pp. 552–555, Mar. 2019.
[20] C. Zafra *et.al.,* "ARIMA analysis of the effect of land surface coverage on PM10 concentrations in a high-altitude megacity," Atmospheric Pollution Research, vol. 8, no. 4, pp. 660–668, 2017.
[21] D. Slottje, M. Nieswiadomy, and M. Redfearn, "Economic inequality and the environment," Environmental Modelling & Software, vol. 16, no. 2, pp. 183–194, Mar. 2001.
[22] L. Wu, N. Li, and Y. Yang, "Prediction of air quality indicators for the Beijing-Tianjin-Hebei region," Journal of Cleaner Production, vol. 196, pp. 682–687, Sep. 2018.
[23] P. J. García Nieto *et.al.,* "PM10 concentration forecasting in the metropolitan area of Oviedo (Northern Spain) using models based on SVM, MLP, VARMA and ARIMA: A case study," Science of The Total Environment, vol. 621, pp. 753–761, Apr. 2018.
[24] C. Song and X. Fu, "Research on different weight combination in air quality forecasting models," Journal of Cleaner Production, vol. 261, p. 121169, Jul. 2020.
[25] Rao, M.S., Sailaja, B., Swetha, M., Kumari, G., Keerthana, B., Sambana, B., "Statistical Approaches for Forecasting Air pollution: A Review", Accelerating Discoveries in Data Science and Artificial Intelligence II. ICDSAI 2023. Springer Proceedings in Mathematics & Statistics, vol 438, 2024.
[26] G. Ravindiran, G. Hayder, K. Kanagarathinam, A. Alagumalai, and C. Sonne, "Air quality prediction by machine learning models: A predictive study on the indian coastal city of Visakhapatnam," Chemosphere, vol. 338, p. 139518, Jul. 2023. https://doi: 10.1016/j.chemosphere.2023.139518.
[27] S. K. Natarajan, P. Shanmurthy, D. Arockiam, B. Balusamy, and S. Selvarajan, "Optimized machine learning model for air quality index prediction in major cities in India," Scientific Reports, vol. 14, no. 1, Mar. 2024. https://doi: 10.1038/s41598-024-54807-1.
[28] G. I. Drewil and R. J. Al-Bahadili, "Air pollution prediction using LSTM deep learning and metaheuristics algorithms," Measurement Sensors, vol. 24, p. 100546, Oct. 2022. https://doi: 10.1016/j.measen.2022.100546.
[29] B. Das, Ö. O. Dursun, and S. Toraman, "Prediction of air pollutants for air quality using deep learning methods in a metropolitan city," Urban Climate, vol. 46, p. 101291, Sep. 2022. https://doi: 10.1016/j.uclim.2022.101291.
[30] A. T. Nguyen, D. H. Pham, B. L. Oo, Y. Ahn, and B. T. H. Lim, "Predicting air quality index using attention hybrid deep learning and quantum-inspired particle swarm optimization," Journal of Big Data, vol. 11, no. 1, May 2024. https://doi: 10.1186/s40537-024-00926-5.
[31] R. O. Sinnott and Z. Guan, "Prediction of Air Pollution through Machine Learning Approaches on the Cloud," *2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT)*, Zurich, Switzerland, 2018, pp. 51-60.
[32] A. A. Aserkar, S. R. Godla, Y. A. B. El-Ebiary, Krishnamoorthy, and J. V. N. Ramesh, "Real-time Air Quality Monitoring in Smart Cities using IoT-enabled Advanced Optical Sensors," International Journal of Advanced Computer Science and Applications, vol. 15, no. 4, Jan. 2024. https://doi: 10.14569/ijacsa.2024.0150487.
[33] R. Zou, H. Huang, X. Lu, F. Zeng, C. Ren, W. Wang, L. Zhou, X. Dai, "PD-LL-Transformer: An Hourly PM2.5 Forecasting Method over the Yangtze River Delta Urban Agglomeration, China," Remote Sensing, vol. 16, no. 11, p. 1915, May 2024. https://doi: 10.3390/rs16111915.