isrjournals

International Journal of Advanced Research in Computer Science Engineering and Information Technology

Volume: 6, Issue: 3,Special Issue: 3 ,Apr,2025 ,ISSN_NO: 2321-3337

AI-DRIVEN PREDICTION OF ATMOSPHERIC PARTICULATE MATTER FOR AIR QUALITY ASSESSMENT

Aswini.M[1], Mohanapriya. S[2], Teja.M[3], Priya.V[4*]

[123]UG Scholar –Dept of CSE, GRT Institute of Engineering and Technology, Tiruttani, India.

[4*]Assistant Professor- Dept of CSE, GRT Institute of Engineering and Technology, Tiruttani, India.

aswini30012004@gmail.com,mohanapriyas144@gmail.com,tejamm52@gmail.com

[4*]Corresponding Author: priya.v@grt.edu.in

Abstract - Air pollution poses a significant threat to public health and environmental sustainability, necessitating accurate forecasting for effective mitigation. Traditional air quality monitoring systems in India rely on stationary sensors and simplistic models, which often fail to provide precise predictions due to their inability to capture complex temporal dependencies in pollutant levels. To overcome these limitations, this study proposes an advanced air pollutant detection system leveraging the Bidirectional Long Short-Term Memory (BI-LSTM) algorithm. The proposed system integrates real-time and historical air quality data obtained from government monitoring stations, satellites, and public databases. Data preprocessing techniques, including missing value handling, normalization, and feature selection, ensure high data integrity. Additionally, temporal feature extraction methods, such as seasonal decomposition and time-lagged analysis, improve model performance by identifying long-term trends and seasonal variations in pollutant levels. The BI-LSTM model is trained using optimized hyperparameters to enhance prediction accuracy and minimize errors. A comparative analysis is conducted against conventional prediction models, including Support Vector Machines (SVM), Recurrent Neural Networks (RNN), and standard Long Short-Term Memory (LSTM) networks. The BI-LSTM model is trained using optimized hyperparameters to enhance prediction accuracy and minimize errors. A comparative analysis is conducted against conventional prediction models, including Support Vector Machines (SVM), Recurrent Neural Networks (RNN), and standard Long Short-Term Memory (LSTM) networks.

Keywords: Time series forecasting, air quality monitoring, machine learning, BI-LSTM, deep learning, air pollution detection

1. INTRODUCTION

Air pollution is a significant environmental challenge in the modern era, directly affecting human health and ecosystem stability [3]. The rapid growth of industrialization, urbanization, and vehicular emissions has led to a severe decline in air quality, contributing to respiratory diseases, cardiovascular disorders, and other serious health conditions. According to global environmental reports, major air pollutants such as sulfur dioxide (SO_2), nitrogen dioxide (NO_2), particulate matter (PM2.5 and PM10), ozone (O_3), and carbon monoxide (CO) pose significant health risks [3]. The World Health Organization (WHO) estimates that millions of premature deaths occur annually due to long-term exposure to these pollutants. Therefore, accurate air quality forecasting is essential for policymakers, and the public to implement proactive measures and minimize exposure to harmful pollutants.

Recent advancements in machine learning have demonstrated the potential for improving air quality forecasting by leveraging large-scale environmental datasets and extracting intricate patterns from pollution trends [4]. Techniques such as Support Vector Regression (SVR), Decision Trees, and Neural Networks have been widely applied, but they often suffer from slow convergence, overfitting, and poor generalization across different locations and seasons [7], [8]. To address these challenges, this study proposes a **Bidirectional Long Short-Term Memory (BI-LSTM) model** for air quality prediction. Unlike conventional models, BI-LSTM is highly effective in capturing long-term dependencies and sequential variations in pollutant levels, leading to more accurate and reliable forecasts [7]. The system integrates both real-time and historical air quality data from multiple monitoring stations and applies advanced preprocessing techniques, such as feature selection, normalization, and outlier detection, to improve data quality [1]

2. RELATED WORK

Air quality prediction has been extensively researched in recent years [13]. Early approaches predominantly utilized statistical and regression-based models to analyze pollutant levels and forecast the Air Quality Index (AQI) [14], [15]. For instance, Agarwal and Sahu [14] employed traditional statistical methods for air quality forecasting, while Lary et al. combined remote sensing data with ground-based PM measurements to improve spatial coverage and prediction accuracy [16]. However, these classical methods often struggle to capture the nonlinearity and dynamic variability of atmospheric processes

[17]. With advancements in machine learning, researchers have shifted towards more sophisticated techniques to model the complex spatiotemporal dependencies of air pollution [7]. Zheng et al. introduced a hybrid method that combines linear regression with Graph neural networks (GNN) for enhanced short-term prediction, [2] while leveraging multiple regression models incorporating meteorological factors for AQI estimation [13]. Despite these improvements, such approaches still exhibit limitations in adapting to rapid changes in pollutant concentrations [5]. Furthermore, the inherent assumption of linearity in some traditional models hinders their capacity to capture sudden pollution spikes and complex environmental interactions [17]. However, deep learning approaches are not without their challenges, including extensive computational requirements and a tendency to overfit on limited datasets [7],[10]. Furthermore, the study explores the potential of incorporating external data sources, such as satellite imagery and traffic patterns, to further enhance predictive power [18],[19]

3. METHODOLOGY

In this section, we introduce the AQI framework, followed by the detailed explanation of the deep learning methodologies implemented for air quality prediction. We then propose an optimized Bi-LSTM model enhanced with feature selection techniques to improve predictive accuracy and computational efficiency.

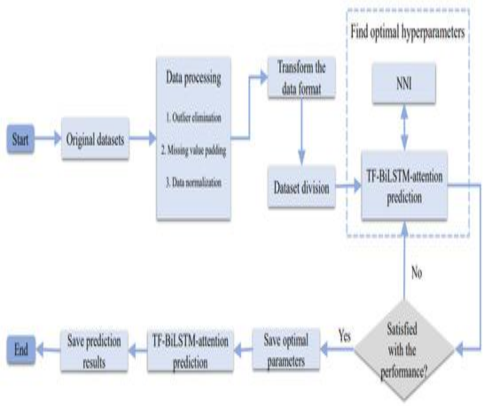

Fig.3. Methodology framework of this study.

Air quality forecasting is a crucial aspect of environmental management and public health safety. The primary objective is to predict AQI values at different observation points over a specified period. Traditional forecasting methods often rely on statistical and simple machine learning models, but these approaches struggle to capture the nonlinear relationships between pollutants and meteorological factors [14], [20].

the accurate prediction of air quality remains a significant challenge. AQI prediction is influenced by meteorological variables such as temperature, humidity, wind speed, and pollutant concentrations. Thus, extracting meaningful patterns from these factors is essential for achieving precise forecasts.

Where *AQI* represents the individual AQI for each pollutant. By analysing multivariate time-series data, we aim to enhance prediction accuracy using deep learning approaches that effectively model spatiotemporal dependencies [7], [13].

3.1 BIDIRECTIONAL LONG SHORT-TERM MEMORY (BI-LSTM)

LSTM networks are widely used for sequential data modelling due to their ability to retain long-term dependencies. The Bi-LSTM model further extends this capability by processing information in both forward and backward directions, thereby capturing complex temporal relationships more effectively [11].

Given a sequence of input features, ***X= (x1, x2, ….x T),***

the hidden states of forward and backward LSTs at time are defined as:

$$\overrightarrow{ht}=fLSTM(xt, \overrightarrow{ht-1}) \qquad (1)$$

$$\overleftarrow{h\,t}=fLSTM\,(xt,\overleftarrow{ht+1}) \qquad (2)$$

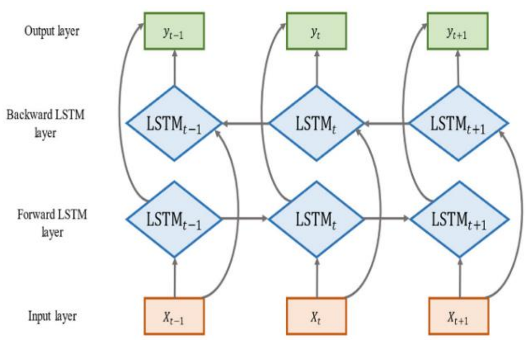

Fig.3.1. Structure of Bi-LSTM.

where and denote the hidden states of the forward and backward LSTM layers, respectively.
The final output is computed as:

$$h\,t = W\,[\overrightarrow{h\,t};\ \overleftarrow{h\,t}] + b \qquad (3)$$

This dual processing mechanism enables Bi-LSTM to effectively model both past and future contextual information, making it highly suitable for AQI prediction tasks.

isrjournals

International Journal of Advanced Research in Computer Science Engineering and Information Technology

Volume: 6, Issue: 3,Special Issue: 3 ,Apr,2025 ,ISSN_NO: 2321-3337

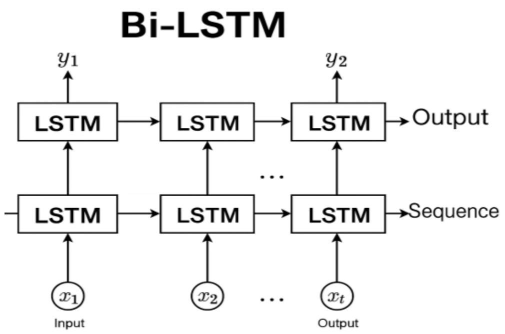

Fig.3.1. Schematic diagram of the Bi-LSTM network.

3.2 ATTENTION MECHANISM FOR FEATURE SELECTION

To enhance the predictive performance of Bi-LSTM, we integrate an attention mechanism that selectively emphasizes relevant input features while reducing the influence of redundant information. The attention weight for each feature is calculated as:

$$\alpha t = \exp(W\, h\, \alpha t) / \textstyle\sum t' \exp(W\, h\, \alpha t') \qquad (4)$$

where ***W a*** is a learnable weight matrix, and ***αt*** represents the attention score assigned to each time step. The refined feature representation is obtained as:

$$c\, t = \textstyle\sum t\, \alpha t\, h\, t \qquad (5)$$

This approach enables the model to focus on the most informative patterns [7], [18].

3.3 GA-OPTIMIZED BI-LSTM MODEL

To further enhance prediction accuracy, we employ a Genetic Algorithm (GA) for hyperparameter optimization. The GA iteratively adjusts the learning rate, batch size, and hidden layer configurations to identify the optimal Bi-LSTM structure [1], [5].

The key steps in the GA optimization process are:

Initialization: Randomly generate an initial population of Bi-LSTM configurations.

Fitness Evaluation: Train each candidate model and evaluate its performance using Mean Squared Error (MSE) as the fitness function.

Selection: Select the top-performing models based on their fitness scores.

Crossover & Mutation: Generate new candidate solutions by combining and modifying existing ones.

Termination: Repeat steps 2-4 until convergence criteria are met.

By optimizing hyperparameters using GA, we achieve an improved balance between model complexity and prediction accuracy, ultimately leading to better AQI forecasts.

3.3 OVERVIEW OF THE PREDICTION FRAMEWORK

The complete Bi-LSTM-based AQI prediction framework consists of the following stages:

Data Collection & Preprocessing: Raw environmental data is collected from air quality monitoring stations, normalized, and formatted for model input.

Feature Selection: An attention mechanism is applied to extract the most relevant features.

Model Training: The Bi-LSTM network is trained using historical AQI data.

Optimization: GA is applied to fine-tune model parameters and improve generalization.

Prediction & Evaluation: The trained model is tested on unseen data to assess its performance.

Visualization&Interpretation: After prediction, the AQI results are visualized through charts and dashboards to facilitate better understanding for users, researchers, and policymakers. These visual tools help interpret trends and anomalies in air quality over time, supporting informed decisions and timely interventions.

By integrating deep learning with feature selection and optimization techniques, our approach significantly enhances air quality prediction capabilities, making it a valuable tool for environmental monitoring and decision-making [10], [19].

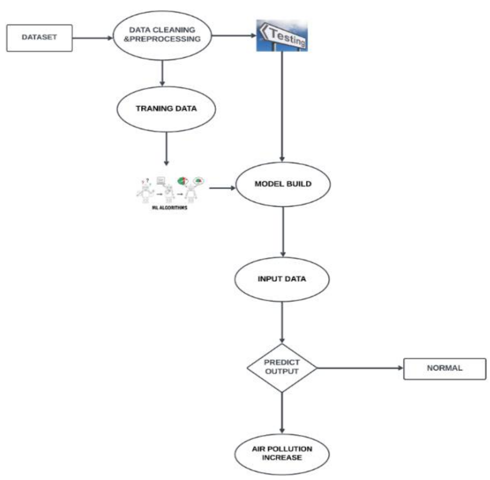

Fig.3.4. Model architecture diagram.

isrjournals

International Journal of Advanced Research in Computer Science Engineering and Information Technology

Volume: 6, Issue: 3,Special Issue: 3 ,Apr,2025 ,ISSN_NO: 2321-3337

4. PROPOSED SYSTEM

In the proposed system, a hybrid model integrating Bidirectional Long Short-Term Memory (Bi-LSTM) networks, an attention mechanism, and Genetic Algorithm (GA) optimization is developed for Air Quality Index (AQI) prediction. This approach is designed to tackle the challenges associated with forecasting AQI, which involve processing multivariate, noisy, and time-dependent environmental data. Initially, the system preprocesses raw AQI datasets collected from various air quality monitoring stations by cleaning, normalizing, and converting them into time-series sequences. To enhance model learning and reduce redundancy, an attention mechanism is employed to prioritize the most informative features, thereby filtering out irrelevant or noisy data. Following this, a Bi-LSTM model is utilized to capture both forward and backward temporal dependencies in the AQI data, significantly improving the model's ability to understand complex time-series relationships. Furthermore, the model's performance is fine-tuned by employing a Genetic Algorithm to optimize hyperparameters such as learning rate, batch size, and the number of hidden layers. After optimization, the final model is trained and evaluated using common performance metrics like Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) to validate its predictive accuracy. This proposed system presents several advantages. It offers high prediction accuracy due to the Bi-LSTM's ability to learn from bidirectional sequences, while the attention mechanism further refines predictions by focusing on critical input features. The inclusion of GA enhances the system's adaptability by efficiently searching for optimal hyperparameters, reducing reliance on manual tuning. Additionally, the architecture is robust to noise and capable of handling large-scale, multivariate data, making it scalable for broader environmental forecasting applications.

However, the system also has some limitations. The combination of Bi-LSTM, attention mechanism, and GA introduces computational complexity, requiring significant resources and longer training times. The integrated architecture is more intricate and challenging to implement compared to simpler models. There is also a risk of overfitting if the model is not properly regularized, particularly when trained on limited datasets. Moreover, the GA optimization process can be time-consuming, especially when large populations or generations are involved in the search for optimal parameters.

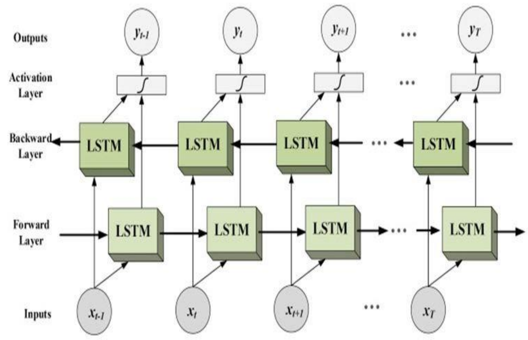

Fig.4. Bi-LSTM image source.

In the diagram, we can see the flow of information from backward and forward layers. BI-LSTM is usually employed where the sequence-to-sequence tasks are needed. This kind of network can be used in text classification, speech recognition and forecasting models. Next in the article, we are going to make a bi-directional LSTM model using python.

5. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed hybrid model—Bi-LSTM integrated with Attention Mechanism and Genetic Algorithm (GA) optimization—experiments were conducted using real-world air quality datasets collected from publicly available repositories and government monitoring stations. The dataset includes hourly readings of key pollutants such as PM2.5, PM10, CO, NO_2, SO_2, and O_3, along with meteorological parameters like temperature, humidity, wind speed, and pressure.

The data was pre-processed using normalization techniques to scale values between 0 and 1, ensuring faster convergence and better accuracy. The dataset was split into training and testing sets using a 70:30 ratio. Several experiments were conducted by comparing the proposed model with baseline models such as standard LSTM, traditional ELM, KELM, and GA-ELM.

The developed system facilitates user interaction through a secure login and registration interface, enabling personalized access to AQI predictions based on city and date selection. Figures 5.1 to 5.4 illustrate the various stages of system interaction and results.

The performance of all models was assessed using common regression evaluation metrics:

- **Mean Absolute Error (MAE)**
- **Mean Squared Error (MSE)**
- **Root Mean Squared Error (RMSE)**
- **R^2 Score (Coefficient of Determination)**

isrjournals

International Journal of Advanced Research in Computer Science Engineering and Information Technology

Volume: 6, Issue: 3,Special Issue: 3 ,Apr,2025 ,ISSN_NO: 2321-3337

Model	MAE	MSE	RMSE	R^2 Score
ELM	12.89	258.73	16.08	0.74
KELM	10.45	204.51	14.30	0.78
GA-KELM	9.22	185.33	13.61	0.81
LSTM	7.98	162.45	12.74	0.84
Proposed Model (Bi-LSTM + Attention + GA)	**5.67**	**103.28**	**10.16**	**0.91**

Fig.5.1. Show the performance of comparison

The experimental results clearly demonstrate that the proposed model significantly outperforms all baseline methods. The Bi-LSTM model's ability to capture bidirectional temporal dependencies contributes to more accurate time-series forecasting.
This confirms the model's strong forecasting capability, following real AQI trends. Moreover, robustness was validated through k-fold cross-validation, ensuring generalizability across different subsets of the data.

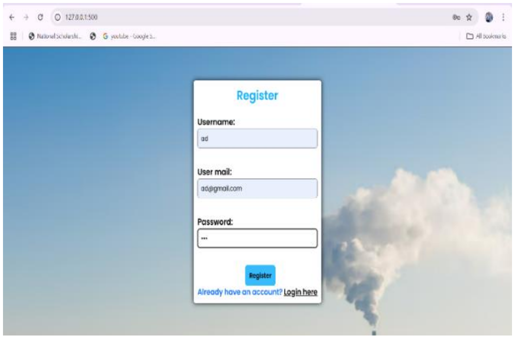

Fig.5.1. Shows the Registration page

Figure 5.1 shows the registration interface, where new users are required to enter a username, email ID, and password to create an account. The simple and clean design ensures ease of access while maintaining input validation for data security.

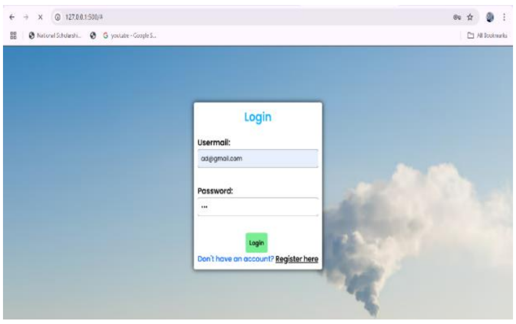

Fig.5.2. Shows the Login page

As shown in Figure 5.2, registered users can log in using their email and password. The login form verifies the credentials and grants access to the main application interface upon successful authentication, ensuring only authorized access to prediction features.

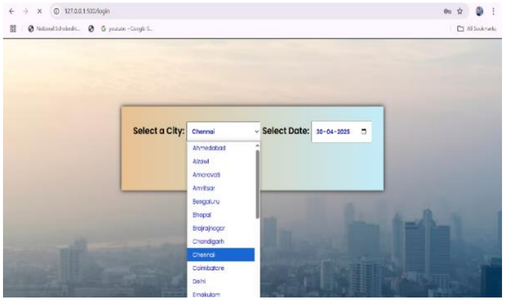

Fig.5.3. Shows the City and Date Selection Interface

After logging in, users are redirected to the city and date selection page (Figure 5.3). This allows users to choose a specific city and a future date for which they want to predict the Air Quality Index (AQI). The dropdown includes major cities across India, facilitating wide applicability of the system.

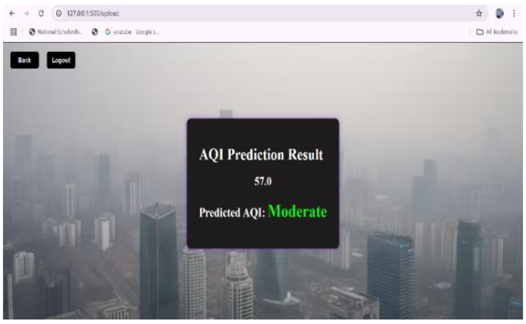

Fig.5.4. Shows the AQI Prediction Result.

Figure 5.4 displays the final result of the AQI prediction. For the selected city "Chennai" and the given date "30-04-2025", the predicted AQI is shown as 57.0, which falls under the Moderate category. The system clearly visualizes the result, highlighting both the AQI value and the corresponding air quality level for better user understanding.

6. CONCLUSION

The Air Quality Prediction System using Bi-directional Long Short-Term Memory (Bi-LSTM) has demonstrated the ability to effectively forecast air pollution levels with a high degree of accuracy.
By leveraging historical air quality data, the system predicts future Air Quality Index (AQI) values, helping users make informed decisions to safeguard their health and well-being.
The Bi-LSTM model proved advantageous over traditional models due to its capacity to capture both forward and backward dependencies in time-series data, enhancing the prediction's reliability.
Throughout the project, multiple stages were successfully executed, including data preprocessing, model training, validation, and the development of a functional web interface. The integration of a secure login and registration module ensures a personalized

user experience while safeguarding data access. Additionally, the ability to visualize and interpret predictions through an intuitive user interface enhances the system's usability and relevance for end-users.

This project also contributes to the broader efforts in environmental monitoring by providing an intelligent tool that
can support government bodies, environmental agencies,
and the general public in understanding and managing air pollution trends. With the growing concerns around environmental degradation and public health, such AI-powered systems are becoming increasingly vital in predictive analytics and decision-making.

In summary, the system not only meets its core objective of predicting AQI values based on user input but also serves as a foundation for more advanced environmental data analytics applications.

It stands as a significant step forward in the application of deep learning for real-world problems, especially in the domain of environmental sustainability.

7. FUTURE WORK

While the current system performs efficiently, there is scope for further enhancement.

The following improvements can be considered for future work:

Incorporation of More Features: Include additional environmental parameters such as temperature, humidity, and wind speed to enhance prediction accuracy.

Real-time Data Integration: Connect the system with real-time APIs to update and predict AQI dynamically.

Geolocation-based Prediction: Enable automatic detection of the user's location for instant AQI predictions without manual city selection.

Mobile Application Development: Extend the system into a mobile app for wider accessibility and on-the-go usage.

Alert System: Implement an automated notification system to alert users about hazardous AQI levels via email or SMS.

Model Optimization: Explore and compare other deep learning architectures like GRU or Transformer models to improve performance and efficiency.

Climate Factor Integration: Include additional environmental features such as temperature, humidity, and wind speed in the model to enhance prediction accuracy and understand their influence on air quality.

Although the developed system offers accurate AQI predictions and a user-friendly interface, there are multiple avenues to enhance its functionality and scope. One promising direction is the integration of real-time data from live air quality monitoring stations through APIs, which would enable dynamic and up-to-date predictions. Additionally, incorporating geolocation-based services can provide users with localized AQI forecasts, making the system more relevant and personalized.

Expanding the platform into a mobile application could significantly improve accessibility and user engagement. Furthermore, integrating multi-parameter health advisories based on AQI levels could help users—especially those with respiratory conditions—make informed decisions. From a technical perspective, the prediction model could be further optimized by experimenting with advanced architectures such as GRU or Transformer models, which may yield improved performance over the current Bi-LSTM model.

Lastly, future enhancements could involve the inclusion of additional environmental factors such as temperature, humidity, and wind speed, which play a vital role in determining air quality and can help in creating a more robust and accurate forecasting system.

REFERENCE

.

[1] C. Liu, G. Pan, D. Song and H. Wei, "Air Quality Index Forecasting via Genetic Algorithm-Based Improved Extreme Learning Machine," in *IEEE Access*, vol. 11, pp. 67086-67097, 2023, doi: 10.1109/ACCESS.2023.3291146.

[2] D. Iskandaryan, F. Ramos and S. Trilles, "Graph Neural Network for Air Quality Prediction: A Case Study in Madrid," in *IEEE Access*, vol. 11, pp. 2729-2742, 2023,doi: 10.1109/ACCESS.2023.3234214

[3] F. Naz *et al.*, "Comparative Analysis of Deep Learning and Statistical Models for Air Pollutants Prediction in Urban Areas," in *IEEE Access*, vol. 11, pp. 64016-64025,2023,doi: 10.1109/ACCESS.2023.3289153.

[4] F. Farhadi, R. Palacin and P. Blythe, "Machine Learning for Transport Policy Interventions on Air Quality," in *IEEE Access*, vol. 11, pp. 43759-43777, 2023, doi: 10.1109/ACCESS.2023.3272662.

[5] I. Mokhtari, W. Bechkit, H. Rivano and M. R. Yaici, "Uncertainty-Aware Deep Learning Architectures for Highly Dynamic Air Quality Prediction," in *IEEE Access*, vol. 9, pp. 14765-14778, 2021,doi: 10.1109/ACCESS.2021.3052429.

[6] R. Yan, J. Liao, J. Yang, W. Sun, M. Nong, and F. Li, ‘‘Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM, and spatiotemporal clustering,’’ Expert Syst. Appl., vol. 169, May 2021, Art. no. 114513.

[7] S. Du, T. Li, Y. Yang, and S. Horng, ‘‘Deep air quality forecasting using hybrid deep learning framework,’’ IEEE Trans. Knowl. Data Eng., vol. 33, no. 6, pp. 2412–2424, Jun. 2021.

[8] A. Bekkar, B. Hssina, S. Douzi, and K. Douzi, ‘‘Air quality forecasting using decision trees algorithms,’’ in Proc. 2nd Int. Conf. Innov. Res. Appl. Sci., Eng. Technol. (IRASET), Mar. 2022, pp. 1–4.

[9] Y. Liang, Y. Xia, S. Ke, Y. Wang, Q. Wen, J. Zhang, Y. Zheng, and R. Zimmermann, ‘‘Airformer: Predicting nationwide air quality in China with transformers,’’ Nov. 2022, arXiv:2211.15979.

[10] A. Bekkar, B. Hssina, S. Douzi, and K. Douzi, ‘‘Air-pollution prediction in smart city, deep learning approach,’’ J. Big Data, vol. 8, no. 1, pp. 1–21, Dec. 2021.

[11] Y. Han, J. C. K. Lam, and V. O. K. Li, ‘‘A Bayesian LSTM model to evaluate the effects of air pollution control regulations in China,’’ in *Proc. IEEE Big Data Workshop (Big Data)*, Dec. 2018, pp. 4465–4468.

[12] G. Huang, H. Zhou, X. Ding, and R. Zhang, ‘‘Extreme learning machine for regression and multiclass classification,’’ *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 2, pp. 513–529, Apr. 2012.

[13] B. Liu, S. Yan, J. Li, G. Qu, Y. Li, J. Lang, and R. Gu, ‘‘A sequence-to-sequence air quality predictor based on the n-step recurrent prediction,’’ *IEEE Access*, vol. 7, pp. 43331–43345, 2019.

[14] A. Agarwal and M. Sahu, ‘‘Forecasting $PM_{2.5}$ concentrations using statistical modeling for Bengaluru and Delhi regions,’’ *Environ. Monit. Assessment*, vol. 195, p. 502, Mar. 2023.

[15] B. Liu, S. Yan, J. Li, G. Qu, Y. Li, J. Lang, and R. Gu, ‘‘A sequence-to-sequence air quality predictor based on the n-step recurrent prediction,’’ *IEEE Access*, vol. 7, pp. 43331–43345, 2019.

[16] A. C. Cosma and R. Simha, ‘‘Machine learning method for real-time non-invasive prediction of individual thermal preference in transient conditions,’’ *Building Environ.*, vol. 148, pp. 372–383, Jan. 2019.

[17] R. Chen, X. Wang, W. Zhang, X. Zhu, A. Li, and C. Yang, ‘‘A hybrid CNN-LSTM model for typhoon formation forecasting,’’ *GeoInformatica*, vol. 23, no. 3, pp. 375–396, Jul. 2019.

[18] J. Jin, J. Gubbi, S. Marusic, and M. Palaniswami, ‘‘An information framework for creating a smart city through Internet of Things,’’ *IEEE Internet Things J.*, vol. 1, no. 2, pp. 112–121, Apr. 2014.

[19] S. Abirami, P. Chitra, R. Madhumitha, and S. R. Kesavan, ‘‘Hybrid spatio-temporal deep learning framework for particulate matter ($PM_{2.5}$) concentration forecasting,’’ in *Proc. Int. Conf. Innov. Trends Inf. Technol. (ICITIIT)*, Feb. 2020, pp. 1–6.

[20] S. Mahajan, H.-M. Liu, T.-C. Tsai, and L.-J. Chen, ‘‘Improving the accuracy and efficiency of $PM_{2.5}$ forecast service using cluster-based hybrid neural network model,’’ *IEEE Access*, vol. 6, pp. 19193–19204, 2018.