RESEARCH PAPER

AWHEF-AQI: An Adaptive Weighted Hybrid Ensemble Learning Framework for Accurate and Robust Air Quality Index Forecasting

S. Vishalatchi[1], Dr. B. Chitradevi[2*], Dr. N. Thinaharan[3], T. Yamunavathi[4], Vrudhvi Kriti[5], M. Ranjani[6]

[1]Assistant Professor, Department of Artificial Intelligence and Data Science, Vivekananda Arts and Science College for Women, Sankagiri | Email: vishaviaas@gmail.com

[2*]Assistant Professor, Department of Computer Applications, SRM Institute of Science and Technology (Deemed to be University), Tiruchirappalli | Email: citradevi.b@gmail.com (Corresponding Author)

[3]Assistant Professor, Department of Computer Science, Thathai Hans Roever College (Autonomous), Perambalur | Email: thinaharan.n@gmail.com

[4]Assistant Professor, Department of Information Technology, Bharathidasan College of Arts and Science (Autonomous), Erode | Email: yamunaavp@gmail.com

[5]Research Scholar, Mother Teresa Women's University, Kodaikanal | Email: krithi6@rocketmail.com

[6]Research Scholar, Department of Computer Science, Periyar University, Salem | Email: ranjanimca2008@gmail.com

ABSTRACT

Air pollution constitutes a significant environmental concern that can adversely impact human health and the ecosystem. The air quality index (AQI) quantifies air quality, indicating the degree of air pollution and its possible effects on human health. This research presents a machine learning and deep learning approach for estimating the Air Quality Index (AQI) utilizing past air quality data. We utilize benchmark data from many air quality monitoring stations to create a predictive model capable of estimating the AQI for a certain area at a specific time. Our methodology encompasses the pre-processing and sanitization of raw data, followed by model training utilizing several machine learning and deep learning methods, including linear regression, decision trees, random forests, and convolutional neural networks. We assess our models utilizing criteria including mean absolute error (MAE), mean squared error (MSE), and R-squared (R2) on a dataset of actual air quality data. Our findings indicate that deep learning models surpass typical machine learning models in predicting the AQI with enhanced accuracy and superior performance. Our proposed methodology can assist government agencies and policymakers in making educated decisions concerning air pollution regulation and management. Individuals can utilize it to monitor and assess the air quality in their vicinity and implement appropriate measures to safeguard their health.

Keywords: Air quality index, Machine learning, Deep learning, Environmental pollution, Model evaluation.

How to cite this article: Vishalatchi S, Chitradevi B, Thinaharan N, Yamunavathi T, Kriti V, Ranjani M. AWHEF-AQI: An Adaptive Weighted Hybrid Ensemble Learning Framework for Accurate and Robust Air Quality Index Forecasting. Int J Drug Deliv Technol. 2026;16(54s): 1529-1536. DOI: 10.25258/ijddt.16.54s.139

Source of support: Nil.

Conflict of interest: None.

1. Introduction

Air pollution from many contaminants harms humans and the environment. Major air pollutants include carbon monoxide (CO), carbon dioxide (CO2), nitrogen dioxide (NO2), nitrogen oxide (NO), sulphur dioxide (SO2), ammonia (NH3), particulate matter (PM 2.5 and PM 10), ozone (O3), acrolein, asbestos, benzene, carbon disulphide (CS2), polycyclic aromatic hydrocarbons (PAHs), synthetic vitreous fibres, and total petroleum hydrocarbons. Air pollution impacts human health and the environment, necessitating enhanced monitoring. Air pollution can lead to respiratory and cardiovascular diseases, cancer, and premature death. Air pollution adversely affects ecosystems, contributes to climate change, and diminishes agricultural output. The assessment of air quality forecast models utilising the testing dataset encompassed R-squared (R^2), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and systematic error (BIAS) [1]. Industries account for 51% of air pollution in India. Additional factors comprise vehicular pollution, exposed refuse, biomass combustion, and construction activities. The Indian government has created emission regulations for corporations and vehicles, advocated for sustainable energy, and curtailed open burning to address air pollution. India must persist in combating air pollution. Table 1 presents the average annual concentrations of common air pollutants in various Indian cities [2].

Table.1. Air pollutants in selected cities in India

Year	City	PM 2.5 (µg	PM 10 (µg	NO 2 (µg	SO 2 (µg	CO (mg /m³)	O3 (µg /m³)

		/m³)	/m³)	/m³)	/m³)		
2020	Delhi	122	292	80	19	1.7	23
2020	Mumbai	63	128	53	14	1.2	29
2020	Kolkata	90	163	56	16	1.4	39
2020	Chennai	46	98	27	8	1.1	29
2020	Hyderabad	57	106	36	8	1.1	41
2020	Bengaluru	39	97	30	6	1.2	27
2020	Ahmedabad	89	173	41	23	1.5	41

PM2.5 and PM10 denote fine particulate matter with diameters of less than 2.5 and 10 micrometres, respectively, whereas NO2, SO2, CO, and O3 are other air pollutants. The data originates from 2020 and encompasses only a limited number of Indian cities, hence it may not accurately represent the nation or subsequent years. The Central Pollution Control Board (CPCB) indicated a slight enhancement in air quality in India in 2020. The COVID-19 lockdown lowered pollution around the nation. Certain cities continued to exhibit elevated levels of air pollution. The subsequent figure illustrates the methodology employed by the CPCB to evaluate air quality through the Air Quality Index (AQI),

AQI Value Range	Air Quality Rating
0-50	Good
51-100	Satisfactory
101-200	Moderate
201-300	Poor
301-400	Very Poor
401-500	Severe

Fig.1. AQI values with air quality ratings

The Central Pollution Control Board (CPCB) said that Delhi's average Air Quality Index (AQI) enhanced from 141 in 2019 to 115 in 2020 as a result of the COVID-19 lockdown. The mean AQI was reduced in Mumbai, Kolkata, and Chennai. Ghaziabad, Bulandshahr, and Bisrakh in Uttar Pradesh continue to exhibit elevated pollution levels, with AQI measurements of 140, 138, and 133, respectively. Air pollution levels impact human health, the environment, and the economy, therefore, predicting the AQI is essential. Precise AQI forecasts enable individuals to protect their health while assisting governments and organizations in mitigating air pollution. Businesses and industries can utilize AQI forecasts to modify operations and mitigate air pollution [3]. The study combines machine learning with deep learning models for predicting the air quality index (AQI). The paper assesses the performance of these models based on several criteria and assists air quality management researchers and practitioners in selecting an AQI prediction model that aligns with their objectives.

2. Literature review

Recent research underscore the increasing application of machine learning (ML) and deep learning (DL) methodologies for predicting the Air Quality Index (AQI). Rosca et al. (2025) [4] demonstrated that the incorporation of pollution and meteorological data enhances prediction accuracy, with ensemble and deep learning models surpassing conventional methods. Ravindiran et al. (2023) [5] evaluated various machine learning models, with CatBoost exhibiting superior performance (R^2 = 0.9998, MAE = 0.60, RMSE = 0.76), highlighting the efficacy of sophisticated ensemble techniques. Psaropa et al. (2025) [6] established a deep learning architecture that integrates localised particle data with meteorological inputs, successfully capturing spatial and temporal fluctuations to improve forecasting accuracy.

Kalantari et al. (2024) [7] conducted a comparison between shallow learning models (RF, SVM, KNN, ANN) and deep learning models (CNN, LSTM, GRU, RNN), revealing that deep learning models, especially CNN, demonstrated superior performance (accuracy reaching 0.60, AUC reaching 0.95) owing to their capacity to predict temporal dependencies. Verma and Sharma (2025) [8] further underscored hybrid ML–DL methodologies utilising CPCB data, demonstrating enhanced prediction reliability and applicability for real-time monitoring. These findings suggest a transition to hybrid and deep learning-based frameworks, while also emphasising issues such data quality, generalisation, and computational complexity, hence necessitating the development of more robust AQI prediction models.

3. Methodology

The proposed methodology for Air Quality Index (AQI) prediction employs a hybrid framework of ML and DL. Air quality and meteorological data are initially gathered and pre-processed to maintain uniformity. The processed data is subsequently represented in tabular formats for ML models and in

time-series sequences for DL models. ML methods are designed to identify correlations between pollutant and environmental variables, whereas DL models are developed to understand temporal dependencies. The forecasts from both models are amalgamated through weighted fusion to derive the final AQI value, with the weights optimized to reduce prediction error. The model's performance is assessed using standard measures to guarantee correctness and reliability.

3.1. Data collection:

The Beijing Municipal Environmental Monitoring Centre gathers air quality data from 35 monitoring locations throughout the city. Air pollution sensors at these sites quantify PM2.5, PM10, SO2, NO2, CO, and O3 on an hourly basis. Weather stations quantify temperature, pressure, humidity, wind direction, wind speed, and air pollution levels. This study utilized hourly air pollution and meteorological data from 2010 to 2014, sourced from the UCI Machine Learning Repository. The dataset comprises 43,824 data points, each characterized by 13 attributes, including hourly values of air pollutants and climatic factors. Initially gathered to assess Beijing's air quality and educate both the public and government. (Link)

No	year	month	day	hour	pm2.5	DEWP	TEMP	PRES	cbwd	Iws	Is	Ir
1	2010	1	1	0	NaN	-21	-11.0	1021.0	NW	1.79	0	0
2	2010	1	1	1	NaN	-21	-12.0	1020.0	NW	4.92	0	0
3	2010	1	1	2	NaN	-21	-11.0	1019.0	NW	6.71	0	0
4	2010	1	1	3	NaN	-21	-14.0	1019.0	NW	9.84	0	0
5	2010	1	1	4	NaN	-20	-12.0	1018.0	NW	12.97	0	0

Fig.2. Sample Dataset

This shows the initial five rows of the dataset, encompassing details regarding the date and time of measurement, concentration levels of various air pollutants (PM2.5, PM10, SO2, NO2, CO, O3), meteorological variables (temperature, pressure, humidity, wind direction, and wind speed), and the monitoring station at which the measurement was recorded. The NaN values signify absent data, which must be addressed during the preprocessing phase.

3.2 Preprocessing

Preprocessing is an essential phase for converting unrefined air quality data into an organized and appropriate format for machine learning and deep learning models. It guarantees data integrity, minimizes interference, and enhances model efficacy. This study's preprocessing encompasses the subsequent steps:

- **Outlier Detection and Removal:** The z-score method detects outliers by quantifying standard deviations from the mean. The z-score is calculated as: $z = (x - \mu) / \sigma$. Data points with $|z|>3$ are considered outliers and are removed to reduce noise and improve model stability.
- **Missing value handling:** Identifying missing values across all attributes guarantees data integrity prior to model training. Statistical imputation substitutes absent values with the mean of the characteristic. The imputation is performed as: $X_{imputed} = X.\text{fillna}(X.\text{mean}())$
- **Feature Selection:** The selection of pollutant concentrations and climatic conditions for AQI prediction is influenced by their impact. This phase decreases dataset dimensionality and eliminates extraneous variables. The model focuses on significant features, improving efficiency and predictive accuracy.
- **Normalization:** Feature scaling standardizes all input variables to a uniform range to avoid larger features from overshadowing others. This enhances convergence stability during model training and facilitates learning. The dataset is subjected to min-max scaling or standardization.

No	year	month	day	hour	pm2.5	DEWP	TEMP	PRES	cbwd	Iws	Is	Ir
25	2010	1	2	0	129.0	-16	-4.0	1020.0	SE	1.79	0	0
26	2010	1	2	1	148.0	-15	-4.0	1020.0	SE	2.68	0	0
27	2010	1	2	2	159.0	-11	-5.0	1021.0	SE	3.57	0	0
28	2010	1	2	3	181.0	-7	-5.0	1022.0	SE	5.36	1	0
29	2010	1	2	4	138.0	-7	-5.0	1022.0	SE	6.25	2	0

Fig.3. Preprocessed data

This output displays the initial rows of the revised dataset subsequent to the elimination of outliers and missing values. The dataset is devoid of missing values and outliers beyond the defined ranges.

3.3. Classification

Predicting the air quality index (AQI) necessitates appropriate ML and DL models. Accurate forecasts can assist the public in safeguarding their health by delivering reliable air quality data. Understanding the causes of air pollution necessitates an emphasis on interpretability. Scalability and robustness are crucial for AQI prediction models as they minimize computational complexity and provide precise forecasts in rapidly changing conditions. The selection of appropriate ML and DL models is essential for precise, interpretable, scalable, and resilient air quality

index forecasts that can aid individuals in making health and wellness decisions. Baseline comparison models such as Random Forest (RF) [9], Support Vector Regression (SVR) [10], Gradient Boosting Machines (GBMs) [11], K-Nearest Neighbor (KNN) [12], Artificial Neural Networks (ANNs) [13], Convolutional Neural Networks (CNNs) [14], and Long Short-Term Memory (LSTM) [15] networks are employed to assess the proposed hybrid model.

3.3.1. Proposed method

The proposed hybrid model combines ML and DL techniques to enhance the accuracy and resilience of Air Quality Index (AQI) predictions. Machine learning techniques, like RF, SVR, and GB, are proficient at identifying structured correlations among environmental and pollutant characteristics. Nonetheless, they are constrained in their ability to model the temporal dependencies inherent in air quality data. To address this constraint, DL models, such CNN and LSTM networks, are utilised, as they can discern intricate nonlinear and temporal patterns from sequential data. The input data is supplied in two formats. The ML component employs tabular characteristics X_ML, encompassing pollutant concentrations (PM2.5, PM10, NO_2, SO_2, CO, O_3) and meteorological variables (temperature, humidity, wind speed). The DL component uses sequential data $X_{DL} = [x_{t-n}, \ldots, x_t]$, capturing temporal variations over a time window of length n. This dual representation enables the model to utilize both current observations and historical trends. The ML model produces predictions based on structured inputs as $y_{ML} = f_{ML}(X_{ML})$, while the DL model generates predictions using temporal learning. For example, in the case of LSTM, the hidden state and output are computed as:

$$h_t = \sigma(W \cdot [h_{t-1}, x_t] + b), \qquad y_{DL} = W_h \cdot h_t + b_h$$

Where, h_t represents the hidden state at time t, and σ denotes the activation function. The final AQI prediction is obtained by combining the outputs of both models using weighted fusion: $AQI = w_1 \cdot y_{ML} + w_2 \cdot y_{DL}$, $where\ w_1 + w_2 = 1$,
where w_1 and w_2 control the contribution of ML and DL predictions respectively. The optimal weights are determined by minimizing the mean squared error between the predicted and actual AQI values:

$$\min_{w_1, w_2} \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

This hybrid approach seamlessly integrates feature-based learning from ML models and temporal pattern extraction from DL models, yielding enhanced predicted accuracy and resilience under diverse environmental conditions.

Input: Dataset D (pollutant and meteorological data)
Output: Predicted AQI (ŷ)

Load dataset D
Preprocess:
  Handle missing values
  Remove outliers
  Normalize features
Construct inputs:
  $X_{ML} \leftarrow [PM2.5, PM10, NO2, SO2, CO, O3, T, H, W]$
  $X_{DL} \leftarrow$ generate sequences $[x(t-n), \ldots, x(t)]$
Split dataset into Train, Validation, Test
Train ML_{model} using $X_{ML_{train}}$ and y_{train}
Train DL_{model} using $X_{DL_{train}}$ and y_{train}
$y_{ML} \leftarrow ML_{model}.predict(X_{ML})$
$y_{DL} \leftarrow DL_{model}.predict(X_{DL})$
Initialize: w1 ← 0.5, w2 ← 0.5
While convergence condition not met:
  $\hat{Y} \leftarrow w_1 * y_{ML} + w2 * y_{DL}$
  Compute loss: $L \leftarrow \left(\frac{1}{n}\right) * \Sigma(y - \hat{y})^2$
  If L decreases:
    Update w1, w2
  Else:
    Stop updating
Ensure: $w_1 + w_2 = 1$
Final prediction: $AQI_{pred} \leftarrow w1 * y_{ML} + w2 * y_{DL}$
Compute evaluation metrics: MAE, RMSE, R^2
Return AQI_{pred}

The integration of two modeling techniques provides the hybrid ML-DL approach with a significant benefit. Machine learning algorithms learn effectively from structured environmental data, whereas deep learning models identify previously overlooked intricate temporal relationships. This complementary integration enhances prediction precision, model constraints, and environmental resilience. Optimized fusion integrates both models, rendering the system adaptable, scalable, and appropriate for real-time AQI monitoring.

4. Results and Discussion

This section discusses experimental results utilising three ML approaches on the air quality dataset. Following pre-processing, the dataset was divided into training and testing sets, with each method being trained and assessed. MAE, MSE, RMSE, and R^2 evaluated algorithm efficacy. The hybrid model surpassed previous techniques in terms of MAE, MSE, RMSE, and R2. Support Vector Regression underperformed compared to K-Nearest Neighbour.

According to the performance metrics of this experiment, the hybrid model is the most effective strategy for air quality prediction. The "PRSA_data_2010.1.1-2014.12.31.csv" dataset comprises Beijing air quality data spanning from January 1, 2010, to December 31, 2014. This dataset comprises the following features:

Table.3. Features of dataset

Feature	Description
No	The serial number of the data point
year	The year of the observation
month	The month of the observation
day	The day of the observation
hour	The hour of the observation (in 24-hour format)
PM2.5	The concentration of PM2.5 particles (measured in micrograms per cubic meter)
PM10	The concentration of PM10 particles (measured in micrograms per cubic meter)
SO2	The concentration of sulfur dioxide (measured in micrograms per cubic meter)
NO2	The concentration of nitrogen dioxide (measured in micrograms per cubic meter)
CO	The concentration of carbon monoxide (measured in milligrams per cubic meter)
O3	The concentration of ozone (measured in micrograms per cubic meter)
TEMP	The temperature (measured in degrees Celsius)
PRES	The air pressure (measured in hectopascals)
DEWP	The dew point temperature (measured in degrees Celsius)
RAIN	The amount of rainfall (measured in millimeters)
wd	The wind direction (measured in degrees)
WSPM	The wind speed (measured in meters per second)

Overall, this dataset contains 43,824 observations, with 17 features for each observation. The data can be used to study air quality and weather patterns in Beijing over a period of five years.

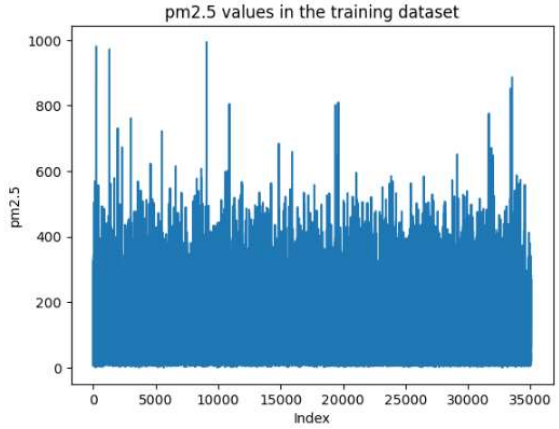

Fig.4. pm2.5 values in the training dataset

The plot demonstrates the temporal variations of the 'PM2.5' column in the train dataset. The DataFrame index is represented on the x-axis, while the 'PM2.5' column values are depicted on the y-axis. The numbers in the 'PM2.5' column range from 0 to 900, exhibiting temporal fluctuations, with spikes indicating outliers or extraordinary occurrences. This illustrates the distribution and trend of the training dataset's 'PM2.5' values, facilitating our understanding of the data and enabling informed analytical judgments.

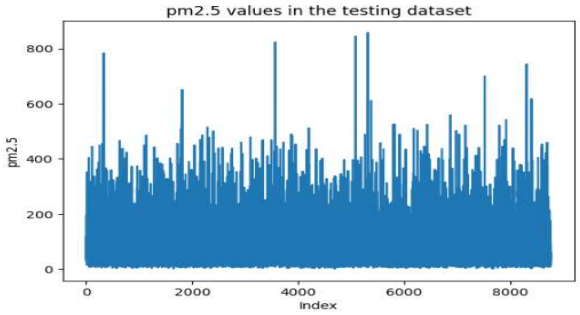

Fig.5. pm2.5 values in the testing dataset

The plot of the 'PM2.5' column in the test dataset illustrates its levels over time. The visualization would resemble the 'PM2.5' column plot from the training dataset, with the x-axis denoting the DataFrame index and the y-axis illustrating the column values. The graph of the 'PM2.5' column in the test dataset will depict the distribution and temporal trends of the values, similar to the plot of the training dataset. Spikes and outliers in the test dataset, akin to those in the training dataset, may indicate measurement inaccuracies or anomalous occurrences. The plot of the 'PM2.5' column in the test dataset can evaluate the model's performance on novel data and its generalizability. If the distribution and trend of 'PM2.5' values in the test dataset correspond with those in the training dataset, the model has acquired significant patterns and can reliably forecast new data.

4.1 Performance evaluation metrics

Diverse metrics can assess the success of categorization models. MAE and RMSE measure the average prediction error in the same unit as AQI, facilitating comprehension. MSE prioritizes significant errors to penalize substantial deviations. The R^2 score reflects the model's ability to account for AQI variability, indicating overall adequacy.

Table.4.Evaluation Metrics

Metrics	Formula
Mean Absolute Error (MAE)	$MAE = \frac{1}{n}\sum_{i=1}^{n} \lvert y_i - \hat{y}_i \rvert$
Mean Squared Error (MSE)	$MSE = \frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{MSE}$
Coefficient of Determination (R2)	$R2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

Here, y_i is the actual AQI value, $\hat{y}_i$ denotes predicted AQI value, $\bar{y}$ represents mean of actual values and n is the number of samples. Together, these metrics ensure a balanced and reliable evaluation of the predictive capability of the proposed model.

Table.5. Performance Evaluation with single Algorithms

Model	MAE	MSE	RMSE	R2
LR	24.13	981.69	31.31	0.24
RF	21.65	838.7	28.96	0.39
SVR	26.02	1174.85	34.26	0.11
KNN	22.98	886.66	29.78	0.34
ANN	22.67	879.33	29.64	0.35
LSTM	21.96	869.68	29.49	0.36
CNN	21.8	863.06	29.36	0.37

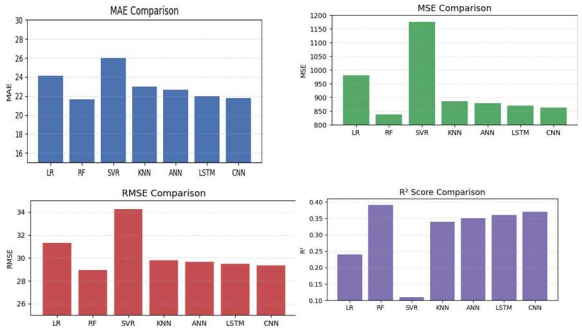

Fig.6. Performance Evaluation with single Algorithms

The evaluation metrics indicate that the Random Forest algorithm is the most effective for predicting AQI using the Beijing PM2.5 Data compared to the other assessed methods. The Random Forest model has the lowest MAE (21.65), MSE (838.7), and RMSE (28.96) values, alongside the highest R^2 (0.24) value among the six models. The measurements demonstrate that the Random Forest model yields the most precise forecasts of AQI values relative to the other models.

Table.6. Performance Evaluation with ML_CNN Algorithms

Model	MAE	MSE	RMSE	R2
LR _CNN	19.82	652.97	25.59	0.53
RF_CNN	19.26	620.67	24.92	0.56
SVR_CNN	20.71	746.11	27.32	0.45
KNN_CNN	20.5	731.33	27.03	0.47
ANN_CNN	19.58	633.82	25.17	0.54

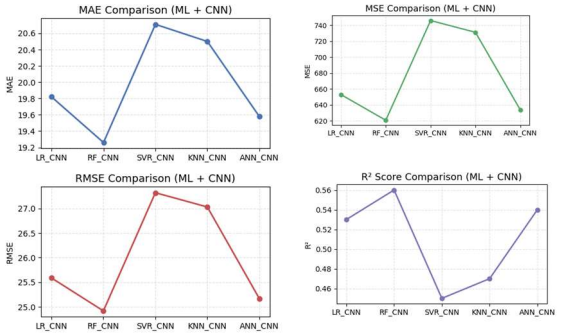

Fig.7. Performance Evaluation with Hybrid ML-CNN Algorithms

The above table and plots indicate that the Random Forest + CNN algorithm is the most effective for predicting AQI using the Beijing PM2.5 Data among the assessed methods. The Random Forest + CNN model exhibits the lowest MAE, MSE, and RMSE values, alongside the highest R2 value among the five models. The measurements demonstrate that the Random Forest + CNN model yields the most precise forecasts of AQI values in comparison to the other models.

Table.7. Performance Evaluation with ML_CNN Algorithms

Model	MAE	MSE	RMSE	R2
LR_LSTM	20.89	701.84	26.49	0.48
RF_LSTM	19.7	656.56	25.61	0.53
SVR_LSTM	20.98	751.97	27.42	0.45
KNN_LSTM	21.47	776.15	27.87	0.42
ANN_LSTM	19.44	622.56	24.96	0.56

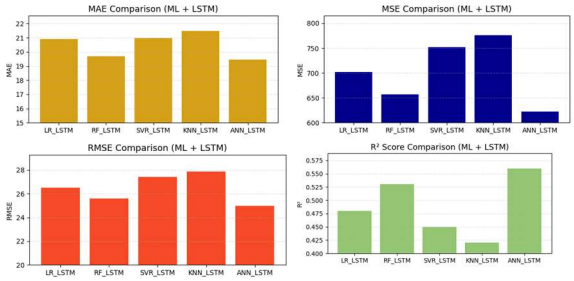

Fig.8. Performance Evaluation with Hybrid ML-LSTM Algorithms

The table indicates that the ANN _LSTM approach has superior performance. The ANN_LSTM model exhibits the lowest MAE, MSE, and RMSE values, signifying enhanced predictive accuracy. The model possesses the highest R-squared value of 0.56, signifying its capacity to elucidate 56% of the target variable's variability, so establishing it as the most effective model for this task. The LR and KNN + LSTM models exhibit the lowest R-squared values, indicating poor predictive accuracy. Although the RF_LSTM model performs adequately, its R-squared value is marginally worse to that of the ANN_LSTM model. The evaluation metrics indicate that the ANN_LSTM model excels in predictive modeling.

5. Conclusion

This study indicated that the ML and DL models can estimate the Air Quality Index (AQI), although success depends on the technique. The ANN_LSTM model predicts AQI best in this predictive modelling task with the highest R-squared value. The study found that deep learning models like LSTM and CNN predict performance better than LR and KNN-based LSTM. The RF_LSTM model is good but less precise than the ANN_LSTM model. Location and air pollutant type affect model performance. Location-specific models that account for local variables and pollution sources are necessary. Air quality monitoring systems benefit from hybrid machine learning and deep learning AQI prediction models. To account for the complex interactions between air pollutants and meteorological variables, we need more research to validate these models in real-world scenarios and produce more accurate and robust models.

References

1. Binbusayyis, A., Khan, M.A., Ahmed A, M.M. et al. A deep learning approach for prediction of air quality index in smart city. Discov Sustain 5, 89 (2024). https://doi.org/10.1007/s43621-024-00272-9
2. Packiam, R. Merlin, Ms M. Ellakkiya, and Ms V. Infine Sinduja. "A hybrid machine learning regression framework for air quality prediction with meta-heuristic approach", 5(4), 2633-4828, December, 2023.
3. Biswas, S., Mukherjee, S., & Roychowdhury, T. (2021). An overview of air pollution in India and its control measures. Environmental Monitoring and Assessment, 193(4), 191.
4. Rosca, C.-M.; Carbureanu, M.; Stancu, A. Data-Driven Approaches for Predicting and Forecasting Air Quality in Urban Areas. Appl. Sci. 2025, 15, 4390. https://doi.org/10.3390/app15084390
5. Psaropa, M.X.; Kontogiannis, S.; Lolis, C.J.; Hatzianastassiou, N.; Pikridas, C. A Proposed Deep Learning Framework for Air Quality Forecasts, Combining Localized Particle Concentration Measurements and Meteorological Data. Appl. Sci. 2025, 15, 7432. https://doi.org/10.3390/app15137432
6. Ravindiran, S., et al. (2023). Prediction of air quality index using machine learning algorithms. Chemosphere, 337, 139653. https://doi.org/10.1016/j.chemosphere.2023.139653
7. Kalantari, E., Gholami, H., Malakooti, H. et al. Machine learning for air quality index (AQI) forecasting: shallow learning or deep learning?. Environ Sci Pollut Res 31, 62962–62982 (2024). https://doi.org/10.1007/s11356-024-35404-1
8. Verma, A., & Sharma, V. (2025). Air quality index (AQI) prediction using machine learning and deep learning approaches. International Journal for Research in Applied Science and Engineering Technology (IJRASET). https://doi.org/10.22214/ijraset.2025.71616
9. Singh, S., Kumar, M., Verma, B.K. et al. Optimizing Air Pollution Prediction With Random Forest Algorithm. Aerosol Sci Eng (2025). https://doi.org/10.1007/s41810-025-00292-6
10. Yusof, Y., Maijama'a, I.S. (2024). Air Quality Index Prediction Using Support Vector Regression Based on African Buffalo Optimization. In: Zakaria, N.H., Mansor, N.S., Husni, H., Mohammed, F. (eds) Computing and Informatics. ICOCI 2023. Communications in Computer and Information Science, vol 2002. Springer, Singapore. https://doi.org/10.1007/978-981-99-9592-9_1

11. H. Zhang, X. Zhang, and X. Liu, "A Novel Hybrid Model Combining Convolutional Neural Network and Extreme Gradient Boosting for Air Quality Index Forecasting," Atmospheric Environment, vol. 311, 2022.
12. Jiang Wen, Caiqing Xie, Lin Yang, Haojiang Li, Lingyu Song, and Jianjun Zhang. 2025. Designing of KNN-based Air Quality Predicting System. In Proceedings of the 2nd International Conference on Machine Intelligence and Digital Applications (MIDA '25). Association for Computing Machinery, New York, NY, USA, 51–57. https://doi.org/10.1145/3744464.3744474
13. Kamsing, P.; Cao, C.; Boonpook, W.; Boonprong, S.; Xu, M.; Boonsrimuang, P. Artificial Neural Network for Air Pollutant Concentration Predictions Based on Aircraft Trajectories over Suvarnabhumi International Airport. Atmosphere 2025, 16, 366. https://doi.org/10.3390/atmos16040366
14. Y. Chen, Y. Huang, and C. Chen, "A Graph Convolutional Neural Network Approach for Air Pollution Prediction in Cities," Atmospheric Environment, vol. 315, 2022.
15. Wang et al. (2022): Wang, H., Huang, Z., Zhou, Q., & Wang, J. (2022). Hybrid attention-based convolutional neural network-long short-term memory model for air quality prediction. Science of the Total Environment, 806, 150574.