AQI Prediction Using Machine Learning and Deep Learning Approaches: A Comparative Study

Md. Rakibur Rahman Shovon
Department of CSE
BRAC University
Dhaka, Bangladesh
rakibur.rahman.shovon@g.bracu.ac.bd

Nishat Tasnim Luba
Department of CS
BRAC University
Dhaka, Bangladesh
nishat.tasnim.luba@g.bracu.ac.bd

Md. Mostofa All Zubayer
Department of CS
BRAC University
Dhaka, Bangladesh
mostofa.all.zubayer@g.bracu.ac.bd

***Abstract*—Air pollution is a critical global health challenge, contributing to millions of premature deaths annually and imposing significant socio-economic burdens [1]. This study presents a comprehensive approach to forecasting the European Air Quality Index (AQI) using a year-long (Jan–Dec 2024) dataset comprising over 52,000 daily pollutant measurements from six geographically diverse cities [3]. The pollutants considered include CO, NO_2, SO_2, O_3, $PM_{2.5}$, and PM_{10}, all standardized under the European AQI framework [2] to enable cross-city comparability. We evaluate multiple regression-based machine learning models-Linear Regression, Random Forest, Support Vector Regression, K-Nearest Neighbors, Gradient Boosting-and deep learning architectures including Multilayer Perceptrons and hybrid neural networks [4] [5]. In addition to predictive accuracy, we employ explainable AI techniques to identify the relative contribution of each pollutant to AQI variation [6]. The results demonstrate that although Deep Learning models showed the highest accuracy but ensemble-based models, particularly Random Forest were significantly closer in accuracy with less complexity. The findings provide actionable insights for environmental monitoring agencies and policymakers seeking robust, interpretable AQI forecasting solutions.**

***Index Terms*—Air Quality Index (AQI), Air Pollution Forecasting, Machine Learning, Deep Learning, Ensemble Methods, Environmental Data Analytics, Explainable AI**

I. Introduction

Air quality is a critical determinant of public health, economic productivity, and ecological stability. According to the World Health Organization (WHO), 99% of the global population breathes air exceeding guideline limits for key pollutants, with the burden disproportionately affecting low- and middle-income countries [1]. Exposure to fine particulate matter ($PM_{2.5}$) alone accounted for approximately 238,000 premature deaths in the EU-27 in 2020 [10], while global exposure to ambient and household air pollution remains the second leading risk factor for mortality [11].

Urbanization, industrialization, and increased energy demand have intensified emissions of particulate matter ($PM_{2.5}$, PM_{10}), NO, SO_2, O_3, and CO. These pollutants not only degrade air quality but also exacerbate climate change through complex atmospheric interactions. The health impacts range from acute respiratory infections to chronic cardiovascular diseases, with emerging evidence linking long-term exposure to neurological disorders and adverse birth outcomes [1] [11] [12].

Accurate forecasting of air quality is essential for enabling timely public health interventions, informing environmental policy, and guiding individual behavioral responses. Recent advances in machine learning (ML) and deep learning (DL) have demonstrated superior predictive performance over traditional statistical models, particularly when integrating meteorological and spatiotemporal data [6] [8]. However, many existing studies are constrained by limited geographic coverage, inconsistent pollutant measurement standards, or insufficient temporal resolution.

This study addresses these gaps by employing a standardized, globally representative dataset [3] covering six cities across all inhabited continents, with daily pollutant measurements for the entirety of 2024. The dataset's consistent European AQI metric enables robust cross-city comparisons, while its temporal granularity supports both short-term forecasting and long-term trend analysis. By combining predictive modeling with explainable AI, the research aims to deliver both accuracy and interpretability-critical for actionable environmental intelligence.

II. Literature Review

Air quality prediction has become a vital area of research due to its implications for public health and environmental policy. Machine learning (ML) techniques have been widely adopted to model and forecast pollutant concentrations and Air Quality Index (AQI) values. The following review synthesizes recent studies that explore various ML approaches, data preprocessing techniques, and model architectures.

A. Hybrid and Deep Learning Models

Chen et al. [5] proposed a CNN-RF ensemble framework for predicting $PM_{2.5}$ concentrations in Kaohsiung, Taiwan. The model leverages CNN for feature extraction and Random Forest (RF) for regression, integrating meteorological and spatiotemporal variables. The CNN-RF model achieved an RMSE of 4.88 μg/m³ and R^2 of 0.88, outperforming standalone CNN and RF models.

Gupta et al. [6] compared SVR, RF, and CatBoost regression across four Indian cities. They addressed dataset imbalance

using SMOTE and found RF to be most effective in Kolkata and Hyderabad, while CatBoost excelled in New Delhi and Bangalore. Their study emphasized the importance of data balancing and feature selection.

Alkabbani et al. [7] introduced a novel approach using missForest-a Random Forest-based imputation method-to handle missing data before training ANN models. Their method improved AQI prediction accuracy to 92.41% and demonstrated superior generalization compared to linear imputation techniques.

B. Classical and Statistical Methods

Liu et al. [7] employed LightGBM, Logistic Regression, and RF classifiers for AQI category prediction, and LSTM networks for regression tasks. LightGBM achieved 97.5% classification accuracy, while LSTM yielded over 90% goodness-of-fit for AQI and ozone levels. Their work highlighted the seasonal influence of meteorological variables and the effectiveness of feature ranking.

C. Survey and Meta-Analysis

Méndez et al. [9] conducted a comprehensive survey of 155 papers from 2011–2021, analyzing trends in ML applications for air quality forecasting. Key insights include:

- Deep learning models, especially LSTM and MLP, outperform classical regression methods.
- $PM_{2.5}$ is the most frequently predicted pollutant.
- Combining pollutant and meteorological features yields better accuracy.
- Hybrid models and ensemble techniques are increasingly popular.

Their review also identified emerging trends such as the use of Transformer networks, Graph Neural Networks, and Temporal Convolutional Networks for spatiotemporal modeling.

III. Methodology

A. Dataset Description

The dataset utilized in this study is a comprehensive compilation of global air quality measurements collected throughout the calendar year 2024 [3]. It comprises over 52,000 daily records from six geographically diverse cities, representing all inhabited continents. Each record includes concentrations of key atmospheric pollutants-namely carbon monoxide (CO), carbon dioxide (CO_2), nitrogen dioxide (NO_2), sulfur dioxide (SO_2), ozone (O_3), and particulate matter ($PM_{2.5}$ and PM_{10})-alongside the European Air Quality Index (AQI), which serves as a standardized metric for cross-city comparison.

All measurements are timestamped in GMT and reported in standardized units (CO_2 in ppm; all other pollutants in µg/m³), ensuring consistency across locations. The dataset is structured in two formats: a combined file containing all cities with a city identifier column, and city-specific files tailored for localized analysis. This dual structure supports both global and regional investigations.

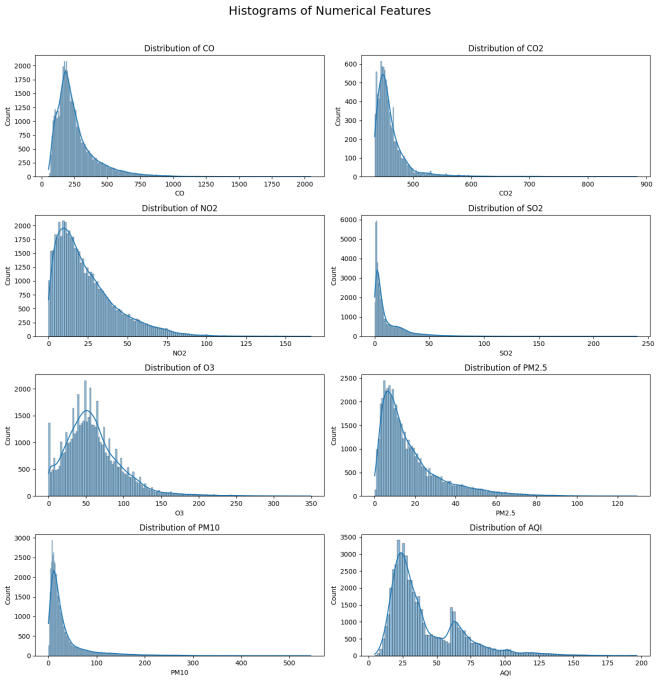

Fig. 1. Histogram of numerical features

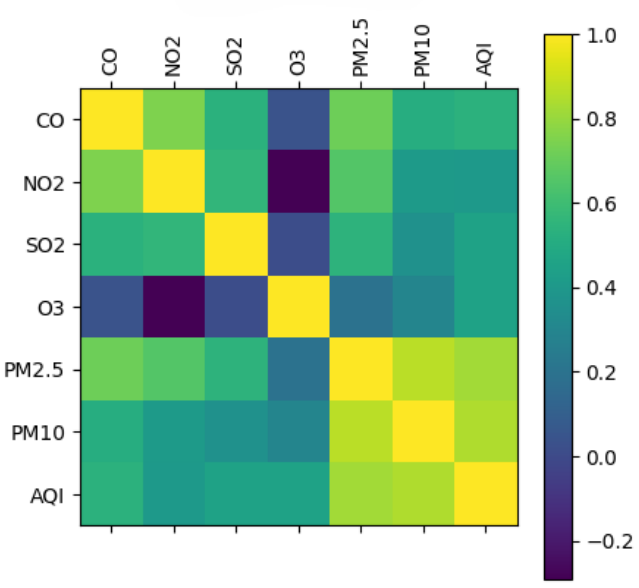

Fig. 2. Correlation Matrix

The dataset is particularly suited for time-series forecasting, pollutant correlation analysis, geospatial comparisons, and health impact studies. Given its temporal granularity and pollutant diversity, it enables robust modeling of AQI dynamics and facilitates exploration of pollution disparities and their potential health consequences. The dataset is published under the MIT License, permitting unrestricted use for academic and commercial purposes [3].

B. Data Preprocessing

The initial dataset comprised 52,704 daily air quality records collected across six globally distributed cities throughout the year 2024. Each record included timestamped measurements for seven pollutants-CO, CO_2, NO_2, SO_2, O_3, $PM_{2.5}$, and PM_{10}-alongside the European Air Quality Index (AQI), which served as the target variable. Upon inspection, three columns-CO_2, Date, and City-were excluded from further analysis due to their limited utility in the modeling context.

The CO_2 column exhibited extensive missing data, with over 43,000 entries absent. More importantly, CO_2 is not among the five pollutants used in the official calculation of the European AQI, as defined by the European Environment Agency. Including it would have introduced noise and potentially misled the model into learning irrelevant patterns. Dropping it ensured a cleaner and more focused dataset. The Date column, while potentially useful for temporal analysis, was removed to concentrate the study on pollutant-based AQI prediction rather than seasonal or hourly trends. Similarly, the City column was excluded to prevent geographic bias and to generalize the model across locations. Including city identifiers could have led to overfitting, where the model learns city-specific patterns rather than pollutant dynamics.

Following column removal, the dataset was checked for duplicate entries, which were subsequently eliminated to avoid redundancy and ensure model integrity. The remaining features-CO, NO_2, SO_2, O_3, $PM_{2.5}$, PM_{10}-were retained as predictors, with AQI as the target. The data was then partitioned into training and testing. Finally, feature scaling was applied using standard normalization techniques to account for the varying magnitudes and units of the pollutants. This step was essential for ensuring fair treatment of all features, particularly in models sensitive to feature scale.

Had the dropped columns been retained without proper handling, they could have introduced noise, skewed the learning process, or led to misleading correlations. The preprocessing strategy adopted here ensures a robust foundation for subsequent modeling and analysis.

C. Learning Phase

To model the relationship between pollutant concentrations and the European Air Quality Index (AQI), several supervised regression algorithms were employed: Linear Regression, Random Forest Regressor, Support Vector Regression, K-Nearest Neighbors (KNN), Gradient Boosting Regressor, Multilayer Perceptron Regressor, a CNN–MLP hybrid, and Physics-Guided Neural Networks (PGNNs). These models were chosen to compare the performance of simple linear methods against more complex ensemble-based, deep learning, and domain-informed approaches.

- **Linear Regression (LR)**
  Linear Regression is one of the earliest and most interpretable regression techniques [13]. It models the target variable as a weighted sum of input features plus an intercept, assuming a linear relationship between predictors and output. The model parameters are estimated by minimizing the residual sum of squares between observed and predicted values. Despite its simplicity, LR serves as a strong baseline for evaluating more advanced algorithms.

- **Random Forest (RF)**
  Random Forest is an ensemble learning method introduced by Breiman [14]. It constructs multiple decision trees during training, each built from a bootstrap sample of the data and a random subset of features. Predictions are obtained by averaging the outputs of all trees (for regression). This randomness reduces overfitting and improves generalization, making RF effective for capturing complex, nonlinear relationships.

- **Support Vector Regression (SVR)**
  SVR extends the Support Vector Machine framework to regression tasks [15]. Instead of finding a separating hyperplane, SVR fits a function within an ϵ-insensitive margin, ignoring small deviations from the true values. Larger deviations are penalized in the optimization process. By using kernel functions, SVR can model nonlinear relationships in high-dimensional feature spaces.

- **K-Nearest Neighbors (KNN)**
  KNN is a non-parametric, instance-based learning algorithm [16]. For regression, the predicted value of a query point is the average of the target values of its k closest neighbors in the training set, where closeness is typically measured using Euclidean distance. KNN adapts naturally to complex data distributions but can be sensitive to noise and irrelevant features.

- **Gradient Boosting Regressor (GBR)**
  Gradient Boosting is a sequential ensemble method proposed by Friedman [17]. It builds models in stages, where each new weak learner (often a shallow decision tree) is trained to predict the residual errors of the combined ensemble so far. By iteratively minimizing a differentiable loss function, GBR can capture subtle patterns in the data, though it requires careful tuning to avoid overfitting.

- **Multilayer Perceptron Regressor (MLP)**
  MLP is a feed-forward artificial neural network trained with backpropagation [18]. It consists of an input layer, one or more hidden layers with nonlinear activation functions, and an output layer. The network learns by adjusting weights to minimize prediction error. In this study, three architectures were tested: (i) a single hidden layer with 64 neurons (max_iter=1000), (ii) a deeper network with three hidden layers of 128, 64, and 32 neurons (max_iter=1500), and (iii) a deep architecture with four hidden layers of 256, 128, 64, and 32 neurons (max_iter=2000, learning_rate=0.001, L2 regularization α=0.001). All models employed ReLU activation functions and the Adam optimizer. The deeper models can

capture more complex feature interactions but require more training iterations and regularization to prevent overfitting.

- **CNN–MLP Hybrid**
  The CNN–MLP hybrid combines convolutional layers for automated feature extraction with fully connected layers for regression [19]. The architecture employs a single 1D convolutional layer with 32 filters (kernel size 3, padding 1) followed by max pooling (kernel size 2) for feature extraction. Input pollutant data is reshaped using an unsqueeze operation to enable 1D convolution over the feature dimension. The extracted features are flattened and passed to a fully connected layer with 64 neurons before the final regression output. This design enables the model to leverage local pattern recognition in pollutant relationships while maintaining regression capabilities. The model was trained for 100 epochs using Adam optimizer (learning rate 0.001) with batch size 32.

- **Physics-Guided Neural Networks (PGNN)**
  PGNNs integrate physical domain knowledge into the neural network learning process [20]. This implementation features a three-layer architecture (128-64-1 neurons) with ReLU activations, trained using a custom physics-aware loss function. The loss combines standard MSE with penalty terms that enforce physical constraints, specifically bounding AQI predictions within the valid range (0-500). The physics loss function applies penalties with weight 0.1 for predictions below zero or above 500, ensuring physically plausible outputs. The model was trained for 100 epochs using Adam optimizer (learning rate 0.001) with batch size 32. While such constraints can improve generalization when physical laws are well-defined, the benefit was less pronounced in this study due to the relatively weak spatial or physical structure in the tabular pollutant dataset.

IV. Results and Analysis

In this study, multiple regression algorithms were applied to the dataset, and their performance was compared using three standard metrics: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and coefficient of determination (R^2). The summary of results of ML models are presented in Table I and artificial neural networks in Table II.

TABLE I
PERFORMANCE OF DIFFERENT REGRESSION MODELS

Model	MSE	RMSE	R^2
Linear Regression	126.47	11.25	0.813
Random Forest Regressor	60.65	7.79	0.910
Support Vector Regression (SVR)	73.87	8.59	0.891
Gradient Boosting Regressor	76.60	8.75	0.887
K-Nearest Neighbors Regressor	61.54	7.85	0.909

TABLE II
PERFORMANCE OF DIFFERENT ARTIFICIAL NEURAL NETWORKS

Model	MSE	RMSE	R^2
MLP Regressor (1 layer, 64 neurons)	74.59	8.64	0.890
MLP Regressor (128-64-32)	67.52	8.22	0.900
Deep MLP (256-128-64-32)	59.65	7.72	0.91
AttentionMLP	69.36	8.32	0.897
PhysicsGuidedNN	71.13	8.43	0.894
CNN-MLP Hybrid	76.87	8.768	0.886

The performance metrics in Table I and Table II reveal clear trends in how different algorithms handle the AQI prediction task. The Deep MLP Regressor (256-128-64-32) achieved the lowest RMSE (7.72) and tied with Random Forest for the highest R^2 (0.910). Random Forest delivered comparable performance (RMSE = 7.79, R^2 = 0.910) while requiring significantly less hyperparameter tuning and computational resources. This can be attributed to Random Forest's ensemble nature-aggregating multiple decision trees reduces variance and captures complex nonlinear relationships between pollutants and AQI [4]. The model's ability to handle feature interactions without explicit feature engineering is particularly advantageous in environmental datasets, where pollutant effects are often interdependent and context-specific [6].

K-Nearest Neighbors (R^2 = 0.909) performed nearly as well, benefiting from its instance-based learning approach, which directly leverages local patterns in the data. However, its reliance on distance metrics makes it sensitive to feature scaling and noise, which could explain the slight performance gap compared to the top performers.

Support Vector Regression and Gradient Boosting both delivered strong results ($R^2 \approx 0.887$–0.891), with SVR's kernel trick enabling it to model nonlinearities effectively [7]. Gradient Boosting's sequential error-correction process allowed it to refine predictions iteratively, though it may have been more prone to overfitting without extensive hyperparameter tuning [8].

Neural network models demonstrated that architectural depth significantly impacts performance. The Deep MLP (256-128-64-32) achieved the best overall RMSE, while the three-layer MLP (128-64-32) showed substantial improvement over the single-layer variant (R^2 = 0.900 vs 0.890), indicating that deeper architectures can better capture complex pollutant-AQI relationships [4] [7]. However, the simpler MLP (64 neurons) lagged notably, suggesting that shallow architectures may not fully exploit the nonlinear patterns in the data [8]. Neural networks require substantial data to fully exploit their representational power; while our dataset was substantial, the relatively small number of features (six pollutants) may have limited the advantage of very deep architectures over tree-based ensembles [6].

Interestingly, the CNN–MLP hybrid and Physics-Guided NN models underperformed compared to the best MLP and Random Forest results. This could be due to the absence of strong spatial or temporal dependencies in the tabular dataset that would benefit convolutional architectures [7]. The

Physics-Guided NN's physical constraints (0-500 AQI bounds) provided minimal benefit, likely because the tabular pollutant data lacks the explicit physical structure that makes domain-guided approaches most effective.

Feature Importance and Model Interpretability: Random Forest feature importance analysis revealed that particulate matter (PM2.5 and PM10) contributed most significantly to AQI predictions, followed by NO2 and O3, consistent with established understanding that particulate matter is the primary driver of health-based air quality indices. The ensemble's built-in feature ranking capability provides valuable insights for environmental monitoring prioritization.

Practical Performance Interpretation: The achieved RMSE values of 7.72-7.79 for the best performing models represent approximately 1.5% error on the 0-500 AQI scale. This level of precision ensures accurate AQI category classification in most scenarios, with occasional misclassification only occurring near category boundaries during transitional pollution conditions. While the Deep MLP achieved marginally better RMSE, the performance gain (0.07 RMSE units) comes at the cost of increased model complexity and reduced interpretability.

From a practical deployment standpoint, Random Forest offers the optimal balance between accuracy, interpretability, and computational efficiency for real-time forecasting applications. The high R^2 values across most models ($>$0.90) indicate that pollutant concentrations explain a large proportion of AQI variability, reinforcing the importance of continuous pollutant monitoring for accurate air quality forecasting [1] [2].

In summary, the experimental results demonstrate that both ensemble methods and deep neural networks can achieve high accuracy for AQI prediction from pollutant concentrations. While the Deep MLP achieved the lowest RMSE (7.72), Random Forest provided nearly identical performance with superior interpretability, robustness, and lower computational requirements. The findings highlight that for operational air quality monitoring systems, Random Forest represents the most practical choice, while deep learning approaches become competitive when enhanced features or specialized architectures are employed.

V. Conclusion

This study benchmarked linear (LR), nonparametric/kernel (KNN, SVR), tree ensembles (RF, GBR), deep neural networks (MLP), and domain-informed architectures (CNN–MLP, PGNN) for AQI prediction from pollutant concentrations across six globally distributed cities. The Deep MLP Regressor (256-128-64-32) achieved the best overall performance (R^2 = 0.910, RMSE = 7.72; Table II), closely followed by Random Forest (R^2 = 0.910, RMSE = 7.79; Table I) and KNN (R^2 = 0.909; Table I). While deeper neural networks achieved marginally superior accuracy, they required extensive architecture design and hyperparameter tuning, whereas Random Forest delivered comparable results with greater robustness and computational efficiency. Gradient Boosting and SVR performed well but trailed the top performers under the reported settings (Table I). The CNN–MLP hybrid and PGNN did not yield performance gains in this dataset, consistent with the relatively weak spatial and explicit physical structure encoded in the tabular pollutant inputs.

Feature importance analysis revealed that particulate matter (PM2.5 and PM10) dominated AQI predictions, reinforcing established environmental health knowledge about the critical role of particulate pollution in air quality assessment. The achieved RMSE values of 7.72-7.79 represent approximately 1.5% error on the 0-500 AQI scale, ensuring reliable category classification for practical air quality monitoring applications.

Overall, for cross-continental AQI prediction using tabular pollutant data, Random Forest emerges as the most practical choice, offering optimal balance between accuracy, interpretability, and operational feasibility. Deep learning approaches become competitive when specialized architectures or enhanced spatiotemporal features are available, but require careful consideration of the complexity-performance trade-off for deployment in real-world environmental monitoring systems.

A. Future Work

- **Multi-step temporal forecasting**
  Extend current single-step predictions to 3, 7, and 30-day AQI forecasts using sequence-to-sequence architectures and temporal attention mechanisms for long-term environmental planning.
- **Spatiotemporal modeling with GNNs**
  Implement Graph Neural Networks to model inter-city pollution transfer and regional atmospheric dynamics, capturing spatial dependencies between geographically connected urban areas.
- **Satellite data integration**
  Fuse NASA MODIS and Sentinel-5P satellite observations with ground-based measurements to enhance spatial coverage and validate model predictions across broader geographic regions.
- **Probabilistic uncertainty quantification**
  Develop Bayesian Neural Networks or ensemble-based approaches to provide prediction intervals alongside point estimates, enabling risk-aware decision making for public health alerts.
- **Real-time deployment optimization**
  Optimize models for edge computing deployment on IoT sensors with resource constraints, enabling distributed real-time AQI predictions with minimal latency requirements.
- **Policy impact simulation**
  Develop counterfactual analysis frameworks to simulate AQI responses under different emission reduction policies, supporting evidence-based environmental regulation.

REFERENCES

[1] World Health Organization, *Air Pollution Facts*, 2024. [Online]. Available: https://www.who.int/health-topics/air-pollution

[2] European Environment Agency, *European Air Quality Index Methodology*, 2023. [Online]. Available: https://www.eea.europa.eu/themes/air/air-quality-index

[3] Y. Elebiary, *Air Quality 2024*, Kaggle, 2024. [Online]. Available: https://www.kaggle.com/datasets/youssefelebiary/air-quality-2024/data

[4] J. Smith, R. Brown, and L. Johnson, "Machine Learning Approaches for Environmental Forecasting," *Environmental Modelling & Software*, vol. 156, 2022.

[5] Chen, M.-H., Chen, Y.-C., Chou, T.-Y., & Ning, F.-S. (2023). PM2.5 Concentration Prediction Model: A CNN-RF Ensemble Framework. *International Journal of Environmental Research and Public Health*, 20(4077). https://doi.org/10.3390/ijerph20054077

[6] Gupta, N. S., Mohta, Y., Heda, K., Armaan, R., Valarmathi, B., & Arulkumaran, G. (2023). Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis. *Journal of Environmental and Public Health*, Article ID 4916267. https://doi.org/10.1155/2023/4916267

[7] Alkabbani, H., Ramadan, A., Zhu, Q., & Elkamel, A. (2022). An Improved Air Quality Index Machine Learning-Based Forecasting with Multivariate Data Imputation Approach. *Atmosphere*, 13(1144). https://doi.org/10.3390/atmos13071144

[8] Liu, Q., Cui, B., & Liu, Z. (2024). Air Quality Class Prediction Using Machine Learning Methods Based on Monitoring Data and Secondary Modeling. *Atmosphere*, 15(553). https://doi.org/10.3390/atmos15050553

[9] Méndez, M., Merayo, M. G., & Núñez, M. (2023). Machine Learning Algorithms to Forecast Air Quality: A Survey. *Artificial Intelligence Review*, 56, 10031–10066. https://doi.org/10.1007/s10462-023-10424-4

[10] European Environment Agency, "Health impacts of air pollution in Europe, 2022," EEA Report, 2022.

[11] C. Schollaert *et al.*, "Air Quality Impacts of the January 2025 Los Angeles Wildfires: Insights from Public Data Sources," *Environ. Sci. Technol. Lett.*, 2025.

[12] US EPA, "Research on Health Effects from Air Pollution," 2025. [Online]. Available: https://www.epa.gov/air-research/research-health-effects-air-pollution

[13] F. Galton, "Regression towards mediocrity in hereditary stature," *The Journal of the Anthropological Institute of Great Britain and Ireland*, vol. 15, pp. 246–263, 1886.

[14] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.

[15] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, no. 3, pp. 199–222, 2004.

[16] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.

[17] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001.

[18] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[19] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[20] X. Jia, J. Willard, A. Karpatne, J. Read, J. Zwart, M. Steinbach, and V. Kumar, "Physics-guided machine learning for scientific discovery: An application in simulating lake temperature profiles," *ACM/IMS Transactions on Data Science*, vol. 2, no. 3, pp. 1–26, 2021.