Journal of Current Science and Technology, January-March 2026 Vol. 16 No. 1, Article 148

ISSN 2630-0656 (Online)

Cite this article: Kuruvilla, J., & Srividhya, V. (2026). Accurate air quality index prediction using MPRKDNN with optimized feature selection. *Journal of Current Science and Technology*, *16*(1), Article 148. https://doi.org/10.59796/jcst.V16N1.2026.148

Journal of Current Science and Technology

Journal homepage: https://jcst.rsu.ac.th

Accurate Air Quality Index Prediction Using MPRKDNN with Optimized Feature Selection

Jiss Kuruvilla[1,2,*], and V. Srividhya[1]

[1]Department of Computer Science, Avinashilingam Institute for Home Science and Higher Education for Women Coimbatore, India

[2]Department of Computer Applications, Muthoot Institute of Technology and Science, Kochi, Kerala, India

*Corresponding author; E-mail: jiss.kuruvilla@gmail.com

Received 26 May 2025; Revised 6 August 2025; Accepted 6 August 2025; Published online 20 December 2025

Abstract

Air quality forecasting is essential for managing environmental and health impacts in rapidly urbanizing regions. The AQI short for Air Quality Index is a standardized measure used to communicate the severity of air pollution based on several pollutant indicators. However, accurately classifying AQI levels remains challenging due to the highly irregular nature of real-world datasets, which often include missing values, noise and redundant variables. Prediction accuracy largely depends on algorithmic complexity and the quality of input data preparation and refinement. To overcome these practical data-related limitations and improve AQI classification, a structured and adaptive model design becomes necessary. This study presents a modular and organized learning framework, Multivariate Piecewise Radial Kernelized Deep Neural Network (MPRKDNN), designed to enhance AQI classification through intelligent preprocessing and targeted feature selection. This process involves estimating missing values using Multivariate Piecewise Constant Interpolation (MPCI) and detecting outliers using the Tietjen-Moore statistical test. Radial Basis Kernelized Quadratic Discriminant Analysis (RBK-QDA) is used to retain the relevant variables and reduce dimensionality. The final output is fed into a deep feed-forward neural network trained using Stochastic Gradient Descent (SGD) for final classification.

The model is evaluated using multicity AQI datasets from India during 2017 to 2023. Comparative studies conducted against baseline deep learning and hybrid models show that MPRKDNN consistently improves classification accuracy, reduces RMSE, and maintains computational efficiency. These results emphasize the importance of integrating structured data preprocessing and kernel-based feature selection to enhance the robustness and interpretability of the AQI prediction system.

Keywords: *air quality index prediction; deep neural network; feature selection; MPCI; RBK-QDA; Tietjen–Moore Test*

1. Introduction

Accurate air quality forecasting has become increasingly important in the context of growing urbanization and its associated environmental challenges. The Air Quality Index (AQI) combines multiple pollutant concentrations such as PM2.5, PM10, and NO_2 into a standardized scale to inform regulatory responses and public awareness (Wu et al., 2024). While several predictive models have been developed, the real-world challenges such as data gaps, outliers, and underlying pollutant effects sometimes limit their effectiveness. Prior work on multi-year PM2.5 and PM10 modeling in Bangkok demonstrated that incorporating key environmental predictors into statistical and seasonal regression frameworks can enhance forecasting accuracy (Ruktamatakul et al., 2025; Sirisumpun et al., 2023).

Previously, models such as ARIMA, SVM, and Random Forest were used for AQI classification and regression. While these models are interpretable, they often struggle to generalize effectively to high-dimensional, nonlinear environmental data (Natarajan et al., 2024). In recent years, the newer studies mostly rely on Deep Learning (DL) and hybrid frameworks, which are more capable of capturing spatiotemporal and multivariate relationships (Wang et al., 2022a). However, in many of these studies, preprocessing and feature selection are either minimally addressed or fully integrated into the network, leading to models that may be less interpretable, slower to train, or prone to overfitting.

Contrary to the assumption that preprocessing and feature selection are always implicit in modern DL architectures, studies show that explicit, domain-aware data refinement improves model generalization and robustness especially in heterogeneous and noisy datasets (Maltare & Vahora, 2023; Wu et al., 2023). For instance, kernelized discriminant techniques and statistical outlier filters have shown measurable improvements in AQI forecasting accuracy.

This study proposes a modular framework, the Multivariate Piecewise Radial Kernelized Deep Neural Network (MPRKDNN), which explicitly incorporates structured preprocessing and feature selection as distinct learning stages. The model integrates Multivariate Piecewise Constant Interpolation (MPCI) for imputing missing values, Tietjen–Moore (TM) testing for outlier detection, and Radial Basis Kernelized Quadratic Discriminant Analysis (RBK-QDA) for feature ranking. A deep neural classifier then assigns the refined inputs to AQI classes across six predefined categories.

The model is validated using multicity AQI data from 2017 to 2023. Evaluation against deep and hybrid baselines highlights the improved performance and interpretability of the proposed method, underscoring the importance of a structured approach to preprocessing and feature selection in environmental data modeling.

1.1 Traditional and Machine Learning Approaches

Previously, AQI prediction was often based on conventional ML models. Decision Trees (DT) and Support Vector Machines (SVM) were used for classification tasks, particularly in studies incorporating optimization techniques such as Termite Fly Optimization and Grey Wolf Optimization to enhance prediction accuracy (Natarajan et al., 2024; Rani & Sampathkumar, 2023). ARIMA, SVR, and Random Forests (RF) were used for regression-based AQI prediction (Imam et al., 2024). These models provided good accuracy, however sometimes faced challenges with non-linearity and scalability. SVM classifiers mostly worked on small to medium datasets but lacked strength in complex environmental scenarios (Janarthanan et al., 2021). Linear models and basic ensemble methods were also tested in spatiotemporal forecasting contexts (Anggraini et al., 2024; Hong et al., 2021). However, these models often exhibited poor generalization in heterogeneous conditions in spite of being effective in certain structured scenarios.

1.2 Deep Learning Models for AQI Prediction

Deep Learning (DL) models especially LSTMs and GRUs gained importance for managing the nonlinear and temporal complexity of AQI data. ADNNet, an attention-based neural network, demonstrated notable gains in classification accuracy (Wu et al., 2024). Comparative evaluations of LSTM, GRU, and their combinations revealed contextual advantages in various forecasting tasks (Zhou et al., 2023; Feng & Zhang, 2023). Bidirectional LSTMs further enhanced sequence learning (Arockia Panimalar & Krishnakumar, 2023), while GRUs offered computational efficiency without major trade-offs in accuracy (Rajakumari & Lanjewar, 2023). Nevertheless, pure DL models often act as black boxes, limiting interpretability and underutilizing domain-specific pollutant relationships.

1.3 Hybrid Frameworks and Data-Aware Modeling

Hybrid models have emerged to improve both adaptability and interpretability. Reg-CLSTM and CNN-ILSTM combined convolutional and recurrent layers to capture spatiotemporal dependencies (Manna & Anitha, 2024; Wang et al., 2022a). Enhanced methods, such as attention-guided LSTMs with QPSO (Nguyen et al., 2024), GAN-based AQI synthesis (Binbusayyis et al., 2024), and CNN-AGU frameworks (Wang et al., 2022b) improved prediction precision. Informer-XGBoost (Shao et al., 2023) and ISSA-LSTM hybrids (Wu et al., 2023) addressed uncertainty and enhanced resolution. While powerful, these models are data-intensive and computationally demanding, posing deployment challenges in real-time settings.

1.4 Feature Selection in AQI Prediction

Effective feature selection is critical in reducing input dimensionality and enhancing model performance. Kernel-based methods, such as RBK-QDA, have proven useful in selecting discriminative pollutant and meteorological variables (Maltare & Vahora, 2023). Other approaches include Stacked Autoencoders (Wang et al., 2021), CEEMDAN-

ARMA decompositions (Sun & Liu, 2022), and IoT-integrated noise modeling (Alnowaiser et al., 2024). Hybrid strategies combining Kernel PCA and mutual information (Liu et al., 2022) or wrapper-based feature elimination (Gopu & Kannan, 2023) further improved generalization and reduced training overhead. Studies also affirm the value of combining filter, wrapper, and embedded methods in deep learning pipelines (Arafin et al., 2025), yielding more interpretable and computationally efficient AQI models. Recent evaluations of statistical feature selection methods such as MRMR, Chi², ReliefF, ANOVA, and Kruskal–Wallis have further shown gains in classification accuracy and reduced model build time (Pechprasarn et al., 2025).

Despite significant progress in deep and hybrid AQI models, most existing studies still embed preprocessing and feature selection implicitly within the network. This limits interpretability, increases training cost, and reduces robustness when applied to heterogeneous real-world datasets containing noise, outliers, and missing values. Therefore, there is a need for a modular AQI prediction framework that explicitly separates preprocessing, feature refinement, and classification to improve generalization and computational efficiency. The proposed MPRKDNN addresses this gap by integrating MPCI, Tietjen–Moore testing, and RBK-QDA as structured components prior to deep learning classification.

2. Objectives of the Study

The primary objective of this study is to develop a structured AQI prediction model that integrates domain-aware preprocessing, effective feature selection, and deep learning classification. Specifically, the study aims to:

- Design a modular framework (MPRKDNN) to improve classification accuracy for AQI categories.
- Apply Multivariate Piecewise Constant Interpolation (MPCI) to handle missing data and enhance data completeness.
- Integrate Tietjen–Moore statistical testing to detect and remove outliers from pollution data.
- Employ Radial Basis Kernelized Quadratic Discriminant Analysis (RBK-QDA) to reduce dimensionality and retain relevant features.
- Evaluate the model using real-world AQI datasets from Indian cities over multiple years, and benchmark it against traditional and deep learning baselines.

These objectives collectively aim to enhance the robustness, interpretability, and scalability of AQI prediction systems for practical environmental monitoring contexts.

3. Methodology

The proposed Multivariate Piecewise Radial Kernelized Deep Neural Network (MPRKDNN) is a structured pipeline designed to enhance AQI prediction by incorporating intelligent data preprocessing, kernel-based feature selection, and optimized classification. The framework consists of the following four modules:

- Preprocessing: Handling missing values and outliers to improve data quality
- Feature Selection: Extracting the most discriminative input variables
- Classification: Employing a deep neural network trained on refined features
- Evaluation: Assessing performance using multiclass classification metrics

This modular pipeline ensures interpretability, generalization, and computational efficiency across varied AQI conditions.

3.1 Preprocessing

Environmental datasets often contain missing values and outliers caused by sensor malfunctions, weather disruptions, or data transmission errors. To address these issues, a two-step preprocessing strategy is adopted.

3.1.1 Missing Value Imputation using Multivariate Piecewise Constant Interpolation (MPCI)

Multivariate Piecewise Constant Interpolation (MPCI) estimates missing values by capturing local continuity across variables and time. Compared to traditional imputation techniques such as mean or k-NN, MPCI preserves the original distribution and structural patterns of pollutants and meteorological features.

3.1.2 Outlier Detection using Tietjen–Moore (TM) Test:

Outliers are detected using the Tietjen–Moore (TM) test, which statistically identifies and removes multi-point anomalies from the dataset. This ensures that the learning model is not misled by spurious fluctuations or sensor faults.

3.2 Feature Selection Using Radial Basis Kernelized QDA (RBK-QDA)

High-dimensional input spaces often lead to overfitting and increase computation time in deep learning models. To address this, feature selection is applied to retain only the most relevant variables while eliminating redundant or noisy dimensions. This study employs a nonlinear, class-aware feature selection strategy Radial Basis Kernelized Quadratic Discriminant Analysis (RBK-QDA) (Maltare & Vahora, 2023). Earlier applications of statistical feature selection in classification tasks have shown improvements in predictive accuracy and stability (Pechprasarn et al., 2024), which supports the inclusion of a discriminative selection stage in this work.

Traditional QDA performs well when features follow normal distributions and possess distinct covariances, but such assumptions are often violated in environmental data. To overcome this, a radial basis function (RBF) kernel is used, through which input features are projected into a higher-dimensional space. In this transformed space, the separation between AQI classes is made more tractable.

The kernel transformation is defined in Eq. (1) as:

$$\phi(x) = \exp\left(-\frac{\|x - x_i\|^2}{2\sigma^2}\right) \tag{1}$$

where:

$\phi(x)$: The transformed version of the input vector x in the kernel space.

x: The input data vector (a row from the dataset).

x_i: A reference sample or another data point in the dataset.

$\|x\text{-}x_i\|^2$: Squared Euclidean distance between vectors x and x_1.

σ: The bandwidth (spread) parameter of the Radial Basis Function (RBF) kernel; it controls the smoothness of the transformation.

exp: Exponential function.

For each class c with N_c samples, the mean and covariance in the kernel space are computed using Eq. (2) and Eq. (3), respectively:

$$\mu_c = \frac{1}{N_c} \sum_{y_i = c} \phi(x_i) \tag{2}$$

and

$$\Sigma_c = \frac{1}{N_c - 1} \sum_{y_i = c} \left(\phi(x_i) - \mu_c\right)\left(\phi(x_i) - \mu_c\right)^{\mathrm{T}} \tag{3}$$

where:

μ_c: Mean vector of all transformed inputs belonging to class c.

Σ_c: Covariance matrix of the transformed inputs for class c.

N_c: Total number of data points in class c.

x_i: Data samples where the class label $y_i = c$.

$\phi(x_i)$: Kernel-transformed version of input x_i.

$(.)^{\mathrm{T}}$: Transpose of a vector/matrix.

A new sample z is classified using the discriminant function in Eq. (4):

$$g_c(z) = -\,0.5 \log|\Sigma_c| - 0.5\left(\phi(z)\text{-}\mu_c\right)^{\mathrm{T}} \Sigma_c^{-1} \left(\phi(z) - \mu_c\right) + \log \pi_c \tag{4}$$

where:

$g_c(z)$: Discriminant score for sample z under class c.

$\phi(z)$: Kernel-transformed version of new test input z.

μ_c: Mean of class c in the kernel space.

$|\Sigma_c|$: Covariance matrix of class c in the kernel space.

$|\Sigma_c^{-1}|$: Determinant of the covariance matrix Σ_c

Σ_c^{-1}: Inverse of the covariance matrix.

π_c: Prior probability of class c_i i.e., proportion of class c in the training set.

log: Natural logarithm function.

Features contributing most significantly to the separation of class-wise discriminant scores are retained. RBK-QDA thus reduces input dimensionality while enhancing predictive power. The proposed structure of the MPRKDNN pipeline is illustrated in Figure 1.

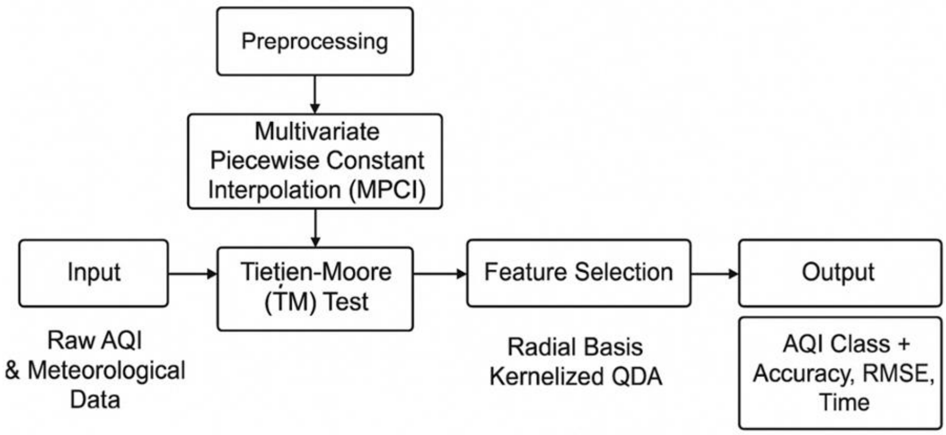

Figure 1 Structural framework of the MPRKDNN

3.3 Deep Neural Network Classification

The selected features are passed through a deep feedforward neural network configured for multiclass AQI classification (Wu et al., 2024; Binbusayyis et al., 2024). Multiple hidden layers are included, each activated by ReLU functions, with dropout applied for regularization. The softmax activation function is employed in the final layer to compute class probabilities across six AQI bands. The network is trained using stochastic gradient descent (SGD) (Tian et al., 2023), minimizing cross-entropy loss to ensure convergence and maintain generalization.

The SGD update rule is given in Eq. (5):

$$w_{t+1} = w_t - \eta \nabla L(w_t) + \gamma (w_t - w_{t-1}) \quad (5)$$

where:

w_t is the weight vector at iteration t
η is the learning rate
γ is the momentum coefficient
$\nabla L(w_t)$ is the gradient of the loss with respect to the weights

The **categorical cross-entropy** loss used for multi-class classification is defined in Eq. (6):

$$L = -\sum_{i=1}^{N} \sum_{c=1}^{C} y_{ic} \log(\widehat{y_{ic}}) \quad (6)$$

where**:**

N: Total number of training instances.
C: Number of classes.
y_{ic}: Ground-truth indicator (1 if instance i belongs to class c, else 0).
$\widehat{y_{ic}}$: Predicted probability that instance i belongs to class c.

The output layer employs a softmax activation function to estimate class probabilities, enabling multi-class classification across AQI categories. Once trained, the model's performance is evaluated using classification accuracy, root mean square error (RMSE), and total inference time-offering a holistic view of its predictive capability and computational efficiency.

3.4 Experimental Setup

To evaluate the proposed MPRKDNN framework, a series of experiments were conducted using real-world air quality data. The goal was to assess the effectiveness of the model in accurately predicting AQI categories under varied conditions and to compare its performance against selected benchmark approaches. The dataset used in this study comprises daily air quality measurements gathered from multiple Indian cities spanning 2017 to 2023. The data were sourced from official platforms such as the Central Pollution Control Board (CPCB) and supplemented with records from open repositories like Kaggle.

After data cleaning and preprocessing to remove incomplete or inconsistent records, the final dataset was structured into three subsets to facilitate model evaluation:

- Set A: Approximately 8,500 records from 2021
- Set B: Approximately 25,000 records from 2019 to 2021
- Set C: Over 50,000 records representing spanning 2017 to 2023

The output variable is the AQI category, classified into six standard bands: Good, Satisfactory, Moderate, Poor, Very Poor, and Severe. All experiments were conducted in Python using TensorFlow and scikit-learn libraries. The models were evaluated using classification accuracy, root mean square error (RMSE), and average inference time per instance to ensure consistent comparisons across various baselines.

4. Results and Discussion

This section presents the empirical validation of the proposed MPRKDNN model. The evaluation focuses on classification accuracy, error rate, prediction efficiency, and comparative performance against four established models: CNN-LSTM, ADNNet, QDA-SVM, and Random Forest. Experiments across three data volumes assessed model scalability and robustness. The results show that combining structured data preparation with effective feature selection can improve AQI classification performance across diverse datasets.

4.1 Accuracy Comparison

Classification accuracy reflects how reliably the model predicts the AQI category for a given day, based on the multivariate input of pollutants and weather variables. This task involves six standard AQI classes: Good, Satisfactory, Moderate, Poor, Very Poor, and Severe. Multiclass classification is challenging in this context due to overlapping pollution bands and inherent class imbalance.

As summarized in Table 1, MPRKDNN demonstrated superior performance across all dataset configurations. On the largest dataset (Set C), it achieved an accuracy of 94.6%, surpassing CNN-LSTM (91.7%), ADNNet (90.5%), and classical models like QDA-SVM (84.2%) and RF (82.7%). The performance advantage of MPRKDNN increased with dataset size, indicating strong generalization capability under varied conditions. The accuracy improvement is likely due to the way the selected features helped the model distinguish between AQI categories, even when the pollution levels were close or overlapping. This trend is illustrated in Figure 2, where the accuracy of all models is plotted across datasets. MPRKDNN's robust feature selection pipeline and non-linear decision boundaries allow it to better separate AQI categories, even in overlapping or imbalanced classes.

4.2 RMSE Comparison

Root Mean Square Error (RMSE) reflects the magnitude of prediction errors. Lower RMSE values indicate more stable, less error-prone predictions. As shown in Table 2, MPRKDNN records the lowest RMSE across all datasets. The RMSE gradually decreases from 0.198 to 0.175 as dataset size increases, further confirming the model's robustness.

MPRKDNN demonstrated steady RMSE improvement as the dataset size increased, suggesting its ability to handle more data without overfitting. This improvement is likely due to effective input cleaning and simplification through missing value handling and feature selection, enabling the model to learn from more reliable data. In contrast, the traditional models exhibited only minor changes in RMSE.

Table 1 Classification accuracy (%) across dataset subsets

Model	Set A (2021)	Set B (2019–21)	Set C (2017–23)
MPRKDNN	93.2	94.1	94.6
CNN-LSTM	90.8	91.5	91.7
ADNNet	89.3	90.1	90.5
QDA-SVM	84.6	85.2	84.2
RF	82.3	83.5	82.7

Table 2 RMSE values for each model across dataset scales

Model	Set A	Set B	Set C
MPRKDNN	0.198	0.183	0.175
CNN-LSTM	0.243	0.226	0.219
ADNNet	0.255	0.241	0.232
QDA-SVM	0.296	0.281	0.293
RF	0.310	0.305	0.310

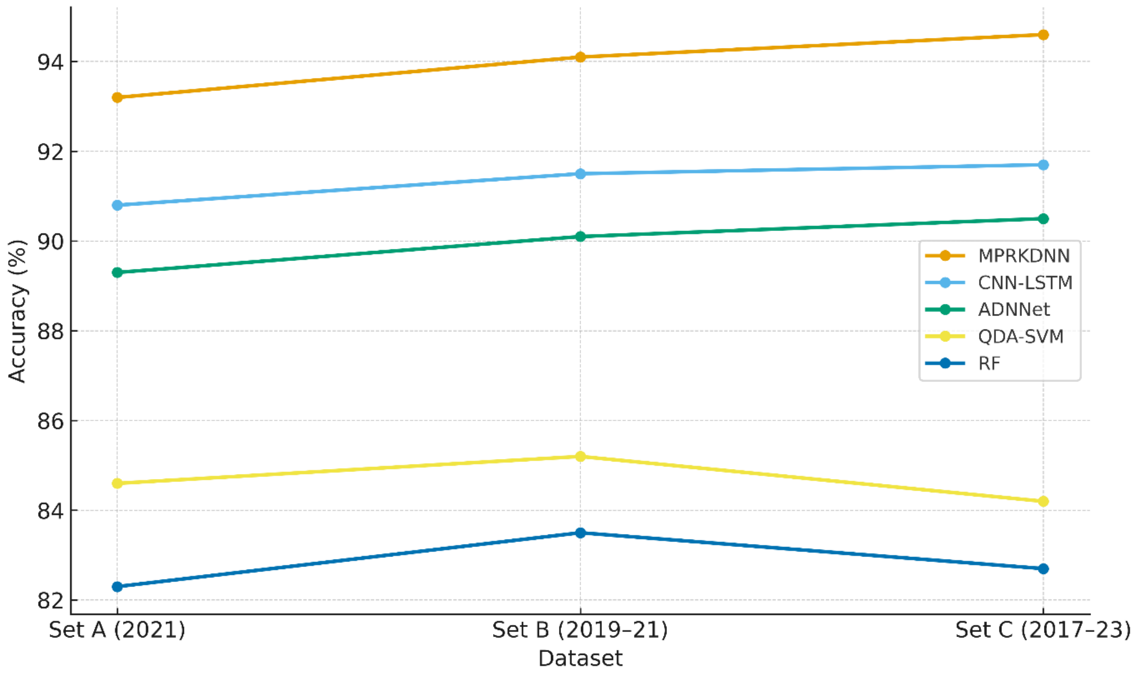

Figure 2 Accuracy of models across Set A, Set B, and Set C

4.3 Prediction Time

Inference time is a critical metric for real-time AQI prediction systems. Table 3 presents the average inference time per record, calculated using Set C. Although RF and QDA-SVM were faster, they performed worse in terms of accuracy and RMSE. MPRKDNN effectively balances speed and accuracy effectively, requiring only 1.73 milliseconds per prediction-making it suitable for real-time AQI systems.

Table 3 Average inference time (ms) per sample

Model	Time (ms)
MPRKDNN	1.73
CNN-LSTM	2.15
ADNNet	2.01
QDA-SVM	0.89
RF	0.67

Although traditional models like Random Forest (RF) and QDA-SVM demonstrated faster inference times, their predictive performance was noticeably lower. In contrast, the proposed MPRKDNN framework strikes a favorable balance between computational efficiency and classification accuracy, making it more suitable for practical deployments where both speed and precision are essential. Inference time remained consistent across different dataset sizes. The values reported in Table 3 correspond to Set C (the full 7-year dataset), as the time variations observed for Sets A and B were minimal and did not significantly affect the overall trend.

In summary, the model gave strong accuracy, lower error, and quick prediction times, which are all useful for practical AQI systems. These improvements stem from effective data cleaning and simplification before training including handling missing values, removing outliers, and selecting the most useful features. This allowed the model to focus on higher-quality inputs, resulting in more reliable predictions across datasets.

5. Conclusion

This study proposed a novel framework-MPRKDNN for AQI classification, centered on structured preprocessing and kernel-based feature selection. By incorporating MPCI and the Tietjen-Moore test, the framework ensures data consistency prior to model training. RBK-QDA further enhances input quality by selecting the most informative features. The model was tested across datasets of increasing size and complexity, consistently outperforming classical and deep learning baselines. It demonstrated reliable generalization, maintained low error margins, and offered near-real-time prediction speed.

These findings underscore the importance of careful preprocessing and feature refinement in environmental forecasting. MPRKDNN represents a step forward in creating interpretable and accurate systems for large-scale AQI monitoring. Future enhancements may include integrating spatiotemporal modeling and sensor reliability estimation to increase the system's adaptability to dynamic urban environments.

6. Abbreviations

Abbreviation	Full Term
AQI	Air Quality Index
MPRKDNN	Multivariate Piecewise Radial Kernelized Deep Neural Network
MPCI	Multivariate Piecewise Constant Interpolation
TM	Tietjen–Moore Test
RBK-QDA	Radial Basis Kernelized Quadratic Discriminant Analysis
RBF	Radial Basis Function
DNN	Deep Neural Network
SGD	Stochastic Gradient Descent
RMSE	Root Mean Square Error
ReLU	Rectified Linear Unit
RF	Random Forest
LSTM	Long Short-Term Memory
GRU	Gated Recurrent Unit
SVM	Support Vector Machine
ML	Machine Learning
DL	Deep Learning

7. CRediT Statement

Jiss Kuruvilla: Conceptualization, Methodology, Software, Formal Analysis, Investigation, Data Curation, Visualization, Writing – Original Draft, Writing – Review & Editing.
V. Srividhya: Supervision, Validation, Writing – Review & Editing, Project Administration.

8. References

Arafin, S. K., Md Ghani, N. A., Rosli, M. M., & Ibrahim, N. (2025). AdjcorT-RBFNN for air quality classification: Mitigating multicollinearity with real and simulated data. *Environment and Natural Resources Journal, 23*(3), 242–255. https://doi.org/10.32526/ennrj/23/20250006

Arockia Panimalar, S., & Krishnakumar, A. (2023). A review of churn prediction models using different machine learning and deep learning approaches in cloud environment. *Journal of Current Science and Technology*, *13*(1), 136–161. https://doi.org/10.14456/jcst.2023.12

Alnowaiser, K., Alarfaj, A. A., Alabdulqader, E. A., Umer, M., Cascone, L., & Alankar, B. (2024). IoT based smart framework to predict air quality in congested traffic areas using SV-CNN ensemble and KNN imputation model. *Computers and Electrical Engineering*, *118*, Article 109311. https://doi.org/10.1016/j.compeleceng.2024.109311

Anggraini, T. S., Irie, H., Sakti, A. D., & Wikantika, K. (2024). Machine learning-based global air quality index development using remote sensing and ground-based stations. *Environmental Advances*, *15*, Article 100456. https://doi.org/10.1016/j.envadv.2023.100456

Binbusayyis, A., Khan, M. A., Ahmed A, M. M., & Emmanuel, W. S. (2024). A deep learning approach for prediction of air quality index in smart city. *Discover Sustainability*, *5*(1), Article 89. https://doi.org/10.1007/s43621-024-00272-9

Feng, H., & Zhang, X. (2023). A novel encoder-decoder model based on Autoformer for air quality index prediction. *Plos One*, *18*(4), Article e0284293. https://doi.org/10.1371/journal.pone.0284293

Gopu, K. L. K., & Kannan, S. (2023). Deep graph neural network with fish-inspired task allocation algorithm for heart disease diagnosis. *Journal of Current Science and Technology, 13*(2), 392–411. https://doi.org/10.59796/jcst.V13N2.2023.1753

Hong, H., Jeon, H., Youn, C., & Kim, H. (2021). Incorporation of shipping activity data in recurrent neural networks and long short-term memory models to improve air quality predictions around busan port. *Atmosphere*, *12*(9), Article 1172. https://doi.org/10.3390/atmos12091172

Imam, M., Adam, S., Dev, S., & Nesa, N. (2024). Air quality monitoring using statistical learning models for sustainable environment. *Intelligent Systems with Applications*, *22*, Article 200333. https://doi.org/10.1016/j.iswa.2024.200333

Janarthanan, R., Partheeban, P., Somasundaram, K., & Elamparithi, P. N. (2021). A deep learning approach for prediction of air quality index in a metropolitan city. *Sustainable Cities and Society*, *67*, Article 102720. https://doi.org/10.1016/j.scs.2021.102720

Liu, Z., Han, H.-G., Dong, L.-X., Yang, H.-Y., & Qiao, J.-F. (2022). Intelligent decision method of sludge bulking using recursive kernel principal component analysis and Bayesian network. *Control Engineering Practice*, *121*, Article 105038. https://doi.org/10.1016/j.conengprac.2021.105038

Maltare, N. N., & Vahora, S. (2023). Air quality index prediction using machine learning for ahmedabad city. *Digital Chemical Engineering*, *7*, Article 100093. https://doi.org/10.1016/j.dche.2023.100093

Manna, T., & Anitha, A. (2024). Hybridization of rough set–wrapper method with regularized combinational LSTM for seasonal air quality index prediction. *Neural Computing and Applications*, *36*(6), 2921-2940. https://doi.org/10.1007/s00521-023-09220-6

Natarajan, S. K., Shanmurthy, P., Arockiam, D., Balusamy, B., & Selvarajan, S. (2024). Optimized machine learning model for air quality index prediction in major cities in India. *Scientific Reports*, *14*(1), Article 6795. https://doi.org/10.1038/s41598-024-54807-1

Nguyen, A. T., Pham, D. H., Oo, B. L., Ahn, Y., & Lim, B. T. (2024). Predicting air quality index using attention hybrid deep learning and quantum-inspired particle swarm optimization. *Journal of Big Data*, *11*(1), Article 71. https://doi.org/10.1186/s40537-024-00926-5

Pechprasarn, S., Srisaranon, N., & Yimluean, P. (2025). Optimizing diabetes prediction: An evaluation of machine learning models through strategic feature selection. *Journal of Current Science and Technology*, *15*(1), Article 75. https://doi.org/10.59796/jcst.V15N1.2025.75

Pechprasarn, S., Suechoey, N., Pholtrakoolwong, N., Tanedvorapinyo, P., & Toboonliang, Y. (2024). Optimizing lung cancer diagnosis with machine learning and feature selection methods. *Journal of Current Science and Technology*, *14*(3), Article 55. https://doi.org/10.59796/jcst.V14N3.2024.55

Peter, R. R. R., & Lanjewar, U. A. (2023). Improving air quality prediction with a hybrid Bi-LSTM and GAN Model. *ECTI Transactions on Computer and Information Technology*, 17(3), 376-386. https://doi.org/10.37936/ecti-cit.2023173.251670

Rani, P. A., & Sampathkumar, V. (2023). A novel artificial intelligence algorithm for predicting air quality by analysing the pollutant levels in air quality data in tamilnadu. *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, *5*, Article 100234. https://doi.org/10.1016/j.prime.2023.100234

Ruktamatakul, S., Kuntamanee, N., Puengsaychol, P., Ruktametakul, S., & Yimprayoon, P. (2025). Comparison of multiple linear regression and periodic models for estimating $PM_{2.5}$ and PM_{10} concentrations in Bangkok. *Journal of Current Science and Technology*, *15*(3), Article 117. https://doi.org/10.59796/jcst.V15N3.2025.117

Shao, Q., Chen, J., & Jiang, T. (2023). A novel coupled optimization prediction model for air quality. *IEEE Access*, *11*, 69667-69685. https://doi.org/10.1109/ACCESS.2023.3293249

Sirisumpun, N., Wongwailikhit, K., Painmanakul, P., & Vateekul, P. (2023). Spatio-temporal PM2. 5 forecasting in Thailand using encoder-decoder networks. *IEEE Access*, *11*, 69601-69613. https://doi.org/10.1109/ACCESS.2023.3293398

Sun, Y., & Liu, J. (2022). Aqi prediction based on ceemdan-arma-lstm. *Sustainability*, *14*(19), Article 12182. https://doi.org/10.3390/su141912182

Tian, Y., Zhang, Y., & Zhang, H. (2023). Recent advances in stochastic gradient descent in deep learning. *Mathematics*, *11*(3), Article 682. https://doi.org/10.3390/math11030682

Wang, B., Kong, W., & Zhao, P. (2021). An air quality forecasting model based on improved convnet and RNN. *Soft Computing*, *25*(14), 9209-9218. https://doi.org/10.1007/s00500-021-05843-w

Wang, J., Jin, L., Li, X., He, S., Huang, M., & Wang, H. (2022a). A hybrid air quality index prediction model based on CNN and attention gate unit. *IEEE Access*, *10*, 113343-113354. https://doi.org/10.1109/ACCESS.2022.3217242

Wang, J., Li, X., Jin, L., Li, J., Sun, Q., & Wang, H. (2022b). An air quality index prediction model based on CNN-ILSTM. *Scientific Reports*, *12*(1), Article 8373. https://doi.org/10.1038/s41598-022-12355-6

Wu, H., Yang, T., Li, H., & Zhou, Z. (2023). Air quality prediction model based on mRMR–RF feature selection and ISSA–LSTM. *Scientific Reports*, *13*(1), Article 12825. https://doi.org/10.1038/s41598-023-39838-4

Wu, X., Gu, X., & See, K. W. (2024). ADNNet: Attention-based deep neural network for Air Quality Index prediction. *Expert Systems with Applications*, *258*, Article 125128. https://doi.org/10.1016/j.eswa.2024.125128

Zhou, H., Wang, T., Zhao, H., & Wang, Z. (2023). Updated prediction of air quality based on kalman-attention-lstm network. *Sustainability*, *15*(1), Article 356. https://doi.org/10.3390/su15010356