Mahamudul Hasan, Sadab Jaowad,
Ronojith Saha Roy Dipta, Samia Shameem Haque,
Abir Hossain, Mainul Hasan,
Anisur Rahman, Mohammad Rifat Ahmmad Rashid,
Sawkat Ali

# DeepBiGRULSTM: ADVANCED DEEP LEARNING ARCHITECTURE FOR PRECISE AIR QUALITY FORECASTING AND MONITORING



**Abstract** *Air pollution remains one of the most critical environmental challenges of the 21st century, with severe implications for public health and urban sustainability. To address the need for accurate and timely air quality forecasting, this study introduces DeepBiGRULSTM, a hybrid deep-learning framework that integrates Bidirectional Gated Recurrent Units (Bi-GRU) with long short-term memory (LSTM) networks. The model leverages both real-time Nowcast and raw concentration values to enhance the reliability of short-term predictions. Using a comprehensive historical dataset from the U.S. Dhaka consulate, incorporating pollutant and meteorological variables, we benchmarked the proposed model against established architectures. such as GRU, LSTM, and Temporal Convolutional Networks (TCN). Experimental results demonstrate that DeepBiGRULSTM consistently outperforms the baselines, achieving the lowest errors with an RMSE of 0.4976 on the training set and 0.7548 on the test set. A key contribution of this work is its integration into a mobile application, providing real-time air quality updates and early warnings across Bangladesh. Beyond forecasting, the system supports health advisory services and can inform public policy on environmental management. Overall, this research delivers a practical and high-precision framework for air quality monitoring, with direct relevance to sustainable urban development and public health protection.*

**Keywords** air quality index (AQI), $PM_{2.5}$, deep learning, LSTM, bidirectional GRU (Bi-GRU), Temporal Convolutional Network (TCN)

**Citation** Computer Science 27(1) 2026: 43–64

## 1. Introduction

Rapid industrial expansion and urban growth have markedly degraded ambient air quality, with significant consequences for human health and overall well-being. Deteriorating air conditions have become a global public-health challenge, as emissions from transportation systems, industrial production, and densely populated settlements contribute to premature mortality on a massive scale. Recent epidemiological studies identify fine particulate matter as one of the top global-risk factors for mortality, with approximately 4.2 million premature deaths annually attributed to its exposure.

Urban centers worldwide are especially vulnerable, where the intensification of industrial processes and urbanization has amplified pollution levels. Bangladesh provides a critical case study in this regard. Dhaka, home to over 20 million inhabitants [28], exemplifies the severe challenges of rapidly expanding cities. According to World Health Organization (WHO) assessments, Bangladesh ranked second globally in terms of poor air quality in 2021 [17]. With a national population of around 165 million, and more than one-third already residing in cities, demographic forecasts suggest that the proportion of urban dwellers will grow substantially if current development trajectories continue [14]. The degradation of air quality in metropolitan areas is fueled by multiple drivers, including the expansion of motorized transport, industrial emissions, accelerated urban sprawl, and persistent reliance on traditional biomass fuels.

Given this context, accurate forecasting of air quality [32] has become an essential tool for environmental protection and public health policy [2]. Predictive systems enable evidence-based decision making for emission regulation, as well as the establishment of early warning mechanisms to safeguard vulnerable populations. To meet these needs, many urban regions have installed real-time monitoring infrastructures capable of measuring pollutants such as $PM_{2.5}$ and related atmospheric variables [10]. For instance, Rahman et al. (2024) introduced AirNet, a web-based platform capable of delivering real-time forecasts for over 23,000 cities using machine learning techniques, including Random Forest and decision tree, which demonstrated strong predictive performance [25]. Despite these advances, the accurate and timely detection of fluctuations in $PM_{2.5}$ concentrations remains a fundamental technical challenge for forecasting models.

Air pollution exerts adverse effects across both short- and long-term exposure windows, disproportionately affecting sensitive groups such as children and the elderly [18]. Acute episodes are linked to respiratory irritation, headaches, and allergic responses, while chronic exposure can result in cardiovascular disease, neurological impairment, cancer, and long-term respiratory illness [23]. Addressing these health impacts requires robust modeling approaches capable of capturing the complex, dynamic interactions between environmental and anthropogenic factors.

Research on air quality prediction has drawn from both statistical and machine learning paradigms. Early studies relied on regression-based models such as Lasso,

Ridge, and ElasticNet, alongside tree-based methods. including gradient boosting, decision trees, and Random Forests [6]. With advances in deep learning, however, more sophisticated architectures have been developed, including recurrent neural networks (RNNs), convolutional LSTMs, one-dimensional CNNs, Bidirectional GRUs, and transfer learning-based BLSTMs [7]. These models have shown superior performance in capturing temporal dependencies in time series data. Building on this trajectory, our study employs a novel hybrid approach that combines long short-term memory (LSTM) networks with Bidirectional GRUs to enhance predictive accuracy and robustness for air-quality index (AQI) estimation.

This research therefore contributes an integrated framework for air quality monitoring and forecasting, leveraging advanced deep -learning methods to address a pressing global concern. The proposed DeepBiGRULSTM model is rigorously validated against multiple baselines and deployed within a mobile application to provide real-time updates and early warnings across Bangladesh. Beyond forecasting, the system supports timely interventions and facilitates the identification of pollution sources, offering practical value for environmental governance and public health protection. The primary contributions of this study can be summarized as follows:

- We propose DeepBiGRULSTM, a hybrid deep-learning architecture that integrates LSTM and Bidirectional GRU layers to capture both long-term dependencies and bidirectional temporal patterns in air quality data.
- A comprehensive evaluation conducted using a large-scale data set from the U.S. Dhaka consulate incorporating meteorological and pollutant information, with rigorous comparisons against GRU, LSTM, and TCN baselines.
- The proposed model achieves state-of-the-art predictive accuracy, with an RMSE of 0.4976 (train) and 0.7548 (test), outperforming existing approaches.
- The model is integrated into a user-friendly mobile application, providing real-time air quality updates and early warning capabilities for Bangladesh.
- Beyond predictive accuracy, the system demonstrates practical utility by supporting public-health decision making and aiding in the identification of key factors driving pollution.



## 2. Background

### 2.1. Air quality index

The air quality index (AQI)—sometimes referred to as the air pollution index (API) or pollutant standard index (PSI)—is a unified framework used internationally to communicate the condition of ambient air relative to prescribed national thresholds. By combining measurements of multiple pollutants into a single indicator, the AQI provides a convenient summary of overall air quality. It serves both as a snapshot of present conditions and as a short-term predictor of potential pollution risks. Importantly, the AQI highlights the health implications of exposure to contaminated

air, particularly those that may appear within a few hours or days, thus offering the public an accessible metric for safeguarding health.

To overcome the limitations of traditional AQI calculations, which are typically based on 24-hour averages that may obscure sudden fluctuations, the U.S. Environmental Protection Agency (US-EPA) developed the Nowcast method [26]. This approach prioritizes the most recent 12 hours of pollution data, placing greater emphasis on current readings, thereby making the index more sensitive to abrupt changes in pollution levels. The Nowcast methodology is particularly relevant for fine particulate matter ($PM_{2.5}$), in which short-term variability can significantly influence health outcomes.

For AQI estimation, the following variables are considered:

$$\begin{aligned}
\text{AQI} &: \text{Air Quality Index} \\
\text{PM2.5} &: \text{Particulate Matter 2.5} \\
\text{NowCast Concentration} &: \text{Predicted concentration of PM2.5} \\
\text{Raw Concentration} &: \text{Observed concentration of PM2.5} \\
\text{PM2.5}_{\text{IAQI_low}} &: \text{Lower limit of the PM2.5 IAQI range} \\
\text{PM2.5}_{\text{IAQI_high}} &: \text{Upper limit of the PM2.5 IAQI range} \\
\text{AQI}_{\text{high}} &: \text{Upper bound of the AQI scale} \\
\text{AQI}_{\text{low}} &: \text{Lower bound of the AQI scale}
\end{aligned}$$

The AQI is determined using three conditional cases:

1. If the raw $PM_{2.5}$ concentration is below the lower threshold:

$$\text{AQI} = 0 \quad \text{if Raw Concentration} < \text{PM2.5}_{\text{IAQI\_low}} \tag{1}$$

2. If the raw $PM_{2.5}$ concentration lies between the lower and upper thresholds:

$$\text{AQI} = \frac{\text{NowCast Concentration} - \text{PM2.5}_{\text{IAQI\_low}}}{\text{PM2.5}_{\text{IAQI\_high}} - \text{PM2.5}_{\text{IAQI\_low}}} \times (\text{AQI}_{\text{high}} - \text{AQI}_{\text{low}}) + \text{AQI}_{\text{low}} \tag{2}$$

3. If the Nowcast concentration exceeds the upper threshold:

$$\text{AQI} = \text{AQI}_{\text{high}} \quad \text{if NowCast Concentration} > \text{PM2.5}_{\text{IAQI\_high}} \tag{3}$$

## 2.2. Air quality assessment

Designing effective mitigation strategies requires systematic monitoring and evaluation of ambient conditions. The AQI integrates information on five key pollutants: carbon monoxide (CO), nitrogen dioxide ($NO_2$), sulfur dioxide ($SO_2$), ozone ($O_3$), and particulate matter ($PM_{2.5}$). In Bangladesh, which serves as our case study, regulatory frameworks are guided by the National Ambient Air Quality Standards (NAAQS). These standards define scientifically established threshold values intended to safeguard both the environment and public health. Table 1 provides the regulatory limits and associated averaging periods for the major pollutants of concern.

**Table 1**
Regulatory Pollutant Thresholds and Averaging Times Under NAAQS

| Pollutant | Limit Value | Averaging Time |
|---|---|---|
| CO | 10 mg/m$^3$ | 8 hours |
| | 40 mg/m$^3$ | 1 hour |
| Pb | 0.5 $\mu$g/m$^3$ | Yearly |
| $NO_x$ | 100 $\mu$g/m$^3$ | Yearly |
| $PM_{10}$ | 50 $\mu$g/m$^3$ | Yearly |
| | 150 $\mu$g/m$^3$ | 24 hours |
| $PM_{2.5}$ | 15 $\mu$g/m$^3$ | Yearly |
| | 65 $\mu$g/m$^3$ | 24 hours |
| $O_3$ | 235 $\mu$g/m$^3$ | 1 hour |
| | 157 $\mu$g/m$^3$ | 8 hours |
| $SO_2$ | 80 $\mu$g/m$^3$ | Yearly |
| | 365 $\mu$g/m$^3$ | 1 day |

## 3. Literature review

Forecasting urban air quality has evolved from conventional statistical regression to deep-learning architectures that better capture temporal and spatiotemporal dependencies. This section synthesizes recent work across five themes: (1) classical and ensemble machine learning, (2) recurrent and hybrid deep networks, (3) transfer learning and multi-resolution modeling, (4) topology-aware graph methods, and (5) exogenous factors, interpretability, and nontraditional sensing. We conclude with the research gap motivating our hybrid BiGRU–LSTM design.

### 3.1. Classical and ensemble machine learning

Early approaches relied on linear and tree-based models to learn relationships between pollutant concentrations and meteorological covariates. For example, a multitask learning (MTL) framework with structured regularization demonstrated that appropriate parameter sharing can simultaneously reduce model complexity and improve forecasts, using hourly data from two stations and RMSE as the primary metric [5,11]. City-scale comparisons in Delhi further indicated that ensemble learners –including Random Forest (RF), eXtreme Gradient Boosting (XGB), AdaBoost, and KNN – often outperform simpler linear baselines across MAE, RMSE, and $R^2$ [8]. A two-stage pipeline combining ANN for short-horizon forecasting with SVM for categorical AQI assessment reported high accuracies when evaluated on multi-site data [16].

### 3.2. Recurrent and hybrid deep networks

Deep-sequence models have consistently improved $PM_{2.5}$/AQI predictions by learning longer-range temporal patterns. Evaluations on the UCI Beijing data set showed that LSTM/GRU variants outperform regression baselines, with a convolutional BiGRU

(CBGRU) variant achieving the lowest RMSE, MAE, and SMAPE among compared models [3, 19]. At larger scales, CNN–LSTM hybrids surpassed classical time-series baselines such as SARIMA on Beijing AQI, reflecting the benefit of joint feature extraction and sequence modeling [31]. Beyond off-the-shelf architectures, studies continue to propose tailored hybrids for specific spatiotemporal characteristics and pollutant targets [9, 24].

### 3.3. Transfer learning and multi-resolution forecasting

Temporal aggregation can degrade accuracy if models fail to preserve short-term variability. A BLSTM with transfer learning reduced RMSE by 36.85% (daily) and 42.58% (weekly) on Guangdong monitoring data, and used IDW interpolation to visualize residuals across space [15]. These results suggest that knowledge transfer across resolutions and locations can stabilize performance when data are sparse or heterogeneous.

### 3.4. Topology-aware and graph-based modeling

Air-quality monitoring networks (AQMN) exhibit spatial autocorrelation and network effects. A residual GCN with BiLSTM (Res-GCN-BiLSTM) aligned with these properties, reducing MAE by $\sim$11% for $NO_2$ and 17% for $O_3$ in Shanghai by fusing topology, auxiliary pollutants, and meteorology [30]. Complementarily, a Bayesian GraphGRU with GraphSAGE (BGGRU) improved $PM_{2.5}$ forecasts in Beijing by jointly modeling spatial neighborhoods and temporal dynamics [13]. Such results highlight the utility of graph signal processing for regional prediction.

### 3.5. Exogenous factors, interpretability, and nontraditional sensing

Accounting for exogenous shocks and interpretability requirements is increasingly important. A deep model incorporating spatial autocorrelation quantified COVID-19 lockdown effects in Wuhan and Shanghai, cutting forecast error by up to 67% and isolating intervention impacts [33]. A variational Bayesian GRU improved variable selection while retaining forecast skill, providing interpretable posterior summaries [12]. Beyond station data, hyperspectral imaging combined with deep learning achieved 85.93% accuracy in classifying pollution levels, showing promise for scalable, low-cost sensing [20]. Bibliometric analyses further document the post-2017 rise of deep learning for $PM_{2.5}$ and AQI tasks, and the COVID-19 shift in time-series forecasting priorities [21].

### 3.6. Process-oriented and neuro-fuzzy pipelines

Process-driven systems have explored optimization and soft computing. An SCC/GA/PSO/PCA/Neuro-Fuzzy pipeline, evaluated in Tianjin, Kaohsiung, and Bogotá, demonstrated consistent MAE reductions from PCA across $PM_{2.5}$, $PM_{10}$, and $O_3$, while ARIMA remained competitive for specific horizons and pollutants [34].

**Table 2**
Summary of recent literature on air quality prediction models

| Study | Methods/Models | Data Set | Pollutants/Features | Key Findings / Results |
|---|---|---|---|---|
| [5] | Multitask learning with structured regularization | Two stations (hourly); meteorology + pollution | PM, met. covariates | Structured regularizer reduced parameters and RMSE |
| [22] | LSTM, SVR, AQI-based modeling | CityPulse EU FP7 (Aarhus, Brasov, 2013–2015) | $O_3$, PM, CO, $NO_2$, $SO_2$, coords., time | LSTM best overall; precision 98%, F1 97% for red alarms |
| [34] | SCC, GA, PSO, PCA, Neuro-Fuzzy | Tianjin (CN), Bogotá (CO) | $PM_{10}$, $PM_{2.5}$, $O_3$, met. | PCA consistently lowered MAE; ARIMA strong for $PM_{10}$ |
| [19] | LSTM, GRU, SVR, GBR, DTR, BGRU, CBGRU | UCI Beijing (43,800 rows) [3] | $PM_{2.5}$ + meteorology | CBGRU achieved lowest RMSE, MAE, SMAPE |
| [16] | Two-stage ANN + SVM | Delhi (37 locations) | Eight AQI pollutants | Median Gaussian SVM reached 97.3% accuracy |
| [15] | BLSTM + transfer learning; IDW viz | Guangdong (3 years, 26,304 hr/station) | $PM_{2.5}$ | RMSE cut 36.85% (daily), 42.58% (weekly) |
| [8] | LR, RF, KNN, Ridge, XGB, AdaBoost | Delhi $PM_{2.5}$ + meteorology | $PM_{2.5}$, temp., RH, wind, etc. | Ensembles (XGB, RF) led across MAE/RMSE/$R^2$ |
| [33] | Deep-model + spatial autocorr. | Wuhan & Shanghai (COVID-19) | $PM_{2.5}$, weather, social | Error reduced up to 67% by modeling lockdown effects |
| [30] | Res-GCN-BiLSTM | Shanghai AQMN | $NO_2$, $O_3$ + topology | MAE down ∼11% ($NO_2$), 17% ($O_3$) vs. baselines |
| [13] | BGGRU (GraphSAGE + Bayesian GRU) | Beijing AQ data | $PM_{2.5}$ | Improved spatiotemporal learning and accuracy |
| [12] | Variational Bayesian GRU | Beijing data sets | $PM_{2.5}$ + covariates | Better variable selection with accurate forecasts |
| [31] | CNN–LSTM vs. SARIMA | Beijing AQI | $PM_{2.5}$, AQI | CNN–LSTM outperformed SARIMA on MAE, RMSE |
| [20] | Hyperspectral + DL (3D CNN AE, PCA, VGG-16) | Custom imaging dataset | Pollution imaging features | 85.93% classification accuracy for levels |

### 3.7. Synthesis and research gap

Table 2 consolidates methods, data sets, targets, and headline results from the representative literature discussed above.

Across methods, three limitations recur. First, many models optimize for either short-horizon responsiveness or multi-resolution stability, but not both; few explicitly integrate real-time *Nowcast* with raw concentrations to retain sensitivity to rapid shifts while preserving calibration. Second, unidirectional sequence models can miss anticipatory cues present in future context, whereas bidirectional encoders remain underused in operational AQI forecasting. Third, deployment and external validity are often secondary: comparatively few systems demonstrate end-to-end integration that spans model training, baseline benchmarking, and public facing delivery at city/megacity scale in South Asia.

Motivated by these gaps, our work develops a hybrid BiGRU–LSTM architecture that (1) fuses ¡ and raw concentration streams, (2) captures both long-range memory and bidirectional temporal structure, (3) is validated against strong baselines (GRU, LSTM, TCN), and (4) is delivered through a mobile application for real-time early warnings in Bangladesh.

## 4. Research methodology

The methodological framework of this study consists of four major phases: data collection, data preparation, deep-learning model development, and evaluation (Figure 1). Each phase is designed to ensure that the proposed model is trained on reliable data, optimized effectively, and validated against strong baselines.

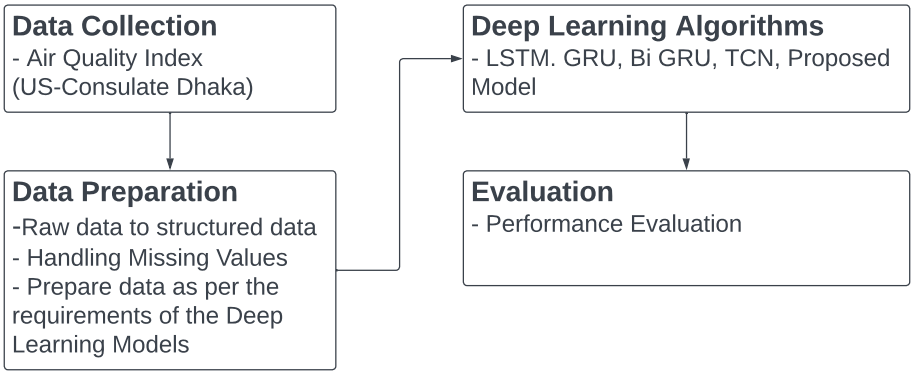


**Figure 1.** Phases of the Research Methodology

## 4.1. Data collection

Data were obtained from the U.S. consulate in Dhaka [1]. The data set contains 53,355 records in eight attributes, four of which are numerical and the remainder categorical. Some negative values were identified in the numerical attributes, necessitating preprocessing before further use. The data set provides both raw pollutant concentrations and derived indicators, enabling the construction of a forecasting pipeline.

## 4.2. Data preparation

Data preparation began by removing invalid AQI entries (e.g., negative values) and renaming attributes for consistency. Figure 2 outlines the workflow for intelligent AQI forecasting. Irrelevant attributes such as Hour, AQI Category, QC Name, and Conc. Unit were excluded to minimize noise.

The predictors were defined as Nowcast and RawConc., while AQI was treated as the target variable. A min–max scaling procedure was applied to normalize inputs into a consistent range, improving convergence stability during training. Data were then reshaped into a three-dimensional tensor to fit the sequential requirements of neural networks. Each tensor instance preserved the temporal relationship between the selected features. These steps ensured the data set was both statistically reliable and structurally suitable for deep learning.

## 4.3. Deep-learning approach

Four sequential deep-learning architectures were employed: LSTM, GRU, Bi-GRU, and TCN. Each was trained on the processed data set and evaluated using Root Mean Square Error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). All implementations were carried out in Python (v3.9). The workflow is summarized in Algorithm 1.

**Algorithm 1** Data Processing and Model Evaluation

```
Load AQI data set
Apply preprocessing (cleaning, scaling, reshaping)
Split data set into training and testing subsets
for each model in {LSTM, GRU, Bi-GRU, TCN} do
    Train using stratified K-fold validation
end for
Evaluate models on test set
Compute RMSE, MAE, and MAPE
```

### 4.3.1. Gated Recurrent Unit (GRU)

The GRU is a recurrent neural network variant optimized for sequential learning [29]. It uses gating mechanisms to manage information flow. The reset gate discards irrelevant past information, while the update gate balances new and historical signals.

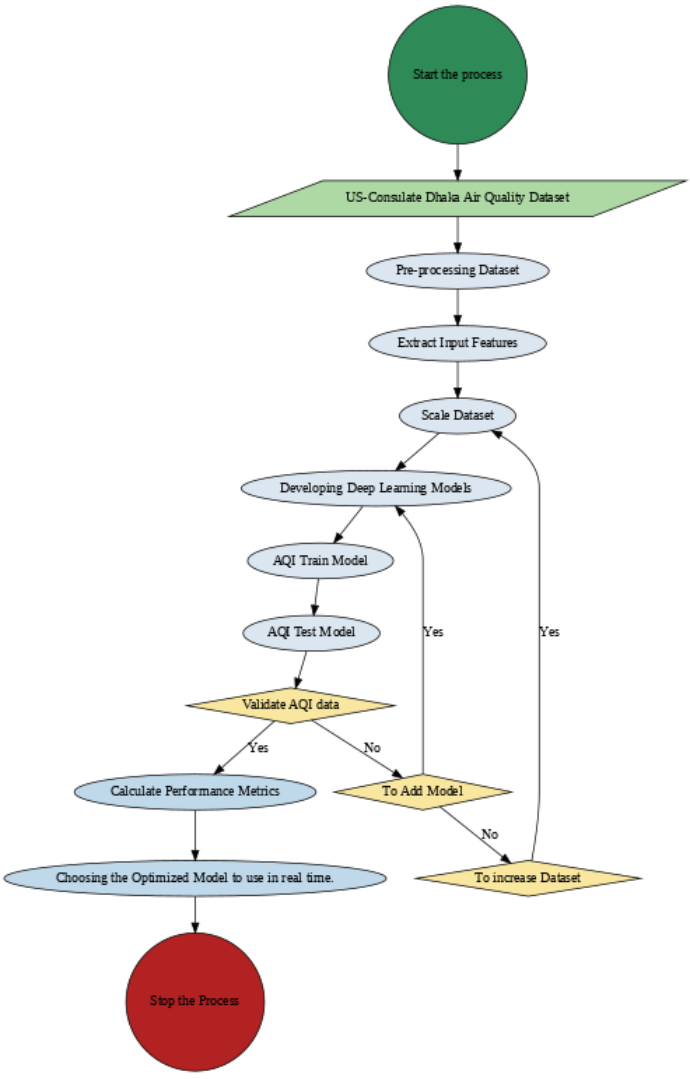


**Figure 2.** Workflow for air quality forecasting

The hidden state ($h_p$) evolves by integrating candidate updates, enabling efficient sequence modeling with fewer parameters than traditional LSTMs. GRUs are well-suited for time-series AQI prediction when computational efficiency and accuracy are both critical.

### 4.3.2. Temporal Convolutional Network (TCN)

The TCN [27] is a convolutional architecture designed for sequence modeling. It employs dilated causal convolutions, allowing the model to capture long-range temporal dependencies without recursion. Batch normalization stabilizes training, while ReLU introduces nonlinear feature learning [4]. The model is trained using the Adam optimizer with MSE loss, yielding stable regression performance for AQI forecasting.

### 4.3.3. Long Short-Term Memory (LSTM)

The LSTM extends RNNs with gates controlling memory flow [25]. The forget gate removes irrelevant past states, the input gate regulates new data entry, and the output gate determines what information is propagated forward. In this study a two-layer LSTM was implemented: the first with 128 units and the second with 64 units, both with L2 regularization. A final dense layer produced the continuous AQI output. Training used Adam optimization with MSE loss

### 4.3.4. Proposed BiGRU–LSTM architecture

The proposed DeepBiGRULSTM architecture integrates long short-term memory (LSTM) and bidirectional gated recurrent unit (Bi-GRU) layers to achieve precise *air-quality index* (AQI) predictions using Nowcast and raw concentration data. This hybrid model leverages the strengths of LSTM for capturing long-term dependencies and Bi-GRU for modeling bidirectional temporal relationships, while L2 regularization is used to prevent overfitting. The architecture is designed to process sequential input data and output a single continuous AQI value, making it suitable for real-time air quality forecasting.

**Model Architecture:**

- **Input layer:** Accepts a 3D tensor $X$ of shape $(\text{batch size}, \text{time steps}, \text{features})$, where the features are Nowcast and RawConc.
- **LSTM layer:** 128 units with ReLU activation, configured to return sequences. L2 regularization ($\lambda = 0.001$) is applied to reduce overfitting.
- **Bidirectional GRU layer:** 64 units with *tanh* activation, processing sequences in both forward and backward directions.
- **Output layer:** A dense layer with a single linear neuron producing the predicted AQI value, $\text{AQI}_{\text{pred}}$.

**Mathematical Formulation:**

*LSTM layer.* Let $X = \{X_1, X_2, \ldots, X_T\}$ represent the input sequence, where $X_p \in \mathbb{R}^2$ is the input vector at time step $p$ (Nowcast, RawConc).

The LSTM computes hidden states $h_p^{\text{LSTM}}$ and cell states $C_p$:

$$f_p = \sigma(W_f \cdot [h_{p-1}^{\text{LSTM}}, X_p] + b_f), \tag{4}$$

$$i_p = \sigma(W_i \cdot [h_{p-1}^{\text{LSTM}}, X_p] + b_i), \tag{5}$$

$$g_p = \tanh(W_g \cdot [h_{p-1}^{\text{LSTM}}, X_p] + b_g), \tag{6}$$

$$C_p = f_p \cdot C_{p-1} + i_p \cdot g_p, \tag{7}$$

$$o_p = \sigma(W_o \cdot [h_{p-1}^{\text{LSTM}}, X_p] + b_o), \tag{8}$$

$$h_p^{\text{LSTM}} = o_p \cdot \tanh(C_p), \tag{9}$$

where $\sigma$ denotes the sigmoid activation. The output sequence is $H^{\text{LSTM}} = \{h_1^{\text{LSTM}}, \ldots, h_T^{\text{LSTM}}\}$.

*Bidirectional GRU layer.* The Bi-GRU processes $H^{\text{LSTM}}$ in both forward and backward directions.

**Forward pass:**

$$r_p^f = \sigma(W_r^f \cdot [h_{p-1}^f, h_p^{\text{LSTM}}] + b_r^f), \tag{10}$$

$$z_p^f = \sigma(W_z^f \cdot [h_{p-1}^f, h_p^{\text{LSTM}}] + b_z^f), \tag{11}$$

$$g_p^f = \tanh(W_g^f \cdot [r_p^f \cdot h_{p-1}^f, h_p^{\text{LSTM}}] + b_g^f), \tag{12}$$

$$h_p^f = (1 - z_p^f) \cdot h_{p-1}^f + z_p^f \cdot g_p^f. \tag{13}$$

**Backward pass:**

$$r_p^b = \sigma(W_r^b \cdot [h_{p+1}^b, h_p^{\text{LSTM}}] + b_r^b), \tag{14}$$

$$z_p^b = \sigma(W_z^b \cdot [h_{p+1}^b, h_p^{\text{LSTM}}] + b_z^b), \tag{15}$$

$$g_p^b = \tanh(W_g^b \cdot [r_p^b \cdot h_{p+1}^b, h_p^{\text{LSTM}}] + b_g^b), \tag{16}$$

$$h_p^b = (1 - z_p^b) \cdot h_{p+1}^b + z_p^b \cdot g_p^b. \tag{17}$$

**Concatenation:**

$$h_p^{\text{Bi-GRU}} = [h_p^f, h_p^b]. \tag{18}$$

*Output layer.* The AQI prediction is generated by a dense layer:

$$\text{AQI}_{\text{pred}} = W_{\text{out}} \cdot h_T^{\text{Bi-GRU}} + b_{\text{out}}, \tag{19}$$

where $W_{\text{out}}$ and $b_{\text{out}}$ are the weights and bias.

**Rationale** This hybrid architecture combines the LSTM's capacity to capture long-term sequential dependencies with the Bi-GRU's ability to model bidirectional temporal patterns. ReLU activation in the LSTM enhances nonlinear feature learning, while tanh activation in the Bi-GRU ensures stable gradient propagation. L2 regularization prevents overfitting, and the linear dense output layer ensures suitability for regression tasks. Experimental results confirm that this design provides superior accuracy and generalization compared to standalone LSTM or GRU models. ression at the output ensures a continuous AQI forecast.

### 4.4. Evaluation

The final phase of the methodology involves assessing the predictive performance of the deep-learning models trained in the previous stage. Model evaluation was carried out using three widely adopted regression metrics: Root Mean Square Error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE). These measures collectively capture error magnitude, robustness to outliers, and relative accuracy, providing a comprehensive basis for comparison. The best-performing variant of each algorithm was selected based on these criteria, and the results were consolidated into a comparative table for systematic analysis prior to deployment.

RMSE quantifies the square root of the average squared difference between observed and predicted values, thus penalizing larger deviations more strongly. This sensitivity to outliers is expressed as:

$$\mathrm{RMSE} = \sqrt{\frac{1}{m}\sum_{j=1}^{m}(y_{\mathrm{actual},j} - y_{\mathrm{pred},j})^2}, \tag{20}$$

where $m$ is the total number of instances, $y_{\mathrm{actual},j}$ the ground truth, and $y_{\mathrm{pred},j}$ the model output.

MAE provides the average magnitude of deviations, independent of direction. Unlike RMSE, it treats all errors linearly, making it less sensitive to extreme values:

$$\mathrm{MAE} = \frac{1}{m}\sum_{j=1}^{m}\left|y_{\mathrm{actual},j} - y_{\mathrm{pred},j}\right|. \tag{21}$$

Together, these metrics provide complementary perspectives: RMSE emphasizes large deviations, MAE offers interpretable average errors, and MAPE contextualizes accuracy in percentage form. This multi-metric approach ensures a robust appraisal of each model's forecasting capability.

## 5. Experimental results and analysis

The predictive performance of the deep-learning models was assessed using Root Mean Square Error (RMSE; Equation 20) and mean absolute error (MAE; Equation 21). The models were evaluated on both training and test data sets, with the test set serving as the basis for validation.

### 5.1. Dataset structure and quality

To establish an initial understanding of the data set a structural overview was conducted, focusing on its size, completeness, and quality. These characteristics are essential for identifying potential preprocessing needs before applying predictive models. The data set consists of 53,355 observations across eight variables. Approximately 11.5% of the entries are missing, but no duplicate records were found, indicating that

redundancy does not present a concern for subsequent analysis. Table 3 summarizes the key structural properties of the data set.

**Table 3**
Data Set Structure and Quality Overview

| Property | Value |
|---|---|
| Number of variables | 8 |
| Number of observations | 53,355 |
| Missing values | 55,213 |
| Percentage of missing values | 11.5% |
| Duplicate records | 0 |

Exploratory data analysis further revealed notable relationships among variables. Figure 3 presents a correlation heat map, showing that both Nowcast concentration and raw concentration are strongly correlated with the air quality index (AQI). These findings support the decision to adopt Nowcast concentration and raw concentration as predictor variables, with AQI designated as the target output.

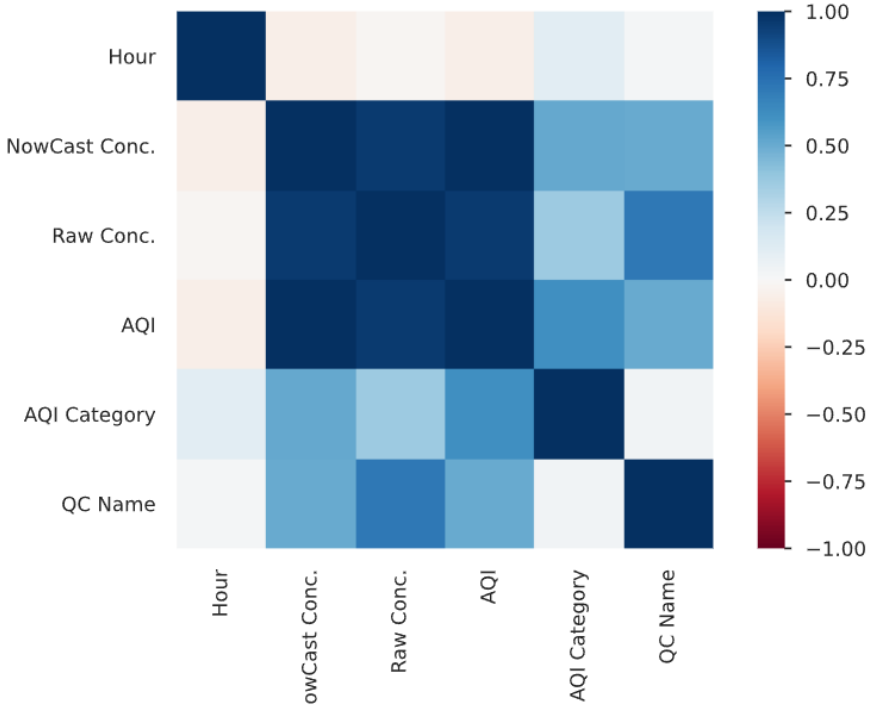


**Figure 3.** Correlation heat map of the variables

## 5.2. Data preparation for model training

Following the structural assessment of the data set, several preprocessing steps were undertaken to ensure data quality and suitability for model development. The data set was first cleaned to address issues related to missing values and then partitioned into training and testing subsets for model evaluation.

The data set contained approximately 11.5% missing values across its variables. Since missing entries can bias model performance and reduce predictive accuracy, a systematic cleaning strategy was applied. Negative or invalid pollutant concentrations were removed, as such values are physically implausible. For remaining missing entries imputation methods were considered. In this study, rows with critical missing values in the target variable (AQI) were discarded to preserve data integrity, while missing values in predictor variables were replaced using mean imputation for numerical attributes and mode imputation for categorical attributes.

To evaluate the predictive performance of the models, the processed data set was partitioned into training and testing sets using an 80:20 ratio. The training set, comprising 80% of the observations, was used to fit the models, while the remaining 20% was reserved for independent testing. This split ensures that the trained models are evaluated on unseen data, thereby providing a realistic measure of their generalization capability. Stratified sampling was employed to maintain proportional representation of air quality index (AQI) levels across both subsets.

## 5.3. Performance on training data

Table 4 presents the results obtained on the training data set. Among the models, Bi-GRU and the proposed BiGRU–LSTM achieved the lowest error values, with BiGRU–LSTM slightly outperforming Bi-GRU in RMSE. LSTM also demonstrated competitive performance, while GRU and TCN exhibited relatively higher error values.

**Table 4**
Model Performance on Training Data Set

| Model | MAE | RMSE |
|---|---|---|
| GRU | 0.8196 | 1.1741 |
| TCN | 3.0253 | 2.3667 |
| LSTM | 0.6631 | 2.0480 |
| Bi-GRU | 0.3458 | 0.5646 |
| BiGRU–LSTM | 0.3615 | 0.4976 |

## 5.4. Performance on test data

The evaluation on the test data set, which validates model generalization, is summarized in Table 5. LSTM achieved the lowest MAE, while BiGRU–LSTM attained the lowest RMSE, confirming its ability to balance accuracy and stability. TCN again underperformed compared to the recurrent models, producing the highest errors.

**Table 5**
Model Performance on Test Dataset

| Model | MAE | RMSE |
|---|---|---|
| GRU | 0.8178 | 1.1741 |
| TCN | 3.0253 | 3.6127 |
| LSTM | 0.4206 | 0.8914 |
| Bi-GRU | 0.4633 | 0.8699 |
| BiGRU–LSTM | 0.4698 | 0.7548 |

## 5.5. Performance analysis

The comparative results reveal clear trends across the evaluated models. GRU showed moderate performance but struggled to achieve low error values relative to its variants. TCN consistently underperformed on both training and test sets, indicating limited suitability for this application.

LSTM delivered strong results on both training and test datasets with low MAE, demonstrating its effectiveness in capturing temporal dependencies. However, Bi-GRU and BiGRU–LSTM provided superior overall accuracy. The Bi-GRU achieved the lowest MAE on the training set, while BiGRU–LSTM consistently produced the lowest RMSE across training and test sets. This indicates that the hybrid architecture not only captures sequential patterns effectively but also generalizes better to unseen data.

In summary, while all recurrent models outperformed TCN, the BiGRU–LSTM hybrid model demonstrated the best trade-off between precision (MAE) and robustness (RMSE), making it the most reliable architecture for AQI prediction in this study.

## 5.6. Model integration

This phase focuses on deploying the trained deep-learning model to a cloud environment and providing end-user accessibility through a mobile application. The integration process involves hosting the model on an Amazon EC2 instance, enabling API-based access and developing an Android application with an intuitive interface.

Amazon EC2 provides a scalable and secure cloud-based infrastructure suitable for hosting machine learning models. The trained model is uploaded to the instance, with all required libraries and dependencies configured. End points are then established to receive input data and return predictions. Leveraging EC2 ensures flexibility, scalability, and accessibility, making the model suitable for real-world deployment.

The model is deployed on the EC2 instance using Flask, which enables the creation of lightweight RESTful APIs. Through these APIs, client applications can send requests containing input data to the deployed model and receive predictions of air quality in real time. Security mechanisms, including authentication and authorization, are incorporated to safeguard data integrity and user privacy during communication.

To make the system accessible to non-technical users, an Android application is developed. The application provides a user-friendly interface where individuals can input relevant parameters, such as date, Nowcast concentration, and raw concentration. The entered data are transmitted to the EC2-hosted model through the Flask API, and the predicted air quality index (AQI) values are displayed to the user. This approach enhances usability and enables individuals to benefit from the model's predictive capabilities without requiring technical expertise.

## 6. Discussion

The comparative evaluation of the models reveals important distinctions in their predictive behavior across data sets. LSTM consistently achieved strong accuracy during both training and testing, confirming its ability to capture complex temporal dependencies within familiar data distributions. However, while effective at recognizing patterns in controlled settings, its performance did not generalize as strongly as the Bi-GRU or BiGRU–LSTM when validation was considered. In contrast, the Bi-GRU demonstrated superior adaptability, achieving lower overall errors on the validation set. The hybrid BiGRU–LSTM model further strengthened this adaptability by combining the long-term memory capacity of LSTM with the bidirectional sequence modeling of Bi-GRU, resulting in the lowest RMSE across both training and validation data.

This divergence highlights a critical insight: models that excel in pattern extraction during training do not necessarily guarantee resilience under data shifts or evolving real-world conditions. The Bi-GRU and BiGRU–LSTM architectures showed stronger robustness and generalization, making them more suitable for deployment in environments where air quality patterns fluctuate dynamically, such as urban areas in Bangladesh. This suggests that robustness to variability, rather than in-sample accuracy alone, should be prioritized when selecting forecasting models for operational use.

From a practical perspective, these findings hold direct implications for real-world deployment. While LSTM can capture established temporal trends, the Bi-GRU and BiGRU–LSTM provide more reliable forecasts under changing conditions, ensuring consistency and stability in predictions. Such adaptability is particularly valuable when integrated into mobile applications, where real-time AQI forecasts can help individuals take preventive health measures. Based on its strong performance and generalization, the BiGRU–LSTM model was selected as the most suitable candidate for application-level deployment.

## 7. Limitations and future work

Despite the promising outcomes, several limitations of this study must be acknowledged. First, the models were trained and validated primarily on data from Dhaka.

Their performance may vary in regions with different climatic, geographic, or anthropogenic characteristics. Future work should expand to multi-city or multi-country datasets to evaluate cross-regional generalizability. Second, the study did not explore hyperparameter optimization in depth. A more systematic investigation of learning rates, batch sizes, and regularization strategies could further improve predictive performance.

Another limitation relates to feature scope. The present models relied primarily on pollutant concentration values as predictors. Incorporating additional exogenous variables, such as meteorological conditions, traffic density, or seasonal effects, may enhance model robustness. Exploring transfer learning strategies could also enable these architectures to adapt efficiently to other time-series forecasting domains, including weather, energy demand, and financial markets.

Finally, deployment considerations highlight the importance of robust system architecture. Building reliable APIs is critical for ensuring smooth communication between the trained model (e.g., hosted on a cloud service such as AWS EC2) and front-end applications. Real-time responsiveness, scalability, and error handling will be central to user adoption. Addressing these aspects in future work will strengthen the utility of deep-learning models for both air quality prediction and broader environmental monitoring applications.

## 8. Conclusion

This study highlights the importance of selecting appropriate deep-learning architectures based on data set characteristics and application-specific requirements in environmental monitoring. Through a comparative evaluation of GRU, LSTM, TCN, Bi-GRU, and the proposed DeepBiGRULSTM, we advanced the understanding of air quality forecasting. The results showed that while LSTM models performed strongly on training and test data sets, the Bi-GRU and the hybrid BiGRU–LSTM demonstrated superior generalization ability, with BiGRU–LSTM achieving the lowest overall error (RMSE of 0.4976 on the training set and 0.7548 on the test set).

The proposed BiGRU–LSTM model represents a methodological contribution by combining the memory retention of LSTM with the bidirectional sequence learning of Bi-GRU. This hybrid architecture proved more robust than standalone models, offering consistent accuracy and stability across both training and validation phases. Importantly, the framework demonstrates how advanced neural architectures can be adapted to practical challenges in environmental data science.

In summary, this research provides a complete methodological pipeline, including data preprocessing, model development, and evaluation, tailored for precise AQI forecasting. As air quality continues to be a pressing global concern, the integration of robust forecasting models with accessible technologies, such as mobile platforms, can play a pivotal role in public health awareness and environmental management. The findings establish a foundation for building more accurate and adaptable prediction

systems, ultimately contributing to informed decision making and improved quality of life in urban environments.

## References

[1] AirNow.gov: The historical archive of U.S. Air Quality Index (AQI) for Dhaka US Consulate, Bangladesh, 2024. https://www.airnow.gov/international/us-embassies-and-consulates/#Bangladesh$Dhaka. Retrieved June 05, 2024.

[2] Akintunde E.A.: Theories and concepts for human behavior in environmental preservation, *Journal of Environmental Science and Public Health*, vol. 1(2), pp. 120–133, 2017. doi: 10.26502/jesph.96120012.

[3] Asuncion A., Newman D., *et al.*: UCI machine learning repository, 2007.

[4] Bai Y.: RELU-function and derived function review. In: *SHS Web of Conferences*, vol. 144, p. 02006, EDP Sciences, 2022. doi: 10.1051/shsconf/202214402006.

[5] Chavda J.: Analyzing the Relationship Between Asthma and Air Impurity in Developed Cities Using Machine Learning Approach.

[6] Chowdhury M.Z.I., Leung A.A., Walker R.L., Sikdar K.C., O'Beirne M., Quan H., Turin T.C.: A comparison of machine learning algorithms and traditional regression-based statistical modeling for predicting hypertension incidence in a Canadian population, *Scientific Reports*, vol. 13(1), p. 13, 2023. doi: 10.1038/s41598-022-27264-x.

[7] Dairi A., Harrou F., Khadraoui S., Sun Y.: Integrated multiple directed attention-based deep learning for improved air pollution forecasting, *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–15, 2021. doi: 10.1109/tim.2021.3091511.

[8] Du S., Li T., Yang Y., Horng S.J.: Deep Air Quality Forecasting Using Hybrid Deep Learning Framework, *IEEE Transactions on Knowledge and Data Engineering*, vol. 33(6), pp. 2412–2424, 2021. doi: 10.1109/TKDE.2019.2954510.

[9] Du W., Chen L., Wang H., Shan Z., Zhou Z., Li W., Wang Y.: Deciphering Urban Traffic Impacts on Air Quality by Deep Learning and Emission Inventory, *Journal of Environmental Sciences*, vol. 124, pp. 745–757, 2023. doi: 10.1016/j.jes.2021.12.035.

[10] Geng G., Xiao Q., Liu S., Liu X., Cheng J., Zheng Y., Xue T., Tong D., Zheng B., Peng Y., *et al.*: Tracking air pollution in China: near real-time PM2. 5 retrievals from multisource data fusion, *Environmental Science & Technology*, vol. 55(17), pp. 12106–12115, 2021.

[11] Hodson T.O.: Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not, *Geoscientific Model Development Discussions*, vol. 2022, pp. 1–10, 2022. doi: 10.5194/gmd-15-5481-2022.

[12] Jin X.B., Wang Z.Y., Gong W.T., Kong J.L., Bai Y.T., Su T.L., Ma H.J., Chakrabarti P.: Variational Bayesian Network with Information Interpretability Filtering for Air Quality Forecasting, *Mathematics*, vol. 11(4), p. 837, 2023. doi: 10.3390/math11040837.

[13] Jin X.B., Wang Z.Y., Kong J.L., Bai Y.T., Su T.L., Ma H.J., Chakrabarti P.: Deep Spatio-Temporal Graph Network with Self-Optimization for Air Quality Prediction, *Entropy*, vol. 25(2), p. 247, 2023. doi: 10.3390/e25020247.

[14] Kabir M.H., Hossen M.N.: Impacts of Flood and its Possible Solutions in Bangladesh, *Journal of Environmental Studies*.

[15] Kang G.K., Gao J.Z., Chiao S., Lu S., Xie G.: Air Quality Prediction: Big Data and Machine Learning Approaches, *International Journal of Environmental Science and Development*, vol. 9(1), pp. 8–16, 2018. doi: 10.18178/ijesd.2018.9.1.1066.

[16] Kumar K., Pande B.P.: Air Pollution Prediction with Machine Learning: A Case Study of Indian Cities, *International Journal of Environmental Science and Technology*, vol. 20, pp. 5333–5348, 2023. doi: 10.1007/s13762-022-04241-5.

[17] Majumder A.K., Patoary M.N.A., Al Nayeem A., Rahman M.: Air quality index (AQI) changes and spatial variation in Bangladesh from 2014 to 2019, *Journal of Air Pollution and Health*, 2023. doi: 10.18502/japh.v8i2.12919.

[18] Makri A., Stilianakis N.I.: Vulnerability to air pollution health effects, *International journal of hygiene and environmental health*, vol. 211(3-4), pp. 326–336, 2008. doi: 10.1016/j.ijheh.2007.06.005.

[19] Mohiuddin A.K.: Chemical Contaminants and Pollutants in the Measurable Life of Dhaka City, 2019. doi: 10.29161/pt.v7.i1.2019.25.

[20] Mukundan A., Huang C.C., Men T.C., Lin F.C., Wang H.C.: Air Pollution Detection Using a Novel Snap-Shot Hyperspectral Imaging Technique, *Sensors*, vol. 22(16), p. 6231, 2022. doi: 10.3390/s22166231.

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

[22] Nie J.Y.: 2017 IEEE International Conference on Big Data: Proceedings. In: *2017 IEEE International Conference on Big Data*, Boston, MA, USA, 2017.

[23] Nikmanesh Y., Mohammadi M.J., Yousefi H., Mansourimoghadam S., Taherian M.: The effect of long-term exposure to toxic air pollutants on the increased risk of malignant brain tumors, *Reviews on Environmental Health*, vol. 38(3), pp. 519–530, 2023.

[24] Niu M., Zhang Y., Ren Z.: Deep Learning-Based PM2.5 Long Time-Series Prediction by Fusing Multisource Data—A Case Study of Beijing, *Atmosphere*, vol. 14(2), p. 340, 2023. doi: 10.3390/atmos14020340.

[25] Rahman M.M., Nayeem M.E.H., Ahmed M.S., Tanha K.A., Sakib M.S.A., Uddin K.M.M., Babu H.M.H.: AirNet: predictive machine learning model for air quality forecasting using web interface, *Environmental Systems Research*, vol. 13, p. 44, 2024. doi: 10.1186/s40068-024-00378-z.

[26] Reff A., Mintz D., Naess L.: The O3 NowCast: US EPA's method for characterizing and communicating current air quality, *USEPA/O3-Nowcast: EPA*, 2019.

[27] S., Y. W., H. Y., Z. H., J. C., Guo W.X.: TAERT: Triple-Attentional Explainable Recommendation with Temporal Convolutional Network, *Neurocomputing*, vol. 567, pp. 185–200, 2021. doi: 10.1016/j.ins.2021.03.034.

[28] Streatfield P.K., Karar Z.A.: Population challenges for Bangladesh in the coming decades, *Journal of health, population, and nutrition*, vol. 26(3), p. 261, 2008. doi: 10.3329/jhpn.v26i3.1894.

[29] Weerakody P.B., Wong K.W., Wang G., Ela W.: A Review of Irregular Time Series Data Handling with Gated Recurrent Neural Networks, *Neurocomputing*, vol. 441, pp. 161–178, 2021. doi: 10.1016/j.neucom.2021.02.046.

[30] l Wu C., d He H., f Song R., h Zhu X., r Peng Z., y Fu Q., Pan J.: A Hybrid Deep Learning Model for Regional O3 and NO2 Concentrations Prediction Based on Spatiotemporal Dependencies in Air Quality Monitoring Network, *Environmental Pollution*, vol. 320, 2023. doi: 10.1016/j.envpol.2023.121075.

[31] Zhang J., Li S.: Air Quality Index Forecast in Beijing Based on CNN-LSTM Multi-Model, *Chemosphere*, vol. 308, 2022. doi: 10.1016/j.chemosphere.2022.136180.

[32] Zhang Y., Bocquet M., Mallet V., Seigneur C., Baklanov A.: Real-time air quality forecasting, part I: History, techniques, and current status, *Atmospheric environment*, vol. 60, pp. 632–655, 2012. doi: 10.1016/j.atmosenv.2012.06.031.

[33] Zhao Z., Wu J., Cai F., Zhang S., Wang Y.G.: A Hybrid Deep Learning Framework for Air Quality Prediction with Spatial Autocorrelation During the COVID-19 Pandemic, *Article*, (1015), 2023. doi: 10.1038/s41598-023-28287-8.

[34] Zhu D., Cai C., Yang T., Zhou X.: A Machine Learning Approach for Air Quality Prediction: Model Regularization and Optimization, *Big Data and Cognitive Computing*, vol. 2(1), pp. 1–15, 2018. doi: 10.3390/bdcc2010005.



## Affiliations

**Mahamudul Hasan**
University of Minnesota, Department of Computer Science And Engineering, Twin Cities, Minnesota, USA, munna09bd@gmail.com

**Sadab Jaowad**
East West University, Department of Computer Science And Engineering, Dhaka, Bangladesh, sadabjaowad24@gmail.com

**Ronojith Saha Roy Dipta**
East West University, Department of Computer Science And Engineering, Dhaka, Bangladesh, rsrdipta1@gmail.com

**Samia Shameem Haque**
East West University, Department of Computer Science And Engineering, Dhaka, Bangladesh, samiashameemhaque@gmail.com

**Abir Hossain**
East West University, Department of Computer Science And Engineering, Dhaka, Bangladesh, hossainabir155@gmail.com

**Mainul Hasan**
East West University, Department of Computer Science And Engineering, Dhaka, Bangladesh, kmhrifat37@gmail.com

**Anisur Rahman**
East West University, Department of Computer Science And Engineering, Dhaka, Bangladesh, anis@ewubd.edu

**Mohammad Rifat Ahmmad Rashid**
East West University, Department of Computer Science And Engineering, Dhaka, Bangladesh, rifat.rashid@ewubd.edu

**Sawkat Ali**
East West University, Department of Computer Science And Engineering, Dhaka, Bangladesh, alim@ewubd.edu

**Received**: 26.01.2024
**Revised**: 25.07.2024
**Accepted**: 21.08.2025