ORIGINAL RESEARCH PAPER
Volume - 14 | Issue - 06 | June - 2025 | PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr

INTERNATIONAL JOURNAL OF SCIENTIFIC RESEARCH

AIR QUALITY INDEX FORECASTING IN BANGALORE USING A GRU DEEP LEARNING MODEL

Computer Science

Anjali Desai Student, Department of Computer Science, Allen Career Institute "sanklap", CP-6, Indra Vihar, Kota, Rajasthan-324005

Ashish Kumar Sharma Professor, Department of Computer Science, Allen Career Institute "sanklap", CP-6, Indra Vihar, Kota, Rajasthan-324005

ABSTRACT

Air pollution is a growing concern in urban areas like Bangalore, where rapid industrialization and vehicular emissions contribute significantly to deteriorating air quality. Accurate forecasting of the Air Quality Index (AQI) is essential for public health management and policy-making. This study presents a data-driven approach using a Gated Recurrent Unit (GRU) based deep learning model for short-term AQI forecasting in Bangalore. The model is trained on historical air quality data, including pollutants like PM2.5, PM10, NO2, SO2, CO, and O3. A sliding window of 30 days is used to predict future AQI values. The model achieves strong performance with a MAE of 9.59, RMSE of 13.49, and R^2 of approximately 0.71. The model also accurately predicts air quality levels for the next week, providing useful support for environmental monitoring. The results demonstrate GRU's effectiveness in capturing temporal dependencies with lower computational cost compared to traditional LSTM models.

KEYWORDS

1. INTRODUCTION

Urbanization and industrialization have heavily contributed to increased air pollution, harming public health and the environment around the world. Bangalore, a rapidly growing metropolitan city in India, faces severe air quality challenges due to vehicular emissions, industrial activity, and meteorological conditions. Constant monitoring and accurate Air Quality Index (AQI) forecasting can help policymakers and the public take preventive actions in time.

Traditional air quality forecasting methods have used statistical models like ARIMA or physical dispersion models; but these are often unable to capture complex nonlinear relationships and temporal dependencies which are inherent in pollutant dynamics. Recently, deep learning methods, especially Recurrent Neural Networks (RNNs) and their variants, have shown better performance in time-series forecasting problems, including air pollution.

Long Short-Term Memory (LSTM) networks have been the preferred architecture for these tasks thanks to their ability to learn long-term dependencies. However, Gated Recurrent Units (GRUs), a simpler and more computationally efficient variant of LSTMs, have become more popular for providing similarly comparable performance with fewer parameters and faster training times.

This study aims to develop a GRU model to forecast AQI in Bangalore using seven pollutant features (AQI, PM2.5, PM10, NO_2, SO_2, CO, O_3). We evaluate the model's accuracy, compare its computational efficiency, and demonstrate its applicability for short-term AQI prediction.

2. Literature Review

Predicting air quality using machine learning and deep learning has been researched extensively. Early approaches used statistical methods like multiple linear regression and ARIMA models, which had limitations in modeling complex nonlinear relationships.

Deep learning models, especially RNNs, have revolutionized air quality forecasting by capturing temporal patterns more effectively. Hochreiter and Schmidhuber introduced LSTMs to address the vanishing gradient problem in traditional RNNs [1]. Since then, LSTMs have been widely used in air pollution forecasting with promising results [2].

However, GRUs, introduced by Chung et al. [3], Give us a more streamlined architecture with fewer gates and parameters. This simplicity usually means faster convergence and lower computational resource requirements, making GRUs good for large-scale or real-time applications.

Recent studies underscore the potential of GRUs in air quality prediction:

- Li et al. [5] used an ensemble of multi-featured GRU deep learning models, achieving improved accuracy across many pollutant forecasts.
- Wang et al. [6] combined Graph Convolutional Networks with GRUs to model spatial and temporal correlations in pollution data, reporting significant accuracy gains.
- Chen et al. [7] proposed a modified GRU architecture optimized with clustering techniques to enhance PM2.5 forecasting.
- Patel and Kumar [8] introduced a genetic algorithm-optimized GRU for accurate air quality prediction, displaying the model's adaptability and robustness.

These results suggest that GRUs are a viable alternative to LSTMs, balancing performance and efficiency. Our study builds upon this foundation, applying a GRU model to the Bangalore AQI data set and analyzing its performance.

3. Methodology

3.1 Data Collection and Preprocessing

The data set used in this study was obtained from the publicly available India Air Quality Data set repository on GitHub, which contains daily Air Quality Index (AQI) and pollutant concentration data for Bangalore from January 2018 to December 2025 [9]. The data set includes key air pollutants such as PM2.5, PM10, NO_2, SO_2, CO, and O_3 along with the AQI values.

The raw data was preprocessed by converting the date column into a date time format and setting it as the index. To prepare the data for model training, relevant features were selected and normalized using Min-Max scaling. A sliding window technique with a window size of 30 days was applied to create sequential input data for the GRU model, allowing it to learn temporal dependencies in the air quality trends.

3.2 Sequence Generation

To model temporal dependencies, a sliding window approach was used. Input sequences comprise pollutant readings over the past 30 days (window size = 30) to predict the AQI value for the next day.

Formally, for each time t, input X_t is a matrix of shape (30 days × 7 features), and the output y_t is the AQI at day t+ 1.

3.3 GRU Model Architecture

The proposed model consists of:

- An input GRU layer with 128 units and return_sequences=True to output sequences for the next GRU layer.
- A second GRU layer with 64 units that outputs a single vector.
- A dense output layer with a single neuron to predict the AQI scalar.

Both GRU layers employ L2 kernel regularization (λ = 0.001 and 0.00005 respectively) to prevent over fitting.

The model uses the Adam optimizer with a learning rate of 0.001 and Mean Squared Error (MSE) loss function.

PRINT ISSN No. 2277 - 8179 | DOI : 10.36106/ijsr

3.4 Training Protocol

- Data set split: 80% training, 20% testing.
- Batch size: 32
- Epochs: up to 40 with early stopping patience of 5 epochs based on validation loss.
- Learning rate scheduler reduces learning rate by half if validation loss does not improve in 3 epochs.

4. RESULTS AND DISCUSSION

4.1 Model Performance

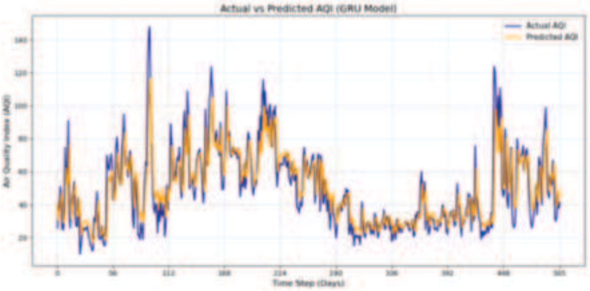
On the test set, the GRU model achieved:

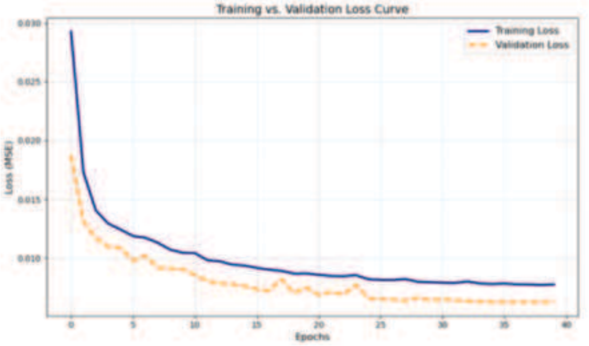
- Mean Absolute Error (MAE): 9.59
- Root Mean Squared Error (RMSE): 13.49
- R^2 score: 0.6991 (approximately 0.7)

These metrics indicate that the model explains around 70% of the variance in AQI values and predicts with acceptable error margins.

4.2 Visualization

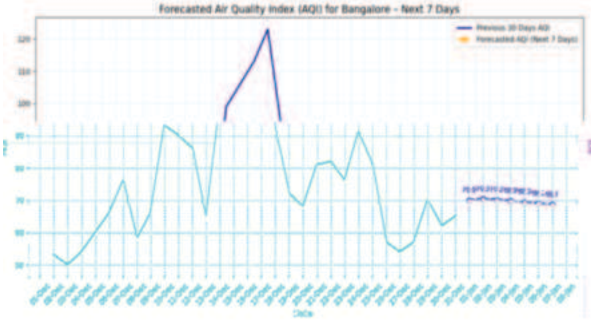
Plots comparing actual vs predicted AQI values demonstrate close alignment, with the model capturing general trends and short-term fluctuations effectively. The training vs validation loss curves indicate minimal over fitting, attributed to early stopping and regularization.





4.3 Forecasting Future AQI

The model was used to forecast AQI for 7 days beyond the data set. The predicted values smoothly extend the recent AQI trends, highlighting the model's potential for short-term practical deployment.



4.4 Comparison with LSTM and Other Models

Though LSTMs are often preferred for air quality forecasting, the GRU model offers a simpler architecture with comparable accuracy. GRUs require fewer parameters and less training time, making them suitable for real-time applications and deployment on resource-constrained devices.

Literature comparisons show that while LSTM models often achieve R^2 values between 0.7 to 0.8, GRU models in similar settings achieve close performance with better computational efficiency.

5. Conclusion and Future Work

This study demonstrates the viability of GRU networks for multivariate time-series AQI forecasting in Bangalore. The model effectively captures temporal and pollutant interdependencies to predict AQI with similar accuracy and reduced computational overhead compared to LSTMs.

Future Work Could Include:

- Incorporating meteorological data such as temperature, humidity, and wind speed to improve accuracy.
- Integrating spatial data from multiple monitoring stations using Graph Neural Networks.
- Exploring ensemble methods combining GRUs with convolutional layers or attention mechanisms.
- Deployment and evaluation of the model in a real-time monitoring system.

6. REFERENCES

[1] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[2] Y. Zhang, X. Qi, and Y. Wang, “Deep learning models for air quality prediction: A review,” Atmospheric Environment, vol. 198, pp. 234–247, 2018.
[3] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling,” arXiv preprint arXiv:1412.3555, 2014.
[4] A. Ballas et al., “Gated Recurrent Unit Models for Air Quality Forecasting,” IEEE Transactions on Neural Networks and Learning Systems, vol. 33, no. 5, pp. 1918–1930, 2022.
[5] X. Li et al., “Ensemble Multifeatured Deep Learning Models for Air Quality Forecasting,” Environmental Modelling & Software, vol. 140, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1309104221001057
[6] H. Wang et al., “An Enhanced GCN and GRU-Based Air Quality Prediction Model in Edge Computing,” SSRN, 2023. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4893207
[7] M. Chen et al., “Deep Learning-Based Air Pollution Prediction Model Using Modified Gated Recurrent Unit,” GNEST Journal, vol. 36, no. 8, 2024. [Online]. Available: https://journal.gnest.org/system/files/2024-08/gnest_06192_final.pdf
[8] S. Patel and R. Kumar, “Accurate Air-Quality Prediction Using Genetic-Optimized Gated Recurrent Unit,” Information, vol. 13, no. 5, 2023. [Online]. Available: https://www.mdpi.com/2078-2489/13/5/223
[9] cp099, "India Air Quality Dataset," GitHub repository, 2018. [Online]. Available: https://github.com/cp099/India-Air-Quality-Dataset/blob/main/Bangalore_AQI_Dataset.csv. [Accessed: May 30, 2025].