Advance Sustainable Science, Engineering and Technology (ASSET)
Vol. 7, No. 3, July 2025, pp. 02503012-01 ~ 02503012-16
ISSN: 2715-4211 DOI: https://doi.org/10.26877/asset.v7i3.1589

Enhanced Air Quality Prediction Using AI: A Comparative Study of GRU, CNN, and XGBoost Models

Kayam Saikumar[1*], Munugapati Bhavana[2], Rayudu Prasanthi[3], Singaraju Suguna Mallika[4], Deepthi Kamidi[5], Naveen Malik[6], Kapil Joshi[7]

[1]Department of Electronics and Communication Engineering, Koneru Lakshmaiah Education Foundation, Hyderabad-500075, Telangana, India.

[2]Department of Computer Science and Engineering, MLR Institute of Technology, Hyderabad 500043, Telangana, India.

[3]ECE Department, Aditya Collage of Engineering and Technology (A), Surampalem, Kakinada, Andhra Pradesh, 533437, India.

[4]Department of CSE, CVR College of Engineering, Telangana 501510, India.

[5]Department of Computer Science and Engineering, Vignan Institute of Technology and Science, Telangana 508284, India

[6]Department of Computer Science and Engineering, Maharishi Markandeshwar (Deemed to be University), Mullana, Ambala, Haryana, India.

[7]Department of Computer Science & Engineering, Uttaranchal Institute of Technology (UIT), Uttaranchal University, Dehraudn 248007, Uttarakhand, India.

*saikumar.kayam@klh.edu.in

Abstract: Weather monitoring is vital due to environmental changes and rising air pollution, which affects health and lifestyles. Accurate air quality prediction models are essential yet challenging due to complex weather-pollution interactions. This study employs explainable deep learning and machine learning techniques—GRU, CNN, and XGBoost—on a custom dataset of 100,000 samples with 15 features, including PM2.5, PM10, humidity, and temperature. Using SHAP for interpretability, the GRU model outperforms others with 98.56% accuracy, 98.43% Recall, and 98.52% True Positive Rate. Temperature, humidity, gases, and pressure are key variables influencing predictions. The proposed model achieves high mAP and precision, surpassing existing methods and demonstrating effective real-time forecasting under diverse weather conditions.
Keyword: GRU-based air quality forecasting, deep learning for AQI, spatiotemporal air pollution modeling, PM 2.5 and AQI.

(Received 2025-01-27, Revised 2025-04-23, Accepted 2025-05-24, Available Online by 2025-06-23)

1. Introduction

The lifestyle associated with global urbanization has significantly affected weather patterns and environmental pollution. Urban expansion, industrial growth, and the rising demand for energy have severely degraded air quality worldwide. Emissions from power plants, factories, transportation, and other sources have consistently introduced pollutants into the atmosphere, contributing to significant health risks such as respiratory illnesses, cardiovascular diseases, and even cancer in extreme cases. Pollutants like particulate matter (PM), particularly PM2.5, have impacted respiratory organs due to their microscopic size and deep penetration. PM2.5 has been estimated to be 2.5 microns or smaller and has been measured using airborne methods. Long-term health issues, such as lung infections and bloodstream complications, have emerged due to the release of harmful chemicals into the atmosphere. Various pollutants, including PM10, SO2, and NO2, have reduced human lifespan and disrupted environmental stability. For public safety, effective air quality management and forecasting systems have become necessary. Governments have taken steps based on recommendations from governing bodies, though these measures have not always been entirely accurate or beneficial for the public. Advanced weather monitoring algorithms have been required to track pollution effectively. On the other hand, deep learning and AI models have been widely adopted to provide fast and accurate solutions for critical problems. As part of this research, CNN with gated recurrent layers has been adapted, incorporating customizations to enhance the model's performance. These modifications, along with a robust dataset, have helped overcome the complexity limitations of existing models. Meteorological elements such as humidity, temperature, and pollutants have been found to significantly impact both the environment and human health. Existing and modern forecasting models have often relied on outdated algorithms, leading to less accurate observations. PM2.5 pollution levels have remained a key factor in calculating the Air Quality Index (AQI), which has been monitored by top countries such as the United States, European nations, and Asian countries. The accuracy, recall, F-measure, and sensitivity achieved by earlier models have been limited. The proposed CNN model has demonstrated the ability to identify dynamic weather conditions accurately. It has utilized daily weather report samples from various sources to assess whether weather conditions are harmful or not. Studies on PM2.5 have helped identify influencing factors such as pressure, temperature, and humidity. The CNN model, with its reliable gated reference, has proven effective in accurately forecasting air quality. Previous models have primarily focused on government data rather than dynamic weather conditions when predicting AQI. Public health officials have observed numerous emerging health issues linked to traditional models. Risk assessment and decision-making have been crucial in weather monitoring applications, and the proposed deep learning-based CNN method has provided a viable solution. Figure 1 clearly explains the air pollution and mental disorder risk assessment model. In this study, the statistical analysis of environmental conditions, including PM2.5, PM10, and SO2 levels, has been performed using the proposed method. Accurately calculating the Air Quality Index (AQI) is possible by analysing how much pollution has occurred in the environment using statistical methods. To effectively control and mitigate air pollution, thereby protecting both human health and the environment, it is essential to understand the factors influencing air pollutants and their evolving trends. This knowledge can be leveraged to evaluate and predict changes in air quality. Additionally, key departments can gain valuable insights into current air quality conditions through air quality forecasting, providing a strong theoretical foundation for research and policymaking. Moreover, policies for reducing and controlling air pollution can be tailored to meet specific local needs. It also offers critical feedback and recommendations to help decision-makers enhance air quality in the future while optimizing costs.

The machine learning models like Support vector machine (SVM), Genetic algorithm (GA), Random Forest Algorithm (RFA) and other models have faced difficulty at weather assessment. The mean average precision (mAP) and true positive rate is less with exiting models. The dynamic dataset can reduces the measures and providing less accuracy and sensitivity. Unbalanced instances and classes in the samples cause misclassification due to label imbalance.

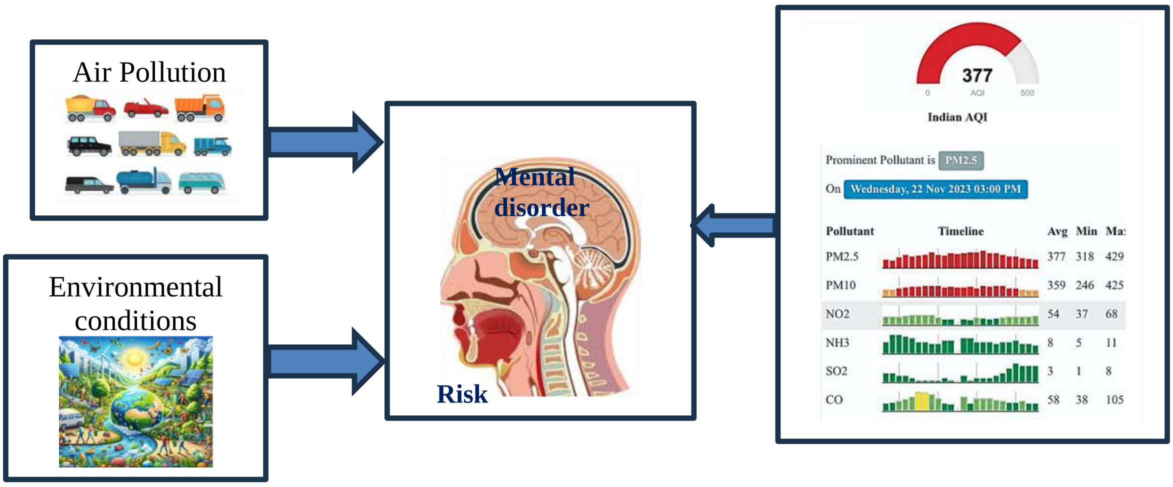


Figure 1: Air quality index and environmetal disorers with Risk assement screen using statistical analysis

The model parameters like batch size, kernel size, epochs and post processing has controlled the model in all diverse usage conditions. In this manuscript section 1 explains introduction of the model, section 2 giving literature of the existing models such as limitations and research gaps. Section3 presents materials and methods of proposed model and section 4 explains results and discussions finally presented the conclusion in section 5.

1.1. Brief details of AQI using ML and DL models

In this section a brief discussion about weather monitoring and pollutant AQI analysis has performed. P. Kumar [3] speculation that more people living in urban areas means more cars on the road, which means more pollution in the air. Overly high concentrations of air pollutants in urban areas pose a serious health risk to city people. Our current understanding of the consequences of exposure to extremely high levels of pollution that are both geographically and temporally confined is limited. Systems of static and sparsely distributed measurement stations provide the backbone of traditional approaches to air best tracking. To develop effective real-time methods for exposure control, To detect pollution hotspots and capture tempo-spatial heterogeneity, these are too costly. The ancient technique is being revolutionised by a new low-cost micro-scale sensing age. making it possible to collect statistics in real-time using capillaries. Their less accurate data raises the question of whether there is value in it. E. D. Schraufnagel [4] conducted research Air pollution is one of the biggest threats to both the environment and human health. Worldwide, exposure to outdoor particulate matter (particulate particles having an aerodynamic diameter < 2.5 μm) is the sixth leading cause of death, with more than 4.2 million deaths and more than 103 million disability-adjusted life years, according to the Global Burden of Disease Report. Both short-term effects, like difficulty breathing or heart palpitations, and long-term effects, which may impact any organ in the body, are possible due to air pollution. It has the potential to bring about or worsen a great deal of unsavoury fitness conditions. Because tiny and ultrafine particles can penetrate dorgans without triggering systemic inflammatory responses, tissue damage may occur promptly as a result of pollution toxicity. Y.-F. Xing [5] carried out studies Air pollution is extremely harmful to ecosystems and people alike. Worldwide, exposure to outdoor particulate matter (defined as particles with an aerodynamic diameter less than 2.5 μm) is responsible for more than 4.2 million fatalities and more than 103 million years of life lost due to disability, as stated in the Global Burden of Disease Report. Both immediate & delayed consequences, such as trouble breathing or irregular heartbeats, which may impact any organ in the body, are possible due to air pollution. It has the potential to bring about or worsen a great deal of unsavoury fitness conditions. Because tiny and ultrafine particles can penetrate organs without triggering systemic inflammatory responses, tissue damage may occur promptly as a result of pollution toxicity. X. Qi [8] investigated the fact that, particularly in China's megacities, air pollution from particulate matter 2.5, Ozone and particle matter 10 are becoming more dangerous for human health. Weather has an impact on air pollution's dispersion as well as concentration, which also have a significant impact on these processes. We look at the connections between Beijing's air pollution levels and weather conditions from 2017 to 2018 in this article. We note that: one meteorological variable has a small impact on pollutant concentrations; temperature-wind velocity, temperature-strain, and humidity-wind velocity aggregates are highly correlated with pollution awareness, suggesting that multiple meteorological variables interact to influence pollutant concentrations; even when considering the same pollutant under the same climatic circumstances, concentrations can fluctuate due to the influence of several meteorological variables. Our research has the potential to improve city control performance and aid in air quality prediction in accordance with current weather conditions. S. Al-Janabi [9] one of the most pressing issues facing the industry right now is the need to study and treat the growing problem of air pollution caused by technological developments. In recent times, there has been a notable rise in the concentration of pollutants in the environment. This research describes an effort to use recurrent neural networks (RNNs) and deep learning techniques to create an intelligent air pollution concentration forecast for the next two days. It's ideal mode of operation is further identified by applying a particle swarm optimisation (PSO) approach. A newly developed model is the Smart Air Quality Prediction Model (SAQPM). Unsupervised learning, in particular long short-term memory (LSTM), and optimisation, in particular PSO, are required. The dataset is then split in half for testing and training, using the ten cross-validation principles. The complete pipeline model has improved through Ros with specific packages. The exiting models facing limitations like less accuracy, model bandwidth and model perseverance.

Table 1. Recent highlights of weather monitoring using DL and ML models analysis.

Ref	Author	Area of research	Methods	Key finding
[31]	Ayoub, A. et.al	Machine learning and deep learning models applied on radioactive dataset to predict contaminated elements.	ML-DL enabled weather forecasting on dynamic dataset.	Accuracy was improved with radioactive dataset.
[32]	Gong, Y. et.al	Deep learning depend temperature monitoring.	CNN and LSTM related temperature monitoring model on dense dataset.	Spatial and temporal features were detected and forecasting the accuracy.
[33]	Ben Bouallègue, Z et.al	DL and AI based weather dynamic nature analysis.	Probable data analysis with DL models.	Potential weather data handling with DL models by complex pipelines.
[34]	Akilan, T et.al	In perception agriculture monitoring the weather and climate.	GRU with CNN model along with IoT for agriculture applications	Weather monitoring benefits and precision in agriculture.
[35]	Sun, Y et.al	High resolution custom data making	Machine and deep learning models on metrological dataset.	Highlighting the effective generation of precise weather data.

The above table briefly explains about various weather monitoring models related to deep and machine learning. The existing models are facing low level instances, mis-classes, overfitting and underfitting issues.

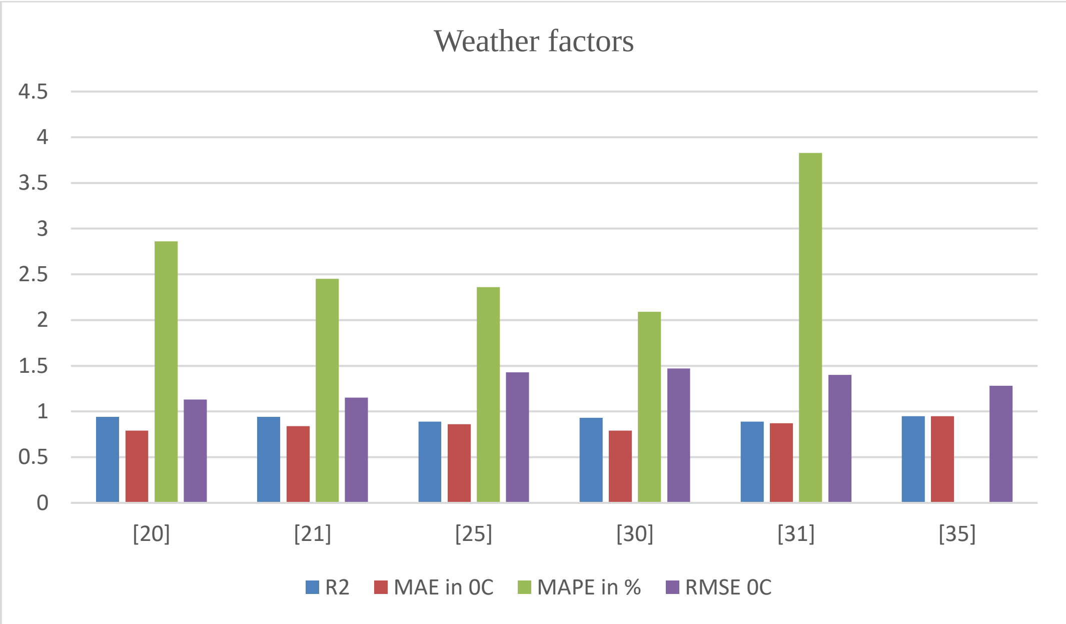


Figure 2. comparative measures examination

The above figure 2 and table 2 clearly explains about various performance and its limited metrics, from this we can conclude that existed models facing issues of dynamic weather dataset.

Table 2. various comparative measures analysis

	R^2	MAE in ^0C	MAPE in %	RMSE ^0C
[20]	0.94	0.79	2.86	1.13
[21]	0.94	0.84	2.45	1.15
[25]	0.89	0.86	2.36	1.43
[30]	0.93	0.79	2.09	1.47
[31]	0.89	0.87	3.83	1.40
[35]	0.95	0.95	3,83	1.28

In this section a brief discussion of literature survey has been collected and limitations of study can be added. The methods like machine learning techniques are unable to work on dynamic dataset as well as cannot give accurate R^2.

2. Methods

In this section a brief analysis on weather monitoring models has been performed as well as proposed one novel model nothing but GRU deep learning technique. While a few of deep learning models do take weather into account when predicting air quality, this information is primarily utilized as input data, and the impact of weather on this prediction has received very little attention in the scientific community. As an example, our current understanding of how weather factors impact air quality forecast using deep learning algorithms is limited. The "black box" nature or lack of explainability that was described earlier is characteristic of deep learning models. The deep learning model's strength lies in its capacity to match complicated data relationships, making it possible to predict air quality using both weather and air quality data. Research into the complicated interplay between air quality forecasts, weather, and related topics remains a challenging and intricate area of study.

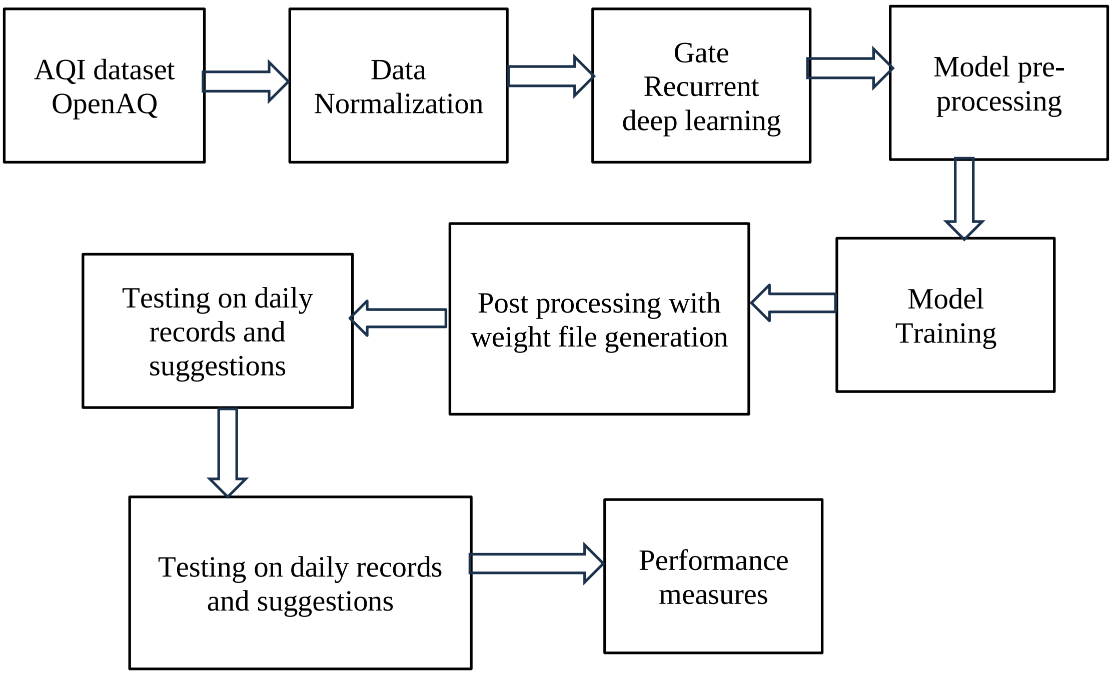


Figure 3. Block diagram of Gate Recurrent Deep learning for AQI statistics analysis while dynamic environmental change conditions

The above Figure 3 explains the process of air quality prediction using deep learning techniques, specifically leveraging Gated Recurrent Units (GRUs). The process begins with the AQI dataset (OpenAQ), which provides historical data on air quality, including key parameters such as PM2.5 levels. Dynamic variables can normalize the dataset using a scaling mechanism, which has improved with uniform convergence during training. The pre-processed dataset can be used well with the gated recurrent CNN framework, which focuses on time analysis, temporal AQI prediction, and health risk management. The preprocessing leads to smooth training, feature extraction, easy further analysis, and making it into a sequence. The model training can handle learning features, historical analysis, pattern identification, and the relationship between variables, such as temperature, pressure, and humidity. The air quality in weather can be estimated using the proposed method and post-processing; the weight file and testing will also be part of the PM2.5 analysis. We used the LSTM model to look at the wrong predictions and classifications and gave feedback to improve the model. The unseen data missed data, and suggestions can be created as synthetic data with generative AI techniques. The predicted air quality index depends on actionable information like results and ML analysis. Model metrics can generate performance measures such as accuracy, recall, sensitivity, and throughput. We can use the iterative feedback loop daily records to update the model and improve its efficiency. We can integrate this process with dataset iterations to improve the quality of feature extraction.

Algorithms for Time Series Forecasting and Air Quality Prediction

Step 1. Distinguishing Time Series: Stationary vs. non-stationary

Determine if the time series is stationary (constant mean and variance) or non-stationary. For non-stationary series, transformations or differencing are applied to achieve stationarity.

Equation for Differencing to achieve stationarity:

$Y\_t' = Y\_t - Y\_\{t-1\}$ ------ (1)

Step 2. LSTM (Long Short-Term Memory)

LSTM networks are designed to handle long-term dependencies in time-series data. They utilize cell states and hidden states to store long- and short-term information.

Equations for LSTM gates:

Forget Gate: $f\_t = \sigma(W\_f \cdot [h\_\{t-1\}, x\_t] + b\_f)$ ------ (2)

Input Gate: $i\_t = \sigma(W\_i \cdot [h\_\{t-1\}, x\_t] + b\_i)$ ------ (3)

Output Gate: $o\_t = \sigma(W\_o \cdot [h\_\{t-1\}, x\_t] + b\_o)$ ------ (4)

Step 3. RNN (Recurrent Neural Network)

Recurrent Neural Networks (RNNs) use internal memory to process sequential data and predict future values.

Equation for RNN hidden state:

$h\_t = f(W\_hh \cdot h\_\{t-1\} + W\_xh \cdot x\_t + b\_h)$ ------ (5)

Step 4. GRU (Gated Recurrent Unit)

GRUs are a simplified version of LSTMs with fewer parameters, making them faster to train. They use update and reset gates for controlling information flow.

Equations for GRU gates:

Update Gate: $z\_t = \sigma(W\_z \cdot [h\_\{t-1\}, x\_t])$ ------ (6)

Reset Gate: $r\_t = \sigma(W\_r \cdot [h\_\{t-1\}, x\_t])$ ------ (7)

Step 5. CNN+LSTM

Combines Convolutional Neural Networks (CNNs) for feature extraction with LSTMs for capturing long-term dependencies.

Feature extraction via CNN: $y\_i = ReLU(W\_i * x + b\_i)$ ------ (8)

Step 6. ARIMA (Autoregressive Integrated Moving Average)

ARIMA is used for modeling time series data with autoregressive and moving average components.

General ARIMA equation:

$$Y\_t = c + \varphi\_1\, Y\_\{t-1\} + \ldots + \varphi\_p\, Y\_\{t-p\} + \varepsilon\_t + \theta\_1\, \varepsilon\_\{t-1\} + \ldots + \theta\_q\, \varepsilon\_\{t-q\}$$

Step 7. Random Forest

Random Forest builds multiple decision trees and aggregates their predictions through majority voting or averaging.

Step 8 Voting Classifier

Combine multiple base models (e.g., SVM, Random Forest, Logistic Regression). Use majority voting (hard voting) or averaged probabilities (soft voting) to predict outcomes. Aggregate the outputs of individual estimators to obtain the final prediction.

Air quality prediction using deep learning models presents both challenges and opportunities. One key drawback lies in the "black-box" nature of deep learning, which makes it difficult to develop reliable models that incorporate weather data. Additionally, there is a limited understanding of how weather conditions influence air quality forecasts when utilizing deep learning approaches. Addressing these challenges, this work highlights the impact of weather on air quality prediction by leveraging explainable deep learning techniques. By elucidating how weather elements such as temperature, humidity, and atmospheric pressure affect predictions, the model achieves improved accuracy and reliability. Applying deep learning in air quality forecasting enhances prediction precision, making it more practical for real-world applications. This enables individuals to plan their activities effectively, take necessary precautions to safeguard their health, and respond promptly to changing air quality conditions. Understanding the current air quality state is essential for implementing appropriate preventative and control measures, ensuring efficient and timely management of air quality concerns. The benefits of this approach are twofold. First, it leads to the development of more trustworthy and accurate deep learning models for air quality prediction. Second, it provides valuable insights for travelers, allowing them to plan their trips more sensibly and take timely precautions to protect their health. This enhanced predictability ensures that people can better manage their exposure to harmful pollutants, improving public health outcomes. The implementation of this system involves several modules. The first module facilitates data exploration by enabling data loading into the system. The processing module reads and preprocesses the data for further analysis. This module also splits the data into training and testing sets to ensure robust model evaluation. The model-building phase includes the application of various algorithms such as LSTM, RNN, GRU, CNN+LSTM, CNN+GRU, ARIMA, Random Forest, KNN-SHAP, MLP, and the voting classifier. These algorithms contribute to generating accurate predictions, with their performance measured through calculated accuracy metrics. Finally, the prediction module displays the forecasted air quality results, enabling users to make informed decisions based on reliable, data-driven insights.

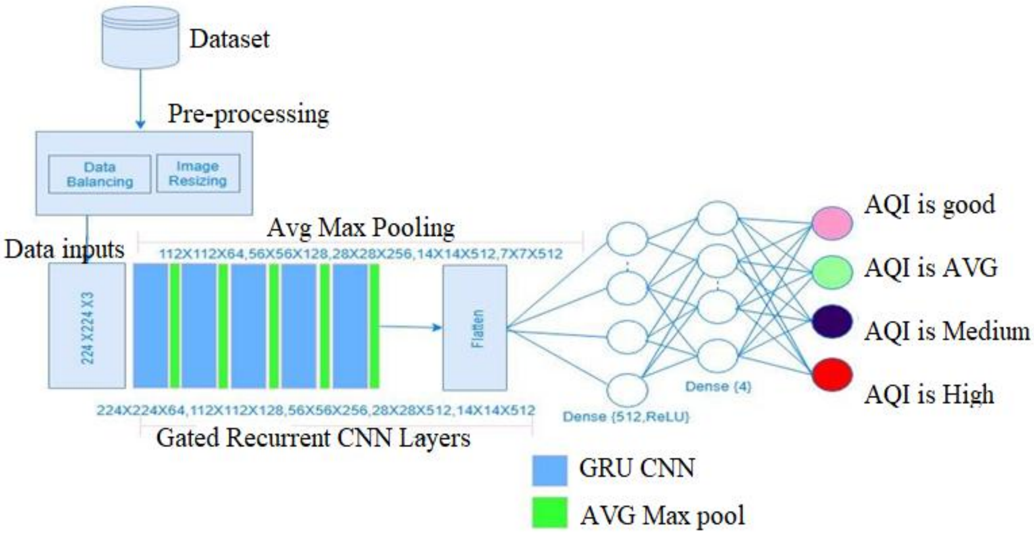


Figure 4. Architectural model of GRU CNN End to End Layers functioning

This figure 4 represents a deep learning pipeline designed for Air Quality Index (AQI) classification, showcasing how data flows through various stages of preprocessing, feature extraction, and classification. The process begins with a dataset, likely containing environmental data such as images or sensor readings related to air quality. Preprocessing is an essential step, involving data balancing to ensure equal representation of all AQI classes (Good, Average, Medium, High) and image resizing to standardize input dimensions (e.g., 224×224×3), making the data suitable for the model and reducing computational complexity.

The pre-processed data serves as input to the model, passing through Gated Recurrent Convolutional Neural Network (GRU-CNN) layers. These layers combine the strengths of Convolutional Neural Networks (CNNs) for extracting spatial features from the input and Gated Recurrent Units (GRUs) for capturing temporal or sequential dependencies, which is particularly beneficial if the data is time-dependent, such as pollution readings over time. The feature maps are progressively refined through multiple layers, where spatial dimensions reduce (e.g., from 224×224×64 to 7×7×512), while the depth increases to capture higher-level abstract features.

To further process the features, the model employs average and max pooling (Avg Max Pooling) to down sample the feature maps. This hybrid pooling strategy ensures the retention of global context through averaging while preserving prominent features using maximum pooling. The output of the pooling layers is then flattened into a one-dimensional vector, which feeds into the fully connected dense layers. The first dense layer, consisting of 512 neurons with a ReLU activation function, processes the extracted features, followed by the final dense layer with 4 neurons. These output neurons correspond to the AQI categories: "Good," "Average," "Medium," and "High," each representing a specific air quality condition ranging from healthy air to hazardous levels.

This architecture effectively combines feature extraction and classification capabilities, leveraging

CNNs for spatial data and GRUs for any sequential dependencies in the input. The hybrid Avg Max Pooling mechanism ensures robust feature preservation, aiding in accurate classification. Such a model can be applied to air pollution monitoring systems for real-time AQI assessment, weather forecasting models to predict air quality trends, and health advisory systems to issue warnings based on AQI levels. Overall, this pipeline is designed to provide an efficient and accurate solution for air quality classification and related applications.

2.1. Dataset

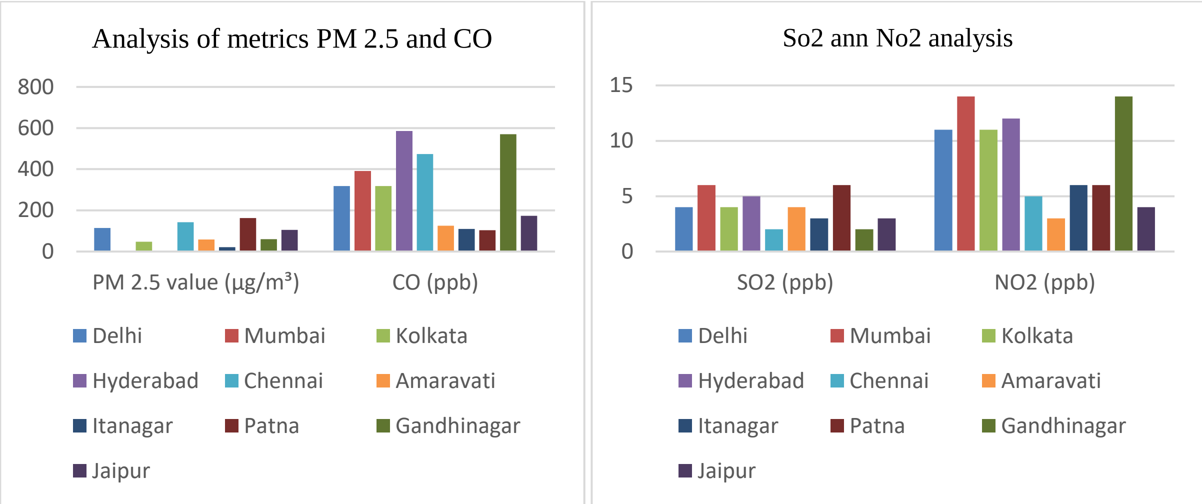
OpenAQ is a platform that provides global air quality data collected from government and research-grade sources. It offers access to data on various air pollutants, including PM2.5, PM10, NO2, CO, O3, and more. This data can be used to monitor and analyze air quality across different regions. For data exploration, tools like Python (with libraries such as Pandas and Matplotlib) or R can be employed for tasks like data cleaning, visualization, and analysis. In the model-building phase, machine learning or deep learning models such as Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), or Random Forest can be applied to predict air quality. Finally, the analysis involves examining the impact of weather parameters such as temperature, humidity, and atmospheric pressure on air pollution levels to gain insights into how these factors influence air quality over time [26].

3. Results and Discussion

In this section a brief results on AQI statistical experimental results were explained related to air quality and weather conditions estimation. The trained model has weight file which was placed on Opensource driveHq cloud. Testing region information and it is daily report has been applied to weight file, then it generates the complete air quality and weather report information dynamically also it will predict the upcoming values.

Table 3. Dynamic dataset on 01/12/2024 some state wise analysis

Region	PM 2.5 value (μg/m³)	CO (ppb)	SO2 (ppb)	NO2 (ppb)
Delhi	115	319	4	11
Mumbai	55	390	6	14
Kolkata	47	319	4	11
Hyderabad	35	586	5	12
Chennai	142	474	2	5
Amaravati	59	126	4	3
Itanagar	21	110	3	6
Patna	163	104	6	6
Gandhinagar	60	570	2	14
Jaipur	105	173	3	4



(a) (b)

Figure 5. weather quality analysis on Daily dynamic 3 (a) PM 2.5, CO 3 (b) SO_2 and NO

The above figures 5(a) and 5(b) clearly illustrate the analysis reports of various states for PM2.5, CO, SO2, and NO metrics. These datasets are instrumental in obtaining accurate weather conditions and AQI (Air Quality Index) reports through the proposed GRU deep learning model. The trained model can be validated using these test samples, providing predictions on potential health risks throughout the day also show in in table 2 And table 3.

Table 4. Weather analysis using timestamp

Region	Health Risk	Time Stamp
Delhi	65	1/12/24 : 12:23:12
Mumbai	51	1/12/24 : 12:23:13
Kolkata	42	1/12/24 : 12:23:14
Hyderabad	37	1/12/24: 12:23:15
Chennai	38	1/12/24: 12:23:16
Amaravati	32	1/12/24: 12:23:17
Itanagar	29	1/12/24: 12:23:18
Patna	39	1/12/24: 12:23:19
Gandhinagar	36	1/12/24: 12:23:20
Jaipur	27	1/12/24: 12:23:21

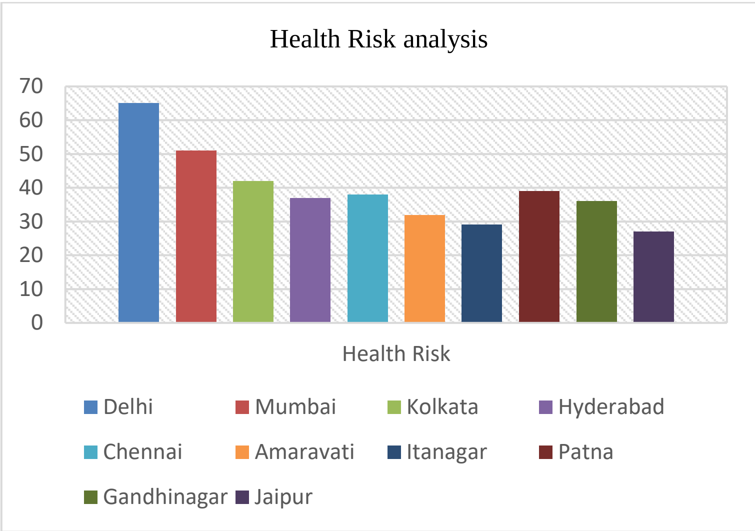


Figure 6. Complete health risk analysis generated by Proposed GRU model with time stamp.

The figure 6 presents a Health Risk Analysis based on air quality data across various cities, with corresponding health risk scores and timestamps for each location. Delhi shows the highest health risk (65), followed by Mumbai (51) and Kolkata (42), indicating poorer air quality compared to other cities. Cities like Jaipur (27) and Itanagar (29) exhibit the lowest health risk, suggesting better air quality. The timestamps indicate the exact moment when the data was recorded, providing real-time insights into the health risks associated with air pollution in each region.

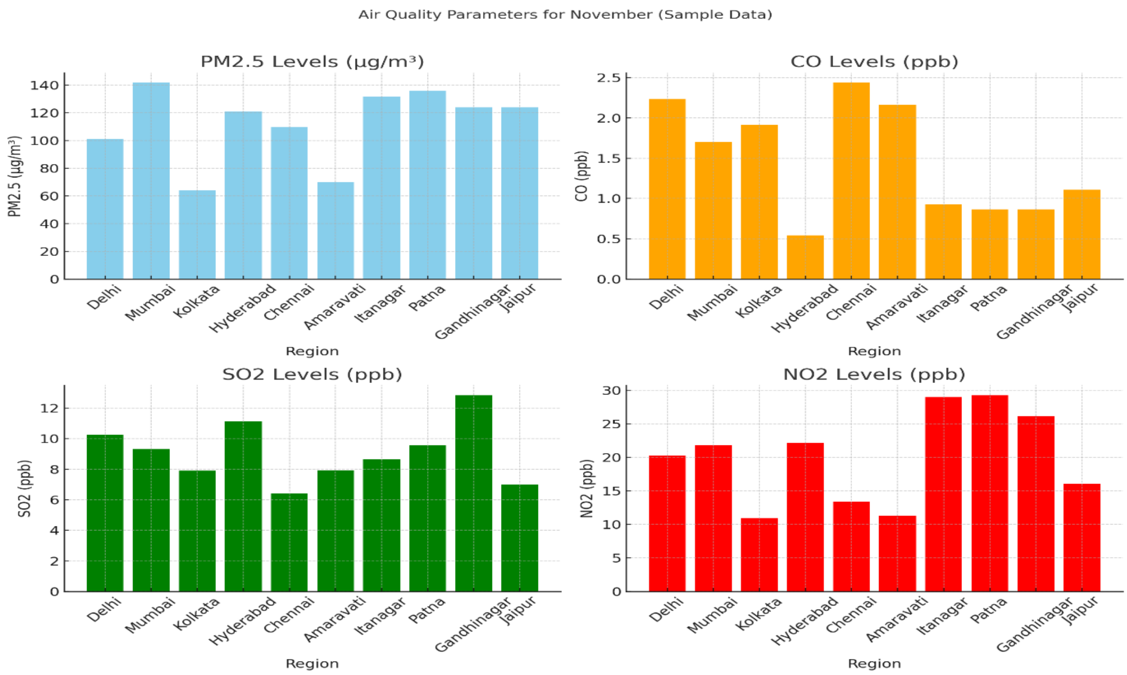


Figure 7. graphs representing hypothetical air quality data for the month of November across different regions in India

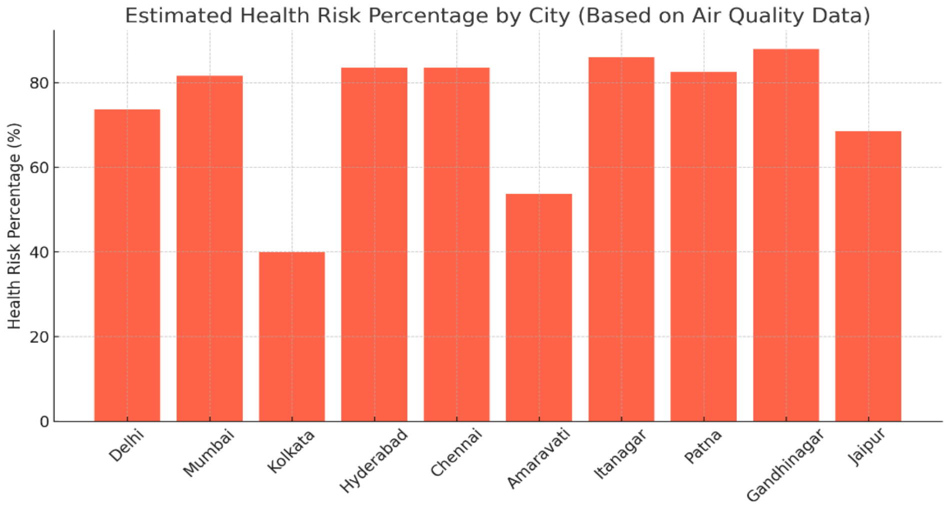


Figure 9. health risk percentage city wise

Figures 7 and 8 illustrate the estimated health risk percentage for various cities based on air quality metrics, including PM2.5, CO, SO2, and NO2 levels, during November 2024. The health risk percentage is calculated using a weighted combination of these pollutants, reflecting the potential impact of air pollution on public health. Cities like Delhi and Mumbai exhibit higher health risks due to elevated pollution levels, while cities such as Itanagar and Jaipur show relatively lower risks. This analysis highlights the varying degrees of air quality-related health risks across different regions. The data underscores the importance of targeted pollution control measures to mitigate health hazards, especially in high-risk areas.

Table 5. Performance measures of proposed model and recent models' comparison

Model	**Accuracy**	**Recall**	**Precision**	**True Positive Rate**	**R^2**	**RMSE ^0C**	**MAPE in %**
CNN	89.23	87.31	89.45	87.23	0.94	2.86	1.13
RCNN	90.24	91.76	91.47	90.28	0.92	2.93	1.16
Xboost	93.76	94.87	94.91	95.26	0.95	2.95	1.19
GRU CNN Proposed	98.56	98.43	97.93	98.52	0.97	1.96	1.10

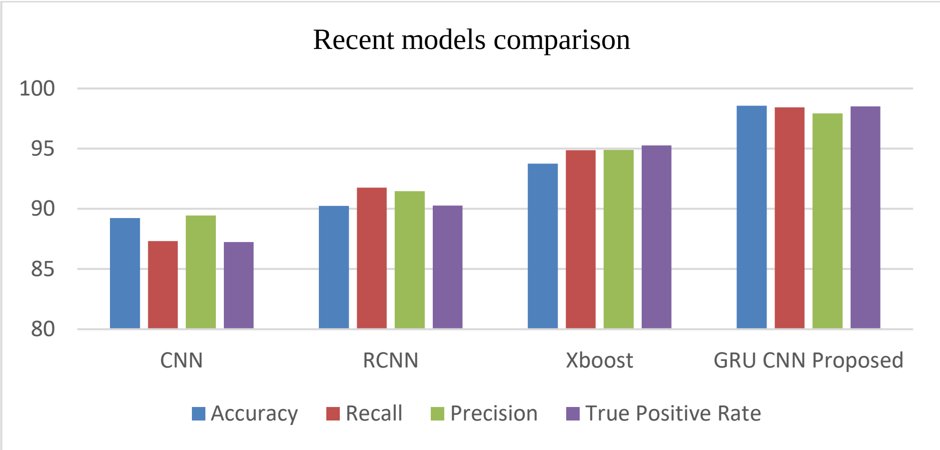


Figure 10. Recent models comparison with proposed model

The above figure 9 and table 4 clearly explains about various deep learning models comparison, in this CNN, RCNN and Xboost techniques were attained less performance analysis. The proposed model GRU CNN attains more improvement in terms of Accuracy, Recall, precision and True positive rate.

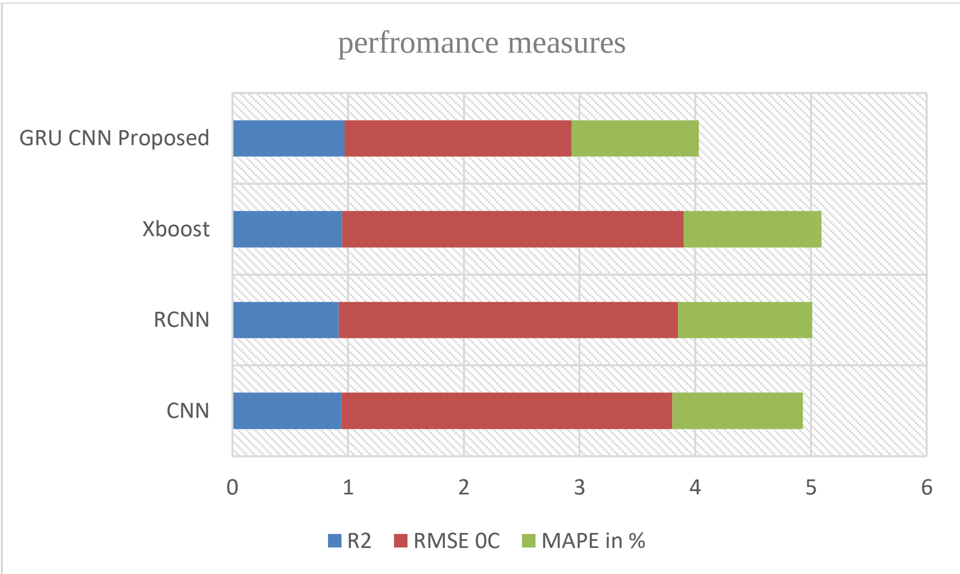


Figure: 10 complex performance measures

Figure 10 briefly explains about coefficient determination, Root mean square and mean absolute error. In this proposed model attains more improvement compared to existing ML, DL and conventional models.

4. Conclusions

This research highlights the critical role of weather conditions in enhancing the accuracy of air quality predictions. Statistical analysis of the Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) models reveals that meteorological factors such as air pressure, temperature, and humidity significantly influence air quality forecasts. The integration of weather data with other air pollutants, such as PM2.5, improves prediction accuracy, with air pressure being the most influential factor, followed by temperature and humidity. The proposed GRU model achieves remarkable performance with an accuracy of 98.56%, recall of 98.43%, precision of 97.93%, and a true positive rate of 98.52%.

These results demonstrate the reliability, interpretability, and effectiveness of explainable deep learning models in air quality forecasting. The findings contribute to more precise and trustworthy air quality predictions, supporting effective pollution management and public health protection. Future research can focus on developing advanced deep learning models incorporating real-time data and additional meteorological factors like wind speed and solar radiation for improved accuracy. Expanding explainable AI techniques and using transfer learning can enhance model adaptability, aiding real-time monitoring and timely health risk mitigation.

Acknowledgements

The authors would like to express their sincere gratitude to Koneru Lakshmaiah Education Foundation, Hyderabad-500075, for providing the necessary support, resources, and academic environment that facilitated the successful completion of this research. We also extend our heartfelt thanks to Mr. Kapil Joshi, from Uttaranchal Institute of Technology (UIT), Uttaranchal University, Dehradun-248007, for his valuable insights, encouragement, and collaborative support throughout this work.

References

[1] H. Kan, R. Chen, and S. Tong, ‘‘Ambient air pollution, climate change, and population health in China,’’ Environ. Int., vol. 42, pp. 10–19, Jul. 2012.

[2] H. Zhang, S. Wang, J. Hao, X. Wang, S. Wang, F. Chai, and M. Li, ‘‘Air pollution and control action in Beijing,’’ J. Cleaner Prod., vol. 112, pp. 1519–1527, Jan. 2016.

[3] P. Kumar, L. Morawska, C. Martani, G. Biskos, M. Neophytou, S. Sabatino, M. Bell, L. Norford, and R. Britter, ‘‘The rise of low-cost sensing for managing air pollution in cities,’’ Environ. Int., vol. 75, pp. 199– 205, Feb. 2015.

[4] E. D. Schraufnagel, R. J. Balmes, T. C. Cowl, S. D. Matteis, S.-H. Jung, K. Mortimer, R. Perez-Padilla, B.

[5] M. Rice, H. Riojas-Rodriguez, A. Sood, D. G. Thurston, T. To, A. Vanker, and J. D. Wuebbles, ‘‘Air pollution and noncommunicable diseases a review by the forum of international respiratory societies’ environmental committee, part 2: Air pollution and organ systems,’’ Chest, vol. 155, no. 2, pp. 417–426,

[6] 2019.

[7] Y.-F. Xing, Y.-H. Xu, M.-H. Shi, and Y.-X. Lian,‘‘The impact of PM2.5 on the human respiratory system,’’ J. Thoracic Disease, vol. 8, no. 1, pp. E69– E74, 2016.

[8] J. J. West, A. Cohen, F. Dentener, B. Brunekreef,T. Zhu, B. Armstrong, M. L. Bell, M. Brauer, G. Carmichael, D. L. Costa, and D. W. Dockery, ‘‘What we breathe impacts our health: Improving understanding of the link between air pollution and health,’’ Environ. Sci. Technol., vol. 50, no. 10, pp. 4895–4904, 2016.

[9] X. Zhang, X. Zhang, and X. Chen, ‘‘Happiness in the air: How does a dirty sky affect mental health and subjective well-being?’’ J. Environ. Econ. Manage., vol. 85, pp. 81–94, Sep. 2017.

[10] X. Qi, G. Mei, S. Cuomo, C. Liu, and N. Xu, ‘‘Data analysis and mining of the correlations between meteorological conditions and air quality: A case study in Beijing,’’ Internet Things, vol. 14, Jun. 2021, Art. no. 100127.

[11] S. Al-Janabi, M. Mohammad, and A. Al-Sultan, ‘‘A new method for prediction of air pollution based on intelligent computation,’’ Soft Comput., vol. 24, no. 1, pp. 661–680, Jan. 2020.

[12] Alimissis, K. Philippopoulos, C. G. Tzanis, and D. Deligiorgi, ‘‘Spatial estimation of urban air pollution with the use of artificial neural network models,’’ Atmos. Environ., vol. 191, pp. 205–213, Oct. 2018.

[13] H. Li, J. Wang, R. Li and H. Lu, "Novel analysis- forecast system based on multi-objective optimization for air quality index", J. Cleaner Prod., vol. 208, pp. 1365-1383, Jan. 2019. Show in Context CrossRef Google Scholar Kumar and P. Goyal, "Forecasting of daily air quality index in Delhi", Sci. Total Environ., vol. 409, no. 24, pp. 5517-5523, Nov. 2011.

[14] W. G. Cobourn, "An enhanced PM2.5 air quality forecast model based on nonlinear regression and back-trajectory concentrations", Atmos. Environ., vol. 44, no. 25, pp. 3015-3023, Aug. 2010.

[15] T. S. Rajput and N. Sharma, "Multivariate regression analysis of air quality index for Hyderabad city: Forecasting model with hourly frequency", Int. J. Appl. Res., vol. 3, no. 8, pp. 443-447, 2017.

[16] Y. Liu, Q. Zhu, D. Yao and W. Xu, "Forecasting urban air quality via a back-propagation neural network and a selection sample rule", Atmosphere, vol. 6, no. 7, pp. 891-907, Jul. 2015.

[17] S. Xiao, Q. Y. Wang, J. J. Cao, R.-J. Huang, W.D. Chen, Y. M. Han, et al., "Long-term trends in visibility and impacts of aerosol composition on visibility impairment in Baoji China", Atmos. Res., vol. 149, pp. 88-95, Nov. 2014.

[18] Y. Qi, Q. Li, H. Karimian and D. Liu, "A hybrid model for spatiotemporal forecasting of PM2.5 based on graph convolutional neural network and long short- term memory", Sci. Total Environ., vol. 664, pp. 1-10, May 2019.

[19] C. Wen, S. Liu, X. Yao, L. Peng, X. Li, Y. Hu, et al., "A novel spatiotemporal convolutional long short- term neural network for air pollution prediction", Sci. Total Environ., vol. 654, pp. 1091-1099, Mar. 2019.

[20] J. Schmidhuber, "Deep learning in neural networks: An overview", Neural Netw., vol. 61, pp. 85-117, Jan. 2015.

[21] Z. Ma and G. Mei, "Deep learning for geological hazards analysis: Data models applications and opportunities", Earth-Sci. Rev., vol. 223, Dec. 2021.

[22] S. Du, T. Li, Y. Yang and S.-J. Horng, "Deep air quality forecasting using hybrid deep learning framework", IEEE Trans. Knowl. Data Eng., vol. 33, no. 6, pp. 2412-2424, Jun. 2021.

[23] Y. Zhou, F.-J. Chang, L.-C. Chang, I.-F. Kao and Y.-S. Wang, "Explore a deep learning multi-output neural network for regional multi-step-ahead air quality forecasts", J. Clean Prod., vol. 209, pp. 134- 145, Feb. 2019.

[24] Y. Jiao, Z. Wang and Y. Zhang, "Prediction of air quality index based on LSTM", Proc. IEEE 8th Joint Int. Inf. Technol. Artif. Intell. Conf. (ITAIC), pp. 17-20, May 2019.

[25] R. J. Kuo, B. Prasetyo and B. S. Wibowo, "Deep learning-based approach for air quality forecasting by using recurrent neural network with Gaussian process in Taiwan", Proc. IEEE 6th Int. Conf. Ind. Eng. Appl. (ICIEA), pp. 471-474, Apr. 2019.

[26] Y.-S. Chang, H.-T. Chiao, S. Abimannan, Y.-P. Huang, Y.-T. Tsai and K.-M. Lin, "An LSTM-based aggregated model for air pollution forecasting", Atmos. Pollut. Res., vol. 11, no. 8, pp. 1451-1463,Aug. 2020.

[27] https://github.com/openaq

[28] Vishnu, T. R., Kumar, K. S., Ahammad, S. K. H., Kumar, G. N. S., Umakanth, N., Rao, M. C., & Krishna, S. (2023). Atmospheric Science Variations of pre-monsoon season related atmospheric parameters over Kakinada region. Journal of the National Science Foundation of Sri Lanka, 555.

[29] Kailasam, S., Achanta, S. D. M., Rama Koteswara Rao, P., Vatambeti, R., & Kayam, S. (2022). An IoT-based agriculture maintenance using pervasive computing with machine learning technique. International Journal of Intelligent Computing and Cybernetics, 15(2), 184-197.

[30] Ahmed, A. A. M., Jui, S. J. J., Sharma, E., Ahmed, M. H., Raj, N., & Bose, A. (2024). An advanced deep learning predictive model for air quality index forecasting with remote satellite-derived hydro-climatological variables. Science of The Total Environment, 906, 167234.

[31] Mohammadshirazi, A., Nadafian, A., Monsefi, A. K., Rafiei, M. H., & Ramnath, R. (2023). Novel Physics-Based Machine-Learning Models for Indoor Air Quality Approximations. arXiv preprint arXiv:2308.01438.

[32] Ayoub, A., Wainwright, H. M., & Sansavini, G. (2024). Machine learning-enabled weather forecasting for real-time radioactive transport and contamination prediction. Progress in Nuclear Energy, 173, 105255.

[33] Gong, Y., Zhang, Y., Wang, F., & Lee, C. H. (2024). Deep learning for weather forecasting: A cnn-lstm hybrid model for predicting historical temperature data. arXiv preprint arXiv:2410.14963.

[34] Ben Bouallègue, Z., Clare, M. C., Magnusson, L., Gascón, E., Maier-Gerber, M., Janoušek, M., ... & Pappenberger, F. (2024). The rise of data-driven weather forecasting: A first statistical assessment of machine learning–based weather forecasts in an operational-like context. Bulletin of the American Meteorological Society, 105(6), E864-E883.

[35] Akilan, T., & Baalamurugan, K. M. (2024). Automated weather forecasting and field monitoring using GRU-CNN model along with IoT to support precision agriculture. Expert systems with applications, 249, 123468.

[36] Sun, Y., Deng, K., Ren, K., Liu, J., Deng, C., & Jin, Y. (2024). Deep learning in statistical downscaling for deriving high spatial resolution gridded meteorological data: A systematic review. ISPRS Journal of Photogrammetry and Remote Sensing, 208, 14-38.