International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

# Result Analysis of Air Pollution Forecasting using Deep Learning based LSTM-GRU Model

[1]Shagufta Perween, [2]Dr. Sushil Kumar
M. Tech. Scholar, Department of Computer Science and Engineering, LNCT, Bhopal[1]
Associate Professor, Department of Computer Science and Engineering, LNCT, Bhopal[2]

**Abstract**
Air pollution has become a critical environmental and public health concern due to rapid urbanization, industrialization, increasing vehicular emissions, and changing climatic conditions. Accurate forecasting of air quality parameters is essential for effective environmental management, public health protection, and timely decision-making. This study presents a deep learning-based approach for air pollution forecasting using Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models. The proposed models are designed to capture complex temporal dependencies and nonlinear patterns in historical air quality and meteorological data. The performance of the LSTM and GRU models is evaluated using standard statistical metrics, including Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and coefficient of determination ($R^2$). The comparative result analysis demonstrates the forecasting capability of both deep learning models and identifies the model that provides more accurate and reliable predictions. The experimental findings indicate that recurrent neural network-based models can effectively learn temporal variations in air pollution levels and provide improved forecasting performance. The proposed framework can support early warning systems, air quality monitoring, environmental planning, and public health management by enabling reliable prediction of future pollution trends.

**Keywords:** Long Term Short Memory (LSTM), Deep Learning (DL), Gate Recurrent Unit (GRU)

## 1. INTRODUCTION

Air pollution is one of the major environmental challenges affecting human health, ecological balance, and overall quality of life across the world. The rapid growth of urban populations, industrial activities, transportation, construction, and energy consumption has significantly increased the concentration of harmful pollutants in the atmosphere. Pollutants such as particulate matter (PM2.5 and PM10), nitrogen dioxide ($NO_2$), sulfur dioxide ($SO_2$), carbon monoxide (CO), and ozone ($O_3$) can cause serious respiratory and cardiovascular problems when their concentrations exceed permissible limits. Therefore, accurate and timely forecasting of air pollution levels is essential for environmental monitoring, public health protection, and effective pollution control strategies. Traditional air pollution forecasting methods generally rely on statistical techniques and conventional machine learning algorithms, which may have limitations in capturing complex nonlinear relationships and long-term temporal dependencies within air quality data.

With the rapid development of Artificial Intelligence (AI) and Deep Learning (DL), data-driven approaches have become increasingly effective for forecasting environmental parameters. Deep learning models can automatically learn complex patterns from large-scale historical datasets containing air quality and meteorological variables. Among various deep learning architectures, Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) networks are particularly suitable for time-series forecasting. LSTM networks use memory cells and gating mechanisms to retain relevant information over long time intervals, making them effective in modeling long-term dependencies in air pollution data. On the other hand, GRU is a simplified recurrent architecture that uses fewer gates and parameters, resulting in reduced computational complexity while maintaining the ability to capture temporal relationships. Both models can analyze historical pollution patterns and learn the dynamic variations of air quality over time.

The performance of air pollution forecasting models depends on their ability to accurately predict future pollutant concentrations under changing environmental conditions. Factors such as temperature, humidity, wind speed, atmospheric pressure, and historical pollutant concentrations can significantly influence air quality. Therefore, incorporating these variables into deep learning-based forecasting models can improve prediction reliability. In this study, LSTM and GRU models are implemented for air pollution forecasting and their performance is comparatively analyzed using standard evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and coefficient of determination ($R^2$). The analysis focuses on identifying the model that provides superior forecasting accuracy and generalization capability. The results of this study can contribute to the development of intelligent air quality monitoring and early warning systems, supporting environmental authorities and the public in taking timely actions to reduce exposure to harmful air pollutants.

### 1.1 Air Pollution and Its Effects

Air pollution refers to the presence of harmful gases, particulate matter, and other toxic substances in the atmosphere at concentrations that can negatively affect humans, animals, plants, and the environment. Major air pollutants include particulate matter such as PM2.5 and PM10, carbon monoxide (CO), nitrogen dioxide ($NO_2$), sulfur dioxide ($SO_2$), ozone ($O_3$), and volatile organic compounds (VOCs). These pollutants are released from various sources, including vehicle emissions, industrial processes, fossil fuel combustion, agricultural activities, construction, and natural events such as forest fires and dust storms. The increasing levels of air pollution are primarily associated with rapid urbanization, industrial development, population growth, and increased energy consumption.

Air pollution has significant short-term and long-term effects on human health. Continuous exposure to polluted air can cause respiratory diseases, asthma, bronchitis, lung infections, and cardiovascular disorders. Fine particulate matter, particularly PM2.5, is capable of penetrating deep into the respiratory system and entering the bloodstream, increasing the risk of serious health complications. Vulnerable groups such as children, elderly people, and individuals with existing respiratory or cardiovascular conditions are particularly susceptible to the harmful effects of air pollution. In addition to human health, air pollution contributes to environmental degradation, including reduced visibility, acid rain, damage to vegetation, soil and water contamination, and climate change. Therefore, continuous air quality monitoring and accurate forecasting of pollutant concentrations are essential for developing effective pollution control strategies and protecting public health. The integration of advanced artificial intelligence and deep learning techniques with air quality monitoring systems can further improve the prediction of pollution trends and support timely preventive measures.

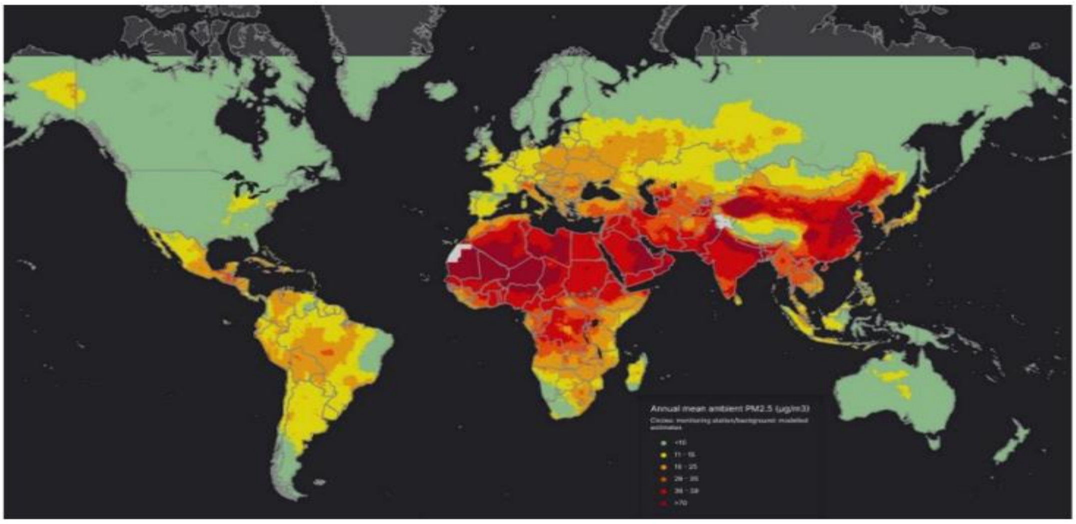


Figure 1: Air Pollution Effect

## 2. PROPSOED METHODOLOGY

The proposed methodology for Air Pollution Forecasting using Deep Learning-based LSTM-GRU Models consists of a systematic sequence of data collection, preprocessing, exploratory data analysis, feature selection, dataset splitting, model development, model training, prediction, and performance evaluation. The overall objective is to develop an accurate forecasting framework capable of learning complex temporal patterns in historical air pollution and meteorological data. The proposed methodology uses two recurrent deep learning architectures, namely Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) and compares their forecasting performance using standard evaluation metrics.

### Data Collection

The first stage involves collecting historical air quality and meteorological data from a reliable dataset. The dataset may contain pollutant parameters such as PM2.5, PM10, $NO_2$, $SO_2$, CO, and $O_3$, along with environmental variables such as temperature, humidity, wind speed, atmospheric pressure, and rainfall. Each observation is associated with a specific timestamp, allowing the data to be treated as a time-series dataset. Historical observations are used to identify temporal relationships and patterns that can help predict future air pollution levels.

### Data Preprocessing

Raw air pollution datasets may contain missing values, duplicate records, inconsistent formats, and abnormal observations. Therefore, preprocessing is performed before training the deep learning models. Missing values are handled using suitable imputation techniques, while duplicate and invalid records are removed. The date and time attributes are converted into an appropriate datetime format and arranged chronologically. Numerical features are normalized or standardized to ensure that variables with different numerical ranges do not negatively affect model training. Data preprocessing improves data quality and enables faster and more stable convergence of the LSTM and GRU networks.

**Exploratory Data Analysis**

Exploratory Data Analysis (EDA) is conducted to understand the distribution and temporal behavior of air pollution variables. Statistical analysis and visualization techniques are used to identify pollution trends, seasonal variations, correlations between pollutants, and relationships between air quality and meteorological parameters. Time-series plots can be used to observe changes in pollutant concentrations over time, while correlation analysis helps identify the most relevant input features for forecasting. This stage provides valuable insights into the characteristics of the dataset before model development.

**Feature Selection and Time-Series Transformation**

Relevant air quality and meteorological variables are selected as input features for the forecasting models. The historical observations are transformed into sequential data using a sliding-window approach. In this process, a fixed number of previous time steps are used to predict the pollutant concentration at a future time step. This transformation is particularly important because LSTM and GRU models are designed to process sequential data and learn dependencies between past and future observations.

**Dataset Splitting**

The processed dataset is divided into training and testing subsets. The training dataset is used to learn the underlying temporal patterns and relationships, while the testing dataset is used to evaluate the forecasting performance of the trained models on previously unseen data. For time-series forecasting, the data are divided chronologically rather than randomly to preserve the natural temporal order and prevent future information from influencing the training process.

**LSTM-Based Air Pollution Forecasting**

The first proposed forecasting model uses a Long Short-Term Memory (LSTM) neural network. LSTM is a specialized recurrent neural network architecture designed to overcome the vanishing-gradient problem commonly encountered in traditional recurrent neural networks. It uses input, forget, and output gates to control the flow of information through memory cells. In the proposed framework, historical air quality and meteorological sequences are provided to the LSTM network. The model learns both short-term and long-term temporal dependencies and generates the predicted air pollution value through fully connected output layers.

**GRU-Based Air Pollution Forecasting**

The second forecasting model is based on a **Gated Recurrent Unit (GRU)** network. GRU is

ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

another recurrent deep learning architecture that uses update and reset gates to control information flow. Compared with LSTM, GRU has a simpler structure and fewer trainable parameters, which can reduce computational requirements while maintaining strong performance for sequential prediction tasks. The GRU model is trained using the same preprocessed dataset and forecasting configuration as the LSTM model, allowing a fair comparison between the two architectures.

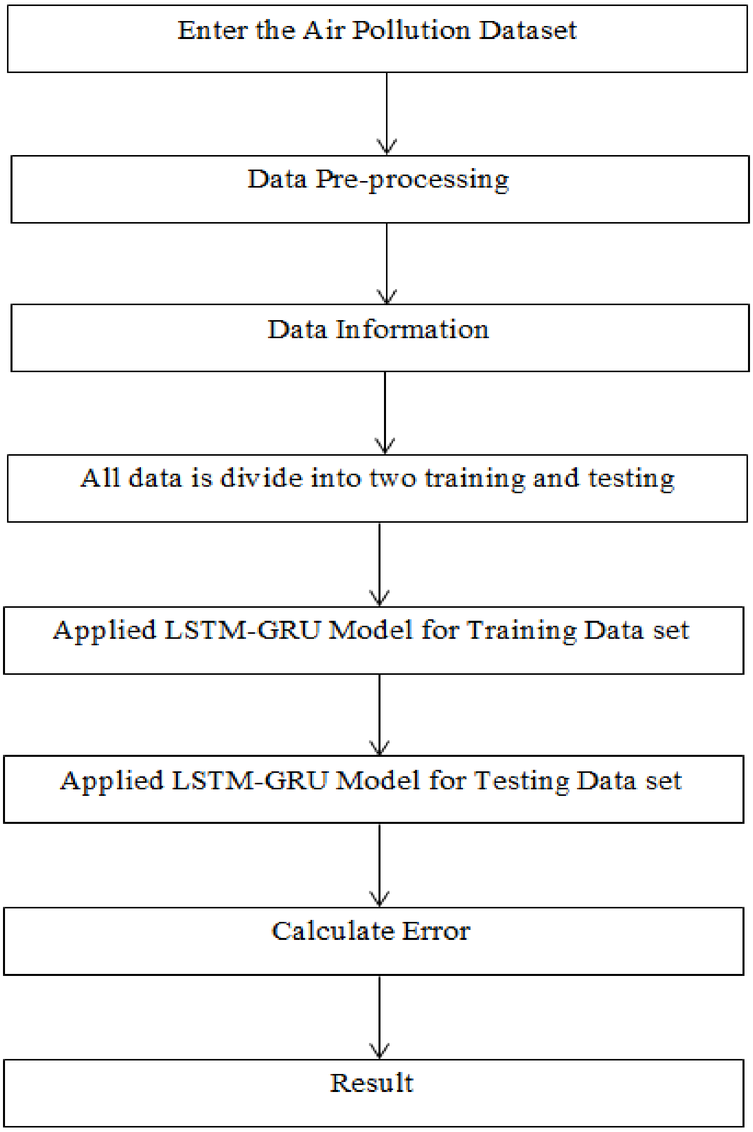



Figure 2: Diagram of Air pollution Prediction

**Model Training and Optimization**

Both LSTM and GRU models are trained using the prepared training sequences. During training, the models iteratively update their weights by minimizing a predefined loss function between the actual and predicted pollution values. The optimizer adjusts the model parameters to reduce prediction errors. Hyperparameters such as the number of hidden layers, neurons, learning rate, batch size, sequence length, and number of epochs can be optimized to

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

improve forecasting accuracy. Techniques such as early stopping can also be applied to prevent overfitting and improve the generalization capability of the models.

**Prediction and Result Analysis**

After training, the LSTM and GRU models are evaluated using the testing dataset. The trained models generate predicted air pollution values, which are compared with the actual observations. Performance metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and coefficient of determination ($R^2$) are calculated to measure forecasting accuracy. Lower MAE, MSE, and RMSE values indicate smaller prediction errors, while a higher $R^2$ value indicates better agreement between predicted and actual values.

**Comparative Performance Evaluation**

Finally, the forecasting performance of the LSTM and GRU models is compared based on their evaluation metrics. Actual and predicted pollution values are visualized using time-series plots to examine how closely each model follows the real pollution trends. The model with lower prediction errors and higher $R^2$ is considered the better-performing model. The comparative analysis helps determine whether the long-term memory capability of LSTM or the simpler gated architecture of GRU is more effective for accurate air pollution forecasting.

## 3. SIMULATION RESULT

Air quality has become a strategic nationwide problem, as in recent years, obscure weather has appeared in several parts of India. Several studies show that particulate matter not greater than 2.5 lm (PM2.5) comprises huge amounts of toxic and detrimental substances [15]. PM2.5 is the predominant air pollutant and has an adverse impact on people's health, i.e., it has a severe impact on the respiratory and cardiovascular systems. Furthermore, PM2.5 has a worse effect on the weather and climate records of a country or a region. Fig. 6 present dataset of AP with 9 attribute.

| | From Date | To Date | PM2.5 | PM10 | NO | NO2 | NOx | NH3 | SO2 | CO | Ozone |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 01-01-2018 00:00 | 01-01-2018 01:00 | 79.04 | 204.97 | 9.43 | 18.69 | 28.12 | 7.62 | 3.75 | 0.54 | 74.45 |
| 1 | 01-01-2018 01:00 | 01-01-2018 02:00 | 74.31 | 169.77 | 9.37 | 17.94 | 27.31 | 7.22 | 2.41 | 0.50 | 74.59 |
| 2 | 01-01-2018 02:00 | 01-01-2018 03:00 | 54.70 | 146.10 | 9.00 | 17.84 | 26.84 | 7.47 | 1.03 | 0.46 | 71.02 |
| 3 | 01-01-2018 03:00 | 01-01-2018 04:00 | 54.82 | 143.33 | 9.30 | 20.66 | 29.96 | 7.93 | 2.19 | 0.78 | 55.91 |
| 4 | 01-01-2018 04:00 | 01-01-2018 05:00 | 63.78 | 148.88 | 14.45 | 34.44 | 48.89 | 8.52 | 1.14 | 0.90 | 28.12 |

Figure 3: Dataset

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |


Figure 4: Exploratory Data Analysis

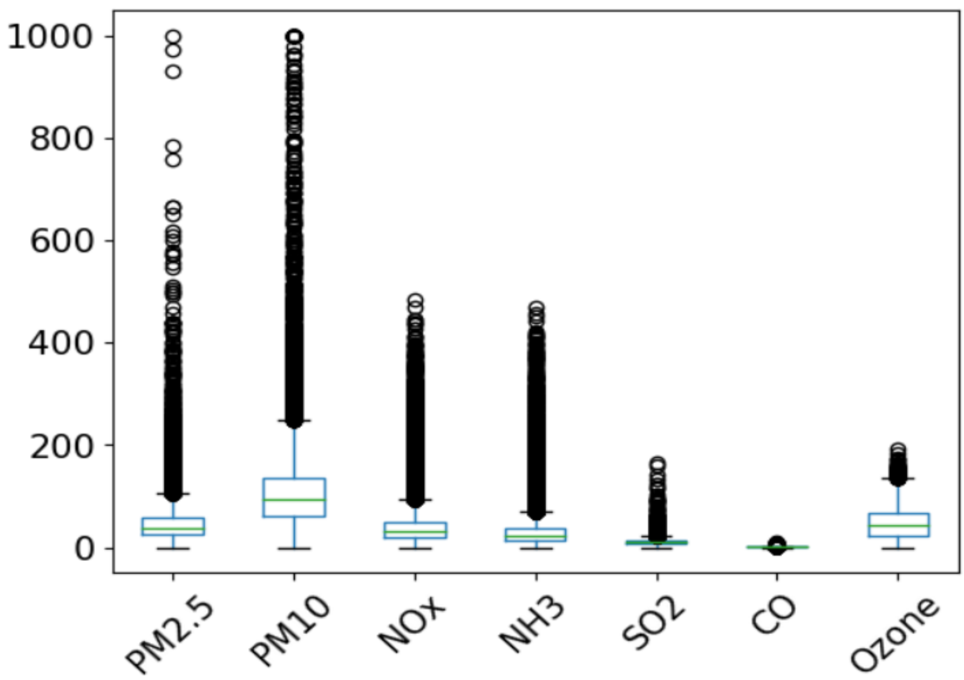
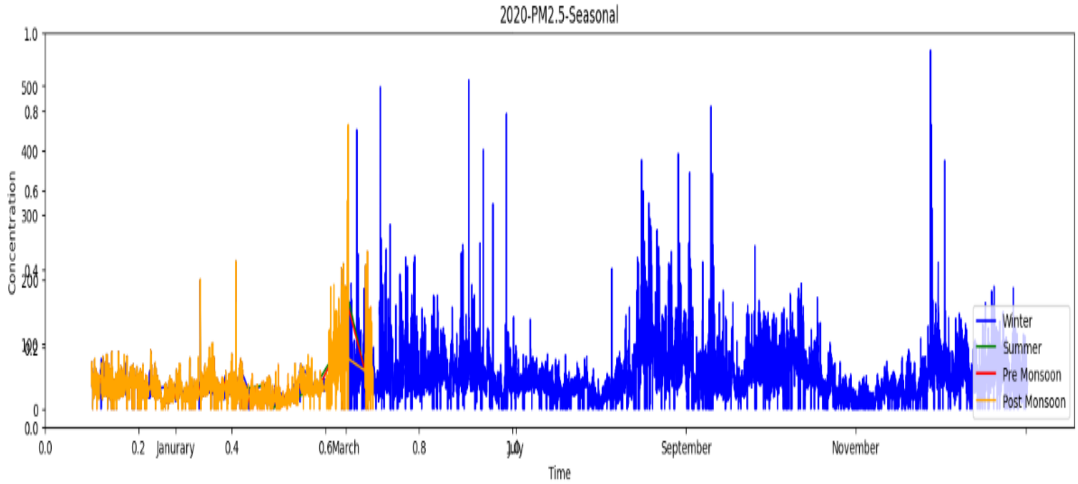



Figure 5: Different Air Quality Index

In atmospheric chemistry and air quality research, the relationship between nitrogen oxides (NOx) and ammonia ($NH_3$) is crucial is present. Whereas NOx is mostly released from combustion processes in automobiles, factories, and power plants, $NH_3$ is mostly released from agricultural activities including fertilizer use and livestock dung.

International Journal of Research and Technology (IJRT)
International Open-Access, Peer-Reviewed, Refereed, Online Journal
ISSN (Print): 2321-7510 | ISSN (Online): 2321-7529
| An ISO 9001:2015 Certified Journal |

In actuality, PM2.5 is a subset of PM10 since PM10 denotes particles with a diameter of 10 micrometers or less, whereas PM2.5 denotes finer particles with a diameter of 2.5 micrometers or fewer. As a result, there is an inherent relationship between the two. Figure 6 represents month and seasonal wise PM2.5.



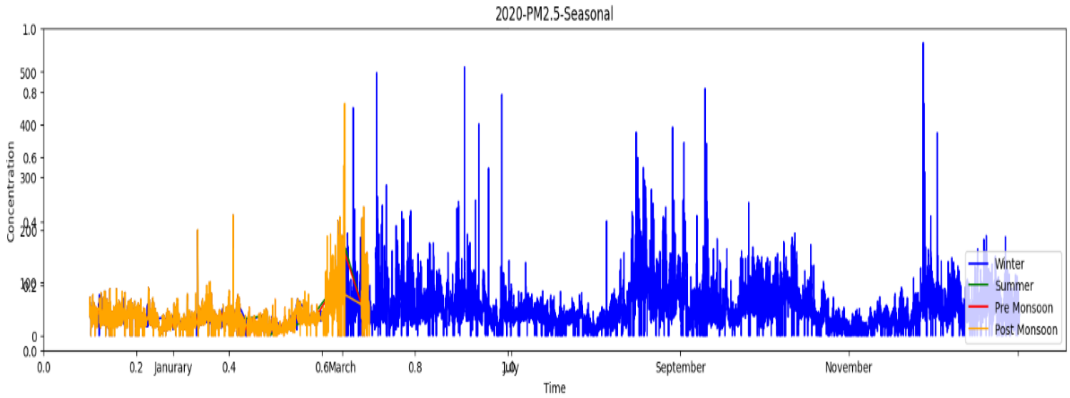




Figure 6: Concentration of PM2.5 for Seasonal Wise

Table 1 displays the results of implemented method in terms of MSE, RMSE, MAE and R Square. SVM method gives an MSE of 997.74, an RMSE of 31.58, an MAE of 14.35 and an R Square of 0.33. LSTM method gives an MSE of 411.70, an RMSE of 20.29, an MAE of 11.64 and an R Square of 0.67. LSTM with GRU method gives an MSE of 370.62, an RMSE of 7.86, an MAE of 11.45 and an R Square of 0.70. Fig. 7 to 8 shows the graphical representation of the different parameter.

Table 1: Comparison Result

| Models | MSE | RMSE | MAE | R Square |
|---|---|---|---|---|
| Previous SVM | 997.74 | 31.58 | 14.35 | 0.33 |
| Previous LSTM | 411.70 | 20.29 | 11.64 | 0.67 |
| Proposed LSTM with GRU | 370.62 | 19.25 | 11.45 | 0.70 |

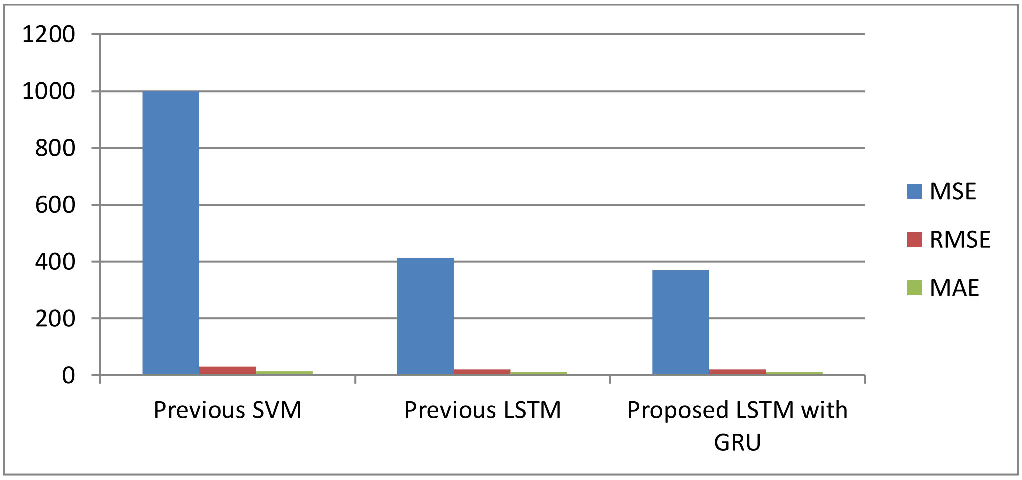



Figure 7: Graphical Error

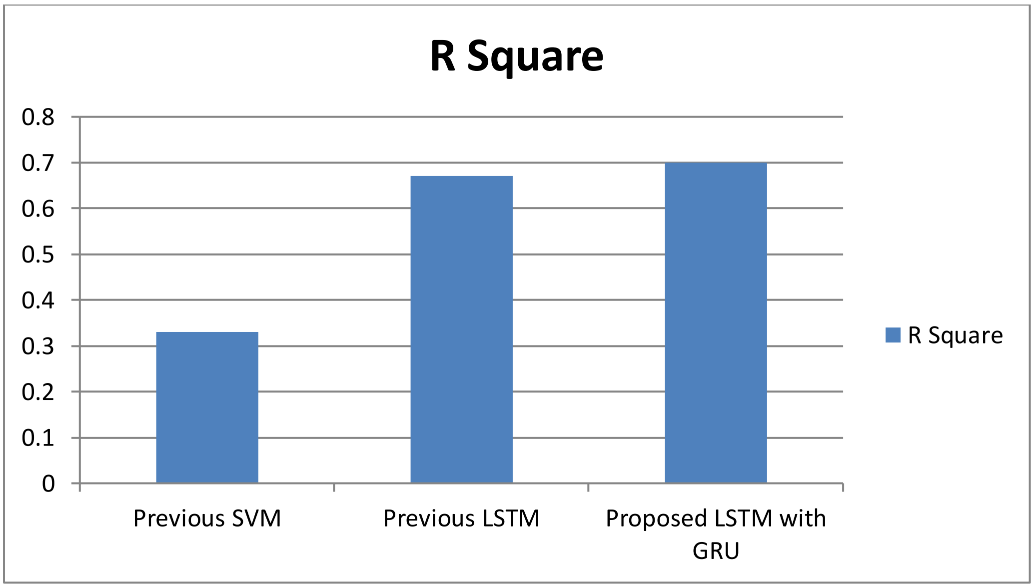



Figure 8: Graphical Represent of R Square

## 4. CONCLUSION

This study presented a deep learning-based approach for air pollution forecasting using Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models. The proposed methodology involved data preprocessing, exploratory data analysis, feature preparation, time-series transformation, model training, prediction, and performance evaluation. The LSTM and GRU architectures were selected because of their ability to learn temporal dependencies and complex patterns from sequential air quality and meteorological data. The forecasting results were evaluated using standard performance metrics, including MAE, MSE, RMSE, and $R^2$, to determine the accuracy and reliability of the proposed models.

The comparative analysis demonstrates that deep learning-based recurrent models can effectively capture the dynamic and nonlinear behavior of air pollution levels. Both LSTM and GRU models provide a promising solution for predicting future pollution trends, while the better-performing model can be identified based on lower prediction errors and higher $R^2$

values. Accurate air pollution forecasting can assist environmental authorities in monitoring pollution levels, issuing early warnings, implementing pollution control measures, and protecting public health. In the future, the proposed framework can be enhanced by incorporating real-time IoT-based air quality data, advanced attention mechanisms, hybrid deep learning architectures, and optimization algorithms. Furthermore, integrating additional meteorological and geographical factors may improve forecasting accuracy and enable the development of intelligent, real-time air quality prediction and early warning systems.

## REFERENCES

[1] Ghufran Isam Drewil and Riyadh Jabbar Al-Bahadili, "Air pollution prediction using LSTM deep learning and metaheuristics algorithms", Elsevier Science Direct, Vol. 24, pp. 01-07, 2022.

[2] Jannah Mohammad and Mohammod Abul Kashem, "Air Pollution Comparison RFM Model Using Machine Learning Approach", IEEE 7th International conference for Convergence in Technology (I2CT), IEEE 2022.

[3] Harshit Srivastava, Goutam Kumar Sahoo, Santos Kumar Das and Poonam Singh, "Performance Analysis of Machine Learning Models for Air Pollution Prediction", International Conference on Smart Generation Computing, Communication and Networking (SMART GENCON), IEEE 2022.

[4] Chakradhar Reddy K, Nagarjuna Reddy K, Brahmaji Prasad K and P.Selvi Rajendran, "The Prediction of Quality of the Air Using Supervised Learning", 6th International Conference on Communication and Electronics Systems (ICCES), IEEE 2021.

[5] Jovan Kalajdjieski, Eftim Zdravevski , Roberto Corizzo, Petre Lameski, Slobodan Kalajdziski, Ivan Miguel Pires, Nuno M. Garcia and Vladimir Trajkovik, "Air Pollution Prediction with Multi-Modal Data and Deep Neural Networks", Remote Sensing, 2020.

[6] Zhang, Y.; Guo, L.; Wang, Z.; Yu, Y.; Liu, X.; Xu, F. Intelligent Ship Detection in Remote Sensing Images Based on Multi-Layer Convolutional Feature Fusion. Remote Sens. 2020, 12, 3316.

[7] Bai, L., Wang, J., Ma, X., and Lu, H., 'Air pollution forecasts: An overview', International journal of environmental research and public health 15(4), 780, 2018.

[8] Balakrishnan, K., Dey, S., Gupta, T., Dhaliwal, R. S., Brauer, M., Cohen, A. J. and Sabde, Y., 'The impact of air pollution on deaths, disease burden, and life expectancy across the states of India: the Global Burden of Disease Study 2017'. The Lancet Planetary Health 3(1), e26-e39, 2019.

[9] Cai, S., Wang, Y., Zhao, B., Wang, S., Chang, X., and Hao, J., 'The impact of the "air pollution prevention and control action plan" on PM2. 5 concentrations in JingJin-Ji region during 2012–2020', Science of the Total Environment 580, 197-209, 2017.

[10] Cascio, W. E., and Long, T. C., 'Ambient Air Quality and Cardiovascular Health Translation of Environmental Research for Public Health and Clinical Care', North Carolina medical journal 79(5), 306-312, 2018.

[11] Huang, M., Zhang, T., Wang, J., and Zhu, L., "A new air quality forecasting model using data mining and artificial neural network", In 6th IEEE International Conference on Software Engineering and Service Science (ICSESS)', pp. 259-262, 2015.

[12] Kang, G. K., Gao, J. Z., Chiao, S., Lu, S., and Xie, G., 'Air quality prediction: Big data and machine learning approaches', International Journal of Environmental Science and Development, 9(1), 8-16, 2018.

[13] Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., and Wu, A. Y., 'An efficient k-means clustering algorithm: Analysis and implementation', IEEE Transactions on Pattern Analysis and Machine Intelligence, (7), 881-892, 2002.

[14] Kelly, F. J., & Fussell, J. C., "Air pollution and public health: emerging hazards and improved understanding of risk. Environmental geochemistry and health", 37(4), 631-649, 2015.

[15] Kemp, A. C., Horton, B. P., Donnelly, J. P., Mann, M. E., Vermeer, M., and Rahmstorf, S., 'Climate related sea-level variations over the past two millennia', Proceedings of the National Academy of Sciences, 108(27), 11017-11022, 2011.