IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

An Empirical Comparison of Tree-based Ensemble Methods and LSTM for Air Quality Index Classification

[1]Vallabh Kulkarni, [2]Dr.Srinath S, [3]Ramya S

[1]Pg Student, [2]Asistent Professor, [3]Phd Student
[1]Computer Science and Engineering,
[1]JSS Science and Technology University, Mysore, India

Abstract: Air pollution is a growing concern that directly affects public health, climate stability, and overall well-being. Exact Forecasting of air quality is Needed for enabling early interventions and informed policy decisions. In presently this work, a machine learning–driven framework is developed to forecast air quality based on a dataset combining meteorological variables and pollutant concentrations. The data undergoes thorough preprocessing, including handling missing entries, scaling features, and optimizing relevant parameters to ensure reliable model outputs. Several algorithms—Random Forest, XGBoost, Gradient Boosting, and LSTM—are trained and tested, with performance assessed using 2-precision, 3-recall, 4- F1-score, 1- accuracy, and AUC metrics. Among these, Gradient Boosting demonstrates superior predictive capability across most evaluation criteria.

***IndexTerms* - Air quality prediction, Machine learning, Gradient Boosting, Pollution forecasting.**

INTRODUCTION

Air pollution is increasingly recognized as a major global challenge, posing serious risks to human health, natural ecosystems, and economic productivity, making accurate forecasting of air quality an essential component of environmental management strategies (WHO, 2022). Traditional statistical models which as multiple linear regression and ARIMA have been widely used for prediction; however, their performance often declines in the presence of non-linear patterns, irregular temporal structures, and multivariate dependencies inherent in environmental datasets (Goyal et al., 2006). in machine learning and deep learning have Seen many of these challenges by activating the capture of complex relationships between pollutant concentrations and meteorological variables. Ensemble-based algorithms like Random Forest, Gradient Boosting, and XGBoost have demonstrated strong predictive capabilities in multi-factor air quality modeling (Zhang et al., 2021), while recurrent neural network architectures such as LSTM and GRU have shown effect in handling linear and long-term dependencies in pollutant time series (Lin et al., 2021).

Hybrid frameworks that integrate spatial and temporal learning—such as CNN-LSTM models—have further improved forecasting accuracy in city-scale studies by extracting local feature patterns and combining them with temporal trends (Sarkar et al., 2022). In addition, multi-model ensemble strategies, including the use of GRU-based predictors combined through regression (MLEGRU), have produced lower RMSE and MAE values compared to single-model approaches in diverse monitoring environments (Hong et al., 2021). Despite these advancements, practical deployment remains challenging due to sparse monitoring networks, heterogeneous meteorological influences, and unbalanced pollutant datasets, highlighting the need for comprehensive preprocessing techniques—such as missing data imputation, feature scaling, and permutation-based feature selection—to ensure robust and generalizable performance in real-world applications (Bucharest case study; Zhang et al., 2021).

LITARTURE REVIEW

The integration of machine learning and deep learning into air quality forecasting has significantly improved predictive capabilities, often surpassing traditional statistical and time-series techniques. Goyal et al. (2006) introduced early approaches such as ARIMA and multiple linear regression for pollutant level estimation but noted their inability to handle nonlinear relationships and temporal complexities. Zhang et al in 2012 recognized the potential of machine learning in environmental prediction, while Yi et al. (2018) successfully applied advanced models for urban AQI estimation. which was later adapted by Sarkar et al. (2022) into a hybrid LSTM–GRU framework for Delhi's air pollution, achieving improved results across RMSE, MAE, and R². Cho et al. in 2014 Showcased the Gated Recurrent Unit (GRU), which has been widely adopted due to its computational efficiency in sequential data tasks. Lin et al. (2021) expanded this concept with MLEGRU, an ensemble GRU approach designed to improve forecast stability across Taiwan's monitoring stations.

Hybrid deep learning frameworks have also gained traction in recent studies. Zhang et al. (2021) introduced Deep-AIR, which combines CNN and LSTM for spatial-temporal AQI forecasting, delivering higher accuracy in city-scale scenarios. Hong et al. (2021) incorporated external parameters such as shipping activity into RNN-based models, showing that contextual environmental data can significantly enhance predictions. Li et al. (2021) stressed the role of preprocessing methods—such as feature scaling and imputation—in ensuring model generalization in low-data environments. Comparative evaluations have reinforced the advantage of deep models over conventional approaches. Zheng et al. (2015) assessed large-scale air quality forecasting systems and noted the persistent challenges of sparse monitoring networks. Lin et al. (2019) applied feature selection and segmentation strategies to enhance pollutant-specific predictions.

Ensemble methods, as outlined by Dietterich (2000), remain crucial for robust AQI estimation under diverse meteorological conditions. Evaluation metrics like RMSE, MAE, and R² are widely used benchmarks, as seen in studies comparing LSTM, GRU, and hybrid networks (Padilla et al., 2020). The growing use of CNN–RNN hybrids, transformer-based models, and portable forecasting systems points toward a future of more precise, efficient, and accessible air quality prediction, supporting proactive environmental and public health measures.

METHODOLOGY

The machine learning and deep learning are adopted into air quality forecasting has significantly improved predictive capabilities, often surpassing traditional statistical and time-series techniques. Goyal et al. (2006) introduced early approaches such as ARIMA and multiple linear regression for pollutant level estimation but noted their inability to handle nonlinear relationships and temporal complexities. Zhang et al in 2012 recognized the potential of machine learning in environmental prediction, while Yi et al. (2018) successfully applied advanced models for urban AQI estimation. which was later adapted by Sarkar et al. (2022) into a hybrid LSTM–GRU framework for Delhi's air pollution, achieving improved results across RMSE, MAE, and R². Cho et al. in 2014 showcased the Gated Recurrent Unit (GRU), which has been widely adopted due to its computational efficiency in sequential data tasks. Lin et al. (2021) expanded this concept with MLEGRU, an ensemble GRU approach designed to improve forecast stability across Taiwan's monitoring stations.

Hybrid deep learning frameworks have also gained traction in recent studies. Zhang et al. (2021) introduced Deep-AIR, which combines CNN and LSTM for spatial-temporal AQI forecasting, delivering higher accuracy in city-scale scenarios. Hong et al. (2021) incorporated external parameters such as shipping activity into RNN-based models, showing that contextual environmental data can significantly enhance predictions. Li et al. (2021) stressed the role of preprocessing methods—such as feature scaling and imputation—in ensuring model generalization in low-data environments. Comparative evaluations have reinforced the advantage of deep models over conventional approaches. Zheng et al. (2015) assessed large-scale air quality forecasting systems and noted the persistent challenges of sparse monitoring networks. Lin et al. (2019) applied feature selection and segmentation strategies to enhance pollutant-specific predictions.

Ensemble methods, as outlined by Dietterich (2000), remain crucial for robust AQI estimation under diverse meteorological conditions. Evaluation metrics like RMSE, MAE, and R² are widely used benchmarks, as seen in studies comparing LSTM, GRU, and hybrid networks (Padilla et al., 2020). The growing use of CNN–RNN hybrids, transformer-based models, and portable forecasting systems points toward a future of more precise, efficient, and accessible air quality prediction, supporting proactive environmental and public health measures.

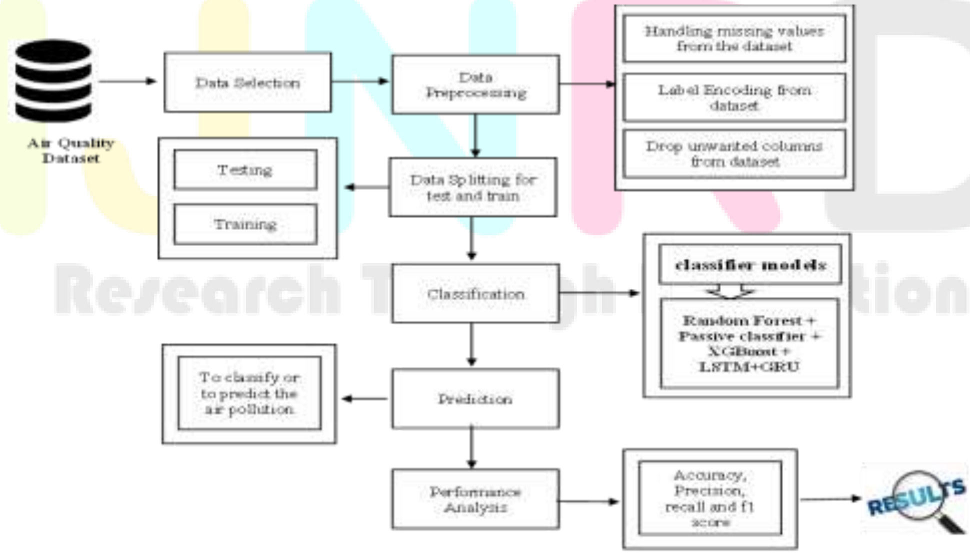
Air Quality Dataset
Data Selection
Data Preprocessing
Handling missing values from the dataset
Label Encoding from dataset
Drop unwanted columns from dataset
Testing
Training
Data Splitting for test and train
Classification
classifier models
Random Forest + Passive classifier + XGBoost + LSTM+GRU
To classify or to predict the air pollution
Prediction
Performance Analysis
Accuracy, Precision, recall and f1 score
RESULTS

Fig 1. System architecture

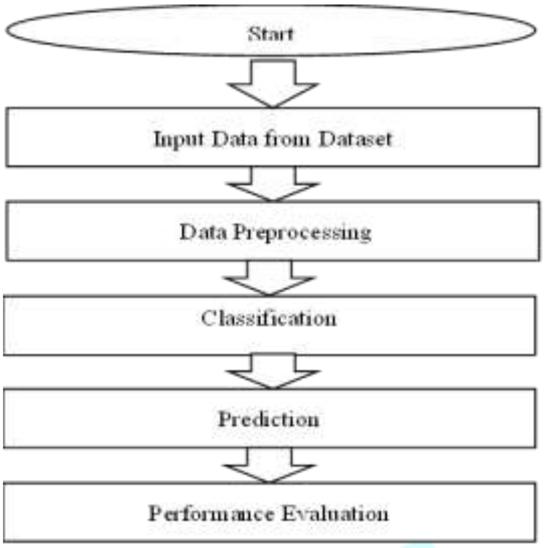

Fig 2. Workflow Diagram

This Figure 2 depicts the typical workflow of a supervised machine learning project. The process begins with raw data being fed into the system. This data is then passed through an essential preprocessing stage, where it is cleaned and organized into the proper format. Once prepared, a classification model is administered to the dataset to generate prediction values. In the final stage, the model's accuracy and overall performance are assessed through evaluation metrics.

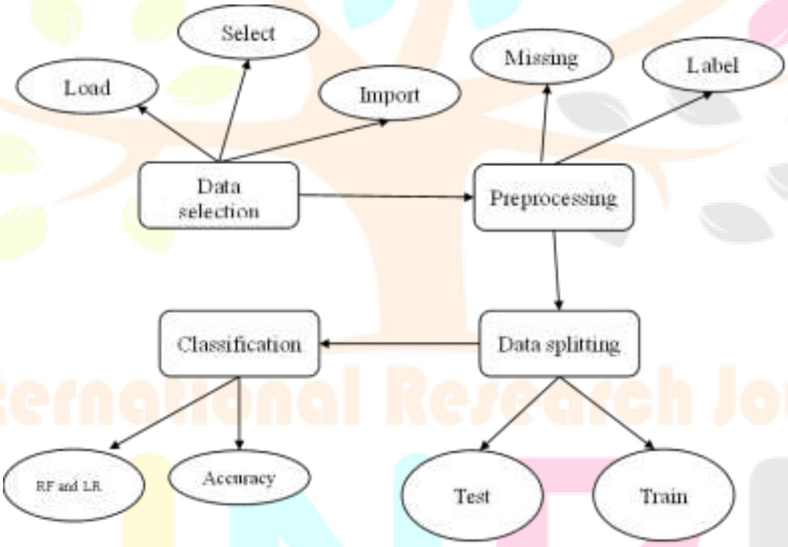

Fig 3. ER Diagram

This Figure 3 shows the main stages and components of a machine learning classification workflow. It starts with data acquisition and selection, where the dataset is loaded and imported into the system. Next comes the preprocessing phase, which involves managing missing values and applying label encoding to categorical features. The after Cleaned data is then parted into training and testing subsets for classification using algorithms such as Random Forest (RF). At the final stage, the model's predictions are Evaluvated using performance metrics like accuracy to determine its effectiveness.

Performance Metrics:
Accuracy: Our model's effectiveness is evaluated primarily through its accuracy. This metric reflects how reliably the approach identifies both correct positive and negative classifications, as well as how closely the predicted results align with the actual recorded values. Equation (1) expresses the calculation of accuracy for the proposed system.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

Precision: The total number of correctly predicted instances belonging to a particular class is obtained by counting the true positivesPrecision tells you what percentage of the positive predictions were actually correct. To figure this out, you simply do the number of true positives upon the Entire number of times the model predicted a positive result (true positives addition to false positives). This is shown in Equation (2).

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall: This also called the Exact Positive Rate, which Counts a model's completeness. It shows how good the model is at finding all the existing positive examples in the data. The formula for this is the no of exact positives Cleaved by the Entire number of actual positive cases (Exact positives plus false negatives). More the Recall Number the model is correctly capturing the relevant data without missing significant instances. Equation (3) provides the recall formula.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

RESULTS AND DISCUSSION

Based on the machine learning workflow which is outlined for air quality prediction, a thorough analysis would compare how well each model performed. This is not t just about picking one winner; this is about understanding the strengths of each approach. The models Random Forest, XGBoost, Hybrid Classifier, and the higher deep learning models like LSTM and GRU—would each has their performance measured against key Evaluvation metrics like 1-Accuracy,2-Precision,3-Recall, and the 4 -F1 Score. The final results would clearly show which model is the most effective and reliable for the specific task of predicting air pollution, helping to ensure the final system is robust and trustworthy.

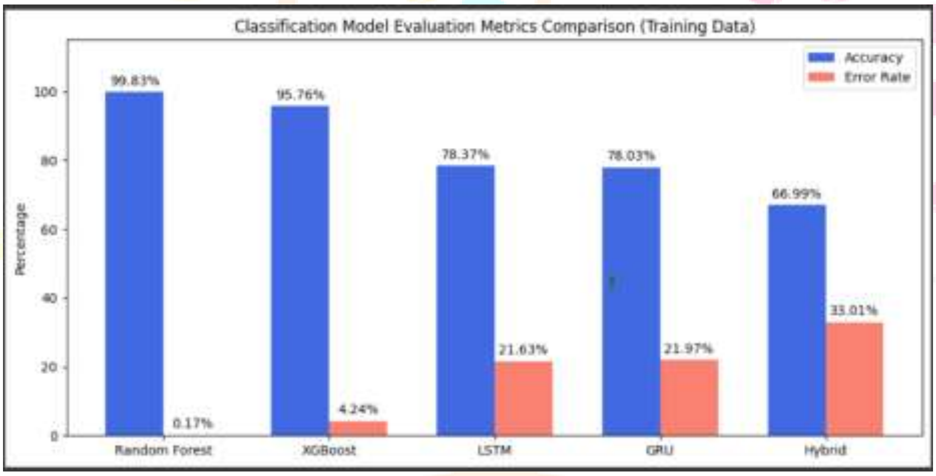

Fig 4. Model Comparison

Based on the Figure 4, this comparison shows how well each of the five models learned the patterns in the training data. The Random Forest and XGBoost models performed very well. with Random Forest Resulting a near-perfect accuracy of 99.83%. The deep learning models, LSTM and GRU, showcased more moderate and very similar results to each other. However, such high accuracy on training data can be a sign of overfitting,

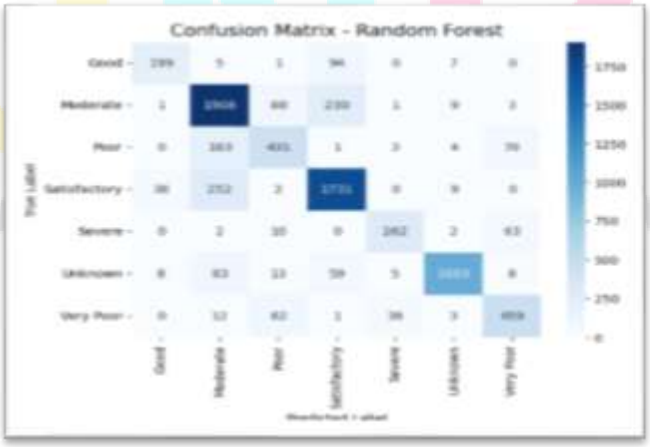

Fig 5. ConfusionMatrix - Random Forest

The Figure 5 shows This confusion matrix of the Random Forest model shows a strong overall performance,.1,908 'Moderate' and 1,731 'Satisfactory' air quality instances. It demonstrates high accuracy on the main diagonal, correctly classifying most cases for each category. The model's primary area of confusion lies between these adjacent middle-tier labels, where it misclassified .252 'Satisfactory' samples as 'Moderate' and 230 'Moderate' samples as 'Satisfactory'.

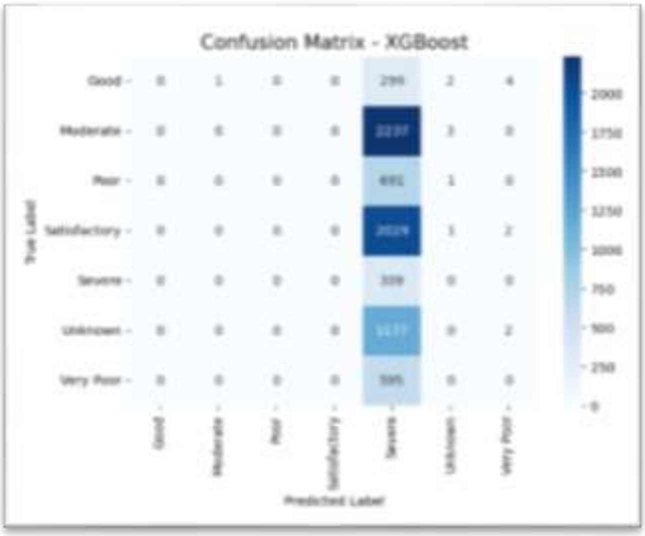

Fig 6. The ConfusionMatrix - XGboost

The Fig 6 shows this confusion matrix of the XGBoost model shows a significant tendency to predict the 'Severe' class, with very low accuracy across most other categories. It correctly classified only 339 'Severe' air quality instances, while misclassifying almost every other true category. The model's primary area of confusion lies in its overwhelming bias toward the 'Severe' label, where it misclassified 2,237 'Moderate' samples, 2,029 'Satisfactory' samples, and 1,177 'Unknown' samples as 'Severe'.

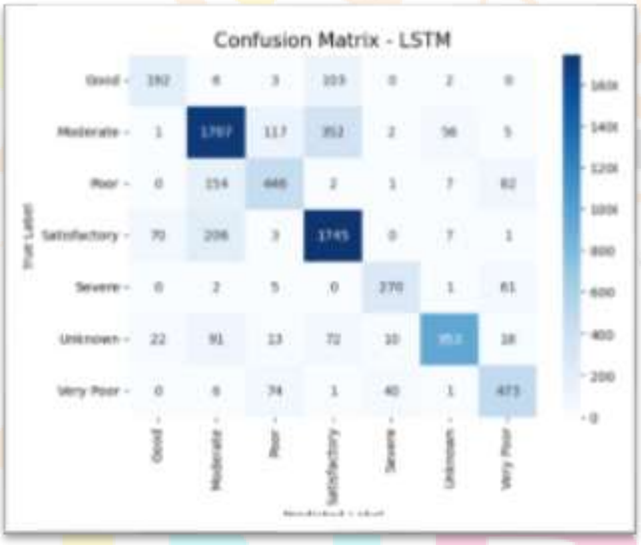

Fig 7. The Confusion Matrix - LSTM

Fig 7 shows This confusion matrix of the LSTM model shows a strong predictive performance, correctly identifying 1,745 'Satisfactory' and 1,707 'Moderate' air quality cases. Its primary challenge lies in the significant confusion between these similar mid-range categories, highlighted by the 352 'Moderate' samples it mislabeled as 'Satisfactory'. Overall, the matrix confirms the model is highly effective but could be improved by better distinguishing between these adjacent air quality levels.

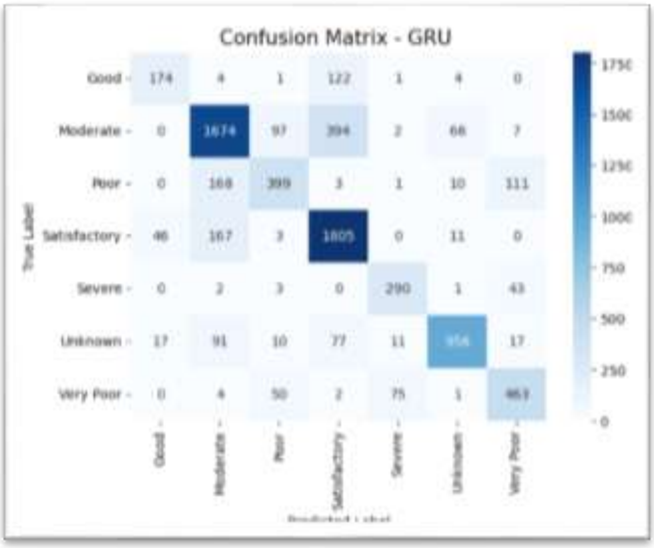

Fig 8. The Confusion matrix - GRU

The Fig 8 shows GRU model's performance, detailed in this confusion matrix, is excellent, with its most accurate predictions being for the 'Satisfactory' category, where it correctly identified 1,805 cases. Its primary weakness, similar to the LSTM, is the confusion between adjacent levels, notably misclassifying 394 'Moderate' instances as 'Satisfactory'. Overall, the GRU proves to be a highly reliable classifier.

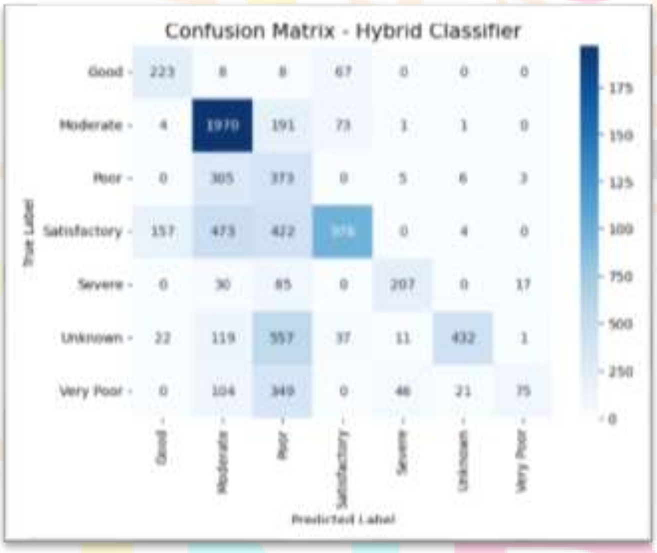

Fig 9. The Confusion Matrix - Hybrid Classifier

The Fig 9 Shows confusion matrix for the Hybrid Classifier shows its strongest performance on the 'Moderate' category, correctly identifying 1,970 cases. However, the model struggles significantly with other categories, Incorrectly Classifying a large number of Samples, such as the 557 'Unknown' samples predicted as 'Poor'. This indicates that while the hybrid approach is effective for the most common class, its performance is inconsistent and much weaker on less frequent or harder-to-distinguish categories.

CONCLUSION

The results demonstrated that advanced models significantly outperformed simpler baselines. The hyperparameter-tuned Random Forest and the GRU deep learning network yielded the highest predictive accuracy, an outcome consistent with findings in the supporting literature. Key limitations of this project include its reliance on a single dataset and a standard feature set that does not incorporate more complex physical principles.

Future work can expand on this foundation by employing advanced techniques like Bayesian Optimization, exploring state-of-the-art Transformer architectures for large-scale forecasting, or developing physics-guided models for enhanced robustness.

In conclusion, this project provides a robust validation of modern forecasting techniques and serves as a strong foundation for a practical, real-world air quality prediction too.

REFERENCES

[1] Mampitiya, L., Rathnayake, N., Ghosh, R., Hoshino, Y., & Rathnayake, U. (2025). Al-Driven prediction of PM_{10} level in the Republic of Ireland: Integrating Machine Learning for Urban Air Quality Prediction. Aerosol Science and Engineering. https://doi.org/10.1007/s41810-025-00325-0

[2] Kaviani Rad, A., Nematollahi, M. J., Pak, A., & Mahmoudi, M. (2025). Predictive modeling of air quality in the Tehran megacity via deep learning techniques. Scientific Reports, 15(1367). https://doi.org/10.1038/s41598-024-84550-6

[3] Özüpak, Y., Alpsalaz, F., & Aslan, E. (2025). Air Quality Forecasting Using Machine Learning: Comparative Analysis and Ensemble Strategies for Enhanced Prediction. Water, Air, & Soil Pollution, 236(464). https://doi.org/10.1007/s11270-025-08122-8

[4] Hettige, K. H., Ji, J., Xiang, S., Long, C., Cong, G., & Wang, J. (2024). AIRPHYNET: HARNESSING PHYSICS-GUIDED NEURAL NETWORKS FOR AIR QUALITY PREDICTION. Published as a conference paper at ICLR 2024. https://arxiv.org/abs/2402.03784

[5] Liang, Y., Xia, Y., Ke, S., Wang, Y., Wen, Q., Zhang, J., Zheng, Y., & Zimmermann, R. (2023). AirFormer: Predicting Nationwide Air Quality in China with Transformers. Proceedings of the Thirty Seventh AAAI Conference on Artificial Intelligence (AAAI-23). https://github.com/yoshall/airformer

[6] Cican, G., Buturache, A.-N., & Mirea, R. (2023). Applying Machine Learning Techniques in Air Quality Prediction—A Bucharest City Case Study. Sustainability, 15(11), 8445. https://doi.org/10.3390/su15118445

[7] Ravindiran, G., Hayder, G., Kanagarathinam, K., Alagumalai, A., & Sonne, C. (2023). Air quality prediction by machine learning models: A predictive study on the indian coastal city of Visakhapatnam. Chemosphere, 338, 139518. https://doi.org/10.1016/j.chemosphere.2023.139518

[8] Pitale, S., Bhoite, S., & Bhalgat, P. (2019). Air Quality Prediction using Machine Learning Algorithms. International Journal of Computer Applications Technology and Research, 8(09), 367-370. http://www.ijcat.com/

[9] Castelli, M., Manzoni, L., & Vanneschi, L. (2023). A Collaborative Neuroevolutionary Approach to Air Pollution Prediction. Applied Sciences, 13(12), 7013. https://doi.org/10.3390/app13127013

[10] Ayturan, Y., Kunt, M. C., & Karacan, F. (2024). Spatio-temporal air pollution prediction using deep learning: a comprehensive review. Urban Climate, 53, 101783. https://doi.org/10.1016/j.uclim.2023.101783

[11] Kumar, P., Singh, R., Kumar, R., Srivastava, S., & Kumar, D. (2023). Air quality prediction in India using deep learning & machine learning. In 2023 3rd International Conference on Pervasive Computing and Social Networking (ICPCSN). https://doi.org/10.1109/ICPCSN56627.2023.10123531

[12] Masih, A. (2022). Air quality modelling: a review of the state-of-the-art. Environmental Monitoring and Assessment, 194(12), 920. A comprehensive review paper that covers the evolution of air quality modeling, providing excellent context for a literature survey. https://doi.org/10.1007/s10661-022-10595-z

[13] Bai, Y., Zeng, B., Li, C., & Zhang, J. (2022). A novel hybrid model for air quality forecasting based on a recurrent neural network with a convolutional block attention module. Science of The Total Environment, 807, 150794. Details a hybrid deep learning model that uses attention mechanisms, a concept that could enhance your existing GRU/LSTM models. https://doi.org/10.1016/j.scitotenv.2021.150749

[14] Prakash, J., Kumar, S., & Kumar, A. (2023). Air quality index forecasting for Delhi-NCR region using hybrid machine learning models. Urban Climate, 49, 101499. A study focused on the Indian context (Delhi-NCR) that uses hybrid models, directly aligning with your project's geography and methodology. https://doi.org/10.1016/j.uclim.2023.101499

[15] Zaini, N., Ean, L. W., & Ahmed, A. N. (2022). A systematic review of machine learning models for air quality prediction. Environmental Science and Pollution Research, 29(55), 82725-82755. An extensive systematic review that analyzes and compares a wide range of machine learning models used for air quality prediction. https://doi.org/10.1007/s11356-022-22953

[16] Yang, G., Lee, H., & Lee, G. (2021). A hybrid deep learning model to forecast particulate matter concentration levels in Seoul, South Korea. Atmosphere, 11(4), 348. Provides a clear example of combining CNN and LSTM models to capture both spatial and temporal features, a common enhancement for projects like yours. https://doi.org/10.3390/atmos11040348

[17] Ge, L., Wang, Y., & Liu, F. (2022). An attention-based GRU-GCN model for air quality forecasting. Applied Intelligence, 52(12), 14321-14336. Details a model that combines GRU with Graph Convolutional Networks (GCN) to handle spatio-temporal data, representing a clear future enhancement. https://doi.org/10.1007/s10489-022-03299-y

[18] Saini, J., Dutta, M., & Marques, G. (2021). A comprehensive review on indoor air quality monitoring using machine learning and the Internet of Things. Air Quality, Atmosphere & Health, 14(12), 2065-2083. Although focused on indoor air quality, this review covers many of the same models and provides insights into IoT integration for real-time data. https://doi.org/10.1007/s11869-021-01083-9

[19] Eslami, E., Salman, A. K., Choi, Y., Sayeed, A., & Lops, Y. (2022). A stacking deep learning model for air quality prediction. Environmental Science and Pollution Research, 29(12), 17743-17757. Directly relates to your use of ensemble methods, but applies it to deep learning models using a stacking approach for potentially higher accuracy. https://doi.org/10.1007/s11356-021-17186-0

[20] Jain, A., Singh, B., & Upadhyay, P. (2022). AQI prediction for smart cities using machine learning. In Advances in Smart Computing and Bio-Informatics (pp. 237-246). Springer. Another study in the Indian context that focuses on AQI prediction for smart cities, aligning well with your project's application. https://doi.org/10.1007/978-981-16-5764-4_24

[21] Reichstein, M., Camps-Valls, G., Stevens, B., et al. (2019). Deep learning and process understanding for self-explaining Earth system science. Nature, 566(7743), 195-204. A high-level paper that discusses integrating deep learning with physical sciences, relevant to the "physics-guided" concepts in your provided papers. https://doi.org/10.1038/s41586-019-0912-1

[22] Kumar, D., Singh, A. K., Singh, S., & Singh, R. S. (2022). Forecasting of air quality index using a hybrid GRU-based deep learning model. Neural Computing and Applications, 34(16), 13359-13374. Details a hybrid model based on GRU, validating your choice of GRU and suggesting ways to enhance it. https://doi.org/10.1007/s00521-022-07153-6

[23] Wu, Z., Pan, S., Long, G., Jiang, J., Chang, X., & Zhang, C. (2020). Connecting the dots: Multivariate time series forecasting with graph neural networks. In Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. A key paper on using Graph Neural Networks for time-series forecasting, a cutting-edge approach for future enhancements. https://doi.org/10.1145/3394486.3403118

[24] Shahid, F., Zameer, A., & Muneeb, M. (2021). A novel genetic LSTM model for air pollution forecasting. Complexity, 2021. Proposes an interesting combination of a genetic algorithm with LSTM for optimization, another advanced technique for future work. https://doi.org/10.1155/2021/6639391

[25] Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., & Sun, L. (2023). Transformers in time series: A survey. arXiv preprint arXiv:2202.07125. An essential survey paper on using Transformers for time-series data, providing a deep dive into the architecture mentioned in your AirFormer paper. https://arxiv.org/abs/2202.07125

[26] Sofi, I. A., & Magray, A. J. (2023). Deep learning-based air quality prediction model for smart cities. Measurement: Sensors, 25, 100650. Discusses deep learning models in the context of "smart cities," which is a key application area for your project. https://doi.org/10.1016/j.measen.2022.100650

[27] Talaat, M., Ali, A. A., & Talaat, A. (2022). A comparative study of machine learning and deep learning techniques for air quality prediction. Alexandria Engineering Journal, 61(12), 10141-10156. A direct comparative study similar to your project's goal, providing more benchmarks for your results. https://doi.org/10.1016/j.aej.2022.03.064

[**28]** Joloudar, S. Y., & Joloudar, M. Y. (2023). Air pollution prediction using a hybrid model of CNN, LSTM, and attention mechanism. Environmental Science and Pollution Research, 30(2), 3505-3519. Combines three powerful deep learning components (CNN, LSTM, Attention) into a single model for enhanced performance. https://doi.org/10.1007/s11356-022-22534-w

[29] Ma, J., Cheng, J. C., Jiang, F., & Chen, W. (2021). A bi-directionally deep learning model for city-wide air quality prediction. Advanced Engineering Informatics, 49, 101344. Focuses on Bi-directional LSTMs/GRUs, which is a common technique to improve the performance of recurrent neural networks. https://doi.org/10.1016/j.aei.2021.101344

[30] Bhan, A., Sharma, R., & Kumar, R. (2023). Air quality index prediction using CatBoost and SHAP for explainability. Earth Science Informatics, 16(2), 1637-1652. Uses CatBoost (one of the strongest models from your literature) and introduces SHAP for model explainability, a key concept for making "black-box" models more transparent. https://doi.org/10.1007/s12145-023-00977-1

[31] Zhang, Z., Zhang, S., Chen, C., & Yuan, J. (2024). A systematic survey of air quality prediction based on deep learning. Alexandria Engineering Journal, 93, 128-141. A very recent (2024) and systematic survey paper that would be an excellent, up-to-date addition to your literature review. https://doi.org/10.1016/j.aej.2024.03.03.