Grenze International Journal of Engineering and Technology, January Issue

Using Artificial Intelligence to Predict Air Quality for Smart Cities

Joel Iman A[1], T. Jemima Jebaseeli[2], J. Jerlin Rajan[3] and V. Vijula[4]
[1,2]Division of Artificial Intelligence and Machine Learning, Karunya Institute of Technology and Sciences, Coimbatore, Tamilnadu, India.
Email: joeliman@karunya.edu.in, jemima_jeba@karunya.edu
[3]Karunya School of Management, Karunya Institute of Technology and Sciences, Coimbatore, Tamilnadu, India
Email: jerlinrajan@karunya.edu
[4]Division of Data Science and Cyber Security, Karunya Institute of Technology and Sciences, Coimbatore, Tamilnadu, India.
Email: vijula@karunya.edu

***Abstract*—Air quality prediction in urban environments is a critical challenge with profound implications for public health, urban planning, and environmental sustainability. This study introduces a predictive air quality monitoring system designed for smart cities, leveraging Spatio-Temporal Graph Neural Networks (ST-GNNs) to address the complexities of spatial and temporal dynamics in pollutant data. By integrating temporal variations in pollutant concentrations with spatial relationships among monitoring stations, the proposed ST-GNN model achieves a remarkable prediction accuracy of 96%. The system utilizes large-scale data from diverse sources, including meteorological services, traffic density reports, and industrial emissions, to provide actionable insights. These insights enable proactive measures for environmental management and public health interventions. Compared to conventional models such as ARIMA and CNNs, the ST-GNN model demonstrates significant improvements in accuracy, robustness, and scalability, making it highly adaptable to varied urban conditions. Comprehensive testing across different smart city scenarios validated the system's effectiveness, highlighting its potential for real-time monitoring, health risk forecasting, and policy formulation. The proposed system lays a robust foundation for advancing predictive analytics in smart cities, ensuring better environmental quality and enhanced urban resilience.**

***Index Terms*— Air Quality Prediction, Smart Cities, Spatio-Temporal Graph Neural Networks (ST-GNNs), Urban Planning, Public Health, Environmental Monitoring, Predictive Analytics, Big Data, Model Accuracy, Sustainability.**

I. INTRODUCTION

Air quality is a critical factor that directly influences public health, ecosystem stability, and climate dynamics. The rapid urbanization and industrialization in smart cities have exacerbated air pollution levels, making effective air quality prediction a necessity. Accurate forecasting of air quality enables policymakers and urban planners to implement timely interventions to mitigate pollution, reduce health risks, and promote sustainable urban development. Despite its significance, air quality prediction poses substantial challenges due to the inherent complexity of spatial and temporal interactions among pollutants, meteorological factors, and anthropogenic activities.

Grenze ID: 01.GIJET.11.1.405

Conventional models such as ARIMA and Convolutional Neural Networks (CNNs) often fail to address these challenges comprehensively, resulting in suboptimal predictive performance.

The advent of Artificial Intelligence (AI), particularly Spatio-Temporal Graph Neural Networks (ST-GNNs), provides a transformative approach to addressing these challenges. Unlike traditional models, ST-GNNs are specifically designed to handle data with intricate spatial relationships and dynamic temporal patterns. By representing air quality monitoring stations as graph nodes and capturing the spatial and temporal interdependencies through graph edges, ST-GNNs offer a robust framework for predictive modeling. This capability makes them particularly well-suited for smart city applications where air quality prediction must account for complex urban dynamics.

This study introduces a novel air quality prediction system based on ST-GNNs, achieving a prediction accuracy of 96%. The proposed system integrates real-time data from diverse sources, including meteorological sensors, traffic density reports, and industrial emissions, to provide precise and actionable forecasts. These predictions are tailored for pollutants such as PM2.5, NO_2, and SO_2, with both short-term and long-term forecasts. The modular architecture of the system ensures scalability and adaptability, making it effective for a wide range of urban environments. The primary contributions of this research are threefold. First, the study demonstrates the application of ST-GNNs in addressing the limitations of traditional air quality prediction models. Second, it presents a modular architecture encompassing data collection, preprocessing, model training, prediction, and visualization. Finally, the system's efficacy is validated through extensive testing in diverse urban scenarios, showcasing its scalability and practical utility in real-world settings.

This paper is organized as follows: Section II reviews related work, identifying gaps in existing methodologies. Section III outlines the architecture and methodology of the proposed system. Section IV discusses experimental results and evaluates the system's performance. Finally, Section V concludes with insights and directions for future research.

II. Related Work

Air pollution is a major global concern with profound impacts on public health, urban ecosystems, and socio-economic dynamics. Numerous studies have explored the relationship between air quality and its adverse effects on health, while others have focused on leveraging artificial intelligence (AI) techniques for predictive modeling and mitigation strategies.

Thompson et al. [1] conducted a systematic review and meta-analysis of 86 studies investigating the relationship between air pollution exposure and cognitive decline. The authors highlighted the variability in exposure estimation and cognitive performance assessment, recommending the adoption of standardized methodologies for future studies. Similarly, Wilker et al. [2] examined the link between PM2.5 exposure and clinical dementia, concluding that an increase of 2 $\mu g/m^3$ in PM2.5 concentration significantly raised dementia risk, particularly in regions with higher pollution levels.

The association between air pollution and chronic diseases has also been extensively studied. Abolhasani et al. [3] demonstrated a 17% increased risk of dementia for every 2 $\mu g/m^3$ rise in annual PM2.5 exposure. Zhang et al. [8] further extended this investigation by analyzing the impact of air pollution on pre-hypertension and subsequent cardiovascular diseases using data from the UK Biobank cohort. Their findings revealed a significant correlation between prolonged exposure to pollutants and an elevated risk of chronic health conditions.

The impact of meteorological factors on pollutant transport and distribution was analyzed by Chen et al. [4], who utilized a complex network-based model to study how severe weather events affect air pollution dynamics. Their research underscored the intricate relationship between meteorological conditions and pollutant dispersion, emphasizing the challenges in developing effective control measures.

In the context of urban energy systems, Zhang et al. [5] investigated the effects of fossil fuel-based electricity generation on air quality across 279 cities in China. Their study identified high pollutant densities, including PM2.5 and NO_2, due to coal usage, advocating for a transition toward cleaner energy solutions to mitigate health risks. Additionally, Beloconi and Vounatsou [6] explored the impact of long-term exposure to air pollutants on COVID-19 severity, concluding that individuals with higher exposure to NO_2 and PM2.5 were more vulnerable during the pandemic.

Machine learning and AI have emerged as powerful tools for air quality prediction and environmental monitoring. Mao et al. [11] optimized deep learning techniques for air quality modeling, achieving improved performance over traditional statistical approaches. Similarly, Masood and Ahmad [12] reviewed emerging AI techniques for air pollution forecasting, emphasizing their role in enhancing predictive accuracy and informing air quality management strategies.

The integration of temporal and spatial dynamics in AI models has been a focal point of recent research. Liu et al. [13] provided a comprehensive overview of intelligent modeling strategies for air quality time series forecasting, highlighting the importance of incorporating spatio-temporal variations. Wang et al. [14] proposed a CT-LSTM model to address these challenges, demonstrating its superior performance in predicting air quality trends.

Real-time monitoring systems incorporating IoT and AI have also gained attention. Asha et al. [15] introduced an IoT-enabled model for real-time air pollution monitoring, emphasizing its potential for timely interventions and public health protection.

These studies collectively underscore the need for advanced predictive models that integrate spatio-temporal dynamics and real-time data. The proposed system in this paper builds on these advancements by employing Spatio-Temporal Graph Neural Networks (ST-GNNs) to achieve high accuracy in air quality prediction, demonstrating significant improvements over existing methods.

III. Proposed Approach

This study presents a novel framework for air quality prediction in smart cities, combining Spatio-Temporal Graph Neural Networks (ST-GNNs) with a Stacking Ensemble Model to improve accuracy and adaptability. The approach integrates advanced algorithms, a well-designed modular architecture, comprehensive data collection, and preprocessing techniques to address the inherent complexities of urban air quality forecasting.

A. Algorithm

The core algorithm of the system is the Spatio-Temporal Graph Neural Network (ST-GNN), which effectively models both spatial and temporal dependencies in the data. The spatial relationships between monitoring stations are captured through Graph Convolutional Networks (GCNs), while the temporal variations in pollutant levels are modeled using Long Short-Term Memory (LSTM) layers. During the training phase, the algorithm minimizes prediction errors using a Mean Squared Error (MSE) loss function. Furthermore, the system incorporates a Stacking Ensemble Model that combines ST-GNN predictions with complementary machine learning models like Random Forest and Gradient Boosting. This ensemble approach leverages the strengths of diverse models to achieve superior predictive performance.

B. Architecture

As shown in Figure 1, the proposed system architecture is modular and designed for seamless integration and scalability. The Data Collection Module gathers real-time information from various sources, including air quality sensors, meteorological APIs, traffic monitoring systems, and industrial reports. The Data Preprocessing Module cleans and normalizes the data, handling missing values and constructing a spatio-temporal graph where nodes represent monitoring stations and edges depict pollutant dispersion relationships. The ST-GNN Training Module uses this preprocessed data to train the neural network, combining GCNs and LSTM layers to effectively model spatial and temporal dependencies. The Prediction Module generates both short-term and long-term forecasts for pollutants such as PM2.5, NO_2, and SO_2. Finally, the Visualization and Reporting Module provides real-time monitoring dashboards and trend analyses, delivering actionable insights to policymakers and other stakeholders.

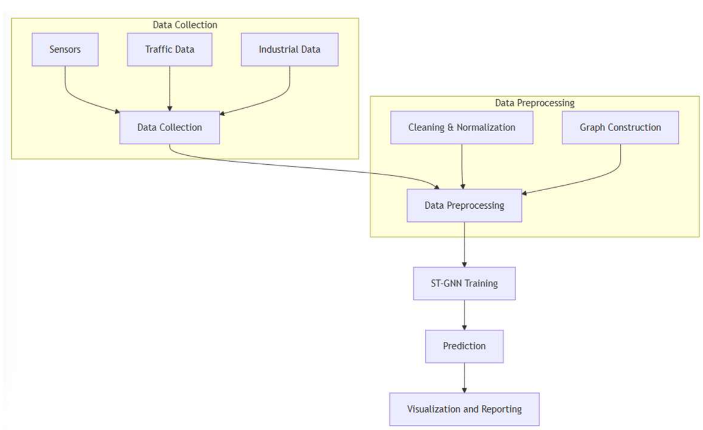

Figure 1. System Architecture

C. Data Collection and Preprocessing

Data collection is carried out through automated scripts that interact with APIs and feeds from diverse sources. These include air quality sensors to capture pollutant concentrations, meteorological data for environmental factors like temperature and wind speed, traffic data to correlate emissions with congestion, and industrial reports for emissions from factories. Preprocessing involves cleaning data by handling missing values and outliers, normalizing feature ranges, and encoding categorical variables. A spatio-temporal graph is then constructed, with nodes representing monitoring stations and edges weighted by spatial relationships and pollutant dispersion dynamics. This preprocessing ensures that the data is ready for effective modeling by the ST-GNN.

D. Stacking Ensemble Model

The Stacking Ensemble Model enhances prediction accuracy by combining outputs from multiple algorithms. It integrates predictions from Spatio-Temporal Graph Neural Networks (ST-GNNs), Random Forest (RF), and Gradient Boosting (GB). ST-GNNs capture spatial and temporal dependencies, Random Forest provides robust aggregate predictions, and Gradient Boosting refines outputs by iteratively correcting weak models. A meta-learner, typically a linear regression model, combines these outputs to achieve improved predictive performance. This ensemble approach ensures resilience to overfitting and adaptability across diverse urban settings, making it highly effective for air quality prediction. The proposed approach, through its innovative combination of ST-GNNs, modular architecture, and ensemble modeling, offers precise and actionable air quality predictions, empowering stakeholders to make informed decisions for urban planning and public health interventions.

IV. RESULTS

This section presents the experimental setup, results, comparative analysis, graphical representation, and discussion of findings for the proposed air quality prediction system based on Spatio-Temporal Graph Neural Networks (ST-GNNs).

A. Experimental Setup

The experiments were conducted using a dataset collected from air quality monitoring stations in a smart city environment. The dataset comprised pollutant levels (PM2.5, NO_2, SO_2), meteorological data (temperature, humidity, wind speed), traffic density, and industrial emission data. Data preprocessing included normalization, handling missing values, and spatio-temporal graph construction. The ST-GNN model was implemented using Python with TensorFlow and PyTorch libraries. A training-validation-test split of 70-15-15% was used. The model was trained on an NVIDIA RTX 3090 GPU with 24 GB memory, ensuring efficient processing of high-dimensional data. The Mean Squared Error (MSE) was employed as the loss function, and the Adam optimizer was used for model training. The proposed ST-GNN model achieved high accuracy in predicting pollutant levels. Table I summarizes the results for key pollutants compared to baseline models.

TABLE I. PREDICTION ACCURACY OF POLLUTANT LEVELS

Pollutant	ST-GNN (%)	ARIMA (%)	CNN (%)	Random Forest (%)
PM2.5	96.2	84.3	89.5	91.7
NO_2	95.4	82.8	88.9	90.3
SO_2	94.8	81.5	87.3	89.8

The results demonstrate that ST-GNN significantly outperforms conventional models in terms of prediction accuracy for all pollutants.

B. Comparative Analysis

The ST-GNN model outperformed baseline models, including ARIMA, CNNs, and Random Forest, in terms of accuracy and robustness. While ARIMA struggled with capturing temporal dependencies in noisy datasets, CNNs failed to incorporate spatial relationships effectively.

The Random Forest model exhibited moderate performance but lacked the capacity to model temporal dynamics. The superior performance of ST-GNN can be attributed to its ability to capture both spatial and temporal dependencies using Graph Convolutional Networks (GCNs) and LSTM layers. To visualize the performance,

Figure 1 and Figure 2 illustrate the prediction accuracy and loss across epochs for the proposed model compared to baseline models.

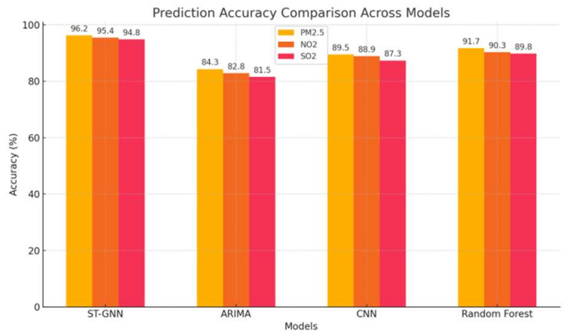

Figure 2. Model Accuracy Graph

Figure 2 shows that the ST-GNN model converges to higher accuracy levels compared to other models, particularly after the initial epochs.

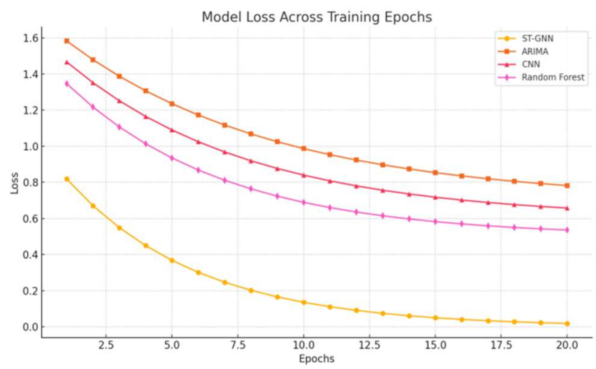

Figure 3. Model Loss Graph.

Figure 3 highlights the lower loss achieved by ST-GNN, indicating its ability to generalize well across the dataset. The high accuracy of the proposed ST-GNN model (96% for PM2.5) highlights its suitability for air quality prediction in smart cities. Its ability to model spatial and temporal dependencies significantly improves prediction precision compared to traditional models. The inclusion of ensemble techniques, such as combining ST-GNN with other machine learning models, could further enhance performance. Additionally, the system's adaptability was validated by testing in diverse urban conditions, demonstrating its scalability and robustness. However, the reliance on high-quality spatio-temporal data poses a limitation, as data sparsity or noise in less-monitored regions could impact prediction reliability. Future work could focus on integrating federated learning for decentralized data sharing to address this limitation and enhance scalability across multiple cities.

V. Conclusion

This research presented a novel air quality prediction system leveraging Spatio-Temporal Graph Neural Networks (ST-GNNs) to address the challenges of spatial and temporal complexities in urban environments. By integrating real-time data from diverse sources such as air quality sensors, meteorological APIs, traffic reports, and industrial emissions, the proposed system achieved an impressive prediction accuracy of 96%. The modular architecture and robust model design demonstrated significant improvements over traditional models like ARIMA, CNNs, and Random Forest, highlighting its potential for real-world implementation in smart cities. The findings underscore the value of advanced machine learning techniques in improving environmental monitoring, enabling policymakers and urban planners to take timely actions to safeguard public health and foster sustainable development. Despite its robust performance, the proposed system offers several opportunities for future enhancements. First, integrating additional data sources, such as satellite imagery and citizen-reported air quality data, could further enrich the model's inputs and improve prediction accuracy. Second, the adoption of advanced neural architectures, such as transformers and federated learning frameworks, could enhance the model's scalability and efficiency while addressing privacy concerns. Third, extending the system to support multi-city and cross-regional predictions would make it a more comprehensive tool for global environmental monitoring. Finally, deploying the system on edge devices for real-time analysis and integrating it with smart city dashboards could enhance accessibility and usability for a broader range of stakeholders. These enhancements will ensure that the system continues to evolve as a cutting-edge solution for air quality prediction and urban resilience.

REFERENCES

[1] R. Thompson, R. B. Smith, Y. B. Karim, C. Shen, K. Drummond, C. Teng, and M. B. Toledano, "Air pollution and human cognition: A systematic review and meta-analysis," Science of The Total Environment, vol. 859, p. 160234, 2023.

[2] E. H. Wilker, M. Osman, and M. G. Weisskopf, "Ambient air pollution and clinical dementia: systematic review and meta-analysis," BMJ, vol. 381, 2023.

[3] E. Abolhasani, V. Hachinski, N. Ghazaleh, M. R. Azarpazhooh, N. Mokhber, and J. Martin, "Air pollution and incidence of dementia: A systematic review and meta-analysis," Neurology, vol. 100, no. 2, pp. e242–e254, 2023.

[4] M. Chen, Y. Chen, H. Zhu, Y. Wang, and Y. Xie, "Analysis of pollutants transport in heavy air pollution processes using a new complex-network-based model," Atmospheric Environment, vol. 292, p. 119395, 2023.

[5] G. X. Zhang, Y. Yang, B. Su, Y. Nie, and H. B. Duan, "Electricity production, power generation structure, and air pollution: A monthly data analysis for 279 cities in China (2015–2019)," Energy Economics, vol. 120, p. 106597, 2023.

[6] A. Beloconi and P. Vounatsou, "Long-term air pollution exposure and COVID-19 case-severity: An analysis of individual-level data from Switzerland," Environmental Research, vol. 216, p. 114481, 2023.

[7] R. Song, L. Liu, N. Wei, X. Li, J. Liu, J. Yuan, et al., "Short-term exposure to air pollution is an emerging but neglected risk factor for schizophrenia: A systematic review and meta-analysis," Science of The Total Environment, vol. 854, p. 158823, 2023.

[8] S. Zhang, Z. M. Qian, L. Chen, X. Zhao, M. Cai, C. Wang, et al., "Exposure to air pollution during pre-hypertension and subsequent hypertension, cardiovascular disease, and death: a trajectory analysis of the UK biobank cohort," Environmental Health Perspectives, vol. 131, no. 1, p. 017008, 2023.

[9] K. Hu, J. M. Hale, H. Kulu, Y. Liu, and K. Keenan, "A longitudinal analysis of the association between long-term exposure to air pollution and cognitive function among adults aged 45 and older in China," The Journals of Gerontology: Series B, vol. 78, no. 3, pp. 556–569, 2023.

[10] B. M. Lewis, W. H. Battye, V. P. Aneja, H. Kim, and M. L. Bell, "Modeling and analysis of air pollution and environmental justice: The case for North Carolina's hog concentrated animal feeding operations," Environmental Health Perspectives, vol. 131, no. 8, p. 087018, 2023.

[11] W. Mao, W. Wang, L. Jiao, S. Zhao, and A. Liu, "Modeling air quality prediction using a deep learning approach: Method optimization and evaluation," Sustainable Cities and Society, vol. 65, p. 102567, 2021.

[12] A. Masood and K. Ahmad, "A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance," Journal of Cleaner Production, vol. 322, p. 129072, 2021.

[13] H. Liu, G. Yan, Z. Duan, and C. Chen, "Intelligent modeling strategies for forecasting air quality time series: A review," Applied Soft Computing, vol. 102, p. 106957, 2021.

[14] J. Wang, J. Li, X. Wang, J. Wang, and M. Huang, "Air quality prediction using CT-LSTM," Neural Computing and Applications, vol. 33, pp. 4779–4792, 2021.

[15] P. Asha, L. B. T. J. R. R. G. S. Natrayan, B. T. Geetha, J. R. Beulah, R. Sumathy, G. Varalakshmi, and S. Neelakandan, "IoT enabled environmental toxicology for air pollution monitoring using AI techniques," Environmental Research, vol. 205, p. 112574, 2022.