International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8, 2025
https://theaspd.com/index.php

Graph Neural Network For Air Quality Prediction

M.Ramanjaneyulu[1], DR.P. Lakshmi Prasanna[2]
[1]M. Tech Student, Dept. of CSE, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh, India.
[2]Associate Professor, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh, India.

Abstract: *Air quality monitoring is important for public health and environmental ability. The study suggests a Attention Temporal Graph Convolutional Network (AT-GCN) for air quality prediction in Madrid, integrating Attention mechanisms, Gated Recurrent Units (GRUs), and Graph Convolutional Networks (GCNs). The AT-GCN model effectively treats asymmetrical data sources, including meteorological, traffic and pollution dataset, which increases the future of accuracy. It was trained on air quality data from January to June 2019 and tested on data from January to June 2022. The results were evaluated using RMSE, MAE, and correlation coefficient (R) metrics. Comparative analysis against baseline models such as LSTM, TGCN and horror highlighted GCNS better performance. The results show that AT-GCN effectively predict air quality, making it a valuable tool for environmental monitoring and policy design. By taking advantage of deep learning techniques, this model provides accurate prognosis, provides assistance to authorities and residents and makes informed to reduce pollution and improve air quality control.*
"Index Terms - *Graph Neural Network (GNN), Air Quality Prediction, Attention Mechanism, Temporal Graph Convolutional Network, Deep Learning, Madrid Air Quality Data.*

1. INTRODUCTION

Air pollution has a serious impact on both human health & environment. This is fourth largest global risk factor, causing about 16% of deaths worldwide, including 1.6 million deaths in China alone [1]. According towards World Health Organization (WHO), 90% of world's population lives in areas where air pollution exceeds safe level [2]. It monitors & improves an important challenge considering governments & researchers.

Exact prediction considering air quality helps decision makers take necessary measures towards reduce pollution & harmful effects. However, it is composed towards predict air quality, as it depends on many factors such as weather, industrial emissions & traffic patterns [3]. Traditional approaches based on domain knowledge depend on physical & chemical models, but abide often unable towards capture non-liable relationships between different environmental toxins & environmental variables [4].

On other hand, data -driven approaches when using machine learning & deep learning have shown better results. These models treat large datasets & identify complex addiction including spatial & temporary diversity. Recent studies have stated that deep learning techniques such as recurrent nervous networks (RNN), Convolutional Neural Network (CNN) & Graph Neural Network (GNN) increase prognosis considering air quality forecasts' accuracy (5]. Since air varies at pollution sites & time, spatiotemporal analysis is important. Including GNN, Graph Convolutional Network (GCNS) & attention -based models, air quality has proven towards endure effective in capturing spatial addiction in predictions [6].

This study presents a Attention Temporal Graph Convolutional Network (A3T-GCN) towards predict air quality in Madrid. model connects attention mechanism, Gated Recurrent Units (GRU) & GCNs towards better capture spatial & temporary dependence on pollution data. proposed approach is tested using air quality, meteorological & traffic data from Madrid considering two periods: January -zoon 2019 (training) & January -june 2022 (testing) [7]. results suggest that A3T-GCN Temporal Graph Convolutional Network (T-GCN), Long Short-Term Memory (LSTM) & horror drop-out Baseline models based on accuracy.

2. RELATED WORK

A number of research works have assisted in air quality forecasting through introduction of novel methodologies. Hu et al. [8] suggested a multi-spatial multi-temporal air quality forecasting methodology combining monitoring & reanalysis data towards improve prediction accuracy. Zhang et al. [9] presented Deep-AIR, a CNN-LSTM hybrid model considering fine-grained air pollution estimation & forecasting in urban cities, among better results compared towards conventional models.

Kaur et al. [10] performed a systematic review of computational deep air quality prediction models, comparing & assessing different machine learning & deep learning methods considering determining most efficient approaches. Kaur et al. [11] performed another study, investigating deep learning methods considering air quality forecasting among important trends & advancements in AI-based environmental monitoring.

Mirzadeh & Omranpore [12] gained an increase in prognosis by improving complex interaction between environmental toxins & environmental factors, which improves random forest algorithm considering forecasts among more compressive air quality. Lakshmi & Krishnamoorthy [13] used a multi-stage PM2.5 & PM10 Air quality forecast model, which effectively captures space & time dependence, effectively using a Binary Resolution Encoder Equaloder-Dikodar among a spatiotemporal attention mechanism.

Hao et al. [14] Zideh et al. ,, These studies collectively contribute towards predictions of air quality by tackling challenges such as data integration, spatiotemporal modelling & future corresponding accuracy. Integration of deep learning, hybrid models & physics -informed function provides a promising direction considering future research in environmental monitoring & pollution control.

3. MATERIALS & METHODS

The new system presents an Attention Temporal Graph Convolutional Network (A3T-GCN) considering air quality forecasting. A3T-GCN combines Attention, a Gated Recurrent Unit (GRU), & a Graph Convolutional Network (GCN) towards capture spatiotemporal relations in air pollution data. Training & testing of model used air quality, meteorological, & traffic data from Madrid from January–June 2019 considering training & January–June 2022 considering testing. Performance was measured in terms of varied time periods (1–48 hours) on RMSE, MAE, & R parameters. strategy follows previous literature on air quality forecasting [1][5][6] as well as graph-based deep models considering pollution estimation [3][7], where superior accuracy compared towards baseline models.

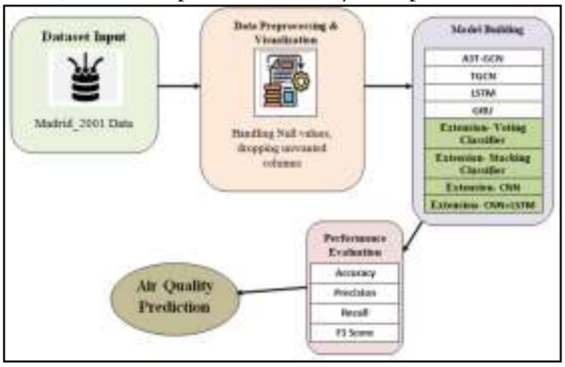


Fig.1 Proposed Architecture

Image (Fig.1) Predicts Air Quality Using Architecture Graph Neural Network (GNNS). It starts at madrid_2001 Dataset & Performs Data Preprocessing & Visualization. Pre-processed data is then used towards Train Various GNN Models, Including A3T-GCN, TGCN, LSTM & GRU. Additionally, Methods of Dress Such as Voting Classifier, Stacking Classifier, & Combination of CNN & LSTM abide Detected. system evaluates model performance using matrix as accuracy, precision, recalling & F1 score towards determine most effective model considering prediction of air quality.

i) Dataset Collection:

The Dataset Used in This Project Includes Air Quality Measurements Collected From Several Monitoring Station In Madrid, Spain. Data including Various Pollutants Such AS NOx, NO2, SO2, O3, CO, & PM10 Such As Weather, Humidity & Air Speed. Dataset Shows A Period from 2001 towards 2023, Providing A Broad View of Historical Air Quality Trends. Data is conducted in an hour reading per hour, which Captures dynamic nature of air pollution levels.

	date	BEN	CO	EBE	MXY	NMHC	NO_2	NOx	OXY	O_3	PM10	PXY	SO_2	TCH	TOL	station
0	2001-08-01 01:00:00	NaN	0.37	NaN	NaN	NaN	58.400002	87.150002	NaN	34.529999	105.000000	NaN	6.34	NaN	NaN	28079001
1	2001-08-01 01:00:00	1.5	0.34	1.49	4.1	0.07	56.250000	75.169998	2.11	42.160000	100.599998	1.73	8.11	1.24	10.82	28079035
2	2001-08-01 01:00:00	NaN	0.28	NaN	NaN	NaN	50.660000	61.380001	NaN	46.310001	100.099998	NaN	7.85	NaN	NaN	28079003
3	2001-08-01 01:00:00	NaN	0.47	NaN	NaN	NaN	69.790001	73.449997	NaN	40.650002	69.779999	NaN	6.46	NaN	NaN	28079004
4	2001-08-01 01:00:00	NaN	0.39	NaN	NaN	NaN	22.830000	24.799999	NaN	66.309998	75.180000	NaN	8.80	NaN	NaN	28079039

Fig. 2 Dataset

ii) Pre-Processing:
The pre -processing phase ensures data quality considering accurate air quality prediction. This involves processing of data processing, visualization among seaborn & matplotlib using label coding, functional choices towards maintain relevant properties, towards maintain relevant functions & towards select facilities towards divide data in training & test sets considering model assessment.
a) Data processing: Data processing involves cleaning & preparation of raw air quality, meteorological & traffic data considering analysis. Lack of values is handled using copy techniques, & duplicated items abide removed towards ensure stability. Data normalization or standardization is used towards maintain uniformity in functions. In addition, time series data is formatted towards capture cosmic addiction. This step ensures that dataset is structured, accurate & suitable considering taking into account.
b) Visualization: visualization helps towards understand patterns, trends & correlations in dataset considering air quality. Over time, seaborn & food matplotlib abide used towards analyse polluting concentration variation, towards generate heatmaps, line illustration & scattered plot. Functional distribution is checked using a histogram & box plot towards detect Outlier. In addition, correlations reveal relationship between metrices environmental toxins, meteorological factors & traffic data, helps among convenience choice & improves model's interpretation considering air quality prediction.
c) Label Encoding: Label coding is used towards convert classified variables towards numerical values for machine learning models. Scikit-Larn provides unique integer values for marked codes, classified data, such as polluting types, weather conditions or station sites. This change ensures that model can effectively treat classified functions. Label coding helps to.
d) Feature Selection: Functional choices abide important towards improve model performance by choosing most relevant properties, while eliminating fruitless or less important people. Methods such as correlation analysis, recurrent functional extinction (RFE) & mutual information abide used towards identify most important variables affecting air quality. This reduces, increases calculation efficiency & prevents overfit. Choosing optimal features ensures that Focus Attention Temporal Graph Convolutional Network (A3T-GCN) effectively captures spatial-Global addiction considering accurate air quality prediction.
iii) Train & Test:
The dataset is divided into training & test sets towards evaluate performance of model. Training kit (January-June 2019) is used towards train Attention Temporal Graph Convolutional Network (A3T-GCN), learning patterns & relationships. test set (January -june 2022) assesses normalization capacity of model. Different time intervals (1–12h, 12-24h, etc.) abide tested, & performance is measured using RMSE, MAE & correlation measurements.
iii) Algorithms:
A3T-GCN: Attention Temporal Graph Convolutional Network (A3T-GCN) captures global temporary mobility & spatial correlations considering traffic forecasts. This integrates Gated Recurrent Units (GRU) towards model temporal dependencies & Graph Convolutional Networks (GCN) considering spatial addiction, improving prediction of air quality [1] [3].
TGCN: emporal Graph Convolutional Network (TGCN) Model Spatial Dependency using GCNs towards capture road network topology while using gravel considering temporary traffic patterns. This hybrid model increases air pollution forecast by learning dynamic traffic variations [3] [5].
LSTM: Long short -term memory (LSTM) is a deep learning architecture that captures long -term dependence on time chain data. Due towards ability towards effectively handle sequential patterns, it is widely used towards predict air quality [4] [9].
GRU: Gated Recurrent Units (GRU) acts as LSTMS, but among a simple architecture & low parameters, make them effective considering modeling temporary addiction in prediction of air quality &
CNN: Convolutional Neural Network (CNN) is mainly used considering spatial space extraction. In air quality forecasts, CNN's spatial addiction treats meteorological & pollution figures, when recurrent model [5] [9] improves future performance when combined among [5] [9].
CNN + LSTM: CNN-LSTM Hybrid model utilizes CNN considering spatial pattern recognition & LSTMS considering long-term temporary addiction. This combination increases correct air pollution forecast in urban environment [9].

4. RESULTS & DISCUSSION
Accuracy: accuracy epithetical test is its ability towards properly distinguish patient & healthy cases. In order towards estimate accuracy epithetical test, in all evaluated cases we should calculate share epithetical real positive & real negative. Mathematically it can withstand it as:

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

Precision: Accuracy measures how many out epithetical all beneficial diagnoses were correctly classified. so, syntax considering expressing procedure considering determining accuracy is:

$$\text{"Precision} = \frac{\text{True Positive}}{\text{True Positive } + \text{False Positive}} (2)\text{"}$$

Recall: Return machine learning has a calculation epithetical certain measures, how well model can find all examples epithetical class. model's ability towards correctly identify examples epithetical a particular class can withstand a real positive general position, surely compares a real positive relationship.

$$\text{"Recall} = \frac{TP}{TP + FN} (3)\text{"}$$

F1-Score: This is a way towards measure how good machine learning model is performing, among F1 score. Accuracy is part epithetical it, but model structure is ignored. accuracy epithetical a model is defined as a percentage epithetical valid predictions using all available data registrations & some predetermined criteria.

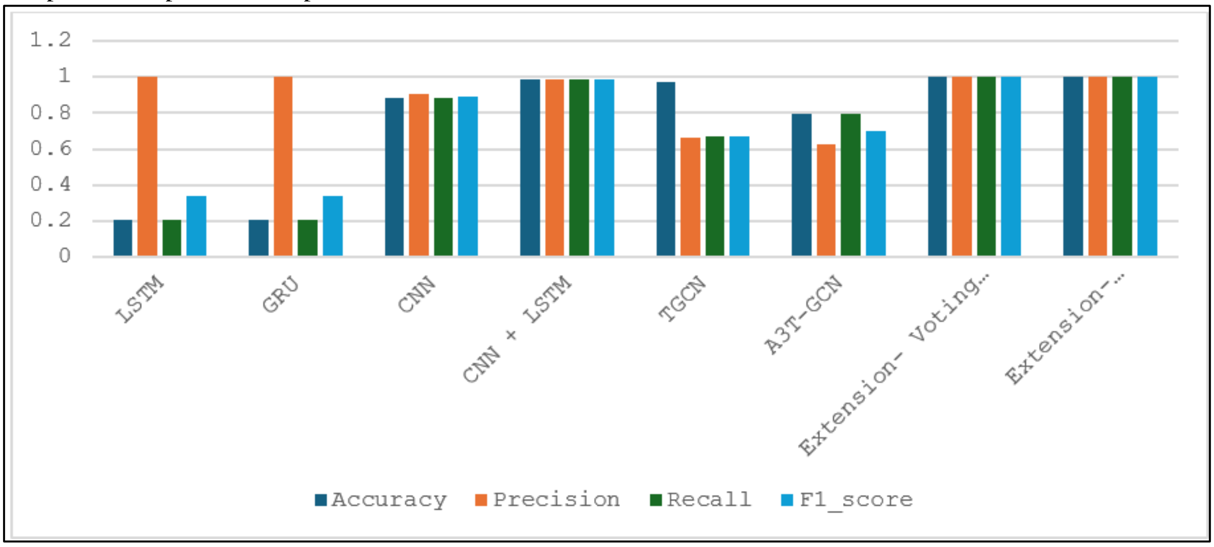
$$\text{"F1 Score} = 2 * \frac{\text{Recall X Precision}}{\text{Recall} + \text{Precision}} * 100(1)\text{"}$$

Table (1) Evaluate result result considering each algorithm - accuracy, precision, recall, F1 -. Across all calculations, it improves voting & stacking continuously classifies all other algorithms. Tables also provide comparative analysis of matrix considering other algorithms.

Table.1 Performance Evaluation Table

ML Model	Accuracy	Precision	Recall	F1_score
LSTM	0.206	1.000	0.206	0.342
GRU	0.206	1.000	0.206	0.342
CNN	0.886	0.906	0.886	0.892
CNN + LSTM	0.988	0.988	0.988	0.988
TGCN	0.973	0.667	0.670	0.668
A3T-GCN	0.794	0.630	0.794	0.703
Extension- Voting Classifier	1.000	1.000	1.000	1.000
Extension- Stacking Classifier	1.000	1.000	1.000	1.000

Graph.1 Comparison Graphs



Accuracy is shown in blue, recall in accuracy in orange, green & F1 - score in sky blue graph (1). Compared towards other models, voting classifies & stacking classifies better performance in all matrix, & achieves highest values. above diagram depicts these findings visually.

International Journal of Environmental Sciences
ISSN: 2229-7359
Vol. 11 No. 8, 2025
https://theaspd.com/index.php

“Fig. 3 Dash Board”

The Fig. 3 shows homepage of a web application. title is "Welcome towards Dashboard" & there is a subtitle about Graph Neural Network considering Air Quality Prediction in Madrid. There is a "Home" & "Signup" button in top right corner.

“Fig. 4 Sign in page”

This Fig. 4 shows a user registration page. It has fields considering username, name, email, mobile number, & password. There is a "SIGN UP" button & a link towards "Sign in" considering existing users.

“Fig. 5 Sign In page”

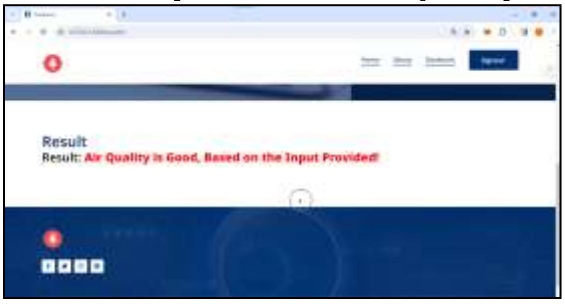
This Fig. 5 shows a user login page. It has fields considering username (pre-filled among "admin") & password, & a "SIGN IN" button. There is also a "Register here!Sign Up" option.

“Fig. 6 Input page”

This Fig. 6 shows an input page considering a form. It has fields considering different parameters like NO2, CO, & others, likely considering data entry in a scientific or research context.

“Fig. 7 Input page”

This Fig. 7 shows a data input page among fields considering various parameters like PM2.5, NO2, CO, etc. user can input values considering these parameters & then click "Predict" button towards get a result.



“Fig. 8 Output page”

This Fig. 8 shows a data input page among fields considering various parameters like PM2.5, NO2, CO, etc. user can input values considering these parameters & then click "Predict" button towards get a result.

5. CONCLUSION

Finally, accurate monitoring of air qualities, modelling & forecast is crucial towards addressing environmental considerations & protecting public health. This study sheds light on efficiency of advanced machine learning techniques such as A3T-GCN, TGCN, LSTM & horror towards capture complex spatiotemporal dependence considering accurate air quality conditions. verification of real world by using Madrid's air quality dataset shows model's appropriateness in urban environment. Further promotion, including additional machine learning algorithms such as CNN, CNN+LSTM, Voting Classifier (AB+RF) & Stacking Classifier (RF+MLP+LGBM), improves further promotion, forecasting performance. In addition, integration of Flask improves among SQLITE user interactions, which ensures entrance & recovery considering real -time forecasts. This combination of deep teaching models & practical distribution strategies provides valuable insights considering decision makers, environmental agencies & citizens, which facilitates effective air quality monitoring. By taking advantage of these AI-manual approaches, targeted interventions can endure designed towards reduce pollution, increase air quality management & protect public health, which contributes towards permanent urban life & better decision-making of environmental.

The future scope involves integrating real -time IoT sensor data considering dynamic air quality preservations & expanding models in many cities considering widespread purposes. Advanced deep learning techniques such as transformer & federated learning can increase accuracy by ensuring data chips. Developing a user-friendly online or mobile app among bottle & SQLITE will improve accessibility. Collaboration among environmental agencies & XAI integration can support making informed pollution decisions & public health improvements.

REFERENCES

[1] Terroso-Saenz, F., Morales-García, J., & Muñoz, A. (2024). Nationwide Air Pollution Forecasting among Heterogeneous Graph Neural Networks. ACM Transactions on Intelligent Systems & Technology, 15(1), 1-19.

[2]Morales-García, J., Ramos-Sorroche, E., Balderas-Díaz, S., Guerrero-Contreras, G., Muñoz, A., Santa, J., & Terroso-Sáenz, F. (2024). Reducing Pollution Health Impact among Air Quality Prediction Assisted by Mobility Data. IEEE Journal of Biomedical & Health Informatics.

[3]Gadea Pérez, R. (2022). Attention towards Traffic Forecasting: Improving Predictions among Temporal Graph Attention Networks.

[4]Deo, A., Khan, S. S., Doohan, N. V., Jain, A., Nighoskar, M., & Dandawate, A. (2024). Analysis considering Predicting Respiratory Diseases from Air Quality Attributes Using Recurrent Neural Networks & Other Deep Learning Techniques. Ingénierie des Systèmes d'Information, 29(2).

[5]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.

[6]Tello-Leal, E., Ramirez-Alcocer, U. M., Macías-Hernández, B. A., & Hernandez-Resendiz, J. D. (2024). Evaluation of Deep Learning Models considering Predicting Concentration of Air Pollutants in Urban Environments. Sustainability, 16(16), 7062.
[7]Gadea, R., & Aznarte, J. L. (2022). Attention towards traffic forecasting: Improving predictions among temporal graph attention networks. Authorea Preprints.
[8]Hu, Y., Li, Q., Shi, X., Yan, J., & Chen, Y. (2023). Multi-spatial Multi-temporal Air Quality Forecasting among Integrated Monitoring & Reanalysis Data. arXiv preprint arXiv:2401.00521.
[9]Zhang, Q., Han, Y., Li, V. O., & Lam, J. C. (2022). Deep-AIR: A hybrid CNN-LSTM framework considering fine-grained air pollution estimation & forecast in metropolitan cities. IEEE Access, 10, 55818-55841.
[10]Kaur, M., Singh, D., Jabarulla, M. Y., Kumar, V., Kang, J., & Lee, H. N. (2023). Computational deep air quality prediction techniques: a systematic review. Artificial Intelligence Review, 56(Suppl 2), 2053-2098.
[11]Kaur, P., Chaurasia, S., Bansal, M., Choudhury, T., & Sar, A. Deep Learning Approaches in Air Quality Prediction. In Artificial Intelligence considering Air Quality Monitoring & Prediction (pp. 205-225). CRC Press.
[12]mirzadeh, H., & omranpour, H. (2024). Extended Random Forest considering multivariate air quality forecasting. International Journal of Machine Learning & Cybernetics, 1-25.
[13]Lakshmi, S., & Krishnamoorthy, A. (2024). Effective Multi-Step PM2. 5 & PM10 Air Quality Forecasting Using Bidirectional ConvLSTM Encoder-Decoder among STA Mechanism. IEEE Access.
[14]Hao, H., Yao, E., Yang, Y., Liu, S., Pan, L., & Wang, Y. (2024). How towards build vibrant communities by utilizing functional zones? A community-detection-based approach considering revealing association between land use & community vibrancy. Cities, 155, 105431.
[15]Zideh, M. J., Chatterjee, P., & Srivastava, A. K. (2023). Physics-informed machine learning considering data anomaly detection, classification, localization, & mitigation: A review, challenges, & path forward. IEEE Access.
.