ITM Web of Conferences **79**, 01048 (2025) https://doi.org/10.1051/itmconf/20257901048
KEIS-2025

Multi-Source Data Fusion using Graph Attention Residual GCN for Urban Air Quality Prediction

Subhra Chakraborty[1], *Mahadev* Kalyanrao Patil[2], *S. G.* Shivaprasad Yadav[3*], *Girija* Vani. G[4], and *Talari* Ashok[5]

[1]Department of Electronics and Communication Engineering, Nitte Meenakshi Institute of Technology, Nitte (Deemed to be University), Bengaluru, India
[2]School of Computer Science and Engineering, R V University, Bengaluru, India
[3]Department of Electronics and Telecommunication Engineering, M S Ramaiah Institute of Technology, Bengaluru, India
[4]Department of Electronics and Communication Engineering, Ballari Institute of Technology and Management, Ballari, India
[5]Department of Computer Science and Engineering, Vignan's Foundation for Science, Technology and Research, A Deemed to be University, Hyderabad, India

Abstract. Air quality forecasting has become essential for sustainable urban management and public health. However, traditional deep spatio-temporal methods often rely on Dynamic Time Warping (DTW), which introduces high computational costs and limited scalability, making real-time forecasting challenging. Therefore, lightweight framework combining Graph Attention Networks and Residual Graph Convolutional Networks (GAT-ResGCN) is proposed for efficient, fine-grained PM2.5. Initially, meteorological time-series data were gathered from India's Air Quality Index dataset and satellite-derived PM2.5, maps from Modern-Era Retrospective Analysis for Research and Applications (MERRA-2) reanalysis dataset. Preprocessing included missing value imputation, Z-score normalization, and temporal alignment. Subsequently, GAT dynamically models spatial dependencies using k-nearest neighbor (k-NN) sparsification and Exponential Moving Average (EMA) updates. Furthermore, Spatio-temporal modelling was achieved through a Dynamic Spatio-Temporal Graph Convolutional Network (DSTGCN). Moreover, it integrates GAT-based graph convolutions, dilated Temporal Convolutional Networks (TCN) for long-range patterns, and temporal attention to highlight critical intervals. In addition, the Residual Network processes remote-sensing images to extract spatial pollutant patterns, which are fused with time-series features using a gated mechanism. The experimental results showed improved accuracy and scalability, achieving a root mean square error of 18.21, a Mean Absolute Error of 11.35, and an R^2 of 0.92.

1 Introduction

In modern times, the rapid growth in global industrialization has led to improvements in living standards by improving new technologies, but this disturbs the quality of air in the atmosphere [1]. The quality of air is more important for human survival, as air pollutants increase respiratory illnesses and cancers and lead to climate change. With the acceleration of urbanization, many difficulties are growing, like traffic congestion and environmental pollution, and timely knowledge of pollution conditions will make changes in increase air quality [2]. Moreover, forecasting plays an important role in preventing air pollution based on timely environmental conditions. However, there are a large number of air quality issues in the environment that cause harmful health illnesses due to air pollutants, such as Carbon Dioxide (CO) and sulfur dioxide (SO2) [3]. Particulate matter (PM2.5) has become a serious problem that affects air quality. Moreover, rural and remote sites contain more pollution owing to high ozone (O3) because of precursors. Furthermore, monitoring the air quality number of ground stations in big cities is low, which causes uneven distributions of stations and significant differences in air quality. Machine learning methods capture only non-temporal relations and suffer from complex forecasting [4]. Conventional methods capture high-dimensional relationships for predicting air quality; however, these methods require a large amount of data and produce a decrease in accuracy [5]. Statistical Machine Learning (ML) and Deep Learning (DL) approaches have been widely applied but suffer from limitations such as noise sensitivity, which cannot fully capture spatial relationships and leads to challenges in handling long and complex sequences [6]. The motivation is that traditional time-series prediction models consider single nodes and consider spatial relationships among multiple nodes in the data, which results in low accuracy when predicting spatio-temporal sequences from multiple source data [7]. Particulate matter (PM2.5) has become a serious problem because of its harmful effects on human health and the surrounding environment. Moreover, PM2.5 material increases owing to traffic congestion to reduce these accurate predictions, and precautions need to be taken [8]. The state-of-the-art method includes time-series prediction methods like autoregressive integrated moving average (ARIMA) and Vector Autoregressive

* Corresponding author: shivaprasad@msrit.edu

(VAR), where it only considers a single node from multiple sources of data that capture non-temporal relations. However, this method utilizes historical data to predict new trends and provides low accuracy in the prediction of spatiotemporal sequences [9]. Subsequently, the modern model graphs long short-term memory with multiple attention heads, which captures the spatial and temporal pattern relationships related to air quality forecasting. Moreover, this model learns complex dependencies as well as correlations and enhances the accuracy, reliability of the data, and leads to overfitting and increases the complexity [10].

The three main contributions of this research are:

- The Graph Attention Network (GAT) is enhanced by k-nearest neighbor (k-NN) sparsification and Exponential Moving Average (EMA) updates. This enables lightweight, adaptive, and stable adjacency learning, making it scalable for real-time forecasting in large monitoring networks.
- A Dynamic Spatio-Temporal Graph Convolutional Network (DSTGCN) was designed by integrating GAT-based spatial aggregation, Temporal Convolutional Networks (TCN) for modelling long-range dependency, and a temporal attention mechanism to emphasize critical time intervals. This combination confirms the accurate capture of both spatial diffusion and temporal evolution of air pollutants.
- The framework incorporates visual feature extraction based on ResNet from Modern-Era Retrospective Analysis for Research and Applications (MERRA-2) remote sensing rasters but fuses them with DSTGCN outputs through a gated fusion mechanism. This adaptive multimodal integration leverages both the station-level dynamics and patterns of large-scale dispersion.

The organization of this paper is ordered as follows. Section 2 provides a literature review, section 3 defines the proposed methodology, section 4 presents the experimental results, and section 5 concludes the research.

2 Literature review

Xia et al. [7] demonstrated multi-station time series data and remote-sensing images for predicting multimodal deep learning air quality. Initially, data were collected from the Beijing dataset, which contained time-series data and remote sensing images. Furthermore, data were preprocessed using a previous time step and converted, cleaned, and normalized images into a Tagged Image File Format (TIFF) for efficient training. Subsequently, the data were fed to the DSTGCN and processed remote-sensing images to ResNet for feature extraction. Furthermore, these spatiotemporal and visual features were sent through a fully connected layer to make predictions for future air quality at many stations. Moreover, this model performs better on individual station prediction tasks and obtains an equally good prediction in advanced multi-station prediction. However, this model depends on limited data sources, such as time series and remote sensing images, and does not use other factors, such as traffic and industry.

Dalal et al. [11] presented an advanced swarm optimization method for optimizing prediction of air quality in smart cities. Initially, data were collected from an air quality monitoring dataset that contained five air pollutant time-series data. Further, data were preprocessed by consolidation, cleaning, and scaling to prepare the Long Short-Term Memory (LSTM) model. Furthermore, a hybrid model was formed where the LSTM model learns sequential relationships, and the Recurrent Neural Network (RNN) model extracts the feature set and dropout used to reduce overfitting. Subsequently, a curiosity-based motivation model was developed to reconstruct the LSTM-RNN network and enhance model prediction. Furthermore, to minimize the prediction error by optimizing weights, Particle Swarm Optimization (PSO) was implemented and further encoded, and SoftMax predictions were made. Moreover, this hybrid model effectively handled both long- and short-term patterns. However, this model relies on high-quality sensor data, so missing or incorrect sensor readings reduce the prediction accuracy.

Han et al. [12] demonstrated a hierarchical semisupervised graph neural network for forecasting of air quality. Initially, quality of air observation data collected from Beijing-Tianjin-Hebei dataset, which contains weather observations such as temperature and wind direction. Furthermore, various spatial and temporal features were constructed from multi-source urban data, and the current quality of air distribution was approximated for unmonitored regions. Later, long-range spatiotemporal dependencies were learned, and spatial sparse air quality was augmented based on historical readings using a hierarchical recurrent graph neural network. The self-supervised hierarchical graph neural network (SSH-GNN) using supervision tasks improves generalization and forecasting prediction. Moreover, the SSH-GNN obtains both the universal topology and context pattern to improve forecasting. However, there are limited monitoring stations that reduce the accuracy of real-time air quality predictions.

Jiang et al. [13] analyzed environmental inequities using machine learning and spatial networks. First, a street-level air quality dataset was collected using sensors that measured air pollution. Further, data were preprocessed by a road network graph where each street segment was a node and connection was an edge, and the data were cleaned, combined, and prepared for analysis. Subsequently, a graph-based light gradient boosting Machie (GBM) machine learning model was trained to estimate PM2.5. Later, after prediction, these values were aggregated within 15 minutes walking range to calculate the exposure levels that people face regularly in daily life. Moreover, this model uses high-resolution street pollution data, which provides a realistic picture of people's exposure. However, data were not collected from all seasonal months.

Martinez et al. [14] demonstrated a smart city framework for modelling urban traffic and air quality.

Initially, the European air quality dataset contained observations from sensors, monitoring stations, and forecasted data. Further, the data were preprocessed, and the traffic sensor data were cleaned using error detection methods and calibrated air quality using a regression model. Later, cleaned traffic was used in a static flow road network, and calibrated air quality along with traffic intensity and weather conditions were used in dynamic models to yield forecasts. The TAQE approach was evaluated in the TRAFAIR project to verify its capability and performance. Moreover, TAQE supports air quality and traffic data in the solo standardization model. However, supports only fixed traffic situations and does not handle remote sensing observations of traffic.

3 Methodology

This study introduces a graph attention-based Residual Graph Convolutional Network (GAT-ResGCN) for efficient and accurate forecasting of PM2.5. Initially, Air Quality Index (AQI) Indian time-series data and MERRA-2 remote-sensing rasters were integrated by providing both localized ground measurements and patterns of large-scale dispersion. Furthermore, the data are preprocessed using missing value imputation, z-score normalization, and temporal alignment, yielding inputs that are clean and synchronized. Subsequently, a GAT with k-NN and EMA updates is employed to generate a dynamic and stable adjacency matrix with minimal computational cost. Furthermore, spatiotemporal features are extracted using DSTGCN, which combines spatial learning based on GAT, TCN, and temporal attention. Simultaneously, ResNet processes satellite rasters to extract the features of visual dispersion. Finally, a mechanism of gated fusion adaptively integrates both modalities, producing robust PM2.5 predictions with improved scalability and accuracy. The workflow of proposed method is illustrated in Figure 1.

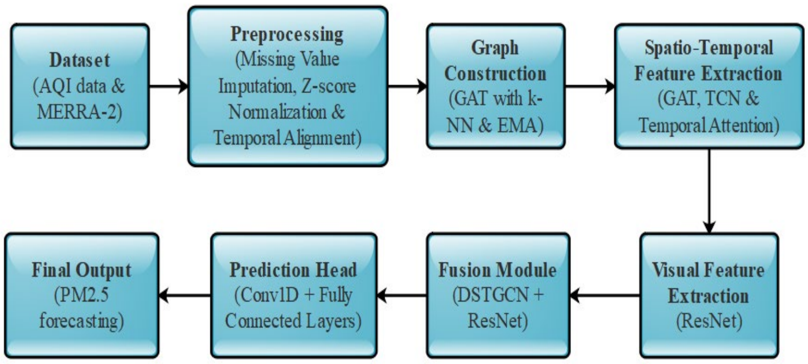

Fig. 1. Work flow of proposed GAT-ResGCN.

3.1 Data collection

To support accurate and scalable Particulate Matter (PM2.5) forecasting, the proposed framework integrates measurements from ground-based time series with satellite-derived remote sensing data. Moreover, for ground monitoring, Air Quality Index (AQI) Data for India (2015-2020) was used, which provides concentrations of hourly pollutants including PM2.5, PM10, NO_2, SO_2, CO, and O_3 along with meteorological variables. These meteorological variables included temperature, humidity, and wind speed across multiple urban monitoring stations. However, this dataset confirms rich spatiotemporal coverage and reflects diverse urban environments. The MERRA-2 reanalysis dataset was used for remote sensing. This dataset is made available by National Aeronautics and Space Administration (NASAs) global modelling and assimilation office [7]. Moreover, MERRA-2 provides gridded atmospheric data, including concentrations of surface-level PM2.5 and related aerosol components, at an hourly resolution. Moreover, these data were extracted and cropped to cover geographic regions that correspond to AQI stations, thereby serving as raster inputs that capture patterns of large-scale dispersions. By combining multi-station time-series with satellite-derived gridded images, the dataset allows the proposed model to simultaneously learn localized trends from observations that are station-level and have broad spatial patterns from atmospheric reanalysis, thereby confirming a comprehensive and reliable basis for dynamic graph construction and predictive modelling.

3.2 Data preprocessing

The preprocessing step s applied to prepare the raw sensor and remote-sensing data for effective model training. Moreover, it confirms the quality of the data by handling missing values, aligning different sources in time, and scaling features to comparable ranges. These operations improve consistency, reduce noise, and enhance the learning capabilities.

3.2.1 Missing value imputation

Air quality and meteorological datasets often contain gaps owing to sensor malfunctions, delays in communication, or maintenance downtime. To ensure continuity in the data, missing values re imputed before training the model. Moreover, for isolated missing entries, a forward-fill strategy (Last Observation Carried

Forward (LOCF)) was applied, as shown in Equation (1).

$$x_t = x_{t-1}, \quad if\, x_t\, is\ missing \tag{1}$$

For short gaps bounded by valid observations, linear interpolation was used to estimate intermediate values, as shown in Equation (2).

$$x_t = x_{t_0} + (x_{t_1} - x_{t_0}) \cdot \frac{t-t_0}{t_1-t_0}, t_0 < t < t_1 \tag{2}$$

Here, xt_0, and xt_1 are the values before and after the missing interval, respectively. Moreover, longer missing segments are either masked or excluded from training to prevent unreliable imputation. This approach preserves temporal consistency and confirms the complete input windows for the forecasting model.

3.2.2 Z-Score normalization

Because air quality and meteorological variables differ in scale and unit, normalization is important to prevent features of high magnitude from dominating the model training. Moreover, here z-score normalization, which is also called standardization, is used to transform every feature into a distribution with zero mean and variance. For a given value x, the normalized form is given by Equation (3):

$$x' = \frac{x-\mu}{\sigma} \tag{3}$$

Where μ signifies the mean and σ is standard deviation of that feature, computed only on the training set to avoid data leakage. Furthermore, the same parameters were applied to the validation and test sets to ensure uniform scaling across all features, accelerate model convergence, and improve the numerical stability during optimization. Moreover, by mapping features into a comparable range, z-score normalization allows the model to learn variations effectively, rather than being biased by absolute magnitudes.

3.2.3 Temporal alignment

The ground-based AQI data and satellite-derived MERRA-2 rasters differed in temporal resolution. To integrate both sources, a unified hourly timeline was established, and station data were kept at their native resolution while satellite rasters were resampled to match. Moreover, for intermediate hours, linear interpolation is applied, as in Equation (4).

$$R_t = R_{t_0} + (R_{t_1} - R_{t_0}) \cdot \frac{t-t_0}{t_1-t_0}, t_0 < t < t_1 \tag{4}$$

Where Rt_0, and Rt_1 are the consecutive valid raster values surrounding the missing timestamp. This ensures that at every hour t, both the station time series X_t and raster image R_t are temporally aligned. Through this synchronization, the multimodal model receives paired inputs by allowing the effective fusion of local station observations with broader patterns of spatial dispersion for accurate PM2.5. As a result, the preprocessing outputs cleaned, standardized, and temporally aligned multimodal datasets comprising continuous multi-station time-series matrices and corresponding gridded PM2.5. These refined inputs were based on robust fusion in the proposed GAT-ResGCN framework.

3.3 Graph construction with graph attention

After preprocessing, continuous, normalized, and temporally aligned multi-station time-series along with synchronized remote-sensing rasters were obtained. Moreover, each monitoring station's preprocessed data are represented as a node embedding that summarizes its pollutant and meteorological conditions within the current time window. Furthermore, a dynamic graph is constructed to capture the evolving interdependencies among these stations. Moreover, a Graph Attention Network (GAT) is employed to build an adaptive and lightweight adjacency matrix directly from node embeddings. For each pair of nodes i and j, attention score was calculated utilizing Equation (5).

$$e_{ij} = LeakyReLU(a^\top [Wh_i \| Wh_j]) \tag{5}$$

Where W represents a matrix of learnable transformations and a signifies the attention vector. Furthermore, the normalized attention coefficient is computed as below Equation (6):

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k\in\mathcal{N}_i} \exp(e_{ik})} \tag{6}$$

Where represents the adaptive edge weights between stations. To avoid fully connected graphs and reduce computational overhead, k-nearest neighbor (k-NN) sparsification retains only the most significant connections for each node. Furthermore, an Exponential Moving Average (EMA) is applied to smooth adjacency updates across consecutive windows, confirming that transient fluctuations do not destabilize the graph structure. Moreover, the resulting adjacency matrix is both dynamic and stable, reflecting current correlations among stations, while remaining efficient enough for forecasting in real time. This representation of the graph becomes the backbone input for DSTGCN, where spatial dependencies that are captured by the graph are combined with temporal modelling to fully exploit spatio-temporal patterns in air quality dynamics.

3.4 Spatio-temporal feature extraction using DSTGCN

Once the dynamic graph is constructed, the next step includes capturing both spatial correlations between stations and temporal evolution of pollutant levels. This is achieved using DSTGCN [15] , which integrates graph attention-based aggregation with temporal sequence modelling.

3.4.1 Spatial aggregation

At each time step, node embeddings are updated using graph attention convolution, as shown in Equation (7).

$$H^{(l+1)} = \sigma(\sum_{j\in\mathcal{N}_i} \alpha_{ij} W H_j^{(l)}) \tag{7}$$

Where $H_j^{(l)}$ represents feature vector of neighbor node j, W is the learnable weight matrix, α_{ij} is attention coefficient from the graph construction step, and $\sigma(\cdot)$ is nonlinear activation. Moreover, this ensures that each station adaptively incorporates information from its most influential neighbors.

3.4.2 Temporal modelling

The spatially updated features are then passed through a TCN to capture the sequential dependencies. A dilated causal convolution is applied, as illustrated in Equation (8):

$$y_t = \sum_{k=0}^{K-1} w_k \cdot x_{t-d\cdot k} \tag{8}$$

Here, d represents dilation factor, and K represents kernel size. Moreover, it expands receptive field by enabling method to capture long-range dependencies without reoccurrence.

3.4.3 Temporal attention

Finally, a temporal attention mechanism assigns importance to each time step by using Equation (9).

$$\beta_t = \frac{\exp(q_t)}{\sum_{k=1}^{T} \exp(q_k)}, H_{ST} = \sum_{t=1}^{T} \beta_t H_t \tag{9}$$

Where q_t is the learned relevance score. Moreover, through this combination of graph attention, dilated convolutions, and temporal attention, the DSTGCN produces spatio-temporal embedding H_{ST} that capture both inter-station pollutant diffusion and time-dependent patterns. Furthermore, these embeddings are forwarded to the next stage of visual feature extraction and fusion for integration with the remote-sensing information.

3.5 Visual feature extraction using ResNet

While the DSTGCN captures spatiotemporal dependencies from the ground monitoring station, it is equally important to model patterns of large-scale pollutant dispersion that are observable from satellite-derived rasters. For this purpose, the proposed framework integrates ResNet to extract high-level spatial features from MERRA-2 PM2.5.

3.5.1 Residual learning

Unlike traditional Convolutional Neural Networks (CNNs), ResNet demonstrates skip connections to address the problem of vanishing gradients in deep architectures. Moreover, a residual block is mathematically presented in Equation (10):

$$y = \mathcal{F}(x, W) + x \tag{10}$$

Where x denotes input block, $F(x, W)$ denotes the transformation learned through stacked convolutional, batch normalization, and layers of ReLU, and y signifies the block output. Moreover, the skip connection confirms that the network learns residual mappings more easily by improving the convergence and accuracy.

3.5.2 Feature extraction

The MERRA-2 rasters $R_t \in R^{H\times W}$ are passed through an initial convolution layer, followed by a sequence of residual blocks. Moreover, each block captures hierarchical spatial features ranging from coarse dispersion patterns to fine pollutant distributions. Global Average Pooling (GAP) is then applied to reduce the spatial dimensions, which is given in Equation (11).

$$h_V = \frac{1}{H\cdot W} \sum_{i=1}^{H} \sum_{j=1}^{W} f_{ij} \tag{11}$$

Where f_{ij} denotes feature activation at location (i, j). The resulting embedded H_V is a compact representation of the remote-sensing image and patterns of pollutant dispersion across the city. Then, the visual feature vector is forwarded to the fusion module, where it is combined with spatio-temporal embedding H_{ST} from DSTGCN to generate robust predictions of multiple models.

3.6 Fusion and prediction

After attaining spatio-temporal embedding H_{ST} from DSTGCN and visual embedding H_V from ResNet, the next step is to integrate both modalities for robust PM2.5 forecasting. To achieve this, a gated fusion mechanism s employed, allowing the model to adaptively balance contributions from time series and remote-sensing features. Moreover, the fusion gate is z is defined as shown in Equation (12).

$$z = \sigma(H_{ST} W_{ST} + H_V W_V + b) \tag{12}$$

Where W_{ST} and W_V indicate learnable weight matrices, b signifies the bias term, $\sigma(\cdot)$ denotes sigmoid activation. Furthermore, the fused fusion representation is computed using Equation (13).

$$H_F = z \odot H_{ST} + (1 - z) \odot H_V \tag{13}$$

This ensures that the model adaptively selects the most informative modality under different conditions such as heavy pollution events or meteorological shifts.

3.6.1 Prediction head

The fused features H_F are passed through a 1-D convolutional layer and Fully Connected (FC) layer to generate final prediction, as presented in Equation (14).

$$\hat{Y}_{t+1:t+\tau} = FC(ReLU(Conv1D(H_F))) \tag{14}$$

Where $\hat{Y}$ denotes the forecasted PM2.5 concentrations for all stations through next τ time steps. This fusion framework leverages both localized time-series dynamics and patterns of large-scale dispersion to produce accurate and reliable forecasts. The final fused output is then optimized during training using the Mean Squared Error (MSE) loss function.

4 Experimental results

Here, a lightweight framework combining Graph Attention Networks and Residual Graph Convolutional Networks (GAT-ResGCN) is presented for efficient prediction of air quality in urban areas by implementing the Python package TensorFlow. For this purpose, software and hardware supplies were used for experimental analysis. The software requirements include operating system windows 11 pro and hardware requirements such as Intel Core i5 processor with 16GB RAM and 3.63 GHz processing speed. To evaluate the proposed GAT-ResGCN, the root mean square error, mean absolute error and R^2 were utilized as the evaluation metrics, which is expressed in Equations (15)-(17):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{15}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{16}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i-\hat{y}_i)^2}{\sum_{i=1}^{N}(y_i-\bar{y})^2} \tag{17}$$

Here, y_i represents the ground truth PM2.5, $\hat{y}_i$ is the predicted value, $\bar{y}$ is the mean of the ground truth values, and N represents the total number of observations.

4.1 Performance analysis

The performance analysis of proposed GAT-ResGCN for multistage wheat yield prediction is performed using various traditional models, namely, self-supervised hierarchical graph neural network SSH-GNN and gradient boosting machine (GBM), with the help of evaluation metrics. The performance evaluation of proposed GAT-ResGCN is presented in Figure 2.

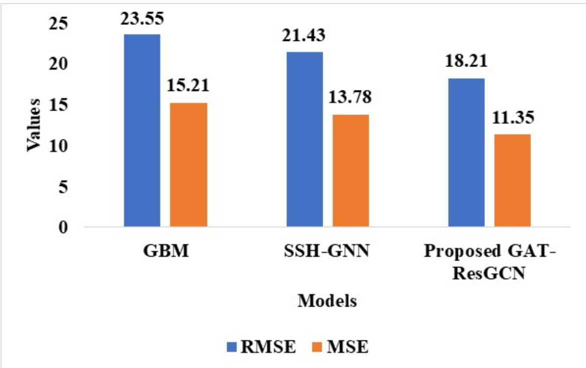

Fig. 2. Performance analysis of proposed GAT-ResGCN.

Figure 2 presents a performance analysis of the proposed GAT-ResGCN with various traditional models, including the GBM and SSH-GNN. The proposed GAT-ResGCN provided better results in metrics like root mean square error (RMSE) and Mean Absolute Error (MAE) of 18.21, 11.35.

4.2 Comparative analysis

The proposed GAT-ResGCN was compared with the existing Dynamic Spatio-Temporal Graph Convolution Network (DSTGCN) model [7]. The proposed GAT-ResGCN was compared with the evaluation metrics RMSE, MAE, and R^2 as shown in Table 1.

Table 1. Comparative analysis of proposed GAT-ResGCN.

Comparative models	RMSE	MAE	R^2
DSTGCN [7]	20.43	12.78	0.90
Proposed GAT-ResGCN	18.21	11.35	0.92

Table 1 presents a comparative evaluation of the proposed GAT-ResGCN with the existing MFA using the evaluation metrics. DSTGCN [7] provided results in metrics such as root mean square error (20.43), Mean Absolute Error (12.78), and R^2 (0.90). The outcomes show that proposed GAT-ResGCN outperformed the existing models by using an attention mechanism (GAT) for spatial feature extraction and residual connections (RES) to improve the performance and training stability, which is well suited for real-time forecasting at a low computational cost. The proposed GAT-ResGCN achieved better results in terms of root mean square error (18.21), Mean Absolute Error (11.35), and R^2 (0.92).

4.3 Discussion

In this research, a GAT-ResGCN is proposed for accurate and efficient PM2.5 forecasting by integrating multi-station time-series with remote-sensing data. The traditional DSTGCN [7] model uses Dynamic Time Warping (DTW) to build dynamic graphs that are effective in capturing temporal similarity. However, they introduce significant computational overhead and limited scalability, and they are less suitable for real-time applications in large networks of monitoring stations. Moreover, approaches based on DTW struggle to maintain stability in adjacency updates, often leading to fluctuating graph structures that reduce robustness. To address these limitations, the proposed framework replaces DTW with GAT, which is enhanced by k-NN sparcification and EMA updates, thereby confirming that spatial dependencies are learned adaptively, efficiently, and stably. Additionally, the integration of TCN with temporal attention improves the ability to capture both long-range dependencies and critical time intervals, which traditional DSTGCN lacks. Furthermore, by combining these improvements with ResNet-based visual feature extraction and gated fusion, the proposed model overcomes the computational and modelling challenges of base approach, thereby obtaining both higher accuracy and real-time practicality.

5 Conclusion

The primary aim of this research was to improve accurate and scalable framework for PM2.5 forecasting that integrates multi-station time-series data with remote-sensing information for urban air quality management in real time. To achieve this, a GAT-ResGCN was designed. In the data collection phase, the AQI India dataset delivered pollutants and meteorological series; however, MERRA-2 rasters obtained large-scale dispersion patterns. Preprocessing was performed for continuity, comparability, and synchronization through missing value imputation, Z-score normalization, and temporal alignment. In contrast to the base DTW method, graph architecture was introduced by integrating a GAT with k-NN sparsification and EMA, which makes adjacency learning more efficient and stable. Spatiotemporal dependencies were designed by a DSTGCN integrating GAT-based convolution, TCN used in long-range patterns, and temporal attention for focusing on critical intervals. Meanwhile, ResNet extracts dispersion features from rasters, which are combined with DSTGCN outputs by applying a gated fusion mechanism for final forecasts. The experimental results showed enhanced accuracy and efficiency, which confirmed the advantages of this hybrid approach. In future work, the extension of additional pollutants, enhancement of interpretability, and scaling of national-level deployments are planned to enhance the overall performances.

References

1. L. Chen, J. Wang, H. Wang, T. Jin, Urban air quality assessment by fusing spatial and temporal data from multiple study sources using refined estimation methods. ISPRS Int. J. Geo-Inf. **11**, 330 (2022). https://doi.org/10.3390/ijgi11060330
2. B. Qi, Y. Jiang, H. Wang, J. Jin, Multi-source PM2.5 prediction model based on fusion of graph attention networks and multiple time periods. IEEE Access **12**, 57603-57612 (2024). https://doi.org/10.1109/ACCESS.2024.3435678
3. X. Chen, Y. Hu, F. Dong, K. Chen, H. Xia, A multi-graph spatial-temporal attention network for air-quality prediction. Process Saf. Environ. Prot. **181**, 442-451 (2024). https://doi.org/10.1016/j.psep.2023.11.040
4. N.H. Van, P. Van Thanh, D.N. Tran, D.T. Tran, A new model of air quality prediction using lightweight machine learning. Int. J. Environ. Sci. Technol. **20**, 2983-2994 (2023). https://doi.org/10.1007/s13762-022-04185-w
5. H. Wang, L. Zhang, R. Wu, Y. Cen, Spatio-temporal fusion of meteorological factors for multi-site PM2.5 predictions: A deep learning and time-variant graph approach. Environ. Res. **239**, 117286 (2023). https://doi.org/10.1016/j.envres.2023.117286
6. M.A. Bhatti, Z. Song, U.A. Bhatti, M.S.M, AIoT-driven multi-source sensor emission monitoring and forecasting using multi-source sensor integration with reduced noise series decomposition. J. Cloud Comput. **13**, 65 (2024). https://doi.org/10.1186/s13677-024-00598-9
7. H. Xia, X. Chen, Z. Wang, X. Chen, F. Dong, A multi-modal deep-learning air quality prediction method based on multi-station time-series data and remote-sensing images: Case study of Beijing and Tianjin. Entropy **26**, 91 (2024). https://doi.org/10.3390/e26010091
8. S. Raj, J. Smith, E. Hayes, Hybrid graph convolutional LSTM model for spatio-temporal air quality transfer learning. Air Qual. Atmos. Health, 1-21 (2025). https://doi.org/10.1007/s11869-025-01713-8
9. Y. Huang, X. Zhu, R. Wang, Y. Xie, S. Fong, A dynamic global–local spatiotemporal graph framework for multi-city PM2.5 long-term forecasting. Remote Sens. **17**, 2750 (2025). https://doi.org/10.3390/rs17162750
10. N.N. Maltare, S. Vahora, Air quality index prediction using machine learning for Ahmedabad city. Digit. Chem. Eng. **7**, 100093 (2023). https://doi.org/10.1016/j.dche.2023.100093
11. S. Dalal, U.K. Lilhore, N. Faujdar, S. Samiya, V. Jaglan, R. Alroobaea, M. Shaheen, F. Ahmad, Optimising air quality prediction in smart cities with hybrid particle swarm optimization-long-short term memory-recurrent neural network model. IET Smart Cities **6**, 156-179 (2024). https://doi.org/10.1049/smc2.12080
12. J. Han, H. Liu, H. Xiong, J. Yang, Semi-supervised air quality forecasting via self-supervised hierarchical graph neural network. IEEE Trans.

Knowl. Data Eng. **35**, 5230-5243 (2022). https://doi.org/10.1109/TKDE.2022.3149815

13. F. Jiang, J. Ma, Environmental justice in the 15-minute city: Assessing air pollution exposure inequalities through machine learning and spatial network analysis. Smart Cities **8**, 53 (2025). https://doi.org/10.3390/smartcities8020053
14. D. Martínez, L. Po, R. Trillo-Lado, J.R. Viqueira, TAQE: A data modeling framework for traffic and air quality applications in smart cities, In International Conference on Conceptual Structures, Springer International Publishing, Cham, September 11 (2022), 25-40
15. B. Liu, Z. Qi, L. Gao, Enhanced air quality prediction through spatio-temporal feature sxtraction and fusion: a self-tuning hybrid approach with GCN and GRU. Water Air Soil Pollut. **23**, 532 (2024). https://doi.org/10.1007/s11270-024-07346-4