Water Air Soil Pollut (2025) 236:459
https://doi.org/10.1007/s11270-025-08103-x

Spatio-Temporal Forecasting using a Hybrid BiGRU-1DCNN Model for $PM_{2.5}$ Concentrations in Delhi, India (2018-2023) Across Multiple Monitoring Stations

Naushad Ahmad · Vipin Kumar

Received: 30 July 2024 / Accepted: 28 April 2025
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2025

Abstract Air quality deterioration, particularly the suspension of particulate matter over large urban areas, has emerged as a significant environmental concern. This issue, exacerbated by urbanization, industrialization, human activities, and climate change, poses serious health risks to populations. The present study proposes a hybrid BiGRU-1DCNN model to predict $PM_{2.5}$ levels in Delhi, India, by leveraging data from multiple monitoring stations. The proposed model incorporates Bidirectional Gated Recurrent Units (BiGRU) and a one-dimensional Convolutional Neural Network (1DCNN) to capture both temporal dependencies and spatial correlations in $PM_{2.5}$ data. The model's performance is evaluated through both single-station (SS) and spatio-temporal correlation (STC) approaches. Results demonstrate that the hybrid BiGRU-1DCNN model outperforms traditional deep learning models in both SS and STC scenarios. Specifically, it achieved a minimal Root Mean Square Error (RMSE) of 15.75, Mean Square Error (MSE) of 248.04, Mean Absolute Error (MAE) of 9.04, and Mean Absolute Percentage Error (MAPE) of 13.31 at the Jawaharlal Nehru Stadium (JNS) station. For comparison, the univariate SS model for the Major Dhyan Chandra National Stadium (MDCNS) station produced an RMSE of 17.31, MAE of 10.03, MAPE of 14.50, and MSE of 299.59. The non-parametric Friedman ranking further corroborated the superior performance of the hybrid BiGRU-1DCNN model, with it achieving the highest ranking across all performance metrics compared to other models. These results highlight the potential of the ST BiGRU-1DCNN model as a robust tool for air quality forecasting and public health risk mitigation in highly polluted urban environments like Delhi.

Keywords Time series forecasting · Earth air quality · Deep learning · $PM_{2.5}$ · Environmental pollutant

N. Ahmad · V. Kumar (✉)
Computer Science and Information Technology, Mahatma Gandhi Central University, Motihari 845401, Bihar, India
e-mail: rt.vipink@gmail.com

N. Ahmad
e-mail: naushad13bhu@gmail.com

1 Introduction

Air pollution, particularly fine particulate matter ($PM_{2.5}$), has become a global environmental and public health crisis. Rapid urbanization, industrialization, and increased human activity in major cities, compounded by climate change, have significantly deteriorated air quality. Air pollution is hazardous to human civilization and one of the most significant environmental issues, attracting global attention (Rincon et al., 2023). Most research indicates that the impacts of air pollution can differ among various age groups within the population. In recent years, poor air quality has resulted from the direct or indirect movement of numerous suspended chemicals across the globe due to urbanization, industrial activities, human behavior, and climate change (Liu et al., 2024). Fog and haze have become more

Published online: 13 May 2025

frequent over the past decade, primarily due to $PM_{2.5}$, which has an aerodynamic diameter of 2.5 microns. Urban congestion has worsened due to the booming economy and the increasing number of vehicles (Xu et al., 2023). Growing economic activity in emerging nations has led to a significant rise in air pollution emissions. $PM_{2.5}$, or tiny particulate matter, represents a substantial public health risk . Since $PM_{2.5}$ air pollution is viewed as a severe threat to human health, practical actions must be taken to mitigate it (Li et al., 2024) . Delhi, India, consistently ranked among the most polluted cities in the world with high $PM_{2.5}$ concentrations pose severe health risks, contributing to respiratory and cardiovascular diseases (Jain et al., 2025; Rotjanabumrung et al., 2023).

The "State of Global Air/2020" report, produced by the Institute for Health Metrics and Evaluation (IHME) and the Health Effects Institute, revealed that in 2019, long-term exposure to ambient $PM_{2.5}$ contributed to 4.1 million fatalities worldwide. Notably, Asia, Africa, and Europe (AAE) accounted for the majority of these premature deaths, totaling 96% of the global figure, with 3,148.3 thousand, 383.4 thousand, and 388.3 thousand fatalities, respectively. These findings highlight the urgent need for action to address air pollution and its harmful effects on global health (Haddaji et al., 2024). Cardiovascular disease (CVD) is the leading cause of death globally, with an astonishing 17.8 million deaths attributed to this condition, according to the Global Burden of Disease Study 2017 (Bodor et al., 2023). Each year, air pollution is responsible for over 4 million deaths worldwide due to chronic pulmonary and heart disorders, as reported by the World Health Organization. The Asia-Pacific region alone accounts for 2.3 million deaths, mainly due to high pollution levels and dense populations. Air pollution poses a significant threat to both the environment and the health of communities around the world (Gilik et al., 2022). Recent studies indicate that approximately 16% of global deaths can be linked to the inhalation of harmful airborne substances. Lower respiratory infections (LRIs) continue to be a major public health concern, as demonstrated by findings from the Global Burden of Disease Study (GBD) 2019, which reported that LRIs led to around 2.49 million deaths, making it the fourth most prevalent cause of death worldwide in 2019. Children are particularly vulnerable due to their higher exposure levels compared to adults (Khaslan et al., 2024).

Several epidemiological studies have linked elevated $PM_{2.5}$ exposure to increased rates of respiratory infections, asthma, chronic obstructive pulmonary disease (COPD), and cardiovascular complications. In Delhi, the situation is particularly concerning. According to a 2020 report by the Global Burden of Disease Study, air pollution contributed to over 54,000 premature deaths in Delhi alone. Moreover, a study by the Indian Council of Medical Research (ICMR) in collaboration with the Public Health Foundation of India (PHFI) noted that long-term exposure to $PM_{2.5}$ in Delhi (Agarwal et al., 2023) significantly increases the risk of ischemic heart disease and stroke. These statistics highlight the urgent need for accurate forecasting models that can help in mitigating exposure risks through timely interventions.

Despite advancements in predicting air pollution, current methods often struggle to accurately forecast $PM_{2.5}$ concentrations, especially in densely populated urban areas like Delhi (Jakhar et al., 2025; Rincon et al., 2023). Traditional models, including statistical approaches and simpler machine learning techniques, frequently overlook the complex spatiotemporal dynamics of air pollution, such as the interaction between temporal trends and spatial distribution. Additionally, these models often fail to effectively capture the geographical variations among different monitoring stations, which can significantly affect $PM_{2.5}$ levels. As a result, there is an increasing need for more sophisticated hybrid models that can integrate both temporal and spatial data to enhance prediction accuracy. Predicting pollutant levels in urban areas is inherently complex and is influenced by three intricate factors: 1) Temporal Dependencies: Current pollution levels are affected by both nearby and distant historical intervals (Wu et al., 2022). The concentration of pollutants at 10:00 a.m. impacts the concentration at 11:00 a.m. Both anthropogenic factors and meteorological conditions follow cyclic patterns, with pollutant time series typically recorded daily, repeating every 24 hours (Rodríguez et al., 2023). 2) Spatial Dependencies: Pollution levels at monitoring stations tend to correlate with those at nearby and even distant stations due to the initial law of geography (Wang et al., 2023; Yu et al., 2023) . 3) External Factors: Various influences, including other pollutants, weather conditions, terrain, and building heights, affect how pollutants are generated, dispersed, and broken down through physical pro-

cesses and reactions. While a few studies have applied machine learning and deep learning techniques to predict $PM_{2.5}$ in Delhi, most rely on single-station data or traditional methods. There is limited research exploring hybrid spatio-temporal models specifically tailored for Delhi's highly variable pollution landscape. Furthermore, comparative studies that benchmark multi-station versus single-station forecasting performance remain scarce.

Early studies applied statistical models such as ARIMA, Multiple Linear Regression (MLR), and Support Vector Regression (SVR). While these models are interpretable, they struggle to capture complex nonlinear and temporal dependencies inherent in air quality data. Deep learning models, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRUs), have shown improved performance over statistical approaches. These models excel in capturing temporal patterns but often lack mechanisms for modeling spatial dependencies. Accurate forecasting of $PM_{2.5}$ concentrations requires models that can effectively capture the complex interplay between temporal patterns and spatial heterogeneity in pollution data (Wang et al., 2024). Traditional approaches, such as purely statistical models or time-series models, often fall short when it comes to understanding the dynamic and interconnected nature of air pollution across both space and time (Wu et al., 2025). This has led to a growing interest in spatiotemporal modeling, which simultaneously accounts for variations in pollution across different locations and over time, thereby offering a more comprehensive and accurate representation of air quality trends (Badawy, 2025). To address these challenges, spatiotemporal modeling has emerged as a robust and innovative approach. Unlike conventional models that treat spatial and temporal dimensions independently, spatiotemporal models integrate both aspects, enabling a deeper understanding of how pollution propagates over time and across different geographical areas (Nakhjiri & Kakroodi, 2024). This approach is particularly beneficial in densely populated and industrialized areas with complex meteorological and topographical features (Elisephane & Ozunu, 2024). Recent studies have demonstrated the efficacy of deep learning architectures for spatiotemporal air pollution forecasting in Indian megacities (Ghosh & Dutta, 2023), reinforcing the need for robust hybrid models that can adapt to complex urban dynamics. Our model's performance surpasses recent ensemble-based approaches developed for urban India, such as the DNN-meteorological hybrid by Sharma et al. (2023), particularly in terms of RMSE and spatial generalization across multiple monitoring stations in Delhi.

To overcome the limitations of isolated temporal or spatial models, spatiotemporal modeling has emerged as a powerful approach that simultaneously captures both temporal dynamics and spatial dependencies across monitoring stations. In a densely populated and geographically diverse city like Delhi, $PM_{2.5}$ levels are not only influenced by historical trends at individual stations but also by pollutant dispersion patterns, traffic flow, and meteorological interactions from surrounding areas. Incorporating spatial correlations alongside temporal sequences enables models to better understand pollutant behavior, leading to more accurate and robust forecasts. In response to these challenges, this study proposes a novel spatiotemporal deep learning framework Hybrid BiGRU-1DCNN for forecasting $PM_{2.5}$ concentrations in Delhi. The model leverages the sequential learning strength of BiGRU and the feature extraction capability of 1DCNN, enabling it to capture both temporal dependencies and spatial correlations. The proposed model is evaluated using data collected from 28 monitoring stations across Delhi, integrating both single-station (SS) and spatiotemporal correlation (STC) strategies. Extensive experimentation demonstrates that the hybrid model consistently outperforms traditional deep learning models across multiple performance metrics.

Objectives of Study

The primary objective of this study is to develop and evaluate a novel hybrid deep learning model BiGRU-1DCNN for accurately forecasting $PM_{2.5}$ concentrations in Delhi, India, by incorporating both spatial and temporal data. This study aims to:

- Capture long-term temporal dependencies and spatial correlations in $PM_{2.5}$ data using a multi-station dataset.
- Compare the performance of the proposed model against conventional deep learning models such as GRU, LSTM, BiLSTM, RNN, and 1DCNN.

- Performance comparison of the proposed model using RMSE, MAE, MAPE, MSE and also validate the effective enhancement..
- Utilize non-parametric statistical analysis (Friedman test) to establish the significance of the proposed model's performance.
- Address current research gaps in $PM_{2.5}$ forecasting specific to the Delhi region.

The following summarises the primary contributions of this study:

Highlights of this Research

- A hybrid ST BiGRU-1DCNN model is proposed for spatio-temporal forecasting of $PM_{2.5}$ concentrations in Delhi, India.
- Experimental results show the proposed model outperforms traditional deep learning models (1DCNN, BiLSTM, GRU, RNN, LSTM) in terms of RMSE, MAE, MSE, and MAPE.
- The proposed model has achieved the best prediction among the models over the RMSE: 15.749 and MAE: 9.043.
- A Friedman non-parametric statistical test confirms the superior performance and robustness of the proposed ST BiGRU-1DCNN model over single-station models.

The organization of the paper is as follows. Section 2 contains the Section 2.1 subsection study area and data collection. The Section 2.2 of EDA contains the statistical analysis, decomposition and autocorrelation function analysis, and spatio-temporal correlation analysis. In Section 3, deep learning models contain the BiGRU, 1DCNN, LSTM, and RNN. The Section 4

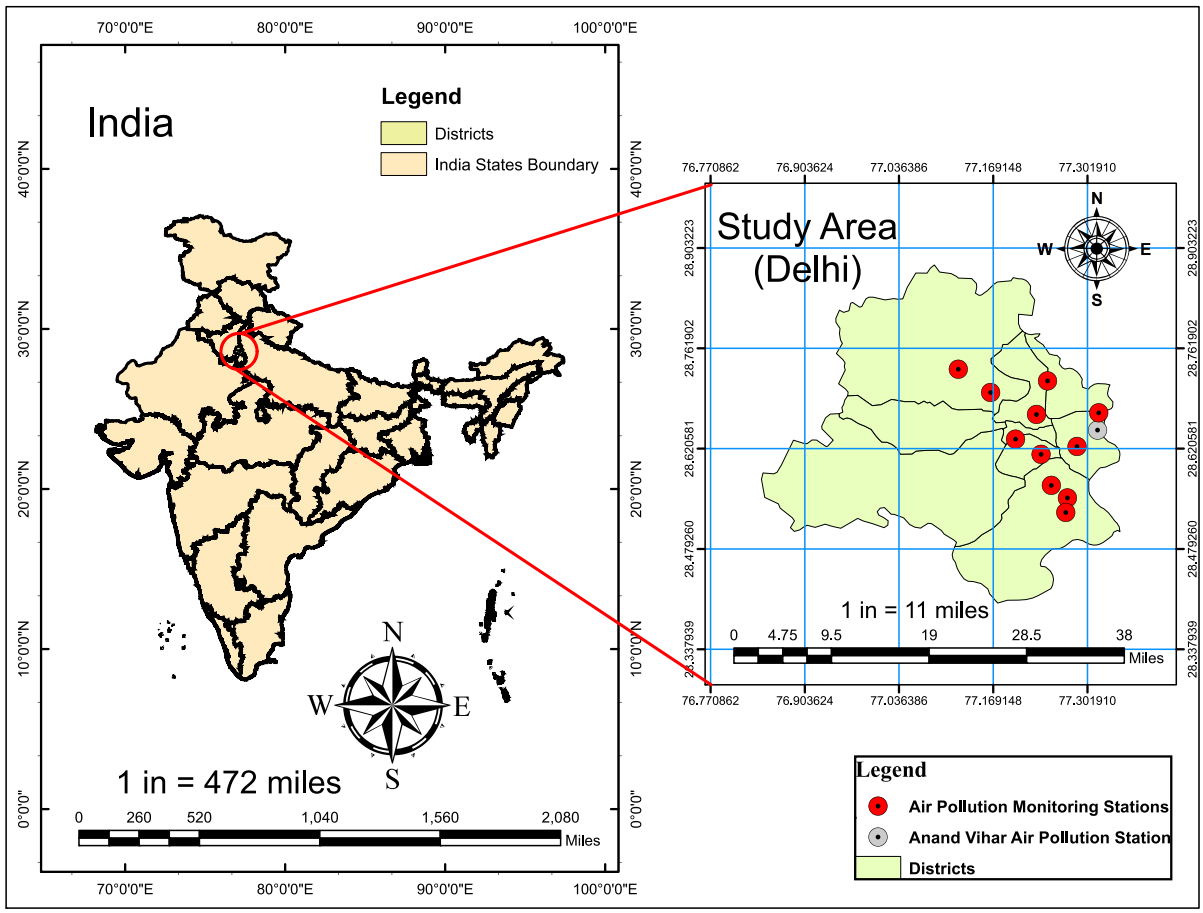

Fig. 1 The study area of selected geographical locations of air pollution monitoring stations Delhi, India

contains the proposed (ST BiGRU-1DCNN) model. In Section 5, the experimental setup contains the Sections 5.1 and 5.2, like hardware and software required and performance measures. The Section 6 section contains the Sections 6.1 quantitative analysis, Section 6.2 Graphical Analysis, Section 6.3 non-parametric statistical friedman ranking analysis, and Section 6.4 one year forecasting $PM_{2.5}$ concentration. The last Section 7 contains the conclusion.

2 Data

The dataset was chosen for its comprehensive spatial coverage of Delhi, allowing for an in-depth spatiotemporal analysis of air quality trends. The CPCB is a government agency responsible for monitoring and regulating air quality across India, and its dataset is considered one of the most reliable sources of air pollution data for the region.

2.1 Study Area and Data Collection

Delhi, the capital of India, spans a geographical area of approximately 1,484 km^2, encompassing 11 districts. As one of the most polluted cities globally, it experiences extreme air quality fluctuations due to vehicular emissions, industrial activities, biomass burning, and meteorological conditions. The study uses air quality data collected from the CPCB monitoring stations spread across the National Capital Territory of Delhi. In total, data from 12 active stations were used, covering diverse urban zones, including residential, commercial, traffic-heavy, and background locations (see Fig. 1). The spatial distribution of these stations allows for robust coverage of Delhi's varying pollution dynamics, which is crucial for effective spatio-temporal modeling.

Hourly data on $PM_{2.5}$ concentrations from 39 air quality monitoring stations in Delhi, India, from 1-Jan-2018 to 6-Aug-2023 were obtained from the Ministry of Environment, Forest and Climate Change. This data encompasses seasonal cycles, including winter, summer, monsoon, and post-monsoon periods. This period was selected due to its ability to capture a comprehensive range of $PM_{2.5}$ concentration variability associated with meteorological shifts, crop residue burning, and festive events of Diwali, significantly influencing Delhi's air quality.

Initially, data were collected from 39 CPCB monitoring stations in Delhi. To ensure data quality and completeness, authors applied strict filtering criteria: removal of stations with more than 10% missing or erroneous data over the study period, and exclusion of stations with irregular time series or inconsistent temporal resolution. After filtering, 28 stations were selected for analysis (see Fig. 2).

The authors employed mean imputation to tackle missing values in the datasets. Mean fill-in method was selected for its simplicity and computational efficiency,

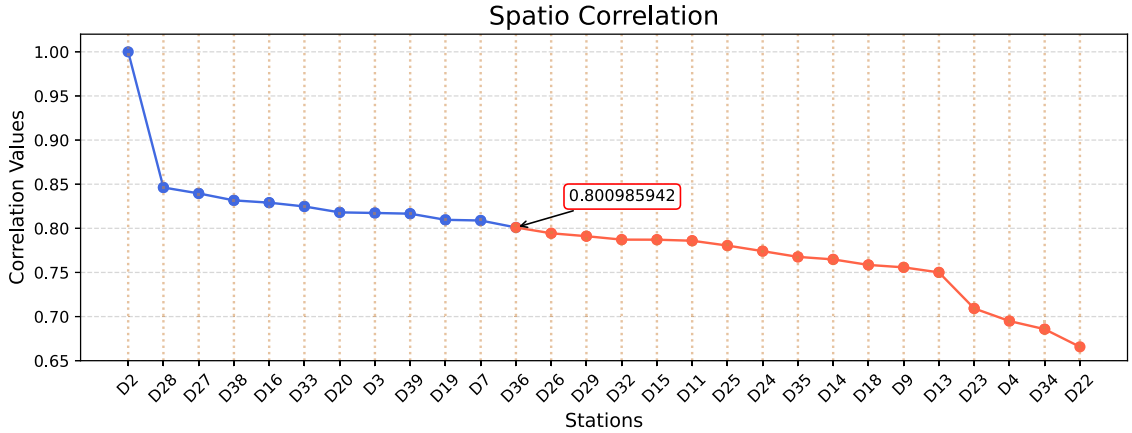

Fig. 2 The selection of station based on spatial correlation to Anand Vihar (D2) where $\alpha \geq 0.80$ correlated station in Delhi, India

which are essential for real-time forecasting (Mariappan et al., 2025). It maintains the temporal structure of the data and prevents distortions that may arise from using values from unrelated time intervals or stations, thereby ensuring the integrity of our analysis.

2.2 Exploratory Data Analysis (EDA)

The authors computed the Pearson correlation coefficients between $PM_{2.5}$ time series from all station pairs to determine spatial relationships among monitoring stations. The authors set a spatial correlation threshold of $\alpha \geq 0.80$ to identify strongly correlated stations (see Fig. 2) . This threshold aligns with standard practice in environmental data analysis, where correlations above 0.8 are typically considered strong and indicate significant similarity in pollutant trends (Draper et al., 2009; Her & Wong, 2020). In the dataset comprising 28 stations, a subset of 12 stations has been identified, which demonstrates a robust positive spatial correlation, with correlation coefficients ranging from 0.80 to 1. This signifies that these 12 stations possess highly interconnected data patterns, with values tending to decline. The high spatial correlation suggests strong interdependence in the data collected at these specific stations, which can benefit various applications and model designs involving spatial analysis and understanding of patterns in the given domain.

Delhi is ranked among the most polluted cities in the world, with Anand Vihar (D2) station being the most polluted city. The author has focused exclusively on Pearson's correlation coefficient values exceeding 0.80 about Anand Vihar $PM_{2.5}$. This study identified 12 stations with the strongest correlation to Anand Vihar $PM_{2.5}$. Table 1 presents statistical summaries for various data columns selected stations (D2, D3, D7, D16, D19, D20, D27, D28, D33, D36, D38, D39). It includes the count, mean, standard deviation, minimum, quartile values (25th, 50th, and 75th percentiles), count 49056, and maximum (max) values for each of these columns. These statistics offer insights into the distribution and central tendencies of the data within each column.

Stationarity Analysis Using ADF and KPSS Tests

To assess the stationarity of the $PM_{2.5}$ time series data, the authors applied the ADF and KPSS tests at each station Table 2. These complementary tests help determine whether the time series exhibits unit roots or trend stationarity, which informs the need for differencing or advanced modeling techniques.

- **ADF Test (H_0: non-stationary)**: For most stations, the ADF test yielded p-values < 0.05, indicating stationarity in the first differences.
- **KPSS Test (H_0: stationary)**: KPSS results for the same stations showed p-values < 0.05, confirming the presence of trend non-stationarity in the original series.

Table 1 The statistical summary of selected stations based on Pearson correlation value $\alpha \geq 0.80$ Delhi, India

Dataset Name	Station Name	Latitude (°N)	Longitude (°E)	Mean	Std Dev	Min	25%	Median	75%	Max
D2	Anand Vihar	28.646835	77.316032	129.31	112.31	0.2	53.25	100.75	158	985
D3	Ashok Vihar	28.66878	77.23001	108.35	104.11	0.25	38.75	76	135.12	989
D7	CPRI MR	28.551201	77.273574	99.94	91.34	0.03	38.48	73.31	128.20	995.98
D16	JNS	28.56893	77.25101	95.25	89.68	0.25	34.5	66.83	121.5	981
D19	MDCNS	28.61257	77.23649	94.11	82.80	0.25	37	68	122	985.75
D20	Mandir Marg	28.6341	77.2005	95.75	84.44	0.27	38	73.5	122.67	977
D27	Okhla Phase-2	28.530785	77.271255	98.57	94.19	1	35.25	67	126	989
D28	Parparganj	28.623748	77.287205	102.47	94.83	0.25	38.25	73	130	986
D33	Rohini	28.732528	77.11992	117.30	109.75	1	41	80.25	151.33	970
D36	Sonia Vihar	28.71601	77.24567	107.45	96.92	0.75	41	78.5	136.75	984
D38	Vivek Vihar	28.67115	77.31772	111.63	104.72	0.67	41	78.5	141	972
D39	Wazirpur	28.699793	77.165453	118.06	106.30	1	46.5	84	146.5	995

Table 2 Stationarity analysis: results of ADF and KPSS tests for assessing time series stationarity

Station	ADF[a]						KPSS[b]					
	ADF Statistic	p-value	1 %	5 %	10%	Stationary	KPSS Statistic	p-value	1 %	5 %	10%	Stationary
Anand Vihar	-12.25	9.56e-23	-3.43	-2.86	-2.56	Yes	0.67	0.01	0.73	0.46	0.34	Seasonal
Ashok Vihar	-10.84	1.53e-19	-3.43	-2.86	-2.56	Yes	0.56	0.02	0.73	0.46	0.34	Seasonal
CRRI Mathura Road	-11.49	4.66e-21	-3.43	-2.86	-2.56	Yes	1.12	0.01	0.73	0.46	0.34	Seasonal
Jawaharlal Nehru Stadium	-10.25	4.32e-18	-3.430	-2.86	-2.56	Yes	0.33	0.1	0.73	0.46	0.347	Trend
MDCNS	-10.50	1.06e-18	-3.43	-2.861	-2.56	Yes	0.24	0.1	0.73	0.46	0.34	Trend
Mandir Marg	-11.53	3.67e-21	-3.43	-2.86	-2.56	Yes	0.71	0.01	0.73	0.46	0.34	Seasonal
Okhla Phase-2	-10.61	5.59 e-19	-3.43	-2.86	-2.56	Yes	0.36	0.09	0.73	0.46	0.34	Trend
Parparganj	-10.76	2.50e-19	-3.43	-2.86	-2.56	Yes	0.39	0.07	0.73	0.46	0.34	Trend
Rohini	-10.68	3.92e-19	-3.43	-2.86	-2.56	Yes	0.54	0.03	0.73	0.46	0.34	Seasonal
Sonia Vihar	-11.06	4.70e-20	-3.43	-2.86	-2.56	Yes	0.31	0.1	0.73	0.46	0.34	Trend
Vivek Vihar	-11.28	1.43e-20	-3.43	-2.86	-2.56	Yes	0.29	0.1	0.73	0.46	0.34	Trend
Wazirpur	-11.09	3.96e-20	-3.43	-2.86	-2.56	Yes	0.671	0.016	0.73	0.46	0.34	Seasonal

Note: Both ADF and KPSS test results suggest that the pollution datasets from all listed stations are stationary

[a] ADF: All stations have very low p-values (close to zero), which are well below the typical significance levels of 1%, 5%, and 10%. This indicates strong evidence against the null hypothesis of a unit root

[b] KPSS: The KPSS statistics for most stations are below the critical values at the 1%, 5%, and 10% levels, indicating that fail to reject the null hypothesis of stationarity for these stations

2.2.1 *Decomposition and Autocorrelation Function Analysis*

Selected stations are highly correlated to Anand Vihar. To presenting the additive decomposition of the Anand Vihar station, which is highly polluted. In the realm of time series analysis, the data from the Anand Vihar station exhibits recognizable characteristics. Firstly, a trend indicates a steady and continuous value increase over a prolonged period. Additionally, a recurring seasonal pattern transpires every 24 units of time, signifying cyclic behavior. Notably, Correlation is still present at lag 100, suggesting long-range dependencies, implying a delayed relationship or impact within the data. The authors conducted a periodicity analysis using additive seasonal decomposition to assess the temporal characteristics of Anand Vihar $PM_{2.5}$ concentrations. These methods identify dominant cycles that are relevant for capturing long-term trends and short-term fluctuations. 24-hour cycles are likely driven by human activity patterns. The absence of structure in the residuals suggests suitability for predictive modeling. Additive decomposition to identify trend, seasonal, and residual components in the time series (see Fig. 3).

The autocorrelation function and partial autocorrelation function were used to assess temporal dependencies in the data (Zhang et al., 2025). The ACF revealed significant correlations at lags up to 100, indicating potential long-range dependencies, while the PACF confirmed these findings by showing a sharp drop after lag 100. The data exhibits both short-term and long-term temporal correlations, making it suitable for modeling with temporal dependencies. Created a plot (see Fig. 4) of the autocorrelation function and the partial autocorrelation function for a subset of 336 observations from the time series data of station Anand Vihar facilitates a more thorough examination of its temporal patterns. In ACF, the authors observe that the waveform and decrement of coefficient values show that the Anand Vihar station has seasonal and trend character-

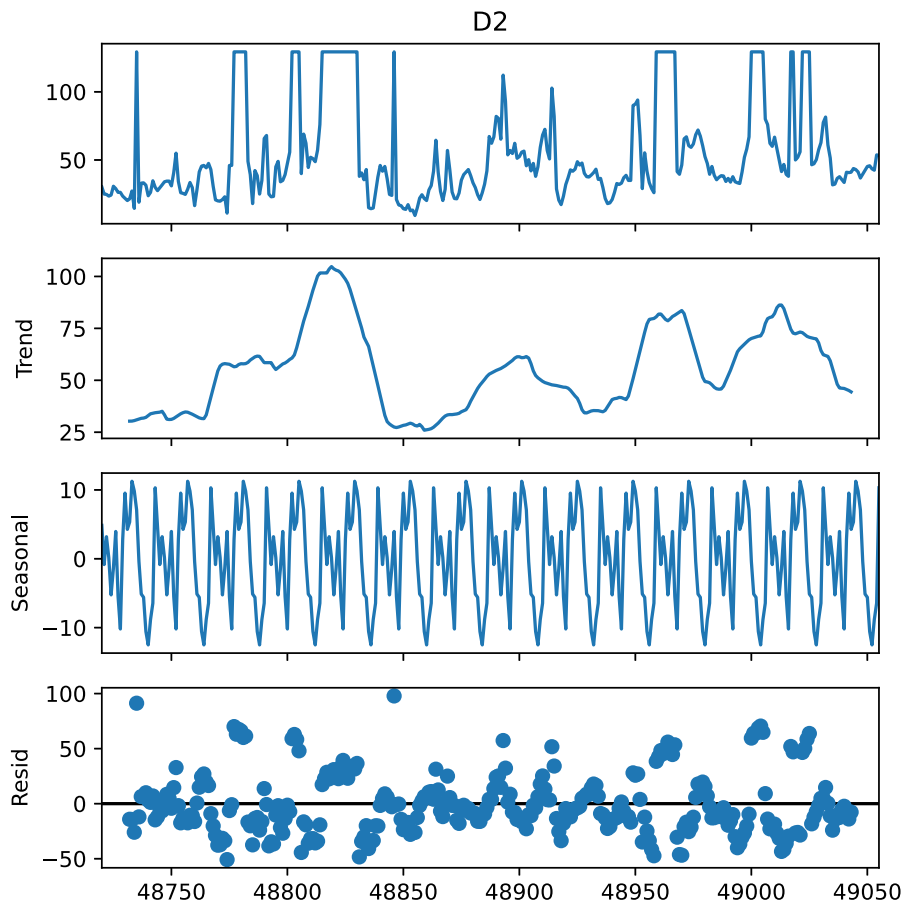

Fig. 3 Anand Vihar (D2) pollution analysis: additive decomposition assessing Delhi's pollution trend and seasonality

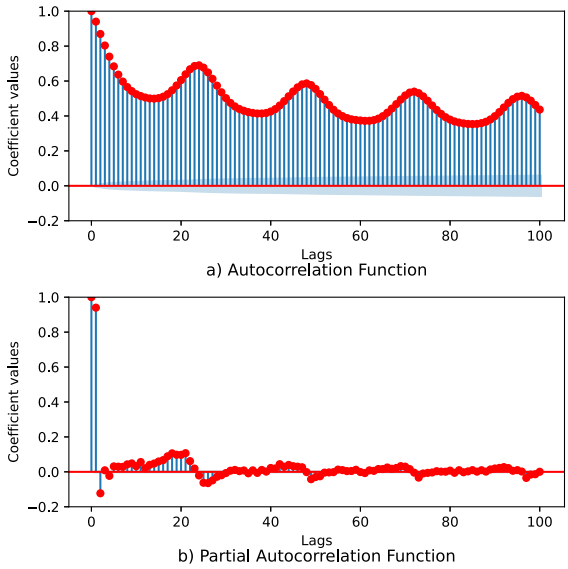

Fig. 4 Exploring Anand Vihar (D2) Station: a) ACF and b) PACF plots analyzing lag effects (0-100) for Delhi's pollution

istics. In PACF, observed the solid seasonal characteristics for lag1 and lag2 values.

2.2.2 Spatio-Temporal Correlation(STC) Analysis

Real-world events are often affected by spatial and temporal factors. Studying air pollution dynamics, sources, health impacts, and trends over time and space requires spatiotemporal analysis (Bahadur et al., 2023). Such analysis is necessary to comprehend their complex dynamics over time and across locations. A univariate model is insufficient for spatiotemporal analysis because it cannot capture complex relationships between variables (Yuan et al., 2022). Multivariate models offer a more comprehensive understanding of the spatiotemporal dynamics by accounting for the interplay between various factors of the same or different types (Sun et al., 2024). Depending on the characteristics and properties of the data at hand, it may become necessary to carry out the spatial and temporal aggregation process. Spatial aggregation consolidates individual data points into more significant regions or grids, facilitating a more comprehensive data analysis within a given spatial context. On the other hand, temporal aggregation entails the creation and establishment of time intervals or seasons, which aids in the organization and examination of data over specific periods. These techniques help identify underlying patterns in the data, authors and analysts can better understand the underlying patterns and trends in the data, leading to more informed decision-making and insightful conclusions. Before embarking on the process of conducting spatiotemporal correlation analysis, it is imperative to engage in a thorough evaluation of spatial autocorrelation. This pivotal preliminary step serves the purpose of comprehending any discernible spatial patterns within the dataset, which must be considered during subsequent analytical procedures.

When examining the temporal autocorrelation coefficients for all 12 selected stations across different time lags, a graphical representation (see Fig. 5) wherein the x-axis denotes the time lags and the y-axis signifies the autocorrelation coefficients. These methods identify

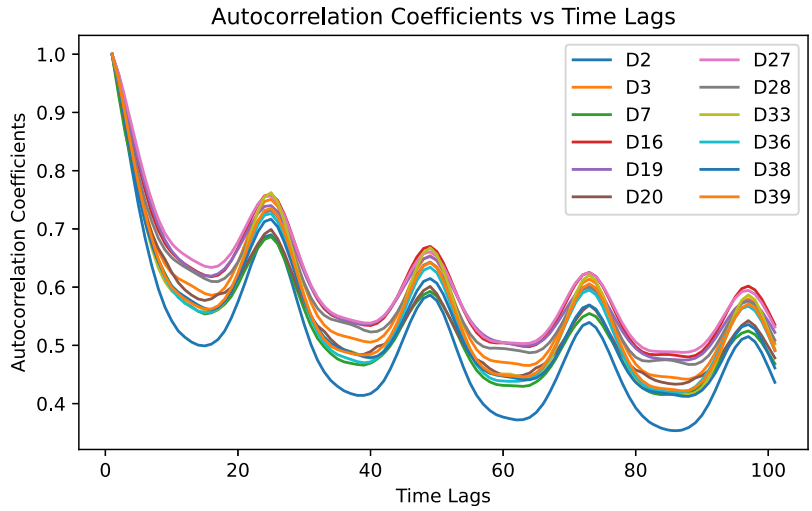

Fig. 5 Temporal autocorrelation coefficients of selected stations to analyze time-dependent relationships

dominant cycles that are relevant for capturing long-term trends and short-term fluctuations. 24-hour cycles are likely driven by human activity patterns. The autocorrelation coefficients quantify the extent to which the data values of each station correlate with their preceding values at various time lags. The autocorrelation analysis reveals both similarities and differences across stations. Stations located in areas with similar traffic patterns or industrial activities tend to exhibit stronger autocorrelation, especially 24-hour cycles. Vivek Vihar (D38) is the least autocorrelated to the other stations shown in (Fig. 5). This graphical representation will demonstrate how the autocorrelation changes as the time lag increases.

Typically, one would anticipate observing a decline in autocorrelation as the time lag increases, with coefficients diminishing and conceivably fluctuating around zero. This indicates how previous data points are related to future data points at different time intervals for the 12 stations.

3 Deep Learning Models

To fairly evaluate the performance of the proposed Hybrid BiGRU-1DCNN model, the authors implemented several traditional deep learning models as baseline architectures. These models were selected due to their widespread use in time-series air pollution forecasting and their ability to capture temporal dependencies. All models were trained under similar experimental conditions, including the same training-validation split, loss function MSE, optimizer Adam, and number of epochs. RNN: Implemented with two hidden layers, each comprising 64 units and `tanh` activation. A dropout layer with a rate of 0.2 was used to prevent overfitting. LSTM: Consists of two stacked LSTM layers with 64 units each and `relu` activation, followed by a dropout of 0.2. LSTM helps capture long-term dependencies in the $PM_{2.5}$ sequences. BiLSTM: Includes two bidirectional LSTM layers with 64 units, allowing both forward and backward temporal learning. All baseline models were implemented using TensorFlow/Keras and trained on the same input features as the proposed model to ensure consistency in comparison.

3.1 Bidirectional-gated Reccurent Unit

The gated recurrent unit is a type of recurrent neural network designed to capture sequential dependencies in data. It differs from traditional RNNs by using gating mechanisms to control the flow of information, helping to mitigate the vanishing gradient problem. GRUs are effective in tasks involving time-series data, such as $PM_{2.5}$ forecasting. The GRU (Dey & Salem, 2017), a simplified version of the RNN similar to the LSTM, incorporates gating mechanisms but with fewer parameters. As a result, it improves its efficacy in acquiring sequential patterns. The BiGRU, on the other hand, operates on input sequences in both the forward and backward directions. This characteristic enables the model to gather information from the preceding and subsequent contexts.

The equations for a BiGRU model with variable descriptions are as follows:

Forward GRU equations:

$$z_t^f = \sigma(W_z^f x_t + U_z^f h_{t-1}^f + b_z^f) \tag{1}$$

$$r_t^f = \sigma(W_r^f x_t + U_r^f h_{t-1}^f + b_r^f) \tag{2}$$

$$\tilde{h}_t^f = \tanh(W^f x_t + U^f (r_t^f \odot h_{t-1}^f) + b^f) \tag{3}$$

$$h_t^f = (1 - z_t^f) \odot h_{t-1}^f + z_t^f \odot \tilde{h}_t^f \tag{4}$$

Backward GRU equations:

$$z_t^b = \sigma(W_z^b x_t + U_z^b h_{t+1}^b + b_z^b) \tag{5}$$

$$r_t^b = \sigma(W_r^b x_t + U_r^b h_{t+1}^b + b_r^b) \tag{6}$$

$$\tilde{h}_t^b = \tanh(W^b x_t + U^b (r_t^b \odot h_{t+1}^b) + b^b) \tag{7}$$

$$h_t^b = (1 - z_t^b) \odot h_{t+1}^b + z_t^b \odot \tilde{h}_t^b \tag{8}$$

Combining both directions:

$$h_t = [h_t^f, h_t^b] \tag{9}$$

Here, x_t represents the input at time step t, h_t denotes the hidden state at time step t, W, U, and b are weight matrices and biases, σ represents the sigmoid function, $\odot$ denotes element-wise multiplication, z_t is the update gate, r_t is the reset gate, and $\tilde{h}_t$ is the intermediate state of the GRU at time step t.

3.2 1-D Convolutional Neural Network

CNNs are good at processing grid-like data (Wang et al., 2021). They can recognize patterns in temporal data, like predicting air pollution or analyzing sensor and biomedical data. CNNs use convolutional and pooling layers to extract hierarchical features and perform convolutions across the temporal dimension of the input sequence (Alzubaidi et al., 2021).

The 1-D Convolutional Neural Network in the equation:

$$Y_i = \sigma\left(\sum_{j=0}^{m-1} w_j \cdot h_{i+j} + b\right) \tag{10}$$

Pooling operation:

$$\text{Pool}_p(Y) = \max_{i=0}^{n/p} Y_{ip:(i+1)p} \tag{11}$$

Here, h_{i+j} represents the input signal at position i+j, Y_i denotes the output of the convolution operation at position i, w_j are the convolutional filter weights, b is the bias term, σ is the activation function LeakyReLU, $\text{Pool}_p(Y)$ represents the pooling operation with a pooling size p, n is the length of the input sequence, m is the size of the convolutional filter/kernel.

4 Proposed Hybrid BiGRU-1DCNN Model

The proposed hybrid model integrates (see architecture Fig. 6) the capabilities of bidirectional gated recurrent units and 1d convolutional neural networks to effectively capture both long-term temporal dependencies and local spatial-temporal features in $PM_{2.5}$ time series data.

The input sequence is first processed by a bidirectional GRU layer with 64 units, allowing the model to learn patterns in both forward and backward temporal directions. To mitigate overfitting, a dropout rate of 0.2 is applied. The output from the BiGRU layer is then passed through a 1d convolutional layer with 64 filters, a kernel size of 3, and ReLU activation to extract local temporal features. A global max pooling layer follows, reducing the feature map size while retaining the most salient information. This is then fed into a dense layer with 64 units and ReLU activation, culminating in a final dense layer with a single output unit for $PM_{2.5}$ concentration prediction. The model is trained using the mean squared error loss function and optimized with the Adam optimizer, with a learning rate set at 0.01. Training is conducted over 200 epochs with a batch size of 64, employing early stopping with the patience of 10 epochs to prevent overfitting. A validation split of 15% is used to monitor model generalization during training.

Let consider $y = \{y_1, y_2, ..., y_n\}^T = \left[\{Y_i\}_{t=1}^n\right]^T$ be a time series vector of n-random variables indes by t-time stamp which denotes the discrete time-space indexed data. In real life, time series data may have missing values, outliers, and non-normalized values, which affect the learning process of time series. Therefore, the preprocessing of time series data is highly required to achieve the required data quality. Let Prep() is the function that handles the missing values, outlier, and standardization of data and can be written in Eq. 12.

$$Prep\left[((y_i)_{t=1}^n)\right]^T \tag{12}$$

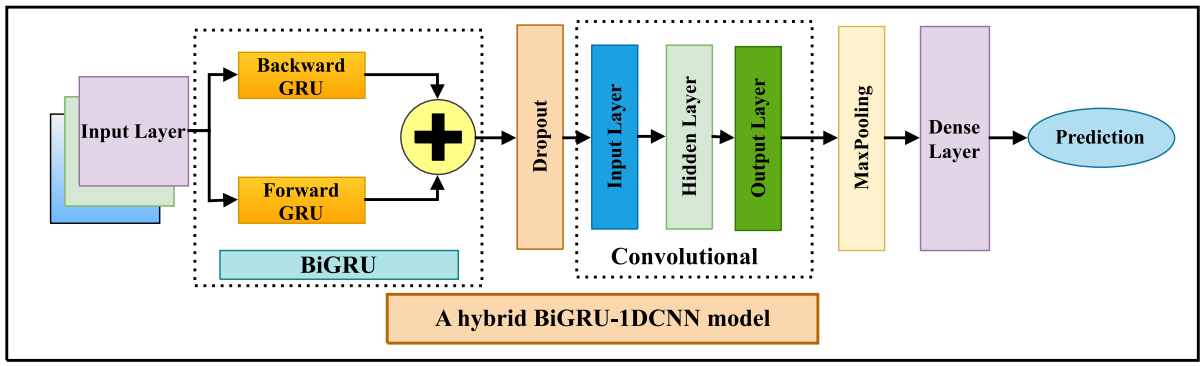

Fig. 6 Architecture of the Hybrid BiGRU-1DCNN model showing the flow of data through the BiGRU and 1DCNN layers

Now, preprocessed time series can be checked for stationarity, which may belong to weak or strict stationarity. The joint distribution is time-invariant for strict stationarity random variables. Then, further investigation of linear or non-linear temporal dependence can be done by autocorrelation. The autocorrelation ρ_i can be obtained as shown in Eq. 13. It measures the strength of linear dependence and direction between y_t and y_{t-j} for j^{th}-order autocorrelation.

$$\rho_i = \frac{Cov(y_t, y_{t-j})}{\sqrt{Var(y_t)Var(y_{t-j})}} \tag{13}$$

where Cov (y_t, y_{t-j}) is called j^{th} order autocovariance which measures the direction of linear dependence between y_t and the coefficient ρ_j value may be utilized to identify the seasional(Se) and trend(Tr) pattern in time series. The ρ_j values may be obtained differently for lag values, i.e., $\mathcal{L}$ =1,2,3,....,k. The series will have pattern, if $\exists\ \rho_j > 0.05\ \forall\ \mathcal{L}$ =1,2,3,...k.

Decomposition of y_i time series: A series sequence decomposition aims to extract the seasonality, trend, and remainder components. The seasonality component explains the regular fluctuation and periodicity for each repetition season, where the season may be denoted by |Se|. Now, the congruent modulo function F: T→ R ($\forall\ y_i \in Y$, i=1,2,3,...) Maybe formalized as:

$$f(t) = mod|Se|, t \in T \tag{14}$$

The optimization problem aims to minimize the error in the model's predictions by adjusting the weights of the hybrid BiGRU-1DCNN architecture. The objective is to minimize the loss function, which quantifies the discrepancy between the predicted $PM_{2.5}$ concentrations and the actual observed values.

So, the decomposition may be written as an optimization process for depreciation of the remainder (X^{Re}) as:

$$min \sum_{t \in T} (X_t^{Re})^2 \tag{15}$$

Subject to:

$$X_t = \left(X_t^{Se}\right) + \left(X_t^{Tr}\right) + \left(X_t^{Re}\right), \quad t \in T \tag{16}$$

$$\sum_{r \in |Se|} X_r^{Se} = 0 \tag{17}$$

$$X_t^{Tr} - \rho \le X_{t+1}^{Tr} \le X_t^{Tr} + \rho, t \in T \tag{18}$$

where X_t is the observed value, X_t^{Se} is the seasonal component, X_t^{Tr} is the trend X_t^{Re} is the residual component at time t and ρ is the smoothness parameter controlling the maximum change allowed between consecutive trend values.

The objective function is the sum of the squared remainder that converse at zero see Eq. 15 The constraint in Eq. 16 adopts the additive decomposition model. The constraint in see Eq. 17 ensures that the seasonality does not reduce boost over the seasons. The last constraint in see Eq. 18 ensures the smooth trend. Apply a Bidirectional Gated Recurrent Unit (BiGRU) to each station's input data see Eqs. 1 to 8:

$$z_t^f = \sigma(W_z^f x_t + U_z^f h_{t-1}^f + b_z^f) \tag{19}$$

the output of BiGRU store in h_t

$$h_t = [h_t^f, h_t^b] \tag{20}$$

Apply a 1D Convolutional Neural Network (1DCNN) to the BiGRU output:

$$Y_i = \sigma\left(\sum_{j=0}^{m-1} w_j \cdot h_{i+j} + b\right) \tag{21}$$

$$\text{Pool}_p(Y) = \max_{i=0}^{n/p} Y_{ip:(i+1)p} \tag{22}$$

Combine the outputs (Y_i)) with fully connected layers for further processing, depending on specific model design. Define a suitable loss function for regression tasks, such as MSE:

$$L = \frac{1}{2N} \sum_{i=1}^{N} (Y_i - P_i)^2 \tag{23}$$

Train the model to minimize the loss L using the training dataset.

$$\hat{Y}_{L_{1,n}}\left(Y_{n+g}\right) = \sum_{j=1}^{n} \varphi_{n,j}(g) Y_{n-j+1} \tag{24}$$

$\hat{Y}_L$ is the final prediction for our proposed (ST BiGRU-1DCNN) model. Where $\hat{Y}$ is the predicted values and n is the data points.

$$\hat{Y}_{n+g} = \sum_{j=1}^{n} \varphi_{n,j}(g) Y_{n-j+1} \tag{25}$$

where the parameter g represents the forthcoming period.

$$\mathcal{E} = \mathcal{E}(Y, \hat{Y}, \mathcal{E}_m) \tag{26}$$

where $\mathcal{E}_m$ are Error measures like RMSE, MAE, MSE, and MAPE shown in Eqs. 27 to 30 (Table 3).

5 Experimental Setup

The flowchart (see Fig. 7) of the proposed work commences by accessing the dataset provided by the CPCB, with a specific focus on $PM_{2.5}$ levels. Step 1 in the flowchart deals with initial data preprocessing tasks, such as handling missing values and accounting for temporal and spatial correlations. Step 2 involves splitting the data into training, validation, and testing subsets. Step 3 is about selecting and comparing various models based on performance metrics to choose the best-performing one. If the losses remain constant, the model is deployed. Hyperparameters were randomly tuned through multiple iterative experiments, balancing performance and training efficiency. Step 4 The final stage entails the evaluation of the model's performance using error evaluation metrics such as RMSE, MAE, MAPE, and MSE. Finally, the process concludes, and the marks the end of the process.

5.1 Hardware and Software Required

The BiGRU-1DCNN model is implemented using the keras and tensorflow frameworks within the jupyter notebook development environment, chosen for their versatility and ease of model development. The model architecture employs the LeakyReLU activation function with an alpha value of 0.7, enhancing the network's ability to learn from a wider range of activations. It utilizes a GRU layer with 100 units to capture sequential dependencies and a 1d-conv layer with 128 filters to extract local temporal features. The model is trained with a batch size of 16 using the nadam optimizer, set with a learning rate of 0.01 to balance convergence speed and stability. The experiments were conducted on a microsoft windows 10 pro system equipped with an Intel core i5-9500 CPU running at 3.0 GHz, featuring six physical cores and six logical processors, and supported by 16 GB of RAM. This configuration provided a capable yet accessible computational environment for model development and evaluation.

5.2 Performance Measures

To assess prediction accuracy, this study employs four commonly used statistical metrics: RMSE, MAE, MAPE, and MSE. Each serves a distinct evaluative purpose. The root mean square error measures the standard deviation of prediction errors, providing an inter-

Table 3 The proposed hybrid approach to utilization of Spatio-temporal BiGRU-1DCNN (ST BiGRU-1DCNN) model to process the sequence

Serial No.	Proposed model Spatio-Temporal BiGRU-1DCNN process sequence
1	Air pollution data set Preparing $\left[((y_i)_{t=1}^{n})\right]^T$ from CBCB, Delhi, India
2	Spatio (Fig. 2), temporal (Fig. 5) correlation analysis of all stations near Anand Vihar (D2)
3	Dataset splitting into training(70%) testing (15%) and validation(15%)
4	Combining forward h_t^f and backward h_t^b direction GRU $h_t = [h_t^f, h_t^b]$ see Eqs. 1 to 8
5	Single dimension convolutional neural network $Y_i = \sigma\left(\sum_{j=0}^{m-1} w_j \cdot h_{i+j} + b\right)$ see Eqs. 10 to 11
6	Minimize the loss L using the training dataset $L = \frac{1}{2N} \sum_{i=1}^{N} (Y_i - P_i)^2$
7	$\hat{Y}_{n+g} = \sum_{j=1}^{n} \varphi_{n,j}(g) Y_{n-j+1}$ where $\hat{Y}$ is the predicted values, n and g is the actual, forecasting data points.
8	$\mathcal{E} = \mathcal{E}(Y, \hat{Y}, \mathcal{E}_m)$ Error measures like RMSE, MAE, MSE, and MAPE shown in Eqs. 27 to 30.

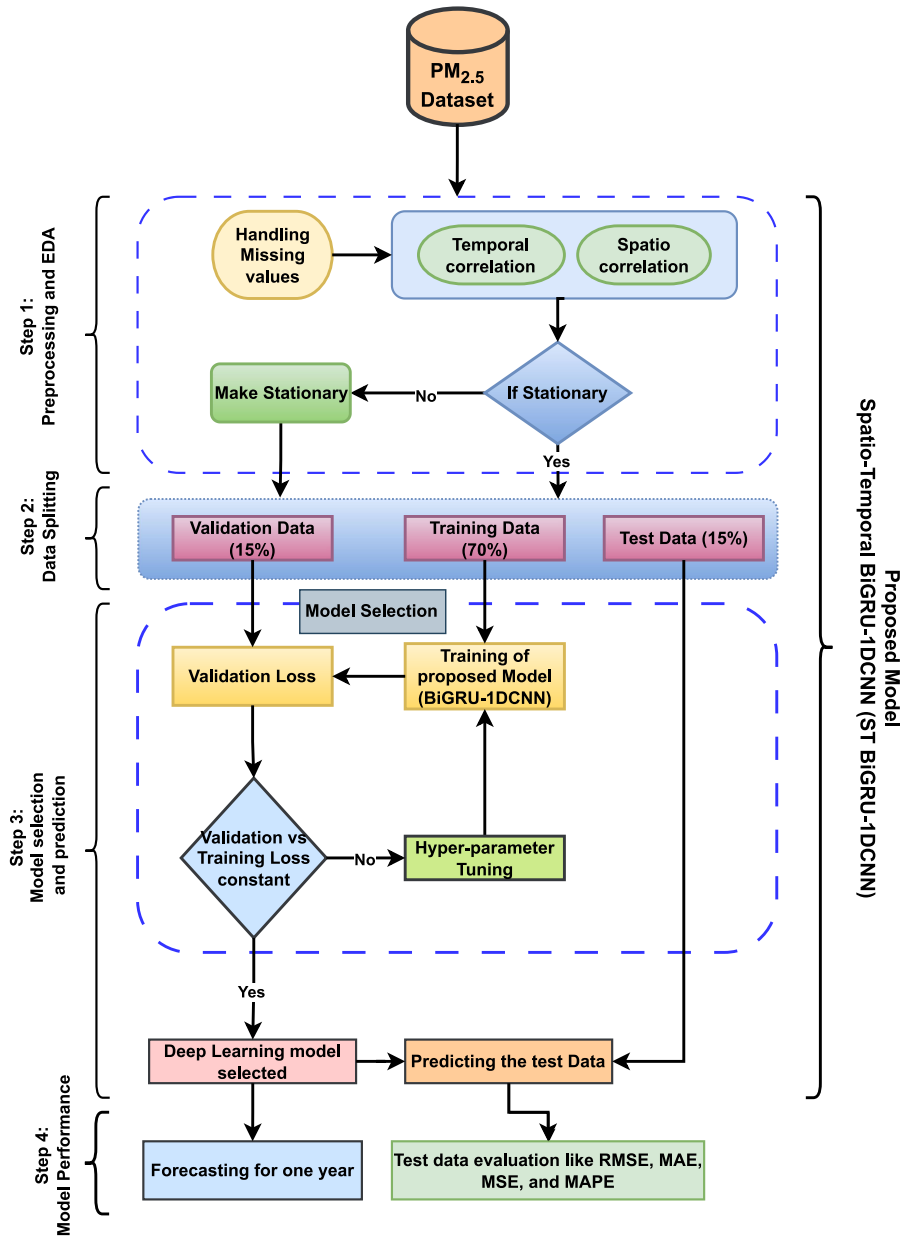

Fig. 7 The workflow of proposed Spatio-Temporal BiGRU-1DCNN (ST BiGRU-1DCNN) model for the $PM_{2.5}$ forecasting

pretable value in the same unit as $PM_{2.5}$ concentrations. The MAE offers an average of the absolute differences between predictions and actual values, yielding a straightforward understanding of model deviation. The MAPE contextualizes this deviation relative to observed values, which is particularly helpful for interpretability across varying pollution levels. While both RMSE and MSE assess prediction accuracy, whereas MSE emphasizes larger errors due to squaring, making it more sensitive to outliers. Employing this combination ensures a balanced and comprehensive evaluation of the proposed and baseline models. Calculate the following performance metrics for each station i to assess prediction accuracy:

The RMSE measures the predicted values ($\hat{Y}_i$) and actual values (P_i) for each station i:

$$\mathrm{RMSE}_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_i[t] - P_i[t])^2} \tag{27}$$

where T is the number of time steps. The MAE measures the predicted values ($\hat{Y}_i$) and actual values (P_i) for each station i:

$$\text{MAE}_i = \frac{1}{T}\sum_{t=1}^{T}|\hat{Y}_i[t] - P_i[t]| \tag{28}$$

The MAPE measures the predicted values ($\hat{Y}_i$) and actual values (P_i) for each station i. To avoid division by zero, a small constant c is added in the denominator:

$$\text{MAPE}_i = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{\hat{Y}_i[t] - P_i[t]}{P_i[t] + c}\right| \times 100\% \tag{29}$$

where c is a small positive constant. The MSE measures the predicted values ($\hat{Y}_i$) and actual values (P_i) for each station i:

$$\text{MSE}_i = \frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_i[t] - P_i[t])^2 \tag{30}$$

6 Results and Analysis

The proposed model across 12 selected stations uses four key metrics, RMSE, MAE, MAPE, and MSE, to assess predictive accuracy and model quality. The spanning of the testing data from October 4, 2022, at 17:00 to August 6, 2023, at 23:00, TableTable offers a detailed evaluation of the model's performance in forecasting $PM_{2.5}$ concentration.

6.1 Quantitative Analysis

Comparison of model performance metrics across different deep learning architectures for single-station and spatio-temporal correlation models. Imp% denotes the percentage improvement in performance by the proposed model over the baseline models for each metric.

6.1.1 Single Station (SS) Based Analysis

The (Table 4) presents the performance measures of the SS Proposed Model across 12 different selected datasets, utilizing four key metrics: RMSE, MAE, MAPE, and MSE to assess predictive accuracy and model quality. Among these datasets, MDCNS stands out as it consistently yields the lowest errors for all four metrics, boasting an RMSE of 17.308647, MAE of 10.029232, MAPE of 14.502798, and MSE of 299.58926, reflecting the highest accuracy in the model's predictions among the examined datasets.

- **RMSE**: The prediction performance of the proposed model got eight stations at the lowest RMSE compared to all 12 selected stations. The BiLSTM got the lowest RMSE for three stations, and the LSTM got the lowest for CPRI MR. The lowest RMSE is 17.308647 for station MDCNS. All the lowest RMSEs are in bold font compared to all baseline deep learning models and the proposed model. The smallest percentage improvement compared to the proposed model to the baseline DL model is BiLSTM for Ashok Vihar, CPRI MR, and JNS stations and LSTM for CPRI MR; otherwise, the proposed model outperforms the overall baseline DL model.
- **MAE**: The utilization of the baseline DL model and the proposed model: out of 12 stations, the proposed model got 10 stations with the lowest MAE score. The BiLSTM and LSTM have the lowest MAE values at Ashok Vihar and CPRI MR. The smallest MAE value is 10.029232 for the MDCNS station in the proposed model. The minor percentage improvement to Ashok Vihar and CPRI MR stations is by applying the BiLSTM, GRU, and LSTM; otherwise, the proposed model outperforms the overall baseline DL model in percentage improvement.
- **MAPE**: The measurement of the performance of MAPE over various stations is compared to the baseline and proposed model. The proposed model got nine stations with the lowest MAPE values out of all selected stations. The baseline LSTM model got the lowest MAPE values with Ashok Vihar, CPRI MR, and Rohini stations. The proposed model has achieved the lowest MAPE value of 14.502798 at the MDCNS station. The analysis of percentage improvement showed the lowest improvement in Ashok Vihar, CPRI MR, and Rohini over various baseline models.
- **MSE**: The performance measure of MSE over various baseline DL models and the proposed model for all selected stations. The proposed model got the lowest MSE values at eight stations of all 12

Table 4 Performance measures of selected stations and percentage improvements over the proposed model

Error	Station	BiLSTM	Imp%	CNN	Imp%	GRU	Imp%	LSTM	Imp%	RNN	Imp%	Proposed
RMSE	Anand Vihar	36.32	**11%**	35.19	**7%**	36.51	**11%**	33.28	**1%**	110.58	**237%**	**32.80**
	Ashok Vihar	**21.56**	-4%	23.014	**2%**	24.20	**7%**	22.69	**1%**	41.54	**84%**	22.56
	CPRI MR	29.45	-2%	30.61	**2%**	31.38	**4%**	**29.328**	-3%	167.33	**455%**	30.14
	JNS	**18.07**	-1%	19.80	**8%**	22.60	**24%**	19.55	**7%**	151.37	**728%**	18.27
	MDCNS	22.26	**29%**	19.21	**11%**	20.34	**18%**	18.30	**6%**	23.56	**36%**	**17.30**
	Mandir Marg	**20.82**	0%	21.97	**5%**	23.13	**11%**	85.72	**310%**	29.82	**43%**	20.89
	Okhla Phase-2	134.89	**588%**	23.29	**19%**	24.00	**23%**	22.12	**13%**	40.38	**106%**	**19.59**
	Parparganj	27.18	**7%**	27.32	**8%**	30.28	**19%**	27.99	**10%**	61.52	**142%**	**25.38**
	Rohini	30.30	**26%**	25.63	**6%**	27.42	**14%**	25.33	**5%**	54.01	**124%**	**24.13**
	Sonia Vihar	27.49	**17%**	24.78	**6%**	27.17	**16%**	24.42	**4%**	178.12	**658%**	**23.48**
	Vivek Vihar	28.11	**4%**	28.33	**5%**	30.83	**14%**	27.77	**3%**	165.81	**514%**	**27.01**
	Wazirpur	27.79	**6%**	27.47	**5%**	29.66	**13%**	27.06	**3%**	36.76	**40%**	**26.26**
MAE	Anand Vihar	21.58	**11%**	20.87	**7%**	21.16	**9%**	19.65	**1%**	93.43	**380%**	**19.47**
	Ashok Vihar	**12.43**	-6%	13.44	**2%**	13.15	-1%	12.66	-4%	21.77	**64%**	13.24
	CPRI MR	16.09	0%	16.57	**2%**	16.94	**5%**	**15.37**	-5%	148.99	**821%**	16.17
	JNS	10.75	0%	11.89	**11%**	12.73	**19%**	11.03	**3%**	125.89	**1073%**	**10.73**
	MDCNS	12.19	**22%**	11.58	**16%**	11.59	**16%**	10.41	**4%**	12.41	**24%**	**10.02**
	Mandir Marg	13.64	**3%**	14.23	**7%**	14.71	**11%**	75.48	**469%**	18.53	**40%**	**13.27**
	Okhla Phase-2	102.36	**849%**	11.94	**11%**	12.29	**14%**	11.30	**5%**	22.55	**109%**	**10.78**
	Parparganj	13.70	**6%**	14.45	**12%**	15.13	**18%**	13.32	**3%**	36.06	**180%**	**12.88**
	Rohini	18.74	**30%**	15.70	**9%**	15.86	**10%**	14.49	**1%**	33.33	**131%**	**14.42**
	Sonia Vihar	15.85	**15%**	15.17	**10%**	15.31	**11%**	14.02	**1%**	154.01	**1015%**	**13.81**
	Vivek Vihar	16.16	**6%**	16.33	**7%**	16.58	**8%**	15.35	**0%**	131.50	**759%**	**15.31**
	Wazirpur	16.27	**5%**	16.47	**7%**	16.69	**8%**	15.73	**2%**	19.61	**27%**	**15.43**
MAPE	Anand Vihar	26.87	**22%**	23.49	**7%**	23.17	**6%**	22.08	**1%**	183.68	**737%**	**21.95**
	Ashok Vihar	22.12	-8%	23.10	-4%	22.70	-5%	**21.03**	-12%	29.74	**24%**	23.94
	CPRI MR	37.21	-5%	36.84	-5%	36.03	-8%	**35.50**	-9%	234.98	**503%**	38.98
	JNS	17.95	**2%**	20.367	**16%**	20.79	**18%**	19.10	**9%**	159.87	**810%**	**17.57**
	MDCNS	17.06	**18%**	16.58	**14%**	16.29	**12%**	15.36	**6%**	16.66	**15%**	**14.50**
	Mandir Marg	26.91	**8%**	27.88	**12%**	27.47	**11%**	253.36	**921%**	36.67	**48%**	**24.81**
	Okhla Phase-2	100.01	**453%**	20.95	**16%**	20.91	**16%**	18.75	**4%**	35.85	**98%**	**18.08**

Table 4 continued

Error	Station	BiLSTM	Imp%	CNN	Imp%	GRU	Imp%	LSTM	Imp%	RNN	Imp%	Proposed
	Parparganj	22.03	**20%**	21.50	**17%**	20.87	**14%**	19.03	**4%**	41.86	**129%**	**18.31**
	Rohini	19.26	**11%**	18.63	**8%**	18.34	**6%**	**17.17**	-1%	42.63	**146%**	17.30
	Sonia Vihar	19.611	**9%**	19.38	**8%**	19.24	**7%**	18.23	**2%**	173.55	**866%**	**17.96**
	Vivek Vihar	25.16	**15%**	23.41	**7%**	23.34	**7%**	21.79	0%	122.98	**464%**	**21.81**
	Wazirpur	20.05	**7%**	19.97	**7%**	20.66	**10%**	20.64	**10%**	21.00	**12%**	**18.73**
MSE	Anand Vihar	1319.79	**23%**	1238.74	**15%**	1333.39	**24%**	1107.62	**24%**	12228.81	**1036%**	**1076.04**
	Ashok Vihar	**465.22**	-9%	529.66	**4%**	585.90	**15%**	515.17	**15%**	1726.15	**239%**	509.29
	CPRI MR	867.42	-5%	937.55	**3%**	984.74	**8%**	**860.16**	**8%**	28001.31	**2982%**	908.55
	JNS	**326.75**	-2%	392.16	**17%**	510.81	**53%**	382.53	**53%**	22915.52	**6758%**	334.12
	MDCNS	495.80	**65%**	369.31	**23%**	413.77	**38%**	334.92	**38%**	555.24	**85%**	**299.58**
	Mandir Marg	**433.88**	-1%	483.07	**11%**	535.42	**23%**	7348.43	**23%**	889.33	**104%**	436.76
	Okhla Phase-2	18197.18	**4639%**	542.47	**41%**	576.43	**50%**	489.41	**50%**	1631.08	**325%**	**384.00**
	Parparganj	738.90	**15%**	746.62	**16%**	917.16	**42%**	783.53	**42%**	3785.90	**487%**	**644.60**
	Rohini	918.55	**58%**	657.39	**13%**	752.16	**29%**	641.86	**29%**	2917.34	**401%**	**582.52**
	Sonia Vihar	755.93	**37%**	614.53	**11%**	738.32	**34%**	596.50	**34%**	31727.52	**5652%**	**551.61**
	Vivek Vihar	790.22	**8%**	803.15	**10%**	950.95	**30%**	771.47	**30%**	27494.25	**3666%**	**730.02**
	Wazirpur	772.32	**12%**	754.86	**9%**	880.15	**28%**	732.74	**28%**	1351.77	**96%**	**690.01**

selected stations. The BiLSTM had the lowest MSE value at Ashok Vihar, Jawaharlal Nehru Stadium, and Mandir Marg stations, and the LSTM achieved CPRI MR stations. In addition, the minimum percentage improvement over various baseline DL models is the proposed model for all selected stations. The minimum percentage improvement stations in BiLSTM are Ashok Vihar, CPRI MR, JNS, and Mandir Marg.

As shown in (Table 4) for single station univariate utilization, the prediction performances of the proposed model are compared with five baseline DL models. The table contains the performance measures of all baseline DL models and the proposed model over all selected stations. In addition, the author used RMSE, MAE, MAPE, and MSE to evaluate the performance of our proposed model over various DL models. The table contains the percentage improvement of performance measures of baseline models over the proposed model. Based on the improvement values, the proposed model is superior to all baseline DL over various performance measures.

6.1.2 Spatio-Temporal Correlation (STC) Based Analysis

The (Table 5) summarizes the performance measures for various datasets with a target STC. Dataset Jawaharlal Nehru Stadium has the lowest error values, demonstrating superior predictive accuracy with an RMSE of 15.749208 and the lowest MSE of 248.03755. Likewise, dataset MDCNS achieves outstanding performance, with the lowest MAE of 9.042538 and a highly competitive MAPE of 13.305757. These findings highlight the reliability and accuracy of JNS and MDCNS in predictive modeling for STC, making them prime choices for this specific task.

- **RMSE**: The prediction performance of the proposed model got nine stations at the lowest RMSE compared to all 12 selected stations. The LSTM has three stations with the lowest RMSE apart. The lowest RMSE is 15.749208 for station JNS. All the lowest RMSEs are in bold font compared to all baseline deep learning models and the proposed model. The smallest percentage improvement compared to the proposed model to the baseline DL model is BiLSTM for CPRI MR stations and LSTM for Ashok Vihar and CPRI MR; otherwise, the proposed model outperforms the overall baseline DL model by a percentage.
- **MAE**: The utilization of the baseline DL model and the proposed model: out of 12 stations, the proposed model got nine stations with the lowest MAE score. The BiLSTM has the lowest MAE values at Ashok Vihar, and the LSTM has Ashok Vihar and Wazirpur. The smallest MAE value is 9.042538 for the MDCNS station in the proposed model. Applying the baseline model, there is a minor percentage improvement to Ashok Vihar, CPRI MR, and Wazirpur stations; otherwise, the proposed model outperforms the overall baseline DL model for all selected stations.
- **MAPE**: The measurement of the performance of MAPE over various stations is compared to the baseline and proposed model. The proposed model got 11 stations with the lowest MAPE values out of all selected stations. The baseline GRU model got the lowest MAPE values with the Ashok Vihar station. The proposed model has achieved the lowest MAPE value of 13.30576 at the Major Dhyan Chandra National Stadium station. In the analysis of percentage improvement, Ashok Vihar showed the lowest improvement over various baseline models.
- **MSE**: The performance measure of MSE over various baseline DL models and the proposed model for all selected stations. The proposed model got the lowest MSE values at ten stations of all 12 selected stations. The LSTM stations with the lowest MSE values are CPRI MR and Okhla Phase-2. In addition, the minimum percentage improvement over various baseline DL models is the proposed model for all selected stations. The minimum percentage improvement stations are CPRI MR to BiLSTM, Ashok Vihar, and CPRI MR to LSTM.

As shown in (Table 5) for spatio-temporal correlation utilization, the prediction performances of the proposed model are compared with five baseline DL models. The table contains the performance measures of all baseline DL models and the proposed models over all selected stations. In addition, the author used different evaluation parameters to evaluate the performance of the proposed model over various DL models. To reflect the highest accuracy in the model's predictions among the examined datasets. The table contains

Table 5 Performance measures and percentage improvements of selected dataset STC for the proposed model compared to traditional DL models (BiLSTM, CNN, GRU, LSTM, RNN) utilizing RMSE, MAE, MAPE, and MSE

Error	Station	BiLSTM	Imp%	CNN	Imp%	GRU	Imp%	LSTM	Imp%	RNN	Imp%	Proposed
RMSE	Anand Vihar	31.52	**3%**	32.40	**6%**	33.05	**8%**	31.38	**3%**	44.18	**45%**	**30.55**
	Ashok Vihar	23.72	**15%**	20.91	**1%**	21.85	**6%**	**20.36**	**-1%**	31.88	**54%**	20.66
	CPRI MR	28.69	-1%	28.85	0%	30.55	**6%**	**28.17**	**-2%**	36.50	**27%**	28.84
	JNS	19.09	**21%**	17.03	**8%**	17.49	**11%**	16.62	**6%**	68.36	**334%**	**15.74**
	MDCNS	18.86	**18%**	17.16	**7%**	17.15	**7%**	17.00	**6%**	28.165	**76%**	**15.99**
	Mandir Marg	19.33	**4%**	20.32	**9%**	19.50	**5%**	19.04	**3%**	26.17	**41%**	**18.56**
	Okhla Phase-2	20.75	**7%**	21.48	**10%**	20.10	**3%**	**19.46**	**0%**	27.93	**43%**	19.47
	Parparganj	30.43	**29%**	24.82	**5%**	26.80	**13%**	28.09	**19%**	168.60	**613%**	**23.63**
	Rohini	25.19	**10%**	24.18	**5%**	24.27	**6%**	26.07	**13%**	42.42	**85%**	**22.97**
	Sonia Vihar	23.04	**6%**	22.87	**5%**	24.42	**12%**	26.48	**22%**	31.92	**47%**	**21.78**
	Vivek Vihar	28.16	**12%**	26.79	**7%**	26.79	**10%**	26.43	**6%**	43.89	**75%**	**25.04**
	Wazirpur	43.89	**31%**	25.61	**4%**	26.37	**8%**	25.30	**3%**	32.68	**33%**	**24.52**
MAE	Anand Vihar	18.53	**2%**	19.28	**7%**	18.83	**4%**	18.36	**1%**	23.33	**29%**	**18.10**
	Ashok Vihar	12.27	-2%	12.26	-2%	11.76	-6%	**11.69**	-6%	15.68	**26%**	12.47
	CPRI MR	**15.53**	-4%	16.21	**1%**	16.92	**5%**	15.60	-3%	20.02	**24%**	16.11
	JNS	10.29	**9%**	10.46	**11%**	10.02	**6%**	9.74	**3%**	55.18	**484%**	**9.44**
	MDCNS	10.18	**13%**	10.32	**14%**	9.59	**6%**	9.53	**5%**	14.75	**63%**	**9.04**
	Mandir Marg	13.00	**8%**	13.27	**10%**	12.77	**6%**	12.61	**5%**	16.20	**34%**	**12.05**
	Okhla Phase-2	11.01	**9%**	11.10	**10%**	10.33	**2%**	10.42	**3%**	13.37	**32%**	**10.13**
	Parparganj	13.98	**20%**	13.31	**15%**	12.60	**8%**	13.06	**12%**	137.36	**1081%**	**11.62**
	Rohini	14.80	**9%**	14.85	**9%**	13.91	**2%**	15.21	**12%**	22.00	**62%**	**13.60**
	Sonia Vihar	13.38	**3%**	14.31	**10%**	13.95	**7%**	14.75	**13%**	16.96	**30%**	**13.03**
	Vivek Vihar	15.33	**8%**	15.41	**9%**	14.54	**3%**	14.38	**1%**	21.45	**51%**	**14.17**
	Wazirpur	17.92	**23%**	15.52	**6%**	14.81	**1%**	**14.48**	-1%	17.16	**17%**	14.61
MAPE	Anand Vihar	21.51	**5%**	22.01	**8%**	21.23	**4%**	20.69	**1%**	23.37	**14%**	**20.47**
	Ashok Vihar	21.34	-6%	21.78	-4%	**21.20**	-7%	21.77	-4%	23.46	**3%**	22.70
	CPRI MR	33.72	0%	35.21	**5%**	36.80	**9%**	33.83	0%	37.36	**11%**	**33.67**
	JNS	18.12	**7%**	18.52	**9%**	17.85	**5%**	17.90	**6%**	140.26	**729%**	**16.92**
	MDCNS	14.88	**12%**	15.20	**14%**	14.37	**8%**	14.39	**8%**	17.75	**33%**	**13.30**

Table 5 continued

Error	Station	BiLSTM	Imp%	CNN	Imp%	GRU	Imp%	LSTM	Imp%	RNN	Imp%	Proposed
	Mandir Marg	26.54	**16%**	26.44	**15%**	25.58	**12%**	24.98	**9%**	28.19	**23%**	**22.93**
	Okhla Phase-2	20.64	**19%**	19.61	**13%**	18.77	**9%**	19.02	**10%**	19.98	**16%**	**17.28**
	Parparganj	19.36	**14%**	20.33	**19%**	18.98	**11%**	19.67	**15%**	147.97	**768%**	**17.04**
	Rohini	17.45	**6%**	18.14	**11%**	16.90	**3%**	17.66	**8%**	22.21	**35%**	**16.41**
	Sonia Vihar	18.09	**3%**	18.95	**8%**	18.48	**5%**	19.84815	**13%**	19.81265	**13%**	**17.54**
	Vivek Vihar	21.41	**9%**	22.22	**13%**	21.04	**7%**	21.45	**9%**	24.71	**25%**	**19.71**
	Wazirpur	22.05	**21%**	19.03	**4%**	18.65	**2%**	18.25	0%	19.78	**8%**	**18.24**
MSE	Anand Vihar	993.57	**6%**	1050.32	**13%**	1092.46	**17%**	985.12	**6%**	1952.51	**109%**	**933.56**
	Ashok Vihar	562.86	**32%**	437.51	**2%**	477.69	**12%**	414.80	**-3%**	1016.37	**138%**	**426.92**
	CPRI MR	823.54	-1%	832.43	0%	933.49	**12%**	**793.74**	-5%	1332.73	**60%**	831.97
	JNS	364.49	**47%**	290.14	**17%**	305.92	**23%**	276.30	**11%**	4673.74	**1784%**	**248.03**
	MDCNS	356.05	**39%**	294.61	**15%**	294.32	**15%**	289.30	**13%**	793.31	**210%**	**255.83**
	Mandir Marg	373.85	**9%**	413.07	**20%**	380.27	**10%**	362.67	**5%**	685.07	**99%**	**344.51**
	Okhla Phase-2	430.84	**14%**	461.5	**22%**	404.03	**6%**	**378.76**	**0%**	780.25	**106%**	379.41
	Parparganj	926.27	**66%**	616.10	**10%**	718.34	**29%**	789.54	**41%**	28428.77	**4990%**	**558.55**
	Rohini	634.78	**20%**	584.86	**11%**	589.11	**12%**	679.91	**29%**	1799.90	**241%**	**528.03**
	Sonia Vihar	530.88	**12%**	523.10	**10%**	596.43	**26%**	701.49	**48%**	1018.97	**115%**	**474.67**
	Vivek Vihar	792.98	**26%**	717.83	**14%**	753.13	**20%**	698.73	**11%**	1926.42	**207%**	**627.08**
	Wazirpur	1026.93	**71%**	656.24	**9%**	695.48	**16%**	640.13	**6%**	1068.13	**78%**	**601.40**

the percentage improvement of performance measures of baseline models over the proposed model. Based on the improvement values, the proposed model is superior to all baseline DL over various performance measures. These findings highlight the reliability and accuracy of Anand Vihar in predictive modeling for STC at 0.80 correlation, making them prime choices for this specific task. Beyond individual station performance, the proposed model consistently outperformed the baseline across all locations, demonstrating robustness and general applicability. Although the Major Dhyan Chandra National Stadium station recorded the lowest RMSE, the STC model achieved substantial error reduction at all stations, indicating that the improvement is not localized but broadly consistent. No significant outliers were observed, and performance improvements were consistent across all stations, reinforcing the robustness of the proposed model.

6.2 Graphical Analysis

6.2.1 Predictive Pattern Analysis (SS +STC)

The actual and predicted DL model and proposed model (see Fig. 8) for the Anand Vihar station testing dataset for SS provides a comprehensive overview of the comparison between observed $PM_{2.5}$ concentration values and predictions generated by a proposed BiGRU-1DCNN and baseline DL with Anand Vihar single station univariate data. The authors plot a total of 144 hours spanning for better visualization from August 1, 2023, at 00:00 to August 6, 2023, at 23:00. These plots offer a detailed evaluation of the model's performance in prediction $PM_{2.5}$ concentration at a diverse array of different baseline DL models. All the baseline DL models are predicted well, except for RNN. The proposed model predicts a similar pattern to actual $PM_{2.5}$ values.

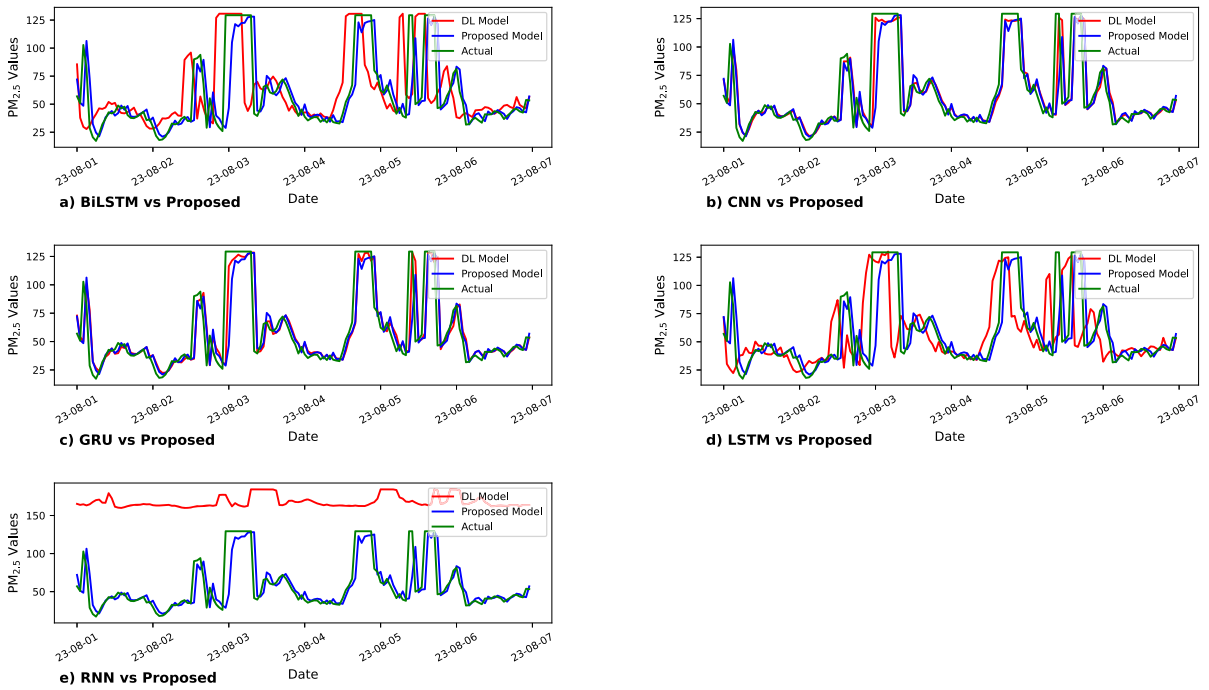

Fig. 8 Single Station $PM_{2.5}$ Predictions at Anand Vihar (01-August-2023 to 06-August-2023). The x-axis represents time and the y-axis denotes $PM_{2.5}$ concentrations ($\mu g/m^3$)

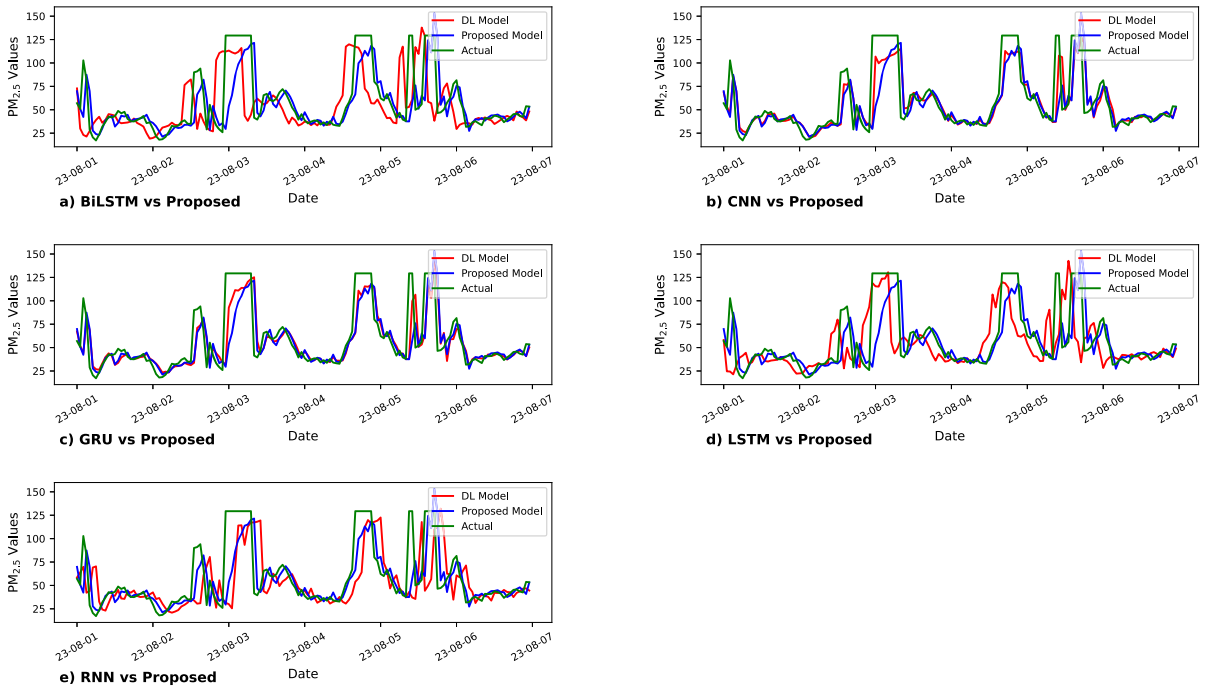

Fig. 9 The spatiotemporal for Anand Vihar (D2) of actual, proposed, and traditional DL model predicted values of $PM_{2.5}$ test set

The line plots (see Fig. 9) for spatio-temporal particular monitoring station Anand Vihar testing dataset offer a comprehensive comparison between observed $PM_{2.5}$ concentration values and predictions generated by a specific proposed model BiGRU-1DCNN and baseline DL model. These plots cover a substantial period, from August 1, 2022, at 00:00 to August 6, 2023, at 23:00, enabling a detailed examination of the model's performance in capturing the spatiotemporal distribution of $PM_{2.5}$ concentration. Each model has unique characteristics and spatio-temporal aspects are considered, providing valuable insights into the reliability and accuracy of the model's predictions across various scenarios. The proposed model's $PM_{2.5}$ prediction is close to the actual values. The red trend represents the baseline line DL model, the blue line represents the proposed model $PM_{2.5}$ prediction, and the green line indicates the actual $PM_{2.5}$ values.

The prediction results at Anand Vihar station, reveal noticeable discrepancies among different deep learning models when compared to the actual $PM_{2.5}$ values. The proposed BiGRU-1DCNN model demonstrates closer alignment with observed trends, particularly during high pollution peaks, compared to other models. Anand Vihar, being a high-traffic and industrial zone, exhibits significant temporal fluctuations and sharp $PM_{2.5}$ spikes, posing challenges for accurate prediction. This emphasizes the impact of localized pollution sources and suggests that future modeling efforts should account for such spatial characteristics to enhance predictive performance in heavily polluted urban hotspots.

6.2.2 Predictive Distribution Analysis SS + STC

The superimposed distribution analysis (see Fig. 10) The single station's univariate Anand Vihar station offers a visual representation of the relationship between observed and predicted $PM_{2.5}$ values over the specified period, from October 4, 2022, at 17:00 to August 6, 2023, at 23:00, focusing on a specific monitoring location Anand Vihar without considering spatiotemporal aspects over utilizing the different baseline DL model.

The coefficient of determination (R^2) was computed using the standard formula $R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$. Negative R^2 values can occur when the model performs worse than the mean baseline, typically in highly noisy or non-stationary data regions. Such values were

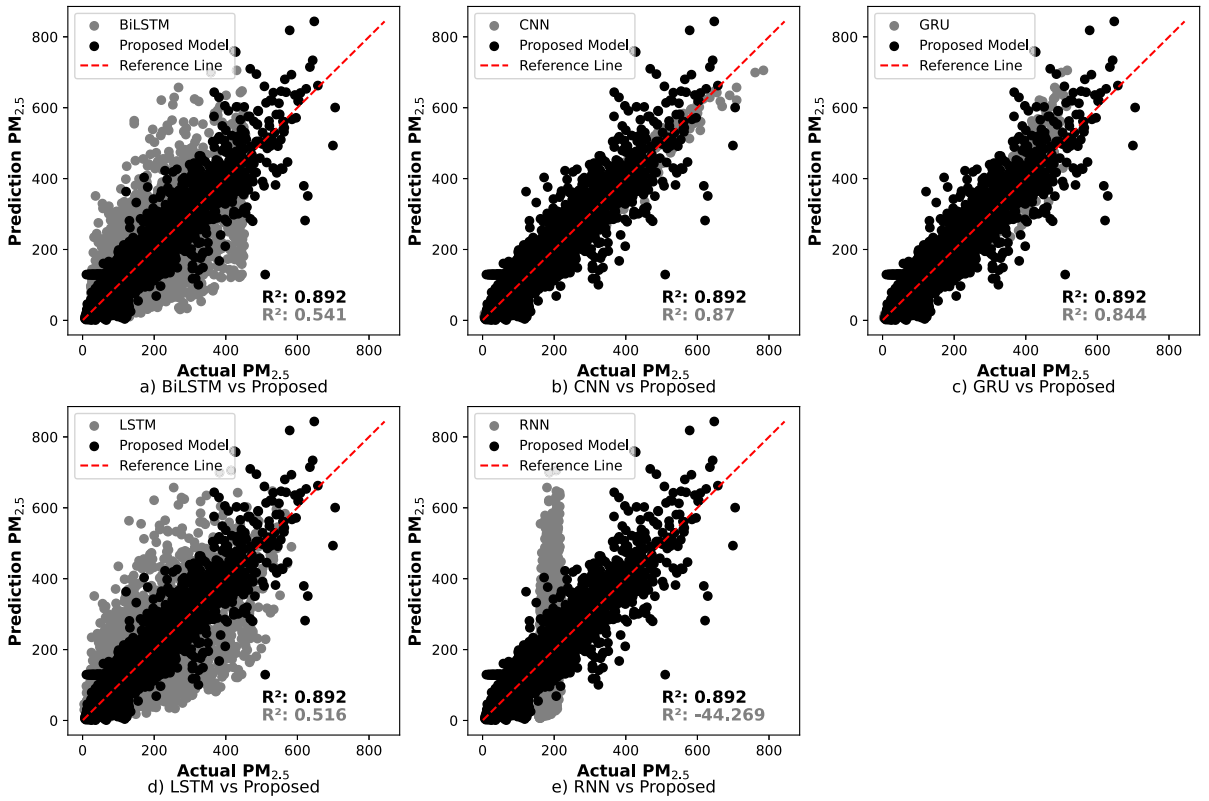

Fig. 10 Predicted vs. Actual $PM_{2.5}$ Concentration at Anand Vihar station. The x-axis represents actual $PM_{2.5}$ concentrations ($\mu g/m^3$), and the y-axis represents predicted values ($\mu g/m^3$) of test set

observed in outlier-prone stations with low spatiotemporal correlation.

Each scattered plot represents the actual values with the proposed model prediction and baseline DL model prediction values for the Anand Vihar station. The R^2 scores of baseline BiLSTM, CNN, GRU, LSTM, and RNN are 0.541, 0.87, 0.844, 0.516, and -44.269, respectively, compared to the R^2 scores of the proposed model, which are 0.892. The grey indicates the baseline DL model, the black indicates the proposed model, and the red indicates the actual reference line. These plots allow a straightforward assessment of the model's performance within this fixed location. The degree of scatter or clustering of data points indicates the accuracy and reliability of the model's predictions at this single station. Analyzing these scattered plots is essential for evaluating the model's accuracy and ability to accurately predict $PM_{2.5}$ concentration at the specified location and over the given time frame.

The superimposed distribution plots (see Fig. 11) for the spatio-temporal data from the Anand Vihar station visually represent the relationship between observed $PM_{2.5}$ concentration values and the predicted $PM_{2.5}$ concentration values from both the baseline deep learning model and the proposed model. This analysis covers the period from October 4, 2022, at 17:00 to August 6, 2023, at 23:00. Each scatter plot displays the actual observed $PM_{2.5}$ values alongside the predicted values from the proposed model and the baseline DL model for the Anand Vihar station. The R^2 scores for the baseline models BiLSTM, CNN, GRU, LSTM, and RNN are 0.474, 0.892, 0.878, 0.489, and 0.457, respectively, while the proposed model achieves an R^2 score of 0.905. In the plots, grey represents the baseline DL model, black represents the proposed model, and red indicates the actual reference line. These plots facilitate a straightforward assessment of the model's performance across various baseline models. The degree of scatter or clustering of the data points reflects the accu-

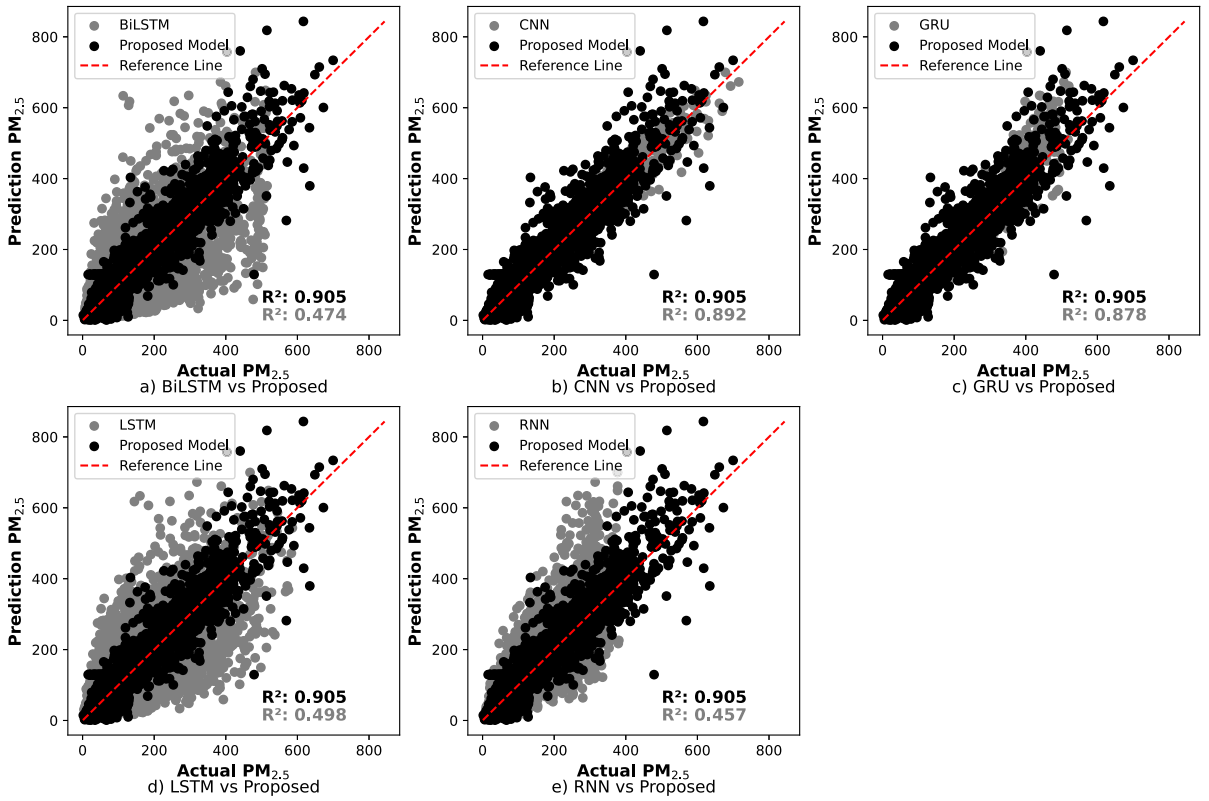

Fig. 11 The scatter plot to STC for all datasets of actual and predicted values of test set

racy and reliability of the model's predictions. A well-fitted model would show data points closely grouped along a linear or consistent trend line, indicating high predictive accuracy. Conversely, more dispersed data points may suggest variations or errors in the model's predictions.

Analyzing these scatter plots is essential for gaining insights into the proposed model's performance and its ability to accurately predict $PM_{2.5}$ concentrations over the specified time frame and datasets.

6.3 Non-parametric Statistical Friedman Ranking Analysis

The table provides a thorough comparison of various deep learning algorithms based on four essential ranking metrics: RMSE Rank, MAE Rank, MAPE Rank, and MSE Rank. These algorithms are evaluated in distinct scenarios, specifically STC and SS. Each algorithm's performance is carefully ranked according to these metrics, with lower values signifying better performance. In the STC scenarios, the proposed ST BiGRU-1DCNN model and LSTM-STC emerge as the top performers, achieving the lowest ranking scores across all four metrics and highlighting their effectiveness. In contrast, RNN-STC is positioned as the least effective among these algorithms. For the SS scenarios, both LSTM-SS and CNN-SS significantly outperform their peers, securing the lowest ranks in the specified metrics, while RNN-SS falls behind, further emphasizing the strengths of the former. This table not only summarizes the algorithm rankings but also clearly delineates the performance differences among them. Friedman ranking of deep learning models and proposed models sample size: $n = 12$ stations, $k = 12$ models based on RMSE, MAE, MAPE, and MSE. The Friedman test yielded statistically significant results indicating meaningful differences in performance rankings across models (Table 6).

The proposed STC and proposed SS got the first ranking in every error parameter measure. The improvement can be attributed to the model's ability to learn

Table 6 The Friedman ranking of all the proposed and traditional deep learning models with 0.80 correlated related to all selected stations

Algorithm	RMSE Rank	MAE Rank	MAPE Rank	MSE Rank
Proposed Model - STC	**1.3333**	**1.8333**	**1.5**	**1.3333**
BLSTM-STC	6	4.8333	5.1667	6
CNN-STC	3.5	5.5	5.9167	3.5
GRU-STC	4.75	3.3333	3.3333	4.75
LSTM-STC	3.25	3	4.4167	3.25
RNN-STC	11	10.9167	9.8333	11
Proposed Model-SS	**4.6667**	**4.8333**	**4.5833**	**4.6667**
BLSTM-SS	7.6667	8.0833	9.0833	7.6667
CNN-SS	7.8333	8.4167	8.8333	7.8333
GRU-SS	9.4167	9.25	8.1667	9.4167
LSTM-SS	6.9167	6.3333	5.6667	6.9167
RNN-SS	11.6667	11.6667	11.5	11.6667

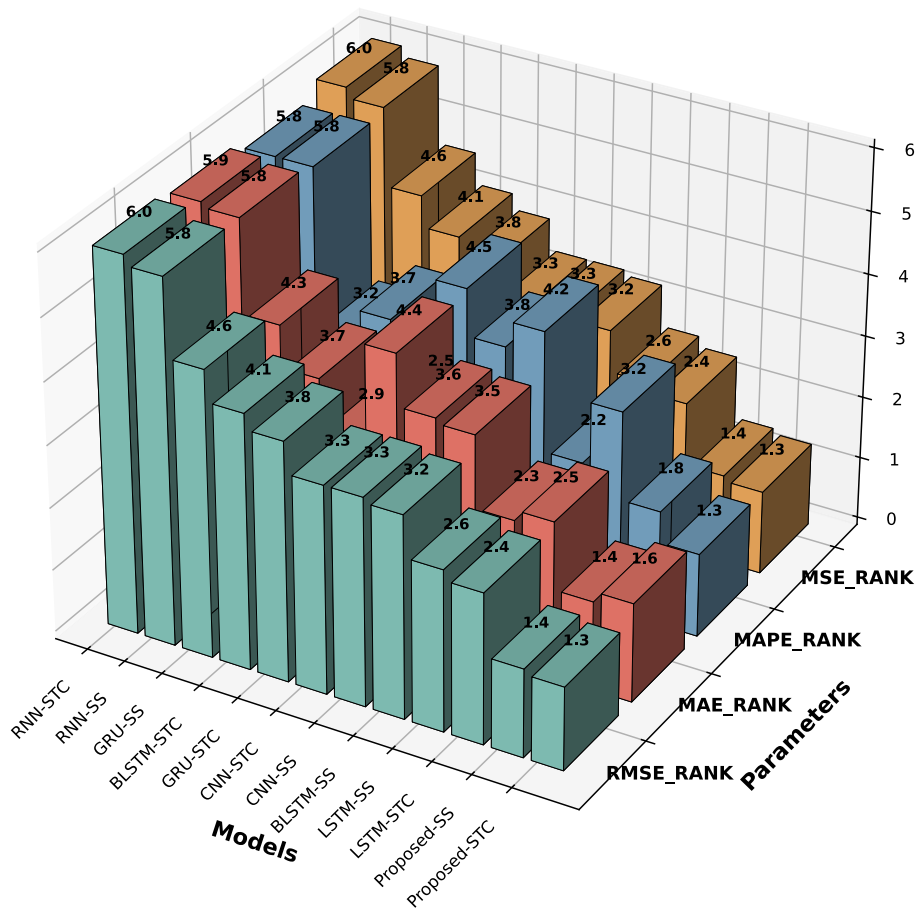

Fig. 12 Friedman method of ranks for model comparison with the different evaluation parameters

not only temporal patterns within a single station but also the spatial dependencies between geographically proximate stations. In densely populated and highly polluted environments like Delhi, nearby stations often exhibit similar pollution trends due to shared sources and meteorological conditions. Capturing these relationships allows the model to generalize better, especially during peak pollution periods or abrupt changes in air quality. This observation is consistent with findings from related studies, which emphasize the role of spatial dependencies in enhancing $PM_{2.5}$ forecasting accuracy (see Fig. 12).

6.4 One Year Forecasting $PM_{2.5}$ Concentration

The forecasting results from the ST BiGRU-1DCNN model, spanning from July 8, 2023, to July 6, 2024, demonstrate its remarkable predictive capabilities, outshining traditional deep learning models. These forecasts offer essential insights into the trends of $PM_{2.5}$ concentrations in Delhi, with February 2024 exhibiting notably higher levels compared to the rest of the year. The median, rather than the mean, is selected to reduce the influence of extreme pollution events or outliers, which are common in air quality data. Compared to the half monthly median balances temporal granularity and noise reduction.It retains essential temporal dynamics while ensuring robustness against skewed distributions and data anomalies, especially useful in long-term forecasts where stability is critical. The median values, indicated by the black dotted line, reveal that the second half of February 2024 surpasses all previous half-month forecasts. The primary contributors to this increase include open waste burning, biomass burning, construction dust, industrial emissions, and stubble burning. The median aggregation method enhances interpretability in visualization (see Fig. 13) and supports a more consistent trend analysis.

Implications and Limitations

The findings of this study underscore the potential of the proposed Hybrid BiGRU-1DCNN model to improve spatio-temporal forecasting of $PM_{2.5}$ concentrations in urban environments like Delhi. By leveraging both spatial correlations between monitoring stations and temporal dependencies in pollution trends, the model demonstrates enhanced predictive accuracy compared to traditional deep learning methods.

Implications The accurate forecasting of $PM_{2.5}$ levels directly affects public health planning, air quality alerts, and policy decision-making. Timely and precise predictions can enable proactive mitigation strategies, such as restricting vehicular traffic or industrial emissions during forecasted high pollution periods. Furthermore, this modeling framework can be adapted to other

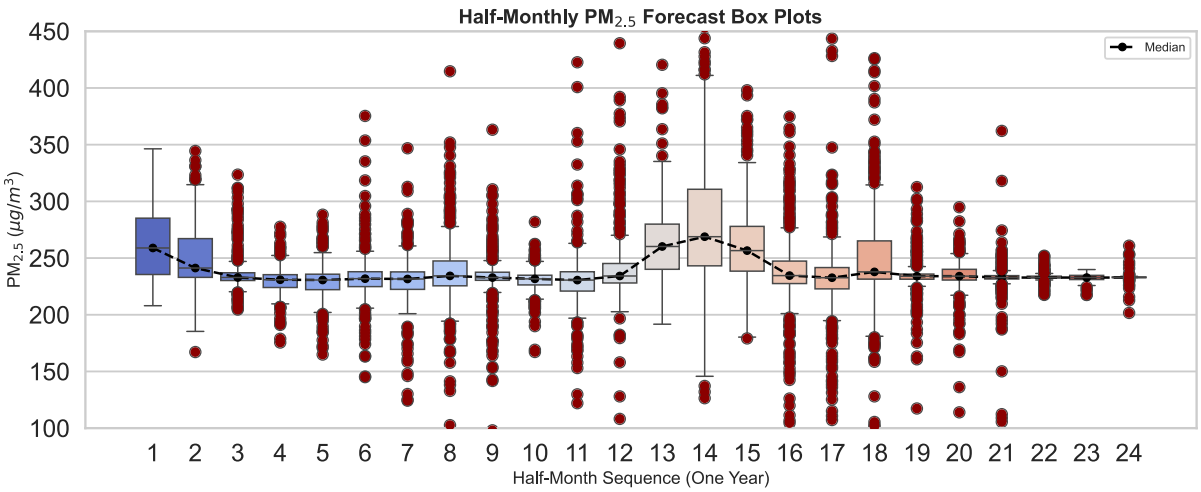

Fig. 13 The one year forecasting $PM_{2.5}$ concentration utilizing proposed STC for Delhi, India

cities with similar air quality challenges, supporting broader environmental management efforts.

Limitations Despite its promising results, the model has several limitations:

- The analysis is limited to a fixed number of monitoring stations 28, and the inclusion of more stations or mobile sensors could improve spatial granularity.
- Meteorological variables were not fully integrated future models could benefit from incorporating wind speed, temperature, humidity, and boundary layer height.
- The model treats spatial and temporal dependencies somewhat independently.
- While the hybrid architecture improved performance, the combination strategy was manually chosen.

Future research should explore multi-modal data fusion, dynamic hyperparameter tuning, and the use of ensembles to address these limitations and further improve long-term air quality forecasting.

7 Conclusion

The hybrid BiGRU-1DCNN model represents a significant advancement in forecasting spatio-temporal $PM_{2.5}$ concentration levels in Delhi, India. By incorporating bidirectional gated recurrent units and a one-dimensional convolutional neural network , this model effectively captures the temporal dependencies and spatial correlations inherent in the $PM_{2.5}$ data. The fusion of these two architectures enables the model to learn sequential patterns while accounting for geographical variations across different monitoring stations. This approach ensures robust predictions for $PM_{2.5}$ concentrations at various locations and time intervals.

In terms of performance, the proposed model achieved impressive results, including the lowest RMSE of 15.75 and MSE of 248.04 at the Jawaharlal Nehru Stadium station. Furthermore, the model attained the lowest MAE of 9.04 and MAPE of 13.31 at the Major Dhyan Chandra National Stadium station. A comprehensive Friedman ranking analysis was conducted, comparing the proposed model with several baseline models. The proposed ST BiGRU-1DCNN model secured the highest rankings across multiple metrics: an RMSE rank of 1.33, MAE rank of 1.83, MAPE rank of 1.5, and MSE rank of 1.33. The results indicate that the proposed ST BiGRU-1DCNN model outperforms other models in terms of RMSE and MSE, demonstrating its effectiveness in forecasting spatio-temporal $PM_{2.5}$ concentrations. Its integration of BiGRU and 1DCNN provides superior accuracy and reliability compared to individual models. From a policy perspective, accurate forecasting can drive more effective air quality management programs. Policymakers can use the model's predictions to implement timely interventions, such as controlling traffic emissions, restricting industrial activities, or promoting public health measures during high-pollution periods. In the long term, the model could contribute to designing sustainable urban planning initiatives to reduce pollution exposure and enhance air quality in Delhi.

Future research holds substantial promise for refining the proposed model, particularly in optimizing its spatiotemporal forecasting capabilities. Key areas of focus will include improving the model's architecture through deeper layers, innovative feature representations, and advanced hyperparameter tuning. Additionally, integrating more granular geographical and meteorological data, such as satellite imagery and ground-level sensor networks, will provide a more comprehensive understanding of the complex factors influencing $PM_{2.5}$ concentrations. By enhancing the model's predictive capabilities, this research aims to mitigate the adverse effects of air pollution on public health in Delhi, contributing to the continued development and relevance of spatiotemporal forecasting models in the field.

Author contributions Naushad Ahmad: Conceptualization of this study, Writing - Methodology, Software, Data curation, Writing - Original draft preparation, Visualization. Vipin Kumar: Validation, Formal analysis, Methodology, Investigation, Writing - Review, Editing, Supervision

Funding Information No funding was received for conducting this study.

Data Availability The datasets and code generated analyzed during the current study are available from the corresponding author on reasonable request. Download data for arcgis plot (https://www.diva-gis.org/gdata).

Code Availability All code was implemented in Python. Interested researchers can request access by contacting the corresponding author.

Declarations

Competing Interests The authors have no conflicts of interest to declare.

Ethics Approval and Consent to Participate Not applicable

Consent for Publication Not applicable

References

Agarwal, S., Praveen, G., Gautam, A. S., Gautam, S., Tiwari, R. N., Kumar, S., et al. (2023). Unveiling the surge: Exploring elevated air pollution amidst the covid-19 era (2019–2020) through spatial dynamics and temporal analysis in delhi. *Water, Air, & Soil Pollution., 234*(12), 1–21. https://doi.org/10.1007/s11270-023-06766-y

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *Journal of big Data., 8*, 1–74. https://doi.org/10.1186/s40537-021-00444-8

Badawy, H. (2025). Recent trends in spatial modeling studies of air quality (2010–2023). *Earth Science Informatics, 18*(2), 193. https://doi.org/10.1007/s12145-024-01667-y

Bahadur, F. T., Shah, S. R., & Nidamanuri, R. R. (2023). Applications of remote sensing vis-à-vis machine learning in air quality monitoring and modelling: a review. *Environmental Monitoring and Assessment, 195*(12), 1502. https://doi.org/10.1007/s10661-023-12001-2

Bodor, K., Szép, R., Keresztesi, Á., & Bodor, Z. (2023). Temporal evolution of pm2. 5, pm10, and total suspended particles (tsp) in the ciuc basin (transylvania) with specific microclimate condition from 2010 to 2019. *Environmental Monitoring and Assessment., 195*(7), 798. https://doi.org/10.1007/s10661-023-11407-2

Dey, R., Salem, F.M. (2017). Gate-variants of gated recurrent unit (gru) neural networks. In *2017 IEEE 60th international Midwest Symposium on Circuits and Systems (MWSCAS)* (pp. 1597–1600). IEEE. https://doi.org/10.1109/MWSCAS.2017.8053243

Draper, C. S., Walker, J. P., Steinle, P. J., De Jeu, R. A., & Holmes, T. R. (2009). An evaluation of amsr-e derived soil moisture over Australia. *Remote Sensing of Environment, 113*(4), 703–710. https://doi.org/10.1016/j.rse.2008.11.011

Elisephane, I., & Ozunu, A. (2024). Emission flux and dispersion analysis of particulate matter (pm10) from mining and industrial areas in rusizi district-rwanda. *Water, Air, & Soil Pollution., 235*(6), 328. https://doi.org/10.1007/s11270-024-07127-z

Ghosh, S., & Dutta, S. (2023). Deep learning-based spatiotemporal prediction of air pollution in indian megacities. *Environmental Technology & Innovation., 32*, 103512. https://doi.org/10.1016/j.eti.2023.103512

Gilik, A., Ogrenci, A. S., & Ozmen, A. (2022). Air quality prediction using cnn+ lstm-based hybrid deep learning architecture. *Environmental Science and Pollution Research,* 1–19. https://doi.org/10.1007/s11356-021-16227-w

Haddaji, B., Atoui, M., Wederni, K., Agoubi, B., & Karroubi, A. (2024). Geothermal water quality index assessment for irrigation purpose with multicomputing modeling coupled with gis: Case of el hamma, southeastern tunisia. *Water, Air, & Soil Pollution., 235*(2), 160. https://doi.org/10.1007/s11270-024-06961-5

Her, Q. L., & Wong, J. (2020). Significant correlation versus strength of correlation. *American Journal of Health-System Pharmacy, 77*(2), 73–75. https://doi.org/10.1093/ajhp/zxz280

Jain, D., Bhatnagar, S., Rathi, V., Sachdeva, K., Tewani, A., & Sharma, G. (2025). Principal component regression approach for measuring the impact of built environment variables on multiple air pollutants in Delhi. *Discover Atmosphere., 3*(1), 2. https://doi.org/10.1007/s44292-025-00029-7

Jakhar, R., Tripathi, S. M., & Yadav, H. (2025). Unraveling air pollution in anand vihar, delhi: insights, impacts, and interventions. *International Journal of System Assurance Engineering and Management,* 1–10. https://doi.org/10.1007/s13198-024-02697-x

Khaslan, Z., Nadzir, M. S. M., Johar, H., Siqi, Z., Sulong, N. A., Mohamed, F., Majumdar, S., Suris, F. N. A., Hawari, N. S. S. L., Borah, J., et al. (2024). Utilizing a low-cost air quality sensor: Assessing air pollutant concentrations and risks using low-cost sensors in selangor, malaysia. *Water, Air, & Soil Pollution., 235*(4), 229. https://doi.org/10.1007/s11270-024-07012-9

Li, X., Liu, J., Liu, Z., Liu, J., & Li, P. (2024). Human health impact and economic effect for pm2. 5 exposure under carbon neutrality. *Water, Air, & Soil Pollution., 235*(9), 1–17. https://doi.org/10.1007/s11270-024-07362-4

Liu, B., Qi, Z., & Gao, L. (2024). Enhanced air quality prediction through spatio-temporal feature sxtraction and fusion: A self-tuning hybrid approach with gcn and gru. *Water, Air, & Soil Pollution., 235*(8), 532. https://doi.org/10.1007/s11270-024-07346-4

Mariappan, Y., Ramasamy, K., & Velusamy, D. (2025). An optimized deep learning based hybrid model for prediction of daily average global solar irradiance using cnn slstm architecture. *Scientific Reports, 15*(1), 10761. https://doi.org/10.1038/s41598-025-95118-3

Nakhjiri, A., & Kakroodi, A. A. (2024). Air pollution in industrial clusters: A comprehensive analysis and prediction using multi-source data. *Ecological Informatics, 80*, 102504. https://doi.org/10.1016/j.ecoinf.2024.102504

Rincon, G., Morantes, G., Roa-López, H., Cornejo-Rodriguez, M. D. P., Jones, B., & Cremades, L. V. (2023). Spatio-temporal statistical analysis of pm1 and pm2. 5 concentrations and their key influencing factors at Guayaquil City, Ecuador. *Stochastic Environmental Research and Risk Assessment., 37*(3), 1093–1117. https://doi.org/10.1007/s00477-022-02310-2

Rodríguez, J., García, M. Á., Pérez, I. A., & Jorquera, H. (2023). Saharan dust contributions to high hourly pm10 concen-

trations at a background station in Southwestern Europe. *Stochastic Environmental Research and Risk Assessment,* 1–17. https://doi.org/10.1007/s00477-023-02479-0

Rotjanabumrung, M., Phosri, A., Sihabut, T., & Neamhom, T. (2023). Short-term effects of biomass open burning related air pollution on outpatient department visits for cardiovascular and respiratory diseases in thailand. *Stochastic Environmental Research and Risk Assessment,* 1–11. https://doi.org/10.1007/s00477-023-02424-1

Sharma, A., et al. (2023). Ensemble-based $pm_{2.5}$ prediction using deep neural networks and meteorological data in Urban India. *Environmental Science and Pollution Research., 30*, 45678–45692. https://doi.org/10.1007/s11356-023-28235-z

Sun, M., Li, F., Li, Y., Chen, J., & Cheng, G. (2024). Assessing the ecological and health risks associated with heavy metals in pm2. 5 based on their potential bioavailability. *Water, Air, & Soil Pollution., 235*(5), 1–14. https://doi.org/10.1007/s11270-024-07118-0

Wang, L., Wang, G., & Gao, A. S. (2024). Exploring heterogeneity and dynamics of meteorological influences on us pm2. 5: A distributed learning approach with spatiotemporal varying coefficient models. *Spatial .,* 100826. https://doi.org/10.1016/j.spasta.2024.100826

Wang, J., Wu, T., Mao, J., & Chen, H. (2023). A forecasting framework on fusion of spatiotemporal features for multistation pm2. 5. *Expert Systems with Applications,* 121951. https://doi.org/10.1016/j.eswa.2023.121951

Wang, K., Ma, C., Qiao, Y., Lu, X., Hao, W., & Dong, S. (2021). A hybrid deep learning model with 1dcnn-lstm-attention networks for short-term traffic flow prediction. *Physica A: Statistical Mechanics and its Applications., 583*, 126293. https://doi.org/10.1016/j.physa.2021.126293

Wu, Y., Lin, S., Shi, K., Ye, Z., & Fang, Y. (2022). Seasonal prediction of daily pm2. 5 concentrations with interpretable machine learning: a case study of beijing, china. *Environmental Science and Pollution Research., 29*(30), 45821–45836. https://doi.org/10.1007/s11356-022-18913-9

Wu, C., Wang, R., Lu, S., Tian, J., Yin, L., Wang, L., & Zheng, W. (2025). Time-series data-driven pm2. 5 forecasting: From theoretical framework to empirical analysis. *Atmosphere., 16*(3), 292. https://doi.org/10.3390/atmos16030292

Xu, S.-Q., He, H.-D., Yang, M.-K., Wu, C.-L., Zhu, X.-H., Peng, Z.-R., Sasaki, Y., Doi, K., & Shimojo, S. (2023). To what extent the traffic restriction policies can improve its air quality? An inspiration from covid-19. *Stochastic Environmental Research and Risk Assessment, 37*(4), 1479–1495. https://doi.org/10.1007/s00477-022-02351-7

Yu, C., Yan, G., Ruan, K., Liu, X., Yu, C., & Mi, X. (2023). An ensemble convolutional reinforcement learning gate network for metro station pm2. 5 forecasting. *Stochastic Environmental Research and Risk Assessment,* 1–16. https://doi.org/10.1007/s00477-023-02564-4

Yuan, K., Wu, K., & Liu, J. (2022). Is single enough? a joint spatiotemporal feature learning framework for multivariate time series prediction. *IEEE Transactions on Neural Networks and Learning Systems.* https://doi.org/10.1109/TNNLS.2022.3216107

Zhang, C., Lv, W., Liu, G., & Wang, Y. (2025). Multidimensional spatiotemporal autocorrelation analysis theory based on multi-observation spatiotemporal moran's i and its application in resource allocation. *Earth Science Informatics, 18*(1), 1–25. https://doi.org/10.1007/s12145-024-01598-8

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.