Deep Learning-based Air Quality Forecasting System for Urban India Using Temporal Neural Networks

Zhijun Tang*

Faculty of Computer Science, Beijing Normal-Hong Kong Baptist University, Zhuhai, Guangdong, China

Abstract. Air pollution poses a serious hazard to public health in Cities of India; Among all sources, those resulting from Energy Consumption dominate the deaths caused by air pollution. Economic growth has intensified air pollution and climate change issues at the same time. Existing prediction methods focus on isolated single-site time-series forecasting, ignoring spatial dependencies and cross-city pollution transport, limiting regional management effectiveness. Based on the hourly air quality data of key pollutants in seven cities across India during 2015-2020, this paper first explores their concentrations and correlations; Then it builds a Deep Learning forecasting system consisting of Long Short-Term Memory (LSTM), a convolutional neural network, and long short-term memory networks (CNN-LSTM) and a transformer to investigate its AQI-prediction performance. LSTM can capture long-term temporal dependencies; CNN-LSTM combines spatial-temporal information model integration; Transformers explore long-range relationships through attention mechanisms. Based on the experimental results of this paper, it can be concluded that CNN-LSTM is more effective than other systems. It also has some advantages over single-LSTM and the transformer.

1 Introduction

Precisely forecasting the air quality is essential for preventing serious public health risks due to long-term exposure of a wide region in Eastern China, which has high energy intensity and thus generates substantial pollutants such as nitrogen oxides (NOx), sulfur dioxide (SO2), and particulate matter (PM2.5). In 2019, the air-pollution-induced premature death exceeded one million; thus, it needs to develop precise prediction models now [1]. In recent decades, rapid economic development in India has caused a more than fivefold increase in carbon dioxide emissions [2]. By 2040, it is expected to have increased another third. The above reasons also cause the problems of air pollution and global warming to worsen at the same time [3]. Therefore, it needs strong Air-Quality Prediction systems. CPCB in India has constructed a three-city per-hour surveillance system with extensive spatiotemporal air-pollution data acquisition capability; currently available prediction models are mostly one-dimensional time-series isolation methods, lacking consideration for vital space interaction mechanisms and the complexity of cross-regional pollution transmission patterns needed by region-based Air Pollution Control.

Deep learning technology can also tackle the problems of Space-Time Prediction Issues. Long Short-Term Memory (LSTM) Networks are excellent in modelling Long-term Time-Varying Dependency and thus serve as the basis for multi-city air-quality forecasting [4-5]. But due to obvious space-time distributions of pollutant concentrations in the Urban Aggregate Areas, spatial features need to be extracted. Convolutional neural networks are good at extracting local spatial features, and a combination of CNN-LSTM for sequential spatiotemporal analysis has also shown some effect [6]. Qiu et al. confirmed that CNN-LSTM has the function of integrating a spatial-time model via a blood-pressure prediction method and offered some examples for application in air-pollution forecasts [7]. In addition, the Structure of the Transformer employs self-attention to enhance the identification capability for remote pollutant correlations precisely [8]. Further improved upon transformer technology to create a double-pass system for precise prediction results [9]. Combine some scales to predict the AQI based on previous changes at multiple sites and regional dispersion patterns.

Therefore, this paper established an all-round AQI prediction system through multiple scales of air pollution monitoring data from the CPCB. The system starts by using LSTM to perform serial city-time series analysis; subsequently adds spatial information represented by CNN-LSTM; Finally, uses Transformers for higher-order temporal model evaluation.

2 Method

2.1 Data description

The air quality dataset used in this paper was sourced from Kaggle and initially presented by the CPCB

* Corresponding author: t330026145@mail.bnbu.edu.cn

https://doi.org/10.1051/itmconf/20268801009

(Central Pollution Control Board) of China [10]. There are 365-hourly Air Quality Observation Records from six cities in India from 2015 to 2020. Nine primary air pollutant concentrations are included in this measurement: PM2.5, PM10, NOx, NO1, NH_3, CO, SO_2, O_3, and others. The combined pollutant set of the full AQI index is used for space-time forecast generation.

2.2 Sample selection

To focus on the 10 major cities that have suffered from severe air pollution most severely in China for this study. The above cities are Ahmedabad, Bengaluru, Delhi, Mumbai, etc. Cleansed the data so that it was uniform and dependable; for example, filled in missing values, eliminated outlier records, etc. Cleaned data consists of 170,031 records.

According to Fig. 1, this study will examine pollutant data, including PM2.5, PM10, NOx, and SO2 at various large Chinese cities. PM2.5/PM10 refers to particulate matter, and the concentration of No $_2$/SO_3 is also different

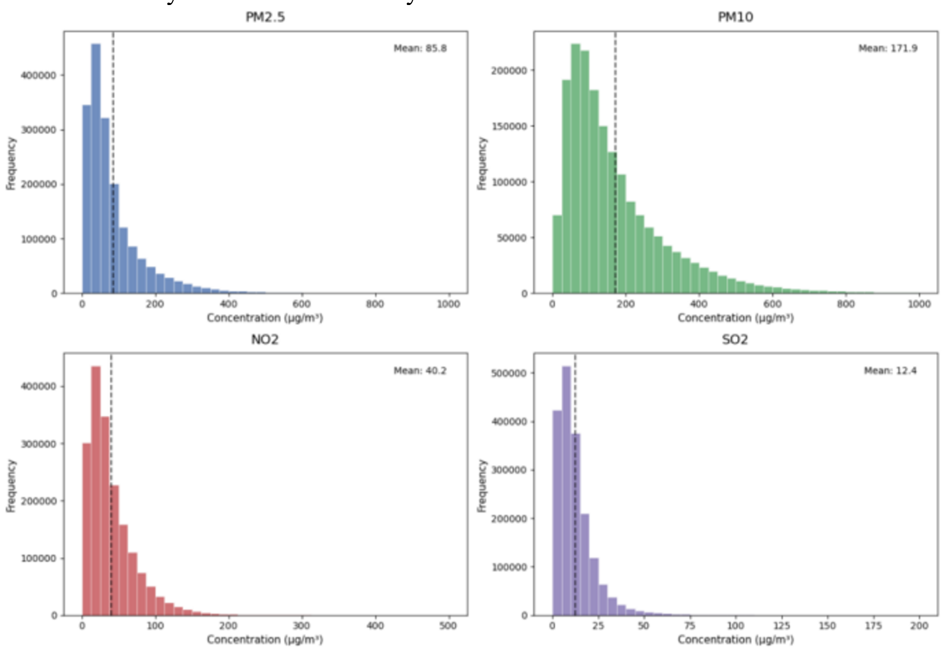


Fig. 1 Distribution of Main Pollutant Concentrations (Picture credit: Original)

Fig. 2 shows the correlation heatmap of key variables: PM2.5 correlates strongly with PM10 (r = 0.87) and AQI (r = 0.80); NO, NO_2, NOx are positively intercorrelated; O_3 is negatively correlated with NOx (r = -0.23) and NO (r = -0.19) (consistent with photochemical reactions) and weakly correlated with AQI (r = 0.04). These confirm the necessity of using all nine pollutants as model inputs.

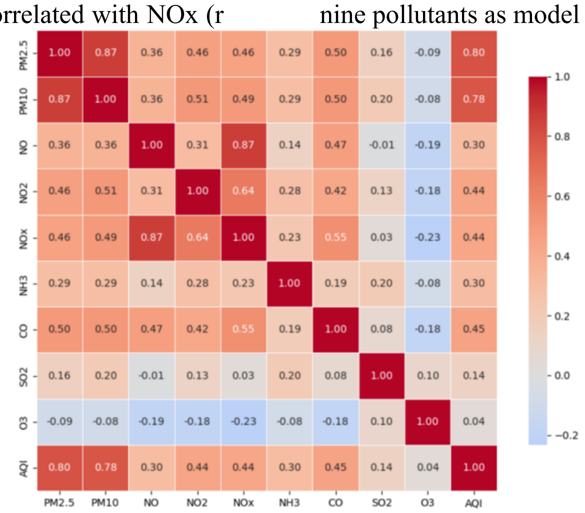


Fig. 2 Correlation Heatmap of Air Quality Variables (Picture credit: Original)

2.3 Method introduction

2.3.1 Basic principles of models

According to the correlation pattern shown in Fig. 2, this paper builds a multi-scale spatial-temporal neural network-based AQI predictor and maintains all nine pollutants as input data. Using an LSTM network to perform temporal processing on one city, overcoming traditional RNN gradient disappearance by adding gates, and capturing spatial-temporal relationships in pollutant histories at specific sites. The trained model based on Delhi's data (16,078 training samples and 4,020 test samples) has two LSTM layers; It is a regularised network with dense connections at the end to reduce overfitting.

Hybrid CNN - LSTM network combining the excellent localisation ability of CNN with improved resolution in feature extraction and pollutant distribution recognition; combine strong long-term memory capability of LSTM to process time series sequence problems.

2.3.2 Experimental design steps

In order to ensure that the prediction has good stability, according to standard design experiment: The chronological order of the dataset is taken into account to divide it into a training set (80%), a testing set (20%) without losing data information; Then take 20% from the training set as the validation Set during training so as not to cause overfitting problems after setting up parameters such as learning rate and batch size in advance; Finally, use the remaining part of the trained sample for the comprehensive evaluation by comparing the prediction performance of different models on a single evaluation point through mean absolute error(MAE), root-mean-square-error(RMSE)and coefficient-of-determination-r2(r^2).

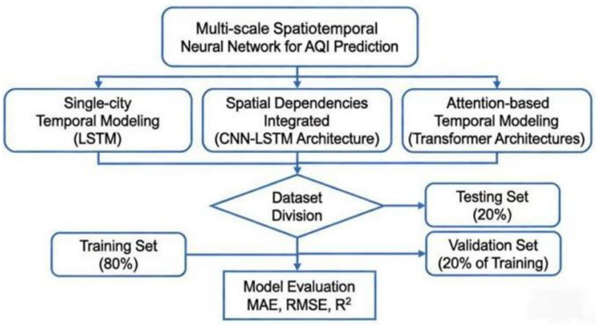


Fig. 3 Overall framework of the multi-scale spatiotemporal neural network for AQI prediction (Picture credit: Original)

3 Result and discussion

3.1 Performance comparison and analysis of LSTM and CNN+LSTM models

3.1.1 Predictive performance evaluation

In total, these were compared with the MAE, RMSE, and R^2 coefficients for all types of data, AQI category prediction accuracy to evaluate the generalization ability and classification capability of this method based on LSTMs in the research. As shown in Table 1 below, the particularised performance data for both model results have been listed side-by-side.

Table 1. Comparison of Models Performance

Metric	LSTM	CNN+LSTM
MAE	24.09	19.53
RMSE	34.26	27.70
R^2	0.9205	0.9480
Category Accuracy	71.2%	77.9%

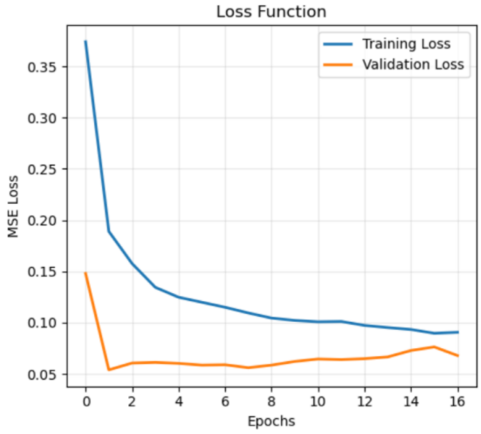
The results of this article's experiment show that LSTMs can discover long-term correlations among pollutants; In addition, they outperform other traditional methods for forecasting AQI in single-city cases. The CNN + LSTM model outperforms the LSTM baseline by a large margin in terms of error indicators, as well as predicting more accurately than before. Clearly, spatial feature extraction has been essential to the problem of air-quality prediction mentioned above. Compared with the one-time point-based approach, this hierarchical architecture obtains a global-space-Time-feature-result simultaneously in a multi-level cascaded manner.

3.1.2 Training dynamics and convergence

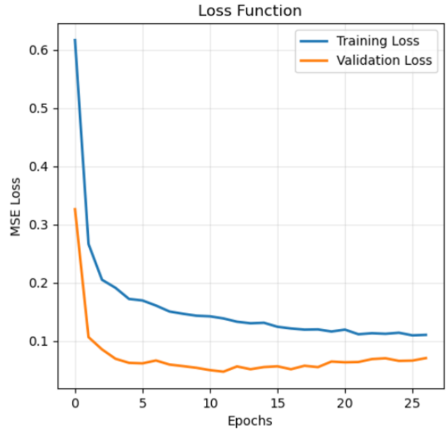
Both of them used early stopping and learning rate decay at plateaus in their training processes. Restore the optimal weight of the LSTM model at this time; The convergence value on the validation set is only 0.0560, indicating stable training with no serious overfitting. The CNN+LSTM model used the same training strategy. The training process stabilised, and the validation loss remained relatively stable at a low level, demonstrating good generalisation performance.

To present intuitively how different the two models have performed in terms of their results, Figs. 3-5 can be compared side by side; the LSTM Model is presented on the left, and the CNN + LSTM on the right; Variation of loss function during training, Comparison between predicted and real value, and distribution of residual, respectively.

During the training process, the training loss and validation loss of both models exhibited an unchangingly declining trend and eventually reached a relatively low level (as shown in Fig. 4); In this case, there was no sign of model over-fitting on either the training or test set. In addition, the training and validation MAE curves of both models also showed the same kind of trend; therefore, they have good generalisation abilities. Among these, the CNN + LSTM model showed better convergence speed and lower final loss.



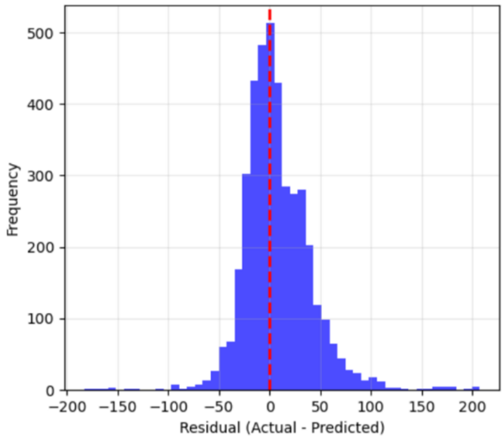
(a)



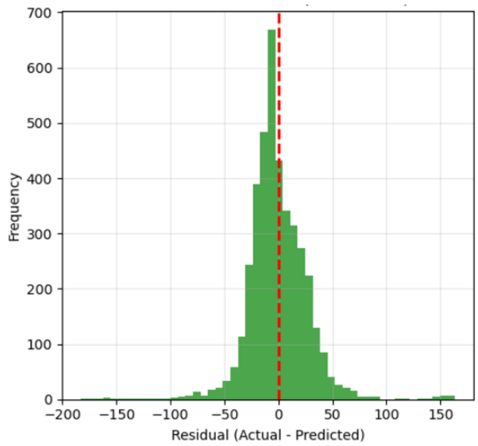
(b)

Fig. 4 Loss function, (a) LSTM; (b) LSTM + CNN (Picture credit: Original)

The residuals' distribution analysis is shown in Fig. 5: The predicted error distributions of both models are close to being normal at the centre around zero; Generally speaking, it falls within a relatively small band and meet people's expectations for an outstanding model. There are no significant system errors or lack of consistency in the predictions for these two models' deviation results. Compared to the LSTM model, there is a more stable error fluctuation pattern in the residual distribution of the CNN+LSTM model



(a)



(b)

Fig. 5 Residual distribution, (a) LSTM; (b) LSTM + CNN (Picture credit: Original)

3.2 Attention-based temporal modeling (transformer)

Based on the space-time frame built with CNN + LSTM, and further researching the ability of the attention mechanism in air quality prediction under the Transformer architecture. Transformer models represent a change in the Paradigm of recurrence and convolutions, replacing them with self-attention mechanisms for capturing global temporal dependencies and feature interactions.

3.2.1 Architecture design and implementation

Transformer models are composed of multiple parts that are suitable for time-series forecasting. Positional encoding is used first to preserve the time-order information; then, several multi-head self-attention blocks are added to simultaneously attend across multiple temporal positions and pollutants' attributes. Each attention head is trained separately on different time dependencies; thus, there are a total of eight in this paper. The attention Mechanism Works As Follows:

Attention(Q,K,V)=softmax(
QKT √dk)V (1)

where Q, K, and V represent query, key, and value matrices derived from input sequences, and d_k denotes the dimension of keys.

The model includes three Transformer encoder layers with feed-forward networks of 256 units, implementing residual connections and layer normalization for stable training. The final output layer employs global average pooling followed by dense layers for AQI prediction. Compared to LSTM's sequential processing, the Transformer's parallel computation significantly reduces training time while potentially capturing longer-range dependencies.

3.2.2 Performance analysis and comparative results

The Transformer model achieved an MAE of 23.60, RMSE of 34.77, R^2 of 0.9181, and category prediction accuracy of 75.4% (Table 2).

Table 2. Transform Model Performance

Metric	CNN+LSTM
MAE	23.60
RMSE	34.77
R^2	0.9181
Category Accuracy	75.4%

While these results represent improvements over the LSTM baseline (+4.2% in accuracy, -2.0% in MAE), they fall short of CNN+LSTM's performance across all metrics.

Fig. 6 Radar Chart Comparison of AQI Prediction Model Performance (Picture credit: Original)

Fig. 6 comparison results provide a reference for building adaptability to the task of air-quality prediction. The Underperformance of Transformers relative to CNN + LSTM can be due to the following reasons for urban air quality data in India:

First, sequence length constraints. The relatively short Input sequence is 24h; therefore, it cannot be fully utilized by transformers to extract information from distant relations, as this requires more than 100time steps for demonstration.

Second, dominance of local spatial correlations. Local spatial correlation patterns of the urban structure of Indian air pollutants (strong neighbourhood effects and topographical restrictions) are better reflected through CNNs' inductive bias rather than Transformers' global attention mechanisms.

Third, limited data scale. Because its dataset is relatively small (only 20,122 examples from Delhi), it may not have performed as well as some highly parameterized ones, such as CNNs and LSTMs, against transformers.

4 Conclusion

By analysing the concentration distributions and correlation features of major atmospheric pollutants in urban areas across India, it was found that PM2.5 serves as the primary agent driving changes in air quality; Nitrogen Oxides have strong intra-correlation among themselves; Ozone hurts most pollutants, also being associated negatively with many others under a photochemical reaction mechanism. Three time-space multi-layered neural networks, namely LSTM, CNN-LSTM, and transformer, were constructed in turn for the evaluation of prediction performance. Among which a combination with spatiotemporal characteristics, such as CNN-LSTM, obtained an outstanding result on this dataset; Thus proving that space feature learning is capable of improving pollutant-air quality forecasting accuracy more pronouncedly for areas studied in India; Although the self-attention based transformer model performed better than the basic LSTMs but still lag behind it mainly because there are localised spatial dependencies in some pollutants inside specific area. As a result of ignoring the spatiotemporal characteristics and lack of cross-border pollutant dispersion models, most current air quality forecasting systems are unsuitable for monitoring work in Indian regions. Based on deep-learning technology, an air-quality-prediction system that can make up for its deficiencies and verify the applicability of time-series neural networks in predicting urban-air quality in India.

Based on this basis for future Optimisation and expansion will be carried out from the following aspects: Firstly, improve the Model Structure by integrating an attention module with a combination of CNN-LSTM networks, adding various types of sources data such as weather and pollutant information, to enhance the Forecasting Accuracy Of serious Pollution Events; Secondly, Expand The Research Scope, Cover Other Regions Outside Dianchi To Achieve prediction scales That Span urban And Suburban Areas Across Different Geographical Locations; Thirdly, Reduce Computational Complexity While Keeping Performance Unchanged Develop A Lightweight Model For Deployment At Runtime More Easily Implemented In Practice Enhance Practical Value.

References

1. M. Gao, Sustained increase in the frequency of concurrent heatwaves, PM2.5, and ozone pollution in future China and India under air pollution control, in Abstract Set of the 36th Annual Meeting of the Chinese Meteorological Society (Hong Kong Baptist University, 2025), 60
2. OECD Publishing, Air Quality and Climate Policy Integration in India Frameworks to deliver co-benefits (OECD Publishing, 2021)
3. Down to Earth (New Delhi, India), Odisha has most contaminated sites in India: CPCB data (2021)
4. X. Shu, L. Zhang, Y. Sun, J. Tang, Host-Parasite: Graph LSTM-in-LSTM for Group Activity

Recognition. IEEE T. Neural Netw. Learn. Syst. **32**, 663–674 (2021)

5. X. Deng, S. Gao, J. Chen, et al., Joint assessment method for lithium battery health state and remaining useful life based on LSTM-XGBoost. Electr. Power Eng. Technol. 1-9 (2026)
6. Z. Qing, F. Zeng, X. Hu, et al., Improved CNN-LSTM external torque prediction algorithm for robotic arm. Modul. Mach. Tool Autom. Manuf. Tech. **2**, 62–67 (2026)
7. S. Qiu, K. Li, P. Tang, et al., Research on blood pressure prediction method based on CNN-LSTM hybrid neural network. Internet Things Technol. **16**(3), 27-31+37 (2026)
8. Z. Ralte, I. Kar, Learn Python Generative AI: Journey from Autoencoders to Transformers to Large Language Models (BPB Publications, 2024)
9. J. Tang, L. Bao, Y. Liang, Wind power prediction method based on dual-tower Transformer model with multi-feature fusion. Electron. Meas. Technol. 1-16 (2026)
10. Vopani, Air Quality Data in India (2015 - 2020) (2019)