

Comparative Study on Predicting Particulate Matter (PM_{2.5}) Levels Using LSTM Models

R. Balamurali

Chennai Institute of Technology

Partheeban Pachaivannan (✉ parthi011@yahoo.co.in)

Chennai Institute of Technology

P. Navin Elamparithi

National Institute of Technology

R. Rani Hemamalini

St. Peter's Institute of Higher Education and Research

Research Article

Keywords: Air Pollution, Deep Learning, LSTM, PM_{2.5}, Real-time air quality data, Regression

Posted Date: April 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-436897/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

In recent times, air pollution has attracted the attention of policymakers and researchers as an important issue. The pollution that contaminates the air that people breathe is from pollutants such as oxides of carbon, nitrogen and sulphur as well as minute dust particles which are smaller than 0.0025mm in diameter. The emissions contain many substances that are harmful to human health when exposed to them for a prolonged period or more than certain levels of concentration. The recent advent of technology in sensors and compact instruments to measure the concentration of pollutant levels with considerable ease. Further, this paper also predicts the air pollution for using multiple Deep Learning models that are variations of the Long Short-Term Memory (LSTM) model. In this research, only PM_{2.5} alone is taken into consideration for prediction. Real-time air quality data were collected at selected places in the study area. It is found that the model prediction data is well matched with the other researchers' results and real-time data.

1. Introduction

Pollution caused by pollutants such as CO, CO₂, NO_x, SO₂ and dust particles with diameters less than 0.0025 mg is one of the leading causes of deaths in India. A study by ¹ estimated that about 1.24 million deaths in India could be attributed to air pollution. The key purpose of this analysis was to identify and classify deep learning models that would be the best method to predict the PM_{2.5} concentrations using the dataset. Data necessary for this experiment were collected from the Central Pollution Control Board of India.

Data was collected over 15 min intervals and the type of artificial Recurrent Neural Network called Long Short-Term Memory was utilized to analyze the data and compare predictions. Data were collected from 3 monitoring stations in the city of Chennai. The locations of these stations are the neighbourhoods of Alandur, Manali and Velachery. These pollution monitoring stations collect a variety of data and exploratory variables were chosen from the gathered data at these stations. Then data was cleaned to have a more ordered and complete dataset to avoid any inaccuracies caused by missing data.

Air pollution has become a major concern in India in recent years, as large parts of the urban population of India are exposed to some of the highest levels of pollution in the world ². World Health Organization estimates that the health effects of air pollution have increased the hazard risks in major cities of India ³. Many cities in India have a population of over 1 million, and some of them rank among the top 10 in the world's most polluted cities. Of the 3 million premature deaths in the world that occur annually due to outdoor and indoor air pollution, the highest number is estimated to have occurred in India. India has many pollution problems, the most severe of which is air pollution.

⁴ developed three machine learning algorithms that predicted the levels of PM_{2.5} using a dataset that included downscaled and uncertainty. ⁵ have applied Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) to predict the concentration of PM_{2.5} and compared results with

other machine learning methods. ⁶ have applied three machine learning models that forecasted PM_{2.5} concentrations and their results showed that the variability was 80% ($R^2 = 0.8$) in the concentrations of PM_{2.5} and 75% of the pollution levels were predicted. ⁷ made an attention mechanism to capture the degree of significance of the effects on future concentrations of PM_{2.5} of the featured states at different times in the past. ⁸ have studied the PM_{2.5} using Interagency Monitoring of Protected Visual Environments (IMPROVE) and Chemical Speciation Network (CSN). They obtained data from these two networks with different operating structures, sampling practices, analytical methods, analytical facilities, and data handling and validation practices and they collected data for 33 months. Further they concluded that the combined method of CSN and IMPROVE dataset will explain better understanding of PM_{2.5} in urban and rural areas.

Magazzino et al established and experimented the relationship among COVID-19-related deaths, economic growth, PM₁₀, PM_{2.5}, and NO₂ concentrations for New York, USA ⁹. They arrived that the PM_{2.5} and NO₂ are the important pollutants and economic growth rate increases pollution level for COVID -19 death rates. A study on Impact of Outdoor and Indoor Meteorological Conditions on the COVID-19 transmission in the Western Region of Saudi Arabia by ¹⁰. They have considered 10 impact outdoor and indoor meteorological parameters for COVID-19 cases. They concluded that highest daily COVID-19 cases when the temperature ranges between 40.71 °C to 41.20 °C. A research on Source analysis of heavy metal elements of PM_{2.5} in canteen in a university in winter was carried by ¹¹. They analysed the indoor and outdoor PM_{2.5} in a canteen of university and found that the PM_{2.5}. At indoor is 99.43 µg/m³ and outdoor is 103.09 µg/m³. Further they found that more than half of the PM_{2.5} penetrates from the adjacent outdoor area in the study location.

¹² developed a novel long short-term memory neural network extended (LSTME) model with spatiotemporal correlations. The authors used hourly PM_{2.5} data from Beijing City and the results showed a mean absolute percentage error (MAPE) of 11.93%. In another study on AQI prediction for Delhi done by ¹³ used a deep recurrent neural network (RNN) model. Their LSTM model achieved good results for concentrations of pollutants. A study carried out for Beijing, China to predict PM_{2.5} using hourly data collected for one year ¹⁴. Their mean R^2 values varied from 0.159 to 0.85 after 216 experiments. ¹⁵ developed a composite system that predicted both PM_{2.5} and PM₁₀. They used Moderate Resolution Imaging Spectroradiometer (MODIS) images, with a 1km spatial resolution and concluded that the LSTM model is best for prediction of PM_{2.5} / PM₁₀. ¹⁶ predicted PM by using two sets of 3-D chemistry-transport model (CTM) simulations and the results index of agreements ranging from 0.62 to 0.79.

A research carried out on PM_{2.5} prediction for Wuhan and Chengdu by ¹⁷ used PM_{2.5} concentration data from 2015-2017. Metrological data were also used in developing the model and better results were achieved because of this. A machine learning method is adopted to predict PM_{2.5} using six-year metrological data ¹⁸. The model has shown that the use of machine learning-based statistical models are important for forecasting concentrations of PM_{2.5} from meteorological data. A study was carried out to predict air quality with time to predict going up to 48 hrs by combining multiple neural networks ¹⁹. This

experiment resulted in excellent performance and outperformed current state-of-the-art methods.²⁰ proposed a deep learning model to predict air quality in South Korea that used Stacked Autoencoders to train and test data. Research work by²¹ used meteorological data to forecast AQI. This is the only study carried for Chennai city and used one of the monitoring stations from which the data used here was collected from.

2. Methods

2.1. About the study area

Chennai is located along the coast of the Bay of Bengal. It is the state capital of Tamil Nadu and the fourth largest metropolis in India. Chennai lies between the latitudes 12°50'49" and 13°17'24" and longitudes 79°59'53" and 80°20'12". It can be counted as a part of the Coromandel Coast along the eastern part of India²². The terrain around Chennai is a flat coastal plain and since it is close to the equator, it is usually humid and hot. The highest temperatures are reached in May-June and are generally around 40°C for a few days and the least temperatures are felt in early January with the recorded temperature of about 20°C throughout the month. Chennai is a major transport hub for road, rail, air and sea transport linking major inland and overseas cities. Chennai is one of India's most prominent educational centres with a range of institutions and research centres. The metropolitan area of Chennai stretches to some 1,189 sq.km.

2.2. About the dataset

The data was collected from the 3 Central Pollution Control Board (CPCB) monitoring stations in the city of Chennai [19]. The stations are located at Alandur, Manali and Velachery and illustrated in Fig. 1. The exploratory variables collected from these locations were the atmospheric pressure (BP), relative humidity (RH), PM2.5 values, wind degree (WD) and wind speed (WS). The data collected was present in 15 min intervals for the period of 00:00, 01 May 2019 to 23:59, 30 April 2020 and each station yielded a dataset containing 35,039 data rows totalling a 105,117 data rows. The missing values of PM2.5 were approximately 78.28%. The data was processed to remove any rows that had empty columns and the data was restricted to rows that had PM2.5 levels of less than 2.5×10^{-4} mg/L. This left the data to be reduced 22,827 data rows as certain elements were missing in all the other rows.

The statistical summary of the dataset has been shown in Table 1. The table provides an insight towards how the dataset is structured. The exploratory variables shown in Table 1 were then used to plot a heatmap. Figure 2 shows the correlation between the different exploratory variables collected in the dataset. Figure S1 shows the first 5 rows of information within the dataset. Figure S2 provides statistics about the dataset such as central tendency, dispersion and shape of the dataset distribution. Figure S3 provides information about the type of data stored within the dataset.

2.3. About DL and RNN

Deep learning is a subset of machine learning methods based on representation learning and artificial neural networks. Learning is of 3 types, namely, unsupervised, supervised or semi-supervised. Deep learning architectures such as deep belief networks, deep neural networks, recurrent neural networks and convolutional neural networks are being applied to speech recognition, natural language processing, computer vision, audio recognition, machine translation, drug design, social network filtering, medical image analysis, bioinformatics, material inspection and board game programs. RNNs are the basis for the LSTMs used in the models. They are a class of Artificial Neural Networks (ANN) that use their internal state to process variable sequence length of inputs. RNNs can be defined as a generalized form of feedforward neural networks. This means that RNNs can use previous outputs as inputs within the model with hidden states as well. RNNs also have the added advantage of being able to compute inputs of varying lengths and the size of the model doesn't change with the size of the input. But the disadvantage of using an RNN is that it is very difficult to train an RNN and it takes a lot of time to train RNNs.

3. Pm2.5 Prediction Using Deep Learning

The deep learning approaches used here were different variations of the artificial recurrent neural network (RNN) called long short-term memory (LSTM). LSTMs were introduced by Hochreiter & Schmidhuber in 1997 and can learn long term dependencies²³. LSTMs have varied uses and multiple ways of implementation. The methods studied here are used to tackle time-series prediction problems and these methods are namely

1. LSTM Network for Regression
2. LSTM for Regression with Time Steps
3. LSTM with Memory Between Batches
4. Stacked LSTMs with Memory Between Batches

A generic LSTM unit has three gates that regulate the flow of information within the unit. These gates are called input, output and forget. All the models had the dataset split into training and testing datasets. Two-thirds of the data was assigned to train the models and the remaining one-third was used to test the models. All the models were trained for both 100 and 1000 epochs.

3.1. LSTM Network for Regression

The network has three layers with the visible layer having one input. The hidden block was made up of 4 LSTM units and the output layer produced a single value prediction. The data from the dataset is then fit into the model and from this the performance of the train and test datasets can be estimated. After this, the model is used to make predictions on both the train and test datasets and from that, the visual skill of the model can be identified.

Fig. 3(a) indicates the PM2.5 values against time for 100 epochs. Green indicates the training dataset and red indicates the testing plot. The RMSE values obtained indicated that the model has an average

error of 0.1552×10^{-4} mg/L for the training dataset and 0.1289×10^{-4} mg/L for the testing dataset. The R^2 values obtained were 0.77 and 0.67 for the training and testing datasets, respectively. Fig. 3(b) shows the LSTM trained on regression for the dataset and the comparison of predicted values (blue) vs the training and testing datasets.

Fig. 4(a) indicates the PM2.5 values against time for 1000 epochs. Green indicates the training dataset and red indicates the testing plot. The RMSE values obtained indicated that the model has an average error of 0.1553×10^{-4} mg/L for the training dataset and 0.1276×10^{-4} mg/L for the testing dataset. The R^2 values obtained were 0.77 and 0.68 for the training and testing datasets, respectively. Fig. 4(b) shows the LSTM trained on regression for the dataset and the comparison of predicted values (blue) vs the training and testing datasets. It can be inferred that running for 100 or 1,000 epochs doesn't create any major differences in results and the model has done a good job in fitting the model for both the training and testing datasets.

3.2. LSTM for Regression with Time Steps

Time steps can be used as inputs to predict the output at the next step. They provide another method in tackling the time series problem. Any point of failure or surge and the conditions that lead up to them are the features that define a time step.

Fig. S4(a) indicates the PM2.5 values against time. Green indicates the training dataset and red indicates the testing plot. The RMSE values obtained indicated that the model has an average error of 0.15×10^{-4} mg/L for the training dataset and 0.1329×10^{-4} mg/L for the testing dataset. The R^2 values obtained were 0.79 and 0.65 for the training and testing datasets, respectively. Fig. S4(b) shows the LSTM trained on regression for the dataset and the comparison of predicted values (blue) vs the training and testing datasets.

Fig. S5(a) indicates the PM2.5 values against time for 1,000 epochs. Green indicates the training dataset and red indicates the testing plot. The RMSE values obtained indicated that the model has an average error of 0.1483×10^{-4} mg/L for the training dataset and 0.1394×10^{-4} mg/L for the testing dataset. The R^2 values obtained were 0.79 and 0.62 for the training and testing datasets, respectively. Fig. S5(b) shows the LSTM trained on regression for the dataset and the comparison of predicted values (blue) vs the training and testing datasets. It can be inferred that running for 100 or 1,000 epochs doesn't create any major differences in results and the model has done a good job in fitting the model for both the training and testing datasets.

3.3. LSTM with Memory Between Batches

LSTM in Python is executed through the Keras deep learning library and the library supports both stateless and stateful LSTMs. The stateful LSTMs provide finer control over the internal state of the LSTM and when the internal state of the LSTM is reset. This can be used to make predictions to by building state over the entire training sequence.

Fig. S6(a) indicates the PM2.5 values against time. Green indicates the training dataset and red indicates the testing plot. The RMSE values obtained indicated that the model has an average error of 0.1602×10^{-4} mg/L for the training dataset and 0.1653×10^{-4} mg/L for the testing dataset. The R^2 values obtained were 0.76 and 0.46 for the training and testing datasets, respectively. Fig. S6(b) shows the LSTM trained on regression for the dataset and the comparison of predicted values (blue) vs the training and testing datasets.

Fig. S7(a) indicates the PM2.5 values against time for 1,000 epochs. Green indicates the training dataset and red indicates the testing plot. The RMSE values obtained indicated that the model has an average error of 0.1582×10^{-4} mg/L for the training dataset and 0.1648×10^{-4} mg/L for the testing dataset. The R^2 values obtained were 0.76 and 0.46 for the training and testing datasets, respectively. Fig. S7(b) shows the LSTM trained on regression for the dataset and the comparison of predicted values (blue) vs the training and testing datasets. It can be inferred that running for 100 or 1,000 epochs doesn't create any major differences in results and the model has done a good job in fitting the model for both the training and testing datasets.

3.4 Stacked LSTMs with Memory Between Batches

Stacked LSTMs are an extension of normal LSTMs which have a single hidden layer. Thereby, stacked LSTMs have multiple hidden layers with multiple memory cells. Stacking LSTM layers make the model deeper and thus justify the usage of the term deep learning.

Fig. S8(a) indicates the PM2.5 values against time. Green indicates the training dataset and red indicates the testing plot. The RMSE values obtained indicated that the model has an average error of 0.1597×10^{-4} mg/L for the training dataset and 0.1724×10^{-4} mg/L for the testing dataset. The R^2 values obtained were 0.76 and 0.40 for the training and testing datasets, respectively. Fig. S8(b) shows the LSTM trained on regression for the dataset and the comparison of predicted values (blue) vs the training and testing datasets.

Fig. S9(a) indicates the PM2.5 values against time for 1000 epochs. Green indicates the training dataset and red indicates the testing plot. The RMSE values obtained indicated that the model has an average error of 0.1595×10^{-4} mg/L for the training dataset and 0.1717×10^{-4} mg/L for the testing dataset. The R^2 values obtained were 0.76 and 0.42 for the training and testing datasets, respectively. Fig. S9(b) shows the LSTM trained on regression for the dataset and the comparison of predicted values (blue) vs the training and testing datasets. It can be inferred that running for 100 or 1,000 epochs doesn't create any major differences in results and the model has done a good job in fitting the model for both the training and testing datasets.

4. Results And Discussions

The collected dataset is divided into two parts: two-third of the data was used to train the model, and the remaining one-third of the data was used to test the performance of the developed models when benchmarking with others. The root-mean-square error (RMSE), and coefficient of determination (R^2) were used to evaluate the performance of the different models presented in this paper.

Four different variations of the LSTM models were compared and tested based on performance. All four models were trained and tested using the same datasets, the PM2.5 mass concentrations from the three air quality monitoring stations in the city of Chennai were present in the dataset and were predicted to evaluate performance. The dataset was cleaned to remove any incomplete data or data that exceeded 2.5×10^{-4} mg/L in providing more uniform predictions and results.

The comparison of the prediction results from the four models are made in terms of RMSE and R^2 in Tables 2 and 3. The time to predict is taken to be 15 min and most prior researchers have not used such a short time to predict for their research. All four DL models used have similar RMSE values for the given dataset. The R^2 seems the same as well for the training dataset, but when it comes to the testing dataset, there appears to be a decrease in R^2 . This difference in R^2 values for the training and testing dataset can be indicative of overfitting by the models.

All the models were trained for both 100 and 1,000 epochs and show very similar results in both cases. The results both in terms of RMSE and R^2 values are very similar and this can be attributed to the fact that any model should be trained till it reaches the minimum error rate and not after that as it may cause overfitting in the model. Also, from the results, it can be said that these are very suitable for predicting urban PM2.5 concentration in the future.

Nonetheless, the research has some drawbacks as emissions have a huge effect on air quality. As emission data are hard to acquire, the data obtained in this paper does not contain emissions from factories and vehicles in the region. This affects the accuracy of the model's prediction. Also, when a sudden increase in pollution because certain accidents occur, the concentration of PM2.5 changes suddenly. Whether the proposed model will forecast this well still needs to be shown.

5. Conclusions

All the models were developed to predict PM2.5 concentrations with the LSTM model that used regression with time steps showing the best results for 100 epochs of training. All four models produce very similar RMSE values for both training and testing datasets. The least difference in RMSE values was in the LSTM with Memory Between Batches variation. While the least values training and testing RMSE values were observed in the LSTM for Regression with Time Steps and LSTM Network for Regression respectively. The R^2 values for these were consistent for the training dataset but varied wildly for the testing dataset. Per this, it can be concluded that the LSTM Network for Regression produced the best results as there was little to no overfitting seen in the model. While these provide some insight onto which models might be appropriate for prediction of PM2.5 values, all of these were trained for 100 and 1,000

epochs with little to no variation in results in terms of accuracy of predictions. Also, there is a necessity to introduce more statistical analysis techniques as well as the introduction of more exploratory variables will improve the model's performance and open new avenues to study new exploratory variables and methods to analyze them. The development of such models is very useful to the city of Chennai, which plays a role as a vital industrial centre and region of economic importance. The models can be used to identify factors that affect air pollution within the city of Chennai and thereby reduce the levels of pollution as well as the impact of air pollution on the inhabitants of the city. The models can also be expanded to different cities around India and the world and thereby improve the quality of life of people around the world.

Declarations

Acknowledgments

The authors would like to thank the government of India and the Central Pollution Control Board for providing access to the data collected and stored in their website.

Author Contributions

R. B (Associate Professor) contributed to problem identification, literature collection and data analysis. P. P (Professor) involved in problem identification, literature collection, data collection, manuscript preparation. P. N. E (B. Tech Student) contributed to data collection, data cleaning, data analysis, model development, interpretation of results. R. R. H (Professor) contributed to data analysis, manuscript preparation.

References

1. Balakrishnan, K. *et al.* The impact of air pollution on deaths, disease burden, and life expectancy across the states of India: the Global Burden of Disease Study 2017. *Lancet Planet. Heal.***3**, e26–e39 (2019).
2. Smith, K. R. Managing the Risk Transition. *Toxicol. Ind. Health***7**, 319–327 (1991).
3. Lippmann, M. Environmental Toxicants: Human Exposures and Their Health Effects, 3rd Edition. *Chromatographia***71**, 555–555 (2010).
4. Di, Q. *et al.* An ensemble-based model of PM_{2.5} concentration across the contiguous United States with high spatiotemporal resolution. *Environ. Int.***130**, 104909 (2019).
5. Huang, C.-J. & Kuo, P.-H. A Deep CNN-LSTM Model for Particulate Matter (PM_{2.5}) Forecasting in Smart Cities. *Sensors***18**, 2220 (2018).
6. Karimian, H. *et al.* Evaluation of Different Machine Learning Approaches to Forecasting PM_{2.5} Mass Concentrations. *Aerosol Air Qual. Res.***19**, 1400–1410 (2019).

7. Li, S. *et al.* Urban PM2.5 Concentration Prediction via Attention-Based CNN–LSTM. *Appl. Sci.***10**, 1953 (2020).
8. Gorham, K. A., Raffuse, S. M., Hyslop, N. P. & White, W. H. Comparison of recent speciated PM2.5 data from collocated CSN and IMPROVE measurements. *Atmos. Environ.***244**, 117977 (2021).
9. Magazzino, C., Mele, M. & Sarkodie, S. A. The nexus between COVID-19 deaths, air pollution and economic growth in New York state: Evidence from Deep Machine Learning. *J. Environ. Manage.***286**, 112241 (2021).
10. Habeebullah, T. M., Abd El-Rahim, I. H. A. & Morsy, E. A. Impact of Outdoor and Indoor Meteorological Conditions on the COVID-19 transmission in the Western Region of Saudi Arabia. *J. Environ. Manage.***184**, 112392 (2021).
11. Niu, Y., Wang, F., Liu, S. & Zhang, W. Source analysis of heavy metal elements of PM2.5 in canteen in a university in winter. *Atmos. Environ.***244**, 117879 (2021).
12. Li, X. *et al.* Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environ. Pollut.***231**, 997–1004 (2017).
13. Bansal, M., Aggarwal, A. & Verma, T. Air Quality Index Prediction of Delhi using LSTM. *Int. J. Emerg. Trends Technol. Comput. Sci.***8**, 59–68 (2019).
14. Yang, M., Fan, H. & Zhao, K. PM2.5 Prediction with a Novel Multi-Step-Ahead Forecasting Model Based on Dynamic Wind Field Distance. *Int. J. Environ. Res. Public Health***16**, 4482 (2019).
15. Wu, X., Wang, Y., He, S. & Wu, Z. PM2.5/PM10 ratio prediction based on a long short-term memory neural network in Wuhan, China. *Geosci. Model Dev.***13**, 1499–1511 (2020).
16. Kim, H. S. *et al.* Development of a daily PM 10 and PM 2.5 prediction system using a deep long short-term memory neural network model. *Atmos. Chem. Phys.***19**, 12935–12951 (2019).
17. Zhang, S. *et al.* LSTM-based air quality predicted model for large cities in China. *Nat. Environ. Pollut. Technol.***19**, 229–236 (2020).
18. Kleine Deters, J., Zalakeviciute, R., Gonzalez, M. & Rybarczyk, Y. Modeling PM 2.5 Urban Pollution Using Machine Learning and Selected Meteorological Parameters. *J. Electr. Comput. Eng.***2017**, 1–14 (2017).
19. Soh, P.-W., Chang, J.-W. & Huang, J.-W. Adaptive Deep Learning-Based Air Quality Prediction Model Using the Most Relevant Spatial-Temporal Relations. *IEEE Access***6**, 38186–38199 (2018).
20. Xayasouk, T. & Lee, H. Air Pollution Prediction System Using Deep Learning. *WIT Trans. Ecol. Environ.***230**, 71–79 (2018).
21. Anurag, N. V., Burra, Y., Sharanya, S. & MG, G. Air Quality Index Prediction using Meteorological Data using Featured Based Weighted Xgboost. *Int. J. Innov. Technol. Explor. Eng.***8**, 1026–1029 (2019).
22. Sivacoumar, R. & Jayabalou, R. Assessment of source contribution to ambient air quality through comprehensive emission inventory, long-term monitoring and deterministic modeling. *Int. J. Environ. Sci. Technol.***16**, 2765–2782 (2019).
23. Hochreiter, S. & Schmidhuber, J. Long Short-Term Memory. *Neural Comput.***9**, 1735–1780 (1997).

Tables

Table 1 Statistical summary of the exploratory variables in the dataset

Statistical Measures	BP (Pa)	RH (%)	PM2.5 ($\mu\text{g}/\text{m}^3$)	WD (deg)	WS (m/s)
Mean	988.028	63.801	40.461	170.322	1.959
Standard Deviation	51.422	18.249	32.170	75.237	1.254
Minimum	751.120	4.450	0.010	0.070	0.020
Maximum	1043.600	100.000	249.680	359.960	10.330

Table 2 Comparison of RMSE and R^2 values across the four prediction models for 100 epochs

Type of LSTM Models	RMSE		R^2	
	Training	Testing	Training	Testing
LSTM Network for Regression	15.52	12.89	0.77	0.67
LSTM for Regression with Time Steps	15.00	13.29	0.79	0.65
LSTM with Memory Between Batches	16.02	16.53	0.76	0.46
Stacked LSTMs with Memory Between Batches	15.97	17.24	0.76	0.40

Table 3 Comparison of RMSE and R^2 values across the four prediction models for 1000 epochs

Type of LSTM Models	RMSE		R^2	
	Training	Testing	Training	Testing
LSTM Network for Regression	15.53	12.76	0.77	0.68
LSTM for Regression with Time Steps	14.83	13.94	0.79	0.62
LSTM with Memory Between Batches	15.82	16.48	0.76	0.46
Stacked LSTMs with Memory Between Batches	15.95	17.17	0.76	0.42

Figures



Figure 1

Location of air quality monitoring stations Note: The designations employed and the presentation of the material on this map do not imply the expression of any opinion whatsoever on the part of Research Square concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitation of its frontiers or boundaries. This map has been provided by the authors.



Figure 2

Heatmap of exploratory variables

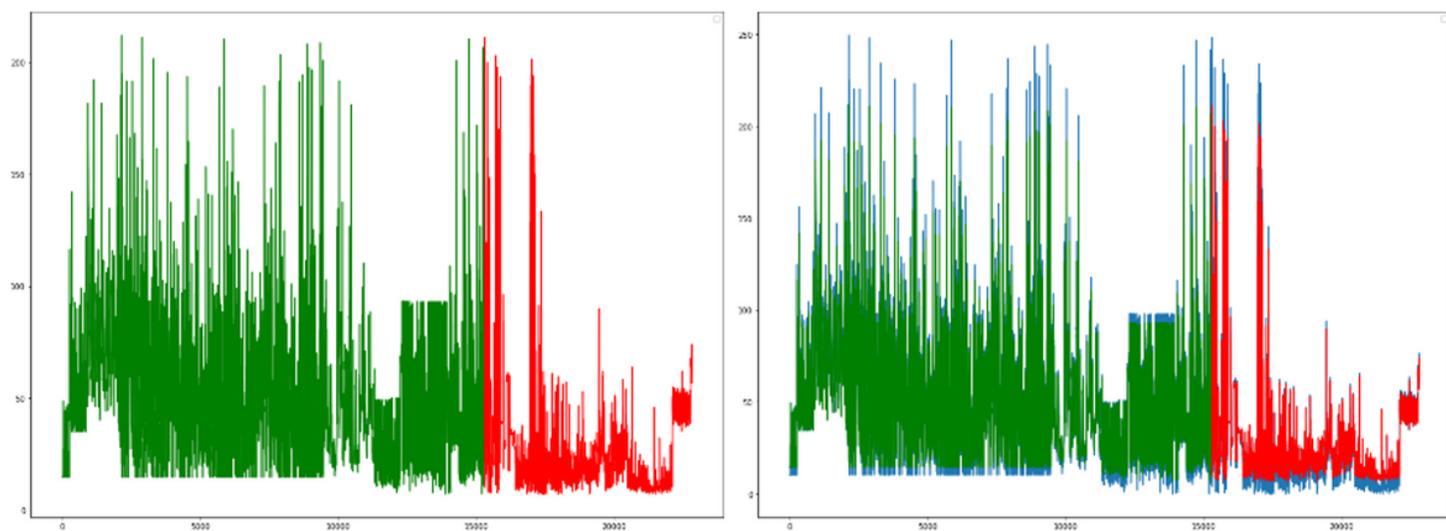


Figure 3

(a) PM2.5 values vs Time (b) Comparison of observed vs predicted PM 2.5 values trained for 100 epochs

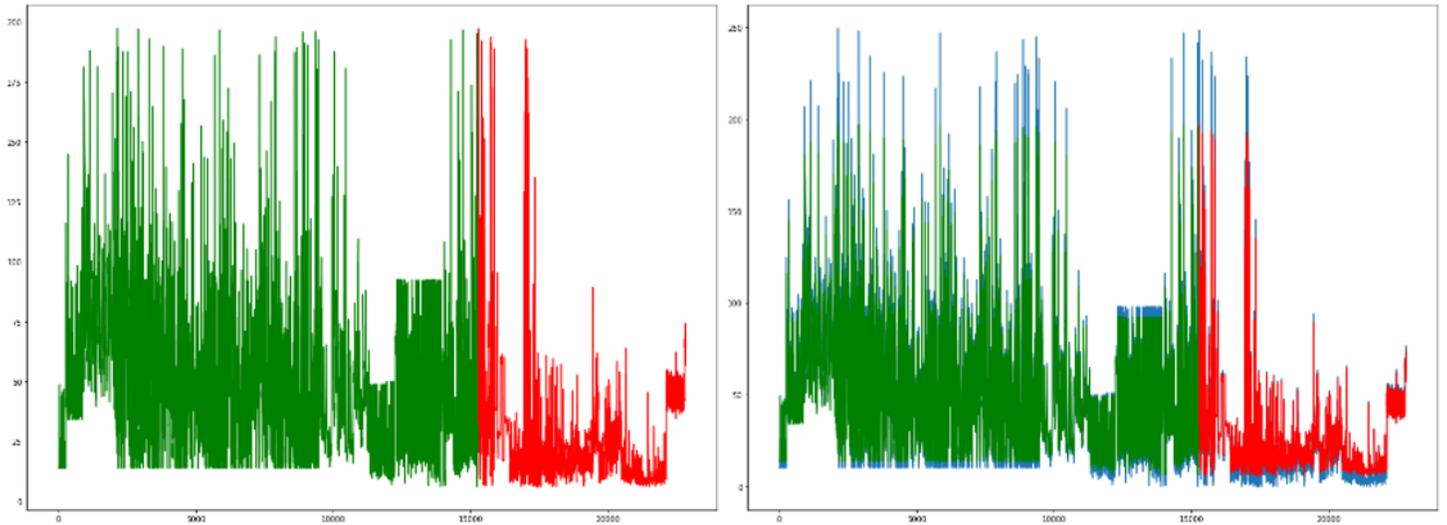


Figure 4

(a) PM2.5 values vs Time (b) Comparison of observed vs predicted PM 2.5 values trained for 1000 epochs

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SupplementaryMaterial.docx](#)