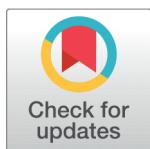




RESEARCH ARTICLE



OPEN ACCESS

Received: 08.08.2020

Accepted: 30.08.2020

Published: 05.10.2020

Editor: Dr. Natarajan Gajendran

Citation: Bamane P, Patil M (2020) A comparative study of different LSTM neural networks in predicting air pollutant concentrations. Indian Journal of Science and Technology 13(35): 3664-3674. <https://doi.org/10.17485/IJST/v13i35.1276>

*Corresponding author.

preetibamane04@gmail.com

Funding: None

Competing Interests: None

Copyright: © 2020 Bamane & Patil. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment ([iSee](#))

ISSN

Print: 0974-6846

Electronic: 0974-5645

A comparative study of different LSTM neural networks in predicting air pollutant concentrations

Preeti Bamane^{1*}, Mangal Patil²

¹ PG Scholar, Dept. of Electronics, Bharati Vidyapeeth Deemed to be University College of Engineering, Pune, Maharashtra, India

² Assistant Professor, Dept. of Electronics, Bharati Vidyapeeth Deemed to be University College of Engineering, Pune, Maharashtra, India

Abstract

Aim/Objective: This study aims to identify the key trends among different types of LSTM networks and their performance and usage for air pollutants ($PM_{2.5}$ and PM_{10}) concentrations prediction. **Methods:** In this study, the extensive research efforts were made for Particulate Matters (i.e., PM_{10} and $PM_{2.5}$) prediction using several LSTM networks, namely Vanilla, Stacked, and Bidirectional. These are trained and tested using air quality data, retrieved from the Central Pollution Control Board (CPCB) of the town Bawana, Delhi. Real-time hourly a data from 2018 to 2020 with nine air pollutants are considered for experimental analysis. We conducted data preparation strategy to select the best features, which improve the quality of the data. An adequate number of experiments are conducted to choose the best hyperparameters using Python package TensorFlow. **Findings:** MSE, MAE, RMSE, and R^2 parameters are used as the statistical criteria for evaluating the model's performances. The numerical experiments revealed that deep neural networks could predict the Particulate Matters ($\mu\text{g}/\text{m}^3$) with high accuracy. We found that Stacked LSTM with minimum MSE, MAE, RMSE, and maximum R^2 works better than the other two methods, i.e., Vanilla LSTM and Bidirectional LSTM for $PM_{2.5}$ and PM_{10} concentrations prediction. The empirical, experimental analysis also shows that Vanilla, Stacked, and Bidirectional LSTM models have comparatively minimum MSE, MAE, RMSE, and maximum R^2 for $PM_{2.5}$ than PM_{10} concentration prediction. **Applications:** With the help of a predictive model, one can find reliable fine concentration prediction information for a particular area. The resultant information on relative performance can help researchers in the selection of an appropriate LSTM algorithm for their studies.

Keywords: Air pollutants; air quality index; PM concentrations; LSTM; TensorFlow; air pollution

1 Introduction

According to World Health Organization data, in India, 1.5 million people died from chronic respiratory and asthma diseases caused by exposure to outdoor air pollution⁽¹⁾. In Delhi, due to the vast industrialization and urbanization, the utmost environmental concern is air pollution in terms of PM concentrations and pollutants consisting of a complex mixture of solid and liquid particles. Hazardous air pollutants are likely to inflame the respiratory tract, may even damage the blood and nervous system, and cause death. Therefore, air quality analysis and prediction have become an extensive area of research.

The government of India has installed pollutant measuring sensors to monitor air pollution, and these sensors regularly capture the air pollutants. The air quality prediction is highly significant for the governmental emergency departments and citizens to implement protective measures and promptly mitigate serious pollution incidents. Concurrently, environmental monitoring using pollutant level prediction is a useful technical medium in executing scientific decision-making for air pollution control and prevention. For instance, Badarpur coal-fired power plant in Delhi has been shut down every winter to reduce the pollution level due to the predicted air quality⁽²⁾.

The Air Quality Index (AQI) is a value used by government organizations, reflecting ambient air quality in the atmosphere for a particular area. According to Indian national ambient air quality standards, AQI is calculated from ambient concentration values of eight major pollutants: particulate matter 2.5 (PM_{2.5}), particulate matter 10 (PM₁₀), carbon monoxide (CO), nitrogen oxide (NO₂), sulfur dioxide (SO₂), ozone (O₃), ammonia (NH₃), and lead (Pb). The AQI computes the overall air quality on a scale with a range of 0 to 500. It is categorized into six remarks: good, satisfactory, moderate, poor, very poor, and severe; these remarks show the possible impact on public health^(3,4).

Different types of LSTM networks have been a dominant method in prediction or forecasting. The AQI and Air pollutant concentration prediction has recently shown a potential application area for these methods.

The stacked LSTM model is used to predict future air pollutant concentrations. The authors have considered features such as meteorological datasets, different factors, including festivals and national holiday traffic, etc. They have predicted pollutants (PM_{2.5}, PM₁₀, CO, NO₂, SO₂, O₃) for the next 1hr, next 6hr, and next 12hr for the major stations Delhi and Agra⁽⁵⁾. The LSTM neural network is implemented for the detection and prediction of PM 2.5 in⁽⁶⁾. The authors have collected data from the Environmental Protection Administration of Taiwan from the year 2012 to 2016. Several kinds of LSTM architectures were used in air pollutants and concentration prediction in the region of Madrid⁽⁷⁾.

The single-layer LSTM network predicted 12 air pollutants on the CPCB data at the Visakhapatnam, India. The authors have compared their results with various SVR kernels⁽⁸⁾. LSTM method is employed for AQI prediction on air pollutant factors and meteorological data⁽⁹⁾. Recurrent Neural Network (RNN), Gated Recurrent Unit (GRU), and LSTM methods are used to forecast the PM₁₀ on time series AirNet data. The paper also analyses predictive performance⁽¹⁰⁾.

The Bi-directional LSTM is constructed to predict air pollutant PM_{2.5}. To increase the accuracy of the model, the authors have included weather information while training the dataset. They have predicted PM_{2.5} for the next 6, 12, and 24 hours at the multiple stations in New Delhi⁽¹¹⁾. Bidirectional LSTM RNN applied to the spatio-temporal interpolation domain to predict PM_{2.5} concentration. The authors have considered daily PM_{2.5} measurements for 2009 in the southeast region of the USA⁽¹²⁾. The air quality and meteorological information, real time-series data, were considered for the pollutant prediction using the RNN LSTM algorithm. Authors have predicted PM_{2.5} for the next 8h, 12h, 16h, 20h, and 24h in South Korea for one year⁽¹³⁾.

The PM concentrations (PM_{2.5} and PM₁₀) predicted using LSTM and deep autoencoder (DAE) methods. The air quality data were collected from 25 stations in Seoul, South Korea, for over three years. Authors have compared the algorithms' performance using different hyperparameters such as learning rate, epochs, and batch size⁽¹⁴⁾. The denoising auto encoder deep network (DAEDN) model is based on LSTM designed and constructed for PM_{2.5} concentration prediction. The five years of air quality data in the Beijing's city were considered for experimental analysis, and the proposed model achieved RMSE and MAE of 15.504 and 6.789, respectively⁽¹⁵⁾. The long short-term memory neural network extended (LSTME) model is proposed to predict air pollutant concentration. The experiment was carried out on hourly air quality data, gathered from 12 monitoring stations in Beijing City over two years and achieved 11.93% of MAPE⁽¹⁶⁾.

Out of several mentioned pollutants, Particulate Matter PM₁₀ and PM_{2.5} are of interest: their diameter is less than 10 and 2.5 micrometers; it can quickly enter the human respiratory system. Hence, the prediction of PM₁₀ and PM_{2.5} plays a vital role in determining air quality. The most cutting-edge deep learning technology is used, RNN based LSTM, which has nonlinear problems processing ability and noise tolerance ability, improving forecasting efficiency and accuracy. Therefore, this method has high theoretical research value. The LSTM network is divided into three basic categories Vanilla, Stacked, and Bidirectional. This study's main objective is to compare the performance of these LSTM networks in terms of air quality prediction and identify efficient ones among them. A comparison is only possible when a standard benchmark on the dataset and the scope is established. Therefore, we chose identical hyperparameters that were implemented on the same data for comparison. We conducted a series of experiments to validate the intended approaches and present evidence to demonstrate the proposed

framework's effectiveness using evaluation parameters.

2 Data description

The dataset used in this study is collected from the monitoring station at Bawana, Delhi, for the period of 4th July 2018 to 30th June 2020⁽¹⁷⁾. The most polluting industrial units operating in Delhi's residential area were relocated to Bawana, North West Delhi; thus, air condition became relatively weak at Bawana. Therefore, air quality analysis is important in the region of Bawana. The dataset contains 17126 instances of hourly pollutants for PM₁₀, PM_{2.5}, NO, NO₂, NOx, NH₃, SO₂, and O₃, provided by the certified analyzer and electronic instruments measured in $\mu\text{g}/\text{m}^3$.

3 Data pre-processing

In the process of conducting air quality prediction on the data, the preprocessing stage plays a vital role⁽¹⁸⁾. Due to some uncontrollable factors, obtained data lacked a certain values, such as timestamps and monitoring data. Figure 1 reports the missing values with the percentage loss present in real-time data. Each factor of the missing values is computed out of 17126 data, and percentage loss is estimated. In the dataset mentioned in Figure 1, the number of missed PM₁₀ parameter, as an essential factor, reaches nearly fourteen hundred, which has a significant impact on the effectiveness of the model. Therefore, constructing and designing reasonable and efficient data is a prerequisite to reduce data noise. It is also a widespread research issue in data science⁽¹⁹⁾.

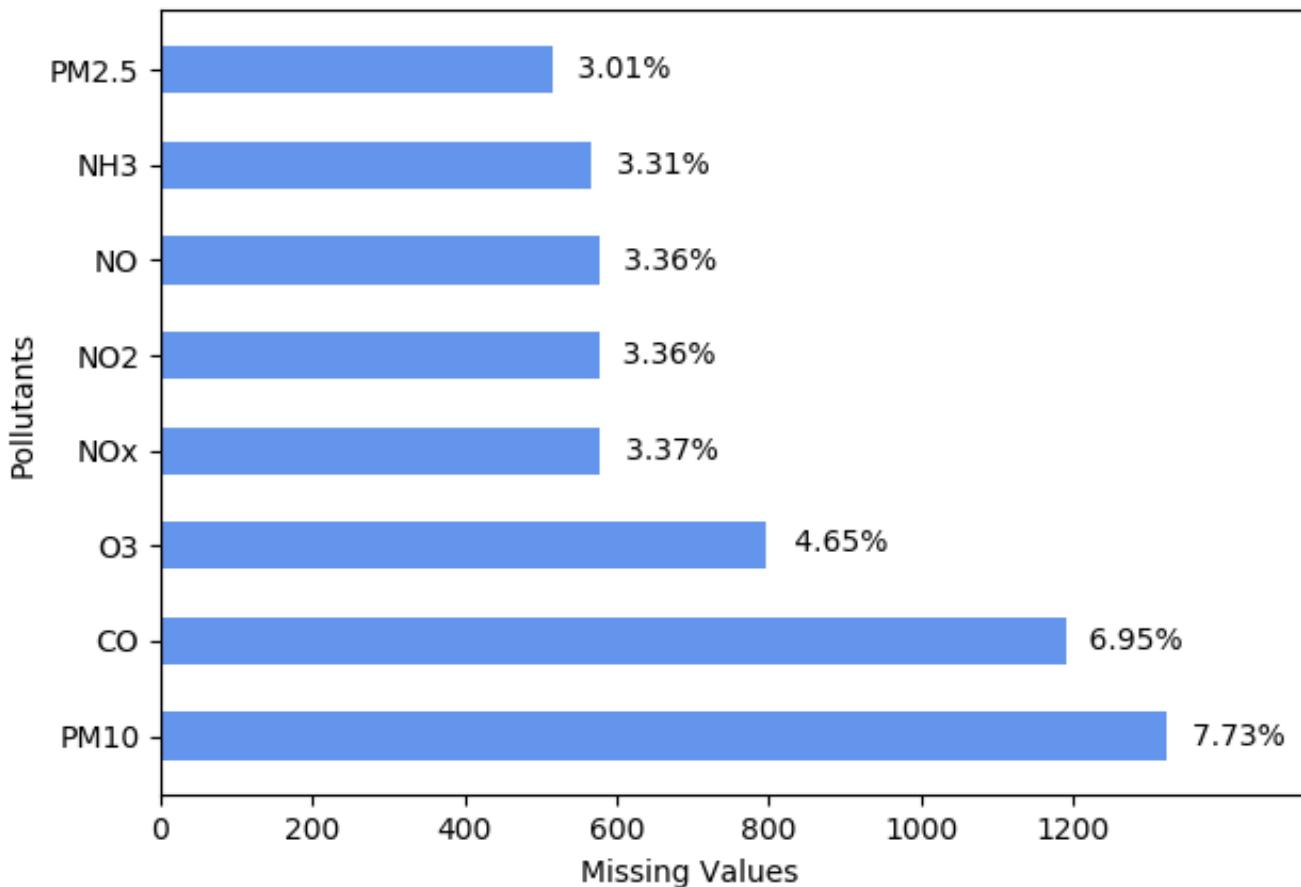


Fig 1. Situation of Missing Data

3.1 Data imputation

In dealing with the data-missing problem, the linear interpolation method is adopted. The missing values in the data field are imputed by interpolating the respective parameter. Equation (1) formulates the linear interpolation as,

$$X_t = \frac{X_s(X_e + X_s)}{n+1} \quad (1)$$

Where X_t indicates the missing value at time t and n indicates time interval between $(X_e + X_s)$.

3.2 Data statistics

Provides descriptive statistics of each factor in the dataset of the city Bawana Delhi, including the statistical properties such as minimum, maximum, and average values. As shown in the Table 1, PM_{2.5} concentration varies from 3 to 991 ($\mu\text{g}/\text{m}^3$), with an average of 121.78 and a maximum of 991. Similarly, the PM₁₀ concentration varies from 2.75 to 995 ($\mu\text{g}/\text{m}^3$), with an average of 235.09 and a maximum of 995. That is the existence of noisy data in the air quality dataset; therefore, data cleaning is required.

Table 1. Statistical table of air quality dataset

	PM _{2.5}	PM ₁₀	NO	NO ₂	NO _x	NH ₃	CO	O ₃	SO ₂
count	17126	17126	17126	17126	17126	17126	17126	17126	17126
mean	121.78	235.09	10.46	37.93	28.69	38.73	1.25	14.85	45.17
std	107.30	161.03	23.31	29.56	29.05	17.11	0.81	8.71	44.38
min	3	2.75	0.2	0.22	1.03	0	0	0.1	0.1
25%	48	117	1.32	16.25	10.28	26.6	0.75	9.28	10.68
50%	88	204	2.5	30.12	18.9	35.62	1.1	13.68	28.72
75%	156.5	303	7.92	50.85	35.3	46.18	1.52	18.05	65.9
max	991	995	318.62	276.88	302.6	138.7	9.53	119	199.9

3.3 Data normalization

When the scale of each factor is different, it could affect the training process, resulting in a decreasing accuracy. Therefore, to facilitate network-training processes and prevent “overfitting” problems, each factor’s original data needed to be normalized on the same scale for better performance. We used Min-Max Normalization to scale the values in the range [0, 1]:

$$\bar{X} = \frac{X - X_{\min}}{X_{\max} - X_{\min}} \quad (2)$$

4 Feature selection

This paper uses spearman’s rank-order correlation to find a monotonic relationship between the Bawana dataset features. The valid range of correlation coefficient is from -1 to +1. The -1 and +1 indicate a strong negative, and positive correlation, respectively, whereas 0 indicates a negligible correlation ⁽²⁰⁾. The formula for spearman’s rank-order correlation is

$$\rho_{r_x r_y} = \frac{\text{cov}(r_x r_y)}{\sigma_{r_x} \sigma_{r_y}}$$

Where r_x , r_y denotes the ranked values of variable x and y, COV stands for covariance and σ indicates standard deviation.

This study refers to Cohen’s standard ⁽²¹⁾ for feature selection. The feature correlation values less than 0.30 are considered weak, hence removed from further processing. Based on investigating in Figure 2, PM₁₀ has a strong positive correlation with PM_{2.5}, moderate positive correlation with NO, NO₂, NO_x, NH₃, and CO, and moderate negative correlation with O₃. Likewise, PM_{2.5} has a strong positive correlation with PM₁₀, and moderate positive correlation with NO, NO₂, NO_x, NH₃, SO₂, and CO. According to Cohen’s standard, SO₂ with 0.29 and O₃ with -0.2 are discarded from the dataset while predicting PM_{2.5} and PM₁₀ concentrations.

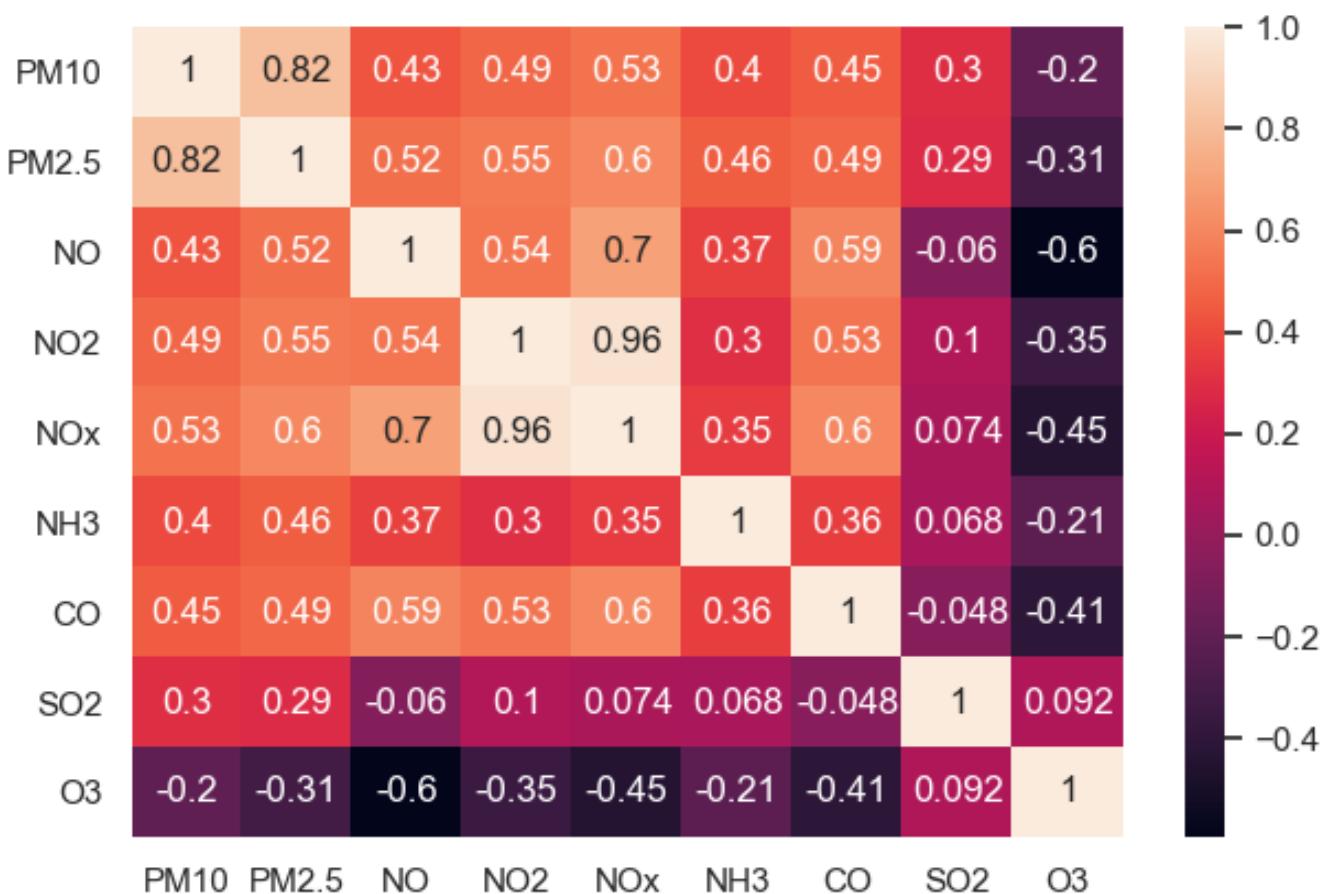


Fig 2. Heat map of the correlation coefficient

5 LSTM neural networks

The ability to memorize sequences of data makes the LSTM a special kind of RNNs capable of catching data from past events and using it for future predictions^{(22), (23)}. In an LSTM, the memorization of earlier stages performed through gates and incorporated memory lines. The basic LSTM unit consists of three filtering gates: the input gate, forget gate, and output gate, which controls the flow of information. This study formulates the problem PM_{2.5} and PM₁₀ concentrations prediction as a time series based problem. The fine pollutant concentrations at time t+1 not only take input data at time t but also take predicted values at time t-1, t-2,..., t-N, where N is the memory length. Because in the approach we have considered, the current pollutant level depends on past pollutant levels. This paper predicts fine PM concentrations using three basic types of LSTM neural networks described as follows:

5.1 Vanilla LSTM

A single layer LSTM network is commonly known as Vanilla LSTM. Vanilla LSTM consists of a single LSTM unit, fully connected hidden layer, and output layer to make a prediction. In this study, vanilla LSTM has a visible layer with one input, a hidden layer with 75 units, and an output layer that produces single value prediction.

5.2 Stacked LSTM

In the multiple hidden layers, each layer comprises the number of memory cells stacked on the top of each other and is called Stacked LSTM. This study stacked two LSTM layers on top of another, where the first LSTM layer returns their whole output sequences, although the second LSTM layer returns the last step its output sequence only.

5.3 Bidirectional LSTM

The Bidirectional LSTM can simultaneously run past and future information in both the direction, i.e., forward and backward. Two hidden states can preserve information from past and future and then join the corresponding output of hidden state to the same output state. Due to the positive and negative storage of time series data, unidirectional LSTM lag problems occurring while training the network can overcome. Consequently, it can provide additional information to the network and improves the correctness of the predictions.

6 Development of LSTM models

The experimental process framework is shown in [Figure 3](#) for predicting PM concentrations using different LSTM neural networks. The process workflow included data processing that processes missing values and generates data for training and testing the model. Using generated input data we trained and evaluated the models. MSE, MAE, RMSE, and R^2 are used to assess the prediction effectiveness of the LSTM networks.

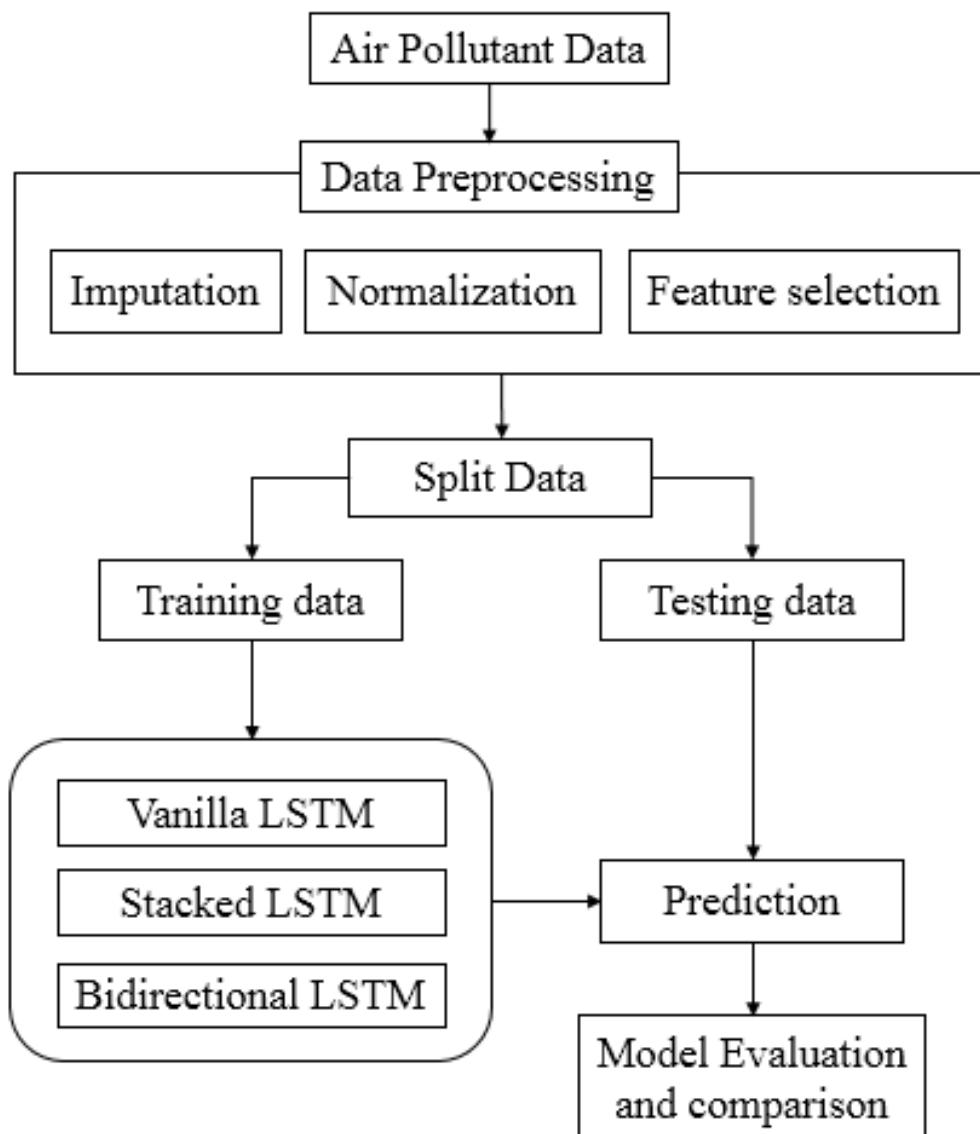


Fig 3. Experimental workflow

In order to develop the PM_{2.5} and PM₁₀ prediction models, the dataset is separated into two sets. Specifically, 1/4th data of the original dataset (12845 samples) is used to learn the model, called the training dataset; 3/4th remaining of the data (4281 samples) is used to justify the model, called the testing dataset.

For simplicity, the number of units or neurons in each LSTM layer taken as 64, dropout of 0.3 used to avoid over-fitting. Moreover, fine-tune hyperparameters are selected, such as Nesterov Adam (Nadam) optimizer, which is a combination of NAG (Nesterov Accelerated Gradient) and Adam⁽²⁴⁾ with a learning rate of 0.001, and loss function is set to mean squared error. The activation function is set to sigmoid, which squeezes a huge input space into a small input space, i.e., 0 and 1. The epochs and batch size are selected for model training chosen from a candidate set of {25, 50, 100, and 200} and {25, 75, 125}, respectively. The more the number of iterations, the more accurate prediction result is, but processing time increases. The most appropriate settings are chosen that yielded the best performance based on several comparative experiments. When the epochs and batch sizes are 100 and 75, respectively, the LSTM model achieved the best performance. Therefore, in this study, epochs and batch sizes are set as 100 and 75, respectively. Finally, the input data reshaped into three-dimension for the architecture of the LSTM network.

The LSTM models are implemented using Python programming language and its packages, namely Pandas, SciKit-Learn, Matplotlib, etc. The LSTM models are built and trained using deep learning libraries called TensorFlow and Keras. All the LSTM algorithms are designed using the same hyperparameters and verified with the same operation environment Windows 8, Intel(R) Core(TM) i3-3110M, CPU @2.40 GHz, 6 GB RAM.

7 Evaluation Parameters

To validate the model's effectiveness, we chose four standard metrics to compare the Vanilla LSTM, Stacked LSTM, and Bidirectional LSTM. The prediction accuracy is determined by the functions Mean Square Error (MSE), and Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and R square (R²) between measured and predicted values. The following acceptable criteria are used for evaluation: (1) low MSE, (2) low MAE, (3) low RMSE, and (4) high R². The mathematical expressions of these functions are given below,

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2 \quad (4)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - y'_i| \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - y'_i)^2} \quad (6)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - y'_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (7)$$

Where y_i denotes the actual values at i^{th} samples and \bar{y}_i denotes the mean of actual values at i^{th} samples, y'_i denotes the predicted values at i^{th} samples and n is the number of samples.

8 Results and discussion

After developing the models, their performance is compared and analyzed using evaluation parameters. Based on the training dataset, the predictive LSTM models (i.e., Vanilla, Stacked, and Bidirectional) are developed as described above. The true and predicted PM_{2.5} and PM₁₀ concentrations are plotted and visualized in Tables 2 and 3. The performance of the developed LSTM models in Table 2 showed that all the models could generate a reasonably accurate concentration estimation.

Table 3 shows the scattering of the actual and predicted values. It can be seen that the trend of the prediction curve follows the trend of the actual curve, and the predicted values have a positive linear relationship with the actual values.

Table 4 calculates indicators of the performance of LSTM techniques for estimating the PM_{2.5} and PM₁₀ concentrations in the training process. Table 5 sorts the ranking of LSTM models. Based on these results, we found that stacked LSTM achieves

the best performance in all evaluation metrics in predicting the concentration. With PM_{2.5} concentration prediction, Vanilla LSTM performed slightly better than Bidirectional LSTM. Whereas with PM₁₀ concentration prediction, Bidirectional LSTM performed marginally better than Vanilla LSTM.

Table 2. Comparison between LSTM networks using prediction performance graphs

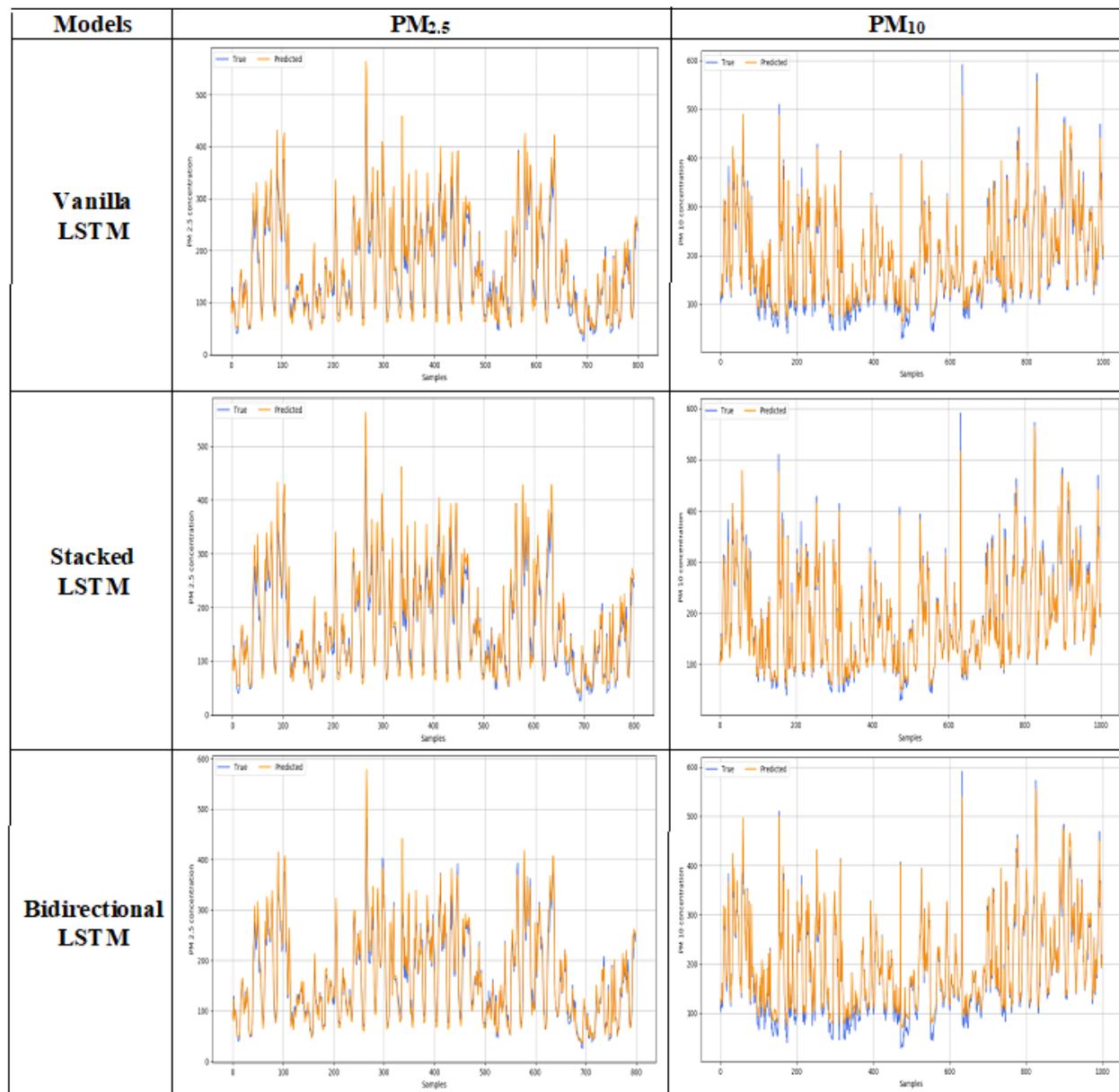
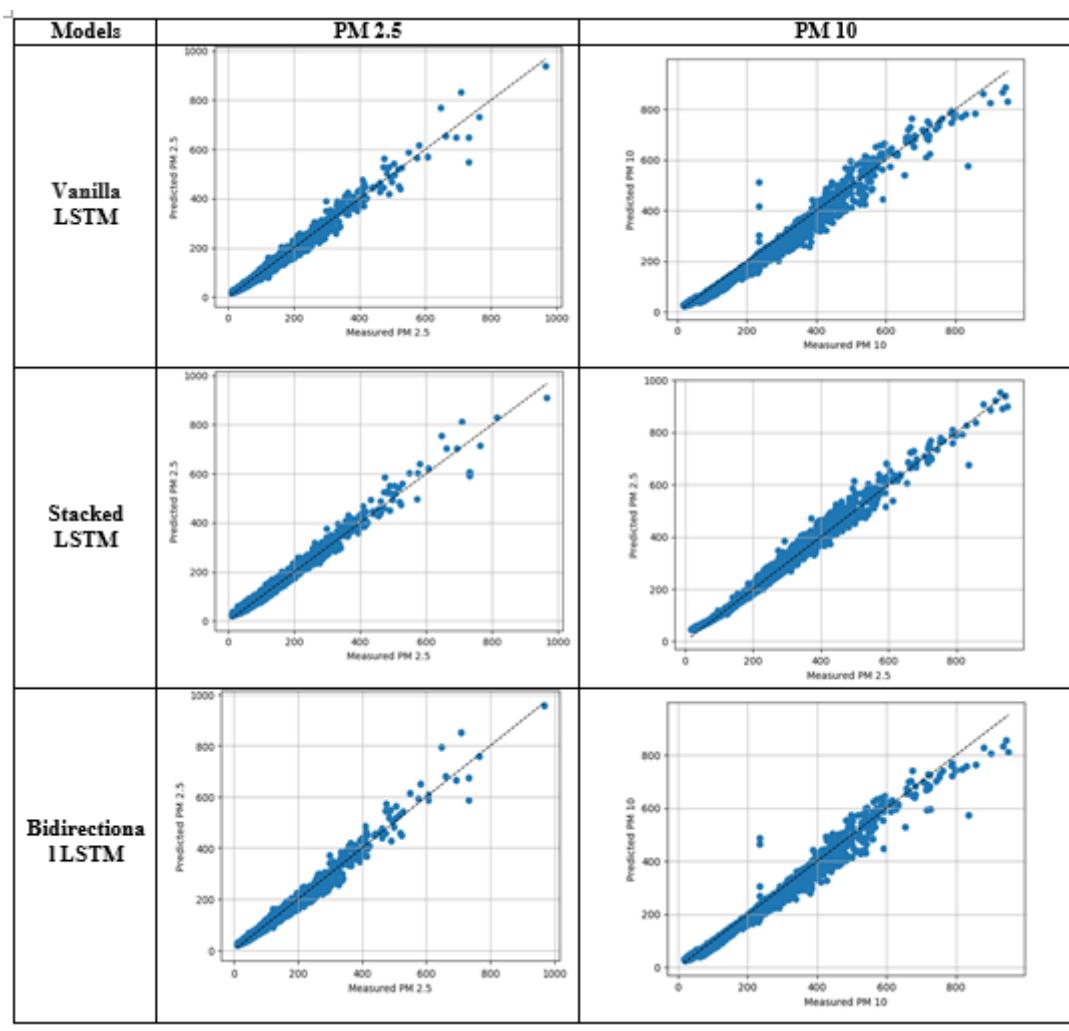


Table 3. Scatter plots of factual values and predicted values

**Table 4.** Comparison among models

Model	MSE		MAE		RMSE		R2	
	PM 2.5	PM 10	PM 2.5	PM 10	PM 2.5	PM 10	PM 2.5	PM 10
Vanilla LSTM	202.522	369.658	10.281	13.025	14.231	19.226	0.977	0.977
Stacked LSTM	173.659	203.727	8.557	8.998	13.178	14.273	0.980	0.987
Bidirectional LSTM	249.022	321.035	11.552	12.468	15.780	17.917	0.972	0.980

Table 5. Ranking of LSTM model based on statistical comparison

Model	Rank for MSE		Rank for MAE		Rank for RMSE		Rank for R2		Tol. Score	Ranking	Sort	
	PM _{2.5}	PM ₁₀			PM _{2.5}	PM ₁₀						
Vanilla LSTM	2	1	2	1	2	1	2	1	8	4	2	3
Stacked LSTM	3	3	3	3	3	3	3	3	12	12	1	1
Bidirectional 1 LSTM	2	1	2	1	1	2	1	2	4	8	3	2

9 Conclusion

This study focused on the comparative performance of different LSTM neural networks in air pollutant concentration prediction. We have drawn number of conclusions as follows: (1) we have handled a problem of multivariate variables and approximate nonlinear mapping very well. (2) Stacked LSTM with minimum MSE, MAE, RMSE and maximum R² works better than other networks for PM_{2.5} and PM₁₀ concentrations prediction. (3) Each LSTM model has comparatively minimum MSE, MAE, RMSE, and maximum R² for PM_{2.5} than PM₁₀ concentration. (4) Feature normalization and feature selection play a major role in making better air quality prediction. This study is limited to predict PM_{2.5} and PM₁₀; as long term prediction task is tedious. Though each LSTM model performs well and has high precision, and strong adaptive ability, huge data in terms of traffic, industry, and meteorological need to be explored in order to build a more comprehensive and effective air quality prediction.

Acknowledgement

The authors would like to thank Indian government, Central Pollution Control Board (CPCB) for providing air quality indices data.

Conflicts of Interest

The authors do not have any conflicts of interest.

References

- 1) Chatterjee P. India takes steps to curb air pollution. *Bulletin of the World Health Organization*. 2016;94(7):481–556. Available from: <http://dx.doi.org/10.2471/BLT.16.2020716>.
- 2) Patel P. Tackling Delhi's Air Pollution Problem. *ACS Cent Sci*. 2019;5:3–6. Available from: <https://doi.org/10.1021/acscentsci.9b00009>.
- 3) Pant P, Lal MR, Guttikunda KS, Russell GA, Nagpure SA, Ramaswami A, et al. Monitoring particulate matter in India: recent trends and future outlook. *Air Quality, Atmosphere & Health*. 2019;12(1):45–58. Available from: <https://dx.doi.org/10.1007/s11869-018-0629-6>.
- 4) Harsh M, Khushboo J. Air Pollution Problem in Delhi. *Journal of Critical Reviews*. 2020;7(10):723–727.
- 5) Chaudhary V, Deshbhrata A, Kumar V, Paul D. Time series based LSTM model to predict air pollutant's concentration for prominent cities in India. and others, editor. 2018. Available from: http://www.philippe-fournier-viger.com/utility_mining_workshop_2018/PAPER1.pdf.
- 6) Tsai Y, Zeng Y, Chang Y. Air Pollution Forecasting Using RNN with LSTM. *IEEE 16th Intl Conf on Dependable, Autonomic and Secure Computing, 16th Intl Conf on Pervasive Intelligence and Computing, 4th Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress*. 2018;p. 1074–1079. Available from: <https://doi.org/10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00178>.
- 7) Navares R, Aznarte LJ. Predicting air quality with deep learning LSTM: Towards comprehensive models. *Ecological Informatics*. 2020;55. Available from: <https://dx.doi.org/10.1016/j.ecoinf.2019.101019>.
- 8) Rao KS, Devi GL, Ramesh N. Air Quality Prediction in Visakhapatnam with LSTM based Recurrent Neural Networks. *International Journal of Intelligent Systems and Applications*. 2019;11(2):18–24. Available from: <https://dx.doi.org/10.5815/ijisa.2019.02.03>.
- 9) Jiao Y, Wang Z, Zhang Y. Prediction of Air Quality Index Based on LSTM. In: IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). 2019;p. 17–20. Available from: <https://doi.org/10.1109/ITAIC.2019.8785602>.
- 10) Athira V, Geetha P, Vinayakumar R, Soman KP. DeepAirNet: Applying Recurrent Networks for Air Quality Prediction. *Procedia Computer Science*. 2018;132:1394–1403. Available from: <https://dx.doi.org/10.1016/j.procs.2018.05.068>.
- 11) Verma I, Ahuja R, Meischeri H, Dey L. Air Pollutant Severity Prediction Using Bi-Directional LSTM Network. *ACM International Conference on Web Intelligence (WI)*. 2018;p. 651–654. Available from: <https://doi.org/10.1109/WI2018.00-19>.
- 12) Tong W, Li L, Zhou X, Hamilton A, Zhang K. Deep learning PM2.5 concentrations with bidirectional LSTM RNN. *Air Quality, Atmosphere & Health*. 2019;12(4):411–423. Available from: <https://dx.doi.org/10.1007/s11869-018-0647-4>.
- 13) Bui TC, Le VD, Cha SK. A deep learning approach for forecasting air pollution in South Korea using LSTM. 2018. Available from: <http://www.arxiv.org/abs/1804.07891v3>.
- 14) Xayasaouk T, Lee H, Lee G. Air Pollution Prediction Using Long Short-Term Memory (LSTM) and Deep Autoencoder (DAE) Models. *Sustainability*. 2020;12. Available from: <https://dx.doi.org/10.3390/su12062570>.
- 15) Qiu Z. Quantitative Testing of micro-cracks by the MFL technique Based on GA-BP neural network. *International Journal of Manufacturing Research*. 2017;12. Available from: <https://dx.doi.org/10.1504/ijmr.2017.10005448>.
- 16) Li X, Peng L, Yao X. Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation. *Environmental pollution*. 2017;231:997–1004. Available from: <https://doi.org/10.1016/j.envpol.2017.08.114>.
- 17) Indian government, Central Pollution Control Board (CPCB). . Available from: <https://data.gov.in/catalog/historical-daily-ambient-air-quality-data>.
- 18) Dhankhar K, Solanki. A comprehensive review of tools and techniques for big data analysis. *International Journal of Emerging Trends in Engineering Research*. 2019;7(11):556–562. Available from: <https://doi.org/10.30534/ijeter/2019/257112019>.
- 19) Chang JC, Hanna SR. Air quality model performance evaluation. *Meteorology and Atmospheric Physics*. 2004;87(1-3):167–196. Available from: <https://dx.doi.org/10.1007/s00703-003-0070-7>.
- 20) McDonald JH. Handbook of Biological Statistics. Baltimore, MD. Sparky House Publishing. 2009.
- 21) Cohen J, Cohen P, West SG, Aiken LS. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences, Routledge. 2013. Available from: <https://doi.org/10.4324/9780203774441>.

- 22) FelixGers A, Schmidhuber J, Cummins F. Learning to Forget: Continual Prediction with LSTM. *Neural Computation*. 2000;12(10):2451–2471. Available from: <https://dx.doi.org/10.1162/089976600300015015>.
- 23) Hochreiter S, Schmidhuber J. Long Short-Term Memory. *Neural Computation*. 1997;9(8):1735–1780. Available from: <https://dx.doi.org/10.1162/neco.1997.9.8.1735>.
- 24) Dozat T. Incorporating Nesterov momentum into Adam. Learning Representations . 2016.