2019 IEEE Fifth International Conference on Big Data Computing Service and Applications (BigDataService)

Real Time Attention based Bidirectional Long Short-Term Memory Networks for Air Pollution Forecasting

Divyam Madaan*
Panjab University
divyam3897@gmail.com

Radhika Dua*
Panjab University
radhikadua1997@gmail.com

Prerana Mukherjee
IIIT Sricity
prerana.m@iiits.in

Brejesh Lall
IIT Delhi
brejesh@ee.iitd.ac.in

Abstract

Approximately 95% of the world's population live in places with unsafe air which leads to serious health hazards for people worldwide. Monitoring and preserving air quality has become one of the most essential activities in many industrial and urban areas today. In view of increasingly alarming environmental pollution problems due to rapid urbanization, ambient air pollution prediction is becoming extremely important. This paper presents an online real time air pollution prediction system at five prominent locations in Delhi utilizing the past historic air quality and meteorological data. We propose a novel end to end sequential modelling framework to predict the air quality by estimating the concentration levels for various pollutants (nitrogen dioxide (NO_2), particulate matter ($PM_{2.5}$ and PM_{10}) and thereby classifying the threat level for them in the next 24 hours. The problem of air pollution prediction system becomes extremely challenging in the presence of unreliable and missing data entries. We also quantify the variations in the pollution concentrations on seasonal basis. We exhaustively evaluate and demonstrate that the proposed approach outperforms several baseline methods for forecasting air pollutants concentration based on data sources obtained in Delhi.

Keywords

Air quality prediction, Real time Pollution forecasting, BiLSTM, attentional selection

I Introduction

One of the most important emerging environmental issues in developing countries is air pollution. Today, the air quality is exponentially degrading due to rampant burning of fossil fuels including factory combustibles, deforestation, vehicles emitting pollution etc. It affects the society, and endangers the survival of life on earth leading to several health hazards such as asthma, lung cancer etc and environmental hazards such as global warming, acid rains, eutrophication and many more.

Delhi, India's third largest city having 14 million inhabitants has led to an explosion in motor vehicle use along with a rising industrial base. The increasing level of pollutants in ambient air in 2016-2018 has deteriorated the air quality of Delhi at an alarming rate. According to World Health Organization (WHO), in 2018 India has 14 out of the 15 most polluted cities in the world which has become a major cause for skin and eye health conditions in the country. Air quality index (AQI) of Delhi is highly dependent on the seasonal variations and is generally reported to be moderate $(101-200)$ AQI level between January to September, and then it significantly deteriorates to seriously poor levels $(301-400)$, Severe$(401-500)$ or sometimes even hazardous(> 500) levels between the months of October to December, due to several factors which includes stubble burning, fire crackers burning during Diwali and cold weather. Even the permissible safe limits for various pollutants cross beyond acceptable levels. Assessments of air pollution in Delhi suggest that vehicular and industrial sources are by large part responsible for emissions of the major critical pollutants, viz SO_2, NO_2, $PM_{2.5}$, PM_{10} and CO. Among these vehicular sources dominate emissions of CO and NO_2. Other sources, primarily those from domestic activities and burning of agricultural wastes dominate emissions of particulate matter ($PM_{2.5}$, PM_{10}) and have a significant impact on human health.

All these above mentioned problems, motivate us to focus the current research study conducted in this paper on air quality monitoring and forecasting in Delhi regions. We show the geographical locations of the five regions (Rohini-North Delhi, Punjabi Bagh-West Delhi, Mandir Marg-Central Delhi, R.K Puram-South Delhi and Okhla-East Delhi) taken into consideration for the current study in Fig. 1. The establishment of a reasonable and accurate forecasting model is of prime importance in order to predict urban air pollution levels and the immediate source of pollution. There is also rising demand for predicting future air quality levels, which can pre-inform government's policy-making bodies to undertake tasks such as maintaining a check on private vehicles emitting exuding amount of harmful emissions, establishing air pollution control towers or anti-smog towers, providing medical attention to the people who suffer from serious respiratory disorders and providing them with free nasal-filters or masks.

Most of the prior work [11], [18], [20] centralize their research studies on the concentration levels of particulate matter such as $PM_{2.5}$ or PM_{10} utilizing the weather and air quality data.

*Equal Contribution

978-1-7281-0059-3/19/$31.00 ©2019 IEEE
DOI 10.1109/BigDataService.2019.00027

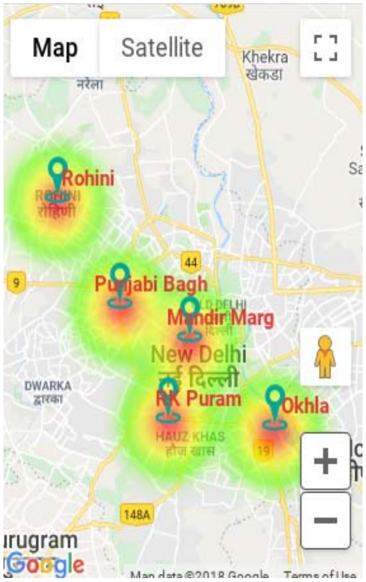

Fig. 1: Map showing air quality monitoring stations of Delhi

However, none of them take into account the serious implications of other pollutants towards the growing health hazards and predicting their concentration levels. We provide a novel holistic end-to-end model which can predict the concentration levels of various pollutants. We also critically analyze the changing trends observed in the level of these pollutants due to the seasonal variations. Due to continuous online training, the models are adaptive and are continuously deployed to take into account the changing trends that may enable to predict the future pollutant concentration values in a more synergistic manner. To the best of author's knowledge, this is the first work that highlights the strength of continuous adaptive learning in respect of air pollution forecasting taking into account both direct (historic air pollution data) and indirect factors (time, meteorological data) that reflect and influence the pollutant's concentration levels. We are able to achieve a significant performance gain of about $\sim 15-20\%$ in accuracy of the proposed system as compared to deploying the same offline trained model and of about $\sim 3-5\%$ in forecasting air pollutants concentration as compared to the reported baseline methods.

The organization of the paper is as follows. In Sec. II we discuss the related work in the area of traditional handcrafted models, sequential modelling methods (including deep learning), wireless network based methods and trend analysis based methods used for temporal forecasting. In Sec. III, we provide an outline of the proposed methodology. We detail the experimental results and analysis in Sec. IV followed by conclusion as discussed in Sec. V.

II Related Work

In this section we provide a detailed overview of the contemporary techniques prevalent in the domains which are closely related to our work and broadly categorize them into the following categories for temporal forecasting.

A. Handcrafted features based techniques

Traditional handcrafted feature based approaches [11] utilize Markovian approaches to predict $PM_{2.5}$ concentration using time series data. Since, Hidden Markov Models (HMMs) suffer from the limitation of short term memory thus failing in capturing the temporal dependencies in prediction problems. To overcome this, authors in [11] highlight a variant of HMMs: hidden semi-Markov models (HSMMs) for $PM_{2.5}$ concentration prediction by introducing temporal structures into standard HMMs. In [18], authors provide a hybrid framework which relies on feature reduction followed by a hyperplane based classifier (Least Square Support Vector Machine) to predict the particulate matter concentration in the ambient air. The generalization ability of the classifier is improvised using cuckoo search optimization strategy. In [19], authors compare and contrast various neural network based models like Extreme Learning Machine, Wavelet Neural Networks, Fuzzy Neural Networks, Least Square Support Vector Machines to predict $PM_{2.5}$ concentration over short term durations. They demonstrate the efficacy of Wavelet Neural Networks over the compared architectures in terms of higher precision and self learning ability. Even meteorological parameters such as humidity and temperature are considered for evaluation. In [16], authors provide an hourly prediction over $PM_{2.5}$ concentration depicting multiple change patterns. They train separate models over clusters comprising of subsets of the data utilizing an ensemble of classification and regression trees which captures the multiple change pattern in the data.
In contrast to the aligning works [11], [16], [18], [19], in this paper we capture the long-term temporal dependencies between the data sources collected using both direct as well as indirect sources and introduce the importance of attention based adaptive learning in the memory networks. Apart from that, we provide a standalone model to predict various pollutants ($PM_{2.5}, NO_2$ and Pm_{10}) over a long period of time in future (next 24 hours).

B. Deep Learning based techniques

In recent times, researchers are predominantly utilizing deep learning based architectures for forecasting expressway $PM_{2.5}$ concentration. In [10], the authors devise a novel algorithm to handle the missing values in the urban data for China and Hong Kong followed by deep learning paradigm for air pollution estimation. This enables in predicting the air quality estimates throughout the city thus, curtailing the cost which might be incurred due to setup of sophisticated air quality monitors. In [14], authors provide a deep neural network (DNN) model to predict industrial air pollution. A non-linear activation function in the hidden units in the DNN reduces the vanishing gradient effect. In [8], authors propose

Long Short Term Memory (LSTM) networks to predict future values of air quality in smart cities utilizing the IoT smart city data. In [9], authors utilize geographical distance and spatiotemporally correlated $PM_{2.5}$ concentration in a deep belief network to capture necessary evaluators associated with $PM_{2.5}$ from latent factors. In [15], authors adopt a Deep Air Learning framework which emphasize on utilizing the information embedded in the unlabeled spatio-temporal data for improving the interpolation and prediction performance.

In this paper, we however predict other pollutants $(PM_{2.5}, NO_2$ and $Pm_{10})$ which are equally important to scale the air quality index to an associated harmful level utilizing the deep learning based paradigm.

C. Wireless networks based techniques

In [2], authors utilize mobile wireless sensor networks (MWSN) for generating precise pollution maps based on the sensor measurement values and physical models which emulate the phenomenon of pollution dispersion. This works in a synergistic manner by reducing the errors due to the simulation runs based on physical models while relying on sensor measurements. However, it would still rely on the deployment of the sensor nodes. In [7], authors utilize node-to-node calibration, in which at a time one sensor in each chain is directly calibrated against the reference measurements and the rest are calibrated sequentially while they are deployed and occur in pairs. This deployment reduces the load on sensor relocations and ensures simultaneous calibration and data collection. In [4], authors propose a framework for air quality estimation utilizing multi-source heterogeneous data collected from wireless sensor networks.

D. Trends and seasonality assessment based techniques

In [12], authors present a detailed survey on the trends of the seasonal variations across Delhi regions and present analysis on the typical rise of the pollution index during certain times over the year. In [13], authors investigate Bayesian semi-parametric hierarchical models for studying the time-varying effects of pollution. In [3], authors study on the seasonality trends on association between particulate matter and mortality rates in China. The time-series model is utilized after smooth adjustment for time-varying confounders using natural splines. In [5], authors analyze status and trends of air quality in Europe based on the ambient air measurements along with anthropogenic emission data.

Relative to these works, we study the seasonal variations at various locations in Delhi and find the analogies for the typical rise in the pollution trends at particular locations utilizing the meteorological data and historic data.

III Proposed Methodology

Fig. 2 demonstrates the framework of the proposed solution.

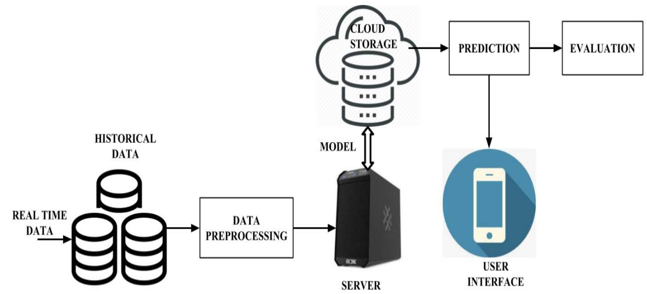

Fig. 2: Proposed framework of air pollution forecasting task

A. Data Collection and Preprocessing

We utilize real time air quality monitoring dataset from **Central Pollution Control Board**[1] database which has been widely used for model evaluation. CPCB has 703 operating stations in 307 cities and 6 Union Territories of India, out of which 78 stations are in Delhi. The dataset contains various direct and indirect parameters including:

1) The air pollutant variables in the air quality data are SO_2, NO_2, $PM_{2.5}$, PM_{10}, CO and O_3.
2) The meteorological parameters include temperature, humidity, wind speed and barometric pressure.
3) Time includes hour of the day, seasons and month of the year.

All these variables of the past 24 hours are collected on hourly basis and extracted features from the collected dataset are used for evaluation of the models for predictions of the concentration of $PM_{2.5}$, PM_{10} and NO_x after every 4 hours for the next 24 hours. Concentrations of all the pollutants are reported in $\mu g/m^3$.

The data collected from the Central Pollution Control Board database had two main issues:

1) The data collected from Central Pollution Control Board database contains missing values at random.
2) The data may contain parameters that are not useful for the model for predictions and may act as noise.

The complications faced due to the above mentioned issues in the data are addressed using data preprocessing which involves the selection of the important features for air quality forecasting and handling the missing values. We perform feature ranking using backward feature elimination, forward feature construction and Random Forests/Ensemble Trees to find the top ranked features for the proposed model. We obtain the temperature and humidity from the meteorological parameters and SO_2, NO_2, $PM_{2.5}$, PM_{10}, CO and O_3 from the air pollution parameters after the feature ranking methods. We also obtain hour of the day and month of the year as yet another important parameters using feature ranking methods. Since, the data consists of missing values at random (MAR) in the dataset for certain variables, it becomes a big challenge

[1] http://cpcb.nic.in/

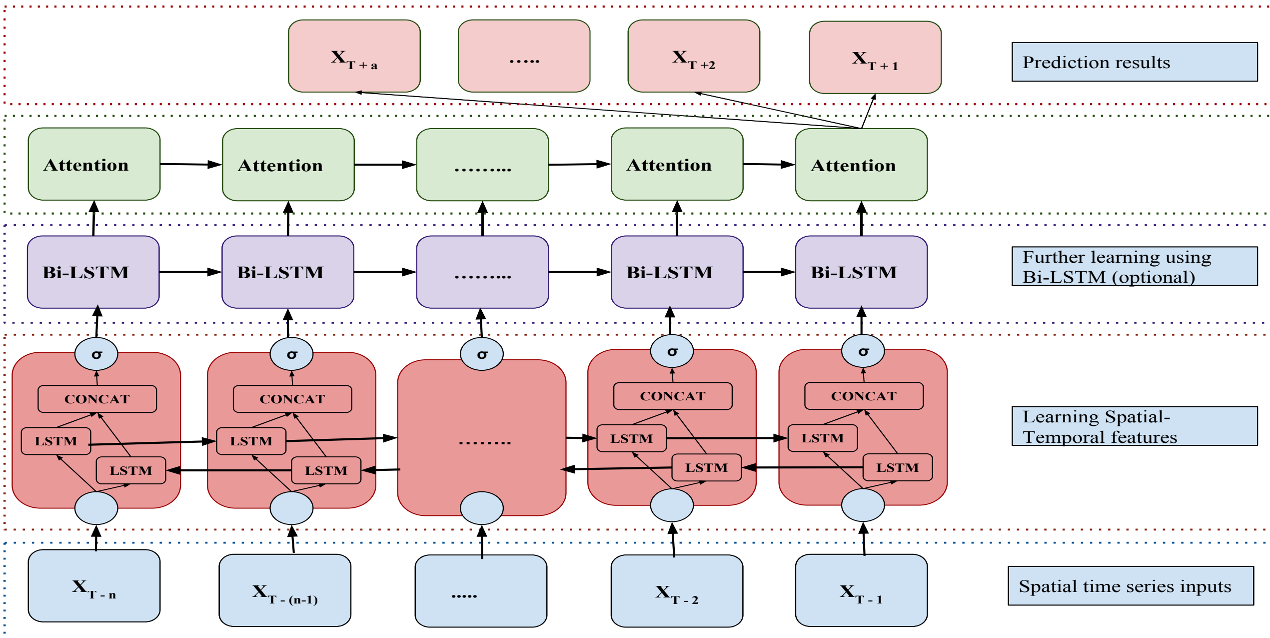

Fig. 3: Attention based BiLSTM model consists of a BiLSTM layer and the attention mechanism layer. $\{x_{T-n}, x_{T-(n-1)}, ...x_{T-2}, x_{T-1}\}$ represent the historical data sequence input to the BiLSTM layer and $\{x_{T+a}, ...x_{T+2}, x_{T+1}\}$ represent the pollution forecasted values. Multiple BiLSTM or LSTM layers are optional.

to handle them in order to obtain accurate forecasting of air pollution. Therefore, we impute the missing values by using the Multivariate Imputation by Chained Equations (MICE) [1]. We then apply normalization techniques to all the features and pollution targets so that the values lie in the range [0,1]. Normalization changes the values of numeric columns in the dataset without distorting differences in the ranges of values by using a common scale.

B. Model and Evaluation

We propose an attention based BiLSTM online network as shown in Fig. 3. We utilize the trained model to forecast the pollutants for the next 24 hours on the test set which are used for the evaluation of different models. Algorithm 1 outlines the proposed adaptive method. By running the algorithm every week on the hourly updated data for each location, the errors of the online adaptive algorithm are minimized. The attention based BiLSTM consists of 4 modules:

1) Input Feature Module: The input of the proposed model is the historical data sequence where each sequence is composed of pollution vectors represented as $P_f = \{p_1, p_2...p_n\}$, meteorological vectors represented as $M_f = \{m_1, m_2...m_n\}$ and time vectors represented as $T_f = \{t_1, t_2...t_n\}$. Finally, a sequence of combination of these vectors $F = \{f_1, f_2, ...f_{n-1}, f_n\}$ is sent to the BiLSTM module as inputs.
2) BiLSTM module: This module consists of a single or multiple BiLSTM layers. For the first BiLSTM layer, the input is the historical data sequence represented as $S_f = \{P_f.M_f.T_f\}$ and the output is $h^1 = \{h_1^1, h_2^1...h_s^1\}$, where $h_t^1 = [\overrightarrow{h_t^1}; \overleftarrow{h_t^1}]$.
3) Attention module: This module consists of the attention layer which takes the output denoted as $h^n = \{h_1^n, h_2^n...h_s^n\}$, from the BiLSTM layer. This module assesses the importance of the representations of the encoding information e_i and computes the weighted sum. The output of the attention function f_{att} is normalized through a softmax function:

$$z_{ji} = f_{att,j}(e_i) \tag{1a}$$

$$w_{ji} = \frac{exp(z_{ji})}{\sum_{k=1}^{E} exp(z_{jk})} \tag{1b}$$

Algorithm 1: Algorithm for proposed adaptive method

INPUTS: Data for each location $\{f_1, f_2, ..f_{n-1}, f_n\}$; learning rate α
Initialize $F(x)$ = BiLSTM model with attention mechanism for N pollutants
for $t \leftarrow 1$ **to** T **do**
  Recieve instance: x_t
  Predict for each pollutant for the next 24 hours
  Recieve the true pollutant value y_t
  Suffer loss: $l_t(w_t)$ which is a convex loss function on both ${w_t}^T x$ and y_t
  Update the prediction model w_t to w_{t+1}
end

TABLE I: Cloud configurations

Services	***Configurations***
Virtal Server	n1-standard-1 (1 vCPU, 3.75 GB memory)
Cloud Storage	Regional Storage Bucket
Machine learning Engine	STANDARD_1

4) Output module: This module is the key module to get the network output. It consists of the fully connected layer which takes the $h^n = \{h_1^n, h_2^n ... h_s^n\}$ as features. It generates the predicted values for the next 24 hours in the case of regression of various pollutants.

C. Cloud architecture

The cloud architecture is composed of three components in our system: Google Virtual Machine, Google Cloud Storage and Google Machine learning Engine. Table I presents the configuration of various components in our cloud architecture. The air quality real time data including the air pollutant's concentration and meteorological data is collected from external sources by crawling the web pages from Central Pollution Control board every hour. This collected data is stored and updated on the Google Cloud storage.

The cloud server based on Google Cloud hosts all the components of our proposed system in Figure 2. The updated data is fetched from Google Cloud storage and is used to adaptively train the machine learning model on Google Machine Learning Engine every week for all the locations. The online trained model w_{t+1} replaces the model w_t which is stored on Google Cloud Storage to minimize the mistakes made by the predictive model. This predictive model is then used to forecast the concentrations of different pollutants at a time interval of every 4 hours for the next 24 hours. These prediction results are then stored back to the Google Cloud Storage for access by the user interface.

IV Evaluation and Analysis

In this section, we evaluate the efficacy and efficiency of the proposed architecture against several baseline models and analyze the experimental results. We design prediction model to forecast the continuous values for different pollutants. Table II represents the summary of our initial dataset on which we evaluate our proposed models with various baseline models. We split 80% of the dataset for training and 20% for test. We select 20% of the training dataset as validation set and allow training to be early stopped according to the mean score error for the validations set. All values are then normalized in the range [0, 1]. The network is built using Keras with tensorflow as backend. For training, batch size is set as 64 and we utilize Adam [6] optimizer with learning rate of 0.001 for gradient descent optimization. Dropout [17] is used as a regularization technique in all the models.

A. Trends in Pollution

Delhi is located at 28.61°N 77.23°E, and lies in Northern India. The area represents high seasonal variation. Delhi is covered by the Great Indian desert (Thar desert) of Rajasthan on its west and central hot plains in its south part. The north and east boundaries are covered with cool hilly regions. Thus, Delhi is located in the subtropical belt with extremely scorching summers, moderate rainfall, and chilling winters.

TABLE II: Details of Dataset

Characterstics	***Values***
Number of Locations	5
Minimum number of samples per location	4, 000
Maximum number of samples per location	30, 000
Average number of samples per location	7, 000
Span of Data Collection	3 years
Features per sample	9
Seasons covered	Summer, Monsoon, Autumn, Winter and Spring
Number of hours per day	24

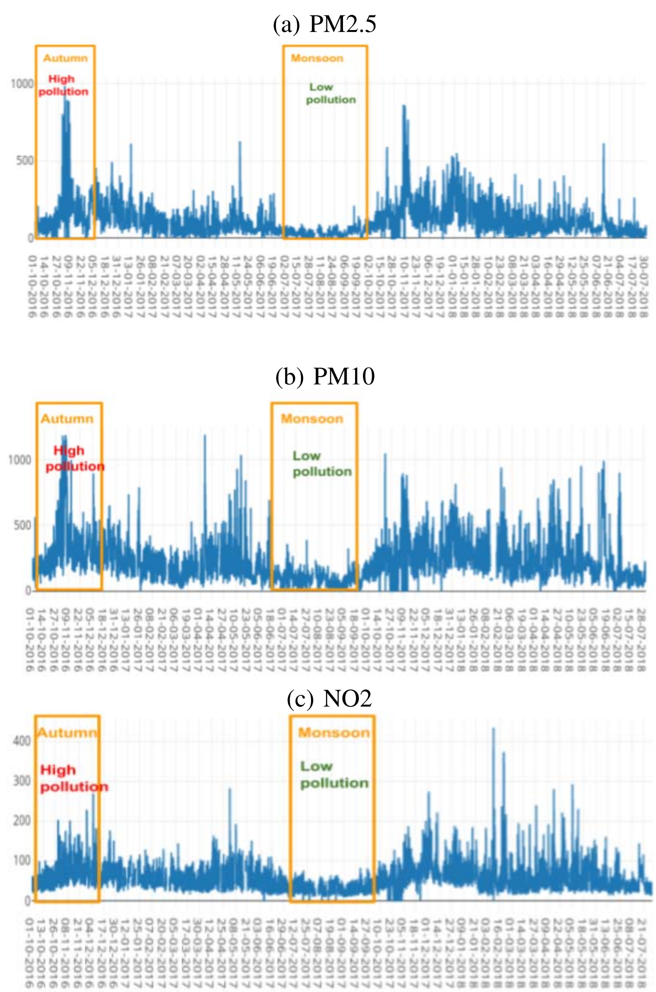

Fig. 4: Seasonal variation in concentration of Mandir Marg, Delhi for (a) $PM_{2.5}$ **(b)** PM_{10} **(c)** NO_2**.** X-axis represent date and Y-axis represent concentration of pollutant.

As the concentrations of each of the air pollutants is directly related to seasonal variation of the atmosphere, it becomes important to study the break-up of 12 months. Winter season includes December and January months, spring season includes February and March months, summer season includes April, May and June months, monsoon season includes July, August and September months while autumn includes October and November months.

Fig. 4 shows the concentration trends for Mandir Marg location in Delhi, in the five seasons (winter, spring, summer, monsoon and autumn), from October 2016 to July 2018. The concentration of Mandir Marg area was comparatively worst in winter and best in monsoon. The following sequence in the decreasing order was observed in the concentration of PM_{10}, $PM_{2.5}$ and NO_2: winter > autumn > spring > summer > monsoon. Decreasing order of concentration implies air quality going from worst to better, it means here that winter has worst air quality and monsoon has best.

High concentration is observed after 4pm due to vehicular emission. The concentration of NO_2 is highest during 8 am - 10:30 am and 4 pm - 7 pm due to the vehicular emission, from the burning of fuel, from emissions from cars, trucks, buses etc.

B. Evaluation on Pollutants Values prediction task

It is a very important task to forecast the continuous values of the pollutants to circumvent the dangers caused by them and to take the necessary actions. To make the predictions, we employ the following prediction models: LSTM (Long short term memory network), LSTM-A (Attention based Long Short term memory network), BiLSTM (BiDirectional Long Short Term memory network), BiLSTM-A (Attention based Bidirectional LSTM Network). In order to evaluate the performance of the various methods for regression, root mean square error (RMSE), and R-squared (R^2) are calculated.

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

$$R^2 = \frac{\sum_{i=1}^{N}(\hat{y_i} - y_{avg})^2}{\sum_{i=1}^{N}(y_i - y_{avg})^2} \tag{2b}$$

where N denotes the number of instances in the test set, $\hat{y}$ are the predicted values, y denotes the actual values and y_{avg} denotes the average of observations in the test set.

Table III shows the Root mean square and R-square evaluation for different methods. It represents the evaluation for the next 4 hours. For all the instances, we use the historical values of the last 24 hours. In general, we observe a performance boost with BiLSTM-A in the predictions as compared to the other baseline models by $\sim 2-6\%$. Random Forest performs well for the forecasting of $PM_{2.5}$ values. The proposed model of BiLSTM-A outperforms in the prediction of all the other pollutants. Fig. 5 shows the prediction results compared with the actual values for the BiLSTM-A model for the next 4 hours. It shows that the proposed model achieves significant improvement in accuracy, especially in the scenarios of sudden change, which clearly demonstrates that the proposed method indeed benefits the temporal predictions.

BiLSTM with attention (BiLSTM-A) is preferred over BiLSTM because when Attention mechanism is applied on top of BiLSTM, it captures periods and makes BiLSTM more robust to random missing values. .

TABLE III: Performance comparison of the proposed model with other baseline models for pollution values forecasting for future 4 hours on the basis of R-squared values and Root mean square error values.

Model	*Pollutants*	*R-square*	*RMSE*
Random Forest	$PM_{2.5}$	**0.35**	**40.69**
	NO_2	0.406	21.12
	PM_{10}	0.425	98.32
LSTM	$PM_{2.5}$	0.31	41.96
	NO_2	0.383	21.52
	PM_{10}	0.445	96.58
LSTM-A	$PM_{2.5}$	0.29	42.52
	NO_2	0.387	21.44
	PM_{10}	0.446	96.49
BILSTM	$PM_{2.5}$	0.30	42.07
	NO_2	0.385	21.47
	PM_{10}	0.442	96.77
BILSTM-A	$PM_{2.5}$	0.310	41.97
	NO_2	**0.417**	**21.08**
	PM_{10}	**0.454**	**96.22**

C. Ablation Studies on real time learning system

The pollution forecasting and source prediction methods discovered in this study can help governments and people take necessary sections. We train the initial model used in the evaluation on the initially collected data. Due to the progressive temporal changes in the concentration of the pollutants, it is necessary to continuously update the collected data and the model, thus resulting into an adaptive model. In this section, we compare the performance of the model on the initial collected data to the performance of the model which is updated every week on the real time hourly updated data after a month of the collection of the initial data. BiLSTM-A model is used for evaluation of both the models with the updated data in Table IV.

TABLE IV: Performance of Initial BiLSTM-A model with initial data compared to Adaptive BiLSTM-A model

Model	*Pollutants*	*R-square*	*RMSE*
Initial model	$PM_{2.5}$	0.310	41.97
	NO_2	0.417	21.08
	PM_{10}	0.454	96.22
Adaptive Model	$PM_{2.5}$	**0.39**	**37.69**
	NO_2	**0.48**	**19.79**
	PM_{10}	**0.51**	**90.38**

Table IV shows substantial improvement of the results after the real time update which is indicative of the relevance of continuous updation of the model with the real time incoming data.

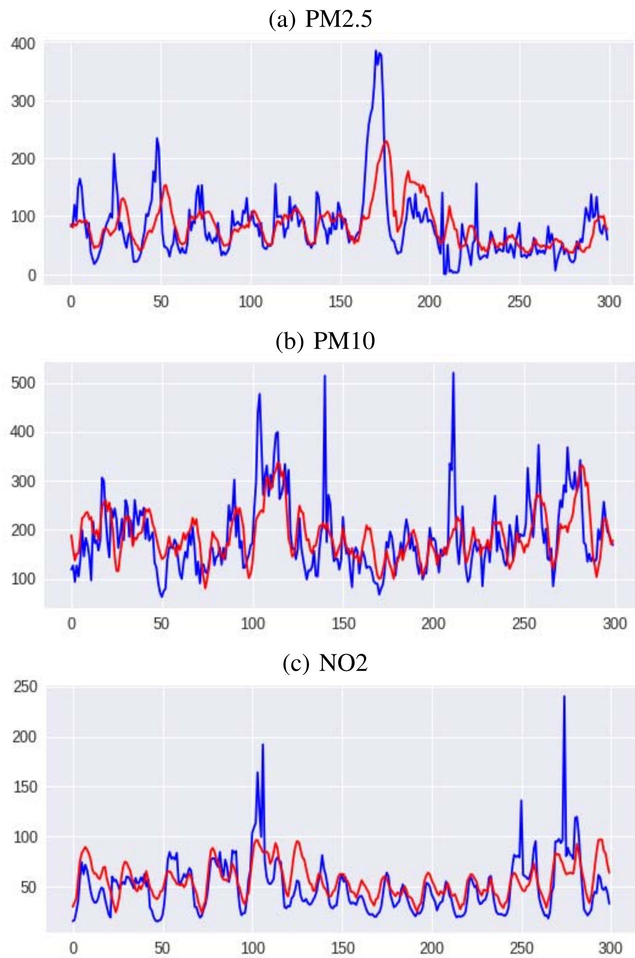

Fig. 5: Comparison between pollution estimates (red line) and actual measurements (blue line) for (a) $PM_{2.5}$ **(b)** PM_{10} **(c)** NO_2**.**Y-axis represents the concentration of pollutant and x-axis represent the sample number.

V Conclusion

Based on the historical and real-time ambient air quality and meteorological data of 5 monitoring stations in Delhi, we inferred the real-time and fine-grained ambient air quality information. We developed a novel end-to-end system to predict the air quality of next 24 hours by predicting the concentration and the level (low, moderate, high) of different air pollutants including nitrogen dioxide (NO_2), particulate matter ($PM_{2.5}$ and PM_{10}) for Delhi. Extensive experiments on air pollution data for 5 locations in Delhi, helped evaluating the proposed approach. The results demonstrated the performance boost with the proposed method over other well known methods for regression models. In future work, we intend to explore more powerful modeling techniques along with the traffic density data, as a way to model the traffic density of the monitored location to get better results.

Acknowledgments

We would like to thank Dr Aakanksha Chowdhery, Google Brain for the inspiring guidance and constant encouragement during the course of this research study. We acknowledge the Marconi Society for their support in the form of research funding from Celestini Program. We would also like to thank Central pollution control board for the dataset which has played a very important role towards the completion of this project.

References

[1] Stuart E. A. Frangakis C. Leaf Azur, M. J. Multiple imputation by chained equations: What is it and how does it work? international journal of methods in psychiatric research, 20(1), 4049. 2011.

[2] Ahmed Boubrima, Walid Bechkit, Hervé Rivano, and Lionel Soulhac. Poster: Toward a better monitoring of air pollution using mobile wireless sensor networks. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, pages 534–536. ACM, 2017.

[3] Renjie Chen, Roger D Peng, Xia Meng, Zhijun Zhou, Bingheng Chen, and Haidong Kan. Seasonal variation in the acute effect of particulate air pollution on mortality in the china air pollution and health effects study (capes). *Science of the total environment*, 450:259–265, 2013.

[4] Cheng Feng, Wendong Wang, Ye Tian, Xirong Que, and Xiangyang Gong. Air quality estimation based on multi-source heterogeneous data from wireless sensor networks. In *Wireless Communications and Networking Conference (WCNC), 2018 IEEE*, pages 1–6. IEEE, 2018.

[5] Cristina BB Guerreiro, Valentin Foltescu, and Frank De Leeuw. Air quality status and trends in europe. *Atmospheric environment*, 98:376–384, 2014.

[6] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *CoRR*, abs/1412.6980, 2014.

[7] Fadi Kizel, Yael Etzion, Rakefet Shafran-Nathan, Ilan Levy, Barak Fishbain, Alena Bartonova, and David M Broday. Node-to-node field calibration of wireless distributed air pollution sensor network. *Environmental Pollution*, 233:900–909, 2018.

[8] İbrahim Kök, Mehmet Ulvi Şimşek, and Suat Özdemir. A deep learning model for air quality prediction in smart cities. In *Big Data (Big Data), 2017 IEEE International Conference on*, pages 1983–1990. IEEE, 2017.

[9] Tongwen Li, Huanfeng Shen, Qiangqiang Yuan, Xuechen Zhang, and Liangpei Zhang. Estimating ground-level pm2. 5 by fusing satellite and station observations: A geo-intelligent deep learning approach. *Geophysical Research Letters*, 44(23), 2017.

[10] Victor OK Li, Jacqueline CK Lam, Yun Chen, and Jiatao Gu. Deep learning model to estimate air pollution using m-bp to fill in missing proxy urban data. In *GLOBECOM 2017-2017 IEEE Global Communications Conference*, pages 1–6. IEEE, 2017.

[11] Yan Kuang David He Serap Erdal Donna Kenskie Ming Dong, Dong Yang. Pm2.5 concentration prediction using hidden semi-markov model-based times series data mining. 36(5):9046–9055, 2009.

[12] Manju Mohan and Anurag Kandya. An analysis of the annual and seasonal trends of air quality index of delhi. *Environmental monitoring and assessment*, 131(1-3):267–277, 2007.

[13] Roger D Peng, Francesca Dominici, Roberto Pastor-Barriuso, Scott L Zeger, and Jonathan M Samet. Seasonal analyses of air pollution and mortality in 100 us cities. *American journal of epidemiology*, 161(6):585–594, 2005.

[14] Yu Pengfei, He Juanjuan, Liu Xiaoming, and Zhang Kai. Industrial air pollution prediction using deep neural network. In *International Conference on Bio-Inspired Computing: Theories and Applications*, pages 173–185. Springer, 2018.

[15] Zhongang Qi, Tianchun Wang, Guojie Song, Weisong Hu, Xi Li, and Zhongfei Mark Zhang. Deep air learning: Interpolation, prediction, and feature analysis of fine-grained air quality. *IEEE Transactions on Knowledge and Data Engineering*, 2018.

[16] Zhigen Shang, Tong Deng, Jianqiang He, and Xiaohui Duan. A novel model for hourly pm2. 5 concentration prediction based on cart and eelm. *Science of The Total Environment*, 651:3043–3052, 2019.

[17] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

[18] Jingyi Sun, Wei; Sun. Daily pm2.5 concentration prediction based on principal component analysis and lssvm optimized by cuckoo search algorithm. 03 2017.

[19] Shan Zhang, Xiaoli Li, Yang Li, and Jianxiang Mei. Prediction of urban pm 2.5 concentration based on wavelet neural network. In *2018 Chinese Control And Decision Conference (CCDC)*, pages 5514–5519. IEEE, 2018.

[20] Yu Zheng, Xiuwen Yi, Ming Li, Ruiyuan Li, Zhangqing Shan, Eric Chang, and Tianrui Li. Forecasting fine-grained air quality based on big data. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '15, pages 2267–2276, New York, NY, USA, 2015. ACM.