Air Quality Prediction in Visakhapatnam using Deep Learning techniques: RNN, LSTM and GRU

K. Srinivasa Rao[1], Dr. G. Lavanya Devi [2], N. Ramesh[3]
Department of Computer Science and System Engineering, Andhra University College of Engineering, Visakhapatnam, India
[1](E-mail: sri.kurapati@gmail.com)
[2](E-mail: lavanyadevig@yahoo.co.in)
[3](E-mail: rameshauvsp@gmail.com)

Abstract— An early warning system for air quality control requires an accurate and reliable forecasting of pollutants in the air. Monitoring and stabilizing air quality has become one of the most essential challenging tasks in the cities today. The quality of the human life in cities is depending upon the quality of ambient air. Hence, air quality assessment and prediction has become an important study area. Traditional approaches depend on numerical methods to estimate the air pollutant concentration and require lots of computing power. To tackle this issue, a tendency to apply the extensively used deep learning techniques on the Visakhapatnam air quality data. The aim of this research paper is to investigate deep learning techniques for air quality forecasting. Hourly meteorological parameters as well as the air pollutant concentration values were used as input to recurrent neural networks and to perform air quality prediction. Deep learning frameworks such as Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models are proposed and compared with state-of-the-art. The experimental verification of the models was conducted on Visakhapatnam air quality data for the period July 1, 2016 to May 17, 2018. The results are encouraging and it was demonstrated that implementation of these techniques could be very effective in predicting air quality. Further, these models may be enhanced by implementing bidirectional mechanism in recurrent layer.

Keywords— *Air quality, air pollutant, prediction, numerical methods, deep learning techniques, recurrent neural networks (RNN), long short-term memory (LSTM) and gated recurrent unit (GRU)*

I. INTRODUCTION

Human beings would not have a society if they destroy the environment. Environmental pollution is an incurable disease. It can only be prevented! Air pollution is one of the major types of environmental pollution. Air pollution is an important environmental issue that has a direct effect on human health (asthma, lung cancer, brain damage, liver damage, allergic reactions wheezing& bronchitis) and ecological balance [1-2]. In general, the Air Quality Index (AQI) is used to distinguish the air quality level. The AQI levels in developed cities like Delhi, Agra (Taj Mahal) are increasing at an alarming pace [3-4]. Hence, analysis and evaluation of air quality is one of the most primary concerns for us today.

Geographically, each and every region has its own notable features. As a reason, better results could be achieved if air quality research is carried out region wise. This gives a broader elevation to features like population, industries located; automobiles in the region etc. and can be studied with more inference. As a reason, this research concentrates to carry out the proposed methodology for the city Visakhapatnam.

The air pollutants, e.g., $PM_{2.5}$, PM_{10}, NO,NO_X, NO_2, NH_3, SO_2, CO, Ozone, Benzene, Toluene, and Xylene as well as meteorological parameters (temperature, relative humidity, wind direction, solar radiation, pressure) have been used in the present study to analyze the air quality in city Visakhapatnam.

Many researchers have been proposed various models for predicting air pollution. The traditional approaches for air quality prediction models use mathematical and statistical techniques. However, these methods were lengthy and inefficient approach for better output prediction [5]. Artificial intelligence (AI) based techniques have been proposed as alternatives to traditional statistical ones on forecasting urban air pollution. There are several AI techniques which have been proposed as feasible and dependable ways for air pollution forecasting, such as artificial neural networks (ANNs), support vector machines (SVMs), and fuzzy logic [6]. Recently, deep learning techniques can extract representative for air quality prediction when compared to shallow methods [7].

This paper presents deep learning based approaches for predicting air ambience in Visakhapatnam. The main contributions of this research paper are:

- RNN, LSTM and GRU based frameworks for forecasting concentrations of air pollutants in Visakhapatnam are proposed.
- Experimental analysis has been carried out on dataset of Visakhapatnam and the proposed models achieved superior results over baseline approaches.

II. RELATED WORK

Deep learning methodology has accomplished outstanding results in sequence data processing, such as image captioning [8], speech recognition [9], house price prediction [10], traffic speed prediction and traffic accident risk prediction [11]. In those tasks, deep learning methods have outperformed conventional machine learning methods. Inspired by this, people are trying to use deep learning representations for example Recurrent Neural Networks (RNNs) and Long Short Term Memory (LSTM), to perform air pollution forecasting. Several studies [12, 13, 14, 15, and 16] have been done to examine the applicability of deep learning techniques in air quality forecasting, and the results demonstrate the advantages of deep learning. Currently, RNN based LSTMs, GRU have been gaining popularity in air pollution forecasting because of their capability to model long term dependencies.

Recently, Zhongang Qi et al. [17] combined deep learning with a spatiotemporal method for interpolation and air quality prediction in Beijing, China. Vikram Reddy et al. [18] accompanied a series of experiments for applying LSTMs to air pollution prediction on the Beijing dataset. However, this model only examined the correlation between meteorological conditions and $PM_{2.5}$ while present work evaluates additional air pollutants information (PM_{10}, NO, NO_X NO_2, NH_3, SO_2, CO, Ozone, Benzene, Toluene, and Xylene) using RNN, LSTM and GRU models.

III. THEORETICAL BACKGROUND AND METHODS

This paper puts forward a RNN, LSTM and GRU based models for forecasting pollutant concentrations by considering temporal sequential data of a particular pollutant.

Problem Formulation

The hypothesis of these models are, given a temporal sequence data of pollutant concentration values and metrological parameters of a particular location, the models will capture the dependencies in the data and predicts the next hour pollutant concentrations. Given, metrological parameters $m = \{m_1, m_2, m_3, m_4, m_5, m_6\}$ and pollutant concentrations x_t, $t = 1...T$ as a paired input $X= \{(m, x_t)\}$, the objective of the proposed models are to recognize patterns and predict x_{t+1}. At each time step, the models comprises of processing input, recurrent and output layers.

Input Layer: This layer generates dense embedding of pollutant concentrations and metrological parameters by concatenation.

Recurrent Layer: This layer generates hidden representations of uncovered sequential features from embedding vector x_t^e. The hidden states h_t are computed by using input from previous time step h_{t-1} and current input x_t^e.

Output Layer: This final layer generates concentration values of next hour time step of the pollutant considered with y_t.

LSTM [12] is a special kind of RNN that improves gradient vanishing problem and has the capability of learning long term dependencies. The proposed architecture, RNN-LSTM is vital to enhance the prediction performance of concentration values.

In contrast to the RNN layer cell module, RNN-LSTM has three gates to preserve long term dependencies.

The computations over states and gates at time step t are defined as follows:

$$i_t = \sigma (W_{xi} x_t + W_{hi} h_{t-1} + b_i)$$

$$f_t = \sigma (W_{xf} x_t + W_{hf} h_{t-1} + b_f)$$

$$o_t = \sigma (W_{xo} x_t + W_{ho} h_{t-1} + b_o)$$

$$c_t = f_t * c_{t-1} + i_t * tanh (W_{xg} x_{t +} b_g)$$

$$h_t = o_t * tanh(c_t)$$

Here, σ is the logistic sigmoid function.

I, f, o and *c* are input gate, forget gate, output gate, and cell activation vectors respectively. All these vectors are of the same size as the hidden vector *h*.

W_{xi}, W_{xf}, W_{xo}, W_{xg} are projections.

W_{hi}, W_{hf}, W_{ho} are recurrent weights.

GRU is a significant kind of recurrent neural networks [19]. It was capable of learning long term dependencies. A well-known variant network based on LSTM is the Gated Recurrent Unit (GRU). GRU is proposed by Cho et al. in 2014 [20] and it is an extension to LSTM network. Compare with LSTM, GRU does not maintain a cell state *c* and use 2 gates instead of 3. It consists of update *(z)* and reset *(r)* gates. They altogether include the balancing flow of data inside the unit. The GRU model is modest and has fewer parameters than LSTM, and has been shown to outperform or keep the same performance as LSTM on some tasks. It is very widespread and capable of processing sequence learning tasks [15]. Mathematically, the relationship between the update and reset gate of RNN-GRU is defined by a set of following subsequent equations:

$$z_t = \sigma (W_z h_{t-1} + W_z x_t)$$

$$r_t = \sigma (W_r h_{t-1} + W_r x_t)$$

$$c = tanh (W_c (h_{t-1} \otimes r) + W_c x_t)$$

$$h_t = (z \otimes c) \oplus (1-z) \otimes h_{t-1}$$

Here, σ is the logistic sigmoid function and *z, r* and *c* are respectively the update gate, reset gate, and cell activation vectors, all of which are the same size as the hidden vector *h*.

The RNN-GRU model was proposed as an alternative to the computationally expensive RNN-LSTM model.

These models are trained for learning parameters by gradient descent optimization technique. The loss function used in this model is mean square error (MSE) function. Finally, training is carried out for fixed number of time steps using back propagation through time (BPTT).

IV. EXPERIMENTAL SETUP

The models are trained with the real time data collected from the Central Pollution Control Board (CPCB), of the city Visakhapatnam for the period July 1, 2016 to May 17, 2018 [21]. This is an industrial city situated in coastal area with bowl type geographical structure located in state of Andhra Pradesh, India [22]. The set of air pollutants and meteorological parameters used in this research study are shown in Tables 1.

Table 1: Set of Air pollutants and metrological parameters

Sno	Parameters	unit
1	$PM_{2.5}$	ug/m3
2	PM_{10}	ug/m3
3	NO	ug/m3
4	NO_2	ug/m3
5	NO_X	ug/m3
6	NH_3	ug/m3
7	SO_2	ug/m3
8	CO	ug/m3
9	Ozone	ug/m3
10	Benzene	ug/m3
11	Toluene	ug/m3
12	Xylene	ug/m3
13	Ambient Temperature	0C
14	Relative Humidity	%
15	Wind Direction	degree
16	Solar Radiation	W/mt2
17	Pressure	mmHg
18	Rain Fall	mm

Data obtained from the aforementioned source has to be preprocessed as noisy data affects the performance of the forecasting model. Missing values add more noise to data [23]. For the proposed model, missing values were imputed by mean values of the respective parameter [24]. Thus, the final dataset considered has 14,379 samples for each pollutant.

The objective of this paper is to develop a framework to predict the air pollutant concentration of the next hour on the basis of the past air quality data. Performance obtained by RNN, RNN-LSTM and RNN-GRU have to be evaluated empirically [25]. The metrics considered for evaluation are Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Determination (R^2) [17].

The dataset randomly divided into two groups: a training set that contains 70% of the original dataset, and the remaining 30% used as a test set for the models. The division dataset into training and test sets might be sensitive to the randomly selected pollutants/ meteorological parameters. Therefore, to ensure that evaluation is not vulnerable to the randomness of the division step, we ran the models ten times, each time with a different partitioning. Cross validation is adopted to minimize the bias at the training phase. The results obtained on evaluation of RNN, RNN-LSTM and RNN-GRU models are compared against variants of base line regression technique Support Vector Regressor (SVR) [26-27]. The values obtained are depicted in Table 2. Comparison was performed with statistical technique, SVR with linear and non liner kernels. It is well known for its performance in fore casting time series data.

The whole outcomes point out that RNN, RNN-LSTM and RNN-GRU models attains higher prediction accuracy when compared to SVR model. Despite the fact that, observed from the Table 2, RNN-GRU model is computationally efficient than an RNN-LSTM network, because of the reduction of gates, it still comes second to RNN-LSTM network in terms of performance. Consequently, RNN-GRU can be utilized when need to train faster and don't have much computation power at hand. Mat plot of compared models are depicted in Figure 1.

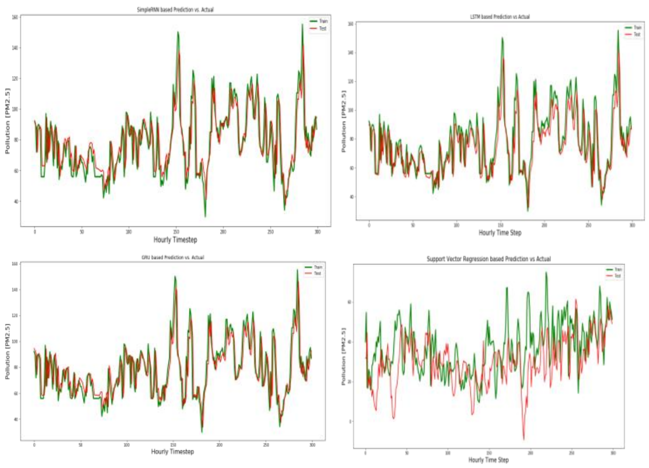

Figure 1: RNN-LSTM-GRU Models vs. SVR model Prediction performance graph

A summary view of the model performance is given in Figure 2.

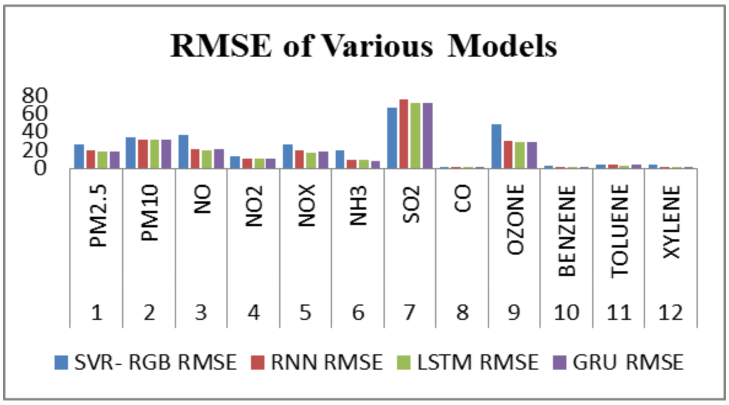

Figure 2: Performance Graph of RNN- LSTM –GRU and SVR models

Table 2: Performance comparison of RNN, RNN-LSTM and RNN-GRU with baseline models

Sno	Pollutant	SVR- KERNEL-RGB			RNN MODEL			RNN-LSTM MODEL			RNN-GRU MODEL		
		RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2	RMSE	MAE	R^2
1	PM2.5	25.7217	661.6040	0.325782	19.467	378.964	0.702520	17.9674	322.8130	0.7046585	18.074	326.677	0.743564
2	PM10	34.0357	1158.4258	0.663535	31.718	1006.026	0.768296	31.2256	975.0380	0.775433	31.010	961.596	0.778529
3	NO	36.0192	1297.3804	-1.459970	20.762	431.053	0.208667	20.0946	403.7929	0.258709	20.236	409.478	0.248276
4	NO2	12.5992	158.7390	0.780514	10.954	119.985	0.802125	10.3570	107.2674	0.823098	10.499	110.235	0.818205
5	NOX	26.3446	694.0378	0.096680	19.086	364.258	0.521605	17.5528	308.1007	0.595359	18.019	324.697	0.573562
6	NH3	19.8518	394.0951	-5.007071	8.532	72.792	0.013927	8.6551	74.9107	-0.014778	8.274	68.455	0.072672
7	SO2	66.4244	4412.1944	-0.648119	74.858	5603.706	-0.558299	71.1063	5056.1058	-0.406021	71.605	5127.307	-0.425820
8	CO	1.4214	2.0203	-3.265139	0.453	0.205	0.457562	0.4660	0.2171	0.425656	0.414	0.171	0.547017
9	OZONE	47.4760	2253.9753	-0.467832	29.695	881.808	0.415469	28.0426	786.3874	0.478722	28.696	823.482	0.454132
10	BENZENE	2.6724	7.1419	0.355471	1.717	2.949	0.665886	1.5907	2.5303	0.713354	1.571	2.466	0.720598
11	TOLUENE	4.4679	19.9622	0.535282	3.597	12.941	0.695934	3.2363	10.4736	0.753900	3.299	10.882	0.744315
12	XYLENE	3.8224	14.6106	0.268886	1.379	1.901	0.660305	1.4448	2.0874	0.627005	1.342	1.801	0.678267

V. CONCLUSION

The construction of a simple and an effective tool for air quality forecasting is highly desirable. This paper proposed RNN, RNN-LSTM and RNN-GRU models given temporal sequence data as input that outputs pollutant concentrations. Real time data of the city Visakhapatnam air pollutant values is considered for experimental study. The performance comparison based on RMSE (Table 2) and the predicted pollutants graphs (Figure 1&2) shows that accuracy and effectiveness of given models. This study indicates that benefits gained from air pollution concentrations, meteorological features and deep recurrent neural networks can bring accurate air pollutants prediction. The proposed models shows significant key results in prediction PM2.5 and remaining pollutants based on historical meteorological data when compared to baseline model.

REFERENCES

[1] Siuli Roy, Anurag D and Somprakash Bandyopadhyay, "Testbed Implementation of a Pollution Monitoring System Using Wireless Sensor Network for the Protection of Public Spaces", Management Information Systems, Indian Institute of Management Calcutta.

[2] Ganganjot Kaur Kang, Jerry Zeyu, Sen Chiao, Shengqiang Lu, and Gang Xie, "Air Quality Prediction: Big Data and Machine Learning Approaches", International Journal of Environmental Science and Development, vol. 9, January, 2018.

[3] https://www.vox.com/energy-and-environment/2017/11/22/16666808/india-air-pollution-new-delhi.

[4] https://www.reuters.com/article/us-health-pollution-india-tajmahal/pollution-turns-indias-white-marble-taj-mahal-yellow-and-green-idUSKCN1IN01S.

[5] Niharika, Venkatadri M, and Padma S Rao, "A Survey on Air Quality forecasting Techniques", International Journal of Computer Science and Information Technologies, vol.5, pp.103-107, 2014.

[6] Min Li and Md. Rafiul Hassan, "Urban Air Pollution Forecasting using Artificial Intelligence-based tools".

[7] Xiang Li et al., "Deep learning architecture for air quality predictions", Springer, 2016.

[8] Oriol Vinyals, Alexander Toshev, Samy Bengio and Dumitru Erhan, "Show and tell: A Neural Image Caption Generator", in Proceedings of the IEEE Conference on Computer Vision and Pattern recognition, pp. 3156–3164, 2015.

[9] Alex Graves, Abdel-rahman Mohamed and Geoffrey Hinton, "Speech Recognition with Deep Recurrent Neural Networks", in Acoustics, Speech and Signal processing, IEEE, pp. 6645–6649, 2013.

[10] Xiaochen Chen, Lai Wei, and Jiaxin Xu, "House Price Prediction using LSTM", https://arxiv.org/abs/1709.08432.

[11] Honglei Ren, You Song, Jingwen Wang, Yucheng Hu and Jinzhi Lei, "A Deep Learning Approach to the City wide Traffic Accident Risk Prediction", https://arxiv.org/pdf/1710.09543.15, April, 2018.

[12] B. Hochreiter and J. Schmidhuber, "Long Short Term Memory", Neural Computation, vol. 9, pp. 1735-1780, 1997.

[13] J. Schmidhuber, “Deep Learning in Neural Networks An Overview”, 2014.

[14] M. Langkvist, L. Karlsson, and A. Loutfi, “A Review of Unsupervised Feature Learning and Deep Learning for Time-Series Modeling”, Pattern Recognition Letters, vol. 42, pp. 11-24, June, 2014.

[15] Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning”, Nature, vol. 521, pp. 436-444, 2015.

[16] I. M. Coelho, V. N. Coelho, E. J. d. S. Luz, L. S. Ochi, F. G. Guimarães, and E. Rios, “A GPU Deep Learning Meta Heuristic based Model for Time Series Forecasting”, Applied Energy, 2017.

[17] Xiang Li et al., “Long short-term memory neural network for air pollutant concentration predictions: Method development and evaluation”, Elsevier Ltd., Environmental Pollution 231, pp. 997-1004, September 2017.

[18] Vikram Reddy et al., “Deep Air: Forecasting Air Pollution in Beijing, China”, 2017.

[19] Gary White, Andrei Palade, and Siobhan Clarke, “Forecasting QoS Attributes using LSTM Networks ”, The proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN), 2018.

[20] Cho et al., “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation”, arXiv preprint arXiv: 1406.1078, 2014.

[21] https://app.cpcbccr.com/ccr/#/caaqm-dashboard-all/caaqm-landing

[22] TVBPS Rama Krishna, MK Reddy, RC Reddy, RN Singh and S Devotta, “Modelling of Ambient Air Quality over Visakhapatnam Bowl Area”,Proc Indian NatnSciAcad72 No.1, pp. 55-61, 2006.

[23] Jiawei Han, Micheline Kamber, and Jian Pei, “Data Mining: Concepts and Techniques”, 3rd Edition, The Morgan Kaufmann Series in Data Management Systems, 2011.

[24] M. Sujatha, G. Lavanya Devi, K. Srinivasa Rao and N. Ramesh, “Rough Set Theory Based Missing Value Imputation”, Cognitive Science and Health Bioinformatics. Springer Briefs in Applied Sciences and Technology. Springer, Singapore, Online ISBN 978-981-10-6653-5, 2018.

[25] Rami Al-Rfou et al., “Theano: A Python framework for fast computation of mathematical expressions”, arXiv preprint arXiv: 1605.02688, 2016.

[26] KostandinaVeljanovska and Angel Dimoski, “Air Quality Index Prediction Using Simple Machine Learning Algorithms”, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS),ISSN 2278-6856,vol 7, Issue 1, January – February 2018.

[27] Bingyue Pan, “Application of XGBoost algorithm in hourly PM2.5 concentration prediction”, IOP Conf. Series: Earth and Environmental Science 113, 2018.

K. Srinivasa Rao completed M.Tech (CST) from Andhra University, Visakhapatnam, India and is currently pursuing his PhD from Andhra University, Visakhapatnam. His area of research includes Artificial Intelligence, Data Analytics, Machine Learning and Deep Learning.

Dr.G.Lavanya Devi is Assistant Professor of the Department of Computer Science & Systems Engineering, A.U. College of Engineering (A), Andhra University, Visakhapatnam, and Andhra Pradesh. She received the Young Engineer Award in 2011 from Institute of Engineers, India, Young Women Scientist in 2016 from the Institute of Bioinformatics & Computational Biology (IBCB), Visakhapatnam, India and Young Faculty Research Fellowship Award in 2018 from MietY, Government of India. 5 Ph.D.’s were awarded under her guidance. She coauthored more than 25 technical research papers in International Journals. She is a life member of IEEE, Computer Society of India. Her current research interests include Bioinformatics, Web mining, Data Analytics, Soft Computing and Deep Learning.

N.Ramesh completed M.Tech (CST) from Andhra University, Visakhapatnam, India and is currently pursuing his PhD from Andhra University, Visakhapatnam. His area of research includes Data Analytics, Soft Computing and Deep Learning.