Air Quality Prediction Using Recurrent Neural Network

Shweta K. Borse
Department of Computer Engineering
R.H.Sapat college of Engineering,
Management studied and research, Nashik
Savitribai Phule Pune University, India
shwetaborse58@gmail.com

Dipak V. Patil
Department of Computer Engineering
R.H.Sapat college of Engineering,
Management studied and research, Nashik
savitribai phule pune university, India
dipakvpatil17@gmail.com

***Abstract*—Pollution effect on humanoid fitness and the environment. Forecast air quality by using machine learning. We developed a method to forecast the Air Quality Index on prior time information about pollution. We develop a model using machine learning methods such as Simple-RNN, Simple-LSTM, and Stack-LSTM. In this paper, we concentrate on Simple-RNN, Simple-LSTM and Stack-LSTM. In this experiment, we observed that Stack-LSTM give better result as compared to Simple-LSTM and simple-RNN.**

***Index Terms*—Air quality index, Simple Long short-term memory, StackLSTM, Simple Recurrent neural network.**

I. Introduction

An air quality index (AQI) is operated by government, public impartial in a manner dangerous airborne presently or in, way mixed it is an estimation to establish. AQI raises, a more part of the public is probable to data of impurity airborne belong on fitness. Air quality index (AQI) is a mathematical ruler used for pleasing the greatest of impurity airborne rises hourly, airborne quantity on public fitness, and the environment. The combination of every pollution in the airborne is evaluated and improved. Consider quantity designed for every pollution named as per a sub-index. The maximum sub-index design historical data distinguished from the AQI used in that period. AQI is similar to a ration that sequence starts from 0-500. The combination of every infection, AQI values is categorized into series allocated to a public fitness warning. Tsai Y. Et al. [11] represented forecast PM2.5 with RNN using LSTM. Authors have used Keras packages for layer (like hidden-layer, Input-layer, output-layer). Keras lid advanced neural structures coded in python, TensorFlow supports these packages. For obtaining neural system, and track RNN with LSTM via TensorFlow. Freeman B. et al. [12] presented RNN is utilizing for airborne concentration estimates of input statistics. Airborne measure the time-series datasets, and contain the information. V.A. et al [13] proposed 3 model RNN, LSTM, GRU. These models conduct three layers (I/P layer with hidden-layer and O/P layers). To take accuracy data split into testing data and training data.

II. Review of Literature

Choubin B. et al. [1] Presented airborne particles are harmful to human health. Particulate matter, different gases are an impact on people and bug, carbon-based, and biological organizations. Airborne value implements used the machine learning method. The authors are applying the random forest and regression tree methods. By applying this method, obtain similar accuracy by using models. Fan J. et al. [2] Proposed three methods as SVM with BAT algorithm, particle swarm optimization (PSO), an estimate of diffuse (Rd) to global (RS) astral pollution. This method is associated with multivariate adaptive regression and (XGBoost). In research, the author obtains SVM to give good accuracy as comparable to XGBoot & multi-variate adaptive reversion spline (MARS). The SVM - BAT method is rising the act of machine learning methods. Y. C. Lin et al. [3] Presented an estimation of airborne quality structures constructed on the Neuro-fuzzy system. The historical dataset constructs fuzzy instruction. Split a dataset into training. Training data distributed a fuzzy cluster according to their characteristics. Fuzzy instructions are extreme quality and constrictions be in good exceptionally. Ding Y. Et al. [4] Proposed airborne smog controls its requirement to forecast airborne quality. Estimation techniques are implemented in earlier times to obtain the accuracy of predicting. In this paper, the authors presented spatial-temporal LSTM to estimate the airborne quality. The method constructed on LSTM. Zhang, D., and Woo, S. S. [5] Presented air quality consumes to rising environmental and fitness worries in South Korea. In explicitly, micro-dust is identifying to cause risky health difficulties to the community. Calculate the micro-dust used for sensors from a fixed spot. Estimate the airborne quality by using sensors, three several cars, and developed the app by using machine learning methods that report airborne quality to the operator. Tian Y. Et al. [6] Proposed to decrease airborne pollution impact on airplane processes everywhere the landing area of the airport. Random forest and Classification method follow supervised learning. RF gives the best accuracy output for the everyday dataset. In the airport, airborne quality, difficulty cracked by machine learning. R. J. Kuo et al. [7] Presented the concept of deep learning. It is machine learning;

it contains a neural network. Deep learning is proposed by RNN. RNN is extorted the layers in classification. Airborne pollution rises are worried not only inside the house, but also in other areas like the stock market, speech recognition, supply chain management, traffic problem, and so on. S. Shanthi and M. Pyingkodi [8] The author presents airborne pollution elements that are harmful to the public like PM, SO2, NO2., machine-learning methods apply to SVM, RF. The author takes the recent year dataset from the Tamil Nadu Control Board. Airborne smog monitoring placed areas. Dixian Zhu et al. [9] Proposed a machine learning method utilizes for airborne quality estimation. Estimate the time-period strength of airborne pollution. In machine learning methods, train data with a huge-scale optimization method. This technique used to increase the execution process. Kok I. et al. [10] Presented the model of IOT developed training in sectors. Such as schooling, manufacturing, and business. Prearranged hyper restraint, gives the output by the experiment. SVR and LSTM method of machine learning. The LSTM method gives better performance than the SVR method. In this paper, we explain a method using machine learning models such as Simple-RNN, Simple-LSTM, and Stack-LSTM. We focus on Simple-RNN, Simple-LSTM, and Stack-LSTM. In this study, we noted that Stack-LSTM gives a better result as compared to Simple-LSTM and simple-RNN. In this experiment, we detected that Stack-LSTM gives a better result as compared to Simple-LSTM and simple-RNN.

III. System Architecture / System Overview

A. Problem Statement

Midair impurity elements arise from the worst environmental germs, the density layer in midair build on the values, the volume of the environment, and integrate or spread these layers.

B. System Architecture

The figure shows phases for an estimate of the AQ by using RNN. Data occupied from stations. 5-fold cross-validation techniques are used for substantiation in estimating the AQI level. RNN is designed for training to estimate the AQI level with airborne pollution statistics. The RNN attempts to estimate the worth of every element. From the dataset, hidden layer per neuron cell (GRU or LSTM) and one as output.

C. Algorithms Used

For taking prediction we used machine learning techniques like a recurrent neural network (RNN), Long short-term memory (LSTM) i.e. Simple-LSTM, and Stack-LSTM. Here we have elaborated on the Simple-RNN, simple-LSTM, and Stack-LSTM.

(a) Simple Recurrent Neural Network (RNN):

The RNN method categorizes the vectors step by step. It passes the prior hidden state in the subsequent phase of the

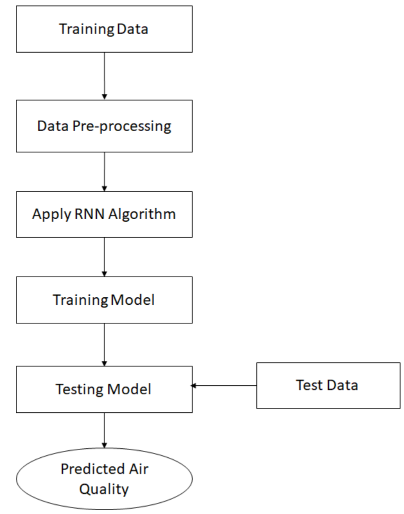

Fig. 1. Block Diagram of air quality prediction Model

classification. The hidden state performance like the neural system. Its storage data on earlier information the system has gotten in the past. Old neural systems do not have any memory. So, they do not receive into justification in earlier contributions when processing the present input. In consecutive datasets, as time-series, the data of the prior period phase are classically applicable for forecasting approximately in the present step. Therefore, a state approximately the prior period phases should be maintained. The midair pollution at time-period t capacity is manipulated by airborne pollution in prior time stages. RNNs have an interior loop, that keeps a state of prior time stages. This state is used for forecasting in the existing time stage. The state is rearranged, a new system is existence processed. In this experimentation, we have used Simple-RNN of the Keras package. We denote an Early-Stopping call-back to break the training, 15 epochs without slight development in the 'validation-loss'. The Model-Checkpoint permits us to store the weights of the greatest model.

1) Start
2) Input: Trained data.
3) Apply prepossessing steps.
4) Apply Recurrent Neural Network.
5) Build a trained model.
6) Apply test data to calculate 'validate-loss' model.
7) Calculate the efficiency of the system.
8) end.

(b) Simple LSTM:

LSTM design maintains a fixed value of the fault as it is posterior, reproduced through the structure. The slope receives less or higher changes posterior over the structure, remains the equivalent. It remembers the short-term fault, across an extensive period (many regressive time stages). We are working to move stage by stage via the procedure of generating an RNN. We use the Keras library to generate deep learning in python.

1) Start
2) Input: Trained data.
3) Apply prepossessing steps (Cleaning data).
4) Apply Long Short-term Memory.
5) Build a trained and test model.
6) Apply test data to calculate the 'validate-loss' model.
7) Calculate the efficiency of the system.
8) end

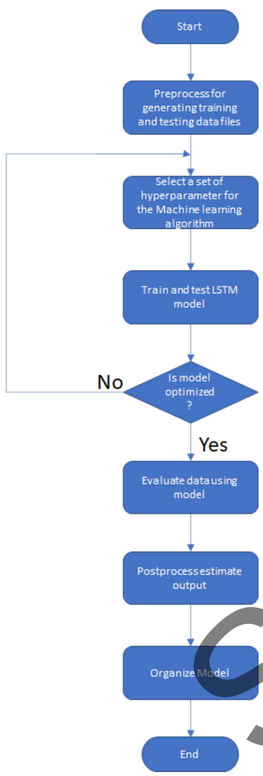

Fig. 2. LSTM Flowchart

(c) Stack LSTM:

We have used Stack LSTM layers on topmost. It makes the structure of the model accomplished by higher-level knowledge of historical representations. The first two LSTMs provide their complete output systems, but the third returns the last stage in its output system, reducing the sequential measurement. Thus, it is Stacking multiple LSTM layers. The model will absorb other concepts of input statistics over-time period. Represent the input information, not at the same time scale.

1) Start
2) Input: Trained data.
3) Apply prepossessing steps.
4) Apply Long Short-term Memory.
5) Add few hidden layers in Long Short-term Memory.
6) Build a trained and test model.
7) Apply test data to calculate the 'validate-loss' model.
8) Calculate the efficiency of the system.
9) end.

IV. SYSTEM ANALYSIS

A. MATHEMATICAL MODEL

The BE-1-2013-2015-aggregated- time series dataset is used for experimentation. As mentioned earlier, we have used RNN, Simple LSTM and StackLSTM for the experimentation. The algorithm used is as follows.

- Input: Air quality data FunctionS
- Pre-possess()-
  Simple-RNN(),
  Simple-LSTM(),
  Stack-LSTM(),
- Testing Phase().

Mathematical Model for RNN:

1) A time-period phase of the i/p is provided to the system.
2) Evaluations its present state used for established current i/p state and prior state.

$$h_{ps} = tanh(W_{rn} * h_{ps-1} + W_{in} * I_{is}) \quad (1)$$

   Where,

   h_{cs} : Present State,

   h_{cs-1} : Prior State,

   I_s : I/p State.

3) The present state (h_{ps}) has changed to the prior state (h_{ps-1}) for the subsequent period phase.
4) Several times periods afford to the difficulty and combined the information from the prior states.
5) Time periods are done, then ending present state is operated to evaluate the output.

$$O_{ot} = W_{ot} * h_{ps} \quad (2)$$

where,

O_{ot}: Output,

W_{ot}: weight at output layer.

6) The manufacture is associated to the actual output, in last produces output and the error.

7) The faults are the backbone to range to the structure to update the weight.

V. RESULT AND DISCUSSION

We use the European environment agency site for taking the BE-1-2013-2015-aggregatedtimeseries dataset. For research, we used 5-fold for the cross-validation technique. The dataset is specifies train data to the RNN for “validation_loss” is an experiment on testing data. The experiment, frequently repeated for Simple-LSTM and stack-LSTM. Table 1 shows the outcomes.

TABLE I
VALIDATION LOSS ON DIFFERENT ALGORITHM

Folds	SimpleRNN Validation-Loss	SimpleLSTM Validation-Loss	StackLSTM Validation-Loss
1	5.79	6.345	2.467
2	6.89	7.98	2.34
3	4.89	5.89	1.67
4	0.91	0.6292	0.3221
5	3.66	5.019	0.3781
Avg	4.428	5.1726	1.4354

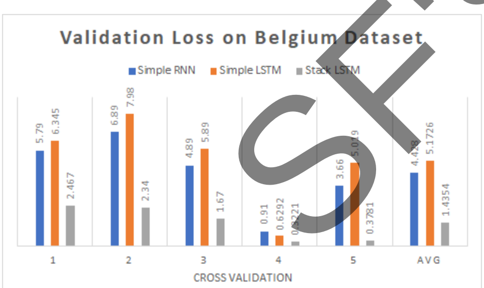

Fig. 3. Validation Loss using different method

Table 1 denoted the initial examination done for the forecast of AQI. Table 1 shows the validation_loss, taking on the BE-1-2013-2015-aggregatedtimeseries.csv dataset. It detects Stack-LSTM gives less validation_loss better as compared to Simple-RNN and SimpleLSTM. Figure 2 so show the outcomes in graphical form. Here we proposed to use StackLSTM, and as per our knowledge, Stack_LSTM performs better as compared to available techniques.

We use Kaggle for taking the data.csv dataset. For research, we used 5-fold for the cross-validation technique. The dataset specifies train data to the RNN for “validation_loss” is an experiment on testing data. The experimentation frequently repeats for SimpleLSTM and stackLSTM. Table 2 shows the outcomes.

TABLE II
VALIDATION LOSS ON DIFFERENT ALGORITHM

Folds	SimpleRNN Validation-Loss	SimpleLSTM Validation-Loss	StackLSTM Validation-Loss
1	0.578	0.0655	0.0485
2	0.058	0.0647	0.0473
3	0.0589	0.0653	0.0476
4	0.0591	0.0652	0.0479
5	0.0589	0.0655	0.0452
Avg.	0.16258	0.06524	0.0473

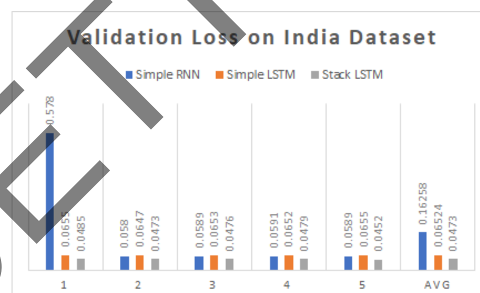

Fig. 4. Validation Loss using different method

Table 2 denoted the initial examination done for the forecast of AQI. Table 2 shows the validation_loss, taking on the data.csv dataset. It detects Stack-LSTM gives less validation_loss better as compared to SimpleRNN and SimpleLSTM. Figure 3 so show the outcomes in graphical form. Here we proposed to use StackLSTM, and as per our knowledge, Stack_LSTM performs better as compared to available techniques.

A. Visualization

We make a glossary of the airborne impurities and dataset quantity, systematic representation, term, and bin boundaries. The bin boundaries build on the scale. In the metadata, we have the organized design for every Sampling Point. We have required that data for design the Sampling Point on the map. Date variables converted into Datetime. We use it later to part the Pandas Dataframe.

1) Data cleaning:

- It takes the records in Data Aggregation Process.
- Eliminate records with the Unit of Airborne Smog Level of the amount.
- Eliminate variables terminated for visualization.
- Eliminate sampling point which has fewer than days of measurement.
- Assign the airborne pollution level to miss values with the importance of the subsequent binding date.

2) Plotting air pollution over time: Stack all of the days for entire sampling points too extensive for the plot. So, we re-sample the information by keeping the end day of every month. The bin boundary utilizes in this paper for experiments. They should generally apply on hourly (arithmetic) mean for NO2, O3, and CO. In the datasets, we only take the daily mean. Display the smog growing over day to day. We use the plugin of Folium for the TimestampedGeoJson. This plugin needs a geojson input structure. We create the function create-geojson-features to convert the information of the data frame.

1) Benzene (C6H6)

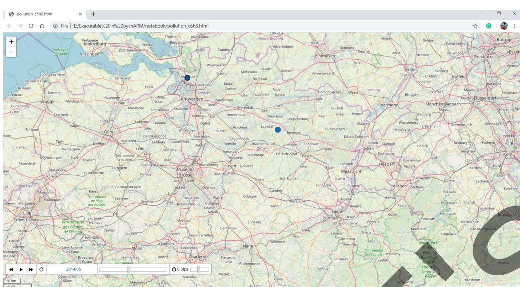

Fig. 5. Visualization air pollution map for Benzene

2) Carbon Monoxide (CO)

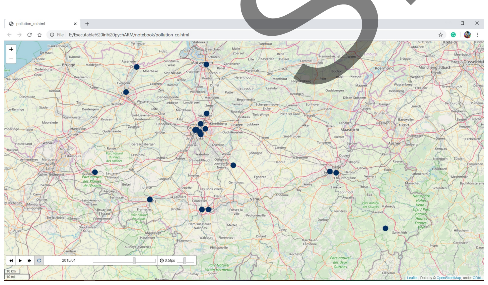

Fig. 6. Visualization air pollution map for Carbon Monoxide

3) Nitrogen dioxide (NO2)

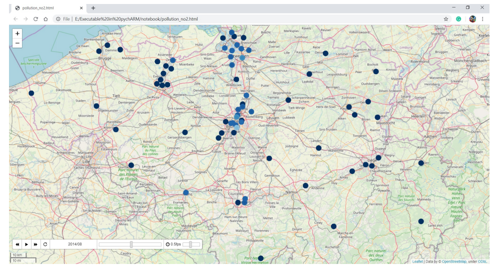

Fig. 7. Visualization air pollution map for Nitrogen Dioxide

4) Ozone (O3)

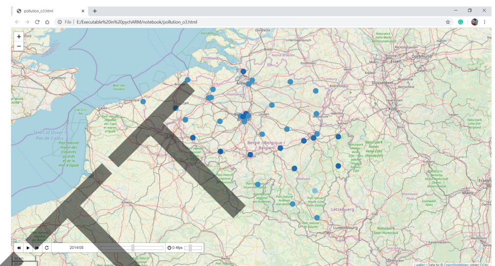

Fig. 8. Visualization air pollution map for Ozone

5) Particulate matter 10 (PM10)

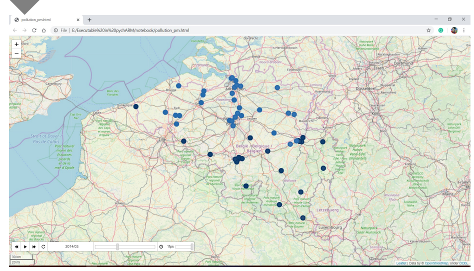

Fig. 9. Visualization air pollution map for Particulate matter-PM10

6) Sulphur dioxide (SO2)

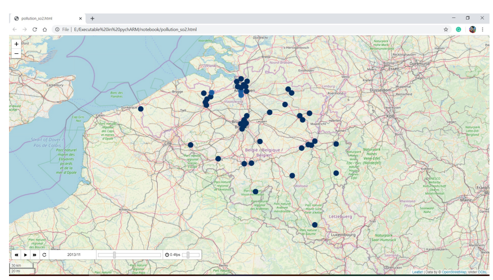

Fig. 10. Visualization air pollution map for Sulphur dioxide

VI. Conclusion

Airborne pollution is most dangerous to human health. Pollution affected public health and the environment. We used RNN, Simple-LSTM, and Stack-LSTM to forecast air quality. In this paper, we investigated with RNN and LSTM. As research, we accomplish that StackLSTM makes better as compared to Simple-LSTM and RNN. We used a RNN and different points designed for an LSTM. The better performance derives from the stack-LSTM involving a small number of hidden layers. There is absolutely a sum of belongings value with the additional examination that could expand the model execution.

- The hourly data used and attempt additional selection approaches than everyday data.
- The statistics about the supplementary impurities as structures to forecast SO2 smog. Maybe supplementary impurities are connected to the SO2 smog.
- Paradigm added structures built on the dates.

As per experimentation, we conclude that StackLSTM performs better as compared to Simple-LSTM and simple-RNN. In the future, we will banquet our process to determine the problem of recurrent neural network estimate and strategy extra metrics used for different observations. Also, we will integrate added exterior issues and discover new processing methods.

Acknowledgment

I have a tremendous pleasure in presenting the project "Air Quality Prediction Using Recurrent Neural Network" under the guidance of Dr. D. V. Patil and PG coordinator Prof. A. S. Vaidya. I am obligated and appreciative of the Head of Department Dr. D.V. Patil for their significant direction and consolation. I might likewise want to thank the Gokhale Education Society's R. H. Sapat College of Engineering, Management Studies, Research, and Nashik-5 for giving the required offices. Web get to and vital books. At last, I must express my sincere, heartfelt gratitude to all the Teaching Non-teaching Staff members of Computer Department of GES-RHSCOE who helped me for their important time, support, remarks, and thoughts.

References

[1] Choubin, B., Abdolshahnejad, M., Moradi, E., Querol, X., Mosavi, A., Shamshirband, S., & Ghamisi, P. (2019). Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. Science of The Total Environment, 134474. doi:10.1016/j. scitotenv. 2019. 134474.

[2] an, J., Wu, L., Ma, X., Zhou, H., & Zhang, F. (2019). Hybrid support vector machines with heuristic algorithms for prediction of daily diffuse solar radiation in air-polluted regions. Renewable Energy. doi:10.1016/j. Rene ne. 2019. 07. 104.

[3] Y. C. Lin, S. J. Lee, C. S. Ouyang, Chih-Hung Wu, Air quality prediction by neural-fuzzy modeling approach, Applied Soft Computing Journal (2019) 105898, https://doi.org/10.1016/j.asoc.2019.105898.

[4] Ding., Y., Li, Z., Zhang., C., and Ma., J., (2019)., Prediction of Ambient PM2.5 Concentrations Using a Correlation Filtered Spatial-Temporal Long Short-Term Memory Model., Applied Sciences, 10(1)., 14. doi:10.3390/app10010014.

[5] Zhang, D., and Woo, S. S. (2019). Predicting Air Quality using Moving Sensors. Proceedings of the 17th Annual International Conference on Mobile Systems, Applications, and Services - MobiSys '19. doi:10.1145/3307334.3328647.

[6] Tian, Y., Huang, W., Ye, B., & Yang, M. (2019). A New Air Quality Prediction Framework for Airports Developed with a Hybrid Supervised Learning Method. Discrete Dynamics in Nature and Society, 2019, 1–13. doi:10.1155/2019/1562537.

[7] Kuo, R. J., Prasetyo, B., & Wibowo, B. S. (2019). Deep Learning-Based Approach for Air Quality Forecasting by Using Recurrent Neural Network with Gaussian Process in Taiwan. 2019 IEEE 6th International Conference on Industrial Engineering and Applications (ICIEA). doi:10.1109/iea.2019.8715113.

[8] S. Shanthi and M. Pyingkodi, Air Quality Index Prediction using Machine Learning Algorithms, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-4, November 2019, DOI:10.35940/ijrte.D5326.118419.

[9] Zhu., D., Cai., C., Yang., T., and Zhou., X., (2018)., A Machine Learning Approach for Air Quality Prediction- Model Regularization and Optimization., Big Data and Cognitive Computing, 2(1), 5. doi:10.3390/bdcc2010005.

[10] Kok, I., Simsek, M. U., & Ozdemir, S. (2017). A deep learning model for air quality prediction in smart cities. 2017 IEEE International Conference on Big Data (Big Data). doi:10.1109/bigdata.2017.8258144.

[11] Tsai., Y.,-T., Zeng., Y.,-R., and Chang, Y.,-S., Air Pollution Forecasting Using RNN with LSTM. 2018 IEEE 16th International Conference on Dependable. Autonomic and Secure Computing., 16th International Conference on Pervasive Intelligence and Computing. 4th International Conference on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech). doi:10.1109/dasc/picom/Datacom/cyberscitec.2018.00178.

[12] Freeman B S , Taylor, G., Gharabaghi, B., and J The. (2018). Forecasting air quality time series using deep learning. Journal of the Air and Waste Management Association, 68(8), 866 – 886. doi-10.1080/10962247.2018.1459956.

[13] V, A., P, G., R, V., & K P, S. (2018). Deep Air Net: Applying Recurrent Networks for Air Quality Prediction. Procedia Computer Science, 132, 1394–1403. doi: 10.1016/j.procs.2018.05.068.