Journal of Advanced Zoology

*ISSN: **0253-7214***

*Volume **44** Issue **S-6 Year 2023** Page **561**:**567***

LSTM-Based Air Quality Prediction

Bhoomika H P[1*], Chandana L[2], Lavanya G M[3], Vishwanath R Hulipalled[4]

[1,2,3,4]School of Computing and Information Technology, REVA University Bangalore, Karnataka
Email: [1]bhoomikahp5@gmail.com, [2]chandanagowda2031.com, [3]lavanyagmlavi@gmail.coml.com, [4]vishwanath.rh@reva.edu.in

**Corresponding author's E-mail: bhoomikahp5@gmail.com*

Article History

Received: 06 June 2023
Revised: 05 Sept 2023
Accepted: 29 Nov 2023

Abstract

In this project we are using Neural networks and Long Short-Term Memory (LSTM) to address the air pollution detection. As each living organism needs fresh and good quality air for every moment, very few of the living things can survive without such air. Increasing industry and populace have end up fundamental contributor for the air pollutants. Over the time, many countries are finding numerous approaches of fighting towards air pollution. The air we breathe every moment causes several health hazards. So we'd like an honest system that predicts such pollutions and is useful in better environment. It leads us to address the advance techniques for predicting the pollution using Air Quality Index. So, here we are predicting air pollution using LSTM and Neural Network techniques for the coming hour mainly on pollutants like ammonia (NH3), lead (Pb), ozone(O3), carbon monoxide (CO), nitrogen dioxide (NO2), sulphur dioxide (SO2) and ecosystem aspects inclusive of temperature, strain, rainfall, wind pace according to minute and wind direction.

Keywords: *Neural networks, Air Quality Index, Recurrent neural network, LSTM.*

1. Introduction

Here Air quality is represented by AQI, which suggests whether the air is breathable or deteriorated in our ecosystem and the health challenges associated with it. This analysis of AQI is composed based on significant air toxins such as, O3, SO2, NO2, and CO. The combination of solid and liquid particles found in air is characterized as Particulate matter. These particles are very fine and studies have found that a there is a close link between PM2.5 exposure and death due to heart and lung disease. Thus, accurately predicting the PM2.5 values is of utmost concern in order to prevent disease and death. This would be a valuable prediction as governments could enact both temporary and permanent measures to reduce pollution for the safety of both the citizens and environment. We considered many techniques for the air quality index computation with accuracy. We use regression to decide for which factor we should give utmost priority and which to ignore. we use LSTM to overcome the problems of recurrent neural networks. Machine learning techniques provide enhanced methodology for analyzing quantitative data sets. It is often difficult to transport sensor data to cloud very often due to network congestion problems and data overloading problems. Processing real time data gives good results but is often a costly operation. The proposed model predicts any major pollutant in air with at most accuracy. The accuracy of the model depends on the data fed from the dataset and error values present in the dataset. We use Recurrent neural networks for sequential facts to carry out such operations, by this method we establish mitigation strategies and systemic action on pollution causing pollutants.

Neural Networks (RNN)-

RNNs works very well with sequence of data as input. Here in recurrent neural networks the output of preceding step is given as input to the present step. The recurrent neural network remembers all the information of the computations using memory. The recurrent neural networks are very much useful in time sequential prediction of data, as it can remember all the information. The only problem with RNNs is it suffers from vanishing gradient problem. And LSTM was created to overcome this for this problem.

LSTM-

Since RNN has limited memory, LSTM appeared as an answer for the RNN. In this, each layer is associated with next layer decreasing the impacts of short memory. LSTM is fairly like RNN yet the tasks inside the LSTM makes it distinctive that is by choosing to keep or fail to remember the data. LSTM utilizes cell and gates like input gate, output gate and forget gate which can handle how the information is processed. The cell state moves relative data to succession chain. Gates contains various activations such as sigmoid activations, tanh activations and relu.

Literature Survey

In the [1] paper IoT based air pollution monitoring and prediction system is proposed. This framework can be used for observing air poisons of a specific region. The proposed system uses combination of IOT and machine learning algorithm which is Recurrent Neural Network more clearly LSTM. Internet of things will connect things with one another by means of internet that mean every thing will have communication capacity. In this work the necessary parameters are noticed distantly utilizing web and the data gathered from the sensors are kept in the cloud and to expand the assessed drift on the program in order to estimate the contamination rate utilizing RNN.

In the [2] paper the main focus is on industrial pollution monitoring by nonlinear Auto Regressive model (NARX) based Artificial Neural Network. The influence of the inclusion of the meteorological variables (WD, WS, T, and HR) is studies to predict concentration of two pollutants (CO & NOx). Selection of the best network is done by determining the performances of NARX model in three phases. But the drawback is the paper only focuses on NOx and CO pollutants which may significantly vary over time and can lead to poor prediction due to other serious pollutants of air.

In the [3] paper main focus is on urban air pollution. Feature based selection to predict the best classifier method for a particular city based on WF-Chem. Examinations are done for 74 urban areas in china, to track down the best model for every city. Test results demonstrates that exactness can be improved by using more features. For strategy perspective, the outcome from consolidated model is superior to the special model but the drawback is Feature selection for datasets containing huge number of features would be a hectic task.

In the [4] paper the Convolution Recurrent Neural Networks Based Dynamic Transboundary Air pollution prediction is presented. Transboundary air pollution is one of the main sources of air pollution in island. The number of air quality monitoring stations in island cities are less than it inland cities The network can simultaneously model the interactions between each atmospheric monitoring data in temporally and spatially. A dynamically asynchronous transboundary pollution prediction approach with convolution recurrent neural networks design is used to do the prediction in island cities much more accurate.

Problem Definition

Air pollution is pressing issue which needs to delt based on various parameters which affect human health and are deemed to be unsafe.

- It is for this multi dimensional problem that we require historical data and analytical solutions.
- So we use data mining techniques and pattern of useful data in extracting accurate knowledge.
- Which will establish a robust data collecting methodology from various places, which can be interpreted for better judgement. By doing so we eliminate sparse and inadequate data.
- This model will help having better network of data available for air pollution index.

Objectives

- This project provides a well-organized method of addressing pollutant concentration regarding pollutants causing Air pollution.
- To predict air quality index to avoid future problems caused due to air pollution.
- The proposed model predicts any major pollutant in air with at most accuracy.

2. Materials And Methods

A*) Data collection*: We collected the datasets from UPC repository. In this case of dataset, we consider attributes such as pm2.5, pm10, temperature, wind direction, dew point, SO2, NO2, CO, O3. The dataset is normalized to contain uniform values within specific range (0 to 1). A clean dataset with no null values. Since machine learning models cannot accept null values.

B) *Data pre-processing*: The dataset is converted from series to supervised dataset and is labelled appropriately. And since attribute such as wind direction are in string format it should be converted to numeric form and this is done using suitable label encoder.

C*) Data visualization*: Different plots are plotted for different contributors of air quality index. Each plot provides observations about the shift and swing of the pollutant concentration in the surrounding air. We use x-axis to represent the number of sample where as y-axis to represent the concentration values in µg/m3. In the below mentioned graphs shows the rise and fall of various pollutants concentration such as SO2, NO2, O3, CO and ecosystem aspects inclusive of pressure, dew point concentration is shown.

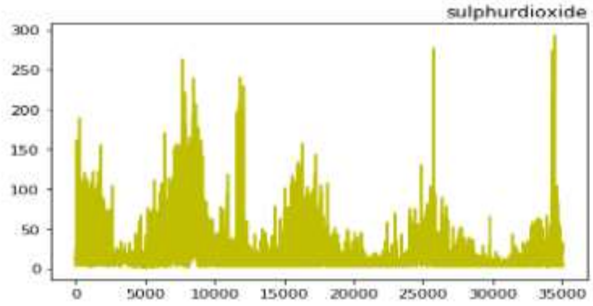


Fig 1. SO2 concentration in air

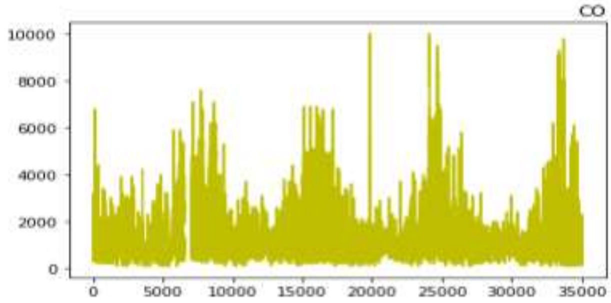


Fig 2. CO concentration

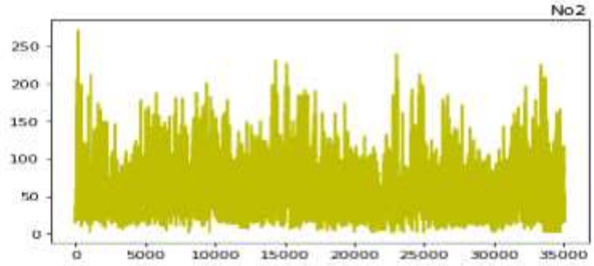


Fig 3. NO2 concentration

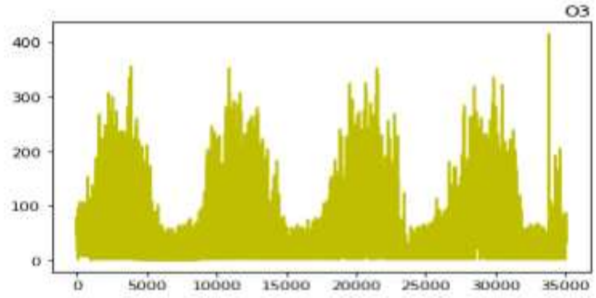


Fig 4. O3 concentration

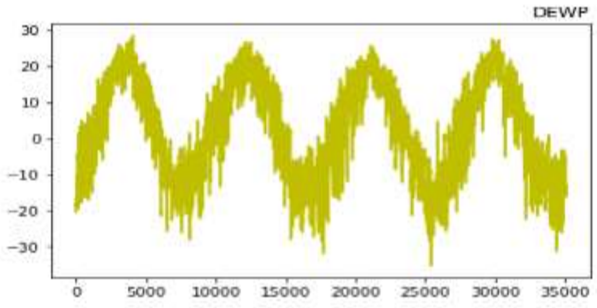


Fig 5. Dew point concentration

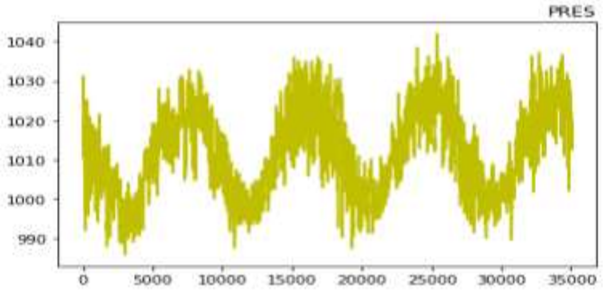


Fig 6. Pressure

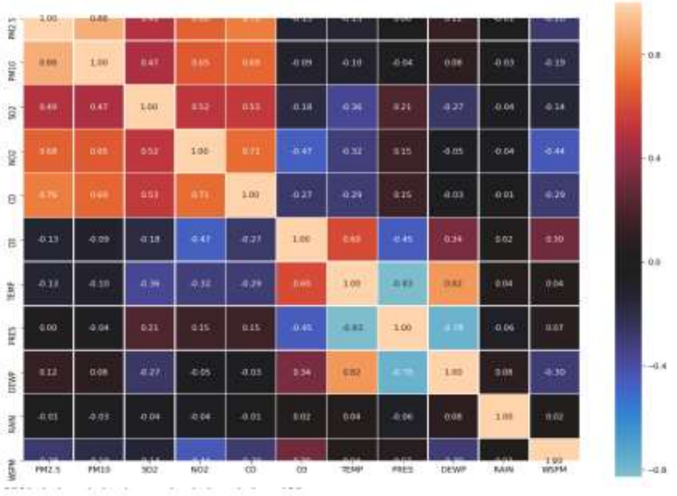


Fig 7. Correlation heatmap representing correlations between various pollutants.

D) *Model fitting*: Here the encoded dataset is split into testing data and training data. Further splitting training and testing datasets into input and output variables. The network parameters is to be set before the preprocessed dataset is provided to the LSTM Model to make predictions. One of the parameters for the neural network is an optimizer which is used change various features such as learnings rates, weights, bias and so on. And there are several types of optimizers available for neural networks, out of which we have opted for Adam optimizer In addition to than set we also set the activation function Relu.

E) *Evaluate performance*: To evaluate the accuracy of prediction for the given model the mean square error is computed. This proposed model gives an error of about 0.1

F) *Expected output*: In the output module, the predicted result is compared with the air quality index ranges of a specific city to determine whether the air is barely, moderately or highly polluted and output is given. In the below mentioned graph x-axis represents epoch and y -axis represents loss.

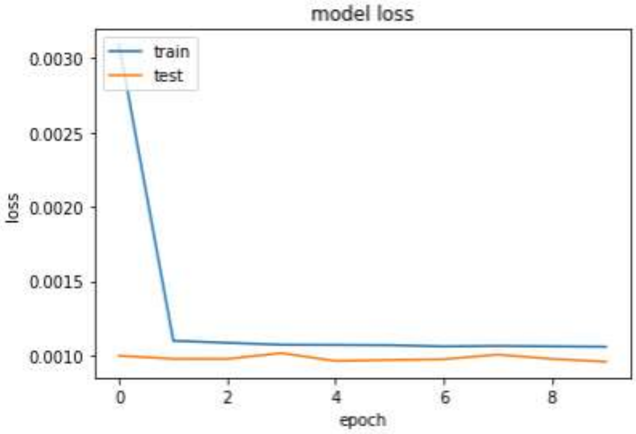
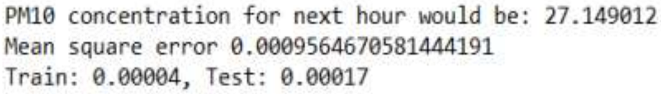


Fig 8. MSE error plot



```
PM10 concentration for next hour would be: 27.149012
Mean square error 0.0009564670581444191
Train: 0.00004, Test: 0.00017
```

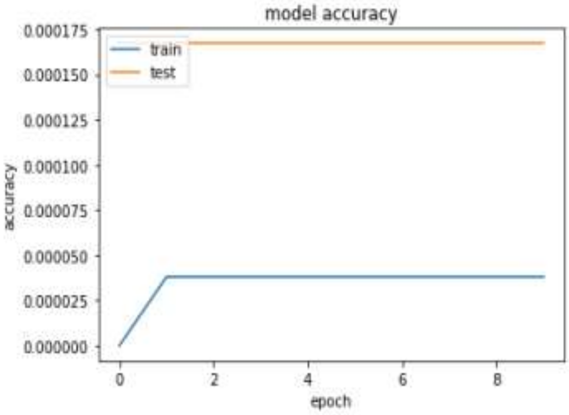
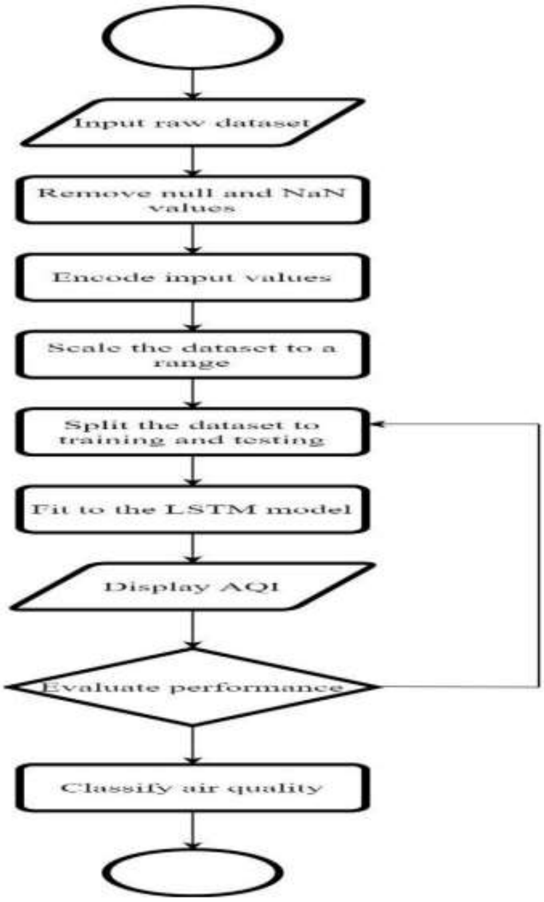


Fig 9. Model accuracy plot

PROCESS CHART



4. Conclusion

Good quality air is the top most priority for all the living organisms. The proposed work uses nominal preprocessing and machine learning techniques and avoids over fitting of data. It is often difficult to transport sensor data to cloud very often due to network congestion problems and data overloading problems. Processing real time data gives good results but is often a costly operation. In this paper, the LSTM Model outputs the concentration of pollutants and predicts Air quality for each day using appropriate datasets, The proposed system can be further trained using abundant dataset values with accurate PM 2.5 concentration values. Here the neural network algorithm is expected to give the maximum accuracy of 92%. In future this can be unified in a real time web application and can also be used for multistep prediction or it can be used by different weather stations [6-11].

References:

[1] Temesegan Walelign Ayele, Rutvik Mehta, "*Air pollution monitoring and prediction using IoT*", IEEE Conference 2018.

[2] Qin, D.; Yu, J.; Zou, G.; Yong, R.; Zhao, Q.; Zhang, B. *A Novel Combined Prediction Scheme Based on CNN and* LSTM for Urban PM2.5 *Concentration.* IEEE Access 2019

[3] Peijiang Zhao, Koji Zettsu*," Convolution Recurrent Neural Networks Based Dynamic Transboundary Air Pollution Prediction*", IEEE conference.

[4] Zhao Wei, Wang Yijie, Bai Xinxin, Rui Xiaoguang, Yin Wenjun, Don Jin, Xia Xi"A *Comprehensive Evaluation of Air Pollution Prediction Improvement by a Machine Learning Method*", IEEE Conference 2015.

[5] Nadjet_Djebbri, Mounira_Rouainia," *Artificial Neural Networks Based Air Pollution Monitoring in Industrial Sites*", IEEE Conference 2017.

[6] Sudhan Murugan Bhagavathi, Anitha Thavasimuthu, Aruna Murugesan, Charlyn Pushpa Latha George Rajendran, A Vijay, Raja Laxmi, Rajendran Thavasimuthu, Weather forecasting and prediction using hybrid C5.0 machine learning algorithm International Journal of Communication Systems, Vol. 34, Issue. 10, Pp. e4805, 2021.

[7] PM Surendra, S Manimurugan, A New Modified Recurrent Extreme Learning with PSO Machine Based on Feature Fusion with CNN Deep Features for Breast Cancer Detection, Journal of Computational Science and Intelligent Technologies, Vol. 1, Issue. 3, Pp. 15-21, 2020.

[8] PK Sadineni, Comparative Study on Query Processing and Indexing Techniques in Big Data, 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), pp. 933-939, 2020.

[9] AH Omar Baabood, Prajoona Valsalan, Tariq Ahmed Barham Baomar, IoT Based Health Monitoring System, Journal of Critical Reviews, Vol. 7, Issue. 4, pp. 739-743, 2020.

[10] Sajay KR, Suvanam Sasidhar Babu, Vijayalakshmi Yellepeddi, Enhancing The Security Of Cloud Data Using Hybrid Encryption Algorithm, Journal of Ambient Intelligence and Humanized Computing, 2019. https://doi.org/10.1007/s12652-019-01403-1

[11] Bindhia K Francis, Suvanam Sasidhar Babu, Predicting academic performance of students using a hybrid data mining approach, Journal of Medical Systems, 43:162, 2019. https://doi.org/10.1007/s10916-019-1295-4