The Journal of Multidisciplinary Research (TJMDR)
Content Available at www.saap.org.in Online ISSN: 2583-0317

Research Article Open Access

A novel approach for air quality prediction using machine learning approach

G.V.P.S.Sruthi[1], M.Lokesh[2], A.Grace[3], S.L.Reddy[4], B.Srinivas Raja*
[1,2,3,4] Department of Electronics and Communication Engineering, Godavari Institute of Engineering and Technology, Andhra Pradesh, India
*Department of Electronics and Communication Engineering, Godavari Institute of Engineering and Technology, Andhra Pradesh, India,

Article History: Received: 11 Sept, 2021 Revised: 17 Nov, 2021 Accepted: 30 Nov, 2021

Abstract
We forecast the air quality by using machine learning to predict the air quality index of a given area. Air quality index of India is a standard measure used to indicate the pollutant (so2, no2, rspm, spm. etc.) levels over a period. We developed a model to predict the air quality index based on historical data of previous years and predicting over a particular upcoming year using machine learning methods.Our model will be capable for successfully predicting the air quality index of any bounded region provided with the historical data of pollutant concentration.In our model by implementing the proposed parameter reducing formulations, we achieved better performance than the standard regression models.This project work can help in constructing air quality, using machine learning methods such as XGBoost, Random Forest (RF) and Convolution Neural Network (CNN).Among these methods, Convolution Neural Network produces a better result in terms of accuracy value of about 91% compared to other algorithms.
Keywords: Air Quality Index , rspm, spm, Pollutant , Machine Learning.

Corresponding Author
B.Srinivas Raja *

1.INTRODUCTION

As the largest growing industrial nation, India is producing record amount of pollutants specifically Co2, pm2.5 etc and other harmful aerial contaminants.

In this project, we calculate the individual index of the pollutant for every available data points and find their respective AQI for the region.

We have designed a model to predict the air quality index of every available data points in the dataset, our model is capable of forecasting the air quality of India in any given area.

By predicting the air quality index, we can backtrack the major pollution causing pollutant and the location affected seriously by the pollutant across India.

2. RELATED WORKS

Yang Gong; Pan Zhang :Since entering modern society, people have paid more and more attention to air quality in order to better help predict the air quality level. This paper proposes an air quality grade prediction model based on the K-nearest neighbor algorithm. Firstly, the historical measurement data of air quality is crawled from the relevant weather website and saved to the local CSV file; then the data is read, and the scatter diagram is used to visually display the 6 characteristics that affect the air quality level evaluation; then the K nearest neighbor algorithm is selected, and the difference is adjusted. Nowadays, air pollution has reached critical levels and the air pollution level in many major cities has crossed the air quality index value as set by the

government. It has a major impact on the health of the human.

Anıl Utku; Umit Can : Today, air pollution in many regions is still above the limits indicated by the World Health Organization. In this study, the prediction of the rate of PM 2.5 , which is an important air pollutantis emphasized. For this purpose, weather prediction models were created using Support Vector RegressionThe model uses the characteristics of nonlinear fitting approximation of BP neural network to solve the problem that air quality has many influencing factors and is nonlinear and difficult to predict.

Hong Zheng; Yunhui Cheng; HaibinLi :Air pollution which is detrimental to people's health is a wide spread problem across many countries around the world. Developing better air quality prediction approaches is an important research issue. Existing methods often focus on the prediction of air pollution concentrations, which is not as intuitive to the public as the air quality levels. In this paper, near future fine-grained air quality level prediction task is explored with a series of machine learning ensemble methods. Included ensemble methods are majority voting, averaging, weighted averaging and 16 different stacking tactics.

3. PROPOSED METHODOLOGY

Existing System :

Random forests or **random decision forests** are an essemble learning method for the purpose of classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean/average prediction (regression) of the individual trees.

The **Gradient Boosting Algorithm(XGB)** is a machine learning technique which is for the regression and classification problems, which produces a prediction model in the form of an ensemble of the weak prediction models, typically decision trees. When a decision tree is the weak learner, the resulting algorithm is called gradient boosted trees, which usually outperforms random forest. It builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function .

Proposed System :

A**convolutional neural network** (**CNN**, or **ConvNet**) is a class of deep neural networks, most commonly applied to analyzing visual imagery. They are also known as **shift invariant** or **space invariantartificial neural networks** (**SIANN**), based on their shared-weights architecture and translation invariance characteristics. They have applications in image and video recognition , recommender systems , image classification , image segmentation , medical image analysis , natural language processing , brain – computer interfaces , and financial time series .

4. BLOCK DIAGRAM

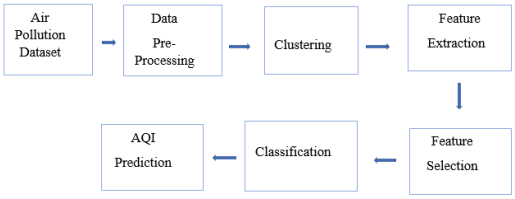

Fig 1. Block Diagram

Convolutional neural networks are composed of multiple layers of artificial neurons. Artificial neurons, a rough imitation of their biological counterparts, are mathematical functions that calculate the weighted sum of multiple inputs and outputs an activation value. When you input an image in a ConvNet, each layer generates several activation functions that are passed on to the next layer.

5 .COMPARATIVE STUDY OF EXISTING AND PROPOSED SYSTEM :

In our project traffic prediction is done by algorithms namely X Gradient Boost (XGB), Random Forest (RF) and Convolution Neural Network (CNN) in terms of accuracy. In the existing system the major drawback is less accuracy but in proposed system we get good accuracy in prediction.

As the linear regression is a regression algorithm, we will compare it with other regression algorithms. One basic difference of linear regression is, LR can only support linear solutions. There are no best models in machine learning that outperforms all others, and efficiency is based on the type of training data distribution.

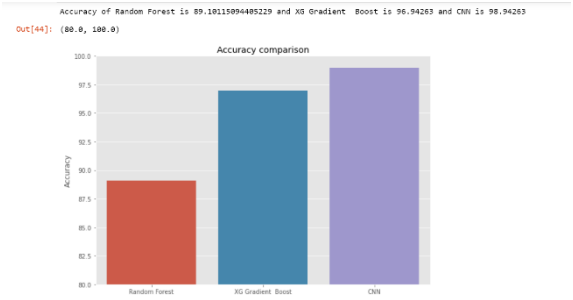

Fig 2 . Accuracy Comparision

6. DATA ANALYSIS

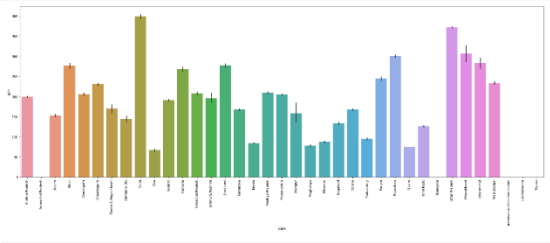

Fig 3. Collected Data of spm

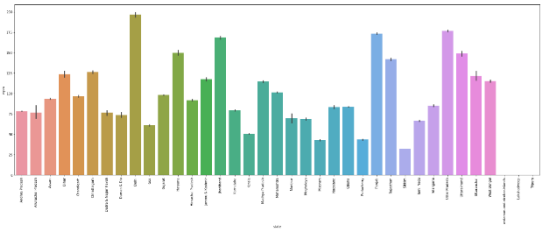

Fig 4 . Collected Data of rspm

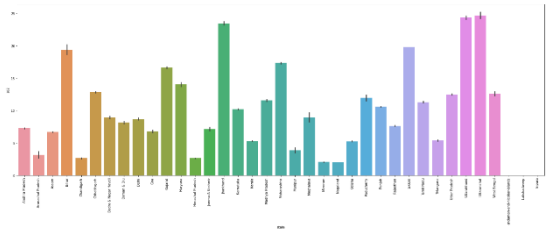

Fig 5 . Collected Data of SO_2

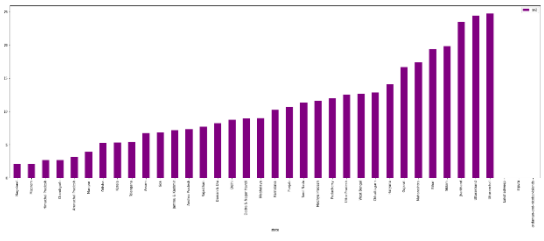

Fig 6 Sorting SO_2Data

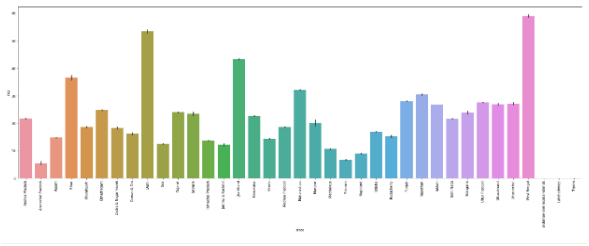

Fig 7 . Collected Data ofNO_2

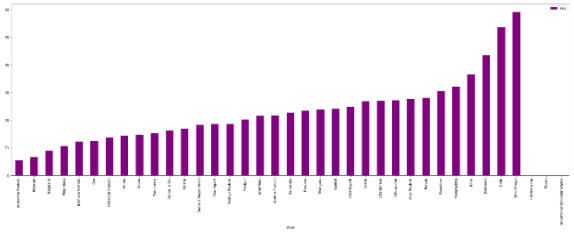

Fig 8 . Sorting of NO_2 Data

7. OUTCOME

	state	location	type	so2	no2	rspm	spm	pm2_5	SOI	Noi	Rpi	SPMi	AQI	AQI_Range
0	Andhra Pradesh	Hyderabad	Residential, Rural and other Areas	4.8	17.4	0.0	0.0	0.0	6.000	21.750	0.0	0.0	21.750	Good
1	Andhra Pradesh	Hyderabad	Industrial Area	3.1	7.0	0.0	0.0	0.0	3.875	8.750	0.0	0.0	8.750	Good
2	Andhra Pradesh	Hyderabad	Residential, Rural and other Areas	6.2	28.5	0.0	0.0	0.0	7.750	35.625	0.0	0.0	35.625	Good
3	Andhra Pradesh	Hyderabad	Residential, Rural and other Areas	6.3	14.7	0.0	0.0	0.0	7.875	18.375	0.0	0.0	18.375	Good
4	Andhra Pradesh	Hyderabad	Industrial Area	4.7	7.5	0.0	0.0	0.0	5.875	9.375	0.0	0.0	9.375	Good

8. AQI RANGE

AQI	Descriptor	General Health Effects
0–50	Good	None
51–100	Moderate	Few or none for the general population
101–200	Unhealthy	Everyone may begin to experience health effects; members of sensitive groups may experience more serious health effects. To stay indoors.
201–300	Very unhealthy	Health warnings of emergency conditions. The entire population is more likely to be affected.
301+	Hazardous	Health alert: everyone may experience more serious health effects

9. CONCLUSION

Since our model is capable of predicting the current data with 91% accuracy it will successfully predict the upcoming air quality index of any particular data within a given region. With this model we can forecast the AQI and alert the respected region of the country also it a progressive learning model it is capable of tracing back to the particular location needed attention provided the time series data of every possible region needed attention.

The air quality information utilized in this project originates from the china air quality checking and investigation stage, and incorporates the normal every day fine particulate issue (PM2.5), inhalable particulate issue (PM10), ozone (O3), CO, SO2, NO2 fixation and air quality record (AQI). The essential perspectives that should be viewed as with regards to gauging of the poison focus are its different sources alongside the components that impact its fixation. It was observed that classification implemented by CNN technique in this research is more efficient compare to existing algorithms as seen in the accuracy and precision.

10 . FUTURE SCOPE

It was observed that classification implemented by CNN technique in this research is more efficient compare to existing algorithms as seen in the accuracy and precision. In future we can do for deep learning approaches.

India meteorological department wants to automate the detecting the air quality is good or not from eligibility process (real time).

11. REFERENCES

[1] Verma, Ishan, Rahul Ahuja, HardikMeisheri, andLipikaDey. ”Air pollutant severity rediction using Bi-directional LSTM Network.” In 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), pp. 651-654. IEEE, 2018.

[2] Figures Zhang, Chao, Baoxian Liu, Junchi Yan, Jinghai Yan, Lingjun Li, Dawei Zhang, XiaoguangRui, and RongfangBie. ”Hybrid Measurement of Air Quality as a 5 Fig. 8. RH w.r.t tin oxide Fig. 9. RH w.r.t C6H6 Mobile Service: An Image Based Approach.” In 2017 IEEE International Conference on Web Services (ICWS), pp. 853- 856. IEEE,2017.

[3] Yang, Ruijun, Feng Yan, and Nan Zhao. ”Urban air quality based on Bayesian network.” In 2017 IEEE 9th Fig. 10. RH w.r.t NO Fig. 11. RH w.r.t NO2 International Conference on Communication Softwareand Networks (ICCSN), pp. 1003-1006. IEEE,2017.

[4] Ayele, TemeseganWalelign, and RutvikMehta.”Air pollution monitoring and prediction using IoT.” In 2018 Second International Conference on Inventive Communication 6 Fig. 12. RH w.r.t Temperature Fig. 13. RH w.r.t CO and Computational Technologies (ICICCT), pp. 1741-1745. IEEE,2018.

[5] Djebbri, Nadjet, and MouniraRouainia. ”Artificial neural networksbased air pollution monitoring inindustrial sites.” In 2017 International Conference on Engineering and Technology (ICET), pp. 1-5. IEEE,2017.

[6] Kumar, Dinesh. ”Evolving Differential evolution method with random forest for prediction of Air Pollution.” Procedia computer science 132 (2018): 824-833.

[7] Jiang, Ningbo, and Matthew L. Riley. ”Exploring the utility of the random forest method for forecasting ozone pollution in SYDNEY.” Journal of Environment Protection and Sustainable Development 1.5 (2015): 245-254.

[8] Svetnik, Vladimir, et al. ”Random forest: a classification and regression tool for compound classification and QSAR modeling.” Journal of chemical information and computer sciences 43.6 (2003): 1947-1958.

[9] NAAQS Table. (2015). [Online]. Available: https://www.epa.gov/criteria-air-pollutants/naaqs-table 3.