Advances and Applications in Mathematical Sciences
Volume 20, Issue 11, September 2021, Pages 2509-2518

COMPREHENSIVE ANALYSIS OF PREDICTING AIR QUALITY USING NEURAL NETWORK MODELS

SUJATA OAK and DEVESH JOHARAPURKAR

Department of Information Technology
Ramrao Adik Institute of Technology
Navi Mumbai, India
E-mail: sujata.oak@rait.ac.in
devesh9j@gmail.com

Abstract

The rise in air pollution is causing health problems to millions of lives. Urbanization, industrialization, increase in number of vehicle and many more are the reasons behind sudden increase in air pollution. Location, time, surroundings and other uncertain things affect the quality of air. To avoid above problems, we are creating a system. There occurs a mild health problem, when we are unprotected to air over a period of time like shortness of breath, eye irritation, skin irritation; all the problems are linked to the level of pollution. In this system, we are focusing on RSPM and AQI levels. We are using neural network learning models like LSTM, CNN. These models will help in the prediction of unhealthy levels of pollution.

I. Introduction

Health care is a major issue in our country since the environmental hygiene is below standard. The air quality of our surrounding is degrading day by day due to increase in pollution. There are many components in air like gases and particles such as SO2, CO2, and PM2.5. SO2 causes shortness of breath, coughing and PM2.5 gets inhaled and irritate and damage alveoli and consequently impair lung functions [6]. If these things we inhale regularly more than expected quantity or breathable level, it might lead to death. There are various reasons that increase the amount of PM2.5, SO2 and other harmful things in air. One of the main reasons for increment of these things is industrialization. The plantation of industries causes exhaustion of

2010 Mathematics Subject Classification: 92B20.

Keywords: Air Quality Index (AQI); Respiratory Suspended Particulate Matter (RSPM); Long Short Term Memory (LSTM); Convolution Neural Network (CNN).

Received September 30, 2020; Accepted October 23, 2020

harmful gases in air through chimneys [7]. There are many air quality monitoring systems installed and setup in many cities throughout the country by the state government and central government to monitor air quality at that particular place. In this system, we are going to predict the air quality by using deep learning (DL) algorithms. The historical data we get from official websites will include some variables such as SO2, CO2, PM2.5, etc. [9] We are using deep learning algorithms because it gives more accurate results than traditional machine learning approach. It trains on much larger dataset than we use in machine learning, it learns patterns in data and analyse it to predict the outcome [10].

II. Related Work

Ibrahim K. OK, Mehmet Ulvi SIMSEK, Suat OZDEMIR proposed a model in 2017, in that they have managed to get data through IOT devices. They have used LSTM. They have divided their model into three sections. In first section, LSTM neural network model is built and experiments conducted for the parameters to give its best LSTM structure and SVR model is trained for evaluating its success. In the subsequent section, a naming unit is made that names information as indicated by the day by day AQI readings. In section three, a choice unit is created which depicts as indicated by the watched and anticipated caution circumstances. Their acquired outcomes represent that the work of the LSTM based forecasting model to IOT information is powerful and satisfying [1].

Ziyue Guan, Richard O. Sinnott proposed a model in 2018, in that they have used three models for the prediction of PM2.5. They have used linear regression, artificial neural network (ANN) and LSTM. Then, they compared these models and came to a conclusion that LSTM gave them the best results. LSTM performed best and was able to estimate high PM2.5 values with reasonable precision. ANN and linear model have their downsides in forecast of excessive PM2.5, they offered sensible performance. [2]

Timothy M. Amado, Jennifer C. DelaCruz, proposed a model in 2018, in this model, they have used various sensors that can measure the temperature, humidity, quantity of gases in air and many other. They have applied five methods to build predictive model for air quality. They have used

KNN, SVM, Random Forest, Naive Bayes and neural network. Among all these methods, neural network gave them the best results [3].

Chavi Srivastava, Shyamli Singh, Jen Amit Prakash Singh proposed a model where they estimate the air pollution in Delhi using various machine learning algorithms such as linear regression, decision tree, support vector regression etc. they have used PM2.5, PM10, SO2,CO2, etc as variables. For evaluation of model they used mse, mae, R2 [4].

Pratyush Singh, Lakshmi Narasimhan T, Chandra Shekar Lakshminarayanan proposed a model in 2019 that uses LSTM network model for various variables such as PM2.5, CO2, etc. They used 50 neurons each in 3 layers of network. They have used root mean squared error (RMSE) for measuring the measure [5].

III. Problem Statement

The rise in air pollution is causing health problems to millions of lives. Urbanization, industrialization, increase in number of vehicle and many more are the reasons behind sudden increase in air pollution. Location, time, surroundings and other uncertain things that affects the quality of air. This system will help in forecasting air quality using neural networks with the help of historical data sets.

IV. Data Collection

In this system, we have gotten our data from the official Maharashtra Pollution Control Board (MPCB) website. As the data is being provided by the government of Maharashtra it is trusted. Data should be complete and there should be no missing values. All the values in dataset should have only one unit for that particular variable. So, data is an important factor in any type of data analytics.

Index Value	Name	Color	Advisory
0 to 50	Good	Green	None
51 to 100	Moderate	Yellow	Unusually sensitive individuals should consider limiting prolonged outdoor exertion
101 to 150	Unhealthy for Sensitive Groups	Orange	Children, active adults, and people with respiratory disease, such as asthma, should limit prolonged outdoor exertion
151 to 200	Unhealthy	Red	Children, active adults, and people with respiratory disease, such as asthma, should avoid prolonged outdoor exertion; everyone else should limit prolonged outdoor exertion
201 to 300	Very Unhealthy	Purple	Children, active adults, and people with respiratory disease, such as asthma, should avoid outdoor exertion; everyone else should limit outdoor exertion
301-500	Hazardous	Maroon	Everyone should avoid all physical activity outdoors.

Figure I. AQI Chart.

The above chart indicates the AQI values and corresponding level of threat to it.

V. Proposed Model

This section describes the work flow of the project and the system architecture. It describes how the data will be used in various stages of the project and the various algorithms and modules that will act upon it.

Extraction. In this system we will be collecting the historical data of air quality from the trusted sources. Maharashtra government site provides this data through their website named Maharashtra Pollution Control Board. We will store this data as a .csv (comma separated values) file.

Input. In the implantation of this system we will first import the libraries which we are going to use such as numpy, pandas. These libraries are going to help us in importing, managing of the data. Then we import the dataset using pandas library.

Pre-processing. Data pre-processing is a procedure of converting raw data into a reasonable format. True information is commonly inadequate, conflicting and is probably going to contain mistakes. In this block, the data gets cleansed, noise and outliers get removed, and missing data is handled and other things. So, this method is used to resolve those issues and prepare it for further processing.

Model Building. We will build LSTM and CNN model. We will give some input variables to the input layer which will be relevant to the outcome. Then the second layer is hidden layer, in hidden layer we have to choose the number of hidden layers as well as number of neurons or units in each hidden layer. Also, we have to specify that if it is going to be the last hidden layer. The last layer is output layer, in which we have to specify output variable. We also have to specify the optimizer in this stage. Analyse the output predicted by this block. And iterate this process by adjusting some values to get the best model.

Visualization of Results. In this block, we are just visualizing the predicted outcome given by the model we trained in the previous block. The results are represented in graphical format and as well as in tabular format.

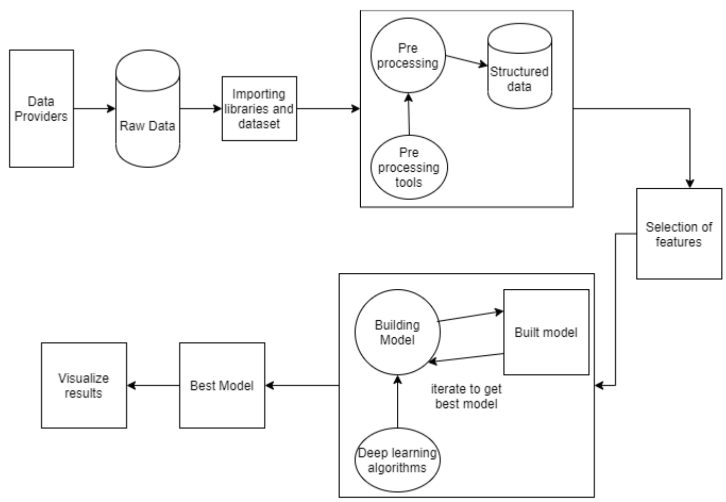

Figure II. Block Diagram 3.

VI. Types of Model

A. LSTM Model

For LSTM model building, we import packages from keras. After that, we have sequential function for initializing, LSTM function for LSTM layers, dropout function for eliminating neurons and a dense layer for output.

B. CNN Model

For CNN model building we need to import packages such as convolutional layer, pooling layer, flatten layer and dense layer. Each layer has its own job. The parameters provided to these are in the next section.

VII. Results and Analysis

The parameters we used to build our model are as below.

For LSTM MODEL for CNN MODEL

Table 1. LSTM Parameters Table II: CNN Parameters.

Hyperparameters	Values	Hyperparameters	Values
Input layer	1	Input layer	1
Hidden layers	3	Filter	64
LSTM units	50	Kernel size	2
Dropout	0.2	Dense layer	3
Output layers	1	Pool size	2
Batch size	32	Verbose	1
Number of epochs	100	Number of epochs	100
Optimizer	Adam	Optimizer	Adam
Loss	Mean squared error	Loss	Mean squared error

We have represented the outcome of model in graphical format as follows:

A. LSTM Model Graph

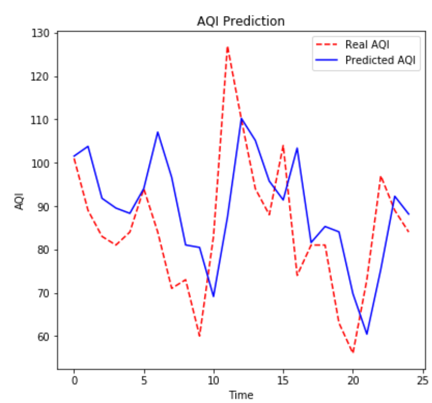

Figure III. Graph of AQI using LSTM.

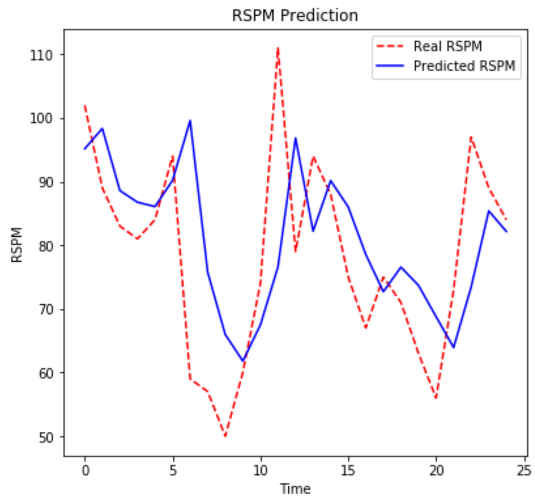

Figure IV. Graph of RSPM using LSTM.

As you can see from the graphs that the LSTM model was able to capture the flow in both AQI readings and RSPM readings.

B. CNN Model Graph

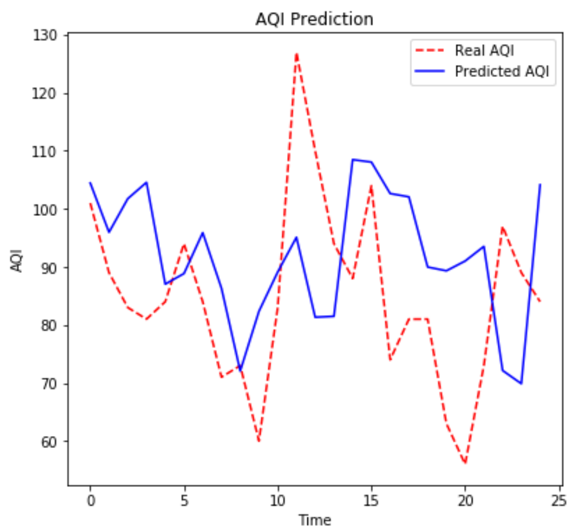

Figure V. Graph of AQI using CNN model.

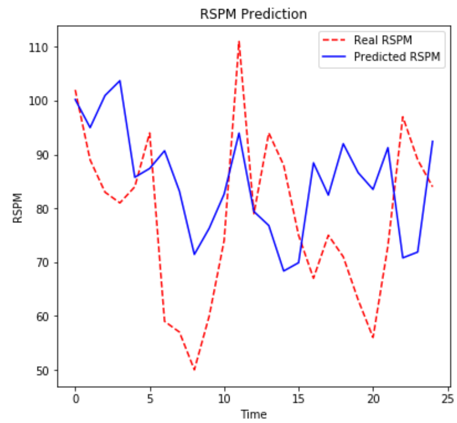

Figure VI. Graph of RSPM using CNN model.

As you can see from the graphs that CNN model has not captured the trend as better as LSTM did. The below table shows the error calculated in model. We have used two different methods of error calculation.

Table III. Error Evaluation.

Prediction Model	Error Evaluation	
	RMSE	MAE
LSTM-for AQI	0.45	0.002
LSTM-for RSPM	0.69	0.002
CNN-for AQI	2.89	0.0295
CNN-for RSPM	0.52	0.0295

The above table gives error values corresponding to that predictive model. We have made a table that compares actual values and the predicted values which we have gotten by respective model.

A. LSTM Table

Table IV: Comparison of LSTM.

AQI Actual	AQI Predicted	RSPM Actual	RSPM Predicted
101	96.13598	102	92.25765
89	87.46323	89	94.28902
83	86.15217	83	84.50819
81	97.286964	81	80.25763
84	86.32033	84	78.69816
94	93.10336	94	81.67512
84	81.9397	59	91.61615
71	101.111725	57	65.84814
73	74.70179	50	52.778763
60	76.96658	60	50.997913
83	87.808846	74	61.89905
127	94.46663	111	74.81721
110	91.60618	79	101.922455
94	81.12796	94	85.627754
88	87.76121	88	89.954346
104	84.86558	75	86.70084
74	82.878204	67	77.69414
81	81.61398	75	70.60411
81	90.085686	71	74.15935
63	95.477036	63	72.16263
56	94.52805	56	67.37898
73	94.99644	73	62.325535
97	73.7042	97	72.03344
89	69.72489	89	84.92557
84	86.887924	84	82.48866

B. CNN Table

Table V. Comparison of CNN.

AQI Actual	AQI Predicted	RSPM Actual	RSPM Predicted
101	84.1923	102	82.158455
89	86.358505	89	84.07325
83	93.557976	83	91.07833
81	94.54599	81	92.22309
84	85.02984	84	86.80616
94	76.95362	94	77.7493
84	70.09761	59	69.1592
71	71.69368	57	71.6137
73	67.920815	50	65.94495
60	75.68939	60	76.24295
83	83.57478	74	84.63607
127	83.02535	111	83.065384
110	77.47689	79	78.87327
94	75.22511	94	75.11769
88	92.39587	88	71.894875
104	78.24991	75	70.409645
74	86.88046	67	80.41003
81	90.87915	75	81.57452
81	90.49584	71	83.14709
63	91.29528	63	74.04281
56	86.71601	56	77.10677
73	80.60192	73	81.56761
97	77.223366	97	77.67498
89	76.094574	89	76.766685
84	75.46063	84	74.70679

The above two tables give a comparison between actual values and predicted values by the models for 25 days.

VIII. Conclusions

This framework proposed an air quality expectation framework for 25 days using neural network models such as LSTM and CNN. As we experimented with the models for predictions; we observed that LSTM model has given better results than the CNN models. For better understanding of the outcome we have provide graphs and comparison table between actual values and predicted values.

References

[1] Ibrahim K. Ok, Mehmet Ulvi Simsek and Suat Ozdemir, Deep Learning for Air Quality Prediction in Smart Cities, IEEE International Conference on Big Data (BIGDATA), (2017).

[2] Ziyue Guan and Richard O. Sinnott, Prediction of Air Pollution through Machine Learning Approaches on the Cloud, IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT), (2018).

[3] Timothy M. Amado and Jennifer C. Dela Cruz, Development of Machine Learning-based Predictive Models for Air Quality Monitoring and Characterization, Proceedings of TENCON 2018-2018 IEEE Region 10 Conference (Jeju, Korea, 28-31 October 2018).

[4] Chavi Srivastava, Shyamli Singh and Jen Amit Prakash Singh, Estimation of Air Pollution in Delhi Using Machine Learning Techniques, International Conference on Computing, Power and Communication Technologies (GUCON), 2018.

[5] Pratyush Singh, T. Lakshmi Narasimhan and Chandra Shekar Lakshminarayanan, Deep Air: Air Quality Prediction using Deep Neural Network IEEE, (2019).

[6] Gaganjot Kaur Kang, Jerry Jeyu Gao and Sen Chiao, Air Quality Prediction: Big Data and Machine Learing Approaches, International Journal of Environmental Science and Development, Vol 9 January (2018).

[7] Jyun-Yu Jiang, Xue Sun, Wei Wang and Sean Young, Enhancing Air Quality Prediction with Social Media and Natural Language Processing, Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Italy, July 28-August 2, (2019).

[8] Brian S. Freeman, Graham Taylor and Bahram Gharabhagi, Forecasting Air Quality Time Series Using Deep Learning, Journal of the Air and Waste Management Association May (24) 2018.

[9] Zheng, Yu, Furui Liu and Hsun-Ping Hsieh, U-Air: When urban air quality inference meets big data, 361 Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM 362 (2013).

[10] Xu and Yi, et al., ADMM without a Fixed Penalty Parameter: Faster Convergence with New Adaptive 414 Penalization, Advances in Neural Information Processing Systems, (2017).