# Air Quality Prediction (IoT) Using Machine Learning

**P. Sardar Maran, Bussu Saikiran Reddy, and Chava Saiharshavardhan**

**Abstract** The idea of the Internet of things (IoT) has become a popular research subject in many fields in recent years, including business, trade and education. To build sustainable urban life, smart cities employ IoT-based infrastructure and applications. IoT enables smart cities to make community residents more aware, responsive and effective by using information and communication technologies. As the number of IoT-based smart city applications increased, the amount of data produced by this application increased tremendously. Effective steps are taken by governments and community stakeholders to process these data and forecast potential consequences to ensure sustainable growth the field of forecasts; recurrent neural network methods were used in big data for many predictive problems. This inspires us to make use of recurrent neural network to predict IoT results. A novel model of recurrent neural network is therefore proposed in this paper to analyze IoT smart city data. Through this research, we have gathered $SO_2$ and $NO_2$ gasses at various locations through Chennai.

**Keywords** Internet of things (IoT) · Air quality · Recurrent neural network · Big data

## 1 Introduction

Internet of things (IoT) is known as daily object interconnection. By linking an increasing number of objects, it enhances the quality of human life. According to projections, IoT will have 50 billion connected devices by 2020. And we are deploying IoT sensors such as $CO_2$ and $NO_2$ in this project to collect the sensor values that can be collected and stored in the server for data analysis. Big data is used for data processing, the Hadoop distributed file system and recurrent neural networks are used for prediction. Apache Hadoop is an open-source software platform for

P. Sardar Maran (✉) · B. S. Reddy · C. Saiharshavardhan
Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, India
e-mail: psmaran@sathyabama.ac.in



P. K. Mallick et al. (eds.), *Advances in Electronics, Communication and Computing*, Lecture Notes in Electrical Engineering 709,
https://doi.org/10.1007/978-981-15-8752-8_59

distributed big data storage and distributed data processing. The Hadoop Distributed file system (HDFS) partitions files into file size-based blocks and distributes the blocks among the cluster nodes.

Recurrent neural networks are fairly primitive electronic neuron networks, based on the brain's neuronal structure. The method records one at a time and learns by comparing their (mostly arbitrary) record prediction to the known actual record. The errors from the first record's initial forecast are fed back to the network and used to change the network's algorithm for the second iteration.

## 2 Literature Survey

Examined that the neural network models are an efficient method for controlling air quality in big cities due to a low level of error estimation by exploring the possibilities of creating an air quality control prognostic method in big cities [1, 2]. Just 1.9% in test series and 1.4% in training series were in the optimally built false prognosis models only 1.9% in testing series and 1.4% in training series. Proposed a model for the estimation of concentrations of inorganic airborne pollutants: $H_2S$–$SO_2$, NO–$NO_2$–$NO_x$, CO ($CO_2$) and PM10 (aerodynamic particulate matter of 10ìm or less) from a risk area (two industrial area-IA) and urban area (UA) from Constanta [3]. A small cumulative absolute error of 0.42 was predicted from simulation result with the actual value in the urban area. It shows the efficacy and validity of the proposed approach in determining the different concentrations of airborne contaminants.

Studied the model evaluation shows that the degree of success in CO concentration forecasting is promising by evaluating a model for predicting CO concentration from 2002 to 2004 [4, 5]. Analyzed Internet traffic data over IP networks through the creation of an artificial neural network (ANN) model based on multi-layer perceptron (MLP) [6]. Levenberg Marquardt Algorithm (LM) and Resilient Backpropagation Algorithm (RP) models were developed for effecive analysis of internet traffic over the network in the time series manner.

Found that for all contaminants using the CNN model, the association between expected and observed data was greater (0.54–0.87) [7]. The CNN model estimated concentrations of $SO_2$ that were higher than those of PM10. Another interesting finding was that all contaminants were expected better in winter concentrations than in summer concentrations [8]. Experiments showed that accurate predictions of missing concentrations of air pollutants can be made using the new approach found in the CNN model by predicting daily concentrations of air pollutants (PM10 and $SO_2$) from the Istanbul air pollution measurement stations in Yenibosna and Umraniye for periods when pollution data was not reported [9, 10]. Historical Ozone component ($O_3$) observations were used to predict air pollution concentrations of $NO_2$ and $SO_2$ in urban areas by using Adaptive Neuro Fuzzy techniques [11]. Proposed algorithm focused primarily on pollution caused by vehicle smoke in urban areas and implemented various fuzzy rules to demonstrate the extent of air pollution induced by vehicle smoke and dust in nearby Jabalpur areas examined the ability of artificial neural network technique in air temperature prediction for daily and monthly ambient [12]. In feed-forward network and Elman network, the mean, minimum, and maximum ambient air temperature was used as the input parameter during the years 1961–2004 [13, 14]. The values of R, MSE and MAE variables in both networks showed that the ANN method is a good model in the prediction of ambient air temperature, whereas the estimates of the mean temperature one day ahead and the maximum temperature one month ahead are more reliable using the Elman network [15].

Computer based models play a vital importance in research and analysis and for decision support systems [16, 17]. For both training and testing stages, the efficiency of the proposed solution is calculated by the mean squared error. Comparison between the functionality of the hybrid ICA-NN and the listed MLP network provides the evidence that the ICA-NN was superior with an reasonable accuracy for figure in terms of reliable performance [18, 19]. ANN approach was applied in problem solving techniques and decision making to improve the environmental factors [20]. Peninsular Malaysia centered on the use of main component analysis (PCA) and artificial neural network (ANN) to predict air pollutant index (API) in the seven selected Malaysian air monitoring stations in the seven-year database in the southern region (2005–2011) [21, 22].

## 3 Motivation

It is concluded that the ANN approach can be used effectively as decision-making and problem-solving strategies for better environmental management [23, 24]. Peninsular Malaysia based on the use of main component analysis (PCA) and artificial neural network (ANN) to predict air pollutant index (API) in the seven-year database selected Malaysian air monitoring stations in the southern region (2005–2011).

## 4 Proposed System

Effective recurrent neural network methods are deployed and an IoT-based hardware is connected to air quality sensors which allow us to collect real-time air quality prediction data. The key idea here is to forecast the air quality values in relation to the location values of the previous year and to build awareness about air pollution. In the proposed system, the air pollution data from the Raspberry pi sensor was collected to the system. In several locations, the data was collected, and finally, the data stored in the MYSQL server. The detailed flow diagram for the proposed systems is shown in Fig. 1.

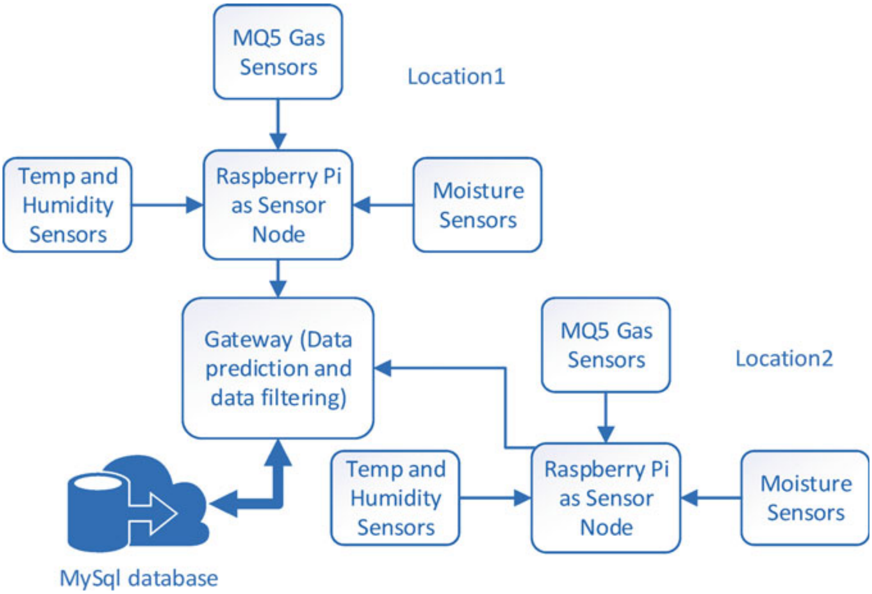


**Fig. 1** Flow diagram of air quality monitoring using IoT

### 4.1 PIC Board

PIC microcontrollers are electronic circuits that can be configured to perform a wide range of tasks. They can be programmed for being timers or managing a production line, and more. This is used for linking the device to the circuit microcontroller as shown in Fig. 2.

### 4.2 Nitrogen Dioxide Sensor

In urban settings, the main source of nitrogen dioxide gas is the burning of fossil fuels. This is most generally related to emissions from motor vehicles in urban areas. Areas with high-density road networks near to large populations like cities and towns are at highest risk of over-exposure. The manufacturing sites would also contain high nitrogen dioxide concentrations. Those include any industry that uses processes such as power plants, electrical utilities and industrial boilers for combustion. Long-term exposure to nitrogen dioxide gas can lead to respiratory problems such as bronchitis and asthma. Aeroqual provides a fixed or portable $NO_2$ sensor for calculating the nitrogen dioxide concentrations for studies of human exposure as well as health and safety at industrial and urban sites. Here, this $NO_2$ sensor is used to measure the

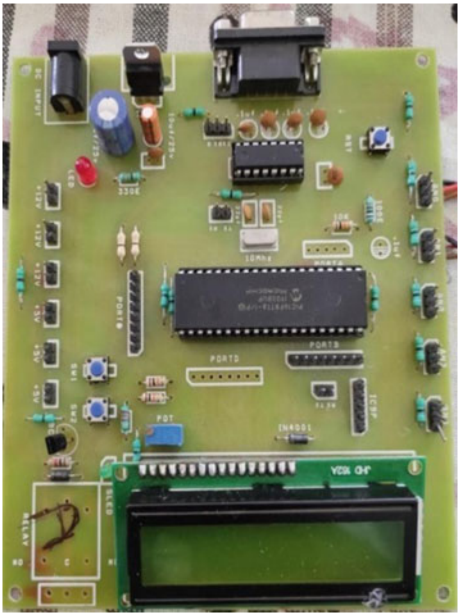

**Fig. 2** PC bord

current position concentration of $NO_2$ gases and then it collects the parameters of $NO_2$ gas in microgram/cubic meter.

### *4.3 Sulfur Dioxide Sensor*

Sulfur dioxide is a colorless gas with a strong odor. It is not combustible but an highly poisonous substance. Sulfur dioxide when mixed with water is highly corrosive sulfuric acid and causes chemical burns. There are two key reasons why control of sulfur dioxide levels is important: because industrial buildings such as power plants emit significant quantities of sulfur dioxide and sulfur dioxide is a highly reactive gas. Sulfur dioxide is heavier than air, so it is measured by PemTech using electromechanical technology. OSHA has set the PEL for sulfur dioxide to be just 5 ppm, as it can be highly harmful to be exposed to only small levels for a short time. Here, this $SO_2$ sensor is used to measure the current position concentration of $SO_2$ gasses and then it collects the $SO_2$ gas data in microgram/cubic meter.

## 5 Implementation

The user interface architecture is implemented using Java coding for the user to pick the data input and location for data analysis. Using this module, all data is analyzed using the big data and recurrent neural network to predict the location's air pollution and then the IOT sensor is interfaced with the Java coding application.

### *5.1 Recurrent Neural Network*

The recurrent neural network is based on the concept of saving a layer's output and feeding this back to input to help predict the year's outcome. Here, the first layer is generated with the product of the sum of the weights and the features similar to the feed-forward neural network. If this is determined, the recurrent neural network cycle begins; it means that each neuron would recall any knowledge that it had in the previous time phase from one-time phase to the next. This makes each neuron function in the output of computations as a memory cell. We need to let the neural network operate on front propagation in this step and remember what information it needs for later use. IoT sensors such as $SO_2$ and $NO_2$ are interfaced with the microcontroller for real-time data collection and storage on a centralized server and then all data is collected from the sensors and the generated data set is stored for analysis in a centralized database using a recurrent neural network to extract the useful data generated for data analysis. All data is stored in a centralized database called MYSQL and then all data is collected from the sensors and the data set for data analysis is generated. We deploy big data-HDFS for the process of data processing and prediction. For data analysis, all sensor values are collected and stored in the server and then all data is processed and analyzed using a neural network to capture the data and predict air pollution at a given location and time. The values are compared with the approved data collection, and the values of the sensors are also processed. Alerted to monitor emissions and safeguard the atmosphere until the data is analyzed.

### *5.2 Comparative Analysis*

Many of them control the air quality in current systems and these systems can not predict the air quality of the gases. And it depends on the quality of the air which affects how you live and breathe. Therefore, it is important to take precautionary measures to forecast air pollution parameters such as sulfur dioxide and nitrogen dioxide. Such parameters can have a direct impact on degree of air pollution and are measured using sensing instruments. In proposed research, IOT-based hardware and sensor interface to detect the existing runtime air pollution of various contaminants in the air and thus deploy the principles of big data and recurrent neural networks help us predict the air quality of gases with values of previous years concerning location.

## 6 Result and Discussion

In this section of results, briefly review the table below to find out the concentration of gasses such as $SO_2$ and $NO_2$ in microgram/cubic meter parameters in some areas of Chennai city. The air pollution concentrations such as $SO_2$ and $NO_2$ were collection for the month of Feb 2018–2019. The real time sensor data of Feb 2020 were compared with the collected concentrations. Table 1 highlighted the standard index about air quality.

The air quality concentrations of $SO_2$ and $NO_2$ from various parts of the city (Anna Nagar, Adyar, Nungambakkam, etc.) were calibrated with the IoT sensor and the data stored in the MYSQL database. Figure 3 shows the sensor calibrated value of $NO_2$ and $SO_2$.

The recurrent neural network is a time series prediction method. In this study, the concentration values of $NO_2$ and $SO_2$ were analyzed as a time series manner and the data was interpreted in MATLAB for training and testing purpose. Finally, Figure 4 shows the comparative analysis of sensor data with the RNN.

**Table 1** Air quality index

| Air quality category | Air quality index |
|---|---|
| Good | 0–50 |
| Satisfactory | 51–100 |
| Moderate | 101–200 |
| Poor | 201–300 |
| Very poor | 301–400 |
| Severe | 401–500 |

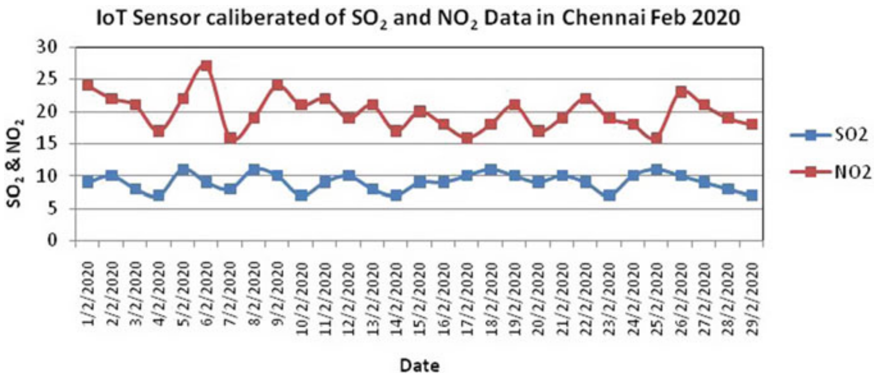


**Fig. 3** $NO_2$ and $SO_2$. concentrations from IoT sensor

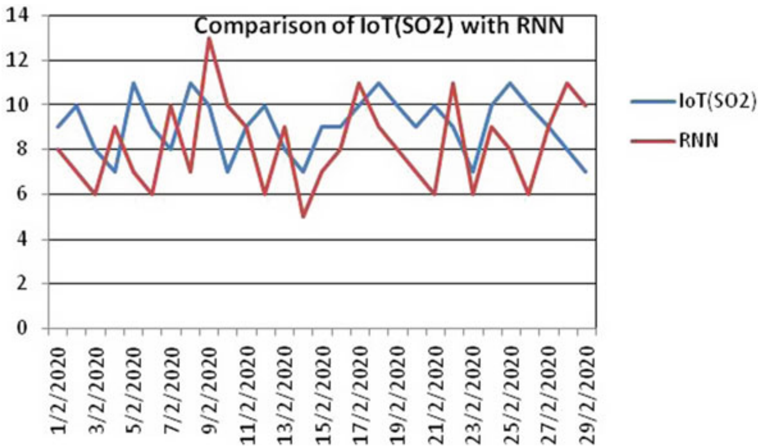


**Fig. 4** Comparative analysis of IoT-$SO_2$ with RNN

## 7 Conclusion

In this study, along with recurrent neural network, we deploy the principle of IOT and big data for successful analysis and prediction of air pollution. Thus, we are forecasting the air quality of various air pollutants for years to come in this phase.

## References

1. Perez, P. (2001). Prediction of sulfur dioxide concentrations at a site near downtown Santiago, Chile. *Atmospheric Environment, 35,* 4929–4935.
2. Anttila, P., Tuovinen, J.-P., & Niemi, J. V. (2011). Primary $NO_2$ emissions and their role in the development of $NO_2$ concentrations in a traffic environment. *Atmospheric Environment, 45*(4), 986–992.
3. Billionnet, C., Sherrill, D., & Annesi-Maisano, I. (2012). GERIE study. Estimating the health effects of exposure to multi-pollutant mixture. *Annals of Epidemiology, 22*(2), 126–141.
4. Kelly, S., Suryadevara, N., & Mukhopadhyay, S. (2013). Towards the implementation of IoT for environmental condition monitoring in homes. *IEEE Sensors Journal, 13*(10), 38463853.
5. Jamal, A., & NabizadehNodehi, R. (2017). Predicting air quality index based on meteorological data: A comparison of regression analysis, artificial neural networks and decision tree. *Journal of Air Pollution and Health, 2,* 27–38.
6. Wang, W., Mao, F., Du, L., Pan, Z., Gong, W., & Fang, S. (2017). Deriving hourly PM2.5 concentrations from Himawari-8 AODs over Beijing–Tianjin–Hebei in China. *Remote Sensing, 9*, 858.
7. Shahbazi, H., Karimi, S., Hosseini, V., Yazgi, D., & Torbatian, S. (2018). A novel regression imputation framework for Tehran air pollution monitoring network using outputs from WRF and CAMx models. *Atmospheric Environment, 187,* 24–33.
8. Nagarajan, G., & Minu, R. I. (2018). Wireless soil monitoring sensor for sprinkler irrigation automation system. *Wireless Personal Communications, 98*(2), 1835–1851.

9. Kezia, S. P., & Mary, A. V. A. (2016). Prediction of rapid floods from big data using map reduce technique. *Global Journal of Pure and Applied Mathematics, 12*(1), 369–373.
10. Nagasankar, T., & Ankaryarkanni, B. (2016). Performance analysis of edge detection algorithms on various image types. *Indian Journal of Science and Technology, 9*(21), 1–7.
11. UN-DESA. (2018). *World urbanization prospects: The 2018 revision.* Department of Economic and Social Affairs.
12. Liu, W., Guo, G., Chen, F., & Chen, Y. (2019). Meteorological pattern analysis assisted daily PM2.5 grades prediction using SVM optimized by the PSO algorithm. *Atmospheric Pollution Research.*
13. Anu, V., Mala, G. A., & Mathi, K. (2014). An overview of RFID data processing techniques. *International Journal of Applied Engineering Research, 9*(21), 8603–8612.
14. Vaniya, S. M., & Bharathi, B. (2016, December). Exploring object segmentation methods in visual surveillance for human activity recognition. In *2016 International Conference on Global Trends in Signal Processing, Information Computing and Communication (ICGTSPICC)* (pp. 520–525). IEEE.
15. Jacob, T. P. (2015). Implementation of randomized test pattern generation strategy. *Journal of Theoretical and Applied Information Technology, 73*(1).
16. Delavar, M., Gholami, A., Shiran, G., Rashidi, Y., Nakhaeizadeh, G., Fedra, K., Hatefi Afshar, S. A. (2019). Novel method for improving air pollution prediction based on machine learning approaches: A case study applied to the Capital City of Tehran. *ISPRS International Journal of Geo-Information, 8*, 99.
17. Xie, H., Ma, F., & Bai. Q. (2009). Prediction of indoor air quality using artificial neural networks. In *Fifth International Conference on Natural Computation (ICNC 2009)*, pp. 414–418.
18. Franklin, R. G., & Muthukumar, B. (2020). A prediction system for diagnosis of Diabetes Mellitus. *Journal of Computational and Theoretical Nanoscience, 17*(1), 6–9.
19. Shyry, S. P. (2014, October). Novel enhanced encryption algorithm for shared key generation. In *Proceedings of the 2014 International Conference on Interdisciplinary Advances in Applied Computing* (pp. 1–7).
20. Chen, X., Liu, X., & Xu, P. (2015). IOT-based on air pollution monitoring and forecasting a system. In *International Conference on Computer and Computational Sciences*, pp. 257–260.
21. Vignesh, R., & Mohana Prasad, K. (2019, September), Cloud-implementation of E-healthcare framework. *International Journal of Recent Technology and Engineering (IJRTE), 8*(3). ISSN: 2277-3878.
22. Maran, P. S. (2019). Meteorological tower wind shear characteristics, vertical wind speed profile and surface roughness analysis near the coastline of Chennai, Songklanakarin. *Journal of Science & Technology 41*(4), 755–760.
23. Saikumar, Ch. V., Reji, M., & Kishoreraja, P. C. (2017). IoT based air pollution monitoring and predictor system on Beagle Bone Black. In *International Conference on Nextgen Electronic Technologies*, pp. 367–370.
24. Hoi, K. I., Yuen, K. V., & Mok, K. M. (2009). Prediction of daily averaged PM10 concentrations by statistical time-varying model. *Atmospheric Environment, 43*(16), 2579–2581.