International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 8 Aug 2023, pp: 572-577 www.ijaem.net ISSN: 2395-5252

Real Time Air Quality Monitoring and Prediction Using Sensing Networks

R.Jayanth Rosario, Mr.S.R.Naresh
K.L.N College of Engineering,Pottapalayam 630612
Associate Professor, K.L.N College of Engineering, Pottapalayam -630612.

Date of Submission: 15-08-2023 Date of Acceptance: 25-08-2023

ABSTRACT: Nowadays air pollution and its harm to human health has become a serious problem in many cities around the world. In recent years, research interests in measuring and predicting the quality of air around people has spiked. Since the Internet of Things (IoT) has been widely used in different domains to improve the quality life for people by connecting multiple sensors in different places, it also makes the air pollution monitoring more easier than before. Traditional way of using fixed sensors cannot effectively provide a comprehensive view of air pollution in people's immediate surroundings, since the closest sensors can be possibly miles away. On modeling the air quality pattern in a given region by adopting both fixed and moving IoT sensors, which are placed on vehicles patrolling around the region. A full spectrum of how air quality varies in nearby regions can be analyzed. The feasibility of our approach in effectively measuring and predicting air quality using different machine learning algorithms with real world data. Our evaluation shows a promising result for effective air quality monitoring and prediction for a smart city application .

EX: Time-series prediction, air quality measurement, machine learning

I. INTRODUCTION

Due to rapid urbanization and industrialization, many countries around the world are facing a critical crisis of air pollution. Air pollution has become a threat to public health and a heavy influential factor on citizen's daily activity. In metropolitan cities in developing countries bothered by problems of air pollution, such as Chennai and Delhi, people usually need to wear a mask before going out. Besides, outdoor activities are also constrained by the intra-day air quality. Air pollution is caused by the presence of different air pollutants. The primary air pollutant gases are nitrogen dioxide (NO2), carbon monoxide (CO), ozone (O3) and sulphur dioxide (SO2) [2]. Another type of air pollutants is air particulate matter (PM). Among them, PM2.5 and PM10 are of particular concerns to people, which refers to atmospheric particulate matter that have a diameter of less than 2.5 µm and 10 µm. These particles can cause many respiratory or cardiovascular diseases [3]. Thus, many cities have built their own air quality monitoring stations and publish the real-time air quality information every hour. As the concern for air pollution increases, its becoming increasingly critical to measure the air quality around people [4], [5], which inform people about when is safe to perform outside activities and help them plan better routes to reach their destinations. Typically, monitoring stations at fixed locations is the conventional approach for atmospheric factor monitoring for a large geographical district .

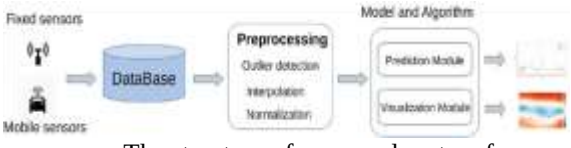
While it is not difficult to implement such fixed sensor based monitoring system, it faces several challenges. First, huge investment is involved in building and deploying monitoring units to cover a large area. Also, it is highly dependent on neighboring environments and tends to be less accurate for farther areas. In areas close to the roads, even small distances can make a huge difference in air quality data measurement from car pollutions. Hence, new ways to collect air quality information in a cheaper and more flexible way and provide detailed air quality prediction is in demand. To address these issues, one possible solution is to make the sensors mobile using Internet-of-Things(IoT). For example, attaching sensors on moving cars or drones proved to be a feasible method [6]. In this work, we developed the IoT devices to monitor air quality. We collected air pollution data by mounting a sensor to a car and moved around the city of Madurai. This data is then preprocessed. One major advantage of using a mobile sensor is that it provides the very first hand air pollution information for an area at a particular time, when the car was moving through there. we can also cover more geographical regions and have more accurate localized information with mobile IoT sensors. While a static fixed sensor can provide

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 8 Aug 2023, pp: 572-577 www.ijaem.net ISSN: 2395-5252

continuous feed of information about a particular area, it is not easy with a mobile sensor. However, this can be minimized by having multiple mobile sensors or assigning smaller coverage area to a mobile sensor.In this work, I propose a hybrid approach, where we deploy multiple static sensors as well as IoT mobile sensors to effectively monitor air quality. The static sensors can provide a holistic view by providing a continuous feed of information. On the other hand, mobile sensors can provide more accurate data about specific areas to reduce the error from static sensors. In this paper, I build a prediction model to utilize the collected data and provide rapid information about the air quality around people. I also analyze and forecast air quality and provide insights to both professional researchers and ordinary users. The main contributions of the work are summarized as follows: • I proposed a hybrid approach to integrate fixed and mobile IoT sensors to measure and predict air quality data. • I demonstrated the feasibility and effectiveness of the approach by analysing the prediction result with different machine models. •I developed a tool to show the relative distribution of the air pollutants with a focus on PM10 and PM2.5, where it provides an intuitive understanding of the air quality around people.

II. RELATED WORKS

To measure the air quality, several monitoring methods have been proposed and utilized. In Zheng et al.'s research [7], they use public and private web services as well as a list of public websites to provide real-time meteorological, weather forecasts and air quality data for their forecasting. Small unmanned aerial vehicles are used in the work of Alvarado et al. [8] as a methodology to monitor PM10 dust particles, where they can calculate the emission rate of a source. With the development of smart city technologies, IoT devices have been shown to be an effective option to collect real time weather, road traffic, pollution and traffic information. Thus, IoT devices are also considered to enable air quality analysis [9]. In addition to the fixed sensors, public transportation infrastructure such as buses has been used to collect air quality data [10]. Also, there is one project [11] engaged the entire community members in collecting data and developed an online air quality monitoring system based on it, which is also called crowdsourcing. Hasenfratz et al. [12] utilized sensor nodes to build a thousand models targeting at different time periods. All these aforementioned methods are either costly or time consuming. In the work, I explore the use of fixed and mobile IoT sensors together to improve the prediction performance, which has not been researched much yet. To meet the increasing query frequency of air quality in real time and also to enable citizens to react instantly to the pollution, there has been a large body of work on building connected monitoring sensor networks to share the current air quality information with them [13]. Garzon et.al presented in [14] an air quality alert service. Their service continuously determines the areas, where the level of certain matter concentration exceeds the preset threshold, and notify users if they entered them. Maag et al. [15] proposed a multi-pollutant monitoring platform using wearable low-cost sensors. Compared with above methods, system can serve the similar functions to end users practically with either fewer sensors or less demand for computation. For prediction, regression models are commonly used in the area of air quality prediction. A multivariate linear regression model for predicting PM2.5 of short-period time is proposed in Zhao's work [16], which includes other gaseous pollutants such as SO2, NO2, CO and O3. As deep learning emerged as an effective method in many applications, time series data of air pollution based on different network models have been also extensively studied and developed. Novel models such as Long Short-Term Memory (LSTM) and Gated Recurrent Unit Network (GRU) have been proved to be a powerful sequential structures in predicting future values of air quality [9], [17] . Yi et al. [18] proposed a deep distributed fusion network to learn the characteristics of spatial dispersion and capture all the influential factors that may have a direct or indirect effect on air quality. These aforementioned technologies fits non-linear models flexibly but usually being short of offering insight to the hidden mechanism. In addition, they have not shown to necessarily outperform classical regression models in many scenarios [19]. There are also a lot of researches concentrate on approaches to model and simulate the pollutants for prediction [20]. With a small amount of data set oriented in the project, I decided to take conventional regression models as the baseline methods because of computation efficiency, while yielding favorable results



The structure of proposed system for monitoring and prediction of air pollution

III. IMPLEMENTATION

The design and implementation of IoT sensor device deployed in the research. The deployment and data collection are performed in Chennai, which is envisioned to be developed as a

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 8 Aug 2023, pp: 572-577 www.ijaem.net ISSN: 2395-5252

smart city. Next, i explain the preliminary processing of the acquired raw data and describe how to store and transmit the collected and cleaned data. Then, further present the user interface to check the collected data for our analysis of overall architecture of our proposed system..

A. IoT SENSOR INSTRUMENT DESIGN

I assembled two types of sensing devices from off-theshelf parts, one for fixed locations and the other type for moving cars. In total, developed three IoT sensor devices, where two of them are deployed in two different fixed locations and the other one is mounted on data collection cas. The subsystems of the air quality monitoring modules are presented and the functions of the sensors are described as following:

Temperature and humidity sensor: We have a single sensor that can measure both temperature and humidity. The humidity sensor provides an accuracy of 2%, whereas the temperature sensor has an accuracy of 0.5˚C. They have measurement ranges of 0 ∼100% and −40 ∼80˚C, respectively.

Micro Dust sensor: This sensor measures both $PM_{2.5}$ and PM_{10}. The range of these measurements is from 0~999.9μg/m^3. The Government of India considers $PM_{2.5}$ andPM_{10} values of over 35μg/m^3 and 100μg/m^3 averaged through a day to be dangerous for human health. Thus, our micro dust sensor covers the entire range that is relevant for human health.

Carbon Dioxide sensor: Our carbon dioxide sensor can measure CO_2 within a range of 0 ∼ 10000ppm, with an accuracy of 5ppm(0 ∼ 2000ppm), 10ppm(2000 ∼ 5000ppm), and 20ppm(5000 ∼ 10000ppm). Note that since in a natural scenario, the proportion of CO_2 is around 0.03%, this level of accuracy is sufficient for our purpose.

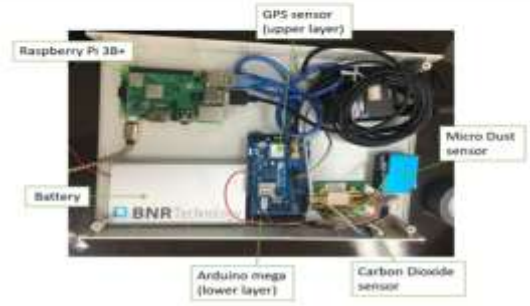
Raspberry Pi 3B+: The Raspberry Pi is connected to LTE using a dongle. Its main function is to process thesensor data and send it over the internet to the cloud server.

Arduino mega: This implements the protocol for sending data over the VoLTE network.

GPS sensor: This GPS sensor is connected to the Arduino, and provides an accuracy of close to 1 m.

Battery: Here use a power bank with a capacity of 7,000 mAh. The overall power consumption of our setup is close to 1A. Thus, setup can run continuously for around 7 hours without a single charge.

(a) Fixed Sensing IoT Device.



(b) Mobile Sensing IoT Device.

Communication Software: I constructed a wireless communication and GPS system to transmit acquired data back to databases for analysis. The geo-tagged data which is stored in Raspberry Pi is transmit over Voice over Long-Term Evolution(VoLTE) once per second.

Database: Design the database to store the collected real time sensor values from fixed as well as mobile IoT sensors. The data fields are: 1) time, 2) GPS location, 3) temperature, 4) humidity, 5) CO_2, 6) PM_{10}, and 7) $PM_{2.5}$, where all the collected values are stored in database .In the areas with weak GPS signals, such as indoors and tunnels, the approximate the value according to the latest neighboring data.

.**Cloud Server and Data Mapping**: Used a cloud server for the system, where the server manages the data and provides an interface for analyzers to check the details of the collected real air pollution value. It allows us to choose sensor number and other fields such as dates. There are two main functions of our web portal: (1) Time flow of the real data of different categories with marked min and max values, and (2) Google Map API is integrated in the website to visualize the traces of the moving cars during the chosen time period. The car follow different paths randomly but tried to cover the entire area as much as possible. All the stored data can be downloaded in the form of Excelspreadsheet for later analysis.

UserInterface(UI): In addition, I developed the User Interface (UI) App so that users can log in our developed APP using their own account and check the air quality data around them immediately. The

International Journal of Advances in Engineering and Management (IJAEM)
Volume 5, Issue 8 Aug 2023, pp: 572-577 www.ijaem.net ISSN: 2395-5252

example of user interface is provided, where APP can measure the real time air quality measurements and display.

Outlier detection: Since sudden changes in the collected data usually means an outlier, thee calculated the discrete differences of measured sensor values along the timeline to detect the outliers. That is, measured samples with a discrete difference beyond the interval [−0.5,2] are removed from the data set.

Interpolation: I choose Gaussian Process Regression (GPR) [22] as our interpolation method because it assists
in reaching the best prediction accuracy in our experiments, and the effect of different interpolation methods.

Data normalization: Since data are measured at different scale, we normalize the sensor measurement between 0 and 1.

SUPPORT VECTOR REGRESSOR (SVR)

The objective of SVR is to determine a hyper-plane in the space generated by mapping training data in its original space to a higher dimensional feature space, and the hyper-plane can minimize the deviation of all sample points from it. Consider the training data set $\{(\mathbf{x}_1,y_1),(\mathbf{x}_2,y_2),...,(\mathbf{x}_m,y_m)\}$, where $\mathbf{x} \in R^n, y \in R$ where m corresponds to the number of training data, then the regression problem can be formulated as:

$$\min_{w,b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{m} l_\epsilon(f(\mathbf{x}_i) - y_i). \quad (2)$$

Here C is a constant, f (**x**) is the hyper-plane represented as f (**x**) = **w**·**x**+b, and l is the cost function which is minimized innumber of input variables yet yielding high accuracy. RFR randomly draws samples from the original dataset with replacement, which is also called bootstrap, and grows an unpruned regression tree for each of the samples, then average the unweighted outputs of multiple decision trees to obtain the final result as follows:

$$\bar{h}(\mathbf{x}) = \frac{1}{K} \sum_{k=1}^{K} h(\mathbf{x};\theta_k), \quad (4)$$

where $h(\mathbf{x};\theta_k)$ is a collection of tree predictors with k = 1,...K, θ_k is random vector, which characterizes the kth RF tree, **x** represents the observed input which are assumed to be independently drawn from the joint distribution (**x**,y). Similarly, x represents time, longitude, and latitude information collected from sensors, and y represents a collected value from air pollutants set CO_2, $PM_{2.5}$.

DATASET

Both the fixed and mobile sensors collect the same format of dataset. The fixed sensors collect air quality data every minute from the three chosen locations in Chennai area. For each fixed sensor, the data collection time periods span all day, basically from morning to night. The mobile sensors, however, collect the air pollution data only a few hours per day, but the whole dataset in general also covers all hours of a day.

The geographical locations of these sensors are presented, where each icon stands for a sensor. The horizontal and vertical lines of the grids are cut according to latitude and longitude, and spaced evenly to grant same size grids. Each collected data instance consists of the sensor box's longitude and latitude, timestamp, temperature, humidity, and concentration value of CO_2, $PM_{2.5}$ and PM_{10}.

The observed time series data of $PM_{2.5}$ and PM_{10} collected from the moving sensors for the entire region are depicted.The averaged the data collected from all the moving sensors at each moment. Along the X axis is the timeline and Y axis represents the pollutants' observed value, and the quantity unit for PM_{10} and $PM_{2.5}$ is $\mu g/m^3$.

IV. PROPOSED SYSTEM

OVERALL PERFORMANCE COMPARISONS WITH DIFFERENT PREDICTION ALGORITHMS

In this section, I used RFR, SVR and GBR to validate the overall performance of our proposed air quality prediction model. I split the entire dataset into non-overlapping training and test pairs, where each individual day is a test dataset and all the prior date forms the training dataset, respectively. The overall performance of different regression algorithms across various test . Values in bold indicates the best prediction in a specific testing. We can see that in general, GB regressor achieved the highest prediction accuracy , while RFR and SVR has marginally better performance .

ACCURACY PERFORMANCE WITH DIFFERENT NUMBER OF GRIDS

For evaluation, we select the test data and data from the training data. Since $PM_{2.5}$ and PM_{10} are our major interest, the focus on the prediction accuracy comparison on $PM_{2.5}$ and PM_{10}, and the chose GBR as the prediction algorithm, as it outperforms the other two methods in the previous evaluation section in general. The counted the number of samples in each of the grids and divide the

number of samples into intervals based on its distribution:

0 ~10, 11 ~20, 20 ~50, 50 ~100, 100 ~200 and above 200. Then, calculate the number of grids in each category and all these grids' RMSE of prediction. At last, averaged the RMSE of all the grids in that specific category.

PERFORMANCE WITH DIFFERENT INTERPOLATIONMETHODS

As discussed before, the collected data is very sparse on the geographical grids in a specific time point and the dispersion characteristics of the fine particles are complex to model. Therefore, different interpolation techniques are examined in the model to fill the missing air pollution data in all the other grids. In order to check whether the interpolation strengthens the prediction, compare three different interpolation methods with the baseline (no interpolation). Since conventional interpolation method Kriging [4], [31] shares the same mean value and confident interval with Gaussian Process Regression (GPR), choose linear interpolation and GPR with different kernels (Gaussian and Cauchy)

V. CONCLUSION

In this paper, I explored a new way to predict immediate air quality around people, by combining fixed and mobile sensors. The experimental results show that the proposed hybrid distributed fixed and IoT sensor system is effective in predicting air quality around the people. In addition, the proposed system can be practically realizable by leveraging public transportation system such as buses as well as taxis to be equipped with IoT sensor devices to measure different areas. The predicted air quality data from the system can be served in various scenarios, such as planing for outdoor activities.

REFERENCES

[1] M. Kampa and E. Castanas, ''Human health effects of air pollution,'' Environ. Pollut., vol. 151, no. 2, pp. 362–367, Jan. 2008.

[2] E. Boldo, S. Medina, A. Le Tertre, F. Hurley, H.-G. Mücke, F. Ballester, and I. Aguilera, ''Apheis: Health impact assessment of long-term exposure to PM2.5 in 23 European cities,'' Eur. J. Epidemiology, vol. 21, no. 6, pp. 449–458, Jun. 2006.

[3] J. Lin, A. Zhang, W. Chen, and M. Lin, ''Estimates of daily PM2.5 exposure in beijing using spatio-temporal kriging model,'' Sustainability, vol. 10, no. 8, p. 2772, 2018.

[4] Y. Jiang, L. Shang, K. Li, L. Tian, R. Piedrahita, X. Yun, O. Mansata, Q. Lv, R. P. Dick, and M. Hannigan, ''MAQS: A personalized mobile sensing system for indoor air quality monitoring,'' in Proc. 13th Int. Conf. Ubiquitous Comput. UbiComp, 2011, pp. 271–280.

[5] D. Zhang and S. S. Woo, ''Predicting air quality using moving sensors (poster),'' in Proc. 17th Annu. Int. Conf. Mobile Syst., Appl., Services, Jun. 2019, pp. 604–605.

[6] Y. Zheng, X. Yi, M. Li, R. Li, Z. Shan, E. Chang, and T. Li, ''Forecasting fine-grained air quality based on big data,'' in Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining KDD, 2015, pp. 2267–2276.

[7] M. Alvarado, F. Gonzalez, P. Erskine, D. Cliff, and D. Heuff, ''A methodology to monitor airborne PM10 dust particles using a small unmanned aerial vehicle,'' Sensors, vol. 17, no. 2, p. 343, 2017.

[8] I. Kok, M. U. Simsek, and S. Ozdemir, ''A deep learning model for air quality prediction in smart cities,'' in Proc. IEEE Int. Conf. Big Data (Big Data), Dec. 2017, pp. 1983–1990.

[9] S. Devarakonda, P. Sevusu, H. Liu, R. Liu, L. Iftode, and B. Nath, ''Realtime air quality monitoring through mobile sensing in metropolitan areas,'' in Proc. 2nd ACM SIGKDD Int. Workshop Urban Comput. UrbComp, 2013, p. 15.

[10] Y.-C. Hsu, P. Dille, J. Cross, B. Dias, R. Sargent, and I. Nourbakhsh, ''Community-empowered air quality monitoring system,'' in Proc. CHI Conf. Hum. Factors Comput. Syst., May 2017, pp. 1607–1619.

[11] D. Hasenfratz, O. Saukh, C. Walser, C. Hueglin, M. Fierz, T. Arn, J. Beutel, and L. Thiele, ''Deriving high-resolution urban air pollution maps using mobile sensor nodes,'' Pervas. Mobile Comput., vol. 16, pp. 268–285, Jan. 2015.

[12] A. C. Rai, P. Kumar, F. Pilla, A. N. Skouloudis, S. Di Sabatino, C. Ratti, A. Yasar, and D. Rickerby, ''End-user perspective of low-cost sensors for outdoor air pollution monitoring,'' Sci. Total Environ., vols. 607–608, pp. 691–705, Dec. 2017.

[13] S. R. Garzon, S. Walther, S. Pang, B. Deva, and A. Küpper, ''Urban air pollution alert service for smart cities,'' in Proc. 8th Int. Conf. Internet Things, Oct. 2018, p. 8.

[14] B. Maag, Z. Zhou, and L. Thiele, ''W-Air: Enabling personal air pollution monitoring on wearables,'' Proc. ACM Interact., Mobile,

Wearable Ubiquitous Technol., vol. 2, no. 1, p. 24, 2018.

[15] R. Zhao, X. Gu, B. Xue, J. Zhang, and W. Ren, ‘‘Short period PM2.5 prediction based on multivariate linear regression model,’’ PLoS ONE, vol. 13, no. 7, 2018, Art. no. e0201011.

[16] J. Ahn, D. Shin, K. Kim, and J. Yang, ‘‘Indoor air quality analysis using deep learning with sensor data,’’ Sensors, vol. 17, no. 11, p. 2476, 2017.

[17] X. Yi, J. Zhang, Z. Wang, T. Li, and Y. Zheng, ‘‘Deep distributed fusion network for air quality prediction,’’ in Proc. 24th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Jul. 2018, pp. 965–973.

[18] K. P. Moustris, P. T. Nastos, I. K. Larissi, and A. G. Paliatsos, ‘‘Application of multiple linear regression models and artificial neural networks on the surface ozone forecast in the greater athens area, greece,’’ Adv. Meteorol., vol. 2012, pp. 1–8, Jul. 2012.

[19] S. Fotouhi, M. H. Shirali-Shahreza, and A. Mohammadpour, ‘‘Concentration prediction of air pollutants in tehran,’’ in Proc. Int. Conf. Smart Cities Internet Things SCIOT, 2018, pp. 1–7.

[20] C. Kim, ‘‘Place promotion and symbolic characterization of new songdo city, South Korea,’’ Cities, vol. 27, no. 1, pp. 13–19, Feb. 2010.

[21] C. E. Rasmussen, ‘‘Gaussian processes in machine learning,’’ in Summer School on Machine Learning. Berlin, Germany: Springer, 2003, pp. 63–71.

[22] L. Breiman, ‘‘Random forests,’’ Mach. Learn., vol. 45, no. 1, pp. 5–32, 2001.

[23] C. Cortes and V. Vapnik, ‘‘Support vector machine,’’ Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.

[24] J. H. Friedman, ‘‘Stochastic gradient boosting,’’ Comput. Statist. Data Anal., vol. 38, no. 4, pp. 367–378, Feb. 2002.

[25] J. Wainer, ‘‘Comparison of 14 different families of classification algorithms on 115 binary datasets,’’ 2016, arXiv:1606.00930. [Online]. Available: http://arxiv.org/abs/1606.00930

[26] J. O. Ogutu, H.-P. Piepho, and T. Schulz-Streeck, ‘‘A comparison of random forests, boosting and support vector machines for genomic selection,’’ BMC Proc., vol. 5, no. S3, p. 11, c. 2011.

[27] A. Liaw and M. Wiener, ‘‘Classification and regression by randomforest,’’ R News, vol. 2, no. 3, pp. 18–22, 2002.

[28] H. Drucker, C. J. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, ‘‘Support vector regression machines,’’ in Proc. Adv. Neural Inf. Process. Syst., 1997, pp. 155–161.

[29] N. Duffy and D. Helmbold, ‘‘Boosting methods for regression,’’ Mach. Learn., vol. 47, nos 2–3, pp. 153–200, May 2002