



Design of IoT-based Air Quality Meter Module and Air Quality Analysis with Machine Learning

Ege Alp Türkyener^{1*}, Savaş Şahin², Sadık Arslan³

^{1*} İzmir Katip Celebi University, Faculty of Engineering and Architecture, Department of Electrical and Electronics Engineering, İzmir, Turkey, (ORCID: 0000-0002-8187-541X), egealpturkyener@gmail.com

² İzmir Katip Celebi University, Faculty of Engineering and Architecture, Department of Electrical and Electronics Engineering, İzmir, Turkey, (ORCID: 0000-0003-2065-6907), savas.sahin@ikc.edu.tr

³ Kentkart Ege Elektronik Sanayi ve Ticaret A.Ş. Department of Hardware R&D, İzmir, Turkey (ORCID: 0000-0003-4057-2030), sadik.arslan@kentkart.com.tr

(3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications June 11-13, 2021)

(DOI: 10.31590/ejosat.957500)

ATIF/REFERENCE: Türkyener, E.A., Şahin, S. & Arslan, S. (2021). Design of IoT-based Air Quality Meter Module and Air Quality Analysis with Machine Learning. *European Journal of Science and Technology*, (26), 364-368.

Abstract

This study proposes an ARM based air quality module placed to public transport vehicles for analyzing the effect of PM2.5 and PM10 particles in the cities in real-time using Internet of Things. The STM32 microcontroller is used for obtaining the data from the PM, humidity, and temperature sensors. The data collected from the sensors are sent to the i.MX6UL microprocessor using RS-485 connected to the internet portal with an Ethernet module. The microprocessor sends the data to the Microsoft Azure Hub in-on-line, and it is also recorded via the computer. The obtained data is analyzed for air quality-meteorological variables and the regression models are implemented via machine learning algorithms. PM2.5, PM10, humidity and temperature data are evaluated with R² test and root mean square error for regression models. The Random Forest algorithm shows better results among other used regression models.

Keywords: air quality analysis, real-time monitoring, machine learning, internet of things, arm-based microprocessor

IoT Tabanlı Hava Kalitesi Ölçer Modülünün Tasarımı ve Makine Öğrenmesi ile Hava Kalitesi Analizi

Öz

Bu çalışma, Nesnelerin İnterneti kullanılarak şehirlerdeki PM2.5 ve PM10 parçacıklarının etkisini gerçek zamanlı olarak analiz etmek için toplu taşıma araçlarına yerleştirilen ARM tabanlı bir hava kalitesi modülünü önermektedir. STM32 mikrodenetleyicisi, PM'den ve nem, sıcaklık sensörlerinden veri elde etmek için kullanılır. Sensörlerden toplanan veriler, ethernet modülü ile internet portalına bağlanan i.MX6UL'ya, RS-485 kullanılarak iletilir. i.MX6UL, verileri çevrimiçi olarak Microsoft Azure Hub'a gönderir ve bilgisayar aracılığıyla da kaydedilir. Elde edilen veriler hava kalitesi-meteorolojik değişkenler için analiz edilmiş ve regresyon modelleri makine öğrenme algoritmaları ile uygulanmıştır. PM2.5, PM10, nem ve sıcaklık verileri R² testi ve regresyon modelleri için ortalama karekök hatası ile değerlendirilmiştir. Rastgele orman algoritması, kullanılan diğer regresyon modelleri arasında en iyi sonucu göstermiştir.

Anahtar Kelimeler: hava kalitesi analizi, gerçek zamanlı izleme, makine öğrenimi, nesnelerin internet, arm tabanlı mikroişlemci.

* Corresponding Author: egealpturkyener@gmail.com

1. Introduction

Air pollution is one of the biggest problems in industrialized cities and causes millions of people to die from its complications therefore, it is important to be aware of the air quality. (Giannadaki, D., 2016). PM2.5 and PM10 are the one of the most harmful particles to humans. Because these particles are small, they can easily go to the lungs and cause future health problems (Brokamp, C., 2017). Therefore, it is important to monitoring and analyzing these particulate matters in real time and inform the decision makers about pollution to prevent the negative impacts that can cause. The air pollution measurement of cities can be made from stationary or mobile stations (Hu, K., Rahman, A., 2017). Air quality taken from fixed points can be measured and transferred using wireless sensor networks in real time. (Yaacoub, E., Kadri, 2013). Many cities have only limited, and stationary air pollution monitors due to the high cost of the multiple stations and this cause lack of understanding the impact of air quality on cities (Snyder, E. G., 2013). Mobile stations take measurements with sensor modules placed on a vehicle. A vehicle can be a car, public transport vehicles, public bikes (Liu, X., Li, B., Jiang, A., 2015). Sensor communication can be done using the internet (Saukh, O., 2012). Using the Internet of Things (IoT), data can be transmitted in real time. The relationship between air particles and humidity-temperature, the regression models are proposed (Kamińska, J. A. 2018). In this study a 32-bit ARM microcontroller-based outdoor air quality system is implemented in order to measure PM2.5, PM10, temperature and humidity. The data collected from the sensors on the public transport vehicles can be sent to the data center in real time using IoT. In addition, impact of temperature and humidity on air quality examined using regression models with machine learning algorithms. Root mean square error (RMSE) and R^2 score tests were used to compare the obtained results. The paper is organized as follows: in Section 2, hardware and software design of the system is presented. Section 3 describes the impact of temperature and humidity on PM2.5 and PM10 with the results from the system. Conclusions are presented in Section 4, and future research directions are outlined.

2. Material and Method

The main components of the system are given as follows: STM32L4 as microprocessor, i.MX6UL processor for transmitting the sensor values to Microsoft Azure cloud, RS-485 communication port for communicating the STM32L4 with i.MX6UL. The implemented system that measures the dust, temperature and humidity values is given in Fig1. And also, the analogy of the system is given in Fig2.

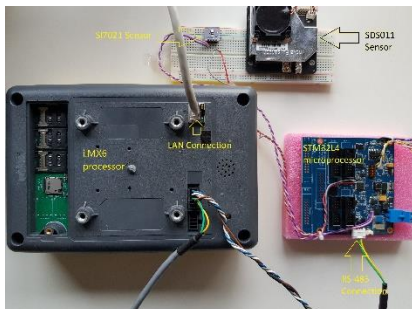


Fig1. The implemented system for real-time remote monitoring

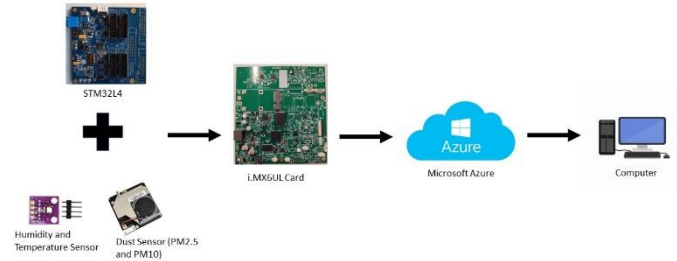


Fig2. The proposed system for real-time remote monitoring

2.1. Transmitting Data

The communication algorithm developed for the system is given in Fig3. STM32L4 microprocessor obtains the data from the sensors through I2C and UART connections and sends it to the i.MX6UL processor through RS-485 protocol. The i.MX6UL is connected to the internet with an ethernet connection. If the data sent through RS-485 is corrupted, it is deleted, and new data is expected to arrive. i.MX6UL sends the transmitted data to Azure Cloud using the MQTT protocol. The data obtained by the cloud, is read through computer and saved to create a regression model in the further steps.

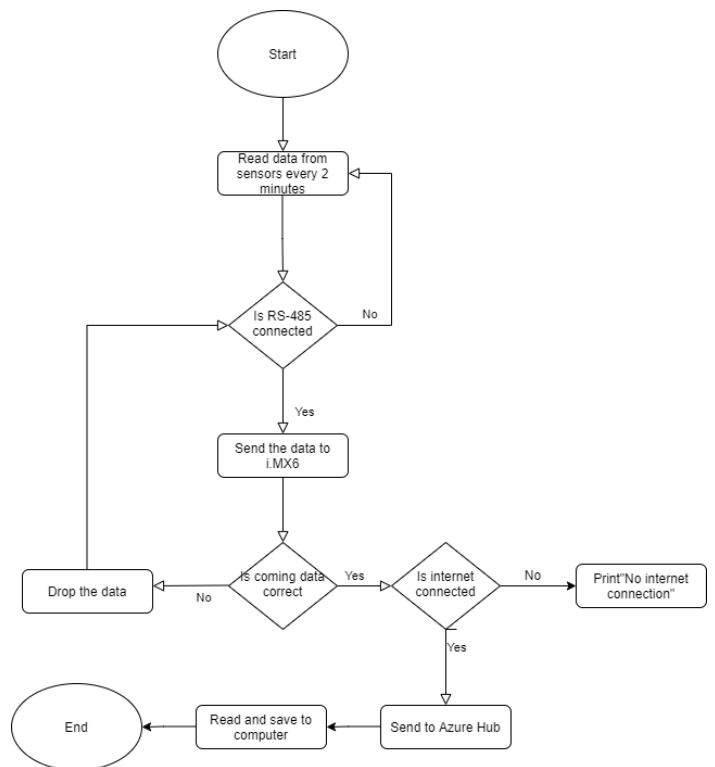


Fig3. Flowchart of the systems communication algorithm

2.2. Sensors

SI7021 sensor used for measuring the humidity and temperature and SDS011 sensor used to measure PM2.5 and PM10 pollutants. SDS011 is a low-cost dust sensor. It detects particle concentration between $0.3\mu\text{m}$ to $10\mu\text{m}$ using principle of laser

Table 1. European Emission Standards for PM2.5 and PM10 (EEA, 2018)

European Air Quality Index	Good	Fair	Moderate	Poor	Very poor	Extremely poor
Particles less than 2.5µm (PM _{2.5})	0-10 µg/m ³	10-20 µg/m ³	20-25 µg/m ³	25-50 µg/m ³	50-75 µg/m ³	75-800 µg/m ³
Particles less than 10µm (PM ₁₀)	0-20 µg/m ³	20-40 µg/m ³	40-50 µg/m ³	50-100 µg/m ³	100-150 µg/m ³	150-1200 µg/m ³

scattering (Budde,M., 2018). European emission standards on PM2.5 and PM10 particles can be seen in Table 1.

SI7021 offers a low-power, high-accuracy and factory-calibrated for both temperature and humidity solution (El Houssaini, D.) It is ideal for a wide range of temperature and humidity applications including high-reliability automotive and industrial systems. SI7021 can operate between -10 and 60 degrees Celsius with an accuracy of ±0.4°C. Temperature and humidity calculation of the sensor can be seen in Eq1 and Eq2.

$$Temperature(C) = \frac{175.72 * TempCode}{65536} - 4 \tag{Eq1}$$

$$\%RH = \frac{125 * RHCode}{65536} - 6 \tag{Eq2}$$

Where Temperature is in degrees Celsius, %RH is the Humidity, TempCode and RHCode are the 16-bit word returned by the SI7021 sensor.

2.3 Internet of Things

The i.MX6UL processor operates on Linux OS. Python programming language is used on i.MX6UL processor to connect to the IoT and transmit the data in real-time to cloud. The processor is used for connecting the whole system to the internet through ethernet connection. The transmitted data can be reached by local computer as in Fig4. Data sensor is stored in .csv format and used for machine learning algorithms.

```

Message received:
Sending message: b'{"temperature": 25.4,"humidity": 45.4,"pm2_5": 23,"pm10": 40}'
Total calls received: 1

Message received:
Sending message: b'{"temperature": 25.4,"humidity": 45.4,"pm2_5": 24,"pm10": 46}'
Total calls received: 2

Message received:
Sending message: b'{"temperature": 25.1,"humidity": 45.2,"pm2_5": 22,"pm10": 40}'
Total calls received: 3

Message received:
Sending message: b'{"temperature": 25.2,"humidity": 45.2,"pm2_5": 24,"pm10": 40}'
Total calls received: 4

Message received:
Sending message: b'{"temperature": 25.2,"humidity": 45.3,"pm2_5": 23,"pm10": 34}'
Total calls received: 5

Message received:
Sending message: b'{"temperature": 25.2,"humidity": 45.3,"pm2_5": 22,"pm10": 37}'
Total calls received: 6

Message received:
Sending message: b'{"temperature": 25.2,"humidity": 45.3,"pm2_5": 23,"pm10": 39}'
Total calls received: 7
    
```

Fig4. Real-Time Monitoring on Computer

2.4 Regression Model

Regression analysis is a process of statistical methods which is used to estimate relationship between dependent and independent variables. The purpose of machine learning algorithms is learning to recognize the pattern with several different methods. The supervised learning method-based machine learning algorithms are used for this study (Ahmad, T., 2018). For example, some of regression models are decision tree, linear regression, support vector machines, random forest, and k-nearest neighbors. The use of machine learning regression models it is aimed to determine the effect of different temperature and humidity conditions on PM2.5 and PM10.

3. Results and Discussion

The effect of different temperature and humidity conditions on PM2.5 and PM10 models are investigated by deriving regression models with machine learning algorithms. The data set is obtained from the different locations in the city. There are 4 different variables in the dataset which are temperature, humidity, PM2.5, PM10. The overall dataset was preprocessed and analyzed thoroughly. Different regression models were used on the dataset. These are linear regression, polynomial regression, K-NN regression, random forest regression and decision tree regression. To evaluate the performance RMSE and R² score are used. See Eq3 and Eq4.

$$RMSE = \sqrt{\frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{m}} \quad (Eq3)$$

where y_i is the output, \hat{y}_i is the value of the estimated output after the prediction of the regression model.

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y})^2} \quad (Eq4)$$

where \bar{y} is the average output.

When data is introduced into machine learning algorithms, the appropriate model for the data was Random Forest shown in Table 2 where it has the lowest RMSE value and the highest R^2 score. The relationship between PM2.5, PM10 and temperature, humidity depicted respectively in Fig5 and Fig6.

Table 2. R^2 score and RMSE values of regression models

Regression Models	RMSE	R^2
Linear Regression	4.88	0.2144
Polynomial Regression	3.50	0.5784
Decision Tree Regression	3.48	0.6665
Random Forest Regression	2.77	0.7529
KNN Regression	3.07	0.7173

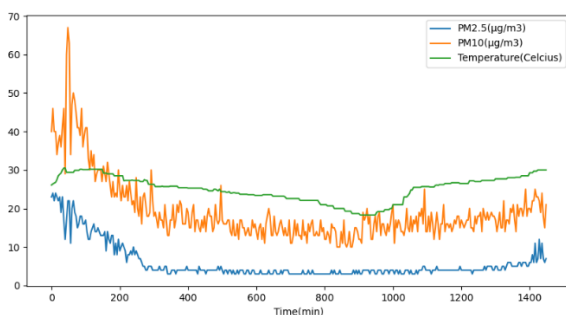


Fig5. PM2.5, PM10 and Temperature versus Time

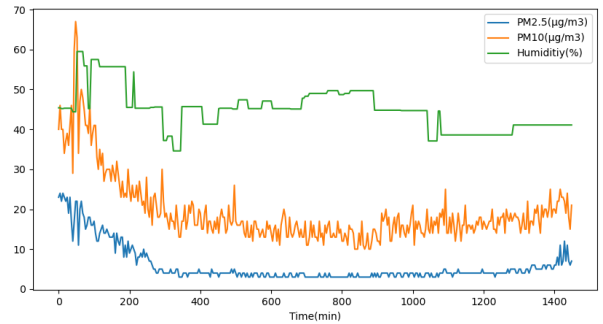


Fig6. PM2.5, PM10 and Humidity versus Time

4. Conclusions and Recommendations

In this study, an ARM based air quality module was designed to be placed to public transport vehicles for obtaining the air quality in the cities in real-time using Internet of Things. Data obtained from the sensors, the effect of different temperature and humidity conditions on the PM2.5 and PM10 regression model is made by using machine learning algorithms. PM2.5 and PM10 were used as pollutant variables. R^2 score and RMSE were performed with different regression models. The test is conducted in Python programming language. Random forest regression was decided as the best regression among the other regression models that were used. With this regression model obtained, it can be seen more clearly how global warming, which causes the air to warm, affects air pollution. As a future work of this study more sensors could be added to the module for better understanding the effect of global warming on the air pollution.

5. Acknowledge

This work is supported by the Scientific and Technical Research Council of Turkey (TUBITAK) under 2209B-Bachelor Final Thesis Focused on Industry Program with Project number 1139B412000704.

References

Ahmad, T., Chen, H., Huang, R., Yabin, G., Wang, J., Shair, J., Azeem Akram, H. M., Hassnain Mohsan, S. A., & Kazim, M. (2018). Supervised based machine learning models for short, medium and long-term energy prediction in distinct building environment. *Energy*, 158, 17–32. <https://doi.org/10.1016/j.energy.2018.05.169>

Air quality in Europe, 2018 Report, European Environment Agency <https://www.eea.europa.eu/publications/air-quality-in-europe2018/download>

Brokamp, C., Jandarov, R., Rao, M. B., LeMasters, G., & Ryan, P. (2017). Exposure assessment models for elemental components of particulate matter in an urban environment: A comparison of regression and random forest approaches. *Atmospheric Environment*, 151, 1-11.

Budde, M., Schwarz, A. D., Müller, T., Laquai, B., Streibl, N., Schindler, G., ... & Beigl, M. (2018). Potential and limitations of the low-cost SDS011 particle sensor for monitoring urban air quality. *ProScience*, 5, 6-12

El Houssaini, D., Khriji, S., Besbes, K., & Kanoun, O. Real Time Temperature Measurement for Industrial Environment.

- Giannadaki, D., Lelieveld, J., & Pozzer, A. (2016). Implementing the US air quality standard for PM 2.5 worldwide can prevent millions of premature deaths per year. *Environmental Health*, 15(1), 1-11.
- Hu, K., Rahman, A., Bhrugubanda, H., & Sivaraman, V. (2017). HazeEst: Machine learning based metropolitan air pollution estimation from fixed and mobile sensors. *IEEE Sensors Journal*, 17(11), 3517-3525.
- Kamińska, J. A. (2018). The use of random forests in modelling short-term air pollution effects based on traffic and meteorological conditions: a case study in Wrocław. *Journal of environmental management*, 217, 164-174.
- Liu, X., Li, B., Jiang, A., Qi, S., Xiang, C., & Xu, N. (2015, June). A bicycle-borne sensor for monitoring air pollution near roadways. In *2015 IEEE International Conference on Consumer Electronics-Taiwan* (pp. 166-167). IEEE.
- Saukh, O., Hasenfraz, D., Noori, A., Ulrich, T., & Thiele, L. (2012, February). Demo Abstract: Route Selection of Mobile Sensors for Air Quality Monitoring. In *9th European Conference on Wireless Sensor Networks (EWSN 2012)* (pp. 10-11).
- Snyder, E. G., Watkins, T. H., Solomon, P. A., Thoma, E. D., Williams, R. W., Hagler, G. S., ... & Preuss, P. W. (2013). The changing paradigm of air pollution monitoring. *Environmental science & technology*, 47(20), 11369-11377.
- Yaacoub, E., Kadri, A., Mushtaha, M., & Abu-Dayya, A. (2013, July). Air quality monitoring and analysis in Qatar using a wireless sensor network deployment. In *2013 9th International Wireless Communications and Mobile Computing Conference (IWCMC)* (pp. 596-601). IEEE.