Air Pollution Monitoring Using WSN Nodes with Machine Learning Techniques: A Case Study

PAUL D. ROSERO-MONTALVO*, *Department of Computer Science and Automatics, Universidad de Salamanca, 37008 Salamanca, Spain, and Department of Applied Sciences, Universidad Técnica del Norte, 100150 Ibarra, Ecuador.*

VIVIAN F. LÓPEZ-BATISTA, *Department of Computer Science and Automatics, Universidad de Salamanca, 37008 Salamanca, Spain.*

RICARDO ARCINIEGA-ROCHA, *Department of Technologies, Instituto Tecnológico Superior 17 de Julio, Urcuquí 100650, Ecuador.*

DIEGO H. PELUFFO-ORDÓÑEZ, *Department of Engineering Corporación Universitaria Autónoma de Nariño, Pasto 520002, Colombia, and Modeling, Simulation and Data Analysis (MSDA) Research Program, Mohammed VI Polytechnic University, Ben Guerir 43150, Morocco.*

Abstract

Air pollution is a current concern of people and government entities. Therefore, in urban scenarios, its monitoring and subsequent analysis is a remarkable and challenging issue due mainly to the variability of polluting-related factors. For this reason, the present work shows the development of a wireless sensor network that, through machine learning techniques, can be classified into three different types of environments: high pollution levels, medium pollution and no noticeable contamination into the Ibarra City. To achieve this goal, signal smoothing stages, prototype selection, feature analysis and a comparison of classification algorithms are performed. As relevant results, there is a classification performance of 95% with a significant noisy data reduction.

Keywords: WSN, air pollution, data analysis

1 Introduction

The different microclimates of the planet are strongly connected. This is due to different factors such as sea currents, weather and moon movement, among others. These variables influence temperature, humidity, atmospheric pressure and precipitation in different continents. In this sense, it becomes a very complex system and any alteration can cause a serious impact on the planet. In recent years, one of the biggest concerns worldwide is the planet's rising temperature. It produces climatic variations that, on the one hand, can generate excessive heatwaves that erode the ground resulting in animal and plant death. On the other hand, aggressive rains generate floods and river overflows, among others [1]. This is mainly due to the uncontrolled industries growth that causes the extermination of forests and generates toxic air and water pollution [2]. These industrialization effects, together

*E-mail: pdrosero@utn.edu.ec

Vol. 00, No. 0,
doi:10.1093/jigpal/jzab005

with urbanization and individual mobility of people, have become a great health risk [3] [2]. Consequently, the World Health Organization (WHO) estimates that one in eight premature deaths is due to air pollution effects [4]. With this, it can be deduced that about 3 million people die from air pollution [5]. The most polluting and detectable gases, on the one hand, is nitrogen oxide (NOx), which is a generic term that refers to a group of highly reactive gases such as nitric oxide (NO) and nitrogen dioxide (NO2) that nitrogen and oxygen contain in various proportions [6]. The main sources of NOx are diesel buses, power generator plants and other industrial, commercial and domestic sources that burn oil fuels [1]. In the atmosphere, nitrogen oxides can contribute to the formation of photochemical ozone (smog or polluting fog) and have health consequences. In prolonged or continuous exposure, the nervous system and the cardiovascular system can be affected, causing neurological and cardiac disorders. On the other hand, for carbon monoxide (CO) [7], the main source is the transportation sector due to the incomplete combustion of gas, petroleum, gasoline and coal. Stoves or heaters are some examples of machines that burn fossil fuels. This type of contamination can lead to health conditions that can include mental confusion, dizziness, headache, weakness and consciousness loss. It also contributes to global warming and can lead to acid rain. Both gases also act as precursors to ozone formation that potentially aggravate climatic conditions [8, 9].

Government entities of each country have made actions to counteract climate change. One of the most important strategies is the implementation of environmental monitoring nodes located in different rural and urban sectors. With this, it is possible to have large volumes of data information to analyze them and propose strategies in the reduction of air pollutants [10]. For this reason, the Internet of Things (IoT) is a fundamental pillar for the deployment of electronic devices that can collect data. However, this technology implementation process must sometimes be installed in sectors that are difficult to access or where the cost of wired data transfer solutions is very expensive. For this propose, the Wireless Sensor Networks (WSNs), allows fulfilling the aim due to its flexibility of use, low power consumption and implementation of wireless communication protocols for its connection to a data storage server. This whole process is made from the use of sensors, which are responsible for acquiring data of the phenomenon to be studied and converting it by means of a transducer to an electrical signal that can be processed [11, 12]. However, the amount of acquired data can be received with a lot of noise. This is due to many reasons, such as the non-linearity of the electronic elements or the wear of the electronic device, among others. For these reasons, the signal must go through a cleaning and selection process to have reliable data on the phenomenon studied. In addition, due to the large area that cities cover, the vehicle density, location of companies and proximity to natural ecosystems, the pollution index is varied in different areas of the cities. The pollution index is varied within the city. For this reason, most applications that display air quality provide only an approximation of what is actually happening. To get the aim, the implementation of a WSN makes it possible to cover large areas of land and have information by sectors that provide real information on what is happening in relation to the air pollution[13]. With this, one of the main characteristics of an integrated system must be adaptability. That is, they can emulate some processing skills that the human brain performs. The same thing implies in some way, the ability to make decisions, learn from external stimuli, adapt to changes or the possibility of executing intelligent mathematical algorithms. Implicitly, it is based on a computational paradigm that receives or processes data to accomplish an assigned task [14].

Among the pollutants mostly analyzed are NO2, sulfur dioxide (SO2), and tropospheric ozone (O3); in the second range are solid particles (PM10), CO, hydrogen (H2) in consideration of the other pollutants analyzed [15, 16]. Studies such as [12, 17, 18, 19] and [20] have developed data collection and monitoring systems using WSN. In addition, [21] presents a single node IoT solution, [16] shows a solution to a web platform and [20] performs data analysis with a database already acquired. With these systems development, the most commonly used machine learning algorithms are Naive Bayes, decision tree, neural networks and decision bearing machine (SVM). However, there are open problems, such as the filtering of data from the sensors, a solution that covers the entire city and provides information differentiated by sectors on existing pollution. Besides, the evaluation in the selection of the classification algorithm that best suits the phenomenon studied.

With the aforementioned, this system shows the development and installation of 13 WSN nodes strategically located in different sectors of the city of Ibarra, Ecuador. The same ones that connect to WiFi networks or 4G cellular network according to the availability of the place. With this, data acquisition is realized by means of sensors to later carry out a stage of pre-processing of the information, where criteria of signal smoothing and feature selection are used. Subsequently, a performance analysis is carried out for the selection of the classification algorithm. For this stage, three types of cases are defined: (i) high levels of pollution, (ii) normally the presence of gases and (iii) absence of emissions. Finally, it presents the information obtained on an IoT server to view the state of the city in real time in relation to the location of the sensor. As relevant results, the performance of the classifier with the k-NN algorithm is 95% under real conditions.

The rest of the document is structured as follows. Section 2 shows the electronic design and location selection strategies for each WSN node. Section 3 presents the proposed data analysis. The results are presented in Section 4. Finally, the conclusions and future works are shown in Section 5.

2 Wireless sensor network design

This chapter shows the design of the WSN electronic system (2.1), and after the design is defined the locations of the WSN nodes and the method of acquisition and data filtering techniques (2.2).

2.1 Electronic design

The WSN design focuses on monitoring the most common gases and air pollutants such as NO_x and CO. In addition, a UV sensor is used to determine the maximum rates of radiation and temperature and relative humidity sensor in order to know its relationship with air pollution. Finally, there are additional elements such as the warning LEDs and the activation button. In some cases, if a direct Internet connection is not possible, there is a GPS module and a micro-SD input to store the data locally. All these data obtained by the sensors are processed by a microcontroller with a wireless network card.

With the established requirements of functionality and accuracy of the system. The following materials are defined. The `NodeMCU` is defined as microcontroller with the sensors `MQ-7` (carbon monoxide), `MQ-135` (carbon dioxide), `ML8511` (UV rays) and `DTH11` (temperature and humidity). With this, one of the main characteristics of an integrated system must be adaptability. That is, they can emulate some processing skills that the human brain performs. The connection diagram of the WSN node is shown in Figure 1.

Once the electronic system is designed, proceed whit the calibration of each of the sensors, especially those that acquire data from ambient gases. In this sense, the `MQ-7` and `MQ-135` sensors have an internal resistance that, in relation to its sensitivity curve as a function of the power of the type $y = ax^b$, must be configured for the detection of CO and NO_x, respectively. To do this, the

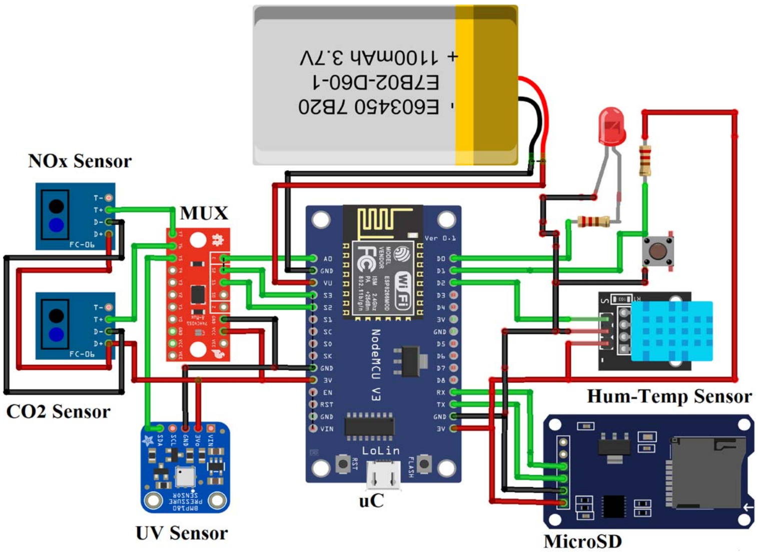


FIGURE 1. Electronic system developed.

equation 1 is used as explained in [22].

$$ppm = \left[\frac{\frac{R_s}{R_o}}{a}\right]^{\frac{1}{b}}, \tag{1}$$

where the values of $a = 20.6690525600$, $b = -0.656039042$ for CO detection $a = 5.5973021420$, $b = -0.365425824$ for NO_x detection with resistance of $R_s = 10k\omega$, $R_o = 100ppm$. For their part, the sensors `DTH11` y `ML8511` have their own libraries for adequate data conversion to the corresponding units. Consequently, sensors do not need to be calibrated.

In this sense, following reliability criteria for each sensor, some recommended performance measures are considered, such as the following: (i) precision, the ability to provide the same reading by repeatedly performing the same experiment (standard deviation); (ii) reproducibility, capacity to reproduce the same results by modifying the initial conditions of the experiment; and (iii) stability, the ability to produce the same output value in a long time. The general results obtained are compiled in Table 1, which corresponds to 10 tests in controlled environments to evaluate the data stability. As can be seen, the data collected from the sensors show an average error of 8% in contrast to those obtained with different available environmental monitoring applications such as Plume, AirQuality and GAIAairstation, among others. Such an error is acceptable enough for implementation purposes.

Because the sensors have slight reading errors and the most reliable data possible is desired. The signal smoothing criterion is used. Since this will eliminate the components in direct voltage and outliers considered as reading noise. In this case, a comparison will be made between the most relevant algorithms: mean, mean filter, Savitzky-Golay, Kalman and Gaussian for the data from the digital sensor [23]. Finally, the sensor with the best signal-to-noise ratio (SNR) will be chosen.

TABLE 1. Sensor performance metrics.

Measure	Sensors			
	MQ-7 (CO)	MQ-135 (NO_x)	DTH11 (temperature)	ML8511 (UV rays)
Precision	9 ±	7 ±	8 ±	7 ±
Reproducibility	It is necessary to wait up 10 seconds for calibration to be done		Adequate	Some reading errors
Stability	4±, variable for each test	2±, variable for each test	Adequate	Adequate

2.2 WSN location and data acquisition

Ibarra city is located north of the inter-Andean region of Ecuador, is a valley crossed by the Tahuando River, southeast of the Yahuarcocha Lagoon and is located at an altitude of 2215 meters above sea level. For the air pollution analysis, three zones are identified in the city. The first zone is the commercial part of the city, the second zone is the residential and educational one and the third zone is located in the suburbs of the city where there are many green mountainous places. In this sense, 5 nodes are installed in zone 1 to have maximum data pollution due to the constant traffic density in the city and 4 in zones 2 and 3 with the aim of having normal pollution data and pollution-free.

The label assignment is defined according to air quality index (AQI) values. This set of values that the AQI can take, is grouped into intervals to which a pattern or characteristic color of the air quality of a given area is associated. For this reason, the label Good (green) is defined to the obtained data in the early morning hours by the nodes of zone 3, Improvable (yellow) to the obtained data by the nodes of zone 2 when there is no vehicular density High and Bad (red) to the obtained data by the nodes located in zone 1 in hours of higher traffic density. This seeks to have a classifier system that shows the AQI assessment for each node in real-time. To do this, the data from each node is sent to an external server for its analysis stage. Where, after 2 months of data acquisition, a matrix is obtained: Matrix $\mathbf{Y} \in \mathbb{R}^{m \times n}$, where m is sample numbers and n represents the attributes of phenomena studied (sensors). Meanwhile, $\mathbf{L} \in \mathbb{R}^{m \times 1}$ is a characteristic vector. In this case, $m = 12000$ y $n = 5$.

3 Data analysis

This section shows the proposed data analysis scheme. Where it is shown in the data cleansing by means of the selection of prototypes (3.1). Subsequently, the selection of characteristics that will determine the variables to be used by the classification algorithm (3.2). Finally, the classification algorithms are shown in relation to their operational criteria (3.3).

3.1 Prototype selection

The prototype selection stage is carried out with the objective of having a training base in each WSN node. In this way, each of them can make their own decision based on their own experience. To do this, it is necessary to significantly reduce the obtained data, taking into account that the NodeMCU microcontroller only has 64k bytes as memory RAM. In this sense, [13] showed that the Condensed Nearest Neighbor (CNN) and Decremental Reduction Optimization Procedures 3

(DROP3) algorithms have a better performance in the elimination of instances and maintaining the intrinsic knowledge of the data. That is why only these algorithms will be used for this work.

3.2 Feature selection

Typically, for a given classification task, a vast number of attributes can be candidates for characterization purposes. However, many of such attributes may be irrelevant or redundant and, consequently, the classification algorithm may suffer from overfitting as well as some important characteristics of supervised learning may be impaired. Thus, the overall classification might not reach the expected performance. Indeed, in many applications involving large data sets, classifiers do not work properly until the unwanted features are mandatorily removed [24].

In such a vein, the feature selection task allows for reducing the number of attributes representing a data set that best fit a model or task of machine learning. There is a wide range of criteria and algorithms that can be used for this task. However, depending on both the nature of data and data analysis goals, some feature selection criteria may result in more suitable. In this work, the feature selection task is carried out through a one-stage approach. The filtering method called `ReliefF` is used for the feature selection itself, which is a target variable driven technique (being highly recommended for multi-label classifications) [25].

3.3 Classification algorithms

Since there are different algorithms, it is necessary to determine the appropriate one that can be presented to the previously acquired data set. Based on refined training, the main task of the system is the identification of environmental pollution through a supervised classification. Due to this, the classification criteria have been taken in relation to the literature review of the subject under development. The same are by distances (k-nearest neighbours), based on models (decision support machines) and deep learning (neural networks).

4 Results

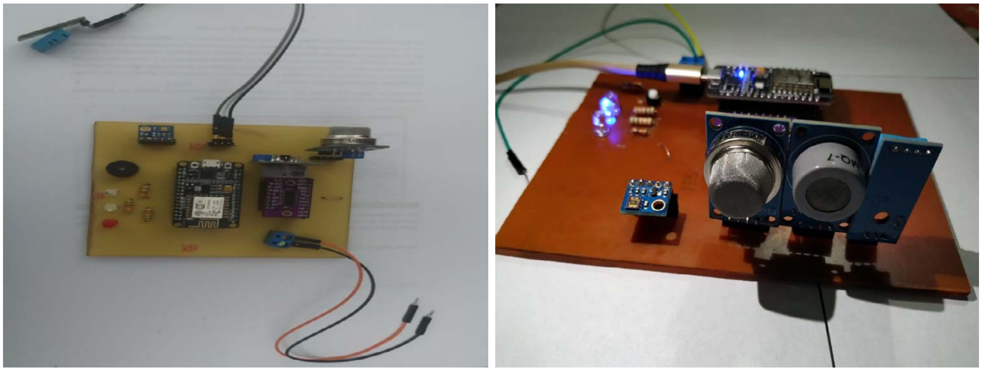
This section presents the overall data analysis results, as well as the systems execution tests. Results are divided into sections: WSN node (4.1), data smoothing (4.2), prototype selection (4.3), feature selection (4.4), classification (4.5) and implementation (4.6).

4.1 WSN node

The WSN node is developed to be implemented in the different sectors of the city of Ibarra. It has acrylic protection for moisture and rain. The two versions of the nodes (WiFi and GPS) are shown in Figure 2.

4.2 Data smoothing

The signal smoothing process uses the aforementioned algorithms and evaluates them with the SNR. With this, it is possible to determine the algorithm better eliminates the components related to noise. It should be noted that the initial SNR of the variables of each sensor are as follows: $\texttt{CO} = 2.12$dB, $\texttt{NOx} = 1.89$, $\texttt{UV} = 1.94$, $\texttt{Temp} = 7.12$ and $\texttt{Hum} = 6.55$. It should be mentioned that for the implementation of the signal smoothing algorithms, sale sizes of k value are used. Table 2 shows SNR results of each variable.

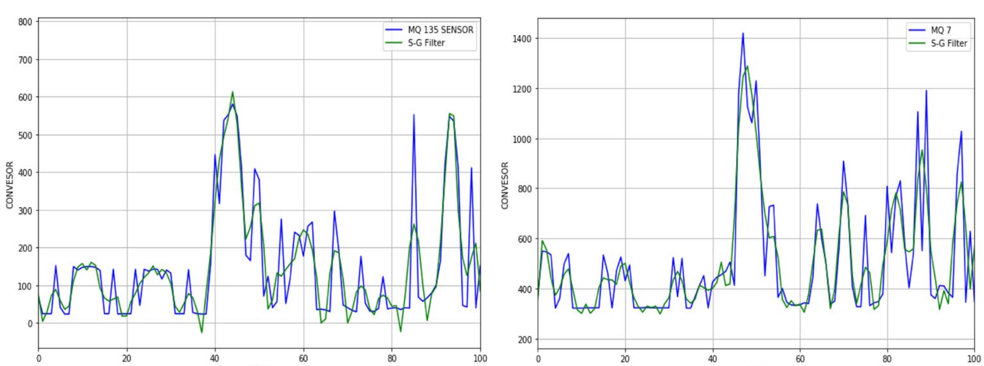


(a) electronic device developed (b) Nodes testing data

FIGURE 2. Nodes sensing data.

TABLE 2. SNR analysis.

Filter	**Config.**	**SNR (dB)**				
		MQ-7	MQ-135	UV	Temp	Hum
Median	k=20	2.20	1.98	2.07	7.20	6.68
Average	k=20	2.15	2.10	2.07	7.35	6.68
Gaussian	k=20, sigma=5	2.20	2.44	2.56	8.12	7.12
Savi-Golay	k=20, pol=4	2.74	2.68	3.01	9.61	8.87



(a) MQ 135 SENSOR (b) MQ 7 SENSOR

FIGURE 3. Data smoothing for MQ7 and MQ135 sensor.

The Figure 3 shows the data smoothing results for MQ 7 and MQ135 sensor

TABLE 3. Prototype selection results.

PS	Remov. inst	Remov. inst %	Time ejec.
CNN	1050	79.16	8.25 s
DROP 3	9200	76.6	42.4 s

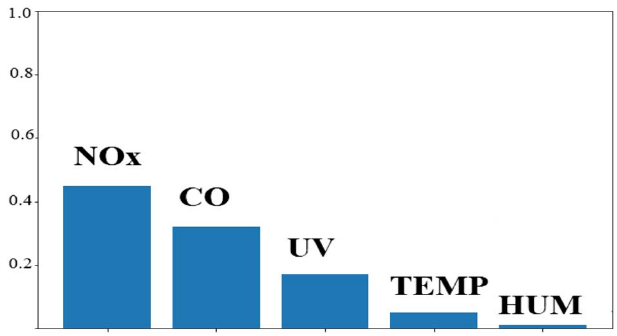


FIGURE 4. Feature selection analysis.

4.3 Prototype selection

In the prototype selection section, the CNN and DROP 3 algorithms were implemented. In Table 3, it can be seen that the algorithm that could eliminate more instances and has a significantly shorter time in execution is CNN.

4.4 Feature selection

With the selection of characteristics based on the `relief` algorithms, it was determined that the relative environmental humidity does not present any relevant information to the classifier. The UV radiation and temperature indices have a certain degree of relevance due to the hours of greatest pollution are relative to morning and afternoon hours. In Figure 4, the relevance of each variable is shown as a percentage.

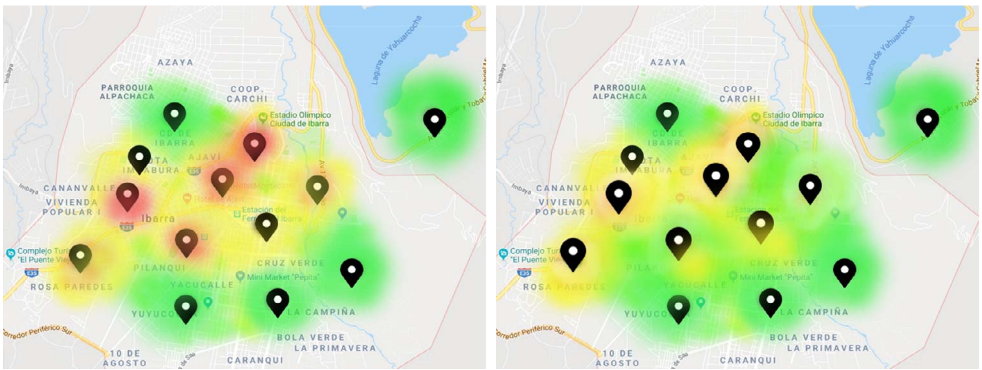
4.5 Classification

In order to determine the classification algorithm, it is to choose the set of cases to induce the classifier. In this sense, it can use the `Holdout` method, which divides the data set into two: training and testing. The test group is used to train the model and the test group to estimate the error rate. The `resampling` method becomes a generalization of the `Holdout` method, since this process is performed multiple times on different samples. With this, the error rate is based on the average of experiments performed. For this reason, the database was divided into ten different ways to train each algorithm and have an average error of each of them to get the different possible metrics from the confusion matrix.

Finally, Figure 6 shows an individual interface created for each node to visualize data acquisition by sensors and the real conditions for one WSN node. Some nodes that are exposed to weather conditions have been reinforced with a structure of agglomerated material that internally has a silicone reinforcement to prevent water seepage.

TABLE 4. Algorithms performance.

Algorithm	Accuracy	Error rate	Time exec.
k-NN	0.95	0.5	54 ms
SVM	0.90	0.1	77 ms
Neural Network	0.96	0.4	268 ms



(a) Nodes sensing data at midday (b) Nodes sensing data at night

FIGURE 5. Nodes sensing data.

(a) Interface developed for each WSN node (b) Node working in zone 3 (no air pollution)

FIGURE 6. WSN nodes implementation in real conditions.

In this sense, two solutions can be defined in relation to the type of system operation. On the one hand, if the decision on the IoT server is wished, it is advisable to use the neural network. Since it can be compiled in a system of great computational benefits. On the other hand, if the decision is locally at each WSN node, k-NN represents better functionality and lower resource costs. In this work, the second criterion is used to search for the individual solution in each node.

4.6 Implementation

Once the entire data analysis process has been completed. The k-NN algorithm is programmed on all WSN nodes. With this, each sample they take from the air quality, they process it and make a decision. This information is sent to the IoT server that receives the data from the sensors to store it with its label and in relation to them, generates a color to show the user the level of contamination. At the moment, the server is not for public use and is investigative in nature. Figure 5 shows the real air pollution measurements for each node on a normal day.

5 Conclusions

The integration of the WSN nodes for the monitoring of air pollution conditions in the city of Ibarra provided important information on the sectors of greatest problem they represent. With this, planning can be made on the implementation of green areas within the city. In addition, it was possible to validate the correct functioning of the system and the way in which the machine learning algorithm adapts to the changes for decision making. In this way, the wireless protocols used (WiFi and 4G) are stable for sending data.

The proposed methodology of data analysis, starting from the data smoothing, it had the correct criteria to provide adequate information to the classifier for the training of its model and the elimination of redundant data through the selection of prototypes and feature selection.

Acknowledgements

This work is supported by the Smart Data Analysis Systems group. http://sdas-group.com/

References

[1] A. K. Saha, S. Sircar, P. Chatterjee, S. Dutta, A. Mitra, A. Chatterjee, S. P. Chattopadhyay and H. N. Saha. A raspberry Pi controlled cloud based air and sound pollution monitoring system with temperature and humidity sensing. In IEEE 8th Annual Computing and Communication Workshop and Conference, CCWC 2018, vol. 2018-January, 2018. doi:10.1109/CCWC.2018.8301660.

[2] D. Wang, E. Duan, Y. Guo, B. Sun and T. Bai. Numerical simulation of the effect of over-fire air on NOx formation in furnace. In International Conference on Materials for Renewable Energy and Environment, IEEE, 2013, pp. 780–783. doi:10.1109/ICMREE.2013.6893790. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6893790.

[3] K. Sujatha, N. P. G. Bhavani and R. S. Ponmagal. Impact of NOx emissions on climate and monitoring using smart sensor technology. In International Conference on Communication and Signal Processing (ICCSP), IEEE, 2017, pp. 0853–0856. doi:10.1109/ICCSP.2017.8286488. URL http://ieeexplore.ieee.org/document/8286488/.

[4] K. Bashir Shaban, A. Kadri and E. Rezk. Urban air pollution monitoring system with forecasting models. *IEEE Sensors Journal*, **16**, 2598–2606, 2016.

[5] A. Maraj, S. Berzati, I. Efendiu, A. Shala, J. Dermaku and E. Melekoglu. Sensing platform development for air quality measurements and analysis. In South Eastern European Design Automation, Computer Engineering, Computer Networks and Social Media Conference

(SEEDA-CECNSM), IEEE, 2017, pp. 1–5. doi:10.23919/SEEDA-CECNSM.2017.8088233. URL http://ieeexplore.ieee.org/document/8088233/.

[6] H. Bae., Basic Principle and Practical Implementation of Near-Infrared Spectroscopy (NIRS).In *Smart Sensors and Systems*. Springer International Publishing, Cham, 2015, pp. 281–302. doi:10.1007/978-3-319-14711-6_12. URL http://link.springer.com/10.1007/978-3-319-14711-6_12.

[7] L. Peng, F. Danni, J. Shengqian, W. MingjieL. Peng, F. Danni, J. Shengqian, W. Mingjie A movable indoor air quality monitoring system. In *The 2nd International Conference on Cybernetics, Robotics and Control (CRC)*. IEEE, **2017**, 126–129, 2017.

[8] A. Quality Expert Group. Air Quality and Climate Change: A UK Perspective. http://webarchive.nationalarchives.gov.uk/20130403220722/http://archive.defra.gov.uk/environment/quality/air/airquality/publications/airqual-climatechange/documents/fullreport.pdf.

[9] P. D. Rosero-Montalvo, J. A. Caraguay-Procel, E. D. Jaramillo, J. M. Michilena-Calderon, A. C. Umaquinga-Criollo, M. Mediavilla-Valverde, M. A. Ruiz, L. A. Beltran, D. H. PeluffoP. D. Rosero-Montalvo, J. A. Caraguay-Procel, E. D. Jaramillo, J. M. Michilena-Calderon, A. C. Umaquinga-Criollo, M. Mediavilla-Valverde, M. A. Ruiz, L. A. Beltran, D. H. Peluffo Air quality monitoring intelligent system using machine learning techniques. In *International Conference on Information Systems and Computer Science (INCISCOS)*. IEEE, **2018**, 75–80, 2018.

[10] A. K. Y. Law. *Development of the Indoor Air Quality Index for Commercial Buildings in Hong Kong*, Thesis, The Hong Kong Polytechnic University, 2003. http://ira.lib.polyu.edu.hk/handle/10397/4193.

[11] Y.-L. Lin, C.-M. Kyung, H. Yasuura and Y. Liu (Eds.), *Smart Sensors and Systems*. Springer International Publishing, Cham, 2015. doi:10.1007/978-3-319-14711-6. URL http://link.springer.com/10.1007/978-3-319-14711-6.

[12] W. Wang, S. De, Y. Zhou, X. Huang and K. Moessner., Distributed sensor data computing in smart city applications. *In IEEE 18th International Symposium on A World of Wireless Mobile and Multimedia Networks (WoWMoM)*, IEEE, 2017, pp. 1–5. doi:10.1109/WoWMoM.2017.7974338. URL http://ieeexplore.ieee.org/document/7974338/.

[13] P. D. Rosero-Montalvo, V. F. López-Batista, D. H. Peluffo-Ordóñez, L. L. Lorente-Leyva and X. P. Blanco-Valencia. Urban pollution environmental monitoring system using iot devices and data visualization: a case study. In *Hybrid Artificial Intelligent Systems*, H. Pérez García, L. Sánchez González, M. Castejón Limas, H. Quintián Pardo and E. Corchado Rodríguez, eds, pp. 686–696. Springer International Publishing, Cham, 2019.

[14] P. D. Rosero-Montalvo, V. F. L. Batista, E. A. Rosero, E. D. Jaramillo, J. A. Caraguay, J. Pijal-Rojas and D. H. Peluffo-Ordóñez. Intelligence in embedded systems: overview and applications. In *Proceedings of the Future Technologies Conference (FTC), 2018*, K. Arai, R. Bhatia and S. Kapoor, eds, pp. 874–883. Springer International Publishing, Cham, 2019.

[15] S. M. Saad, A. M. Andrew, A. Y. M. Shakaff, A. R. M. Saad, A. M. Y and A. Zakaria Kamarudin. Classifying sources influencing indoor air quality (iaq) using artificial neural network (ann). *Sensors*, **15**, 11665–11684, 2015.

[16] D. Tudose, T. A. Pătraşcu, A. Voinescu, R. Tătăroiu, N. Ţăpuş Mobile sensors in air pollution measurement. In *The 8th Workshop on Positioning, Navigation and Communication*, 2011, pp. 166–170. doi: 10.1109/WPNC.2011.5961035.

[17] G. B. Fioccola, R. Sommese, I. Tufano, R. Canonico and G. Ventre. Polluino: An efficient cloud-based management of IoT devices for air quality monitoring. In *IEEE 2nd International*

Forum on Research and Technologies for Society and Industry Leveraging a Better Tomorrow (RTSI), IEEE, pp. 1–6, 2016. doi:10.1109/RTSI.2016.7740617. URL http://ieeexplore.ieee.org/document/7740617/.

[18] N. Kafli and K. Isa. Internet of Things (IoT) for measuring and monitoring sensors data of water surface platform. In *The 7th International Conference on Underwater System Technology: Theory and Applications (USYS)*, IEEE, 2017, pp. 1–6. doi:10.1109/USYS.2017.8309441. URL http://ieeexplore.ieee.org/document/8309441/.

[19] S. Kumar and A. Jasuja. Air quality monitoring system based on IoT using Raspberry Pi. In *The 2017 International Conference on Computing Communication and Automation (ICCCA)*, IEEE, pp. 1341–1346, 2017. doi:10.1109/CCAA.2017.8230005. URL http://ieeexplore.ieee.org/document/8230005/.

[20] R. W. Gore and D. S. Deshpande. An approach for classification of health risks based on air quality levels. In *The 1st International Conference on Intelligent Systems and Information Management (ICISIM)*, pp. 58–61, 2017. doi:10.1109/ICISIM.2017.8122148.

[21] A. Kamilaris, A. Pitsillides, F. X. Prenafeta-Bold and M. I. Ali. A web of things based eco-system for urban computing - towards smarter cities. In *The 24th International Conference on Telecommunications (ICT)*, pp. 1–7, 2017. doi:10.1109/ICT.2017.7998277.

[22] Hwsensor MQ7 technical data. https://www.hwsensor.com/.

[23] P. Kowalski and R. Smyk. Review and comparison of smoothing algorithms for one-dimensional data noise reduction. *International Interdisciplinary PhD Workshop (IIPhDW)*, **2018**, 277–281, 2018.

[24] M. Peker, A. Arslan, B. Şen, F. V. Çelebi and A. But. A novel hybrid method for determining the depth of anesthesia level: Combining relieff feature selection and random forest algorithm (relieff+rf). In *International Symposium on Innovations in Intelligent SysTems and Applications (INISTA)*, pp. 1–8, 2015. doi: 10.1109/INISTA.2015.7276737.

[25] Y. Zhai, W. Song, X. Liu, L. Liu and X. Zhao. A chi-square statistics based feature selection method in text classification. In *The 9th International Conference on Software Engineering and Service Science (ICSESS)*, **2018**, 160–163, 2018.

Received 10 February 2020