An AI-based Ventilation KPI using embedded IoT devices

Francisco Maciá-Pérez, Senior, IEEE, Iren Lorenzo-Fonseca, José-Vicente Berná-Martínez

***Abstract*—The air ventilation of enclosed premises has a direct impact on the occupants' well-being. If not properly regulated, the air ventilation can originate a multitude of diseases and pathologies. The present study proposes a new KPI (ventilation KPI) adapted to Smart Cities. It is especially designed for academic environments (Smart Universities) in which community members spend a long time gathered in classrooms, seminars, laboratories, etc. The ventilation KPI (or KPIv) was designed to support decision-making and is based on the estimation of the number of occupants of an enclosed space and the accumulation of existing CO2. Two AI techniques are proposed to perform these estimations, specifically, two regressive neural networks. The resulting models, together with the KPI were implemented through the development of value-added services for the University of Alicante's Smart University platform. The network models were designed to be embedded within the built IoT device prototypes. These prototypes are small and inexpensive. They act as intelligent sensors and are connected via a low consumption and emission network (LoRa). The case study showed that it is possible to take advantage of the pre-existing services and resources of these platforms, and to validate the KPIv.**

***Index Terms*— key performance indicator; Smart University; AI; ANN; embedded sensors; IoT**

I. Introduction

INDOOR Air Quality (IAQ) evaluation is also a notable indicator that measures the well-being of people occupying a common enclosed space [1]. This metric has become even more essential in the context of the COVID pandemic. The situation has revealed the need to be prepared for such contingencies and that temporary measures are far from satisfactory. Reliable, swift, affordable and long-lasting solutions are required.

This necessity to control a room's number of occupants and its ventilation is particularly relevant in academic environments. The common practice is that students and teachers share classrooms together over long periods of time. Indoor air quality is of course affected by outdoor air quality, but this quality can be significantly worsened by a range of independent factors. IAQ can be affected by a number of specific pollution sources and can have short-term as well as long-term negative effects on human health.

The immediate effects of indoor pollution include allergic reactions, headaches, dizziness, and fatigue. But the effects of long-term exposure can lead to serious respiratory illness and heart disease [2]. Moreover, aerosols accumulated in closed premises are the major transmitters of viral diseases.

To this end, a range of technological tools have emerged today that facilitate the monitoring and control of indicators. For example, IT platforms supporting Smart City developments are widespread and they are used, to a greater or lesser extent, on a global scale and in university environments are adopted as a Smart University platform. The data collected by these intelligent platforms is the basis of the metrics which–converted into KPIs–are used to evaluate the services and quality of life in cities [3].

Due to the characteristics of educational environments, Indoor Air Quality, and more specifically classroom ventilation indicators are necessary to measure the well-being of both students and teachers [4]. For all the above, this work proposed to enrich Smart University platforms with a new KPI to control classroom ventilation (KPIv) that is effective, quick to implement and deploy, sustainable over time, affordable, and low-maintenance. The KPIv is also based on artificial intelligence algorithms that enrich the information obtained from the current Smart University platform metrics data. Using AI, we can fill in uncertain information or anticipate risky situations where the allowed limits will be exceeded. The proposed KPIv uses CO_2 measurements and the institution's existing knowledge of its environment: the number of connected Wi-Fi network users, the structure and organisation DB, standard room specifications and classroom volumes, etc. All this information is available in the university's specific management applications and enables to evaluate the number and characteristics of a room's occupants and ventilation efficiency.

The KPIv proposal was implemented via a case study performed at the University of Alicante University. It was incorporated into the university's SmartUA platform architecture. When the validation phase will be completed, the prototype will be transferred over the current year to ten public universities, which are part of an initial consortium to deploy a common, nationwide Smart University platform.

The remainder of this paper is organized as follows. Section II presents a review of the technique's state of the art. Section III explains the research methodology, the development of the proposed KPIv (section 3.1) and the metrics it is composed of: Wi-Fi connections (section 3.2) and CO2 concentration (section 3.3). Section IV evaluates this methodology. Section V then concludes this paper.

Received: 27 October 2022. *Corresponding author: José-Vicente Berná-Martinez.*

Francisco Maciá-Pérez, Iren Lorenzo-Fonseca and José-Vicente Berná-Martínez are with the University of Alicante, ES03690 San Vicente del Raspeig, Alicante, Spain (e-mail: {pmacia, iren.fonseca, jvberna}@ua.es).

II. Background and Related Work

Concentrations of particles and gases in an enclosed space have been shown to be harmful to occupants [5]. This accumulation is directly related to the number of occupants and the room's ventilation. Several studies have shown that indoor CO2 concentrations can serve as markers of an environment's odour load due to the presence of its occupants if no other polluting sources (other than humans) are present. It is important to note, however, that human breathing CO_2 emissions are linked to that of other human metabolism.

A range of works [6] have searched for adequate methods to obtain a CO2-based occupancy control, in order to estimate the number of people present in a room. Most of them consider a person's individually-generated CO_2 level. The amount of CO_2 emitted by a person has in fact been defined in different studies [7]. Equation 1 allows to define the amount of carbon dioxide generated by a healthy person on a normal diet:

$$g[L/s] = 0.000569 * RQ * BMR * M \qquad (1)$$

Where: RQ: breathing ratio
BMR: basal metabolic rate in Megajoules(MJ)/day
M: metabolic activity expressed in MET
MET: the approximate amount of oxygen consumed per minute, by a person in a resting state

This measure, as can be observed (eq. 1), is highly conditioned by different factors. The first is the amount of energy required by the body to perform the most basic functions to maintain life, called Basal Metabolic Rate (BMR). It is dependent on age and sex. The second is the Breathing Rate (RQ), which some authors set at 0.85, while others, following the recommendations of [7], use 0,83. A third factor is metabolic activity (M) which also influences the indicator. Studies have found different measures for this metric according to the work or action being carried out. In addition, different metrics have been defined to determine this value in school-age children, where children's weight and height also play a role.

They have concluded that CO_2 monitoring is a good measure for both presence control and ventilation control and a black box model could thus be a good solution in this field. Artificial Neural Network (ANNs) tools have shown to be powerful and robust for both classifications and calculating regressions [8]. Over the years, ANN instruments have been widely used in multiple contexts for both linear and nonlinear relationships between input and output variables. In addition to their general advantages, a number of studies have used ANN techniques for occupancy control and in the domain of indoor air quality.

III. Methodology

The KPIv is based on the combined information of a room's occupancy and CO_2 concentration that allows to measure the room's ventilation. Deducing the number of occupants of a given space is feasible and viable using technological tools. This task is also facilitated by the presence of a functioning Smart University platform. Moreover, a room's CO_2 concentration can be measured using a network of sensors. Taking the above into account a KPI expresses a ratio based on the information of both the number of occupants and the CO_2 concentration in a room, and this KPIv allows the management of ventilation and permits Smart University tools to provide safe premises for the community. The KPIv is therefore perfectly aligned with the objectives of these types of smart platforms.

3. 1. Developing KPI ventilation (KPIv)

The number of occupants is drawn from the number of devices connected to the classrooms' Wi-Fi access points, along with other available information, such as the type of teaching, the time of day and the type of degree. All this information combined allows a more accurate determination of the number of people based on the number of Wi-Fi connections.

The number of occupants of a given enclosed space (#Personwifi) was thus calculated using the expression (eq. 2):

$$\#Person_{wifi}=f(num_{wifi\_conn}, type_{degree}, type_{teaching}, time) \qquad (2)$$

Where f() is a function that estimates the number of classroom occupants based on the defined parameters.

In addition, the CO_2 concentration value helps to estimate the number of people sharing the room ($\#Person_{co2}$) following previous studies that controlled presence based on CO*2* concentration [6]. To determine the number of occupants, we also proposed to use the room's dimensions, broken down into surface area by height, type of degree, season in the year, type of subject and the academic year of the students present. Based on these input values, the number of occupants was calculated using an ANN model. The model had been previously trained with these data, but under conditions of little or no ventilation (eq. 3). In this way, the result obtained for #Personco2 based on measurements under good ventilation conditions would always be lower than the actual number of occupants.

$$\#Person_{CO2}=g(concentration_{CO2}, area_{classroom}, height_{classroom}, type_{degree}, season_{year}, type_{subject}, year_{academic}) \qquad (3)$$

Where g() represents a function that estimates the number of co-existing people in the classroom based on the defined input parameters.

Using these definitions, the KPIv can be expressed as a function of a ratio between the CO_2 concentration-based estimate of number of occupants ($\#Person_{co2}$) and the Wi-fi-based estimate of number of occupants ($\#Person_{wifi}$). The objective was to keep the CO_2 concentration-based estimate always below the Wifi-based estimate, which would produce a measure of adequate ventilation. In this way, the Smart University platform dashboards should detect a problem in rooms with a KPIv greater than or equal to 1. To provide a more complete measurement, an alarm value was assigned to the CO2 concentration detected in the classrooms ($Alarm_{co2}$). Once this value is reached, the KPI takes on a negative value (i.e., greater than 1) regardless of the measurements obtained through the metrics.

Based on all the above, the KPIv could be expressed as (eq. 4):

$$KPI_v = \begin{cases} V_n\,, & if\ concentration_{CO2} \geq Alarm_{co_2} \\ \dfrac{\#\mathrm{Person}_{\mathrm{CO}_2}}{\#Person_{wifi}}, & otherwise. \end{cases} \quad (4)$$

3. 2. Metric wifi connections: #Person$_{wifi}$

The control of the number of people was carried out via the control of the number of Wi-Fi connections to the room's access points [9]. The data regarding the number of connections to an Access Point (AP) can be obtained almost immediately and provides, alone, valuable information for the Smart University platform. However, the number of AP connections in a classroom does not directly correspond to the number of people in the classroom. First of all, base connections exist owing to devices that are constantly connected to the AP, regardless of the people who are in the classroom. This is the case for example of the classroom's desktop computers, printers, sensors or smart devices. In addition, one person can account for more than one connection to the AP if they have multiple devices connected at once: laptop, mobile phone, tablet, smart watch, etc.

Also relevant are the parameters that influence the estimation of the number of people based on Wi-Fi connections. To select the most significant parameters, a survey was conducted with experts, based on an observational analytical study, using multiple-answer questions with one open item. As a result, a number of parameters were determined that affect the fluctuation of the estimation of the number of occupants according to the connections to the PA. These parameters were: time, type of teaching and type of university degree. These latter parameters were thus incorporated into the final model.

A basic ANN model was selected taking into consideration the need to obtain a reliable but also lightweight model, which could be embedded in an IoT device with limited resources; it also had to have been previously supported by several previous studies [8]. These models have evolved over the years. A number of tools today facilitate their use and allow obtaining lightweight models, specifically designed to be embedded in devices. In this way, we proposed that the function f() be a Regression Neural Network which, based on the defined input parameters, estimates the number of people present in the classroom. Table 1 shows the input analysis together with the discretisation or normalisation of the nominal data where necessary.

To select the ANN capable of predicting the number of people based on Wi-Fi connections (f()), we used a set of 50,000 input data that was selected to be representative of the broadest range of possibilities. The dataset was normalised using each set's mean and standard deviation. Additionally, based on the experimentation results, due to data convergence difficulties, we decided to convert the dataset's nominal fields into a one-hot, which increased the number of input variables. This pre-processing allowed to significantly reduce the errors compared to the raw data. The dataset was randomly divided into an 80:20 ratio to generate the training and test sets.

Table 1. Processing of input data.

Input Variables	**Type**	**Normalisation nominal values**	**Possible values**
Number of Wi-fi Connections	Discrete/ Numeric	-	x € N
Time	Discrete/ Numeric	(Hour not counting minutes)	x € {0, 1, 2, ..., 23}
Type of Teaching	Discreet/ Nominal	practical = 0; theory = 1; workshop = 2; laboratory = 3; seminar = 4	x € {0, 1, 2, 3, 4}
Type of University Degree	Discreet/ Nominal	economic = 0; science = 1; health sciences = 2; law = 3; education = 4; philosophy and humanities = 5; engineering = 6; Architecture = 7	x € {0, 1, 2, 3, 4, 5, 6, 7}

Once the dataset was ready, relu() was used as an activation function for the ANN model generation process. The decision was based on both the training speed and the fact that many libraries are optimised for use, which makes it easier to use the ANN in an embedded IoT device. Standard metrics for regressions such as Mean Square Error (MSE) were used to evaluate loss and Mean Abs Error (MAE) as a measure of accuracy. In addition, the early stop technique was applied to the training to avoid overfitting, and the weights were initialised following a uniform distribution between -r and +r where r=sqrt(3/m)

The training and analysis of the results were performed to model the f() function and to obtain the number of occupants according to the Wi-fi connections (#Personwifi). This led to the obtention of an ANN with an input layer of 15 neurons (once the nominal data was converted to one-hot), two fully connected hidden layers of 20 and 12 neurons, respectively, and an output layer with the result of the regression.

3. 3. CO_2 metric: #Person$_{co2}$

As in the case of the previous metric, the CO_2 concentration raw data, alone, already generates value for the Smart University platform. Different studies have succeeded at establishing CO_2 levels limits that can be harmful to humans in indoor environments. It is thus highly relevant to install sensors capable of measuring the concentrations of this gas in order to design alert mechanisms [2].

Regardless of their basic advantages, taking into account the previous studies [6, 8], the number of occupants of a given space (#Person$_{co2}$) can also be predicted based on CO_2 readings, along with other related variables (g()), such as the volume of the premises or the age of the attendees (eq. 3). This function will act as a numerator of our KPIv (eq. 4), providing an estimate of the number of a room's occupants.

To perform the estimation and to determine the relationship between the different variables involved, we opted for the application of an ANN model that was trained to this end. We proposed the use of an ANN with supervised learning for embedding the model into an IoT device and also due to the efficiency and effectiveness that has been demonstrated in previous research.

This article has been accepted for publication in IEEE Embedded Systems Letters. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/LES.2023.3238284

The training was conducted using the data of a range of CO_2 concentrations and varying numbers of people in closed premises where the model only contemplated the worst scenario: rooms with little or no ventilation. In this way, the CO_2-based measurement of the number of occupants of a ventilated room ($\#Person_{co2}$) would always give a lower value than the number of occupants detected by Wi-Fi ($\#Person_{wifi}$) connections. That is, once the g() function model is trained with zero ventilation values, when it receives the input values of a ventilated room's CO_2 concentration, the prediction of the number of people will be lower than the number of real people. This is because it will use its learning and predictably, the ventilation-detected CO_2 measurement will correspond to a smaller number of people if an unventilated room is used as a reference. Due to the above, the measure of KPIv < 1 would signify adequate premise ventilation.

Table 2. Input data model CO_2

Input Variables	Guy	Normalisation nominal values	Possible values
CO_2 concentration	Discrete/ Numeric	-	x € N (ppm)
Classroom area	Continuous/ Numeric	-	x € R x>0 (m2)
Classroom height	Continuous/ Numeric	-	x € R x>0 (m)
Type of University Degree	Discreet/ Nominal	economic = 0; science = 1; health sciences = 2; law = 3; education = 4; philosophy and letters = 5; engineering = 6; Architecture = 7	x € {0, 1, 2, 3, 4, 5, 6, 7}
Season of the year	Discreet/ Nominal	autumn = 0; winter = 1; spring = 2; summer = 3	x € {0, 1, 2, 3}
Type of course	Discreet/ Nominal	degree = 0; postgraduate = 1: personal training = 2	x € {0, 1, 2}
Academic year	Discrete/ Numeric	-	x € N x € {0, 1, 2, 3, 4, 5}

Table 2 shows the analysis of the inputs and their nominal data discretisation or normalisation where necessary. The code developed for training and testing is available in the public repository [10].

VI. Conclusions

Smart University platforms are capable of monitoring, controlling and even managing the proper functioning of a campus. They have a direct impact on the university community. Indoor air quality and classroom ventilation play a major role in the well-being of community members, especially in universities and educational environments generally. Space occupancy and CO_2 concentrations have proven to be useful metrics for ventilation management. The fact of being provided with indicators that evaluate the measures taken is therefore naturally very useful for the university's governing bodies and management teams. The indicators also strengthen the security of the entire community.

Based on this hypothesis, a ventilation KPI (KPIv) was proposed for Smart University platforms. It was based on a ratio between the estimation of the number of occupants according to CO_2 concentrations and the estimation of the number of occupants according to Wi-fi connections. To perform these estimates, neural network models were used for regressions due to their demonstrated effectiveness in the case of similar problems, presenting heterogeneous input data and no predefined data correlation.

The proposal was implemented using the KPIv of the University of Alicante's Smart University platform (SmartUA). The tests carried out using the selected datasets supported the effectiveness of the proposed ANN models. The service usage statistics and the results of the surveys conducted with the university's senior management further supported this efficiency.

In the short term, we plan to associate ventilation quality certificates with the campus buildings. This certificate would be dynamically updated with the measurements taken. In a more sustained way over time, the longer-term objective is to enrich the Smart University with new university-specific KPIs that can be used as governance and management tools.

References

[1] Zivelonghi, A., & Lai, M. (2021). Mitigating aerosol infection risk in school buildings: the role of natural ventilation, volume, occupancy and CO2 monitoring. Building and Environment, 204, 108139.
[2] Bidilă, T., Pietraru, R. N., Ioniţă, A. D., & Olteanu, A. (2021, May). Monitor Indoor Air Quality to Assess the Risk of COVID-19 Transmission. In 2021 23rd International Conference on Control Systems and Computer Science (CSCS) (pp. 356-361). IEEE.
[3] Petrova-Antonova, D., & Ilieva, S. (2018, August). Smart cities evaluation–a survey of performance and sustainability indicators. In 2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA) (pp. 486-493). IEEE.
[4] Miranda, M. T., Romero, P., Valero-Amaro, V., Arranz, J. I., & Montero, I. (2022). Ventilation conditions and their influence on thermal comfort in examination classrooms in times of COVID-19. A case study in a Spanish area with Mediterranean climate. International Journal of Hygiene and Environmental Health, 240, 113910.
[5] Saidan, M. N., Shbool, M. A., Arabeyyat, O. S., Al-Shihabi, S. T., Al Abdallat, Y., Barghash, M. A., & Saidan, H. (2020). Estimation of the probable outbreak size of novel coronavirus (COVID-19) in social gathering events and industrial activities. International Journal of Infectious Diseases, 98, 321-327.
[6] Yang, S., Huang, Z., Wang, C., Ran, X., Feng, C., & Chen, B. (2021). A real-time occupancy detection system for unoccupied, normally and abnormally occupied situation discrimination via sensor array and cloud platform in indoor environment. Sensors and Actuators A: Physical, 332, 113116.
[7] Johnson, D. L., Lynch, R. A., Floyd, E. L., Wang, J., & Bartels, J. N. (2018). Indoor air quality in classrooms: Environmental measures and effective ventilation rate modeling in urban elementary schools. Building and Environment, 136, 185-197.
[8] Vanus, J., Gorjani, O. M., Dvoracek, P., Bilik, P., & Koziorek, J. (2021). Application of a New CO_2 Prediction Method Within Family House Occupancy Monitoring. IEEE Access, 9, 158760-158772.
[9] Mehmood, U., Moser, I., Jayaraman, P. P., & Banerjee, A. (2019, April). Occupancy estimation using WiFi: A case study for counting passengers on busses. In 2019 IEEE 5th World Forum on Internet of Things (WF-IoT) (pp. 165-170). IEEE.
[10] Macia F, Lorenzo I, Berná JV. kpi_ventilation. Github. 2022. https://github.com/irenfonseca/ventilation-kpi/blob/main/kpi_ventilation.ipynb