ARTIFICIAL INTELLIGENCE ENABLING DIGITAL TWINS IN EXISTING BUILDINGS

Eng. Luca Rampini
Prof. Fulvio Re Cecconi

Department of Architecture, Built environment and Construction engineering, Politecnico di Milano, **Italy**

ABSTRACT

In response to the pandemic caused by the COVID-19 virus, many countries adopted control and mitigation measures (e.g., lockdown, social distancing, intermittent attendance, etc.) that significantly increase the amount of time people spend inside the built environment. Nevertheless, the need to resume activities in attendance and ensure the continuity of businesses that cannot be stopped has led to the adoption of measures to reduce the risk of contagion, mainly involving increased ventilation of indoor environments. In this context, the management of Indoor Air Quality became crucial to ensure occupants' health while optimizing thermal comfort and buildings' energy demands. In the scientific literature, one of the most explored solutions to this optimisation problem is the adoption of a digital twin (DT) triggered by the broader diffusion of sensors and technological systems. DTs have three main domains: the physical one (the asset), the virtual one (the digital counterpart), and the cyber-physical (sensors and devices to send data to the digital domain) that binds the two together. However, installing a proper number of sensors in existing buildings may be too expensive or impossible due to technical problems. This study introduces a methodology to take advantage of DTs in existing buildings where technical and economic reasons hinder the deployment of a fully developed cyber-physical domain, replacing hard sensors with soft/virtual ones deploying Artificial Intelligence (AI) techniques. AI can leverage the cyber-physical part of DTs by retrieving useful information on occupant behavior from even few available data, thus allowing for a fully functional DT even in existing buildings with a limited number of installed sensors. The methodology has been tested on a school building in Italy, where a DT has been deployed. In the case study, low-cost and easy-to-install sensors are used to monitor CO_2 levels within various rooms. Many control and mitigating protocols adopted during the pandemic set thresholds and limits for CO_2 concentrations, requiring increased ventilation once a certain CO_2 threshold is reached, with consequent repercussions on indoor comfort and energy consumption. In this study, the opening of windows is automatically detected with AI techniques by examining CO_2 temporal trends, thus automatically recognizing building users' behaviors. Consequently, systems can be properly managed to limit energy consumption and thermal discomfort (e.g., turning off the heating while the window is on). Therefore, the introduced methodology enables a DT approach, overtaking difficulties often encountered in installing sensors in existing assets, with obvious benefits in terms of health, energy consumption, and economic savings.

Keywords: Digital Twin, Artificial Intelligence, Neural networks, School building

INTRODUCTION

The recent pandemic has increased the amount of time spent inside buildings due to containment measures based on social distancing and lockdowns. However, most buildings present critical flaws in terms of ventilation and filtration for indoor spaces since they are set for bare minimums and not designed for infection control [1]. In this scenario, the attention has shifted towards the assessment and improvement of Indoor Air Quality (IAQ) [2]. In particular, increased indoor ventilation can reduce air contamination with virus-laden aerosols such as SARS-CoV-2 [3]. Regulations and standards frequently assume CO_2 concentration as an indicator of IAQ to evaluate the occupancy level and the indoor environment healthiness [4]. Since most buildings do not contain significant internal sources of CO_2, apart from occupants, CO_2 might be considered a surrogate for exhaled breath [5]. Thus, monitoring indoor concentrations of CO_2 has become crucial to influence behaviors aimed at better IAQ management [6].

Recently, spurred by the widespread diffusion of wireless sensors and devices – the so-called Internet of Things (IoT) -, a new concept for data-driven management of physical assets has emerged: the Digital Twin (DT). According to [4], DTs have three main elements: the physical artifact that has to be managed, the virtual model used to store all the information required, and the connection linking the two together. This connection is characterized by exchanging data, information, and knowledge between the physical and cyber counterparts, enabled by the diffusion of IoT sensors and advanced analytics technologies [7]. In the traditional approach, DTs perform better if the amount of data generated is vast; however, in existing assets that characterize most of the built environment, installing a significant number of sensors is not always possible due to several limitations (accessibility, cost, etc.). Therefore, for the management of the Built Environment, a common approach is to create a Soft DT, a digital replica characterized by a low level of geometrical information but still able to collect real-time data from the indoor building spaces and make analyses [8]. In this context, the introduction of Artificial Intelligence (AI) techniques might help understand information from a limited set of data. AI is a generic term that refers to any algorithm that can observe its environment, learn, and make intelligent actions or propose decisions based on knowledge and experience [9]. Many different techniques fall under this definition, but at the moment, Machine Learning (ML) techniques are the most widely used.

Specifically, this research investigates a case study from a primary school in Italy, where low-cost, easy to install sensors are adopted to monitor CO_2 levels in different classrooms. Many control and mitigating protocols set thresholds and limits for CO_2 concentrations during the pandemic, requiring increased ventilation once a specific CO_2 limit is reached. The increased air circulation, especially in Italian school buildings, is done mechanically or manually by opening the windows, impacting indoor comfort and energy consumption. In this study, the opening of windows is automatically detected with ML techniques. By analyzing data sequences, the modern algorithm can understand users patterns and help identify the motivations that trigger virtuous behavior or the problems that hinder the successful implementation of mitigation procedures [10]. Consequently, heating systems can be adequately managed to limit energy consumption and thermal discomfort (e.g., turning off the heating while the window is on).

The introduced methodology enables a soft-DT approach, overtaking difficulties often encountered in installing sensors in existing assets, with obvious benefits in terms of health, energy consumption, and economic savings.

MATERIALS AND METHODS

The case study investigated in this research is a primary school building located in Milan, Italy (Figure 1). Specifically, the research focused on three classrooms that have the characteristics shown in Table 1.

a) b)

Figure 1. The case study building: a) the view from the street with the main entrance and b) a satellite view (image from maps.google.it)

Inside the building, the CO_2 concentrations are collected by sensors connected through the IoT paradigm: defining a shared data model for each of the sensors installed in the building and using a web-based communication protocol [11]. The network is composed of fourteen multi-purpose SAF TEHNIKA ARANET4 sensors deployed across the classrooms. The data gathering has been going on since December 2020; however, the case study will focus on the period from December 2020 to March 2021. The retrieved data needed some cleaning work, especially for the school closure period, which showed minor anomalies, probably due to power supply failures. The sensors measured CO_2 concentration every 30 seconds; however, for this research, the granularity of the measurements was set to two minutes (applying the mean) in order to obtain a more efficient training of the model.

Table 1. Sensorized spaces in the case study school

Classroom	level	students	Age [years]	Area [m^2]	Windows [m^2]	Floor	Window opening [min/h]	Facing
Quarta A	primary	17/19	9-10	55.62	6	second	13	garden
M_Arancione	nursery	24/26	3-5	51.73	8	first	8	garden
Mininfanzia	kindergarten	24/26	0-3	43.99	7	ground	12	street

Figure 2 shows the CO_2 trend in a typical school day on 17th December 2020 (from 8 a.m. until 4 p.m.). The CO_2 concentrations are different among the classrooms, so it is difficult to derive a general law that guesses the trend of concentrations. In fact, CO_2

levels depend on several features that influence the outcome (students, area, age, etc.). In addition, the available data has no labels, making classical supervised learning algorithms challenging to apply to determine window openings.

Nonetheless, by looking at the trend of CO_2 concentrations, it is possible to guess the opening of windows when there is a sharp drop in CO_2 levels (except at lunchtime when the decline is due to classroom emptying). Therefore, the identification of opened windows can be treated as a "system fault" in the CO_2 trend.

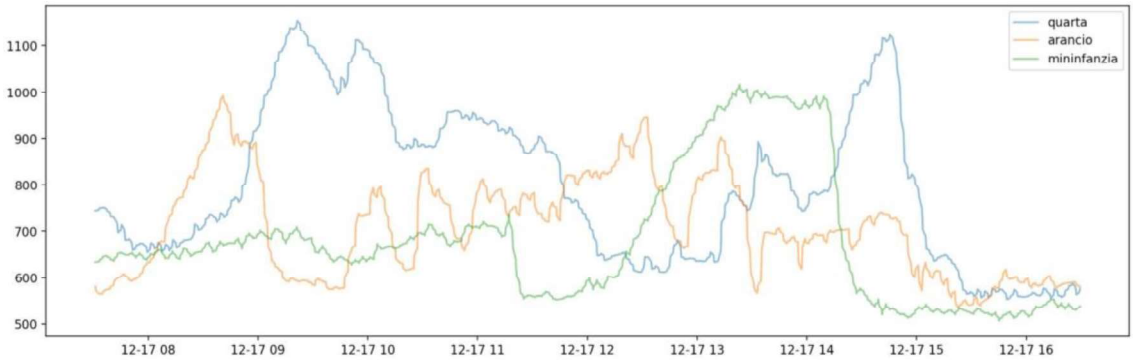


Figure 2. CO_2 concentrations collected on 12/17/2020 in three classrooms

Identifying anomalous data instances with respect to the whole dataset is usually referred to as "collective anomaly" [12]. For such a problem, one typical approach is to tweak the usage of time series forecasting: we can fit a model to a specific period before and predict the value after. The actual value is then compared to see if it falls into the prediction interval. The model is validated using the forward chaining method (see Figure 3) [13]. In the first step, the model is trained on the last n-data points (blue) and validates the prediction on the next m-data points (orange). Then, sliding the n+m training/validation window in time (the blue part in the middle plot), the model is again validated using the following m-data points.

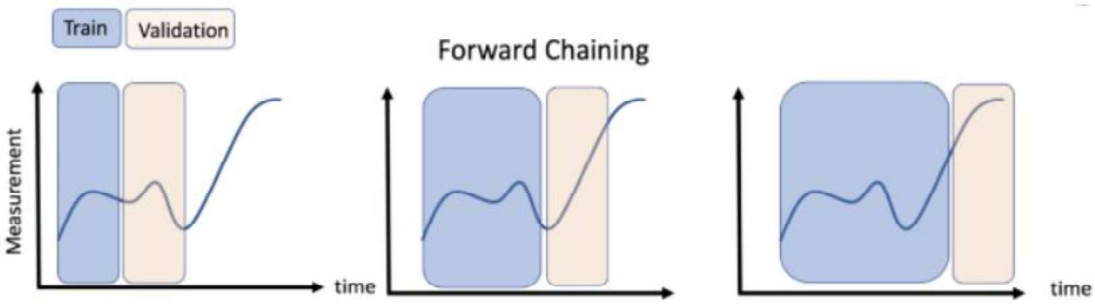


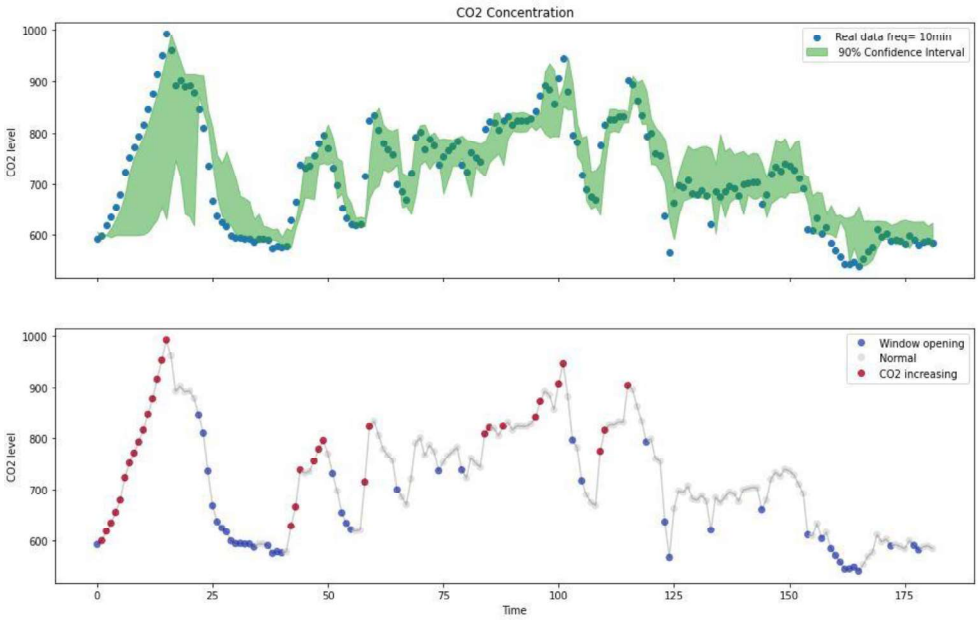
Figure 3. Concept of forward chaining. The features and targets are extracted from the time series using a sliding window [14]

In addition to the main regression model, two more quantile regressors are trained with different significance levels to predict the upper and lower bounds of the prediction interval, with which it is possible to assess whether the actual value is above or below the interval band.

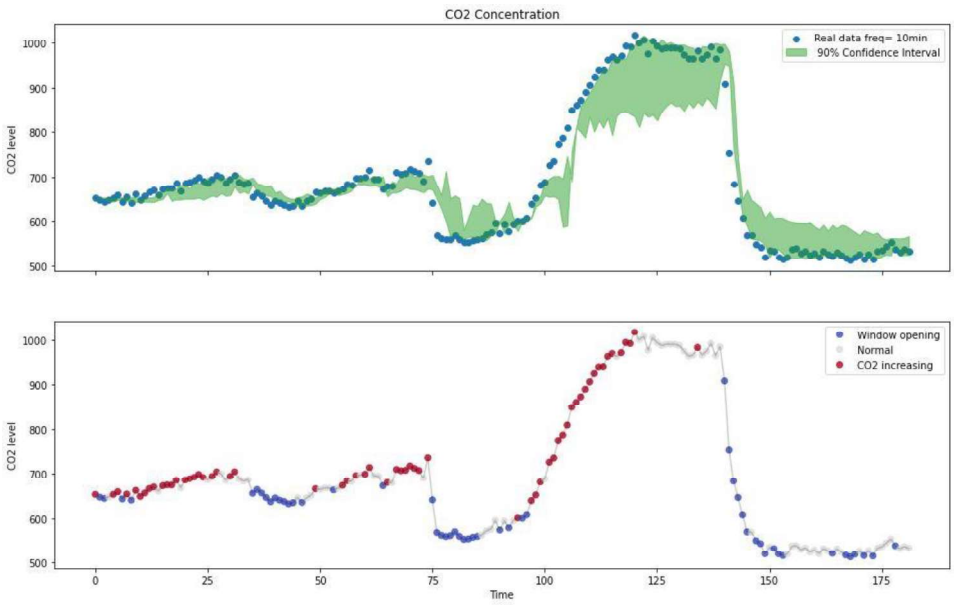
RESULT

The described method is applied to the collected dataset, giving the results shown in Figure 4. The dataset is sampled every 10 minutes, and the features are extracted and fed into three Gradient Boosting Regressor models (one is the main and two are quantile regressors) using the Python library Scikit-Learn. The significance levels are chosen so that the prediction level represents a 90% confidence interval (the green area in Figure 4). The dots outside the green area are detected as anomalies and flagged based on whether they are growing or declining. In particular, a decreasing succession outside the forecast interval, corresponds to the opening of the windows.

a) Arancio



b) Mininfanzia



c) Quarta

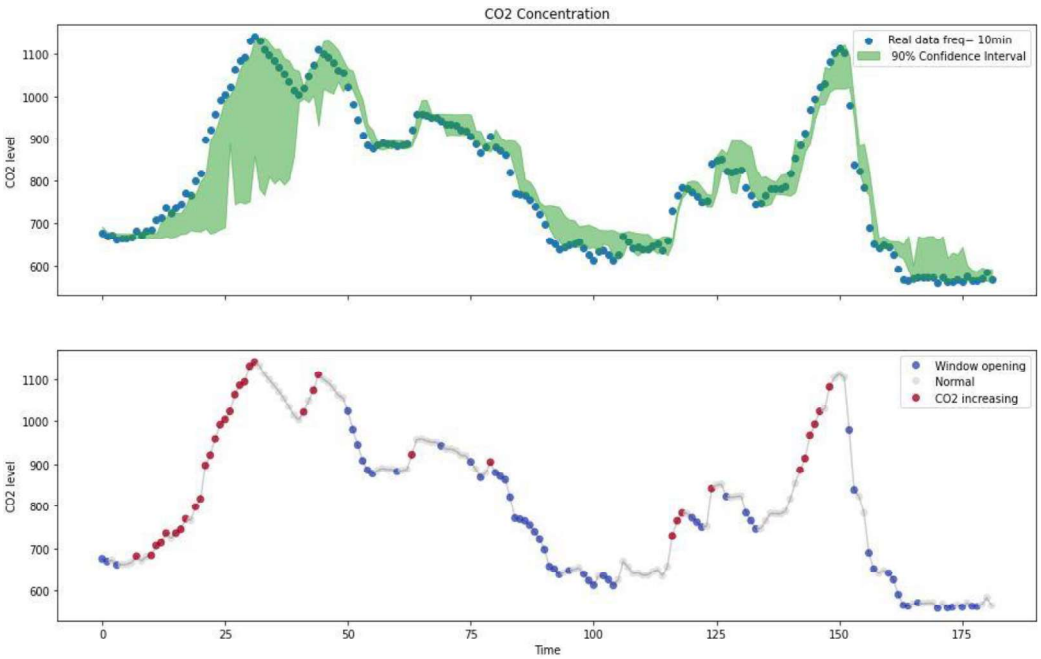


Figure 4. Flagging results with the CO_2 concentrations time-series data (sampled every 10 minutes) collected on December 17th 2020, by applying the quantile regressor method. Top: data and prediction interval; Bottom: data with flags showing above or below the prediction interval. The plot refers to a) Arancio, b) Mininfanzia, and c) Quarta classrooms

The model shows its ability to detect anomalous decreases in CO_2 concentrations and tag them accordingly. Although the collected data does not come with labels, the estimation seems to commit false positive errors(i.e., a decline of CO_2 level not caused by opening windows) rather than false negative errors (i.e., the windows were opened, but the model could not detect it, based on the CO2 measures). These results are positive because it is possible to save energy every time the CO2 level decreases by linking the model to the systems management. Therefore, cases where the pupils and the staff leave the room for the lunch break are also covered.

DISCUSSION AND CONCLUSION

This study uses modern AI techniques to create a soft-DT that supports decision-making processes that meet precise target objectives (e.g., minimizing the infection risk; maintaining indoor comfort parameters within standard ranges; minimizing the energy consumption during health, safety, and environment procedures). Specifically, in this research, an IoT system formed by low cost and easy to install sensors is implemented to monitor CO_2 concentrations as proxy data for airborne contagion risk. In a cold climate, such as the winter period in Northern Italy, opening windows, although conducive to the mitigation of infection risk, can significantly reduce internal thermal comfort and, at the same time, increase energy consumption. In this scenario, the introduced ML model can understand when windows are opened in compliance with safety procedures. This automatic process might help to find the right compromise between safety, comfort, and economic aspects. For instance, the heating can be turned off if windows are opened, or an audible signal can indicate when windows can be closed once sufficient time has passed to mitigate the risk.

The case study shows the model's ability to understand the opening of windows, although the data collected are not labeled, limiting the model choices and performances. The soft DT can prevent and adjust uncomfortable and unhealthy classroom conditions with great benefits for pupils, staff, and facility managers. In addition, if some labels are provided in future research, it could be possible to adopt supervised or semi-supervised solutions that might be more accurate and easy to achieve.

Beyond this quantile regression model, deep learning models such as Long Short Term Memory (LSTM) might be able to achieve better performance. LSTM is a specialized artificial Recurrent Neural Network (RNN) that is the state-of-the-art choice when dealing with larger and massive datasets. However, a more straightforward approach, like the one presented here, is more feasible for soft DT applications since it does not take much time and effort to set up.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to PhD Paolo Bellagente, prof. Angelo Ciribini, PhD Stefano Rinaldi and prof. Lavinia Tagliabue for providing the data used in the case study.

REFERENCES

[1] J. G. Allen and A. M. Ibrahim, "Indoor Air Changes and Potential Implications for SARS-CoV-2 Transmission," JAMA - Journal of the American Medical Association, vol. 325, no. 20. American Medical Association, pp. 2112–2113, 25-May-2021, doi: 10.1001/jama.2021.5053.

[2] M. Awada et al., "Ten questions concerning occupant health in buildings during normal operations and extreme events including the COVID-19 pandemic," Build. Environ., vol. 188, p. 107480, Jan. 2021, doi: 10.1016/j.buildenv.2020.107480.

[3] L. Dietz, P. F. Horve, D. A. Coil, M. Fretz, J. A. Eisen, and K. Van Den Wymelenberg, "2019 Novel Coronavirus (COVID-19) Pandemic: Built Environment Considerations To Reduce Transmission," mSystems, 2020, doi: 10.1128/msystems.00245-20.

[4] J. F. Meadow et al., "Indoor airborne bacterial communities are influenced by ventilation, occupancy, and outdoor air source," Indoor Air, vol. 24, no. 1, pp. 41–48, 2014, doi: 10.1111/ina.12047.

[5] S. N. Rudnick and D. K. Milton, "Risk of indoor airborne infection transmission estimated from carbon dioxide concentration," Indoor Air, vol. 13, no. 3, pp. 237–245, Sep. 2003, doi: 10.1034/j.1600-0668.2003.00189.x.

[6] J. Heydon and R. Chakraborty, "Can portable air quality monitors protect children from air pollution on the school run? An exploratory study," Environ. Monit. Assess., vol. 192, no. 3, pp. 1–16, Mar. 2020, doi: 10.1007/s10661-020-8153-1.

[7] R. Rosen, G. Von Wichert, G. Lo, and K. Bettenhausen, "About the importance of autonomy and digital twins for the future of manufacturing," IFAC-PapersOnLine, vol. 48, no. 3, pp. 567–572, May 2015, doi: 10.1016/j.ifacol.2015.06.141.

[8] F. Re Cecconi, E. De Angelis, L. C. Tagliabue, S. Rinaldi, P. Bellagente, and A. L. C. Ciribini, "Soft Digital Twin for IEQ enabling the COVID risk mitigation in educational spaces," in SDEWES 2021, 2021, pp. 1034-1,1034-12.

[9] M. (Ed. . Craglia et al., "Artificial Intelligence - A European perspective." EUR 29425 EN, Publications Office, Luxembourg, 2018, ISBN 978-92-79-97217-1, doi:10.2760/11251, JRC113826., 2018, doi: 10.2760/11251.

[10] A. Tealab, "Time series forecasting using artificial neural networks methodologies: A systematic review," Futur. Comput. Informatics J., vol. 3, no. 2, pp. 334–340, Dec. 2018, doi: 10.1016/J.FCIJ.2018.10.003.

[11] Q. Huang and K. Kieffer, "An intelligent internet of things (IoT) sensor system for building environmental monitoring," J. Mob. Multimed., 2019, doi: 10.13052/jmm1550-4646.15122.

[12] L. Bontemps, V. L. Cao, J. McDermott, and N.-A. Le-Khac, "Collective Anomaly Detection Based on Long Short-Term Memory Recurrent Neural Networks," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 10018 LNCS, pp. 141–152, Nov. 2016, doi: 10.1007/978-3-319-48057-2_9.

[13] F. Bacchus and F. Kabanza, "Using temporal logics to express search control knowledge for planning," Artif. Intell., vol. 116, no. 1–2, pp. 123–191, Jan. 2000, doi: 10.1016/S0004-3702(99)00071-5.

[14] "Time Series Machine Learning Regression Framework | by Pourya | Towards Data Science." [Online]. Available: https://towardsdatascience.com/time-series-machine-learning-regression-framework-9ea33929009a. [Accessed: 19-Oct-2021].