Applied Thermal Engineering 261 (2025) 125122

Contents lists available at ScienceDirect

Applied Thermal Engineering

journal homepage: www.elsevier.com/locate/apthermeng

Research Paper

Design, comparison and application of artificial intelligence predictive models based on experimental data for estimating carbon dioxide concentration inside a building

Vincenzo Ballerini [a], Paolo Valdiserri [a,*], Dorota Anna Krawczyk [b], Beata Sadowska [b], Bernadetta Lubowicka [b], Eugenia Rossi di Schio [a]

[a] Alma Mater Studiorum – University of Bologna, Department of Industrial Engineering DIN. Viale Risorgimento 2, 40136 Bologna, Italy
[b] Bialystok University of Technology, Faculty of Civil Engineering and Environmental Sciences, Wiejska Street 45E, 15-351 Bialystok, Poland

ARTICLE INFO

Keywords:
Predictive models
Artificial intelligence
CO_2 estimation
Microcontroller
IAQ
Public buildings

ABSTRACT

This paper describes applications of artificial intelligence to estimate indoor carbon dioxide levels, even in the absence of a direct CO_2 sensor, based on simply obtained indoor parameters. The aim is to develop a novel universal model that is easy to use in a wide range of applications and easy to replicate. The presented predictive models are based on experimental data collected via affordable microclimate stations. They are built with Wi-Fi-enabled microcontrollers similar to Arduino, designed to monitor environmental factors such as temperature, humidity, human presence, atmospheric pressure, and carbon dioxide levels. The experimental data used to train and test the models are specifically collected for this analysis in a primary school classroom in Bialystok, Poland, through the homemade stations that gather data minute-by-minute. Linear and non-linear models include machine learning models such as SVM, random forest, decision tree, and neural networks. Among the various models tested, the random forest method, which relied solely on temperature, humidity, and human presence measurements, produced the most accurate results, achieving an R-squared value of 0.89. The use of temperature, humidity, and presence sensors, which are more affordable than CO_2 sensors, highlights the novelty of the present analysis and the cost-effectiveness of predictive modelling in environmental monitoring.

1. Introduction

Occupants of buildings are continuously subjected to various indoor environmental stimuli, including factors related to thermal conditions or air quality [1]: proper Indoor Air Quality (IAQ) is crucial for the well-being of buildings' users. Recently, we have observed a tendency to reduce energy consumption for cooling and heating, often achieved by sealing the building envelope and decreasing the ventilation air exchange rate per hour (ACH). While this trend may ensure the appropriate indoor temperature, it may result in an undesirable increase in carbon dioxide concentration. This aspect is delicate and essential, especially regarding educational buildings, where often CO_2 concentrations higher than the limit occur [2]. Various solutions were proposed to improve the IAQ, starting from the use of plants that have air-purifying properties not only outside a building [3] but even inside [4,5], or by setting the maximum occupied space of 2.3 m^2 per person [6]. Angelova et al. [7] found that it could decrease the CO_2 concentration by changing the indoor temperature in classrooms; however, it could influence thermal comfort. The in-situ estimation of dynamic air change rates for various window-opening configurations based on occupant metabolic CO_2 emissions was proposed by Schreck et al. [8].

IAQ and heating energy consumption are intricately linked in densely occupied spaces like classrooms. Labihi et al. [9] conducted an experimental simulation study in a university classroom in Rennes, and they concluded that upgrading the controlled mechanical ventilation (CMV) with a double flow ensures air quality without increasing heating consumption. Loreti et al. [10] ran a simulation study on an Italian school applying different strategies to enhance energy performance. Regarding the installation of heat recovery, they conclude that it improves indoor air conditions and comfort, but the energy cost is increasing. To reduce the economic cost regarding the use of CMV, it is crucial to correctly associate renewable energy sources, such as photovoltaic panels, and optimize the module distribution (i.e., orientation row distance, tilt angle, etc.) [11]. Evaluating different modelling

* Corresponding author.
E-mail address: paolo.valdiserri@unibo.it (P. Valdiserri).

https://doi.org/10.1016/j.applthermaleng.2024.125122
Received 14 September 2024; Received in revised form 11 November 2024; Accepted 1 December 2024
Available online 7 December 2024

Nomenclature

Symbols

D (m)	Diameter
DT (K)	Temperature difference
$PRES_i$ (°C)	Presence of people (Boolean value)
RH (%)	Relative humidity
$t_{5,i}$ (°C)	Air temperature at 0.05 m from the floor
$t_{120,i}$ (°C)	Air temperature at 1.20 m from the floor
t_{air} (°C)	Air temperature
t_{bg} (°C)	Black-globe temperature
t_{mr} (°C)	Mean radiant temperature
U (W/(m^2 K))	Total heat transfer coefficient
U_{doors} (W/(m^2 K))	Transmittance of the doors
U_{walls} (W/(m^2 K))	Transmittance of the walls
$U_{windows}$ (W/(m^2 K))	Transmittance of the windows
$y_{CO2,day,i}$	Estimated daily CO_2 concentration
ε (−)	Emissivity
$\Delta\tau_i$ (s)	Time difference between two consecutive measurements
Subscript $_i$	*i-th* value of the considered variable

Acronyms

ACH	Air Exchange Rate
ADC	Analog To Digital
AI	Artificial Intelligence
ANN	Artificial Neural Networks
CMV	Controlled Mechanical Ventilation
CO_2	Carbon dioxide
IAQ	Indoor Air Quality
IoT	Internet of Things
HVAC	Heating, Ventilation and Air Conditioning
ML	Machine Learning
MPC	Model Predictive Control
NTC	Negative Temperature Coefficient
R^2	Determination Coefficient
RMSE	Root Mean Squared Error
SHS	Smart Healthy Schools
SVM	Support Vector Machine
USB	Universal Serial Bus

approaches for predicting levels of CO_2 is crucial for the proper designing of HVAC systems and maintaining good IAQ. Models based on different indoor parameters, such as indoor temperature, pressure, number of users, ACH were developed by Lawrence&Braun [12], Liu et al. [13], Škrjanc&Šubic [14], Krawczyk et al. [15,16]. The comprehensive study of Ma et al. [17] compared the most common thermal comfort models, and their variables related to steady-state and adaptive models. Many authors have experimentally measured the air quality parameters [18,19], and recently, attention has been paid to the use of low-cost sensors or equipment, as by Gonzales Rivero et al. [20], Ballerini et al [21], Tryner et al. [22]. A review of the application of low-cost sensing technology for indoor air quality monitoring is presented by Sà et al. [23].

Researchers and designers have increasingly applied artificial intelligence (AI) in building design, usage, and proper maintenance. Liu et al. [24] developed an applicable model of the insulation material considering the moisture absorption and desorption process. Sari et al. [25] developed a predictive model utilizing machine learning (ML) techniques that considered energy efficiency, indoor environmental quality, and site planning to design buildings. Yussuf et al. [26] explored using AI to enhance energy efficiency throughout various stages of the building lifecycle, including building design, construction, operation and control, maintenance, and retrofit. Alsalemi et al. [27] introduced an economical, high-performance Internet of Energy platform that collects, measures, and processes data on energy usage, temperature, illuminance, humidity, and occupancy within a space to improve occupants' consumption behaviour. Various researchers developed predictive models for heating or cooling performance to achieve the effective system [28–30]. Zivelonghi & Giuseppi found the idea of Smart Healthy Schools (SHS) as a paradigm in education buildings that merge together IAQ, IoT (Internet of Things) and AI to provide optimal control of the IAQ [31]. Tagliabue et al [32] proposed using IAQ data gathered by IoT sensors to activate the control of the indoor conditions according to the occupancy rate in the educational building located in the Smart Campus of the University of Brescia. Since experimental monitoring is not always possible, over the last few years, many artificial intelligence applications have been introduced in energy systems [33]. Various other studies focused on using Artificial Neural Networks (ANN) and IoT to reduce energy consumption and improve energy management in buildings e.g. Chou&Truong [34], A. Thangamani et al. [35], Selvaraj et al [36], Ngarambe et al. [37] proposed AI as a tool for predicting thermal comfort in buildings. Merabet et al. [38] showed known applications of ANN and paid attention to the necessity of future studies regarding the use of AI for human comfort and energy-efficiency management in buildings. Zhang et al. [39] proposed an explainable Artificial Intelligence model to predict energy usage and greenhouse gas emissions of residential buildings. Model predictive control (MPC) was the subject of an investigation by Yao et al. [40] and by Afram and Janabi-Sharifi [41], which showed various kinds of modelling techniques and optimization methods used.

However, a search in scientific databases has shown a limited number of studies related to the use of Artificial Neural Network (ANN) techniques in the field of carbon dioxide concentration. Ahn et al. [42] developed a hybrid model for forecasting the varying indoor CO_2 concentration levels in a residential building by controlling the ventilation rates of a heat recovery system. Baghoolizadeh et al. proposed models for CO_2 emissions developed using the GMDH type of artificial neural network (ANN-GMDH) in different studies. In [43] they used five design variables (cooling and heating set point temperature of the air conditioning thermostat, the level of the residents' clothes for the hot and cold seasons and the amount of fresh air transferred into the building by the ventilation system). In [44] they introduced twelve variables (the level of occupants' clothing, the air velocity, the set-point temperature of the cooling and heating, the volume of fresh air transported into the building, total heat transfer coefficient (U), thermal absorbance, visible absorbance and solar absorbance for roof and external wall, window optical properties and occupant's activity level). Most existing studies have used readings from CO_2 sensors to predict IAQ, while others developed CO_2 models based on several variables connected with building parameters, HVAC systems, etc. Moreover, in the existing literature, the non-replicability to different locations or contexts often arises as a limitation of existing studies.

In contrast, this current work focuses on estimations of indoor CO_2 levels, even in the absence of a direct CO_2 sensor, based on easily obtained indoor parameters, to develop a universal model, easy to use in a wide range of applications. The affordable microclimate station that employs Arduino-like Wi-Fi microcontrollers to monitor environmental parameters, including temperature, humidity, human presence, atmospheric pressure, and carbon dioxide levels was proposed. Predictive models are employed to estimate CO_2 concentration: both linear and non-linear ones, including machine learning models such as SVM (Support Vector Machine), random forest, decision tree, and neural network. The analysis and modelling, presented in Section 4, are conducted in a Matlab environment [45] using the Statistics and Machine

Learning Toolbox [46]. The results are compared with experimental data, to assess the quality of the research. Results of such predictive models, based on easily known values such as temperature, humidity and people presence in rooms that make them general-purpose, easily replicable, and relatively cheap tools, are crucial for a proper ventilation profile of rooms to maintain an optimal IAQ.

2. Experimental campaign

First, let us describe the experimental campaign performed both to collect data and to double check the predicted results. We equipped a school class with measurement sensors placed in two different boxes. Each box contains two microcontrollers (Arduino Uno [47] and NodeMCU [48]) and some sensors, namely two NTCs, a DHT11 humidity sensor, and a BMP280 pressure and temperature sensor. In this configuration, the Arduino Uno is responsible for reading the temperature values from the NTCs through its internal 10-bit ADC (analogue to digital converter) and the measurements from the DHT11 and BMP280. The NodeMCU, on the other hand, saves the data obtained by Arduino Uno via serial connection to an onboard SD card (connected to the NodeMCU via a logic level converter) and transmits the collected data via Wi-Fi.

Fig. 1 shows the schematic of the implemented data acquisition system. The two boxes, to which the sensors are connected, locally save the data collected from the sensors every 60 s on an SD card (as a backup measure) and simultaneously transmit the data via Wi-Fi to a Google Sheet [49] using a Google Script created for this purpose.

The system is portable due to its compact size and USB power supply. Moreover, the system is low-cost. Details regarding the composition of the boxes and their positioning inside a classroom are provided in Appendix 1. The installed sensors, the variables measured by each, and the accuracies provided by datasheets or obtained following laboratory calibration are listed in Table 1.

The research was conducted in an educational building (primary school) in Bialystok, Poland, with the following geographic coordinates: longitude 23°15′23″E and latitude 53°11′83″N. A detailed description of the building and the local weather and climate conditions is provided in Appendix 2.

Table 1
Data about the sensors employed.

Sensors	Measured values	Range	Accuracy
NTC (negative temperature coefficient) [50]	temperature	0–27 °C	±0.4 K
DHT11 [51]	humidity	20–90 %	± 5 %
BMP280 [52]	temperature	−10–85 °C	± 0.5 K
	pressure	300–1100 hPa	± 1.0 hPa
PIR 503 [53]	presence (−)	–	–
CO2 sensor [54]	temperature	−10–60 °C	± 0.5 K
	humidity	0–95 %	± 3 %
	CO_2 concentration	0–10000 ppm	±50 ppm + 2 % of the measured value

The experimental campaign covered part of the winter season, from January 26, 2024, when the sensors were installed, to March 12, 2024.

3. Experimental results

Let us first discuss the trend of carbon dioxide emissions registered during the experimental campaign.

In Fig. 2(a), the trend of carbon dioxide concentration during the staff in-service days is reported, i.e., corresponding to a period when no children are present in the school, only the staff. This period extends from the installation of the sensors on January 26 to February 4, 2024, aligning with a frame of the typical Polish winter vacation. The figure shows that on Saturdays and Sundays (January 27–28 and February 3–4), the concentration detected by the sensor is below 500 ppm. However, on the other days within the considered timeframe, despite the lack of continuous presence in the classroom and with the classroom's access door from the main corridor being closed, the carbon dioxide concentration increases due to leakages from the rise of CO_2 in other rooms and hence in the corridor due to the presence of the staff. A peak concentration is observed in the evening, with values exceeding 510 ppm, whilst a minimum is noted early in the morning between 6:00 a.m. and 7:00 a.m. due to the absence of people in the building

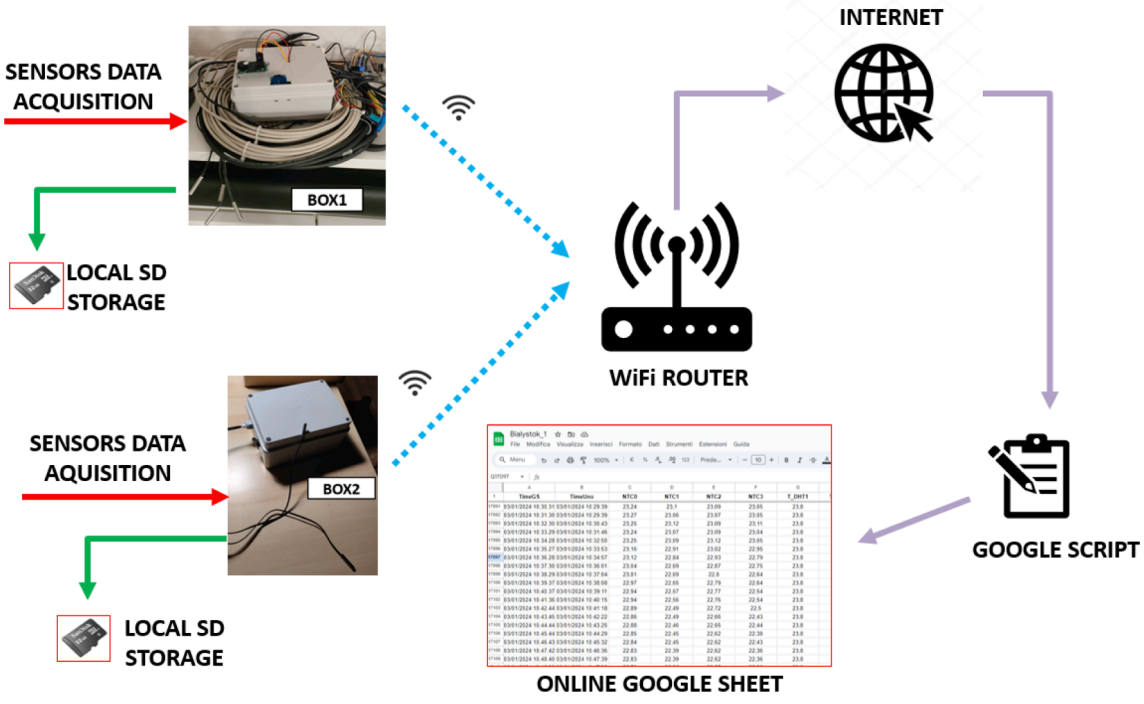

Fig. 1. Scheme of the implemented acquisition devices.

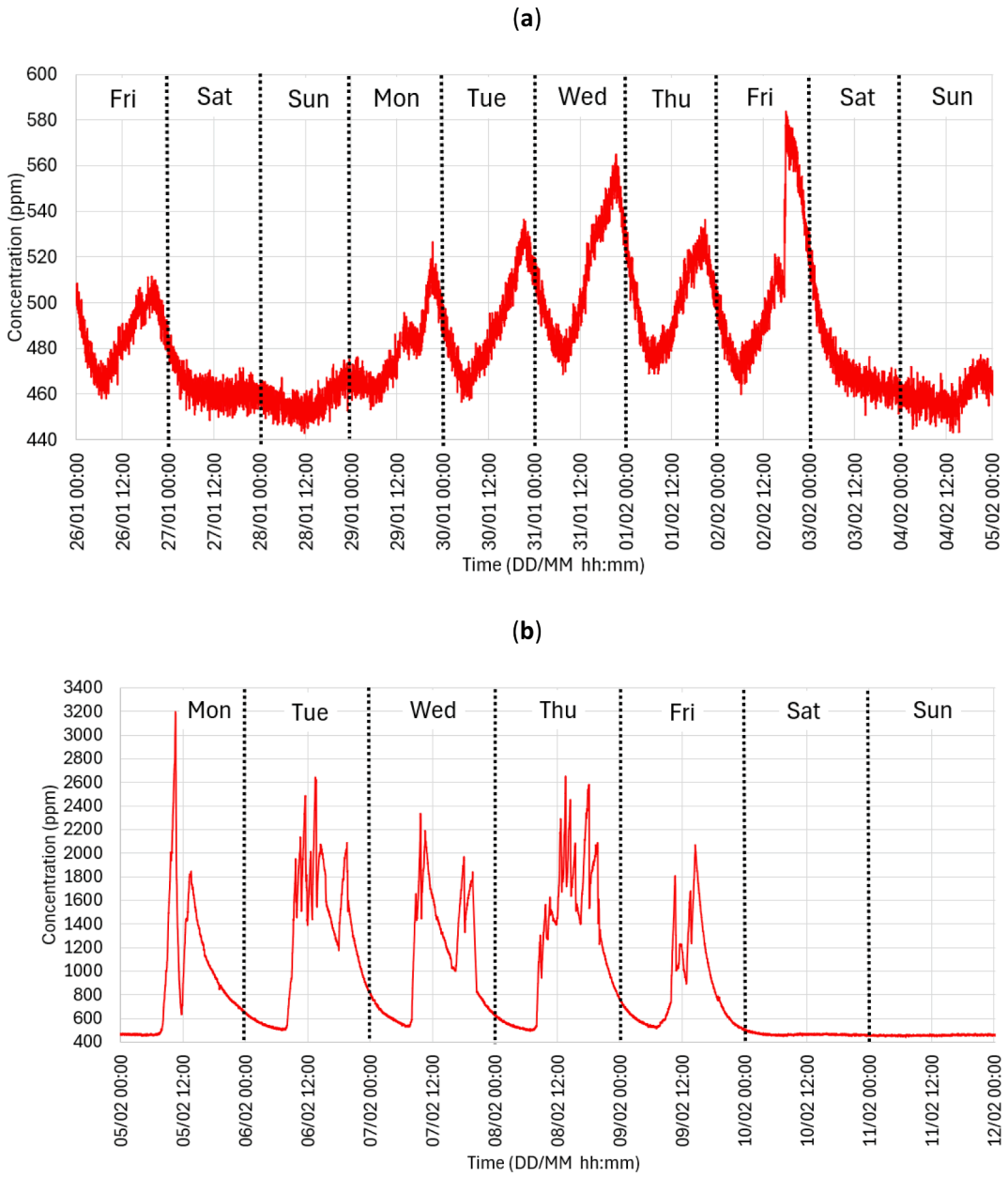

Fig. 2. Carbon dioxide concentration during staff in-service days (**a**) and during a typical working week (**b**).

overnight. Finally, the entrance of staff members on Friday, February 2, is observed, likely for the classroom preparation for the following week's lessons: during the afternoon of February 2, there is an increase from 510 ppm to 580 ppm in a few minutes. In Fig. 2(b), the detected CO_2 trend during a typical working week, when classes are held in the analyzed classroom, is shown. Fig. 2(b) displays a characteristic trend from Monday to Friday, with concentration peaks during the lessons up to 3200 ppm. Different relative maxima and minima are visible each day, mainly due to opening the windows for room ventilation purposes. The starting CO_2 value in the morning is, nonetheless, close to 500 ppm.

Moreover, in Fig. 3, the minute-by-minute CO_2 trend on February 7th is displayed alongside the presence of people within the room detected by the presence sensor. Notably, high concentration values are observed in the morning from 8:00 a.m. to 11:00 a.m. and, in the afternoon, from 4:30 p.m. to 8:00 p.m., which corresponds to the presence of people in the classroom. In fact, the presence of people moving in the classroom is detected by the PIR sensor and shown with a yellow band in Fig. 3. Additionally, a decrease from about 2100 ppm to 1000 ppm in 5 h is noted during the central hours of the day when the room is unoccupied, and the windows are closed. Fig. 3 also clearly shows the drastic reductions in CO_2 levels achieved when the classroom windows are opened. Nevertheless, the carbon dioxide values are exceedingly high throughout most of the working day, indicating poor air quality.

Fig. 4(a) shows the trends of humidity inside the classroom during a period without lessons, i.e., during the in-staff days, in the period between the end of January and the beginning of February, while Fig. 4(b) represents the trend of relative humidity detected during a typical week when lessons are held inside the classroom.

In detail, Fig. 4(a) shows a stable humidity trend for the entire period except for Friday afternoon, corresponding to the entry of staff members into the classroom to prepare it for the following week's lessons: the presence of people is associated with a significant increase in humidity from values around 35 % to 45 % (this is probably due to endogenous activities that involve a significant increase in relative humidity, such as floor washing). A humidity value close to 35 % is observed for the rest of the reported period.

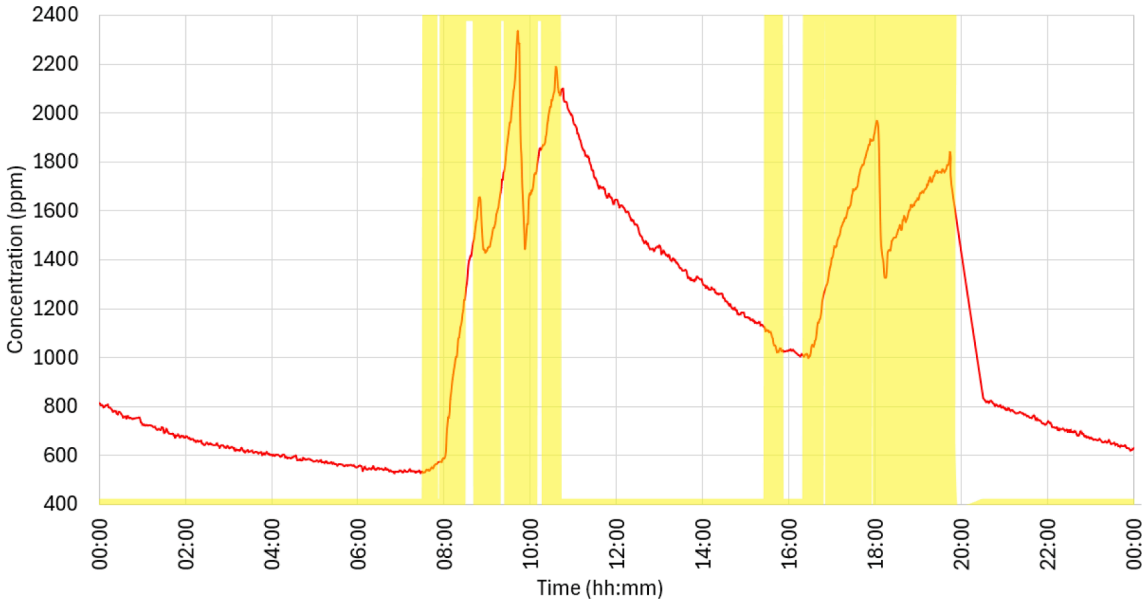

Fig. 3. Carbon dioxide concentration February 7th, a typical working day; the yellow area represents the presence of people moving in the classroom detected by the PIR sensor. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

On the other hand, Fig. 4(b) shows a vastly different humidity trend. Every day (from Monday to Friday, when lessons are held) features a highly variable humidity trend with peaks reaching 50 %. These increases are due to the presence of people. The humidity trend in the classroom during the day shows maxima and minima, and the latter is essentially due to windows opening for air exchange. Finally, a stabilization of the relative humidity value (approximately 30 %) is observed during the weekend when the classroom is not occupied.

Fig. 5 shows the temperature of the air and the mean radiant temperature. The air temperature is measured at 1.20 m from the ground, and the data are collected every minute; the mean radiant temperature is calculated using Eq. (1) [55], which is usually employed to correlate the black-globe temperature (the temperature measured by a sensor placed inside a black hollow sphere with known emissivity) and the air temperature to the mean radiant temperature under very low air velocity conditions.

$$t_{mr} = \left(\left(t_{bg} + 273\right)^4 + \frac{0.25 \cdot 10^8}{\epsilon} \left(\frac{\left| t_{bg} - t_{air} \right|}{D} \right)^{0.25} \cdot \left(t_{bg} - t_{air}\right) \right)^{0.25} - 273 \tag{1}$$

In **Eq.** (1), D represents the diameter of the sphere, which is 0.15 m in this case, ε represents the emissivity of the sphere considered equal to 0.95, t_{bg} is the temperature detected inside the globe thermometer (placed at 1.9 m from the ground). In comparison, t_{air} is the average air temperature from the two NTCs at 2.75 m and 1.20 m. It was decided to use an intermediate air temperature to estimate the mean radiant temperature at the same height as the globe thermometer (1.90 m) due to the impossibility of placing an additional temperature sensor near the globe. As is shown in Fig. 6, stratification in the room is limited between 1.20 m and 2.75 m from the floor (usually less than 0.5 K), so this choice should have a limited impact on the calculated mean radiant temperature value.

Fig. 5(a), refers to the period when only the staff was present inside the school and there were no lessons in the considered classroom; it shows up until February 2 an air temperature and a mean radiant temperature below 18 °C, compatible with a reduction of the set-point inside the classroom. On February 2, however, an increase in temperature was observed with values close to 19 °C for both air temperature and mean radiant temperature: this is compatible with the entry of staff members into the classroom to prepare it for future lessons and with a possible increase in the set-point. Focusing on Fig. 5(b), an increase in average air and radiant temperatures is observed with maximum values of 23.5 °C and minimum values of 18 °C. The values of air temperature and mean radiant temperature are also similar during a typical week of lessons, with the mean radiant temperature generally higher than the air temperature but with a limited difference, less than 0.5 K, which is also attributable to the different mounting heights of the black-globe sensor (1.90 m from the floor) compared to the NTC at 1.20 m from the floor. Moreover, during Saturday and Sunday, a progressive lowering of both temperatures is observed with a stabilization around 20.5 °C, indicating that probably during the weekend, the heating system's set-point is above 20 °C.

In Fig. 6, the temperature trends detected in the classroom on February 20, a typical working day, are shown. The temperature at different heights from the floor (0.05 m, 1.20 m, and 2.75 m), the mean radiant temperature, the temperature near the window, and the temperature detected by the sensor placed outside the building are reported. A stratification between the floor and the ceiling of 1 K is observed during the nighttime hours (20 °C at a floor level and 21 °C at a ceiling level), while the mean radiant temperature and the temperature measured at 1.20 m from the floor are found to be between the two aforementioned values. Significant fluctuations in temperatures near the floor (black line) are observed, especially following windows opening for air exchange and the temperature near the window (minimum temperature reached 16 °C). The temperature near the window is also lower than the temperature detected at floor level during the night, while it shows considerable fluctuations, up to 24.5 °C due to the position of the NTC: it is placed near a window and above a radiator. In this position, the temperature during the day shows these peaks due to external irradiation and the operation of the radiator. The outside temperature, on the other hand, is found to be between 1.5 and 9 °C.

Finally, in Fig. 7, the daily mean values of CO_2 throughout the entire day and during the typical 12 h of classroom occupation are reported. An occupancy factor for the classroom is also presented, where a value equal to 1 corresponds to occupancy for the 12 h of the day. It is observed that the CO_2 values during the 12 h of occupation are much higher than those averaging over 24 h; the values for the 12 h on working days generally range between 1000 and 2000 ppm. Furthermore, a correlation between the presence factor and the average measured CO_2 is noted. However, this correlation is not always visible:

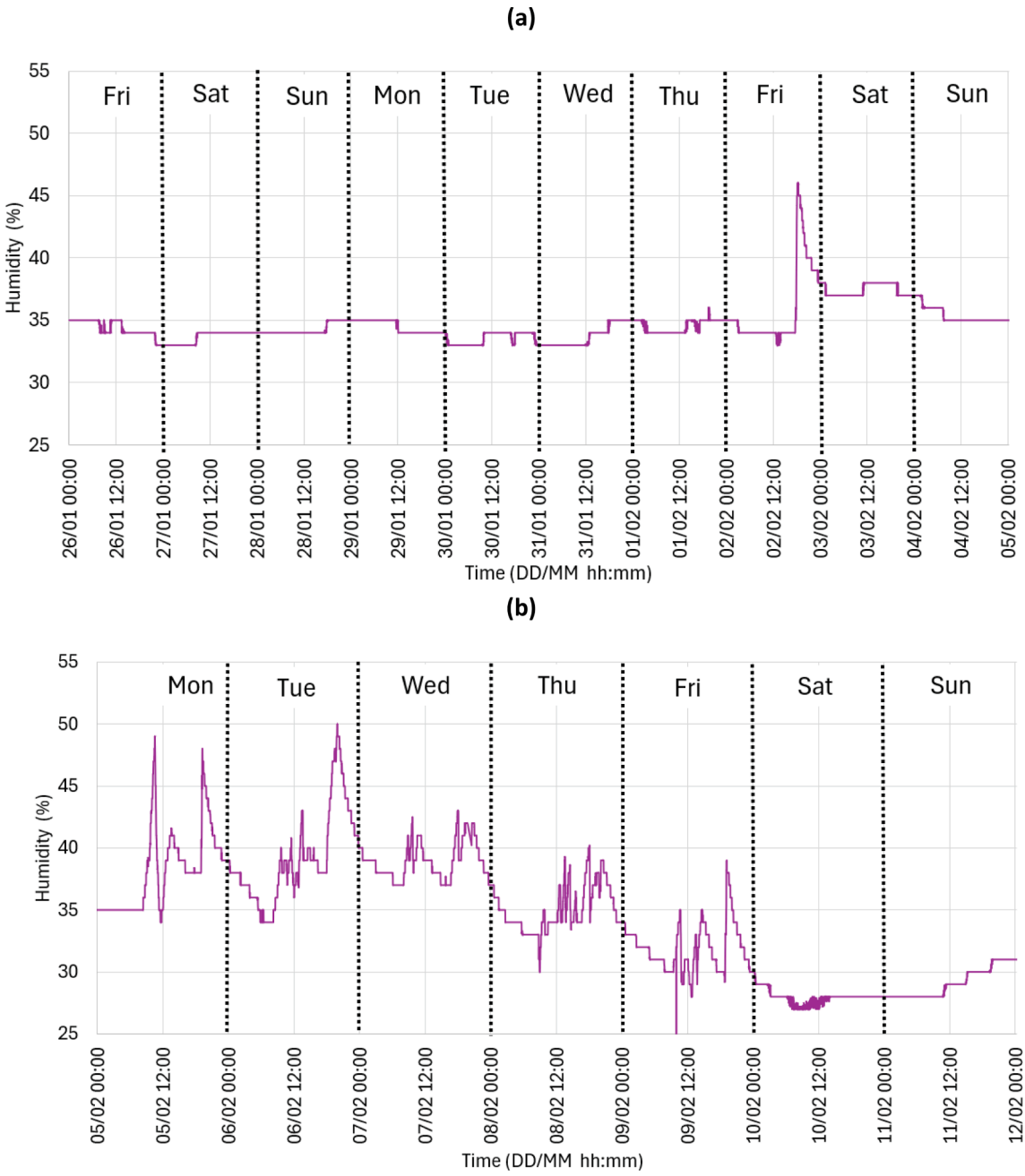

Fig. 4. Relative humidity during staff in-service days (**a**) and during a typical working week when lessons are held in the classroom (**b**).

for example, considering February 21 and February 22, there is a quite different occupancy factor, but a very similar average daily CO_2 value. Finally, on Saturdays and Sundays and generally, when the classroom is not occupied, the averages over 12 h and 24 h are similar; it is also noted that on February 24, the classroom was occupied for a total period of fewer than two hours.

4. Estimation of CO_2 concentration in the environment using predictive methods

This section will present some predictive methods for determining carbon dioxide levels without carbon dioxide meters. These methods estimate CO_2 based on environmental data collected by other sensors located within the classroom and then compared with the experimental data. Firstly, before introducing the predictive methods, let us discuss some qualitative correlations between the measured carbon dioxide concentration and the other measured values (temperature, humidity, and presence) will be presented.

4.1. *Instantaneous CO_2 concentration estimation*

In Fig. 8(a), the trends of carbon dioxide concentration, humidity, and presence are shown on the same graph, and in Fig. 8(b), the trends of CO_2 and the temperature difference *DT* between the NTC located near the floor and at 1.20 m from the floor are displayed. From the mentioned graphs, a correlation between the CO_2 trend and the detected relative humidity value in the classroom is visible: the presence of people inside the room leads to an increase in CO_2 and also an increase in humidity due to respiration and perspiration. Moreover, it is observed that the CO_2 trend is also influenced by the presence of people in the classroom. From Fig. 8(b), a peculiar trend between *DT* and CO_2 concentration is noted: as *DT* increases, indicating the opening of windows for

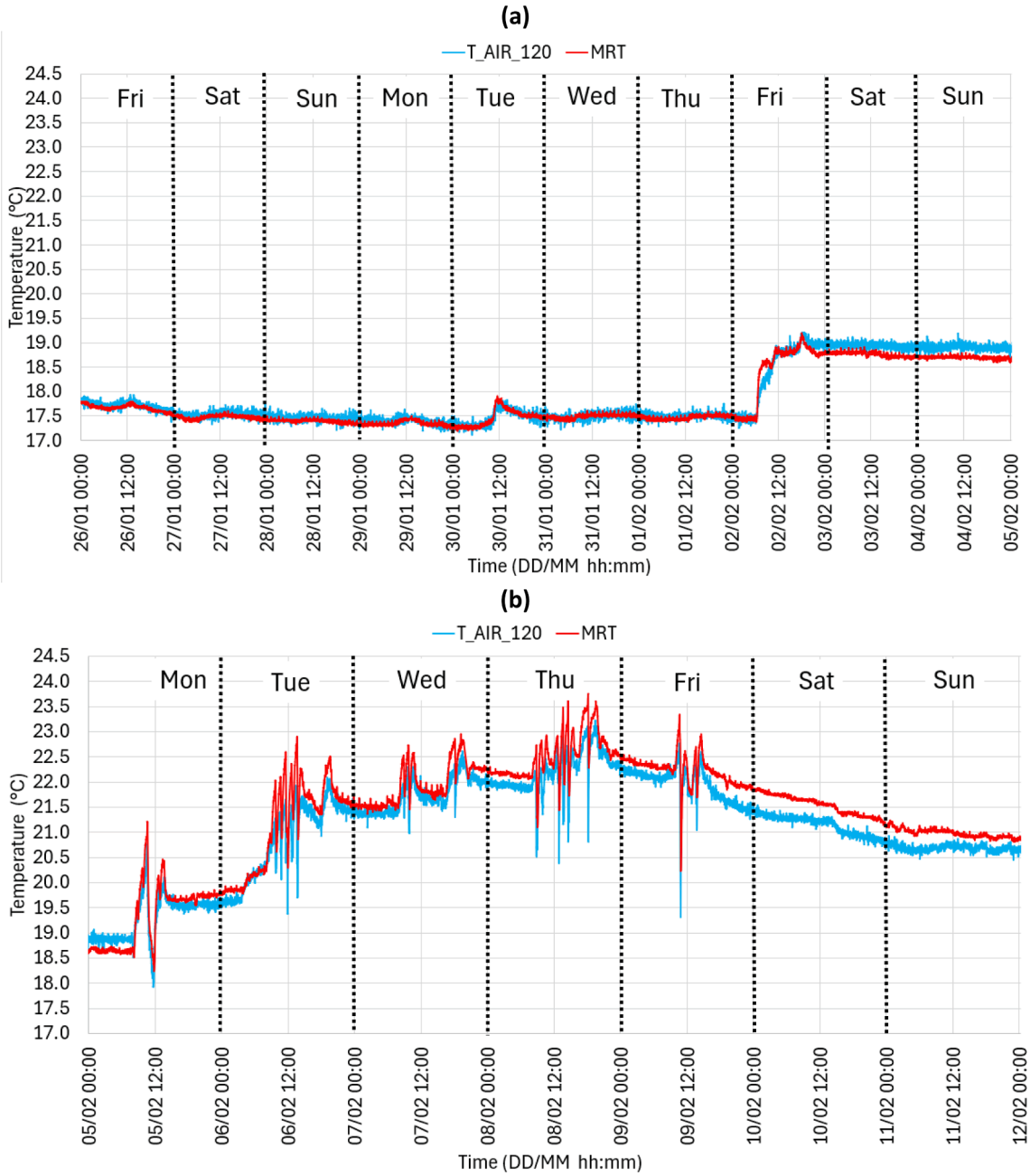

Fig. 5. Air temperature at 1.20 m from the floor "T_AIR_120" and mean radiant temperature "MRT" determined during staff in-service days (**a**) and during a typical working week when lessons are held in the classroom (**b**).

ventilation, there is a corresponding decrease in CO_2 concentration.

Therefore, we developed predictive methods for assessing the CO_2 concentration without a CO_2 measuring sensor based on detected temperature, humidity, and presence values. If the predictive model is sufficiently accurate, it could be easily replicable and lead to a reduction in hardware costs: a device consisting of a microcontroller paired with a CO_2 sensor indeed has a higher cost compared to an acquisition system that includes a microcontroller, two temperature sensors, one humidity sensor, and a presence sensor.

The predictive models for CO_2 determination are based on variables that utilize data of temperature at 0.05 m from the floor ($t_{5,i}$), temperature at 1.20 m from the floor ($t_{120,i}$), humidity (RH_i), and presence ($PRES_i$), where the notation i represents the i-th measurement, which occurs at intervals of about 60 s as indicated. The new variables introduced based on the measured data are indicated in equations from (2) to (6):

$$x_{P,i} = \sum_{k=i-50}^{i} \frac{PRES_k}{50} \tag{2}$$

$$x_{H,i} = RH_i \tag{3}$$

$$x_{T,i} = t_{120,i} - t_{5,i} \tag{4}$$

$$x_{DT,i} = x_{T,i} \sum_{k=i-16}^{i} \frac{1}{(t_{5,k}/16)\cdot\Delta\tau_k} \tag{5}$$

$$x_{DH,i} = \sum_{k=i-60}^{i} \left(\frac{1}{x_{H,k}/60}\right) \sum_{k=i-5}^{i} \frac{x_{H,k}}{5\cdot\Delta\tau_k} \tag{6}$$

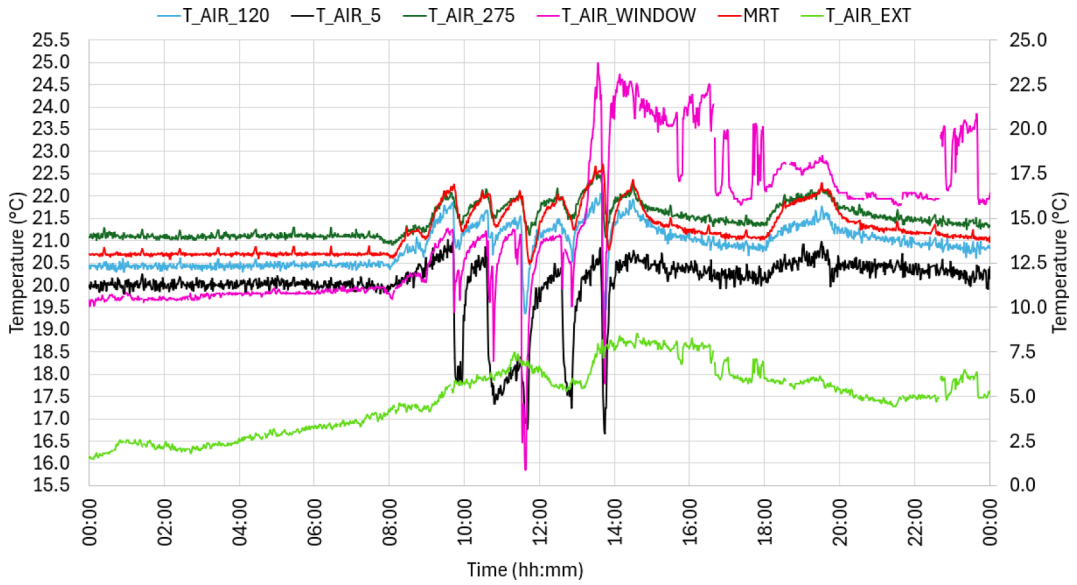

Fig. 6. Temperature trends during February 20 (typical working day) at different heights from the floor (0.05 m, 1.20 m and 2.75 m from the floor), mean radiant temperature and air temperature near the window "T_AIR_WINDOW"; On the secondary axis on the right is also reported the external air temperature "T_AIR_EXT".

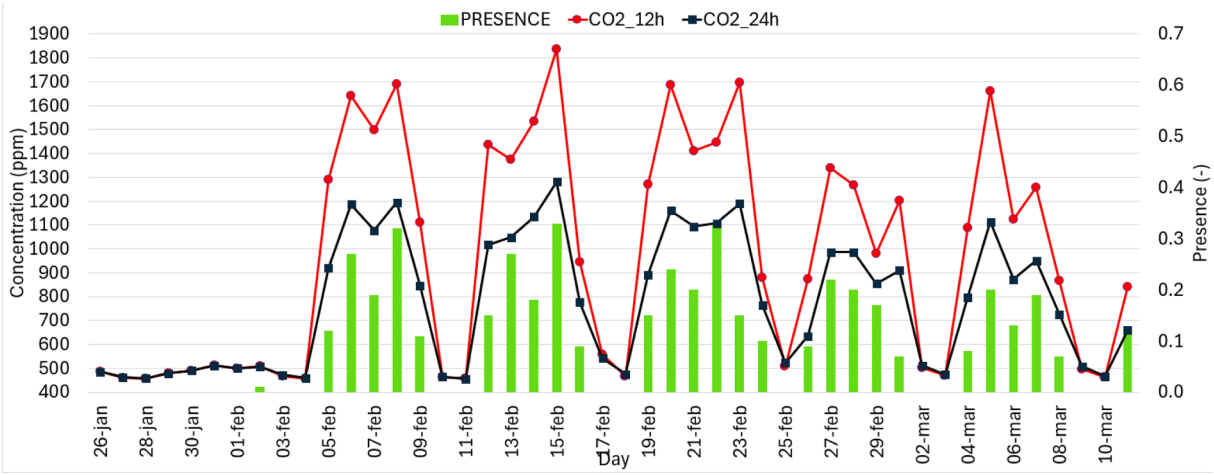

Fig. 7. Daily carbon dioxide values averaged over 24 and 12 h (8:00 a.m. – 8:00 p.m.) and daily occupancy factor.

In the aforementioned, $\Delta\tau_i$ represents the time difference in seconds between the *i*-th measurement and the previous one. The values of **eqs. (2)-(4)** represent respectively the average presence value in the previous 50 measurements (approximately in the previous 50 min), the relative humidity detected from the *i*-th measurement, and the temperature difference at 1.20 m from the floor and near the floor, respectively. The parameters are determined by eqs. (5) and (6), and consider the variation over time of the humidity and the temperature difference detected in the latest measurements (16 measurements for the temperature difference and 60 for the humidity).

To estimate the CO_2 concentration inside the classroom, six predictive methods were used that take as input the expressions reported in eqs. (2)-(6). The best fitting for each method was determined using MATLAB and the "Statistics and Machine Learning Toolbox" [45,46]. The first two methods analyzed were linear and non-linear regression models; the correlation coefficients and root mean square errors are reported in Table 2. These two models (indicated with identifiers ID1 and ID2) are the only ones with an analytical expression; the other regression models used are machine learning methods: SVM, random decision forests, decision trees, and neural networks. The main characteristics of the aforementioned predictive systems are as follows:

- SVM (Support Vector Machine): SVMs are models used for classification and regression. In the case of non-linearly separable data, SVMs use kernel functions to transform the input space into a higher-dimensional space where linear separation can be carried out [56].
- Decision trees are predictive models that map the features of objects to conclusions about target values. They use a tree structure with various levels, where the internal nodes represent tests on the features, branches represent the test outcomes, and each tree leaf represents a class or value [57].
- Random decision forests combine the predictions of numerous decision trees to improve generalization and reduce the risk of overfitting. Each tree in the forest is constructed from a random sample of the training data set, and the split points in each tree are chosen from a random subset of the features [58].
- Neural networks are systems of algorithms composed of nodes (neurons) organized into layers: an input layer, one or more hidden layers, and an output layer. They require substantial data for training and can be opaque regarding interpretability [59].

The main parameters characterizing the predictive methods indicated with ID 3–6 in Table 2 are as follows:

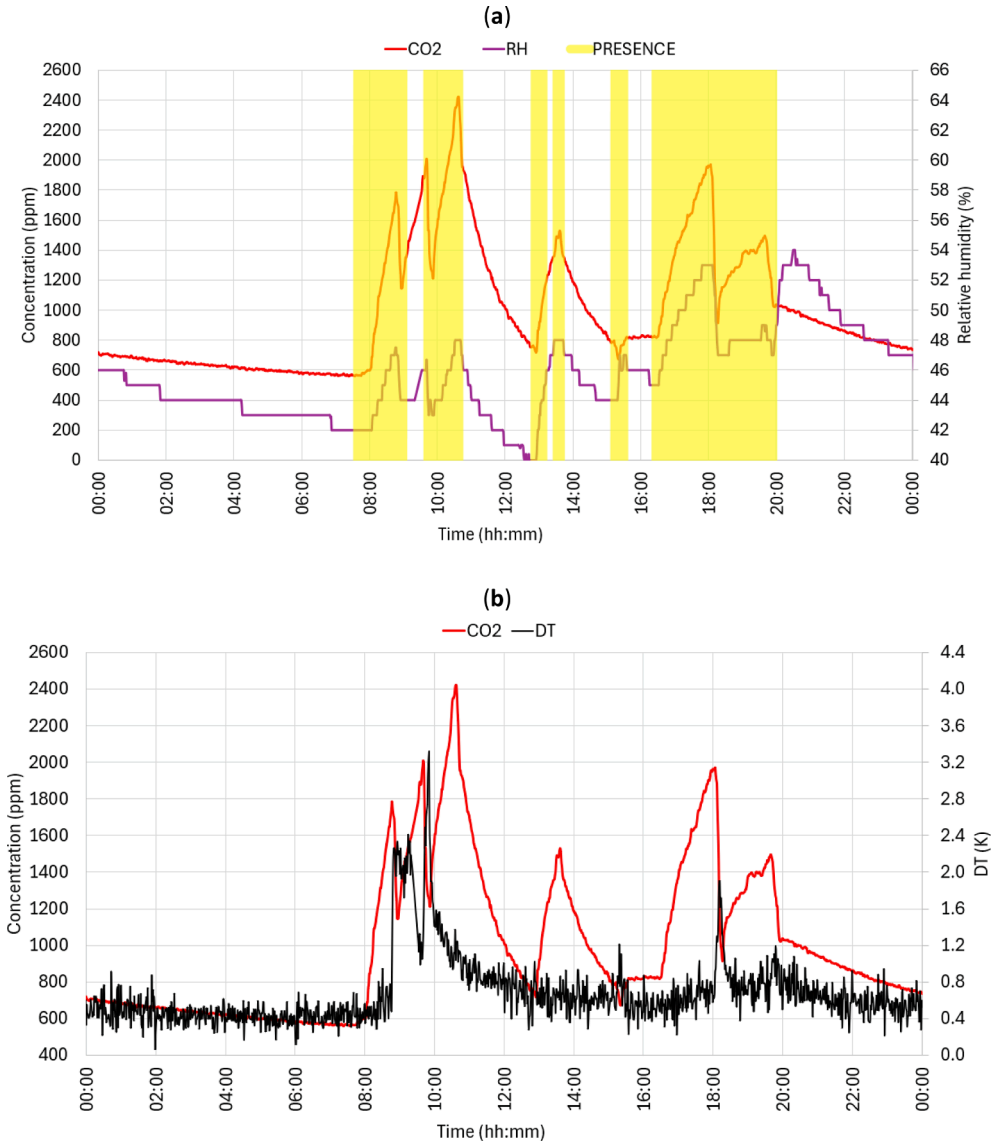

Fig. 8. Trends of carbon dioxide concentration compared to relative humidity and presence of people (yellow area represents the presence of people in the classroom) (**a**), and temperature difference *DT* between 1.20 m and 0.05 m from the floor (**b**) on February 28th (typical working day). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 2
Correlation coefficient and root mean squared error related to the six predictive methods analyzed. Note: [a)] refers to the case where the training and test sample is considered (100% of the measurements taken during the data acquisition period), while [b)] refers to the indices calculated only on the test data set (the last 30% of the dataset).

ID	Type	Expression	*RMSE* (ppm)	R^2
1	linear	$\beta_1 + x_{H,i}\,\beta_2 + x_{T,i}\,\beta_3 + x_{P,i}\,\beta_4$	241	0.732
2	non-linear	$\beta_1 + x_{H,i}\,\beta_2 + x_{T,i}\,\beta_3 + x_{P,i}\,\beta_4 + x_{T,i}x_{T,i}\beta_5 + \beta_6\,\exp(\beta_7\,x_{H,i} + \beta_8\,x_{T,i})$	231	0.753
3	SVM	–	190.7[a)]/ 195.7[b)]	0.830[a)]/ 0.797[b)]
4	Random forest	–	162.9 [a)]/ 141.7[b)]	0.878 [a)]/ 0.893[b)]
5	Decision tree	–	189.7 [a)]/ 183.5[b)]	0.830 [a)]/ 0.821[b)]
6	Neural network	–	197.8 [a)]/ 212.1[b)]	0.821 [a)]/ 0.761[b)]

- A Gaussian function was considered for the SVM model;
- Minimum leaf size of 25, to reduce the overfitting for the Decision Trees model;
- 100 trees for Random Decision Forest, to increase the accuracy and limit overfitting problems;
- 25 Hidden layers for neural networks.

Moreover, the methods were trained on 70 % of the available data (approximately 65,400 data for CO_2 values from experimental measures, and corresponding variables determined through **Eqs. (2)-(6)**) and tested on the entire data set and the remaining 30 % of the data, as indicated in Table 2.

Table 2 also reports the values of the root mean squared error (*RMSE*) and the coefficient of determination (R^2) obtained from the application of the six models, whilst Table 3 shows the values of the coefficients for linear regression (ID1) and nonlinear regression (ID2).

From the values reported in Table 2, it is observed that the R^2 value is 0.732 for the linear model, while it is 0.75 for the nonlinear model, with a limited percentage increase of the latter compared to the former (2.8 %): thus, these regression models lead to very similar estimates of CO_2

Table 3
Values of the coefficients indicated in Table 2 for the linear and non-linear methods.

	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
Linear	−303.68	24.606	84.729	1349	–	–	–	–
Non-linear	274.19	−10.556	200.96	1261.3	−60.952	87.029	0.052874	0.12496

concentration in the classroom, as also shown in Fig. 9(a) for March 7. This figure highlights the similar behaviour and values of the two predictions, expecially for the two peaks during the classroom occupation. However, both predictions tend to underestimate the overall distribution, mainly the higher peak.

Focusing on the neural predictive model and the decision tree, a remarkably similar determination coefficient value is observed on the entire data set (0.821 and 0.830, respectively); the values related to the two predictive methods are exposed in Fig. 9(b). In comparison to Fig. 9 (a), the neural network and the decision tree models fit the measured values better, especially for the occupancy period and the morning peaks due to the presence in the classroom. However, a slight underestimation is still present during the afternoon, when no occupancy together with some oscillations.

Finally, considering the SVM and random forest predictive methods, an R^2 value of 0.830 and 0.878 is observed on the entire data set, particularly 0.893 on the test set (including March 7, reported in Fig. 9). Moreover, observing the *RMSE*, it is lower for the random decision forest model (141.7 on the test set and 162 on the complete data set). Fig. 9(c) displays the predictive models for random forest and SVM. From the observed data, it is clear that the best correction is obtained with the random forest model compared to all others; it indeed has an increase in the determination coefficient of 22.0 % compared to the linear regression model (0.893 vs 0.732); similarly, a decrease in *RMSE* is observed compared to the linear case by 41.2 % (141.7 vs 241). The underestimation during the afternoon period of no occupation tends to reduce significantly.

Fig. 10 shows the prediction of the random forest model (ID4) compared to the base measured values. Thus, a correlation was found between the measured values (temperature difference, humidity, and presence of people) and the predicted CO_2 values in the classroom in the best model (ID4) with a determination coefficient R^2 of 0.893 and an *RMSE* of 141.7 ppm, against a range of measured CO_2 values between (480 ppm and 2830 ppm).

4.2. Daily CO_2 concentration estimation

In the previous section, estimates of CO_2 concentration in the classroom were determined with minute-by-minute granularity; these estimates have an 'operational purpose', meaning they can also be used to implement locally and in real-time control logics on a controlled mechanical ventilation system for air exchange in the environment. In this

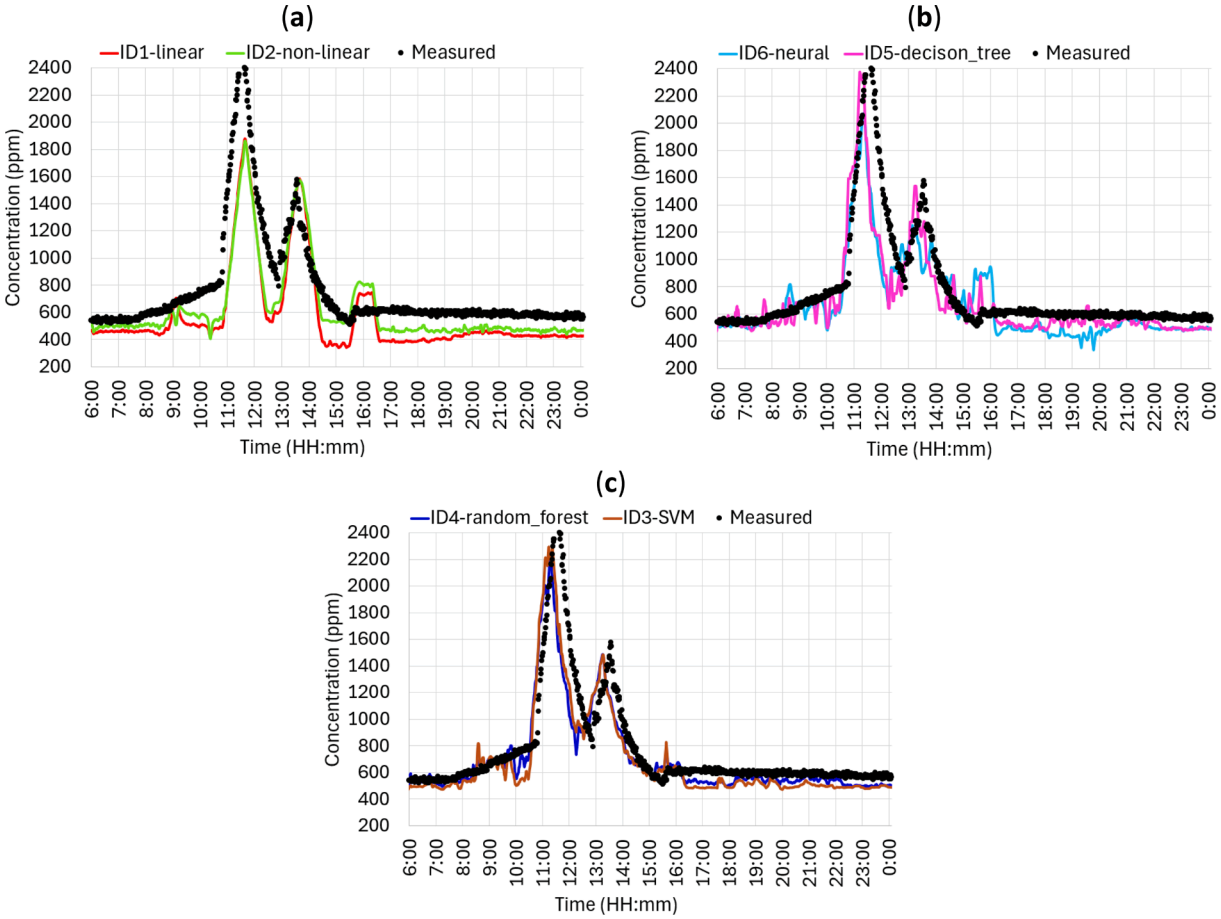

Fig. 9. Representation of the approximations through the predictive models previously described for a typical day of the heating season (March 7); in (**a**), the forecasts from the linear and non-linear predictive models are reported, in (**b**) the predicted values from the model based on neural networks and decision trees are shown, while in (**c**) the forecast models based on SVM and random forests are reported.

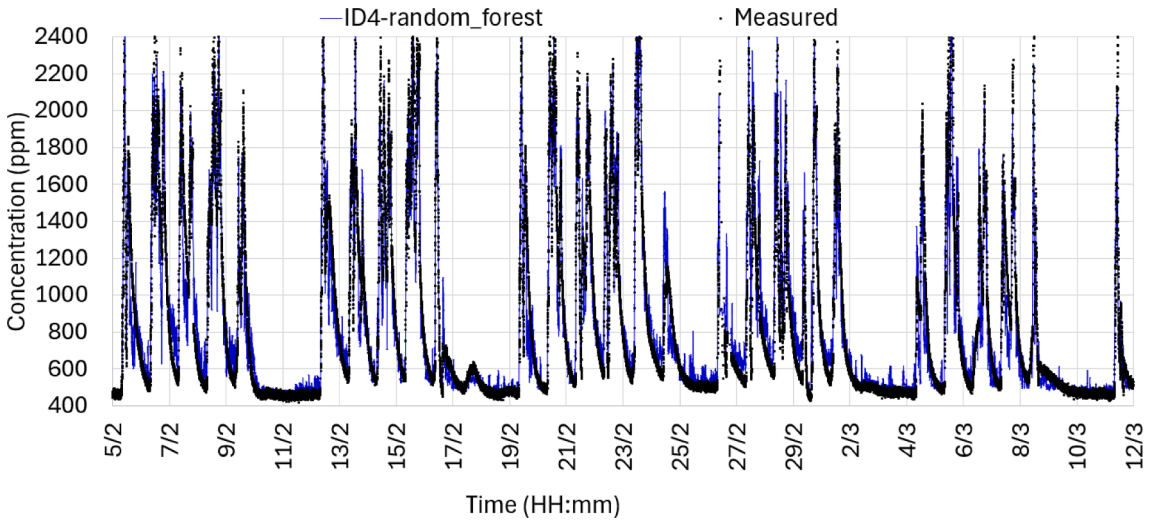

Fig. 10. Representation of the values predicted by the random forests' method compared to those measured between February 5 and March 12.

section, however, we will design and present a predictive model for analyzing CO_2 concentration in the classroom daily based on other measured parameters; thus, these estimates will have a purely analytical purpose and will be used for predicting daily carbon dioxide concentration to analyze air quality in the classroom. A linear correlation was determined between the CO_2 concentration during the typical occupancy hours of the classroom $y_{CO2,day,i}$ (generally 12 h during Monday to Friday, between 8:00 a.m. and 8:00p.m.) and the daily average temperature values at 1.20 m above the ground, $t_{120,day,i}$, at 0.05 m above the floor $t_{5,day,i}$ during the 12 h of occupancy, the presence factor over the 12 h, $p_{day,i}$, and the average relative humidity $RH_{day,i}$ over the 12 h:

$$y_{CO2,day,i} = \alpha_1 RH_{day,i} + \alpha_2 t_{120,day,i} + \alpha_3 t_{5,day,i} + \alpha_4 p_{day,i} + \alpha_5, \quad (7)$$

where the values of the coefficients are reported in Table 4.

This correlation shows an R^2 of 0.725 and an *RMSE* of 199 ppm; it can be observed that the value of the correlation index is close to that determined in the linear case with minute-by-minute granularity (ID1). Fig. 11 shows the measured daily concentration values and the values estimated with the linear predictive model for the actual occupancy days of the classroom.

5. Conclusion

The present study develops, compares, and assesses the accuracy of predictive models for estimating indoor carbon dioxide concentrations, even without a direct CO_2 sensor. An experimental campaign was conducted to determine environmental parameters such as temperature, humidity, human presence, atmospheric pressure, and carbon dioxide levels. For this purpose, homemade and low-cost microclimate stations that utilize Arduino-like Wi-Fi microcontrollers are installed in a primary school classroom without mechanical ventilation in Bialystok, Poland. All the stations were previously calibrated for accurate data collection. The recorded data, collected minute-by-minute, covers part of the winter season, from January 26 to March 12, 2024; lessons were suspended during the first period (1 week), and pupils started school on February 5.

All the collected data are used i) to develop and evaluate the accuracy of predictive models for estimating indoor CO_2 concentrations from previous carbon dioxide measurements and ii) to predict the daily behaviour of the carbon dioxide from the other measured parameters (temperature, relative humidity, and human presence). In the first case, linear and non-linear models were employed, including machine learning models such as SVM, random forest, decision tree, and neural network. The analysis and modelling are conducted in a Matlab environment using the Statistics and Machine Learning Toolbox. All cited methods are trained on 70 % of the available data, and the results are compared with the experimental data. In the last case, this is an affordable solution to reduce system cost setup through low-cost sensors.

To better emphasize the novelty of the paper, the following conclusions are drawn:

1. The random forest model yields the most accurate correction compared to all the other models. In fact, it is observed that the R^2 value for the random forest model is 0.87 and 0.830 for the SVM. Regarding the other employed models, the neural predictive model and the decision tree have similar R^2 coefficients: 0.821 and 0.830, respectively. R^2 is 0.732 for the linear model and 0.75 for the nonlinear model; moreover, these two regression models lead to very similar estimates of CO_2 concentration in the classroom;
2. Machine learning, compared to over deep learning, ensures that the model can run efficiently on some microcontrollers with limited resources. This approach demonstrates a replicable and cost-effective solution for environmental monitoring, validated through comparisons with actual CO_2 sensor data;
3. The combination of temperature, relative humidity and human presence measurements are suitable for predicting the daily carbon dioxide level inside a classroom. This leads to conclusions like those of employing dedicated CO_2 sensors. In this case, the installation costs are also significantly reduced.

A limitation of the present analysis may arise from the positioning of the sensor. Indeed, we aim to investigate the possible effect of the positioning of temperature sensors on the quality of prediction models in future works. Moreover, future aspects of this work will include installing a mechanical ventilation system in a classroom, which will work on the correlation obtained following a comfort analysis.

Table 4
Coefficients related to equation (7), a model for estimating daily concentration in the classroom.

α_1	α_2	α_3	α_4	α_5
16.158	−73.219	149.47	1305.7	−1276.2

Declaration of competing interest

The authors declare that they have no known competing financial

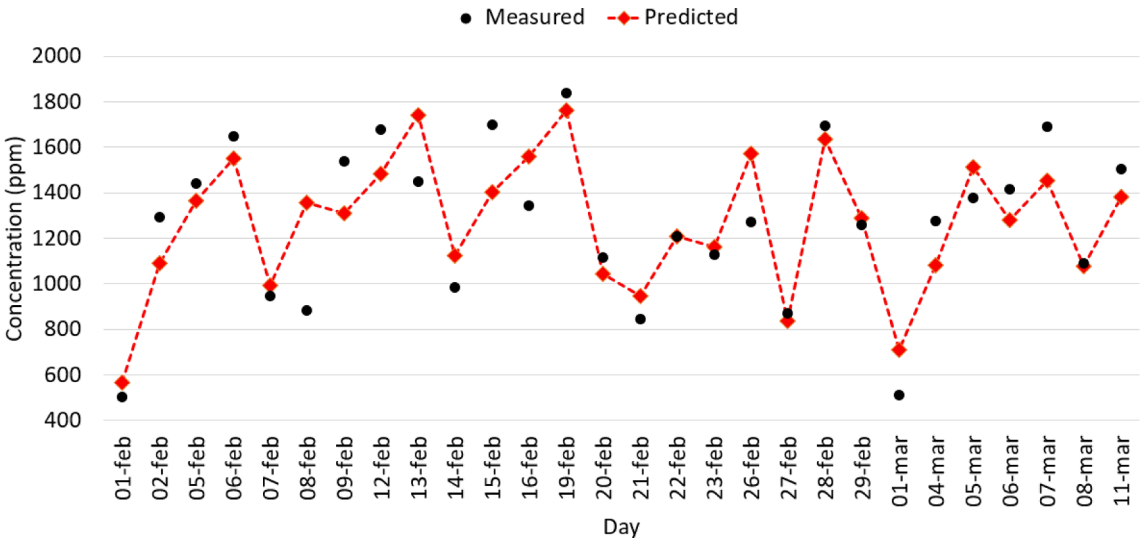

Fig. 11. Estimating daily CO_2 concentration: measured values and values predicted with the linear regression given by **Eq.** (7).

interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research was carried out as a part of work no. WZ/WB-IIL/2/2023 at the Bialystok University of Technology and was financed from the research subvention provided by the Minister responsible for science.

Appendix 1

In this appendix, details regarding the experimental equipment and the positioning of the sensors are provided.

Fig. A1.1 displays the interior of BOX2 (for brevity, the interior of BOX1, which is similar, is not shown). The figure shows the connections between the two microcontrollers (Arduino Uno [48] and NodeMCU [49]) and the sensors, namely two NTCs, a DHT11 humidity sensor, and a BMP280 pressure and temperature sensor. In this configuration, the Arduino Uno is responsible for reading the temperature values from the NTCs through its internal 10-bit ADC (analog to digital converter) and the measurements from the DHT11 and BMP280. The NodeMCU, on the other hand, handles saving the data obtained by Arduino Uno via serial connection to an onboard SD card (connected to the NodeMCU via a logic level converter) and transmitting the collected data via Wi-Fi. Fig. A1.1 shows the connections of the NTC and DHT11 sensors and the 5V USB power supply on the left. The system shown is, therefore, portable due to its compact size and USB power supply. Moreover, the system shown is low cost.

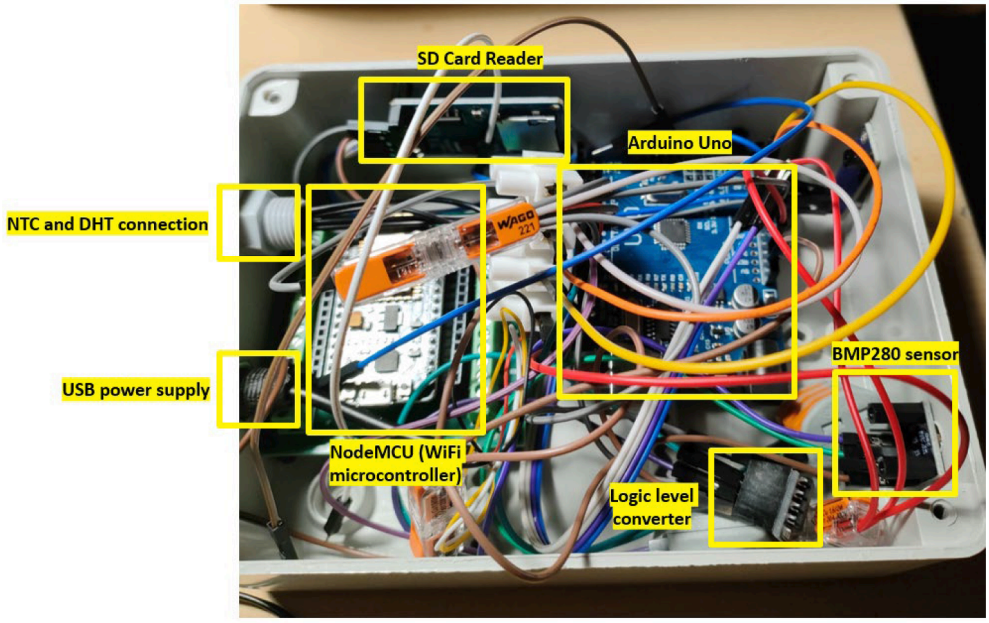

Fig. A1.1. Connections between sensors and microcontrollers in BOX2.

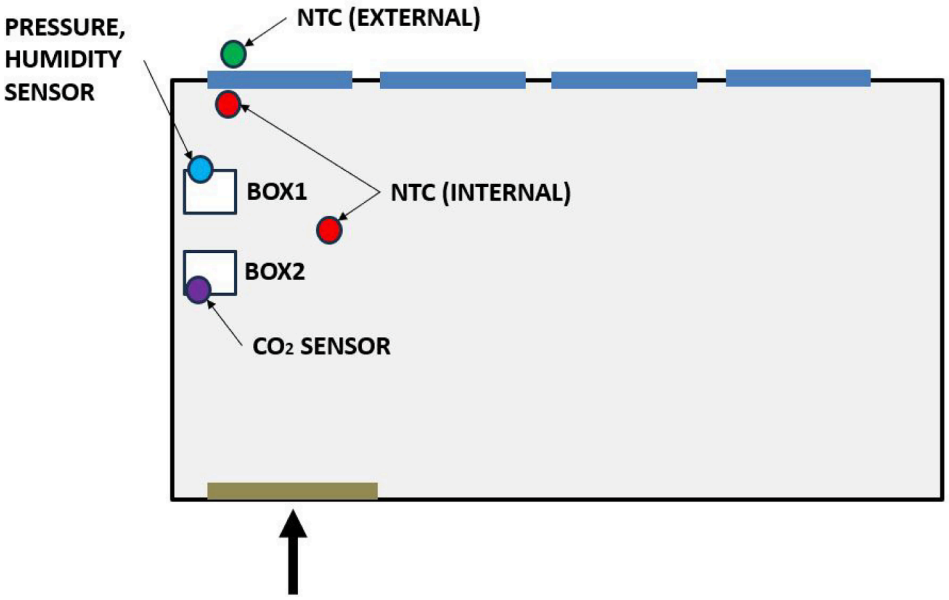

Fig. A1.2. Layout and position of the sensors within the classroom.

The sensors were installed in a south-facing classroom of the school. The room's layout is shown in Fig. A1.2, while pictures of the installed sensors are reported in Fig. A1.3. The small size of each box is evident. The arrow at the bottom highlights the entrance to the classroom.

Precisely, one NTC sensor was placed outside the classroom to measure the external temperature. The carbon dioxide sensor and the motion sensor were installed on a piece of furniture inside the classroom at a height of 1.90 m from the floor. Three NTC sensors were installed at different heights (indicated in Fig. A1.2 with a red dot between BOX1 and BOX2): specifically, the heights of the NTC are 0.05 m, 1.20 m, and 2.75 m from the floor; another box was placed near the window above the radiators (visible in Fig. A1.3(a)). For calculating the mean radiant temperature, an NTC was also placed inside a black globe with a diameter of 0.15 m, and then positioned at a height of 1.90 m from the floor.

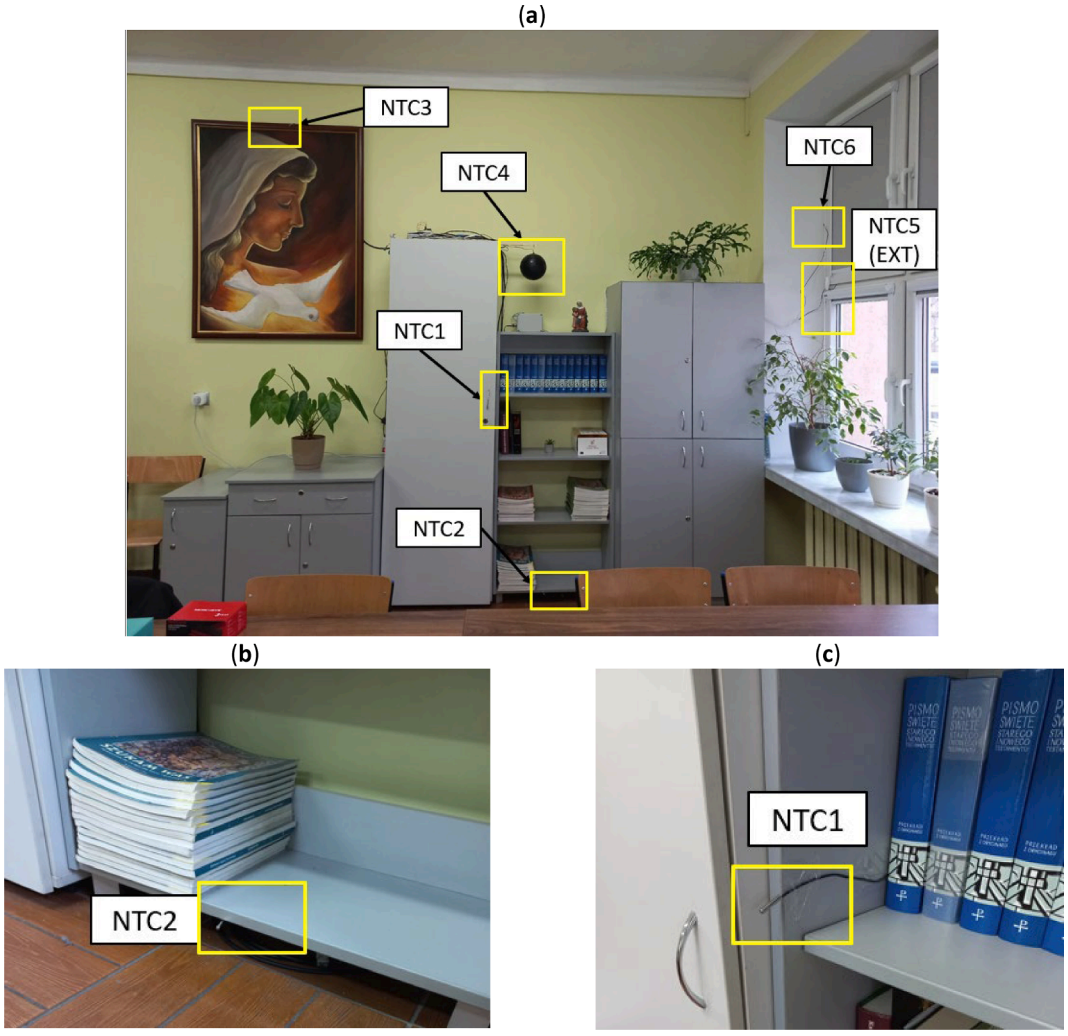

Fig. A1.3. View of the sensors positioned in the classroom (**a**); NTC placed next to the floor (**b**) and NTC placed at 1.20 m from the floor (**c**).

Appendix 2

In this appendix, details regarding the building subject to experimental activity and climate conditions are provided.

The research was carried out in an educational building located in Bialystok (a city in north-eastern Poland, which is the coldest part of Poland). The geographic coordinates are, respectively: longitude 23°15′23′′E and latitude 53°11′83′′N. The location of the building and its immediate surroundings are shown in Fig. A2.1 **(a)**. The classroom where the research equipment is placed is on the ground floor in the southeastern part of the building (marked in red in Fig. A2.1). The building was constructed in the 1950s. As shown, the north-west building side is connected to the adjacent gym building, constructed in later years than the education building.

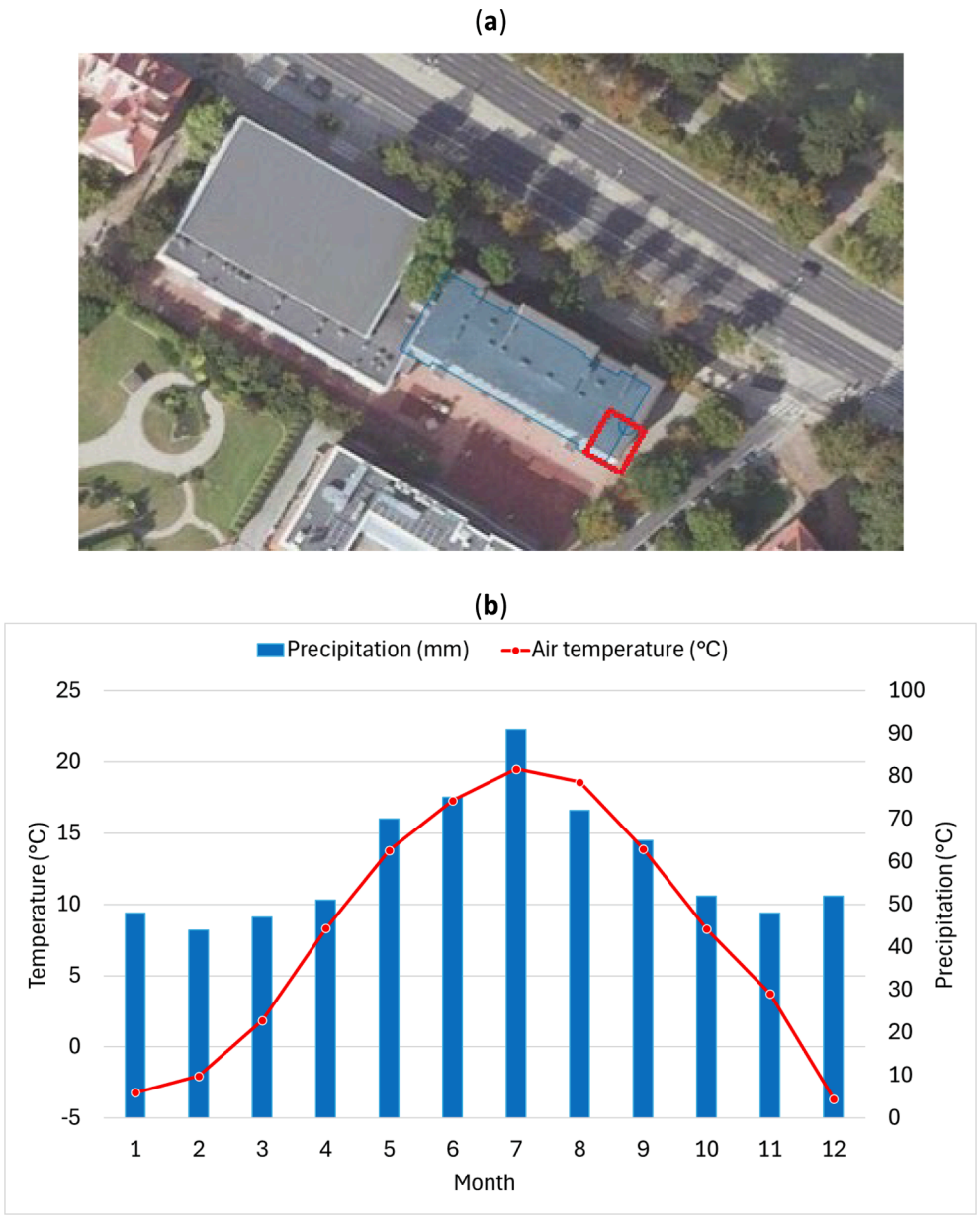

Fig. A2.1. Location of the educational building and the classroom where the research was carried out (**a**); Weather by month in Bialystok (1991–2021) [60] (**b**).

The building includes three overground floors and a basement. The useful floor area is 2,681.70 m^2 with a volume of 8,389 m^3. The building was built in a traditional style. The external walls are made of 58 cm thick ceramic brick. The roof is flat, partially unventilated (above the small gymnasium) and partially ventilated (above the rest of the building). In 2002, the building underwent thermal modernization, among others, the external walls were insulated with a layer of 14 cm thick Styrofoam. In general, the thermal transmittance coefficients of external partitions correspond to the requirements in force in Poland at that time ($U_{walls} = 0.23$ W/(m^2K)). The entire window and door joinery was replaced ($U_{windows} = 1.30$ and 1.50 W/(m^2K); $U_{doors} = 2.50$ W/(m^2K)). New windows with PVC frames are not equipped with air inlets. Fresh air is supplied into the building through the gaps obtained by thinning the rubber gaskets on the fragments of rebates of the window sashes to the frames. The building uses a two-pipe central heating system supplied with heat from the municipal heat network. There is no mechanical ventilation; only natural ventilation is used in all rooms.

Weather conditions in Bialystok are characterized by a moderate and chilly climate with a significant amount of rainfall during the year, even in the driest months. According to the Köppen-Geiger climate classification, this weather pattern is identified with the category of Dfb. According to meteorological records in Fig. A2.1 **(b)**, the average annual temperature in Bialystok is 8.2 °C, and the annual rainfall is 715 mm.

The lowest precipitation is recorded in February, only 44 mm, while the most precipitation falls in July – 91 mm. In terms of temperature, the warmest month of the year is July. The average temperature during this period reaches 19.5 °C. The lowest average temperature of the year is recorded in January and is −3.3 °C. It was observed that the difference in rainfall between the months with the lowest and highest levels of precipitation is 47 mm. Average temperatures vary by 22.8 °C throughout the year.

In Białystok, the highest number of daily hours of sunshine are recorded in June. There is an average of 11.01 h of sunshine per day and 330.2 h of sunshine throughout June. The lowest number of daily hours of sunshine is recorded in January. The number of hours of sunshine is 1.75 per day and 54.38 throughout the month, respectively [61].

More details about the climate conditions in Bialystok are given in [62]. During the measurement campaign, from January 26 to March 12, 2024,

the average temperature in Bialystok was 4.1 °C. The amount of precipitation for this period was 84.7 mm [63].

Data availability

Data will be made available on request.

References

[1] G. Chinazzo, R.K. Andersen, E. Azar, V.M. Barthelmes, C. Becchio, L. Belussi, S. Wei, Quality criteria for multi-domain studies in the indoor environment: Critical review towards research guidelines and recommendations, 109719, Building Environ. 226 (2022), https://doi.org/10.1016/j.buildenv.2022.109719.

[2] C. Heracleous, A. Michael, Experimental assessment of the impact of natural ventilation on indoor air quality and thermal comfort conditions of educational buildings in the Eastern Mediterranean region during the heating period, 100917, J. Build. Eng. 26 (2019), https://doi.org/10.1016/j.jobe.2019.100917.

[3] C. Ciacci, N. Banti, V. Di Naso, F. Bazzocchi, Green strategies for improving urban microclimate and air quality: a case study of an Italian industrial district and facility, 110762, Building Environ. 244 (2023), https://doi.org/10.1016/j.buildenv.2023.110762.

[4] H. Parhizkar, R.A. Khoraskani, M. Tahbaz, Double skin façade with Azolla; ventilation, indoor air quality and thermal performance assessment, J. Clean. Prod. 249 (2020) 119313, https://doi.org/10.1016/j.jclepro.2019.119313.

[5] C. Jung, J. Awad, Improving the IAQ for learning efficiency with indoor plants in university classrooms in Ajman, United Arab Emirates, Buildings 11 (2021) 289, https://doi.org/10.3390/buildings11070289.

[6] S.S. Korsavi, A. Montazami, D. Mumovic, Ventilation rates in naturally ventilated primary schools in the UK; Contextual, Occupant and Building-related (COB) factors, 107061, Build. Environ. 181 (2020), https://doi.org/10.1016/j.buildenv.2020.107061.

[7] R.A. Angelova, R. Velichkova, D. Markov, P. Stankov, The influence of the air temperature on the CO2 emissions by occupants indoors, IOP Conf. Ser.: Earth Environ. Sci. 952 (2022) 012012, doi: 10.1088/1755-1315/952/1/012012.

[8] C. Schreck, S. Rouchier, A. Foucquier, F. Machefert, E. Wurtz, In situ air change rate estimation from metabolic CO2 measurement. Summer experimental campaign in a single-family test house, 111646, Build. Environ. 259 (2024), https://doi.org/10.1016/j.buildenv.2024.111646.

[9] A. Labihi, Y. Benakcha, A. Meslem, P. Byrne, F. Collet, Improving the ventilation of a classroom to achieve heating energy savings and better indoor air quality, Build. Serv. Eng. Res. t. 45 (2024) 275–291, https://doi.org/10.1177/01436244241233756.

[10] L. Loreti, P. Valdiserri, M. Garai, Dynamic simulation on energy performance of a school, Energy Procedia 101 (2016) 1026–1033, https://doi.org/10.1016/j.egypro.2016.11.130.

[11] C. Ciacci, N. Banti, V. Di Naso, F. Bazzocchi, Evaluation of the cost-optimal method applied to existing schools considering PV system optimization, Energies 15 (2022) 611, https://doi.org/10.3390/en15020611.

[12] T.M. Lawrence, J.E. Braun, Evaluation of simplified models for predicting CO2 concentrations in small commercial buildings, Build. Environ. 41 (2006) 184–194, https://doi.org/10.1016/j.buildenv.2005.01.003.

[13] T. Lu, A. Knuutila, M. Viljanen, X. Lu, A novel methodology for estimating space air change rates and occupant CO2 generation rates from measurements in mechanically-ventilated buildings, Build. Environ. 45 (2010) 1161–1172, https://doi.org/10.1016/j.buildenv.2009.10.024.

[14] I. Škrjanc, B. Šubic, Control of indoor CO2 concentration based on a process model, Autom. Constr. 42 (2014) 122–126, https://doi.org/10.1016/j.autcon.2014.02.012.

[15] D.A. Krawczyk, A. Rodero, K. Gładyszewska-Fiedoruk, K., A. Gajewski, CO2 concentration in naturally ventilated classrooms located in different climates—measurements and simulations, Energy Build. 129 (2016), 491–498, Doi: 10.1016/j.enbuild.2016.08.003.

[16] D.A. Krawczyk, M. Żukowski, Experimental verification of the CO2 and temperature model, Int. J. Vent. 19 (2020) 127–140, https://doi.org/10.1080/14733315.2019.1592333.

[17] N. Ma, D. Aviv, H. Guo, W.W. Braham, Measuring the right factors: a review of variables and models for thermal comfort and indoor air quality, 110436, Renew. Sustain. Energy Rev. 135 (2021), https://doi.org/10.1016/j.rser.2020.110436.

[18] R.M. Almeida, M. Pinto, P.G. Pinho, L.T. de Lemos, Natural ventilation and indoor air quality in educational buildings: experimental assessment and improvement strategies, Energ. Effi. 10 (2017) 839–854, https://doi.org/10.1007/s12053-016-9485-0.

[19] I. Attar, N. Naili, N. Khalifa, M. Hazami, M. Lazaar, A. Farhat, Experimental study of an air conditioning system to control a greenhouse microclimate, Energy Convers. Manage. 79 (2014) 543–553, https://doi.org/10.1016/j.enconman.2013.12.023.

[20] R.A. González Rivero, L.E. Morera Hernández, O. Schalm, E. Hernández Rodríguez, D. Alejo Sánchez, M.C. Morales Pérez, ..., A. Martinez Laguardia, A low-cost calibration method for temperature, relative humidity, and carbon dioxide sensors used in air quality monitoring systems, Atmosphere, 14 (2023), 191, doi: 10.3390/atmos14020191.

[21] V. Ballerini, C. Biserni, G. Fabbri, P. Guidorzi, E.R. di Schio, P. Valdiserri, The use of arduino and PID control approach for the experimental setup of HVAC temperature testing, J. Robot. Control (JRC) 5 (2024) 482–489, https://doi.org/10.18196/jrc.v5i2.20915.

[22] J. Tryner, M. Phillips, C. Quinn, G. Neymark, A. Wilson, S.H. Jathar, J. Volckens, Design and testing of a low-cost sensor and sampling platform for indoor air quality, 108398, Build. Environ. 206 (2021), https://doi.org/10.1016/j.buildenv.2021.108398.

[23] J.P. Sá, M.C.M. Alvim-Ferraz, F.G. Martins, S.I. Sousa, Application of the low-cost sensing technology for indoor air quality monitoring: a review, 102551, Environ. Technol. Innov. 28 (2022), https://doi.org/10.1016/j.eti.2022.102551.

[24] K. Liu, H. Niu, Y. Wang, Moisture absorption and desorption characteristics and prediction model analysis of building thermal insulation materials, 123196, Appl. Therm. Eng. 248 (2024), https://doi.org/10.1016/j.applthermaleng.2024.123196.

[25] M. Sari, M.A. Berawi, T.Y. Zagloel, R.W. Triadji, Machine learning model for green building design prediction, IAES Int. J. Artif. Intell. (IJ-AI) 11 (2022) 10.11591/ijai.v11.i4.pp1525-1534.

[26] R.O. Yussuf, Omar S. Asfour, Applications of artificial intelligence for energy efficiency throughout the building lifecycle: an overview, Energy Build. 305 (2024) 113903, 10.1016/j.enbuild.2024.113903.

[27] A. Alsalemi, Y. Himeur, F. Bensaali, A. Amira Smart sensing and end-users' behavioral change in residential buildings: An edge-based internet of energy perspective IEEE Sens. J. 21 (2021), 27623-27631, 10.1109/JSEN.2021.3114333.

[28] S. Wu, Y. Sun, F. Wang, Z. Ma, R. Zhao, D. Huang, A prediction model of air-source heat pump system performance with frost-retarded heater, 123315, Appl. Therm. Eng. 248 (2024), https://doi.org/10.1016/j.applthermaleng.2024.123315.

[29] W. Lu, C. Ma, D. Liu, Y. Zhao, X. Ke, T. Zhou, A comprehensive heat transfer prediction model for tubular moving bed heat exchangers using CFD-DEM: validation and sensitivity analysis, 123072, Appl. Therm. Eng. 247 (2024), https://doi.org/10.1016/j.applthermaleng.2024.123072.

[30] L. Reichembach Pizzatto, C.A. Richter Nascimento, N. Mendes, An empirical model of a split-type inverter air conditioner for building energy simulation, 121714, Appl. Therm. Eng. 236 (2024), https://doi.org/10.1016/j.applthermaleng.2023.121714.

[31] A. Zivelonghi, A.ndro Giuseppi, Smart Healthy Schools: An IoT-enabled concept for multi-room dynamic air quality control, Internet of Things and Cyber-Physical Systems, Volume 4, 2024, Pages 24-31, ISSN 2667-3452, Doi: 10.1016/j.iotcps.2023.05.005.

[32] Lavinia Chiara Tagliabue, Fulvio Re Cecconi, Stefano Rinaldi, Angelo Luigi Camillo Ciribini, Data driven indoor air quality prediction in educational facilities based on IoT network, Energy and Buildings, Volume 236, 2021, 110782, ISSN 0378-7788, Doi: 10.1016/j.enbuild.2021.110782.

[33] S.A. Kalogirou, Applications of artificial neural networks in energy systems a review, Energy Convers. Manage. 40 (1999) 1073–1087, https://doi.org/10.1016/S0196-8904(99)00012-6.

[34] J. Chou, N. Truong, Cloud forecasting system for monitoring and alerting of energy use by home appliances, Appl. Energy 249 (2019) 166–177, https://doi.org/10.1016/j.apenergy.2019.04.063.

[35] A. Thangamani, L.S. Ganesh, A. Tanikella, A.M. Prasad, Issues concerning IoT adoption for energy and comfort management in intelligent buildings in India, Intelligent Build. Int. 14 (2022) 74–94, https://doi.org/10.1080/17508975.2020.1838253.

[36] R. Selvaraj, V.M. Kuthadi, S. Baskar, Smart building energy management and monitoring system based on artificial intelligence in smart city, Sustain. Energy Technol. Assess. 56 (2023) 103090, 10.1016/j.seta.2023.103090.

[37] J. Ngarambe, G.Y. Yun, M. Santamouris, The use of artificial intelligence (AI) methods in the prediction of thermal comfort in buildings: energy implications of AI-based thermal comfort controls, 109807, Energ. Buildings 211 (2020), https://doi.org/10.1016/j.enbuild.2020.109807.

[38] G.H. Merabet, M. Essaaidi, M.B. Haddou, B. Qolomany, J. Qadir, M. Anan, D. Benhaddou, Intelligent building control systems for thermal comfort and energy-efficiency: a systematic review of artificial intelligence-assisted techniques, 110969, Renew. Sustain. Energy Rev. 144 (2021), https://doi.org/10.1016/j.rser.2021.110969.

[39] Y. Zhang, B.K. Teoh, M. Wu, J. Chen, L. Zhang, Data-driven estimation of building energy consumption and GHG emissions using explainable artificial intelligence, 125468, Energy 262 (2023), https://doi.org/10.1016/j.energy.2022.125468.

[40] Y. Yao, D.K. Shekhar, State of the art review on model predictive control (MPC) in heating ventilation and air-conditioning (HVAC) field, 107952, Build. Environ. 200 (2021), https://doi.org/10.1016/j.buildenv.2021.107952.

[41] A. Afram, F. Janabi-Sharifi, Theory and applications of HVAC control systems–a review of model predictive control (MPC), Build. Environ. 72 (2014) 343–355, https://doi.org/10.1016/j.buildenv.2013.11.016.

[42] K.U. Ahn, D.W. Kim, K. Cho, D. Cho, H.M. Cho, C.U. Chae, Hybrid model for forecasting indoor CO2 concentration, Buildings 12 (2022) 1540, https://doi.org/10.3390/buildings12101540.

[43] M. Baghoolizadeh, M. Rostamzadeh-Renani, S.A.H.H. Dehkordi, R. Rostamzadeh-Renani, D. Toghraie, A prediction model for CO2 concentration and multi-objective optimization of CO2 concentration and annual electricity consumption cost in residential buildings using ANN and GA, 134753, J. Clean. Prod. 379 (2022), https://doi.org/10.1016/j.jclepro.2022.134753.

[44] M. Baghoolizadeh, M. Rostamzadeh-Renani, M. Hakimazari, R. Rostamzadeh-Renani, Improving CO2 concentration, CO2 pollutant and occupants' thermal

comfort in a residential building using genetic algorithm optimization, 113109, Energ. Buildings 291 (2023), https://doi.org/10.1016/j.enbuild.2023.113109.
[45] MATLAB. (2024). version 2023a. Natick, Massachusetts: The MathWorks Inc. Available online: http://www.mathworks.com/products/matlab (accessed on September 15, 2024).
[46] Statistics and Machine Learning Toolbox. (2024). version 12.5. Natick, Massachusetts: The MathWorks Inc. Available online: http://www.mathworks.com/products/statistics (accessed on September 15, 2024).
[47] Arduino Uno R3. Datasheet available online: https://docs.arduino.cc/resources/datasheets/A000066-datasheet.pdf (Accessed on September 15, 2024).
[48] NodeMCU Lua Lolin Modulo V3 ESP8266 ESP-12F WIFI. Datasheet available online: https://cdn.shopify.com/s/files/1/1509/1638/files/NodeMCU_Lua_Lolin_V3_Modul_mit_ESP8266_12E_Datenblatt.pdf?342081239282763366 (Accessed on September 15, 2024).
[49] Google docs. Google LLC. https://docs.google.com/ (Accessed on September 15, 2024).
[50] : Conrad Electronic SE. NTC Datasheet. Datasheet available online: https://www.mouser.com/datasheet/2/758/DHT11-Technical-Data-Sheet-Translated-Version-1143054.pdf (Accessed on September 15, 2024).
[51] : Mouser Electronics. DHT11Humidity and temperature sensor. Datasheet available online: https://www.mouser.com/datasheet/2/758/DHT11-Technical-Data-Sheet-Translated-Version-1143054.pdf (Accessed on September 15, 2024).
[52] : BOSCH, Pressure sensor BMP280. Datasheet available online: https://www.bosch-sensortec.com/media/boschsensortec/downloads/datasheets/bst-bmp280-ds001.pdf (Accessed on September 15, 2024).
[53] : HC-SR501 PIR Motion Detector. Datasheet available online: https://www.mpja.com/download/31227sc.pdf (Accessed on September 15, 2024).
[54] E+E Elektronic. Digital Sensor Module for CO2, Temperature, Humidity and Ambient pressure mod. EE894. Datasheet available online: https://www.epluse.com/fileadmin/data/product/ee894/datasheet_EE894.pdf (Accessed on September 15, 2024).
[55] : EN ISO 7726:2001. Ergonomics of the thermal environment - Instruments for measuring physical quantities. International Organization for Standardization, Geneva, 2001.
[56] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, A. Lopez, A comprehensive survey on support vector machine classification: Applications, challenges and trends, Neurocomputing, 408 (2020), 189-215, Doi: 10.1016/j.neucom.2019.10.118.
[57] B. De Ville, Decision trees, Wiley Interdiscip. Rev. Comput. Stat. 5 (2013) 448–455, https://doi.org/10.1002/wics.1278.
[58] D. Denisko, M.M. Hoffman, Classification and interaction in random forests, Proc. natl. acad. sci. 115 (2018) 1690–1692, https://doi.org/10.1073/pnas.1800256115.
[59] A. Abraham, Artificial neural networks, Handbook Measur. Syst. Des. (2005), https://doi.org/10.1002/0471497398.mm421.
[60] https://en.climate-data.org/europe/poland/podlaskie-voivodeship/bialystok-1031/ (Accessed on September 14, 2024).
[61] M. Kottek, J. Grieser, C. Beck, B. Rudolf, F. Rubel, world map of the köppen-geiger climate classification updated, Meteorol. Z. 15 (2006) 259–263, https://doi.org/10.1127/0941-2948/2006/0130.
[62] V. Ballerini, B. Lubowicka, P. Valdiserri, D.A. Krawczyk, B. Sadowska, M. Kłopotowski, E.R. di Schio, The energy retrofit impact in public buildings: a numerical cross-check supported by real consumption data, Energies, 16 (2023), 7748, https://doi.org/ 10.3390/en16237748.
[63] Data Bank of the Institute of Meteorology and Water Management. http://www.imgw.pl/ (Accessed on September 14, 2024).