



**HAL**  
open science

# Investigation of Input Feature Combinations Considering Occupant Behavior for Modelling Indoor Air Temperature in a Classroom

Abderrahmane Soufi, Abhinandana Boodi

## ► To cite this version:

Abderrahmane Soufi, Abhinandana Boodi. Investigation of Input Feature Combinations Considering Occupant Behavior for Modelling Indoor Air Temperature in a Classroom. BS 2023, IBPSA, Sep 2023, Shanghai (CH), China. <hal-04243765>

**HAL Id: hal-04243765**

**<https://hal.science/hal-04243765v1>**

Submitted on 15 Jan 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

# Investigation of Input Feature Combinations Considering Occupant Behavior for Modelling Indoor Air Temperature in a Classroom

Abderrahmane Soufi<sup>1</sup>, Abhinandana Boodi<sup>2</sup>

<sup>1</sup>CESI LINEACT Immeuble Le Quatrième Zone Aéroportuaire de Montpellier Méditerranée, 34130 Mauguio

<sup>2</sup>CESI LINEACT 230 Rue Rolland Garros, 29490 Guipavas

## Abstract

This study investigated the performance of artificial neural networks and random forests with various combinations of input variables in modelling indoor air temperature in a classroom. The data collection methodology was designed to investigate key input parameters, including indoor air data, classroom occupancy, and occupant behavior factors such as windows, doors, blind operation, and occupant interaction with HVAC control. The random forest models exhibited the highest accuracy (determination coefficient of 0.99) compared with artificial neural networks (determination coefficient of 0.80). The results also show that the occupant behavior data are important parameters for model performance. Combinations with limited occupant behavior data exhibited high root-mean-square errors.

## Highlights

- Simpler machine learning algorithms were developed for classroom temperature prediction.
- Corridor and outdoor temperatures are the environmental parameters that influence the indoor temperature the most.
- The number of occupants, door opening/closing, window opening/closing, and blind operation data are essential for model performance.
- The RF model outperformed the ANN in indoor temperature prediction.

## Introduction

In France, the residential and tertiary building sectors account for 49% of energy consumption and 1/4th of greenhouse gas emissions. In this context, the Tertiary Decree established goals for the building sector to address climate change by reducing energy consumption by 40%, 50%, and 60% by 2030, 2040, and 2050, respectively. These goals can be achieved by enhancing a building's operational efficiency while ensuring occupant comfort (Tchang (2022)). The trade-off between occupant comfort and energy consumption reduction is a challenging task in building design and energy management (Boodi et al. (2018)). An effective strategy is to optimize the building design and operation strategies by utilizing indoor temperature predictions.

Traditionally, indoor temperatures have been predicted using simulation tools (white boxes), which require input parameters based on specific building characteristics and

occupant behaviors. The simulation tools consider the occupants deterministically through occupancy scenarios, which increases the uncertainties between the actual and predicted temperatures. An alternative to simulation tools is modeling using data-driven approaches (black boxes). In contrast to white-box models, these models offer faster predictions and do not require an explicit physical understanding of the building (Wang et al. (2022); Amasyali and El-Gohary (2018)).

In previous studies, black-box models were developed to predict the indoor temperatures in different buildings. Ramadan et al. (2021) compared the ability of different models to predict the indoor temperature in an unoccupied closed room of 9 m<sup>2</sup> in a university laboratory without windows. The accuracy of these models was evaluated using the root mean square error (RMSE) and the coefficient of determination (R<sup>2</sup>). The methods used were an artificial neural network (ANN), multi-layer perception (MLP) structure, multiple linear regression (MLR), decision tree (DT), random forest (RF), extra trees (ET), gradient boosting (GB), and extreme gradient boosting (XGB). They found that the ANN (RMSE = 0.081 and R<sup>2</sup> = 0.99965) and ET (RMSE = 0.159 and R<sup>2</sup> = 0.99864) performed better at predicting the indoor temperature. Fang et al. (2021) used long short-term memory (LSTM) to predict indoor temperatures in rooms of institutional buildings. They used weather data (outdoor temperature and solar radiation) and outdoor temperature (CO<sub>2</sub> concentration and room temperature setpoint) as input variables. The performances of the three different architectures (LSTM–dense, LSTM–LSTM, LSTM–dense–LSTM models) were evaluated using the RMSE, R<sup>2</sup>, mean absolute scaled error (MASE), and mean normalized absolute error (MNAE). They found that the LSTM–dense model performed better for short-term prediction (48 h ahead).

Alawadi et al. (2020) compared the accuracy in terms of the R-coefficient and RMSE of different methods for forecasting indoor temperature in university buildings. Among these methods, artificial neural networks, linear regression, RF, and regression trees have been used. They used a dataset of the HVAC system (status (on/off), set-point temperature, and percentage of humidity attached to the central air-conditioning flow) and weather data (temperature, relative humidity, and solar radiation). The results revealed that the ET regressor achieved the highest average accuracy (0.97%) and performance (0.058%).

Mateo et al. (2013) compared different modeling methods to predict the indoor temperature in the rooms of tertiary buildings. They used an autoregressive model (AR), robust MLR, MLP, extreme learning machine (ELM), and nonlinear autoregressive exogenous multi-layer perceptron (NARXMLP). The datasets included weather data (temperature and relative humidity) and indoor data (setpoint temperature and total thermal power). These methods were evaluated using mean absolute error (MAE). The best overall performing method was NARXMLP with an MAE of 0.11 °C. Delcroix et al. (2021) developed autoregressive neural network models to predict indoor temperatures in commercial buildings. The datasets included the outdoor temperature, relative occupancy, and state of the heating system (on/off). They reported an  $R^2$  of up to 0.824 and RMSE of 1.11 °C. Feng et al. (2022) compared two methods—MLR and RF regression—to predict the indoor air temperature in a teaching building. They used outdoor conditions, occupancy features, and spatial design variables as the input characteristics. They found that RF regression performed better in predicting the indoor temperature.

Most of the previous related studies were performed without considering occupant behavior, which has a significant impact on indoor temperature, particularly in high-occupancy-density environments such as educational classrooms. Occupants can affect the thermal and energy performance of a building through their presence and control of various building systems such as lighting, doors, windows, and HVAC systems (Zambrano et al. (2021); Belazi et al. (2022); Azar et al. (2020); Carlucci et al. (2020)).

For an effective prediction of these models, precise data must be integrated with the behavior of the occupant. In this study, we developed a model to simulate and predict the indoor temperature of a classroom using an ANN and an RF. This study investigated the combination of inputs related to occupants, such as windows, doors, blind operations, and occupant interactions, with HVAC controls.

## Data collection and methodology

The study was conducted in a higher education building located in the Montpellier Airport area in southern France. The study was conducted between autumn 2022 and winter 2023 (September–January). The study area was a classroom located on the ground floor with a south-facing orientation that provides direct solar radiation in the afternoon. The classroom had a surface area of 53 m<sup>2</sup>, and it had three windows with electric blinds, two doors (one of the doors is always closed), and a heating/air-conditioning system with two grilles of blowing located in the ceiling (Figure 1). Doors, windows, and blinds are elements that the occupant can act on. Occupants can also turn on the HVAC system to adjust the set temperature.

### Data collection

Physical monitoring and occupant observations were used for data collection. These methods include the measurement of indoor and outdoor environmental parameters (air temperature and relative humidity) and

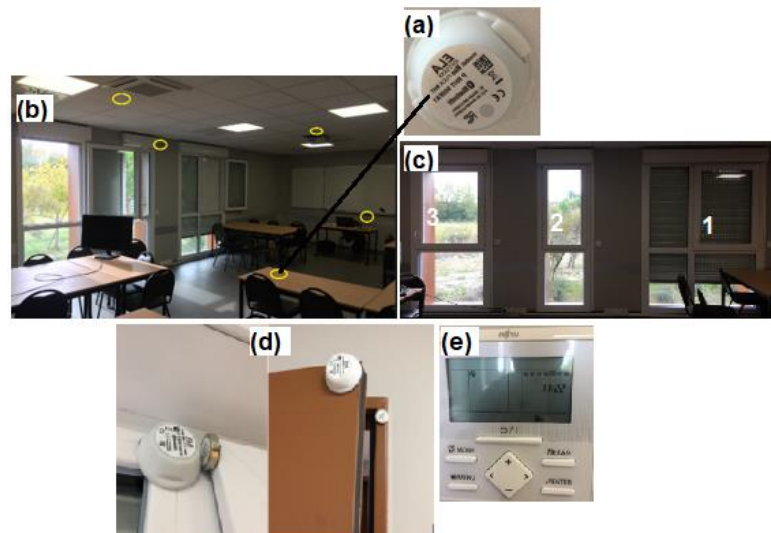


Figure 1: Data collection in classroom (a) RHT sensor (b) Installed sensors in the classroom (c) : windows and blinds (d) magnetic sensor (e) HVAC remote control

occupant interactions with building systems (window and door status: open/closed, blind state, and thermostat/air-conditioning adjustment). Some of these features are shown in Figure 2. Five sensors were installed in the classroom to measure both temperature and relative humidity. The average of these two parameter values was considered the classroom temperature.

- Weather data: Weather data were collected from a Meteo France station located 1 km from the investigated buildings. The station provided hourly observational values for the main weather elements. In this study, the outdoor air temperature, relative humidity, wind speed, and solar radiation were collected as potential control variables to indicate different outdoor conditions.
- Indoor air data in the classroom: The indoor air temperature and relative humidity in the classroom were monitored using a Blue Puck RHT wireless sensor. Six RHT sensors were placed at different locations: one in the center, two on the ceiling next to grilles, one on the carpentry of one of the windows, and one in the corridor outside the classroom. Indoor parameters were recorded at ten minutes intervals.
- Classroom occupancy characteristics: The number of occupants was determined hourly (morning and afternoon) by counting and surveying (attendance sheets). The usage schedules of the classroom were 8:30–18:00. The number of occupants varied from 0 to 31.

Occupant behavior monitoring:

- Windows and door operation: The states of doors and windows (open or closed) were monitored using a Blue Puck Mag wireless sensor, which is a magnetic sensor (see Figure 1(d)) that detects the openings of

doors and windows. The states of the door and windows were recorded at ten minute intervals.

- **Blind operation:** The window-blind closing rate was determined by visual observation. The closing rates were 0%, 25%, 50%, 75%, and 100%. Observations were conducted throughout the day in the morning and afternoon at 1 h intervals. Observations showed that the occupants opened their blinds in the morning as soon as they arrived in the classroom. The occupants interacted with the blinds, particularly in the afternoon between 1 pm and 3 pm, when the solar radiation was the highest.
- **Occupant interaction with HVAC controls:** The state of the heating/air-conditioning system was determined to be off or on every hour (in the morning and afternoon) using the HVAC remote control (see Figure 1(e)). Air conditioning was used until the beginning of November, and heating was used from the end of November.

### Artificial Neural Network

An ANN is a type of machine learning model composed of three layers: an input layer, one or more hidden layers, and an output layer (Figure 3). Each layer constitutes a network of connected nodes, and the connections between the nodes are weighted to aid the model to learn and predict the future. A few popular activation functions that can be used in ANNs are ReLU, sigmoid, and Tanh. Adam and stochastic gradient descent are two common optimization algorithms that can be used to train models. ANNs are effective machine-learning models that can be applied to many different tasks, including regression and classification. They are effective in handling large and complex datasets such as occupancy and building datasets (Mohandes et al. (2019); Krogh (2008)).

ANNs have various architectural forms. This study adopted the MLP architecture. The input layer receives the input data and the output layer generates the final prediction or decision of the model. This type of ANN comprises multiple layers of interconnected neurons. The majority of computations in the model are performed by hidden layers, which are between the input and output layers. The weight of the corresponding neurons in the hidden layer is multiplied by the neurons in the input layer

during the ANN data mapping process, and the sum of these weights is then applied with a bias to the neurons in the output layers. When used with a single hidden layer and adequate number of neurons, the MLP architecture provides good accuracy in approximating functions (Ruano et al. (2006)).

For temperature forecasting, ANNs are used to forecast the indoor temperature of a building to optimize the energy efficiency of the system and guarantee occupant comfort. Occupant behavior and outdoor and indoor conditions can be used as input parameters for the model, and the prediction time is influenced by the building's thermal inertia and energy regulation system (Gonzalez and Zamarreno (2005); Huang et al. (2013)). The hyperparameters of the ANN are selected through grid-search optimization. The following configuration was used to develop the ANN model: four hidden layers with 32 units in each layer, the mean squared error (MSE) as the loss function, and Adam as the optimization function combined with sigmoid and ReLU activation functions, which are suitable for regression problems, and 100 epochs.

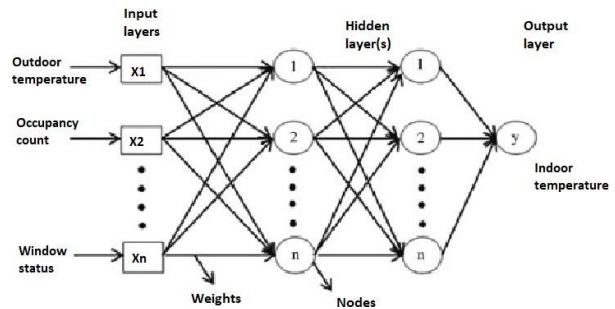


Figure 3: ANN Architecture.

### Random Forest Algorithm

An RF is a machine learning model that combines the predictions of multiple DT models to perform more accurate and stable predictions. Figure 4 shows the modeling procedure. It accomplishes predictions using different subsets of training data to train multiple DT models, averaging the predictions from all trees, and then performing the final prediction. By using a subset of features that are randomly selected when training each tree, RF can increase the predictive accuracy of the model and decrease overfitting. By adjusting hyperparameters such as the maximum depth and minimal number of samples necessary to split a node, users can also regulate the complexity of the trees. Overall, RF is a robust and popular machine learning algorithm that can handle large and complex datasets and has good performance (Biau and Scornet (2016); Schonlau and Zou (2020); Kanthila et al. (2022) Fawagreh et al. (2014)).

### Performance metrics

In the modeling process, the performance of the models was evaluated using five metrics: MSE, RMSE,  $R^2$ , MAE, and symmetric mean absolute percentage error (sMAPE) (Ciulla and D'Amico (2019); Ahmad and Chen (2018)):

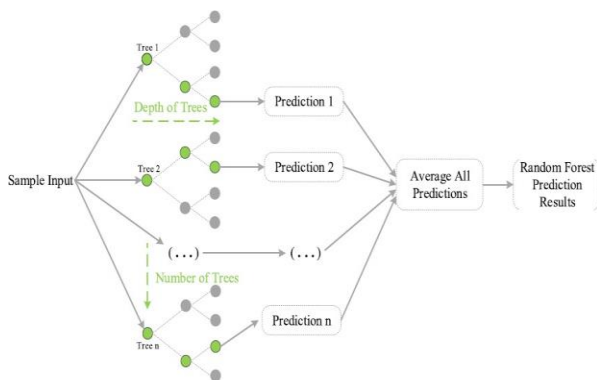


Figure 4: Modeling process of random forest (Chen et al. (2022)).

The MSE is the average squared difference between model prediction and target value.

The RMSE is the difference between the value predicted by the model and the measured value.

$R^2$  is an indicator of the quality of model prediction.  $R^2$  range from 0 to 1. The closer it is to 1, the more the results of the modelling match the data collected.

For the MAE, the smaller the value, the better the model's performance.

The sMAPE is a measurement of accuracy based on percentage errors.

$$MSE = \frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2 \quad (1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2} \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (X_i - Y_i)^2}{\sum_{i=1}^N (X_i - X_a)^2} \quad (3)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |X_i - Y_i| \quad (4)$$

$$sMAPE = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i - X_i|}{(X_i + Y_i) / 2} \quad (5)$$

where

$X_i$  is the  $i$ -th expected output;

$Y_i$  is the  $i$ -th predicted output;

$X_a$  is the average of the whole desired output; and

$N$  is the number of the identification set samples.

## Features Analysis

The ANN and RF models used the collected data on weather, building conditions, and occupancy rates as input variables. In this study, we used 70% of the dataset as the training dataset and the remaining 30% to test the accuracy. A sensitivity analysis was developed to determine the basic impact of each variable to validate the effects of the model variables. The input and output data used in the sensitivity analysis were identical for the RF and ANN model training. Both models were tested for multiple input combinations to verify their impact and to select an appropriate combination. Figure 5 shows the correlation between the data points. Temperature data, such as corridors and outdoors, had strong positive correlations with indoor temperatures, whereas the

opening/closing rate of blinds had negative correlations. Based on these correlation data, and after approximately 1000 simulations of different combinations, we selected a few combinations (Table 2) to show the results.

## Results and discussion

The proposed models were compared using multiple performance indices. The results were composed of both a training set and test metrics. Thus, model robustness and overfitting could be verified.

A performance comparison between the results indicated that it is not always necessary to have many input variables, but a combination of both environmental conditions and occupancy-behavior data is necessary. Tables 3 and 4 show the combinations used to evaluate both the ANN and RF model performances. The performance of the ANN was similar for all combinations during training; however, the testing performance had a high variation for different combinations. However, the RF model exhibited consistency in both training and testing performances. The RF model outperformed the ANN for all combinations.

The results also showed that occupant behavior data are an important parameter for model performance. Combinations with limited behavior data such as combinations 1–7 had a high RMSE for both training and testing. Furthermore, as shown in Table 4, combination 10 had better performance than the others. Combination 10 had 10 input features, which indicated that the remaining features have less impact. Combinations 11 and 12 contained 14 and 22 input features, respectively. The  $R^2$  value for combinations 9–12 was approximately 0.99; however, the MSE, RMSE, and MAE had different values. When behavior data such as blind operation, occupant number, door open/close, and window state were considered, the model had a better  $R^2$  and RMSE. The RF model exhibited errors of approximately the same order of magnitude for both the training and testing datasets. These results indicated that the list of inputs used was sufficient to explain the variations in the indoor temperatures in the classroom.

We have also observed that in the absence of corridor and outdoor temperature features, the model performed very poorly by showing RMSE of more than 1.5 °C. Similarly, when the relative humidity values were used for training and testing, the model performed poorly, indicating that the window and door statuses influenced this parameter. Thus, detailed information about occupant space and system utilization patterns, as well as the number of occupants, is critical during data collection to improve model performance.

We demonstrated that simple machine algorithms, such as an ANN and RF, are effective for modeling and predicting the indoor temperature of a classroom. The dataset size, number of occupants, and choice of input features significantly affect the model performance. All the outcomes examined here provide evidence for this.

## Conclusion

The main objective of this study was to investigate the performance of a simple machine learning approach for modeling indoor temperature in a classroom and the influence of input features on model performance. A comparative study was conducted between the two ML models, ANN and RF. We also studied and compared different input-feature combinations. The results proved that simpler machine learning algorithms perform well for classroom temperature prediction. However, if we compare the performance metrics between the ANN and RF for both the testing and training datasets, the RF is prominent. The results signify that with proper data collection, the processing is the same for model fittings, and even simple standard models perform well. The proposed model is simple to develop, requires few input variables, and is computationally efficient.

Based on the simulation results of the feature impact analysis, the temperatures in the corridor and outdoors are more influential than other environmental parameters. Humidity is the least effective variable. When occupancy behavioral features are used with temperature features, the model performs better. This indicates that parameters such as door opening/closing, window opening/closing, and blind operation data are essential for model performance when applied to buildings in bitter climates. Furthermore, the model may not perform well when all input features are used. Some may be irrelevant to output-feature dynamics. Thus, we recommended that a feature impact analysis be conducted to determine the most suitable features. It can also be used to reduce IoT system costs by installing sensors for only the most influential parameters. Future studies should explore more algorithms and generalizations, and improve the robustness of the algorithm. Occupant and environmental variables have a significant impact on the model. Therefore, data collection must be conducted carefully. Studies predicting occupant behavior are still lacking to their heterogeneous nature (Kanthila et al. (2021)). This aspect requires further research.

## References

- Ahmad, T. and H. Chen (2018). Short and medium term forecasting of cooling and heating load demand in building environment with data-mining based approaches. *Energy and Buildings* 166, 460–476.
- Alawadi, S., D. Mera, M. Fern´andez-Delgado, F. Alkhabbas, C. M. Olsson, and P. Davidsson (2020). A comparison of machine learning algorithms for forecasting indoor temperature in smart buildings. *Energy Systems*, 1–17.
- Amasyali, K. and N. M. El-Gohary (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews* 81, 1192–1205.
- Azar, E., W. O’Brien, S. Carlucci, T. Hong, A. Sonta, J. Kim, M. S. Andargie, T. Abuimara, M. El Asmar, and R. K. Jain (2020). Simulation-aided occupant-centric building design: A critical review of tools, methods, and applications. *Energy and Buildings* 224, 110292.
- Belazi, W., S.-E. Ouldboukhitine, A. Chateaneuf, and A. Bouchair (2022). Thermal modeling of the occupied multi-zone buildings taking into account the uncertainties of occupant behavior. *Case Studies in Thermal Engineering* 33, 101978.
- Biau, G. and E. Scornet (2016). A random forest guided tour. *Test* 25, 197–227.
- Boodi, A., K. Beddiar, M. Benamour, Y. Amirat, and M. Benbouzid (2018). Intelligent systems for building energy and occupant comfort optimization: A state of the art review and recommendations. *Energies* 11 (10), 2604.
- Carlucci, S., M. De Simone, S. K. Firth, M. B. Kjærgaard, R. Markovic, M. S. Rahaman, M. K. Annaqeeb, S. Biandrate, A. Das, and J. W. Dziejczak (2020). Modeling occupant behavior in buildings. *Building and Environment* 174, 106768.
- Chen, Y., M. Guo, Z. Chen, Z. Chen, and Y. Ji (2022). Physical energy and data-driven models in building energy prediction: A review. *Energy Reports* 8, 2656–2671.
- Ciulla, G. and A. D’Amico (2019). Building energy performance forecasting: A multiple linear regression approach. *Applied Energy* 253, 113500.
- Delcroix, B., J. L. Ny, M. Bernier, M. Azam, B. Qu, and J.-S. Venne (2021). Autoregressive neural networks with exogenous variables for indoor temperature prediction in buildings. *Building Simulation* 14, 165–178.
- Fang, Z., N. Crimier, L. Scanu, A. Midelet, A. Alyafi, and B. Delinchant (2021). Multi-zone indoor temperature prediction with LSTM-based sequence to sequence model. *Energy and Buildings* 245, 111053.
- Fawagreh, K., M. M. Gaber, and E. Elyan (2014). Random forests: From early developments to recent advancements. *Systems Science & Control Engineering: An Open Access Journal* 2(1), 602–609.
- Feng, J., Z. Zhou, and W. Li (2022). Associating indoor air temperature with building spatial design and occupancy features: A statistical analysis on university classrooms. *Building and Environment* 216, 109009.
- Gonzalez, P. A. and J. M. Zamarreno (2005). Prediction of hourly energy consumption in buildings based on a feedback artificial neural network. *Energy and Buildings* 37(6), 595–601.
- Huang, H., L. Chen, M. Mohammadzahari, E. Hu, and M. Chen (2013). Multi-zone temperature prediction in a commercial building using artificial neural network model. *2013 10th IEEE International Conference on Control and Automation (ICCA)*, 1896–1901. Hangzhou (China), , June 12-14, 2013

- Kanthila, C., A. Boodi, K. Beddiar, Y. Amirat, and M. Benbouzid (2021). Building occupancy behavior and prediction methods: A critical review and challenging locks. *IEEE Access* 9, 79353–79372.
- Kanthila, C., A. Boodi, K. Beddiar, Y. Amirat, and M. Benbouzid (2022). Building occupancy detection using machine learning-based approaches: Evaluation and comparison. *IECON 2022–48th Annual Conference of the IEEE Industrial Electronics Society*, pp. 1–6.
- Krogh, A. (2008). What are artificial neural networks? *Nature Biotechnology* 26(2), 195–197.
- Mateo, F., J. J. Carrasco, A. Sellami, M. Millan-Giraldo, M. Domínguez, and E. Soria-Olivas (2013). Machine learning methods to forecast temperature in buildings. *Expert Systems with Applications* 40(4), 1061–1068.
- Mohandes, S. R., X. Zhang, and A. Mahdiyari (2019). A comprehensive review on the application of artificial neural networks in building energy analysis. *Neurocomputing* 340, 55–75.
- Ramadan, L., I. Shahrouh, H. Mroueh, and F. H. Chehade (2021). Use of machine learning methods for indoor temperature forecasting. *Future Internet* 13(10), 242.
- Ruano, A. E., E. M. Crispim, E. Z. Conceição, and M. M. J. Lucio (2006). Prediction of building's temperature using neural networks models. *Energy and Buildings* 38(6), 682–694.
- Schonlau, M. and R. Y. Zou (2020). The random forest algorithm for statistical learning. *The Stata Journal* 20(1), 3–29.
- Tchang, N. (2022). Décret Eco-énergie Tertiaire: Méthodologie et enjeux à l'échelle d'un patrimoine immobilier. *Annales Des Mines-Réalités Industrielles* 2, 66–68. Cairn/Softwin.
- Zambrano, J. M., U. F. Oberegger, and G. Salvalai (2021). Towards integrating occupant behaviour modelling in simulation-aided building design: Reasons, challenges and solutions. *Energy and Buildings* 253, 111498.

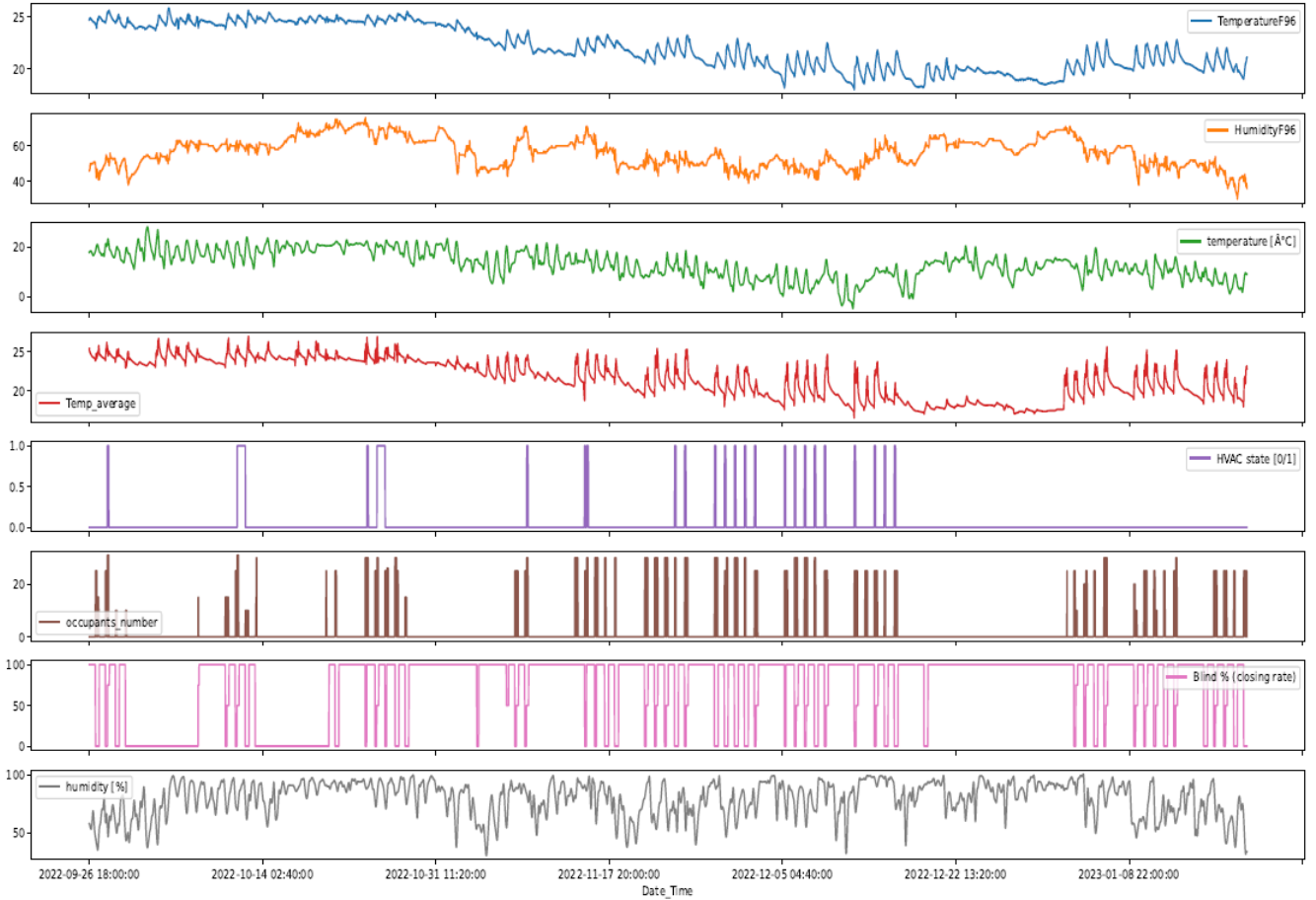


Figure 2: Classroom data used for the model training.

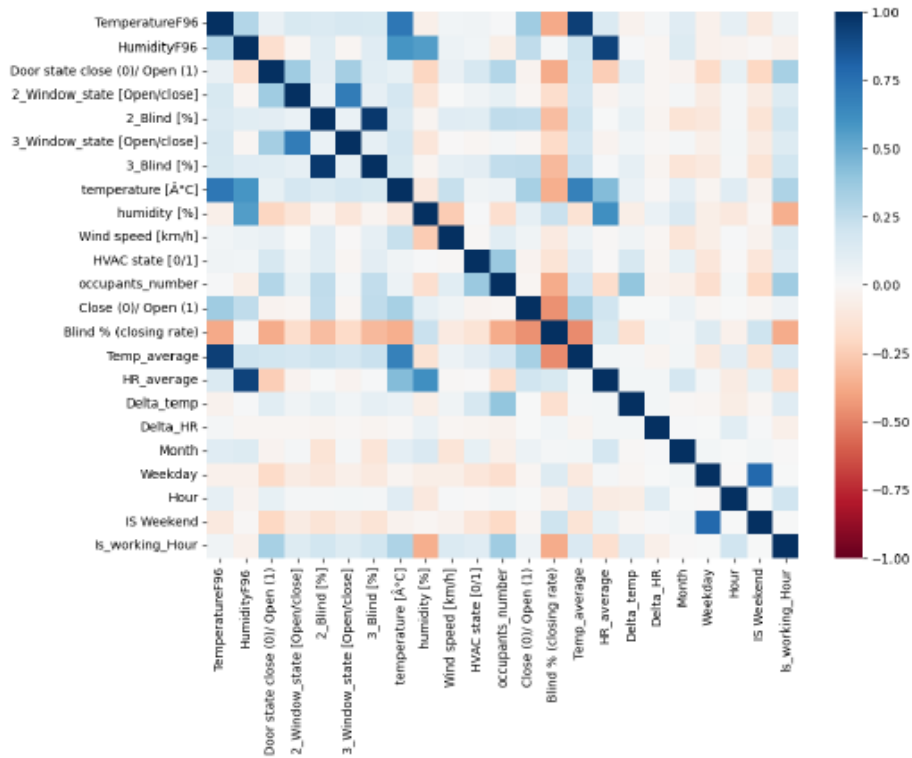


Figure 5: Input feature relations and correlations

Table 1: Input features.

Input	Unit	Description
Time	YYYY-MM-DD HH:MM:SS	Time including year, month, date, hour, minutes and seconds
Month	[-]	Month index
Weekday	[-]	Day index in a week (0–6)
Hour	[-]	Hour index in a day (0–23)
Is Weekend	Boolean	Working or non-working day (TRUE/FALSE)
Is working hour	Boolean	Working hour (TRUE/FALSE)
Temperature difference	°C	Difference between temperatures at time t and t+1
Relative humidity difference	%	Difference between relative humidity at time t and t+1
Relative humidity	%	Indoor relative humidity measured using sensors (averaged values from 5 sensors)
Occupancy	In numbers	Occupancy numbers present in the classroom
Temperature exterior	°C	Temperature of outdoor environment
Relative humidity exterior	%	Relative humidity of outdoor environment
Wind speed	km/h	External wind speed collected
HVAC status	Boolean	HVAC system condition (ON/OFF)
Door status	Boolean	Door status (Open/Close)
Window status (1,2,3)	Boolean	Three windows status (Open/Close)
Window blind status (1,2,3)	%	Percentage of window blinds open (5 states: 0, 25, 50, 75, 100)
Temperature corridor	°C	Temperature of the corridor - TemperatureF96
Humidity corridor	%	Corridor relative humidity measured using sensors - HumidityF96

Table 2: Input features combinations.

Combination N°	Combinations
Comb 1	TemperatureF96 (corridor), HumidityF96 (corridor)
Comb 2	TemperatureF96, occupants number
Comb 3	TemperatureF96, Blind % (closing rate)
Comb 4	TemperatureF96, temperature outdoor
Comb 5	TemperatureF96, temperature outdoor, occupants number
Comb 6	TemperatureF96, HumidityF96, temperature outdoor, occupants number
Comb 7	TemperatureF96, temperature outdoor, humidity [%], occupants number, Wind speed [km/h]
Comb 8	TemperatureF96, HumidityF96, Door state close (0)/ Open (1), 2Window state [Open/close], 3Window state [Open/close]
Comb 9	TemperatureF96, HumidityF96, 2Window state [Open/close], 3Blind [%], temperature outdoor, humidity [%], Wind speed [km/h], Month
Comb 10	TemperatureF96, HumidityF96, Door state close (0)/ Open (1), 2Blind [%], 3Blind [%], temperature outdoor, humidity [%], Wind speed [km/h], Month, humidity indoor
Comb 11	TemperatureF96, HumidityF96, Door state close (0)/ Open (1), 2Window state [Open/close], 2Blind [%], 3Window state [Open/close], 3Blind [%], temperature outdoor, humidity [%], Wind speed [km/h], HVAC state [0/1], occupants number, Month, humidity indoor
Comb 12	all features listed in Table 1

Table 3: ANN model performance evaluation for different combinations.

Combinations	Training performances					Testing performances				
	MSE	R <sup>2</sup>	RMSE	MAE	sMAPE	MSE	R <sup>2</sup>	RMSE	MAE	sMAPE
comb1	0.565	0.867	0.752	0.491	0.022	0.962	0.681	0.980	0.736	0.051
comb2	0.436	0.897	0.660	0.454	0.020	0.683	0.773	0.826	0.620	0.042
comb3	0.515	0.879	0.717	0.531	0.024	1.281	0.575	1.132	0.962	0.053
comb4	0.570	0.866	0.755	0.595	0.027	1.656	0.451	1.287	1.079	0.063
comb5	0.447	0.895	0.668	0.453	0.020	0.944	0.687	0.971	0.789	0.048
comb6	0.441	0.896	0.664	0.508	0.023	0.943	0.687	0.971	0.793	0.045
comb7	0.598	0.859	0.773	0.506	0.023	1.179	0.609	1.086	0.854	0.056
comb8	0.508	0.880	0.712	0.461	0.021	0.895	0.703	0.946	0.740	0.044
comb9	0.500	0.882	0.707	0.498	0.022	0.838	0.722	0.915	0.692	0.046
comb10	0.430	0.899	0.655	0.459	0.021	0.684	0.773	0.827	0.672	0.036
comb11	0.390	0.908	0.625	0.417	0.019	0.580	0.807	0.762	0.583	0.035
comb12	0.342	0.919	0.584	0.413	0.019	0.863	0.713	0.929	0.772	0.045

Table 4: RF model performance evaluation for different combinations.

Combinations	Training performances					Testing performances				
	MSE	R2	RMSE	MAE	sMAPE	MSE	R <sup>2</sup>	RMSE	MAE	sMAPE
comb1	0.185	0.969	0.431	0.244	0.130	0.502	0.915	0.708	0.410	0.131
comb2	0.329	0.944	0.573	0.395	0.129	0.450	0.926	0.671	0.446	0.127
comb3	0.325	0.945	0.570	0.389	0.130	0.451	0.924	0.671	0.445	0.128
comb4	0.067	0.988	0.260	0.154	0.130	0.442	0.925	0.665	0.405	0.124
comb5	0.041	0.993	0.204	0.118	0.131	0.276	0.953	0.526	0.315	0.130
comb6	0.441	0.896	0.664	0.508	0.023	0.943	0.687	0.971	0.793	0.045
comb7	0.020	0.996	0.143	0.068	0.131	0.137	0.977	0.370	0.178	0.132
comb8	0.131	0.978	0.362	0.198	0.131	0.381	0.935	0.617	0.344	0.132
comb9	0.008	0.998	0.092	0.041	0.131	0.054	0.991	0.233	0.112	0.131
comb10	0.006	0.998	0.078	0.035	0.131	0.035	0.994	0.188	0.085	0.128
comb11	0.006	0.998	0.082	0.034	0.131	0.046	0.992	0.216	0.094	0.126
comb12	0.005	0.999	0.072	0.030	0.131	0.041	0.993	0.203	0.085	0.126