



The COVID-19 impact on air condition usage: a shift towards residential energy saving

Muhammad Saidu Aliero¹ · Muhammad Fermi Pasha¹ · Adel N. Toosi² · Imran Ghani³

Received: 29 July 2021 / Accepted: 26 November 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

The enforcement of the Movement Control Order to curtail the spread of COVID-19 has affected home energy consumption, especially HVAC systems. Occupancy detection and estimation have been recognized as key contributors to improving building energy efficiency. Several solutions have been proposed for the past decade to improve the precision performance of occupancy detection and estimation in the building. Environmental sensing is one of the practical solutions to detect and estimate occupants in the building during uncertain behavior. However, the literature reveals that the performance of environmental sensing is relatively poor due to the poor quality of the training dataset used in the model. This study proposed a smart sensing framework that combined camera-based and environmental sensing approaches using supervised learning to gather standard and robust datasets related to indoor occupancy that can be used for cross-validation of different machine learning algorithms in formal research. The proposed solution is tested in the living room with a prototype system integrated with various sensors using a random forest regressor, although other techniques could be easily integrated within the proposed framework. The primary implication of this study is to predict the room occupation through the use of sensors providing inputs into a model to lower energy consumption. The results indicate that the proposed solution can obtain data, process, and predict occupant presence and number with 99.3% accuracy. Additionally, to demonstrate the impact of occupant number in energy saving, one room with two zones is modeled each zone with air condition with different thermostat controller. The first zone uses IoFClima and the second zone uses modified IoFClima using a design-builder. The simulation is conducted using EnergyPlus software with the random simulation of 10 occupants and local climate data under three scenarios. The Fanger model's thermal comfort analysis shows that up to 50% and 25% energy can be saved under the first and third scenarios.

Keywords Occupancy detection and estimation · Environmental sensing · Indoor comfort · A training dataset

Introduction

The adoption of the Movement Control Order (MCO) during the COVID-19 pandemic has impacted global residential energy consumption patterns. A statistical study conducted on residential electric appliance utilization online indicates occupants' poor integration in appliance control.

Consequently, the hourly usage of residential appliances increased significantly during the MCO period and after the MCO from March to June 2020, especially an air conditioner that doubled during MCO and increased by 5 hours after the first MCO (F. Shahzad et al. 2020a, b; K. Shahzad et al. 2020a, b; Shahzad et al. 2021; Shakoor et al. 2020) (see Fig. 1), while utilization of the majority of the appliances considered decreased significantly after the MCO period (Fareed et al. 2020; Iqbal et al. 2020). This indicates that the pandemic opens on to a surge in the continued use of the air conditioner despite a partial resumption of office activities.

Integration of occupancy information in HVAC system control must be optimized to fully exercise energy consumption control on the HVAC system to eliminate energy waste (Aliero et al. 2021). Recently, researchers have applied direct sensing technologies such as cameras, wearables, and environmental sensing alongside machine learning techniques to

Responsible Editor: Lotfi Aleya

✉ Muhammad Saidu Aliero
muhammad.saidu@monash.edu

¹ School of Information Technology, Monash University, 47500 Subang Jaya, Malaysia

² Faculty of Information Technology, Monash University, Clayton, Australia

³ Virginia Military Institute, Lexington, VA, USA

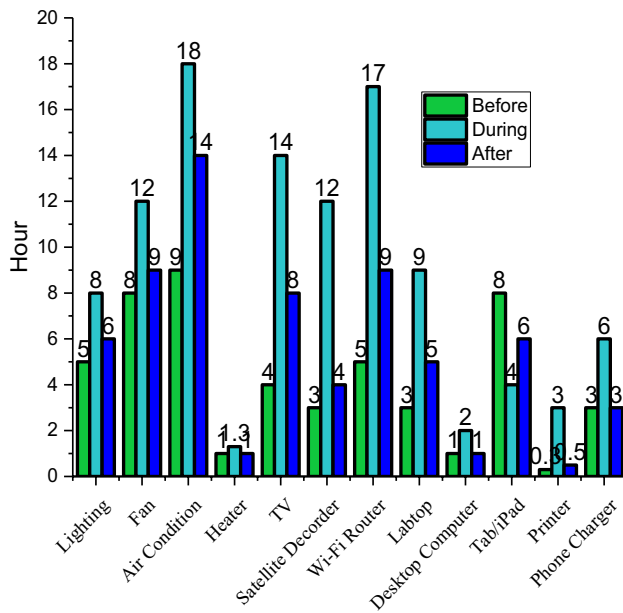


Fig. 1 Home appliance energy consumption before, during, and after MCO

enable occupancy-based control on existing and new HVAC system infrastructure to reduce energy consumption waste (Ahmad et al. 2021). Many of the camera-based (Abade et al. 2018; Cao et al. 2018; Ahmad et al. 2021) and wearable (Castro et al. 2017) solutions so far proposed can obtain training datasets without user interaction with solid performance. Although direct sensing methods have proven reliable performance for occupancy estimation problems, they still present serious issues, including privacy challenges, complex processing, expensive hardware, and installation feasibility (Cao et al. 2018). Environmental sensing is considered as a viable alternative to overcome the issues and limitations of direct sensing by measuring the effects of occupant presence in the building, such as changes in temperature, humidity, CO₂ concentration, among others.

Despite many existing occupancy estimation solutions in environmental sensing methods, there is still a lack of attention paid to generating a publicly accessible robust dataset that can be easily used to compare different machine learning algorithms. The existing solutions (Brennan et al. 2015; Schwee et al. 2019; Szczurek et al. 2017) have made their datasets publicly accessible. Still, most of the datasets are poorly documented with occupant ground truth such as occupant number range, which is considered key attributes to estimate the building occupation. This study proposed a novel multi-wireless device data model that incorporates environmental sensing to obtain indoor condition data and camera sensing to obtain occupant's numbers for training purposes. The primary objectives of this research are to collect quality datasets that can

be used to compare different machine learning algorithms for occupancy estimation problems to reduce air condition unnecessary energy consumption. The key contributions in this study include:

- collection and quality verification of indoor environmental conditions and occupant datasets
- evaluating machine learning algorithm on the proposed framework in living room
- modeling single room with two zones to monitor and improve air condition energy efficiency

To achieve this task, only a random forest regressor algorithm is used to verify dataset quality during data collection and occupancy detection and estimation tasks, although different machine learning algorithms could be integrated easily into the proposed approach. The performance evaluation measures show a high F-score value and minimum mean square error compared to the existing single environmental sensing baseline design (Candanedo and Feldheim 2016). Furthermore, air condition energy usage was monitored for 24 hours using a programmable thermostat, which reveals the excess usage of air condition by an average of 5 hours. To improve the current state, a one-room with two zones is modeled using DesignBuilder. The first zone depicts the HVAC programmable thermostat (Meana-Llorián et al. 2017). While in the second zone, occupants number variables are added in addition to variables proposed in (Meana-Llorián et al. 2017) to illustrate the random nature of the building occupants in a typical living room. The results analysis shows that the occupancy-based adaptive control can optimize energy usage by an average of 50% if the occupant participates in a demand response program and 25% without significantly compromising the comfort level (keeping the room temperature within the range of 19–21 °C).

It is essential to note that this study primarily focuses on estimating occupancy in enclosed rooms based on indoor environmental factors. The outdoor environmental factors are beyond the scope of this study. In addition, because temporal dependency in the data is not considered, the indoor environment is regarded as static. Furthermore, while the proposed solution design considered integrating into other building management services (such as intrusion detection, emergency system, and energy management), the actual integration with such systems is also beyond the scope of this study.

The rest of the study is organized as follows. “Literature review” section presents the related work on indoor occupancy detection and estimation. “Proposed framework” section presents the proposed framework. “Model training and testing” section provides results on occupancy

detection and estimation experiment. Subsequently, “Occupancy and air condition usage monitoring” section provides results on energy-saving potential simulations. “Conclusion” section presents research conclusions and future work.

Literature review

Integration of occupancy detection and estimation application in a control system is essential to support demand control ventilation. The integration has been achieved through various forms, including occupancy detection, occupancy estimation, occupancy recognition, and occupancy activity monitoring (Ahmad et al. 2021). Table 1 summarizes the existing studies on occupancy-based demand control ventilation.

The environmental sensing approach predicts room occupation by measuring indoor conditions. The sensor modality can be modified in many indoor sensing applications, including multi-sensing technology to observe concentrations of volatile organic compounds in the air. Cameras (infrared and optical cameras) are used alongside machine learning to carefully analyze and capture image frames for occupancy detection and estimation in commercial and residential buildings. The fusion modalities are considered to differentiate human occupancy and other object emitting thermal heat in the environment and support night vision prediction. The camera-based approach can accurately handle binary and multi-class occupancy predictions, up to 96% and 26% energy-saving potential.

Wearables and acoustic approaches use a product of tasks completed by other systems, which can track the occupancy location. ML model can obtain signal intensity

from statically positioned beacons in a target space to obtain a fine-grained occupant location and achieve the location accuracy of five meters. Activation of specific sensors with established positions has previously been used in passive infrared and acoustic sensors to obtain occupancy and location details using a heterogeneous sensing network. In these studies, a multimodal data fusion and deep learning method were employed to estimate occupancy.

Proposed framework

Dataset quality is a key factor for most real-time applications to drive effective modeling. A collaborative intelligence between environmental sensing and camera can support a collection of high-quality datasets. Datasets used are collected in residential building settings in a living room in a house consisting of five different rooms located at the Taman Teratai, Johor, Malaysia, with a tropical climate year-round with average temperatures starting from 25 to 30 °C throughout the year. The living room is being designed for occupant’s gatherings activities such as resting, eatery, watching TV, and other social gatherings. The indoor condition data is collected using sensors installed in the area of interest (see Table 2) to monitor indoor environmental conditions such as temperature, light illuminance, relative humidity, and CO₂ concentration.

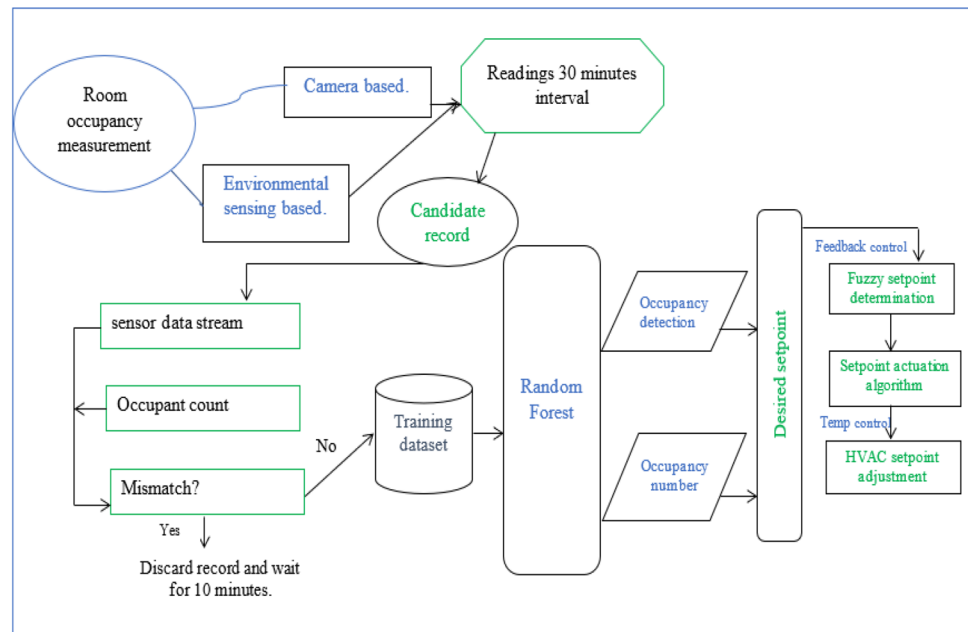
A high-level overview of the overall framework proposed in Fig. 2, consisting of the room occupancy measurement (camera and environmental sensing approaches), quality assessment of candidate record, recorded training dataset, and random forest model, performs prediction and sends the output value to the fuzzy system. These outputs are used to drive or generate the appropriate setpoint temperature for

Table 1 Occupancy-based demand control ventilation

Occupancy input	Technology	Study	Limitations
Occupancy detection	Environmental sensing	(Brennan et al. 2015; Dogan et al. 2020; Hänninen et al. 2017; Sarwar et al. 2021)	Poor quality training dataset, prone to external noise
Occupancy estimation	Environmental sensing	(Chen and Zhou 2020; Fareed et al. 2020; Gruber et al. 2014; Shahzad et al. 2021)	
Occupancy count/estimation	Camera	(Aryal and Becerik-Gerber, 2019; Cao et al. 2018; Waqar Saeed et al. 2021; Wei-long Ding et al. 2021a)	Processing power, poor coverage of the scene, occupancy overlapping, and privacy challenges
Occupancy identity	Acoustic	(Huang 2018; Kim et al. 2020; Salamone et al. 2017; Wang et al. 2021; Wei-long Ding et al. 2021b; Wu and Wang 2019; Zhang et al. 2020)	Poor quality training, background noise, and difficulty in modeling
Occupancy activity	Wearable	(Barut et al. 2020; Castro et al. 2017)	Device installation, limited device installation capacity, privacy concern
	Passive infrared	(Han et al. 2011; Sheikh Khan et al. 2021)	High false alarm, does not provide additional occupancy information, and poor quality of training dataset

Table 2 The various sensor data sources

Sensor	Description	Uncertainty	Unit	Data record
Temperature	Measure indoor temperature	1 °C	Degree Celsius	60-s interval
Relative humidity	Measure indoor relative humidity	±5%	Percentage	60-s interval
CO ₂	Measure indoor CO ₂ concentration level	300–1000 ppm: ±120 ppm	Parts per million (ppm)	60-s interval
Light	Measure illuminance indoor light levels	10–2000 lux range	Lux	60-s interval

Fig. 2 Proposed occupancy detection and estimation framework

the HVAC system using a fuzzy inference engine depending on the occupant's choice or behavior. Each variable in the dataset record stream (carbon dioxide, temperature, humidity, light intensity, humidity ratio, and occupant) essentially contains feature correlated to the occupant. Each time the sensors measure the readings of the occupant surrounding, the corresponding occupant number is captured by a camera and recorded in a dataset sheet. But the quality of these record streams can be influenced by a specific event, such as a door open during entrance or exit from the building. Each time such an event happens, it takes time for the record stream to rebuild the actual level that tallies with the occupants in the building.

The study adopted Candanedo and Feldheim (2016) for environmental sensing and Cao et al. (2018) for the camera. To ensure the quality dataset is recorded, 30-min interval is set between the previously registered record stream and candidate record.

Let denote dataset stream captured by sensors at a point in time to be $record(t)$ with a list of features $f_1 \dots f_n = record(t)$. Therefore, $f_1(record(t)) \dots f_n(record(t))$ is related to the occupant's presence in the building at a point in time t , where $f_i(t) \in (\hat{f}_i, \check{f}_i)$. Let denote occupant number with label $l(t)$ captured by the camera corresponding with the dataset stream obtained by environmental sensors. Therefore, the

complete dataset stream record collected for training at a time can be denoted as $r(t) = (record(t), l(t))$. At this point, dataset stream with no corresponding label $l(t)$ assigned is known as candidate record and a dataset stream with the corresponding label $l(t)$ assigned is known as a recorded record. Let a dataset of recorded records represent p . The objective is to verify the quality of record before been record in p to avoid a poor p by ensuring sensors data stream match corresponding occupant label $l(t)$ that was assigned by the camera using the rule-based classifier. In this study, a decision trees' algorithm proposed in (Chen and Zhou 2020) is deployed to simplify and optimize the analysis of the (if-then) rules.

Data collection and processing

The dataset collection starts on April 1st, 2021, to April 28th, 2021, using continuous readings. The only dataset with full-day readings and more than three streams' columns in a raw is considered. Additionally, records are swapped to avoid revealing occupancy schedules when datasets are published, as reported in Schwee et al. (2019) that CO₂ concentration can be deanonymized for susceptible privacy attacks. For odd days (Sunday, Tuesday, and Thursday), the two

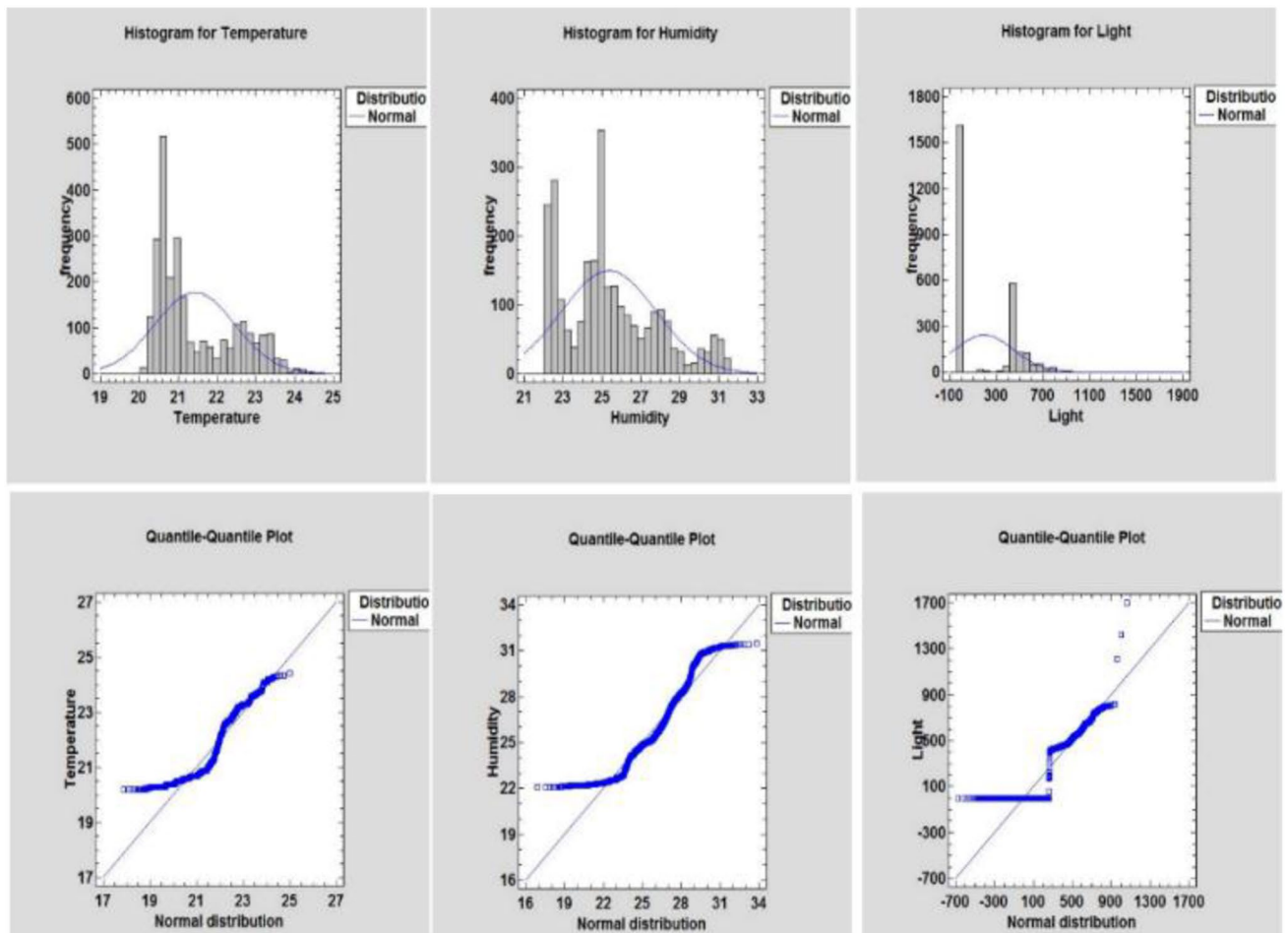


Fig. 3 Temperature, humidity, and light dataset normality check

consecutive rows' streams are randomly swapped, while for the even days (Saturday, Monday, and Wednesday), the first two rows' streams are swapped sequentially. Even though it is not considered in a recent study (Schwee et al. 2019), the study decided to introduce and compute the humidity ratio from the original dataset stream to improve occupancy estimation accuracy.

Although the graphical representation for assessing normality requires a great deal of expertise to prevent incorrect interpretations, the data for graphic interpretation is usually presented in histograms or Y and X vectors. According to Gregorutti et al. (2016), suppose Y is the variable that depends on the regression matrix of variables X . If $X(x_1, x_2, x_3, \dots, x_n)$ are jointly normal, then Y is said to be conditionally on X and $\mu = f(X)$ is normally distributed vector. Therefore Y and μ can be expressed as:

$$\begin{aligned} Y|X &\sim N(\mu = f(X), \sigma^2) \\ \mu = f(X) &= (\beta_0 + \beta_1 * x_1 + \beta_2 * x_2 + \dots \beta_n * x_n) \end{aligned} \quad (1)$$

The graphical presentation of the normality distribution of the sample dataset is conducted using the Q-Q plot (see Figs. 3 and 4).

The analysis indicates that the dataset points do not fully follow standard distribution consist of slight variance, requiring data analysis at this stage to achieve a Gaussian distribution. After manual inspection of the unfitted points, it was concluded that the skew is not caused by inaccurate sensor readings or recordings but is spontaneously created and is not inherently a concern and cannot affect the model prediction results. The distributions of unfitted points appear in all variables, with more extreme values in the CO_2 and occupancy variables. According to several experiments, about 1 in 340 observations in a regular distribution would be at least three standard deviations apart from the mean (Zittis 2017). However, in smaller datasets, random chance can contain extreme values. In other words, producing odd values naturally is routine, and there is nothing wrong with these data points. Thus, even though they are rare, they are a natural part of the data distribution.

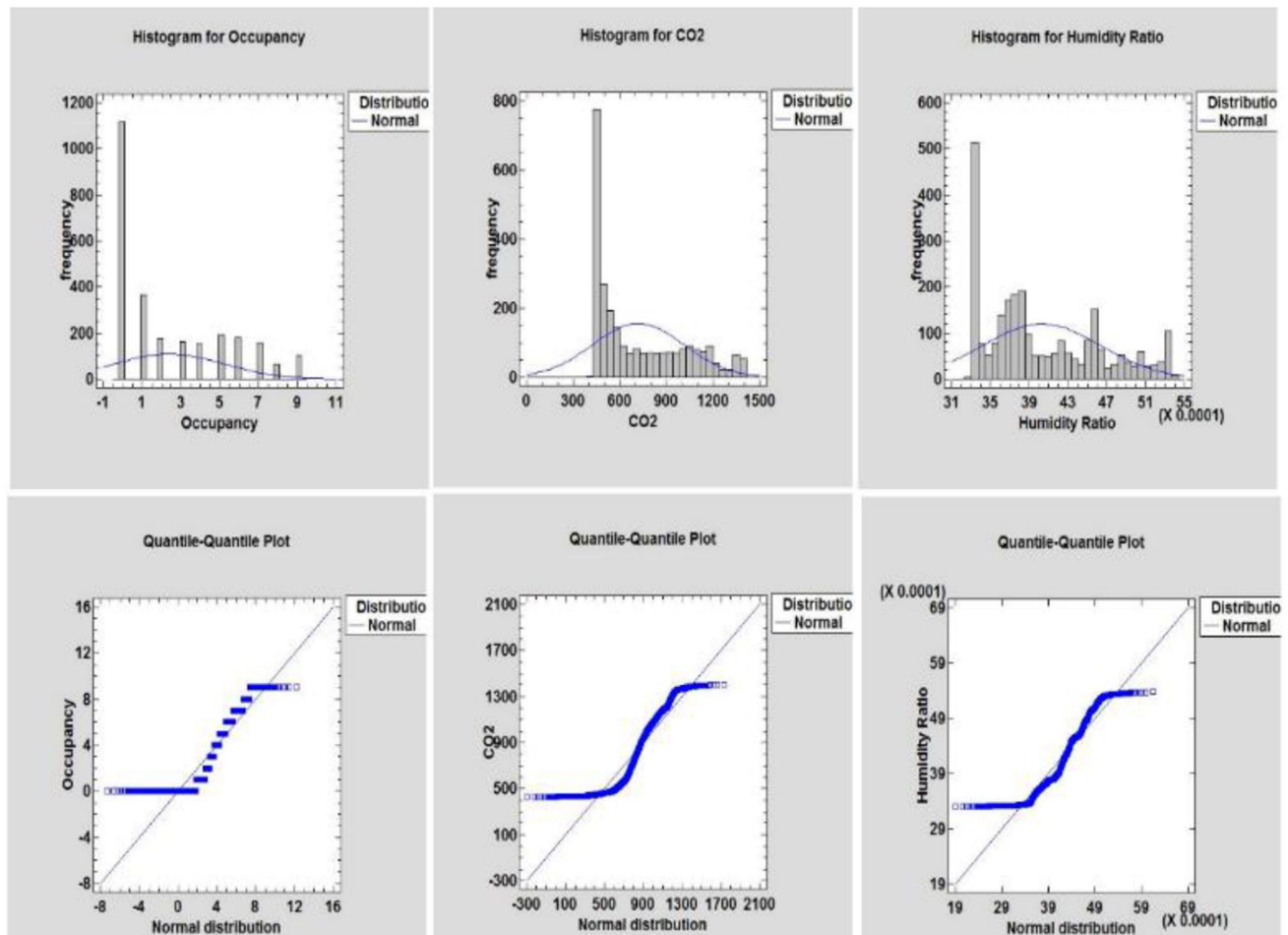


Fig. 4 Occupancy, CO₂, and humidity ratio dataset normality check

This study uses Pearson's product-moment coefficient (PPMC) metric for generating a correlation coefficient value. PPMC measures the strength of dependency between the variables x and y when given a set of paired (x,y) values between -1 and $+1$ (Candanedo and Feldheim 2016). Fig. 5 present the computed PPMC values using six variable parameters with values vary from -1 to 1 indicating a heavy positive correlation label shaded with white background color, followed by 0.9 shaded with red background color and so forth to 0.00 and -0.00 shaded with a green background color indicating a weak correlation between the variables. Predictors that are not correlated with predicting variables at all variables or with weak correlation values are most likely candidates to remove from the model using variable permutation importance measure known as feature selection. Furthermore, it is recommended that if two variables are highly correlated, only one of them should be considered to simplified models, and simpler models are easier to understand.

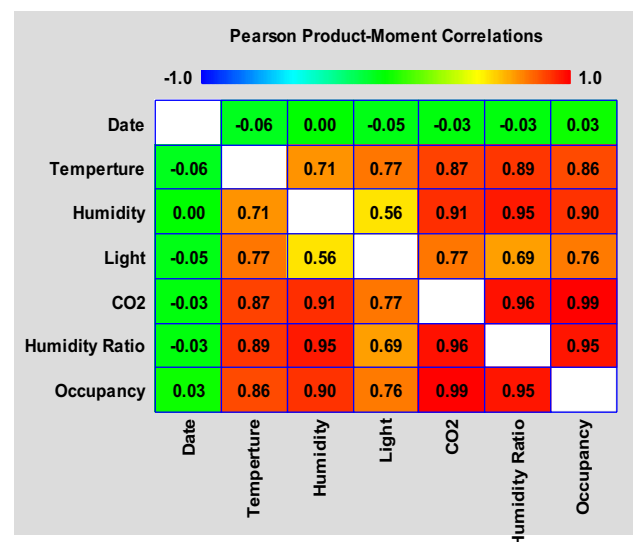


Fig. 5 Measured correlation values of the variables

Variable feature selection

Feature engineering is essential in developing ML models, which requires removing features with weak correlation before deploying the dataset sample into the model for evaluation. A variable importance measure metric in Gregorutti et al. (2016) is considered to remove uncorrelated variables parameters. The theory in Gregorutti et al. (2016) suggest for predicting variable Y and predictors $X = (X_1, \dots, X_p)$ be a vector of random variables. The rule \hat{f} in regression setting for predicting variable Y is a function that can be measure using the values in \mathbb{R} . The prediction error of \hat{f} can be defined by $\mathcal{R}(\hat{f}) = \left[\left(\hat{f}(X) - Y^2 \right) \right]$ and object is to calculate the conditional expectation $f(x) = \mathbb{E}[Y|X = x]$. Similarly, let $\mathcal{D}_n = \{(X_1, Y_1), \dots, (X_n, Y_n)\}$ be a set of learning of n replications of (X, Y) where $X_i = (X_{i1}, \dots, X_{ip})$. Since the true prediction error of \hat{f} is unknown in practice, observation of a test dataset ($\bar{\mathcal{D}}$) is considered for prediction, and therefore, $\bar{\mathcal{D}}$ can finally be presented as:

$$\bar{\mathcal{D}} : \hat{\mathcal{R}}(\hat{f}, \bar{\mathcal{D}}) = \frac{1}{\bar{\mathcal{D}}} \sum_{i: (X_i, Y_i \in \bar{\mathcal{D}})} Y_i - \hat{f}(Y_i - \hat{f}(X_i))^2 \quad (2)$$

Permutation variable importance is a model inspection technique in Breiman (2001) that has shown proficiency in non-linear estimators like our model and therefore adopted in this study. The technique considered predictors $X_i X_j$ as the critical predicting Y from (see equation 2). If the link between the feature $X_i X_j$ and Y is broken, the increase in prediction error score may be observed. The score value in the model reflects how much the model is dependent on the feature. This methodology has the advantage of being model agnostic, allowing it to be measured several times with various function permutations. To demonstrate this model, Breiman (2001) randomly permute the observations of the $X_i X_j$.

Formalizing the statistical permutation value calculation is as follows: define a group of out-of-bag samples $\{\bar{\mathcal{D}}_n^t = \mathcal{D}_n \setminus \bar{\mathcal{D}}_n^t, t = 1, \dots, n_{tree}\}$. Let $\{\bar{\mathcal{D}}_n^{tj}, t = 1, \dots, n_{tree}\}$ represent permuted out-of-bag samples by randomized permutations of the j -th variable's values in each out-of-bag subset. The variable X_j 's statistical permutation value is defined as:

$$\hat{I}(X_j) = \frac{1}{n_{tree}} \sum_{t=1}^{n_{tree}} \left[\hat{\mathcal{R}}(\hat{f}_t, \bar{\mathcal{D}}_n^{tj}) - \hat{\mathcal{R}}(\hat{f}_t, \bar{\mathcal{D}}_n^t) \right] \quad (3)$$

This quantity is the statistical equivalent of the permutation importance measure $\hat{I}(X_j)$ recently formalized by Zhu et al. (2015). Let $(X_j) = (X_1, \dots, X_j', \dots, X_p)$ be the random vector such that X_j' is an independent replicate of X_j that is also independent of Y and all other predictors, and the permutation significance measure is provided by:

Table 3 Predicting variable versus independent variable correlation index

Variables	Correlation index
Occupancy + Date	0.03
Occupancy + Temperature	0.86
Occupancy + Humidity	0.90
Occupancy + Light	0.76
Occupancy + CO ₂	0.99
Occupancy + Humidity Ratio	0.95
Occupancy + Occupancy	1

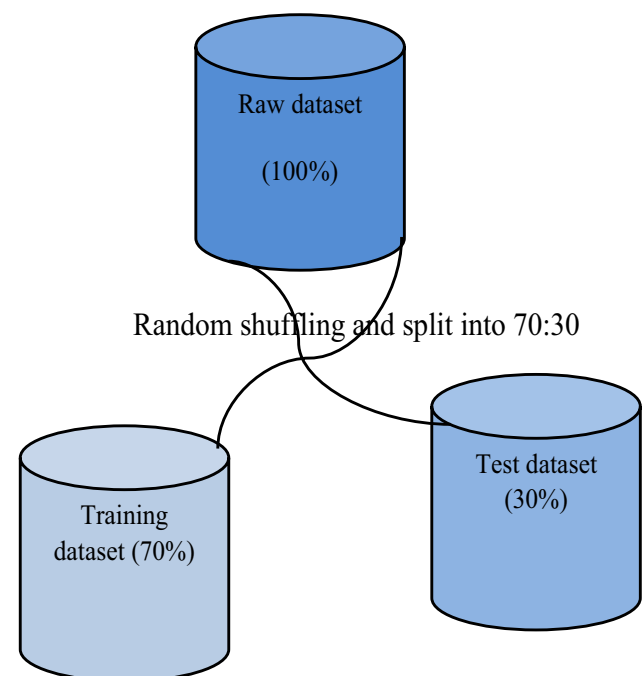


Fig. 6 Ratio of training and test dataset

$$I(X_j) = \mathbb{E} \left[(Y - f(X_{(j)}))^2 \right] - \mathbb{E} \left[(Y - f(X))^2 \right] \quad (4)$$

In the expression of $\hat{I}(X_j)$, the permutation values of X_j mimics the identical and independent duplicate of the distribution of (X_j) in $I(X_j)$. Thus, equation 4 can compute the correlation index value of predicting variable and independent variable as presented in Table 3.

The predictor's correlation index in relation to predicting variable is computed and displayed in Table 3 to visualize weak correlation values. As presented in Table 3, the variable predictor Date demonstrates a weak correlation index and is removed from the original dataset. The remainder of the variables can feed the model to train the machine learning model and measure its accuracy against the test dataset.

Model training and testing

Typically, datasets are split in form of the train and test ratio during the model training when ML algorithms are employed to make predictions on data to measure their prediction performance. The technique is straightforward and quick for assessing model prediction performance on various ML methods and choosing among the optimal methods that fit the model prediction problem. The technique entails shuffling and splitting the original dataset into training and testing in a ratio, such as 70:30 see (Fig. 6). The first portion, known as the training dataset, is used to match the model. The second portion, known as the test dataset, is used as input to variables' dataset to feed the model to test prediction and measure the prediction outcomes.

Candidate model

Random forest (RF) model is chosen for inquiry to further measure its performance for both binary classification (occupancy detection) and regression prediction problems (occupancy estimation). This model is not very complex; its existing implementation and recent developments have shown solid performance in many areas. The implementations in this work use the scikit-learn Python library, and details about default algorithm settings can be found in the library documentation ("The Python Standard Library" 2021).

RF is a collection of various decision trees that are applied sequentially from a root (parent) node to a terminal (or child) node to predict the behavior described by trained data (Rodriguez-Galiano et al. 2014). This technique provides several conditional rules as easy as comparing a sensor reading to a threshold to match data samples by related traits. Each decision tree employs bootstrap sampling, also known as bagging (Breiman 1996), which essentially used two-thirds of the training samples for prediction and the remainder for evaluation of prediction accuracy for both deep or very deep trees. This implies each tree in RF is working against the same target but is given separate portions of the training data to learn from. The outcomes from all the trees are added together to generate the results. These rules influence how the models handle bias and uncertainty in their forecasts. The binary classifier uses CO₂ as predicting variable to tell whether a room is occupied or vacant. The performance evaluation for RF binary classifier is presented in Table 4. The number of decision trees that the model can use to match the data is generated from classification and recursive splitting from the dataset during analysis.

The evaluation is carried out to test and verify the performance of the model against new data. This is essential in many cases, especially on the control system where the model's performance can exhibit different behavior on the

Table 4 RF binary occupancy prediction results using CO₂ data

Score bin	Positive rate	Negative rate	Fraction above threshold	Accuracy	F1 score	Precision	Recall	Negative precision	Negative recall	Cumulative AUC
(0.900,1.000)	1064	1	0.570	0.987	0.988	0.999	0.978	0.970	0.999	0.000
(0.800,0.900)	9	1	0.576	0.991	0.992	0.998	0.986	0.981	0.997	0.001
(0.700,0.800)	0	0	0.576	0.991	0.992	0.998	0.986	0.981	0.997	0.001
(0.600,0.700)	4	1	0.578	0.993	0.994	0.997	0.990	0.986	0.996	0.003
(0.500,0.600)	6	2	0.583	0.995	0.995	0.995	0.995	0.994	0.994	0.005
(0.400,0.500)	0	0	0.583	0.995	0.995	0.995	0.995	0.994	0.994	0.005
(0.300,0.400)	4	1	0.585	0.996	0.997	0.995	0.999	0.999	0.992	0.006
(0.200,0.300)	1	5	0.589	0.994	0.995	0.990	1.000	1.000	0.986	0.013
(0.100,0.200)	0	13	0.596	0.987	0.989	0.978	1.000	1.000	0.969	0.029
(0.000,0.100)	0	755	1.000	0.583	0.736	0.583	1.000	1.000	0.000	0.999

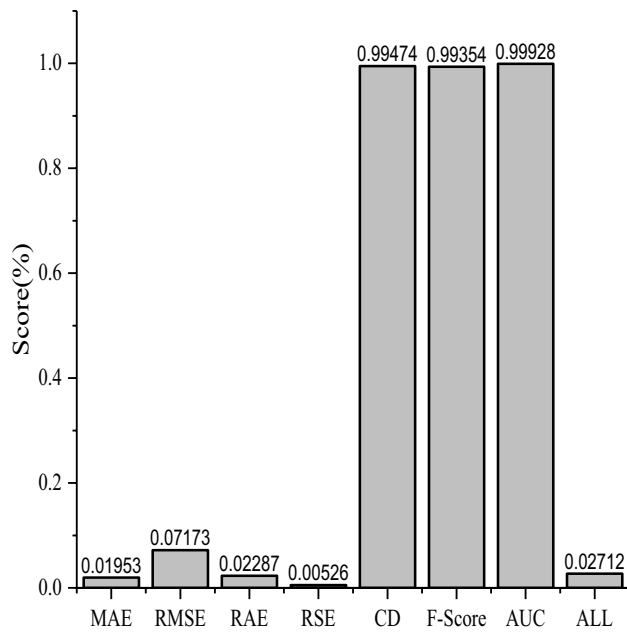


Fig. 7 Occupancy estimation model performance evaluation

new dataset. Therefore, the scoring bin in Table 4 holds the dataset record split into a training and a testing dataset. The binary classifier prediction performance on ranges from 58.3 to 99.6% accuracy, 73.6 to 99.7% F1 score, 58.3 to 99.9% precision, and 97.8 to 100% recall.

Occupancy detection using binary classification cannot provide enough information about room occupancies which is essential to optimize energy proportional to the number of occupants in the building. For this reason, the proposed RF model performs dataset recursive splitting and regression to estimate room occupancies using five indoor conditions. Fig. 7 represents the model performance analysis using different performance measures.

Typically, the single metric alone cannot provide enough information for model performance. Therefore, other members of metrics are considered including:

$$\text{F-Score} = 2 \times \left(\frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \right) \quad (5)$$

Ultimately, it is essential to have an overall metric to trade-off the precision and the recall performance by measuring a single grade value score. Thus, it makes sense to merge the precision and recall metrics (see equation 5). The overall F-score performance for RF is excellent, achieving up to 99.3% (see Fig. 7).

Mean absolute error (MAE) refers to the magnitude of the difference between the model prediction observation and the actual value of that observation which is calculated for the whole group, and therefore, the error between the

actual value and the predicted value is 0.0193 (see Fig. 7). Mathematically MAE can be presented as:

$$\text{MAE} = \left(\frac{\sum_{i=1}^n \text{abs}(y_i - \lambda(x_i))}{n} \right) \quad (6)$$

Root mean square error (RMSE) demonstrates how far projections differ from observed true values. Whether during testing or cross-validation, the residual difference between prediction and ground truth for each data point is referred to as RMSE. The RMSE in RF model is 0.071 (see Fig. 7). Mathematically RMSE can be presented as:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n \|y(i) - \hat{y}(i)\|^2}{N}} \quad (7)$$

Relative squared error (RSE) is used to evaluate model efficiency by comparing it to that of a basic predictor. The RSE in RF performance is 0.0052 (see Fig. 7). Mathematically RSE can be presented as:

$$E_i = \left(\frac{\left(\sum_{j=1}^n P_{ij} - T_j \right)^2}{\sum_{j=1}^n \left(T_j - \bar{T} \right)^2} \right) \quad (8)$$

Relative absolute error (RAE) is expressed as a ratio when a mean error (residual see equation 9) is opposed to errors produced by a negligible or naive model. The model RAE is 0.022 (see Fig. 7). Mathematically RAE can be presented as:

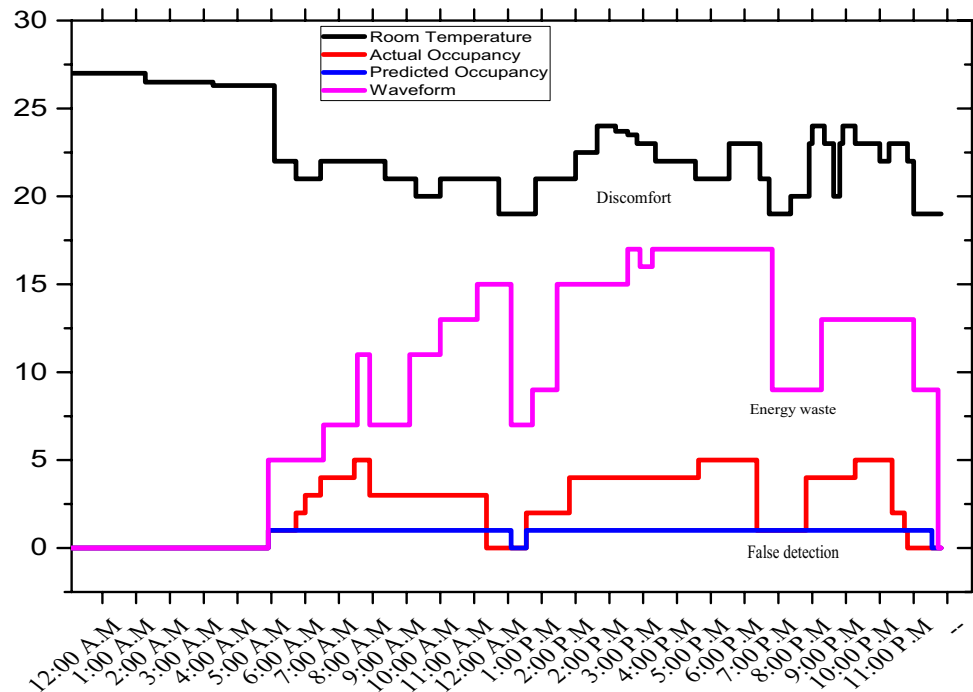
$$E_i = \frac{\sum_{j=1}^n |P_{ij} - T_j|}{\sum_{j=1}^n |T_j - \bar{T}|} \quad (9)$$

Coefficient of determination (CD), also referred to as R2, clarifies how much a model will perform when replicating observed results. It provides information on the probability of possible events occurring within the expected outcomes. The R2 value produced by RF model is 99.4% (see Fig. 7). Mathematically CD can be presented as:

$$\text{R2} = \frac{n(\sum xy) - (\sum x)(\sum y)}{\left[n \sum x^2 - (\sum x)^2 \right] \left[n \sum y^2 - (\sum y)^2 \right]} \quad (10)$$

Average log loss (ALL) uses to evaluate the model prediction efficiency based on the likelihood of a record being categorized in a particular class and then assign the data point as one of two classes (1 or 0) depending on whether the probability exceeded a threshold value. The proposed RF model achieved R2 score of 0.027 (see Fig. 7). Mathematically ALL can be presented as:

Fig. 8 24-h data on indoor occupants, temperature, and air condition operation



$$\text{logloss} = \frac{1}{n} \sum_{i=1}^n \text{logloss}_i \quad (11)$$

where:

$$\begin{aligned} \text{logloss}_i &= -[y_i \ln p_i + (1 - y_i) \ln (1 - p_i)] \\ \text{logloss} &= -\frac{1}{n} \sum_{i=1}^n [y_i \ln p_i + (1 - y_i) \ln (1 - p_i)] \end{aligned}$$

Occupancy and air condition usage monitoring

The main objective of binary control is to ensure air condition is turned off when the room is vacant to avoid unnecessary energy consumption. Figure 8 plots the information obtained during the room condition monitoring campaign. Notice in Fig. 8 from 12:00 to 5:00 AM, the living room is vacant, air condition is off, and room temperature is high. Because of the challenge to obtain the aggregated energy consumption data in a time series as there are many mixtures of energy consumption from various appliances and built-in facilities, a watt analyzer is attached to a power outlet to monitor the waveform of air condition power consumption in a time series.

Notice from 6:00 AM occupancy keeps on rising from 1 to 5 and drops at 8:00 AM. From 11:00 AM to 1:00 PM, 6:00 to 8:00 PM, and 11:00 PM, the room is vacant, but the programmable controller predicts that the room is occupied during this time. This is because the controller

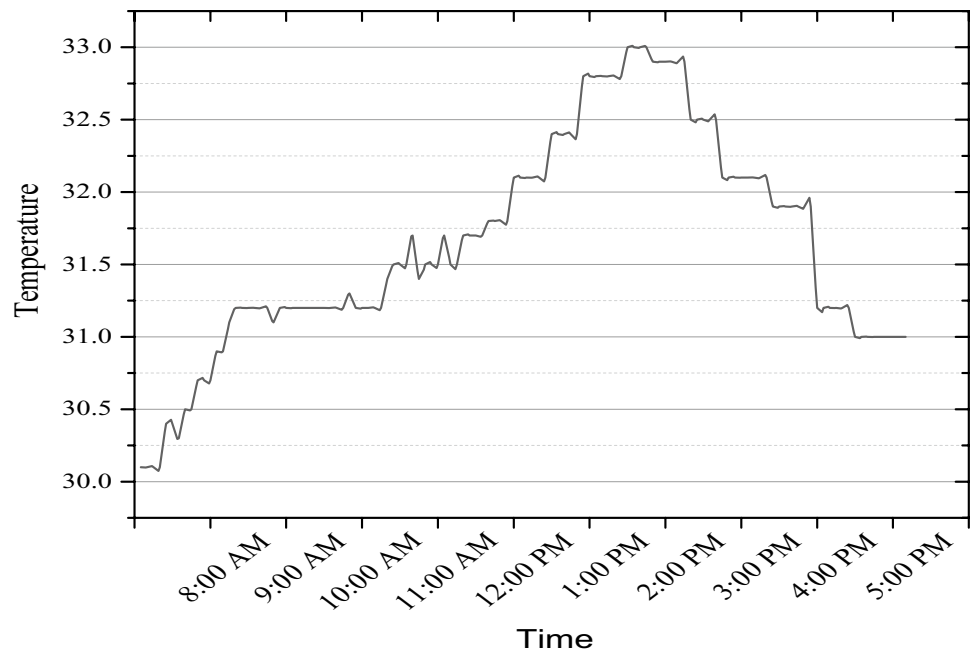
works based on the occupancy schedule assumption, which is modified due to MCO policy. Besides MCO policy, further investigation reveals that indoor CO₂ is used as an indicator of the room being occupied. Thus, when no occupants leave the building, it takes a long time for CO₂ to disappear, which also contributes to the poor performance of the controller. This results in excess energy consumption for an average of 6 h a day. The total excess energy consumption depends on room temperature and rate of air condition operation.

Furthermore, notice the rise of a waveform of air condition operation from 1:00 to 7:00 PM. This is because, during this period, the outdoor temperature is high, affecting the function of an air conditioner to maintain the desired occupant comfort. However, despite the increase in air condition operation, discomfort can be observed during this period.

Occupancy estimation-based air condition control

DesignBuilder is used to model two separate zones, each attached with a modified HVAC system of 10 kW, using RTP tariff to demonstrate the effectiveness of occupancy estimation on air condition control. The first zone uses the controller proposed IoFClima in Meana-Llorián et al. (2017), and the second zone uses the proposed controller. The final model is sent to EnergyPlus together with local weather. The controller's goal is to adjust the setpoint temperature to attain the desired comfort based on the predicted

Fig. 9 Ambient temperature condition in Johor, Malaysia



number of occupants in the building. Desired temperature, energy consumption, and comfort satisfaction achieved are used as evaluation indices to evaluate the performance of the proposed model under the following three scenarios:

- The occupants are worried about the energy consumption cost and are knee to participate in the demand response programs. So, the controller adjusts the setpoint temperature remarkably, especially during very hot and extreme temperatures, to attain the desired comfort as possible.
- The occupants worried about the thermal comfort satisfaction. Thus, the controller slightly adjusts the setpoint temperature.
- The occupants balance thermal comfort satisfaction and energy consumption cost. Accordingly, the controller reacts moderately to trade-off energy consumption with thermal comfort; the setpoint changes are not as remarkable as in scenario 1 and slightly as in scenario 3.

Setpoint temperature control

It is important to note that 19 °C is considered the desired setpoint with ten constant occupants throughout the experiments. Fig. 9 represents the time series room temperature in Johor Bahru, Malaysia, with air-conditioning turned off. Notice in Fig. 9, from 8:00 to 9:00 AM, the temperature is considered hot, from 10:00 to 11:00 AM and 3:00 to 5:00 PM are considered very hot, and 12:00–2:00 PM is considered extremely hot. The controller's objective is to adjust the setpoint according

to the number of occupants and temperature conditions to achieve the desired comfort without risking the occupants leaning.

Notice in Fig. 10, both controllers keep the indoor temperature within the range of 24 °C. This is because the occupant's choice is to save energy and avoid the aggressive air condition operation. However, notice in Fig. 11 and Fig. 12, the proposed controller was able to adjust setpoint temperature to make indoor temperature closer to the desired comfort. Except during extremely hot temperatures, the controller finds it challenging to maintain. Notice the performance of the proposed controller is better than IoFClime in stabilizing the room temperature and waveform frequency is lower than IoFClime. The proposed controller is designed with a feedback loop that updates the controller about the room temperature condition. Whenever the room temperature goes closer to the desired comfort, the speed of the air conditioner is reduced by adjusting the setpoint temperature, unlike the IoFClime that used deadband setpoint control.

Comfort analysis

The interpretation of thermal comfort analysis focused only on a proposed controller. The scenarios considered and the result is presented in Fig. 13. Nineteen degrees Celsius is chosen as desired indoor comfort for the ten occupancies. Every 30 min, the occupancy feedback on comfort feeling is analyzed in the zone, passed to the controller through a feedback loop, and expected to react based on the information received. Notice in the

Fig. 10 Scenario 1

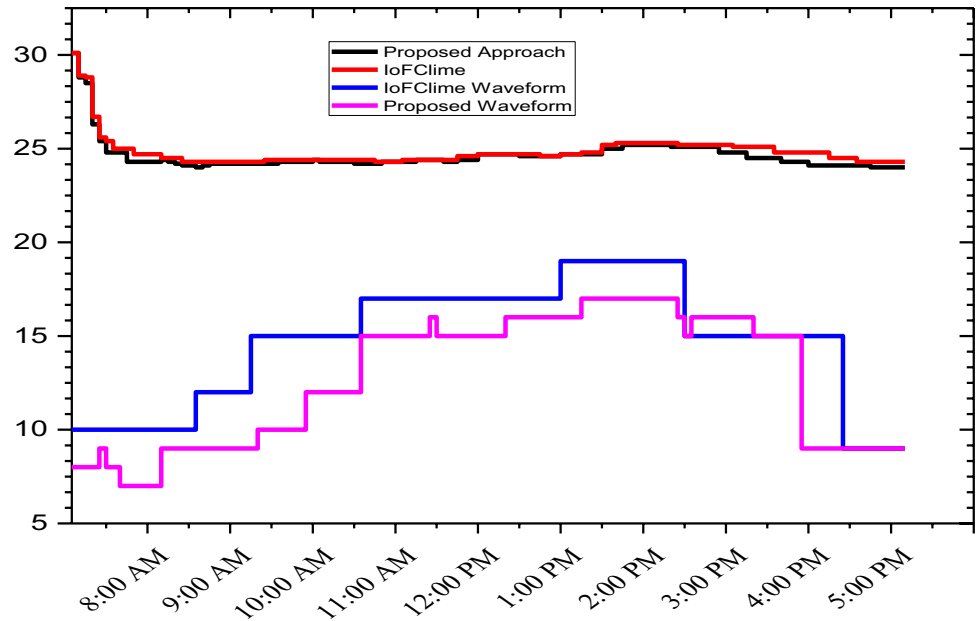
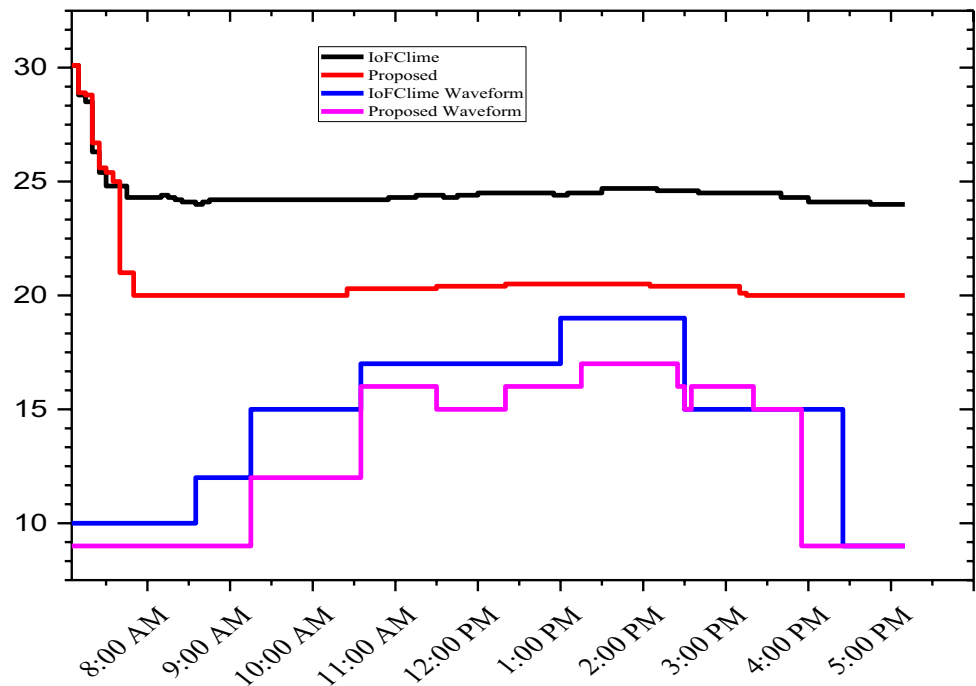


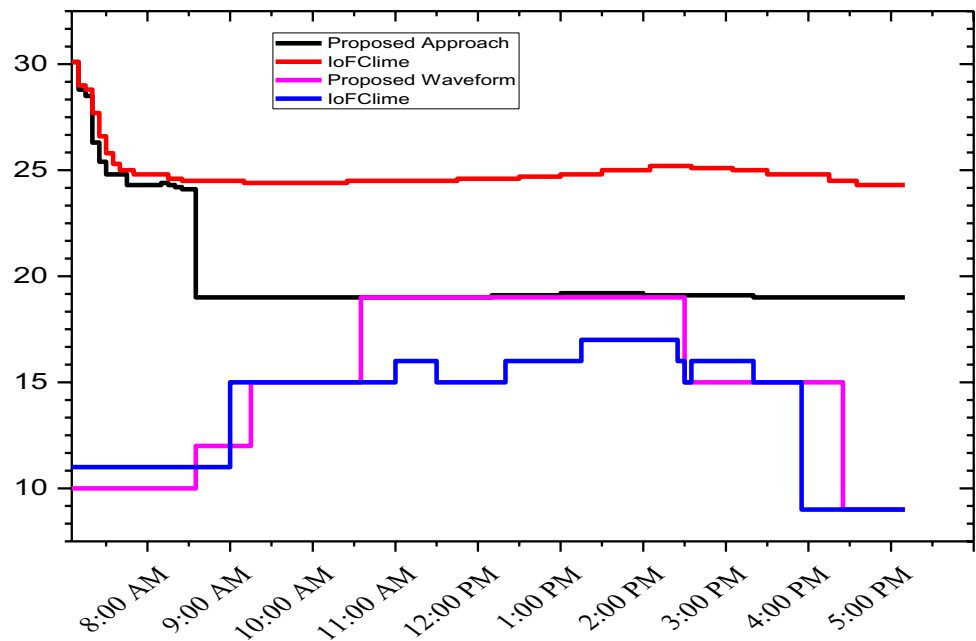
Fig. 11 Scenario 2



first scenario, the thermal comfort perception is as follows: 10% report feeling normal, 35% report feeling warm, 40% report feeling hot, and 15% report feeling very hot throughout the ventilation activities. In the second scenario, 88% report feeling normal, and 12% report warm comfort throughout the cooling activities. In the third scenario, 78% report average comfort while 22% report warm comfort through the cooling activities.

Energy consumption

In scenarios 2 and 3, the electricity tariff variation has no greater influence on energy consumption as occupancy gives thermal comfort preferences over energy usage. The air conditioner must maintain zone temperature relatively closer to the initialized setpoint (desired temperature). During the peak hour, compressor is turned on and stays on until the zone temperature drops to the closer level as the desired

Fig. 12 Scenario 1

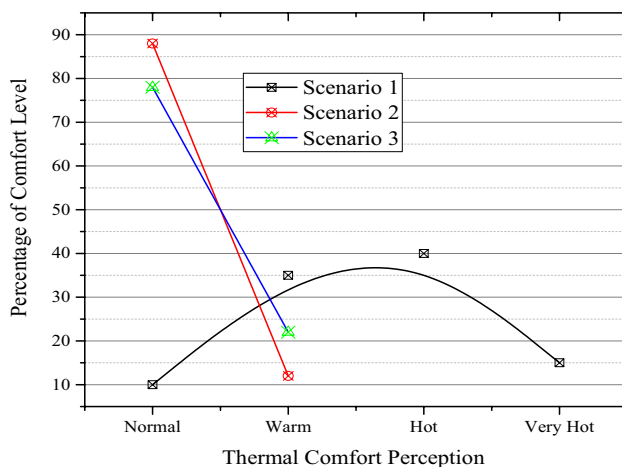
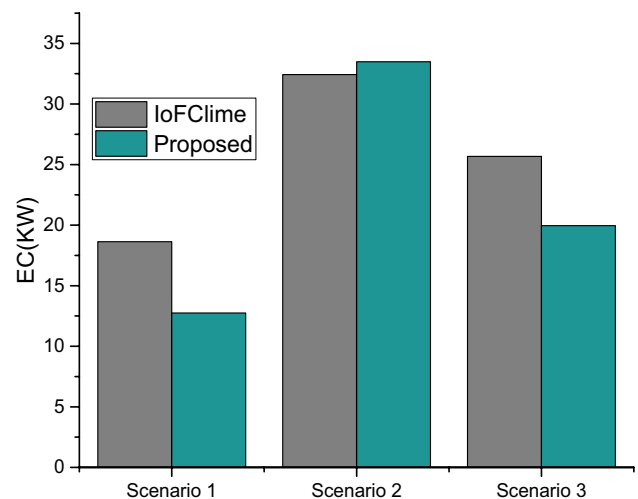
temperature is set on the thermostat. The compressor turns off after the desired temperature is reached, and the zone temperature rises again. A cycle of the air conditioner is determined by the compressor turning on and off.

An air conditioner's cycle time is when it takes to operate to keep the zone at the desired temperature. As demonstrated in Fig. 11 and Fig. 12, the compressor works significantly for longer periods to lower the zone temperature, increasing the cycle time. The existing approach's energy consumption is 18.6328 kW, 32.4284 kW, and 25.6724 kW, while the proposed approach consumed 12.7467 kW, 33.4933 kW, and 19.96 kW in three settings, as indicated in Fig. 14. The energy consumed in scenarios 1 and 3 for the existing controller is much greater than the proposed

approach's energy consumed. Since our approach uses a feedback loop, it provides continuous operation (long cycles); hence, it consumes less energy by 50% in the first scenario and 25 % in the third scenario than the existing IoFCIime technique.

Conclusion

Indoor occupancy information and modeling have been considered as key contributors for building energy efficiency, and research highlights the requirements for further investigation in this application area. Particularly occupancy prediction status in the buildings can save an average of 50%

**Fig. 13** Analysis of occupant thermal comfort perception of the proposed approach**Fig. 14** Comparison of energy consumption under three settings

unnecessary energy consumption, thereby reducing the demand for building energy generation and consumption.

Research on various technologies has been researched, including cameras, wearables, and passive infrared sensors, to modify or replace existing building infrastructure to optimize building energy usage. These efforts can be classified into direct approaches based on technologies such as cameras and wearables that solely rely on direct contact with occupants to report the presence. On the other hand, the indirect or environmental sensing approach inferred occupants' status by measuring the effects of occupant presence in the building. These include changes in temperature, humidity, CO₂ concentration, among others. Even though the direct approach has proven a solid performance for counting the occupancy, environmental sensing is considered to be a viable alternative to overcome the weakness of the direct approach that presents serious issues, such as privacy, complex processing, expensive hardware, and installation feasibility.

Despite many solutions in environmental sensing to overcome the problem of occupancy detection and estimation, much attention has not been paid to publicly accessible robust datasets that enable the comparison of different algorithms to be easily compared. The proposed framework that uses collaborative intelligence obtained a high-quality dataset, and RF model is trained to count occupant numbers in the building. The proposed solution is tested in the living room with a prototype system, and the model's prediction results obtained achieved a performance accuracy of 99.3%.

Three different thermostat settings on air condition were used to control the indoor comfort using a two-zone ventilation system to keep indoor comfort consistent with the number of occupants. The simulation is conducted using EnergyPlus software with local climate data. The Fanger model's thermal comfort analysis demonstrates poor comfort performance in the first scenario, satisfactory performance in the third scenario, and remarkable performance in the second scenario. Furthermore, it demonstrates up to 50% of energy can be saved in the first scenario (if the occupant is willing to compromise his comfort) and 25% in the third scenario (if the occupant wishes to balance energy consumption with thermal comfort).

Acknowledgements The authors wish to thank Monash University Malaysia for providing a valuable resource to carry out this study.

Author contribution All the authors participate and contributed to experiments, writing, and organizing the content of the paper.

Conceptualization and experiment: Muhammad Said Aliero and Adel N. Toosi. Literature review and conclusion: Muhammad Fermi Pasha. Resource and proofreading: Imran Ghani and Muhammad Fermi Pasha.

Data availability <https://github.com/MSAliero/Occupancy-Measurmnts-Data>.

Declarations

Ethics approval and consent to participate Graduate research, Monash University obtains ethical approval.

Competing interests The authors declare no competing interests.

References

- Abade B, Perez Abreu D, Curado M (2018) A non-intrusive approach for indoor occupancy detection in smart environments. *Sensors* 18(11):3953. <https://doi.org/10.3390/s18113953>
- Ahmad J, Larijani H, Emmanuel R, Mannion M (2021) Occupancy detection in non-residential buildings – a survey and novel privacy preserved occupancy monitoring solution. *Appl Comput Inform* 17(2):279–295. <https://doi.org/10.1016/j.aci.2018.12.001>
- ALiero MS, Qureshi KN, Pasha MF, Jeon G (2021) Smart home energy management systems in internet of things networks for green cities demands and services. *Environ Technol Innov* 22:101443. <https://doi.org/10.1016/j.eti.2021.101443>
- Aryal A, Becerik-Gerber B (2019) A comparative study of predicting individual thermal sensation and satisfaction using wrist-worn temperature sensor, thermal camera and ambient temperature sensor. *Build Environ* 160. <https://doi.org/10.1016/j.buildenv.2019.106223>
- Barut O, Zhou L, Luo Y (2020) Multitask LSTM model for human activity recognition and intensity estimation using wearable sensor data. *IEEE Internet Things J* 7(9):8760–8768. <https://doi.org/10.1109/jiot.2020.2996578>
- Breiman L (2001) Random Forests. *Mach Learn* 45. <https://doi.org/10.1023/A:1010933404324>
- Breiman L (1996) Bagging predictors. *Mach Learn* 24:123–140
- Brennan C, Taylor GW, Spachos P (2018) Designing learned CO₂-based occupancy estimation in smart buildings. *IET Wireless Sensor Systems* 8(6):249–255. <https://doi.org/10.1049/iet-wss.2018.5027>
- Candanedo LM, Feldheim V (2016) Accurate occupancy detection of an office room from light, temperature, humidity and CO₂ measurements using statistical learning models. *Energy Build* 112:28–39. <https://doi.org/10.1016/j.enbuild.2015.11.071>
- Cao N, Ting J, Sen S, Raychowdhury A (2018) Smart sensing for HVAC control: collaborative intelligence in optical and IR cameras. *IEEE Trans Industr Electron* 65(12):9785–9794. <https://doi.org/10.1109/tie.2018.2818665>
- Castro D, Coral W, Rodriguez C, Cabra J, Colorado J (2017) Wearable-based human activity recognition using an iot approach. *J Sens Actuator Netw* 6(4):28. <https://doi.org/10.3390/jsan6040028>
- Chen Y, Zhou Y (2020) Machine learning based decision making for time varying systems: parameter estimation and performance optimization. *Knowl-Based Syst* 190:105479. <https://doi.org/10.1016/j.knsys.2020.105479>
- Dogan B, Ben Jebli M, Shahzad K, Farooq TH, Shahzad U (2020) Investigating the effects of meteorological parameters on COVID-19: case study of New Jersey United States. *Environ Res* 191:110148. <https://doi.org/10.1016/j.envres.2020.110148>
- Fareed Z, Iqbal N, Shahzad F, Shah SGM, Zulfiqar B, Shahzad K, Hashmi SH, Shahzad U (2020) Co-variance nexus between COVID-19 mortality, humidity, and air quality index in Wuhan, China: new insights from partial and multiple wavelet coherence. *Air Qual Atmos Health* 13:673–682. <https://doi.org/10.1007/s11869-020-00847-1>

- Gregorutti B, Michel B, Saint-Pierre P (2016) Correlation and variable importance in random forests. *Stat Comput* 27(3):659–678. <https://doi.org/10.1007/s11222-016-9646-1>
- Gruber M, Trüschel A, Dalenbäck J-O (2014) CO₂ sensors for occupancy estimations: potential in building automation applications. *Energy Build* 84:548–556. <https://doi.org/10.1016/j.enbuild.2014.09.002>
- Han J, Choi CS, Lee I (2011) More efficient home energy management system based on ZigBee communication and infrared remote controls. *IEEE Transactions on Consumer Electronics* 57(1):85–89. <https://doi.org/10.1109/TCE.2011.5735485>
- Hänninen O, Canha N, Kulinkina AV, Dume I, Deliu A, Mataj E, . . . Egorov AI (2017) Analysis of CO₂ monitoring data demonstrates poor ventilation rates in Albanian schools during the cold season. *Air Qual Atmos Health* 10(6): 773–782. <https://doi.org/10.1007/s11869-017-0469-9>
- Huang Q (2018) Occupancy-driven energy-efficient buildings using audio processing with background sound cancellation. *Buildings* 8(6):78. <https://doi.org/10.3390/buildings8060078>
- Iqbal N, Fareed Z, Shahzad F, He X, Shahzad U, Lina M (2020) The nexus between COVID-19, temperature and exchange rate in Wuhan city: new findings from partial and multiple wavelet coherence. *Sci Total Environ* 729:138916. <https://doi.org/10.1016/j.scitotenv.2020.138916>
- Kim J, Min K, Jung M, Chi S (2020) Occupant behavior monitoring and emergency event detection in single-person households using deep learning-based sound recognition. *Build Environ* 181:107092. <https://doi.org/10.1016/j.buildenv.2020.107092>
- Meana-Llorián D, González GarcíaPelayo G-Bustelo CBC, Cueva Lovelle JM, García-Fernandez N (2017) IoFClime: the fuzzy logic and the internet of things to control indoor temperature regarding the outdoor ambient conditions. *Futur Gener Comput Syst* 76:275–284. <https://doi.org/10.1016/j.future.2016.11.020>
- Rodríguez-Galiano VF, Chica-Olmo M, Chica-Rivas M (2014) Predictive modelling of gold potential with the integration of multi-source information based on random forest: a case study on the Rodalquilar area, Southern Spain. *Int J Geogr Inf Sci* 28(7):1336–1354. <https://doi.org/10.1080/13658816.2014.885527>
- Salamone F, Belussi L, Danza L, Galanos T, Ghellere M, Meroni I (2017) Design and development of a wearable wireless system to control indoor air quality and indoor lighting quality. *Sensors* 17(5):1021. <https://doi.org/10.3390/s17051021>
- Sarwar S, Shahzad K, Fareed Z, Shahzad U (2021) A study on the effects of meteorological and climatic factors on the COVID-19 spread in Canada during 2020. *J Environ Health Sci Eng*, 1–9. <https://doi.org/10.1007/s40201-021-00707-9>
- Schwee JH, Johansen A, Jorgensen BN, Kjaergaard MB, Mattera CG, Sangogboye FC, Veje C (2019) Room-level occupant counts and environmental quality from heterogeneous sensing modalities in a smart building. *Sci Data* 6(1):287. <https://doi.org/10.1038/s41597-019-0274-4>
- Shahzad F, Shahzad U, Fareed Z, Iqbal N, Hashmi SH, Ahmad F (2020a) Asymmetric nexus between temperature and COVID-19 in the top ten affected provinces of China: a current application of quantile-on-quantile approach. *Sci Total Environ* 736:139115. <https://doi.org/10.1016/j.scitotenv.2020.139115>
- Shahzad K, Farooq TH, Doğan B, Zhong Hu L, Shahzad U (2021) Does environmental quality and weather induce COVID-19: case study of Istanbul, Turkey. *Environ Forensic*: 1–12. <https://doi.org/10.1080/15275922.2021.1940380>
- Shahzad K, Shahzad U, Iqbal N, Shahzad F, Fareed Z (2020b) Effects of climatological parameters on the outbreak spread of COVID-19 in highly affected regions of Spain. *Environ Sci Pollut Res Int* 27(31):39657–39666. <https://doi.org/10.1007/s11356-020-10551-3>
- Shakoor A, Chen X, Farooq TH, Shahzad U, Ashraf F, Rehman A, Sahar NE, Yan W (2020) Fluctuations in environmental pollutants and air quality during the lockdown in the USA and China: two sides of COVID-19 pandemic. *Air Quality, Atmosphere & Health* 13(11):1335–1342. <https://doi.org/10.1007/s11869-020-00888-6>
- Sheikh Khan D, Kolarik J, Anker Hviid C, Weitzmann P (2021) Method for long-term mapping of occupancy patterns in open-plan and single office spaces by using passive-infrared (PIR) sensors mounted below desks. *Energy Build* 230:110534. <https://doi.org/10.1016/j.enbuild.2020.110534>
- Szczurek A, Maciejewska M, Pietrucha T (2017) Occupancy determination based on time series of CO₂ concentration, temperature and relative humidity. *Energy Build* 147:142–154. <https://doi.org/10.1016/j.enbuild.2017.04.080>
- The Python Standard Library (2021) [Online] From; <https://docs.python.org/3/library/>. Accessed 6 Dec 2021
- Wang C, Jiang J, Roth T, Nguyen C, Liu Y, Lee H (2021) Integrated sensor data processing for occupancy detection in residential buildings. *Energy Build* 237:110810. <https://doi.org/10.1016/j.enbuild.2021.110810>
- Waqar Saeed ZA, Jehangiri AI, Mohamed N, Umar AI, Ahmad J (2021) A fault tolerant data management scheme for healthcare internet of things in fog computing. *KSII Trans Internet Inf Syst* 15(1):35–57. <https://doi.org/10.3837/tiis.2021.01.003>
- Wei-long Ding X-JZ, Bin Xu, Yan Xu, Chen K, Wan Z-X (2021a) The interactive modeling method of virtual city scene based on building codes. *KSII Trans Internet Inf Syst* 15(1):195–215. <https://doi.org/10.3837/tiis.2021.01.005>
- Wei-long Ding X-JZ, Bin Xu, Yan Xu, Chen K, Wan Z-X (2021b) The interactive modeling method of virtual city scene based on building codes. *KSII Trans Internet Inf Syst* 15(1):74–89. <https://doi.org/10.3837/tiis.2021.01.012>
- Wu L, Wang Y (2019) A low-power electric-mechanical driving approach for true occupancy detection using a shuttered passive infrared sensor. *IEEE Sens J* 19(1):47–57. <https://doi.org/10.1109/jsen.2018.2875659>
- Zhang H, Zhang Z, Gao N, Xiao Y, Meng Z, Li Z (2020) Cost-effective wearable indoor localization and motion analysis via the integration of UWB and IMU. *Sensors* 20(2):344. <https://doi.org/10.3390/s20020344>
- Zhu R, Zeng D, Kosorok MR (2015) Reinforcement learning trees. *J Am Stat Assoc* 110(512):1770–1784. <https://doi.org/10.1080/01621459.2015.1036994>
- Zittis G (2017) Observed rainfall trends and precipitation uncertainty in the vicinity of the Mediterranean, Middle East and North Africa. *Theoret Appl Climatol* 134(3–4):1207–1230. <https://doi.org/10.1007/s00704-017-2333-0>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.