10.48047/jocaaa.2025.34.07.13

An Approch to Building Energy Management Using Smart Control And Machine Learning

Shahidu Ibrahim
Waziri Umaru Federal Polytechnic, Birnin Kebbi, Nigeria
shahiduibrahim@wufpbk.edu.ng

Ibrahim A. Bello
Waziri Umaru Federal Polytechnic, Birnin Kebbi, Nigeria
ibrahimabellobk@wufpbk.edu.ng

Abubakar Bashir
Waziri Umaru Federal Polytechnic, Birnin Kebbi, Nigeria
abubakarbashar4@gmail.com

Corresponding Author: Shahidu Ibrahim
[shahiduibrahim@wufpbk.edu.ng]

Abstrack
Energy use in homes has been impacted, particularly by HVAC systems, as a result of conventional control (manual ON/OFF). It has been acknowledged that occupant prediction and detection play a major role in increasing building energy efficiency. Over the past ten years, a number of strategies have been put out to increase the accuracy of occupancy estimation and detection in buildings. One useful method for identifying and estimating building occupants during unknown behaviour is environmental sensing. However, the research shows that the low calibre of the training dataset employed in the model results in relatively poor environmental sensing performance. This study employed machine learning algorithm collect consistent and reliable datasets about occupancy that can be used for cross-validation. Although different technique might be readily incorporated into the suggested approach, this study uses various machine learning technqiues suggest best candidate for occupancy prediction which in turn provide the potential for higher energy saving. The findings show that the suggested approch can collect data, compute it, and make reliable prediction (up to 99.3%). Additionally, a room with two zones is modelled, each with air conditioning and a separate thermostat controller, to show how the number of occupants affects energy efficiency. IoFClime is used in the first zone, and a design-builder is used to modify IoFClime in the second zone. Using EnergyPlus software, ten tenants are randomly simulated together with local climatic data in three different situations. According to the Fanger model's thermal comfort study, the first and third scenarios can save up to 50% and 25% of energy, respectively.

1.0 Introduction

One crucial component that supports the sustainability of the building industry is energy [1]. Buildings already consume a large portion of the world's energy, and as more electrical appliances are installed, the proportion is predicted to rise in the future years [1]. The majority of energy used in buildings is needed to create a comfortable and healthy atmosphere, which greatly affects the health of our surroundings. The authors in [2] demonstrated a significant turning point that is closely related to the ongoing rise in population in the Asian area and the growing size of the building industry. the findings of an earlier study, initiatives have been praised to enhance the offer that satisfies energy needs for over 1.3 billion people and many industries [1]. The past several years have seen an unusual rise in the ongoing deployment of energy-consuming devices like hair dryers, televisions, ovens, HVAC (heating, ventilation, and air conditioning) systems, and so on. A number of researchers have used historical energy usage statistics to promote energy-efficiency through various approaches over the years. Innovative smart grid strategies that manage energy supply, demand, and usage across multiple sectors can result in greater energy savingsBecause of this, both the construction and transport sectors' total energy usage growth dropped down to less than half of what it was twenty years ago, and the sector's energy consumption growth has also stagnated. Regardless of modern advances in technology, the need for non-combusted energy remains steady, particularly in the petroleum sector as a source of energy. In many construction use and infrastructures, including smart buildings, interior security, evacuation, building activity, and demand management systems, occupancy rate prediction is crucial [3]. Today, a variety of methods are used in buildings to measure the precise amount of occupants, including passive infrared, cameras, lighting, Wi-Fi, Bluetooth, and environmental monitoring (such as temperature, relative humidity, and CO2) [4[]. Occupant estimation offers the potential to cut needless building energy use by over 60%, according to studies. According to a number of publications, study on occupancy assessment using images and videos is accelerating due to the quick development of machine learning (ML) and computer vision [3, 5–9]. The two primary components of the study's methods are installed internal cameras and cameras at the room entry. Cameras at the inside entryway were set up for overview video capture in the study in [3]. The ML background subtraction (BS) technique was used in this research to compute the number of occupants based on counts and eliminate noise from the frame or region of interest. Surveillance cameras are used as boundary sensors in a study in [10] that detects occupants using an ML histogram of orientated gradient (HOG) classifier based on the ML support vector machine (SVM). The event-based optimisation (EBO) theory was also applied in the study to improve the estimation outcomes. In order to determine which way to enter or exit rooms, this kind of approach typically uses motion detection and tracking.

In order to reduce plug load energy consumption and user strain, an occupancy localisation research was suggested that utilised Plug-Mate in conjunction with IoT technology for occupation-driven plug load handling that used intelligent plug load automation [11]. The proposed system uses an intelligent plug load awareness feature to determine plug load type information, automates plug load based on high-quality occupancy information obtained from a nonintrusive indoor localisation system, and supports various patterns and a customised interface [11]. Using information from beacons positioned within the premises, a similar method used Bluetooth Low Energy (BLE) technology, which has the potential to be very useful in locating an occupancy [12].

Despite the fact that a variety of machine learning techniques are frequently employed in building occupancy detection [13]. The Random Forest (RF) algorithm, on the other hand, is widely popular due to its adaptability and simplicity of use, which enable it to effectively handle problems pertaining to both regression and classification, making it a valuable tool for a range of machine learning predictions. Furthermore, it has demonstrated outstanding abilities in managing complex datasets and preventing overfitting, especially in residential buildings with occupants that don't follow a defined routine [14]. A comparison of the performance of building occupancy prediction using various machine learning techniques is presented in research in [2, 15]. The findings show that the RF approach produces fewer false positives and a greater prediction accuracy. Because of its adaptability to random data and ability to forecast room occupancy based on carbon dioxide, studies in [16,17] select RF as a potential model above another ML model.

Environmental sensing employing occupancy data and interior metrological circumstances to create a model that initiates an event to regulate appliance performance is getting more interest in building occupancy estimation development [18]. The phrase "indoor environmental quality monitoring" refers to a comprehensive idea that addresses thermal comfort or indoor air quality and is further explained in [19]. In order to autonomously regulate the lighting and air conditioning system in a four-person household building, another study suggested multiplicative manufacturing techniques that use a microcontroller to integrate temperature, CO2, and relative humidity sensors along with a few other modules [20]. A related study [21] showed how energy prices affect user safety, comfort, and health. According to recent studies, smart indoor lighting has quickly gained popularity as a way to enhance human life [16]. To regulate indoor lighting needs, the authors in [22] proposed using occupancy sensors, dimmable light-emitting diodes, and photodetectors that employ microcontroller units. In recent years, these appliances have been the main focus of study [23]. In order to address energy waste, appliances may practically be made to greatly optimise their energy use by including occupancy parameters into the control system [1]. Recently, direct sensing methods for smart energy optimisation have gained popularity employing devices like wearables, cameras, or passive infrared [24, 25]. However, considerable interaction with occupancy was necessary for most wearable [26, 27] and camera-based [3, 15, 24] solutions. Despite being a dependable solution for sophisticated processing, hardware costs, placement feasibility, and occupancy prediction, direct sensing still has significant drawbacks, such as privacy concerns [3].

As a counterpart to earlier occupancy prediction techniques, environmental sensing is now used in several smart buildings to address privacy issues and other issues. To forecast building occupancy, environmental sensing analyses environmental factors such variations in CO2, temperature, or humidity. The lack of publicly accessible data for environmental sensing was noted in a prior study [2], which proposed a method for creating a high-quality dataset in order to create a comprehensive model that can predict solid occupancy and, ultimately, lower HVAC energy consumption.

Additionally, previous study findings revealed that lighting appliances were the primary energy consumers in buildings, limiting their options to HVAC systems alone. Furthermore, when there are beyond seven occupant space, the dependability of their approach tends to deteriorate. In order to enhance the quality of occupancy-related information utilised for training purposes, this study investigated and evaluated existing methods for smart building occupancy prediction using a multi-modal data fusion strategy inside the smart home ecosystem. In contrast to earlier methods, the suggested combination of data approach employs a parametric classifier to filter noisy sensor readings without being unduly intrusive using rule-based decisions, confirm the quality of occupancy-related data, and guarantee an excellent fit between the room occupancy and occupancy-related data in the room. Furthermore, in contrast to recent studies, a lighting control parameter is integrated with HVAC system control [29, 30]. The occupancy prediction duties were handled by a Random Forest (alternative machine learning algorithms might be added) in order to balance thermal comfort and HVAC energy consumption. Several assessment indicators were used to compare the occupancy prediction efficiency outcomes to the baseline design [31]. Using specially created surveys for respondents of various region, on-site interviews with occupiers regarding their experiences of thermal comfort in various temperature settings were used to analyse the thermal perception of the occupancy. When compared to the prior method, the results show that the suggested controller can save up to 45% of energy consumption while maintaining a satisfactory level of comfort.

Predicting indoor building occupancy with the aim of reducing energy consumption and enhancing occupant thermal comfort with little intrusiveness is the study's main contribution. The structure of the paper is as follows. The earlier studies on interior occupancy prediction and recognition are covered in Section 2. Section 3 presents the information and techniques. The results of energy-saving potential scenarios and occupancy recognition and evaluation models are presented in Section 4. Section 6 wraps up the study results and future research, while Section 5 offers remarks on the findings of the study.

2.0 Related Work

The literature on intelligent systems for energy management in homes and smart building occupancy prediction is reviewed in this section. It draws attention to common methods for occupancy prediction. Current developments in this area have focused on adding functionality that enables residents to participate in data collection for model training, monitor building energy use, schedule or adjust building energy use profiles, and communicate with the grid through utility allocations like demand responses and self-fault reports. Prior research concentrated on advancing and using smart building technology to improve energy savings using techniques like fusion mechanisms or the multimodal strategy.

Numerous methods, such as wearable occupant tracking, voice identification, facial recognition, along with tracking indoor occupancy tasks, can be used to achieve the connectivity [24]. The previous suggestions for occupancy prediction techniques are presented in this section. This section classified the reviewed study into predictive approaches and predictive approaches.
Utilising the generation and analysis of occupant details as well as indoor climatic data to assess the likelihood of being present in space, the predictive technique seeks to simplify the handling of HVAC activities. Occupant data can be interactively gathered in real-time by sensors at the occupant prediction point, or it can be substantially generated through occupant predetermined schedule operations [1]. Conversely, the dynamic predictive technique determines the presence or number of occupants in the building by using real-time data from placed sensors inside the occupancy surrounds. When creating a model that detects room occupancy, the occupant schedule is sometimes seen as a desirable parameter [1]. Identifying actual occupancy to minimise false-negative findings caused by PET or stationary objects and estimating the total number of occupants in the area to modify comfort temperature seat points are challenges faced by most of the dynamic predictive control currently in use.

The research [40] employs a CO2PIR sensor-based method to identify occupancy in the room. An automated controller was designed to flags the HVAC system to modify the temperature set point, guaranteeing occupants desired comfort. Technologies like Infrared and CO2 sensors are frequently used to forecast occupants in indoor spaces so that HVAC control is made easier. These sensors are highly effective at detecting abrupt changes in an occupied environment. Their main issue, though, is the absence of extra data to detect human occupancy, which may result in a high energy consumption when PET or other fixed items are present in the interior.
Using a Gaussian Mixture Model, the occupancy detection technique in [35] effectively counts the overall amount of occupants and recognises the presence of humans. Despite large false positives caused by nearby ambient noise, the recommended method was able to achieve a respectable level of precision. Similarly, the system requires continuous feedback from all interior occupancies in order determine occupation for HVAC ventilation regulation. In order to handle with noisy inference from undesirable sources, the research [36,42] that aimed to increase the performance of [35] proposed a background cancellation procedure that disregards the

sound frequency level of the selected desired sound level threshold. The idea relies on the background cancellation algorithm's capacity to identify the intensity of sound frequencies, which, in comparison to [35], minimises HVAC equipment consumption of energy by 3.54% and occupancy forecast false alarms by 11–12%. The study's occupancy projection is inaccurate even though a background cancellation approach was used. According to a study [3], utilising two cameras simultaneously could enhance the detection of building occupancy. Occupancy prediction often relies on camera-based image and video processing using open-source human-computer vision libraries. To determine total occupancy numbers, numbering or tagging of indoor objects is also commonly used [43]. Single-camera occupancy tracking was used in [43] to control indoor ventilation in the lecture hall using an estimate of occupancy assessment procedure. The study in [32] uses an open CV library template and a naive Bayesian algorithm to forecast the anticipated number of occupants. The analysis of data from experiments shows that the experiment demonstrated a substantial amount of occupancy prediction and identification, along with a false positive because of the students' non-linear line of sight as they competed or wandered around the area of interest. Inappropriate occupant recognition during the overlapped period of entry and exit within the study area is a major challenge for this study. In order to detect human occupancy in space, research in [44–47] proposed combining passive infrared cameras with optical cameras. By using a single camera, the likelihood of false alarms is reduced and detection reliability is ensured. Using a support vector machine technique and a computer vision human template, pixels were extracted and analysed to produce potential detection thresholds. Using a single camera is intended to reduce the likelihood of false alarms while ensuring detection reliability. Using a support vector machine technique and a computer vision human template, pixels were extracted and analysed to produce feasible detection limits. To identify the kind of occupancy found, a naive Bayesian algorithm classification process was used in both cameras. The technique is vulnerable to false results when one camera quality is superior to the other since occupancy detection in this way is only allowed if the cameras reflect identical occupancy type. Otherwise, the detection is disregarded and deemed unclear. According to a suitable threshold that need to be applied for occupancy prediction, a research in [3] proposed a method to lessen this difficulty in [48] by extracting image pixels for training using the Naive Bayesian algorithm. The concept was to lower the threshold of the camera with a negative forecast by 30% for a possible match in the event of an occupancy detection contradiction. Compared to the prior study, the experimental finding assessment demonstrates an increase of 12% relative to occupancy estimation and a 5% energy savings.

To lessen false alarms in disputed sensors, comparable research efforts that merge occupancy data gathered by PIR sensors and cameras were recommended in [44, 46] to regulate HVAC ventilation requirements and preserve occupancy thermal comfort, these methods use occupancy tagging using binary object recognition to monitor occupancy indoors. The chair has thermal sensors to measure the frequency of heat generated by occupation and electromagnetic radiation.

Real-time occupant information collected by installed CO2, PIR, and movement detection sensors is gathered using a sensor fusioc n technique utilised in [20, 38, 42]. The data is then parsed for behavioural assessment using K-Nearest Neighbour. In the indoor pre-existing trained test data template of the SVM threshold, machine learning was used to track and categorise the occupancy type. In order to control HVAC ventilation demands, the controller uses CO2 sensors to monitor occupants rates as concentration levels rise and PIR and motion sensors to give information about occupancy activity inside the building.

In order to determine the number of occupants in an interior space and regulate the ventilation needs of HVAC equipment to reduce needless energy use, the research in [49] uses a light sensor counter and CO2 as a stream of bytes. In order to track the inside occupancy number and measure the number of people entering and leaving, a light sensor was installed on the door entrance. In order to monitor a warm breath of occupancy as a sign of occupancy present status in the interior space, the CO2 sensor was mounted on the ceiling. When compared to conventional thermostat control, the laboratory's study of the testing results indicates that HVAC equipment has a large potential for energy savings.

3.0 Material and Method

With an opportunity to significantly contribute to intelligent construction technologies that allow for dynamic, responsive ways to managing building appliances properly, IoT sensor data has been a vital actor. Therefore, the integration of sensors, including occupancy sensors, with smart building control systems lays the groundwork for more intelligent, efficient appliance control and energy savings. Additionally, this work uses sensor data to build an occupancy prediction model utilising the machine learning approach. The research process and technique for the suggested approach are shown in Fig 1.

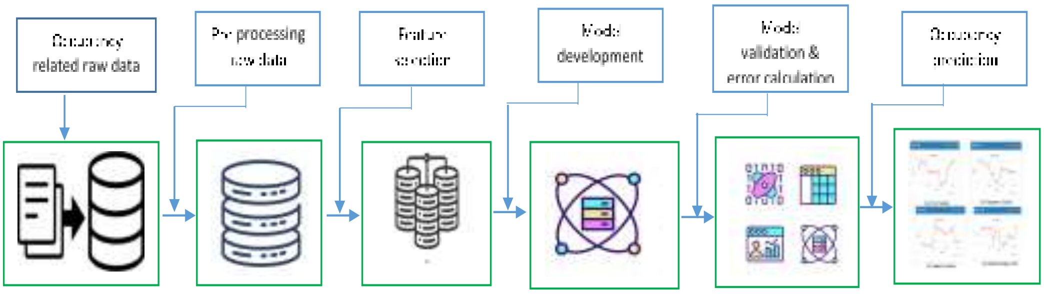


Fig 1. Research Methodology

3.1 Occupancy-related raw data

The office block at Federal Polytechnic Binin Kebbi, Nigeria, is used as a case study in this research. The building material's thicknesses and thermal properties are shown in Table 1. These features aid in assessing the occupants' steady and dynamic behaviour. The sitting room's function is to accommodate social events where people may eat, relax, and watch television. The sensors were put in place to monitor a number of indoor environmental factors, including as CO2 concentrations, temperature, lighting, and relative humidity. In order to confirm that the values matched the sensor readings, the comings and goings in the sitting area were also recorded. The experimental setup is described in more depth in [2].

Table 1. Room thermos-physical properties.

Properties	Material	c (J/Kg·K)	(W/m·K)	Thickness (cm)
	Tuff	**650**	**1.5**	**10**
Wall	Brick	1000	0.11	18
	Polystyrene	1600	0.028	8
Ground Floor	Concrete	650	0.43	
	Stoneware flooring	650	1.25	1.3
	Igloo	650	0.07	8
	Gravel		1.1	1
	Screed ordinary concrete	650	1	5
Ceiling	Hollow-core concrete	650	0.7	25
	XPS polystyrene panel	650	0.4	8
	Bricks tuff	650	0.5	5

The sensors utilised to gather occupancy-related data in the building through contact with residents regarding the problem's context are described in Table 2. Our approach used a camera to address the labelling problem that arises in the calculation of the amount of occupants in trained learning techniques, which are widely used in numerous contexts [14, 44, 50-52]. This was motivated by the difficulties in labelling occupancy values via collaborative learning data collections [28].

Table 2. Sensors used in the study

Sensor	Description	Ambiguity	Unit	Record
Temperature	Measure indoor temperature	1 °C	Degree Celsius	60-second interval
Relative Humidity	Measure indoor relative humidity	±5%	Percentage	60-second interval
CO2	Measure indoor CO2 concentration level	300–1000 ppm: ±120 ppm	Parts per million (ppm)	60-second interval
Light	Measure luminance	10–2000 lux	Lux	60-second

in the building	range	interval

3.2 Data pre-processing

One of the main elements influencing the prediction model's performance is the dataset's quality. An interactive method was previously utilised in [28] to offer occupancy estimates during data collection in order to address real-world privacy issues and inaccurate values used in model training. However, in order to enable self-estimation or labelling, this study uses ambient sensing and camera sensing in place of the manual way (interactive methodology) during data collection. When the prediction model is implemented in the building, the privacy issue is reduced because this method simply employed camera processing for data collecting. Data pre-processing involves the following three processes.

3.2.1 Sensor fusion

Figure 2 provides a broad overview of the suggested framework, which consists of the candidate record, dataset training, and machine learning model (random forest) that results in the fuzzy system. These findings are delivered to a fuzzy inference engine to govern HVAC operation by regulating or creating the appropriate set point temperature based on the occupant's decision or habits. There is an occupant-related feature for each parameter in the dataset log stream. Data logs provide the entire occupancy count. However, a specific event, like a door opening while arriving or leaving, might have an impact on the content of these dataset streams. Restoring the interior conditions to the actual level that corresponds to the building's residents takes some time when such an event takes place. This study used [3] for the camera and [31] for environmental sensing.

An interval of five minutes was established between the last recorded record and the subsequent candidate record in order to guarantee the dataset record's dependability.

For modelling to be effective, occupancy data must be dynamically assessed. The fusion method supplements ambient sensor data with camera data. When compared to the ambient sensing methodology alone, dataset collection utilising this method has demonstrated high performance for cross-validation of several ML algorithms.

As a result, this study suggested a multimodal occupancy prediction framework (see Fig 2) that uses a multi-layer perceptron regression technique for model training and combines a camera approach with an environmental sensing strategy to facilitate dataset collecting and quality assessment.

10.48047/jocaaa.2025.34.07.13

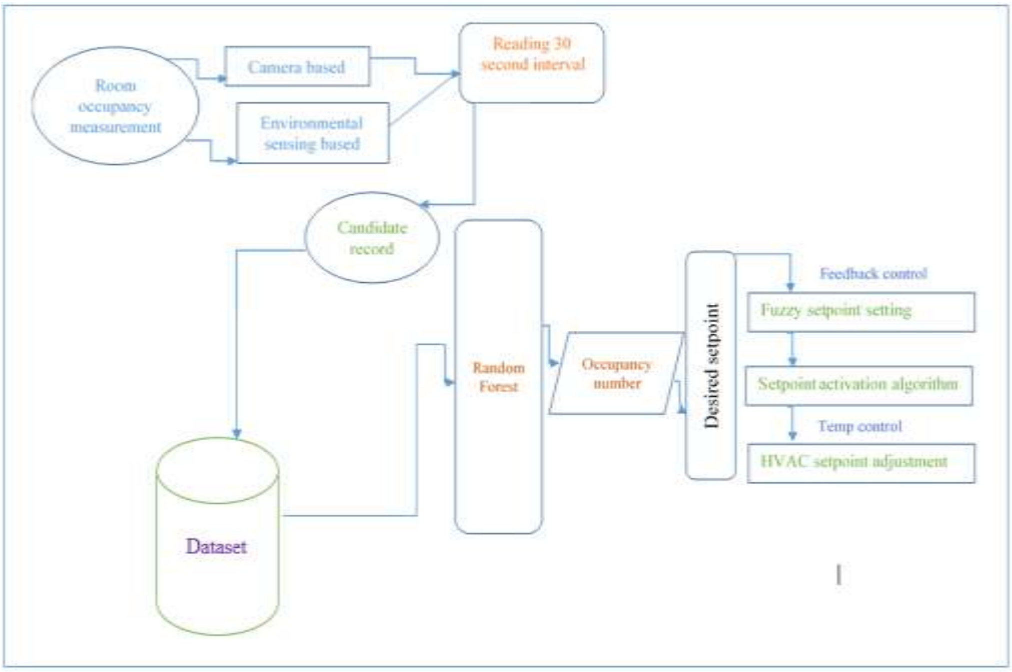


Fig. 2. The Proposed framework for occupancy prediction

The data reading from the sensors must be confirmed prior to storage in order to guarantee the quality and dependability of training data, which subsequently supports the model prediction when it is used for testing. Thus, the following approach can be used to address this issue.
One of the key factors that smart thermostats employ to control interior comfort and energy consumption is occupancy number. An interactive method was previously utilised in [28] to offer occupancy numbers during data collection in order to address real-world privacy issues and inaccurate values used in model training. However, in order to enable self-estimation or labelling, this study uses ambient sensing and camera sensing in place of the manual way (interactive methodology) during data collection. Only environmental sensors are employed when the prediction model is implemented in order to read the indoor environmental parameters and provide an accurate occupancy number forecast. Additionally, before being stored, the quality of the data record that the sensors read is checked during data collection, which subsequently validates the model forecast. Thus, the following approach can be used to address this issue.

Let dataset records measured by sensors every time be $\text{recorded}(t)$ consisting of certain features$f_1 \ldots f_n = \text{record}(t)$. Thus, $f_1(\text{record}(t)) \ldots f_n(\text{a record})$ correlates with the status of occupancy at time t, where $f_i(t) \in (\tilde{f}_1\tilde{f}_1)$. let occupancy count represent labels measured by a thermal camera that tally with the corresponding indoor data recorded by environmental sensors. Thus, the comprehensive dataset records measured to train the prediction model are represented as $r(t) = (\text{record}(t)), l(t)$). Any record in the dataset that has no corresponding label recorded is considered a candidate record (required validation) and any record with a corresponding label recorded is considered as a recorded record.

Allow the donation of recorded records. The following are the main obstacles to this kind of interactive learning: (1) how to prevent a candidate record from being recorded in the dataset? Given that a camera automatically gathers it (2), how can the quality of the label (t) that the camera assigns to datasets be evaluated? To offer the ground truth, this procedure is carried out based on a distinct response from the camera-predicted occupancy label. Every time an environmental record matches data captured by cameras, it is deemed legitimate and included to the dataset; if not, it is deleted. Throughout data collection, this procedure is repeated with the use of a parametrised classifier. Decision trees were utilised in this study's parameterised classifier to examine the candidate record utilisation of (if-then) rules. A predefined classifier structure with adjustable parameters based on the input data is used by this parameterised classifier. To match and solve the intended tuning problem, this classifier can be adjusted in the final structure based on each new data set and how much it varies from the previous one. Reducing the discrepancy between the classifier's estimated and real indoor occupancy is one of the goal functions. In the last stage, we integrate the results from each modality and apply a combination of rule-based conclusions and biassed alternatives.

The last step of multimodal techniques involved combining the results of each modality and executing a mix of weighted possibilities to confirm the data quality using rule-based judgement. The decision tree and parameterised classifier are used to evaluate the quality of the data during data collection, and the camera's prediction of the occupants' label is used to automate the occupant input value rather than the occupants' manual input. To determine and examine the error distribution for the suggested method, the effective data were trained and modelled using a random forest.

3.2.2 Normality of the data distribution

Moreover, checking the normality distribution of data is essential to determine whether the data is normally distributed. However, the graphical depiction for evaluating normalcy necessitates a high level of knowledge to avoid erroneous readings. X and Y vectors are commonly used to show data for graphic analysis. According to [53], suppose Y is the parameter that relies on the regression of X. If $X(x_1, x_2, x_3, \dots x_n)$ are correlated, then Y is considered to be dependent on X, and $\mu = f(X)$ is a scattered vector. Hence Y and μ can be presented

$$Y|X \sim N(\mu = f(X), \sigma^2) \quad (1)$$

$$\mu = f(X) = (ß_0 + ß_1 * x_1 + ß_2 * x_2 + \cdots ß_n * x_n)$$

The normality distribution of the recorded dataset is illustrated in a graphical presentation using the Q-Q plot test (see Fig 3 and 4).

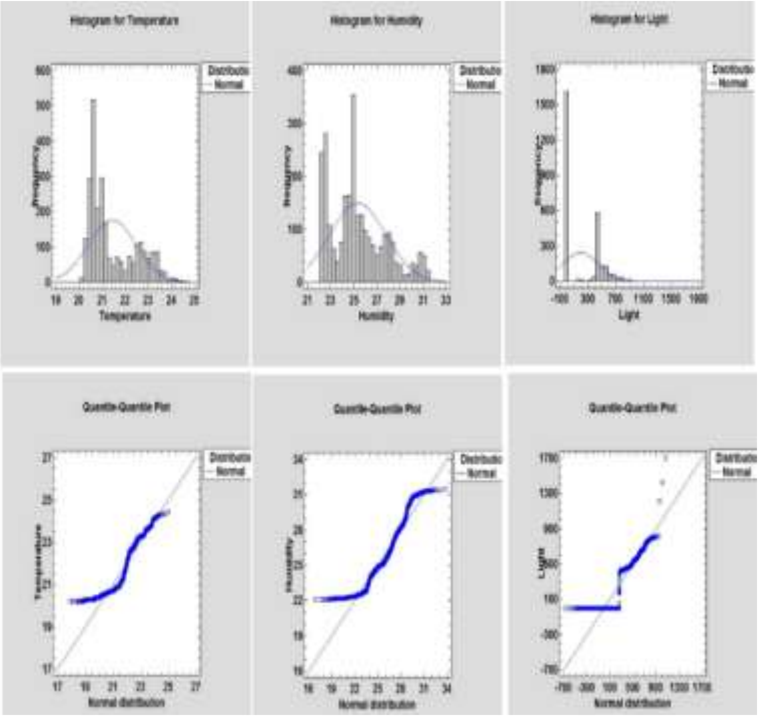


Fig.3. Temperature, humidity and light

10.48047/jocaaa.2025.34.07.13

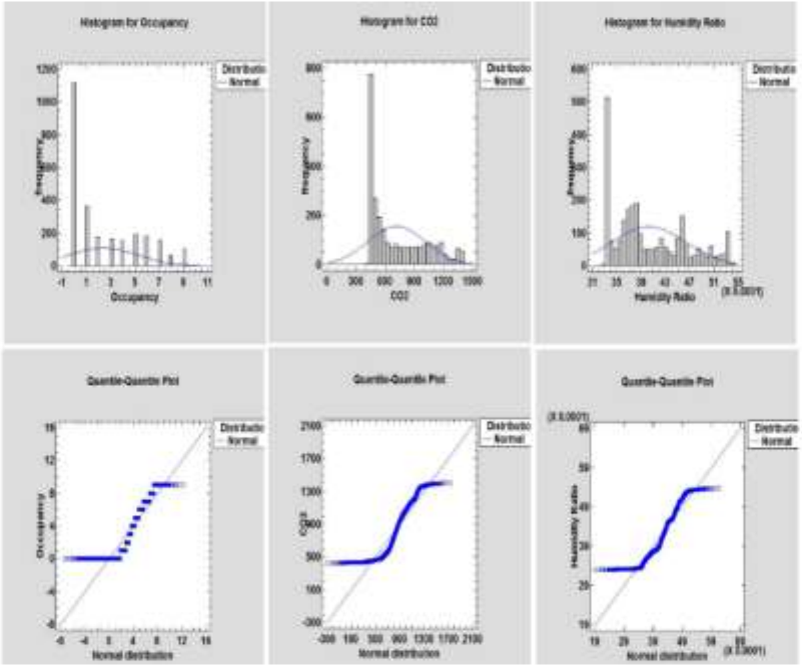
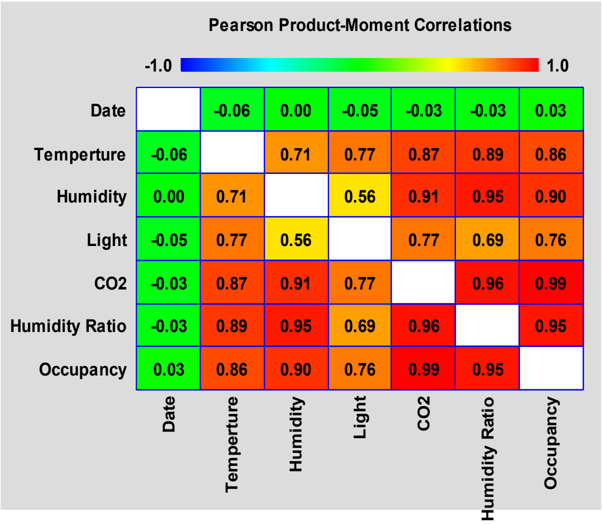


Fig.4. Occupancy, CO2, and humidity ratio

The distribution of datasets varies somewhat. Following physical examination of the mismatched plots, it was determined that the skew was caused by natural creation rather than sensor malfunction, which was not a problem and could not affect the accuracy of the model's predictions. Unfitted point distributions are seen for all variables, with the most severe values found for CO2 and occupancy. In a normal distribution, one out of every 340 observations may deviate from the mean by at least three standard deviations [54].

3.2.3 Data correlation

Finding predictors that have a high correlation with predicting variables is crucial before feature selection [14]. The correlation value in this study is calculated using the Pearson Product-Moment value (PPMC) metric (see Fig 5). When given a set of related (x, y) values between -1 and +1, PPMC assesses the level of dependence between the variables y and x [6]. The calculated PPMC values for a total of six variables, whose values range from -1 to 1, are shown in Fig. 5. A label peak correlation is represented by a value of 1, followed by strong correlations (0.9), moderate correlations (0.8), and extremely weak correlations (0.00 and -0.00 on a green backdrop). Through variable permutation significance, parameters that have no connection to predictors or a poor correlation index may be removed from the predicting model's development. Moreover, in order to simplify models, it is recommended that just one of two components be studied if they are highly associated.



	Date	Temperture	Humidity	Light	CO2	Humidity Ratio	Occupancy
Date		-0.06	0.00	-0.05	-0.03	-0.03	0.03
Temperture	-0.06		0.71	0.77	0.87	0.89	0.86
Humidity	0.00	0.71		0.56	0.91	0.95	0.90
Light	-0.05	0.77	0.56		0.77	0.69	0.76
CO2	-0.03	0.87	0.91	0.77		0.96	0.99
Humidity Ratio	-0.03	0.89	0.95	0.69	0.96		0.95
Occupancy	0.03	0.86	0.90	0.76	0.99	0.95	

Fig.5. Variable correlation values

3.3 Feature selection

Variable feature identification is critical in the development of machine learning models because it necessitates the removal of features with a low association before the model training evaluation. In [38], the variable significance measure metric is evaluated to eliminate uncorrelated variable parameters. According to the theory in [38], predictors $X = (X_1, \dots, X_p)$ should be an array of parameters or variables for forecasting variable Y. In a regression situation $\hat{f}$, the formula for variable predictor Y is a function measured with numbers in $\mathbb{R}$. The error in the estimation of $\hat{f}$ can be calculated with $\mathcal{R}(\hat{f}) = [(\hat{f}(X) - Y^2)]$ and the goal is to measure the expectation $f(x) = \mathbb{E}[Y|X = x]$. Moreover, Let $\mathcal{D}_n = \{(X_1, Y_1), \dots (X_n, Y_n)\}$ be set data of n replications of (X, Y) where $X_i = (X_{i_1}, \dots, X_{i_p})$. Since the positive prediction error of $\hat{f}$ is not known in the test dataset $(\bar{\mathcal{D}})$. Hence $\bar{\mathcal{D}}$ can be represented as:

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

Variable permutation significance proposed [40] demonstrated competence in many nonlinear predictors similar to the proposed model and was thus used in this research. The method took into account predictors $X_i X_j$ as the key predicting Y. If the connection between the features $X_i X_j$ and Y is disrupted, the error score in the prediction may rise. The model's score number shows the extent to which the predictor relies on the collated features. This technique offers the advantage of being model-independent, enabling it to be tested multiple times with different function permutations [40]. Arbitrarily permute the data of the $X_i X_j$ to illustrate this model.

The statistical permutation can be measured as follows: express the set of samples of out-of-bag $\{\bar{\mathcal{D}}_n^t = \mathcal{D}_n \backslash \bar{\mathcal{D}}_n^t,\ t = 1, \dots, n_{tree}\}$. Let $\{,\ t = 1, \dots, n_{tree}\}$ donate permuted out-of-bag samples by randomly permuting of the $j - th$ variable's values for every out-of-bag subset. The variable X_j's statistical permutation value is defined as:

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

This quantity is the statistical equivalent of the permutation importance measure $\hat{I}(X_j)$. Recently formalized by Zhu [55]. 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 which can be measured by:

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

The permutation values of X_j in the expression of $\hat{I}(X_j)$, replicate the exact and independent copy of the pattern of distribution of (X_j) in $\hat{I}(X_j)$.

3.4 Model development

ML approaches are employed to achieve solid prediction efficacy, and datasets are often split into train and test ratios during model building. The process for evaluating the effectiveness of several machine learning techniques for model prediction and selecting the most effective approach to deal with the advancement of model prediction is straightforward and effective. Rearranging and dividing the dataset into a 70:30 ratio are steps in the process (Fig 6). The first ratio, which is often referred to as the training dataset, is used to match the model. In order to give the variables dataset for testing and evaluating the forecast, the second ratio—also referred to as the test dataset is supplied into the model as input.

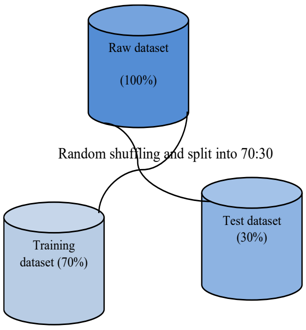


Fig.6. The ratio of training and test dataset

10.48047/jocaaa.2025.34.07.13

3.5 Proposed flowchart

To guarantee that thermal comfort is maintained within predetermined comfort boundaries, the IoT thermostat gathers the occupancy number, sets the intended set point, and gives the controller instructions to maintain the desired set point (see Fig 7). Based on known environmental data (e.g., indoor CO2, temperature, and humidity), our first work in [29] suggested a novel adaptive controller for HVAC systems or IoT thermostats that can forecast optimal temperature bounds. In order to compare the current interior temperature with the desired temperature and determine whether to modify the HVAC set point temperature, the controller asks the IoT thermostat for the current indoor temperature and occupancy every five minutes (see Fig 7).

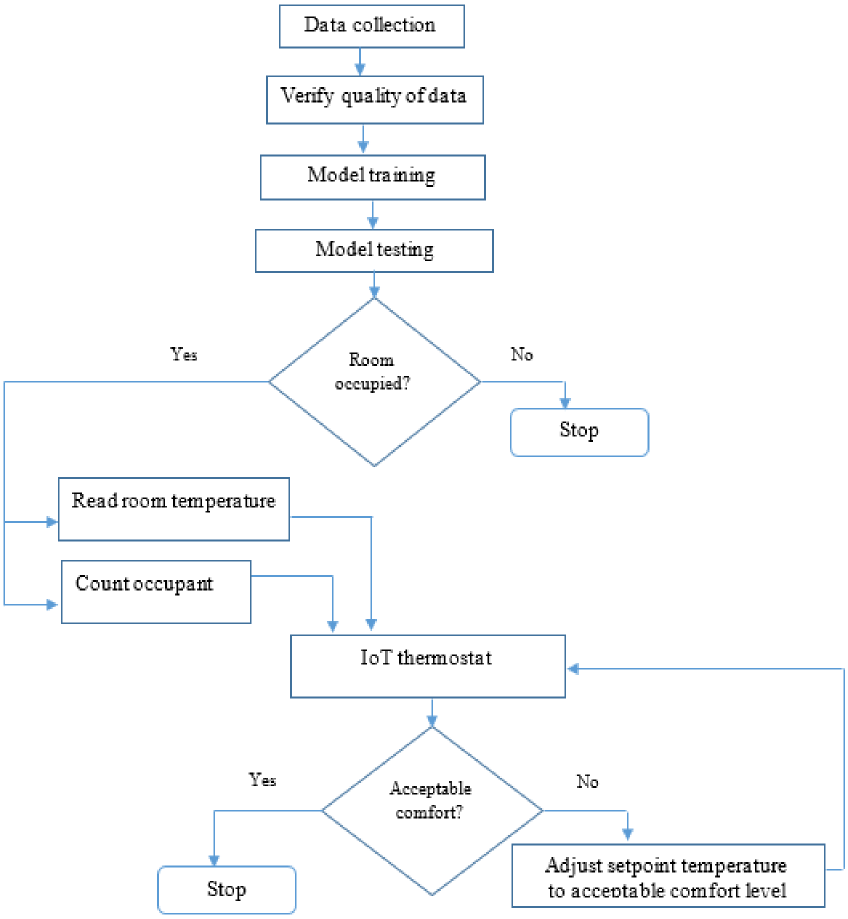


Fig 7 Flowchart of the proposed controller

In most smart thermostat modes, direct control is used. This may be easily accomplished by adjusting set-points by 5 degrees from the desired temperature, which forces the thermostat to turn on when the temperature falls below comfort limits or off when the temperature rises above them. In order to keep comfort satisfaction within the intended bound, the controller then takes note of the HVAC status. Comparable research [3, 56,57] determines that 5 degrees is a sufficient margin to handle the headband and maintain room conditions until the subsequent control interval. The process keeps on until the specified response time is up, at which point the controller restores the smart thermostat's regularly intended comfort level, allowing it to function in the standard deadband-based setpoint mode.

3.6 Hardware and software layout

As seen in Fig. 8, the adaptive controller suggested in this work is made up of three parts. The initial component consists of sensors placed inside the perimeter of interest that gather occupancy and surrounding data. The sensors transfer the data via serial connection to an Arduino microcontroller, which then uses a wi-fi-based microcontroller (ESP/011 to analyse the data and send it to the IoT cloud ThingSpeak over the internet). The

10.48047/jocaaa.2025.34.07.13

fuzzy rule controllers were created for occupancy detection and estimation processes in the second component, which is ThingSpeak coupled with Matlab. The third part is an intelligent controller that determines whether to accept or deny the request to turn lights and HVAC equipment on and off based on signal information from ThingSpeak.

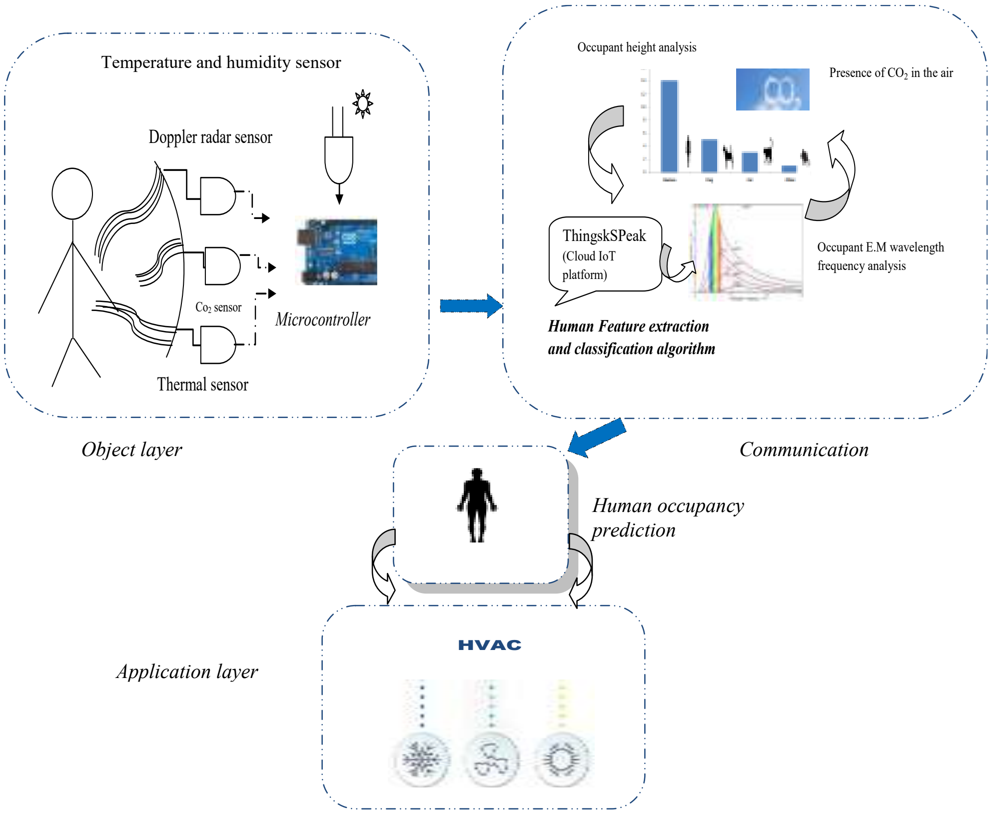


Fig.8 Proposed smart controller

In order to perform occupancy body dimension analysis, including occupancy height prediction, sensors, including ultrasonic sensors, were placed perpendicular to occupancy entrance in the region of interest, as shown in Fig. 8. In order to carry out the system's control logic, temperature and humidity sensors track changes in the interior climate in addition to CO2 concentration. The lamp had a 10W light bulb dimmer for energy usage and was attached to a photoresistor that was placed around 1.5 meters from the ground. Since the main objective of [14] is HVAC energy consumption and thermal comfort study using EnergyPlus, photoresistor or lux sensor was not taken into consideration.

4.0 Machine Learning Models

In order to better understand the parameters in the machine learning framework for prediction, this area has been selected for investigation. These models are well-known and frequently serve as efficiency baselines, despite being simpler and less exciting than several of the more notable recent improvements in this field. These models have an additional benefit in that neural network libraries serve them well as basic choices for numerous applications outside occupation prediction. The scikit-learn Python library is used in both of the methods in this study, and information regarding the predefined settings of the algorithms is available in the library documentation (Abade, Perez Abreu, & Curado, 2018). This section's remaining content provides a high-level summary of the selected machine learning models and how their outcomes are anticipated.

4.1 Naïve Bayes classification

Among many efficient and effective classification techniques is naive bayesian (NB). The technique is based on the Bayesian Theorem of Chance, first proposed by Reverend Thomas Bayesian (Bouhamed, Amayri, & Bouguila, 2022). The idea states that historical data and present-day circumstances influence the likelihood of a hypothesis. It is a technique for figuring out how new information affects the likelihood that a presumption is

10.48047/jocaaa.2025.34.07.13

true. It has been used for a wide range of purposes. The result of the model prediction using NB indicate solid performance (see Fig 5).

4.2 Artificial neural network

Artificial Neural Networks (ANNs) are structures inspired by biology that are used to model and estimate problems where sample data is employed to forecast a variety of parameters during training. The neural net approach uses a set of variables that are both independent and dependent to train the model that processes the data. Each neurone in these networks is an individual. Typically, particular learning rules are used to determine the strengths of links between neurones. A neural net with two hidden layers, each containing the same combination of neurone numerals, was assessed using the dataset. The result obtained using ANN tend to be more reliable than NB prediction (see Fig 5).

4.3 Logistic regression

Logistic regression (LR) predicts a dependent variable in logistic settings with a dependent variable with two potential values output and one or various independent variables. The independent variables are evaluated using the dataset, typically using a maximum-likelihood calculation to decide which is appropriate in predicting depending on the variable. The performance of LR prediction is same as NBC prediction results (see Fig 5).

4.4 Random Forest

Random Forests (RF) are 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 (Breiman, 2001). This technique provides several conditional rules that can be 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 [42], 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. As can be seen in Fig 5, the RF classifier is evaluated to verify its performance prediction on new data. The prediction accuracy obtained RF are better than NB and ANN (see Fig 5).

4.5 Support Vector Machine

The Support Vector Machine (SVM) algorithm does not require the same assumptions as the LDA model to make predictions. This approach operates by locating the boundary that maximizes the difference between the groups to be divided, which is always achieved in a high-dimensional space. The boundary is discovered by fitting the data samples with a chosen kernel function, which informs the relationship of neighboring data. In this study SVM produces lower prediction accuracy in comparison with other ML techniques employed (see Fig 5).

4.6 Evaluation metrics

Usually, more than one metric is used to measure the performance of the model to confirm its performance across new data. As a result, other measures are taken into account, such as AUC: Area under the ROC curve refers to a metric used to measure aggregate performance across the entire categorization criteria. It can be used to interpret the probability that the rates of a random positive sample may be larger than those of a random negative sample.

Precision: Precision is measured in an unbalanced classification issue containing two classes as a total of correct detection known as True Positives (TP) divided by the total correct detection and incorrect detection known as False Positives (FP). The equation can be represented

$$Precision \frac{TP}{TP+FP} \qquad (5)$$

Recall: Recall is measured in an unbalanced classification problem containing two classes as recall can be determined by dividing the total of correct detection by the total correct detection and unpredicted detection (False Negatives). The equation can be represented

$$Recall \frac{TP}{TP+FN} \qquad (6)$$

10.48047/jocaaa.2025.34.07.13

F1-Score: is the metric used to harmonize the value of precision and recall. The equation can be represented

$$\mathrm{F1-Score} = 2 \times \left(\frac{\mathrm{Precision} \times \mathrm{Recall}}{\mathrm{Precision} + \mathrm{Recall}}\right) \tag{7}$$

4.7 Results of the RF model

To obtain the results, the outputs from each tree are joined together. The models' handling of bias and uncertainty in their predictions is influenced using these guidelines. To determine whether a room is inhabited or unoccupied, the binary classifier employs CO2 as a predictive variable.

The efficiency of the RF classifier for occupancy room occupancy detection is shown in Table 3. The score bin represents recursive splitting analysis from the dataset sample that provides the set decision strategies that prediction used to fit the data, as shown in Table 3.

Table 3 RF model prediction using binary classification using CO2 data

Score Bin	Cumulative AUC	F1 Score	Precision	Recall	Negative Precision	Negative Recall	Accuracy
(0.900,1.000)	0.001	0.813	0.983	0.719	0.792	0.999	0.837
(0.800,0.900)	0.013	0.867	0.999	0.741	0.721	0.982	0.849
(0.700,0.800)	0.027	0.821	0.965	0.756	0.732	0.964	0.855
(0.600,0.700)	0.030	0.823	0.976	0.773	0.753	0.960	0.865
(0.500,0.600)	0.047	0.881	0.932	0.784	0.704	0.940	0.867
(0.400,0.500)	0.073	0.861	0.926	0.788	0.778	0.908	0.860
(0.300,0.400)	0.169	0.845	0.865	0.804	0.837	0.793	0.833
(0.200,0.300)	0.364	0.877	0.763	0.936	0.912	0.579	0.808
(0.100,0.200)	0.644	0.789	0.665	1.000	1.000	0.297	0.706
(0.000,0.100)	0.941	0.754	0.583	1.000	1.000	0.000	0.583

The evaluation has been conducted to confirm the effectiveness of the proposed method on new data, which is crucial, especially for systems that use ON/OFF control, because the model's performance might change based on both new and historical data. Consequently, the training and testing bin ratios are separated from the dataset record in Table 4. While the accuracy ranges from 74% to 86%, the precision ranges from 70% to 99.6%, the recall ranges from 29% to 99%, and the classifier prediction performance varies from 75% to 88% for the FI score.

4.8 Comparative analysis

This section is on the evolution of benchmark dataset selection in the field of building occupancy prediction. The examination of four popular and published occupancy prediction and machine learning datasets is presented in this section. The authors' disclosed performance measures served as the basis for the collection of these datasets. Researchers may easily access these datasets. The size of the records and the quantity of variables in these databases vary. We have found that a number of studies have employed them to empirically assess the selection, composition, ranking, and suggested procedures (refer to Table 4). We must do a performance study of the four public web service datasets prior to recommending the effectiveness of our framework for building occupancy prediction. As a result, choosing the best dataset can improve the suggested framework's effectiveness. The purpose of person correlation coefficient analysis is to ascertain how predictors and predicting factors are related. In a similar manner, dataset characteristics were extracted using a normality test using a Q-Q plot to lessen dataset parity. Using random data, this chapter also helps identify potential machine learning models that can accurately estimate building occupancy.

Table 4 Benchmark databases extracted from the literature

10.48047/jocaaa.2025.34.07.13

Approach	Technologies	Technique	Accuracy (%)
[2]	Camera and sensors	Machine Learning	89-99
[29]	Sensors	Machine Learning	79-85
[14]	Camera and sensors	Machine Learning	76-99
Proposed approach	Camera and sensors	Machine Learning	89-99.6

Both research [14] and [2] use a similar methodology, as shown in Table 5, with the suggested study obtaining prediction accuracy of 89-99% and 76-99%, respectively. Following the required procedures to clean up the authors' datasets, our model's accuracy increased to 99.6%. Additionally, a comparable method (using only sensors and no camera) was suggested by [29] and produced a prediction of 79–85%. Our prediction model produced the same outcome as the authors' because of the discrepancy (outlier) in the dataset they presented. Furthermore, as the occupancy number exceeds seven, the accuracy of the [29] decreases. In conclusion, our research shows that the multi-modal data fusion technique is currently receiving greater attention in recent literature in an effort to enhance the dataset quality, which is crucial for confirming the prediction model's effectiveness.

5.0 Building Occupation and HVAC System Analysis

Binary thermostats are mostly used to turn the HVAC system on and off while no one is around. The air conditioner was off and the temperature was higher during our observation while the room was unoccupied. Between 8:00 a.m. and 5:00 p.m., the study took indoor temperature readings (see Fig. 9). It was found that the temperature rose considerably between 11:00 and 2:00 in the morning, making the discomfort level higher than it was during any other time.

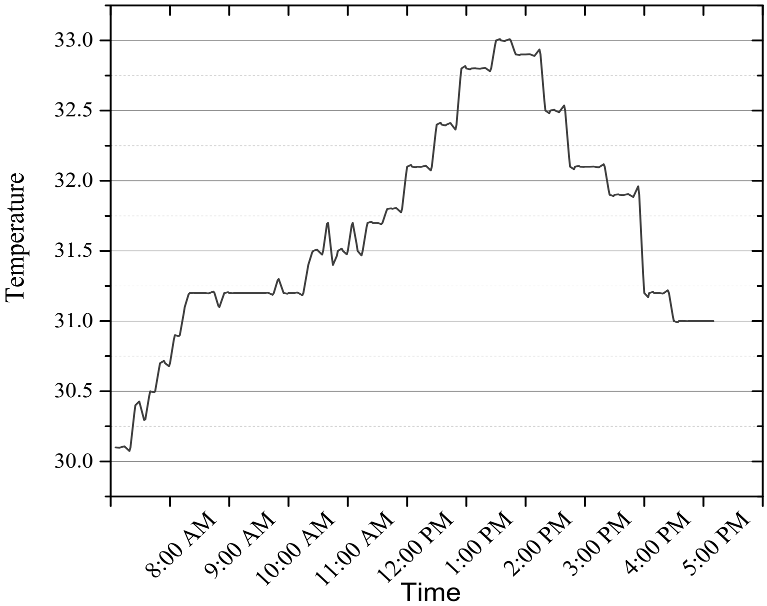


Fig.9 Room temperature with air condition OFF

Because the current controller operates on a set timetable, it is susceptible to false alarms, particularly when tenants' plans or schedules clash with the control policy, which may lead to wasteful energy usage. Our approach controls HVAC operation by predicting the space's occupancy, so the system only turns on when the room is occupied. This suggests that the temperature of the room and the speed at which the air conditioner runs to maintain a comfortable level have an impact on the total amount of energy used.

5.1 Room temperature control

As can be seen in Fig 9 from 8:00: A.M to 9:00 A.M and 4:00 PM to 8:00 PM, the temperature is “Hot. “From 9:00 PM to 11:00 AM, and 3:00 PM to 4:00 PM is considered to be “Very Hot”. From 11:00 PM to 3:00 PM, the level of temperature is “Extremely Hot” when the air conditioner is not operating.

10.48047/jocaaa.2025.34.07.13

In order to maintain occupant comfort at a sufficient level, the suggested controller's objective at any given time is to lower the temperature as much as feasible. By achieving the appropriate comfort level for the interior temperature, the suggested controller demonstrates exceptional performance. The inside temperature is shown in Fig. 10 when the air conditioner is running. Figure 10 illustrates how the controller managed to keep the room temperature between 20.5°C and 16.5°C throughout the air conditioning system's operation. This demonstrates that, with the exception of extremely hot and high conditions, the suggested controllers fairly raise the internal temperature to the appropriate comfort level.

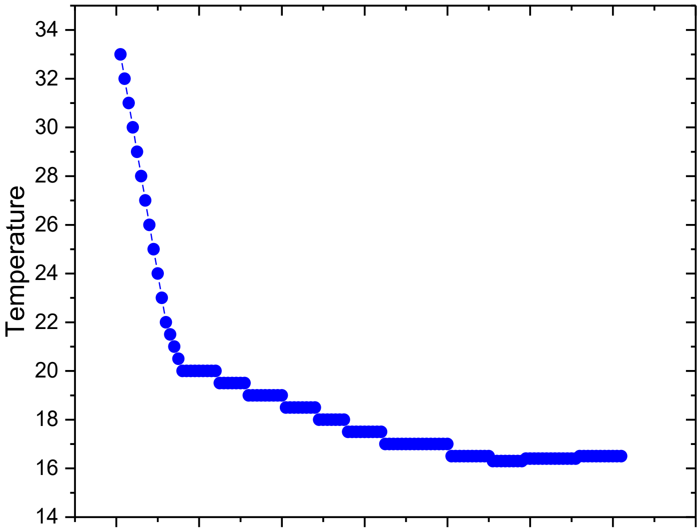


Fig.10 Proposed smart controller

5.2 Thermal Comfort Analysis

In order to prevent radiation or emissions from surfaces or objects, the data logger is positioned in the centre of the room and allowed to hang freely. Data loggers were also mounted on walls and poles so that the devices could be stacked on top of each other. The data logger was secured with double-sided tape at a height of two meters on a wall or pillar; this tape was fastened to walls and poles to prevent the loggers from falling to the ground. The data logger is positioned far enough away from windows, roofs, and the sun to prevent heat, radiation, and cold sources. The installation of measurement equipment in the building is indicated by the loggers' setup to record temperatures at 30-second intervals.

Using ISO 7730, ISO 10551, and ISO 8996 as guidelines, this configuration will allow our research design to incorporate on-site interviews with occupants regarding their experiences of thermal comfort in various temperature settings. These interviews will be conducted using a questionnaire designed for occupants of different ethic group in Nigeria, including Hausa, Yoruba, Igbo, and Fulani. Adults between the ages of 24 and 60 make up the residents. Subjective factors and thermal perception of thermal feeling were recorded using the questionnaire. Following the inquiry, the seven-point ISO 77300 scale very hot, hot, warm, neutral, cool, cold, and extremely cold was chosen.: “How are you feeling now?” A Rayman software available at https://www.urbanclimate.net/rayman/ was utilized to compute PET based on the occupant's feedback gathered. A total of 184 questionnaires were considered valid for the length of the investigation. The average PET readings for the participant are displayed in Table 5.

Table 5 PET values for all participant

Nationality	PET,(°C)			
	Maximum	Medium	Minimum	Aplitude
Hausa	41	25	12	29
Yoruba	56	30	12	44
Igbo	40	26	12	28
Fulani	52	28	12	24

While the majority of the respondents (50%) of Hausa abd Fulani ethic group belive the weather is cold at the same temperature level (see Fig 13), a larger percentage (55%) of Yoruba and Igbo ethic group believe the weather is neutral in the PET with a range of 16°C. This suggests that Youruba and Igbo seem to have lower PET levels and a greater tolerance for cold. At temperatures exceeding 16°C to 24°C, most ethnicities report feeling neutral (see Fig 12 and Fig 13). Additionally, all ethicnicity have very similar thermal sensations of the temperature in the 20°C–24°C PET range, suggesting that hot weather occurs when PET levels are higher than 24°C.

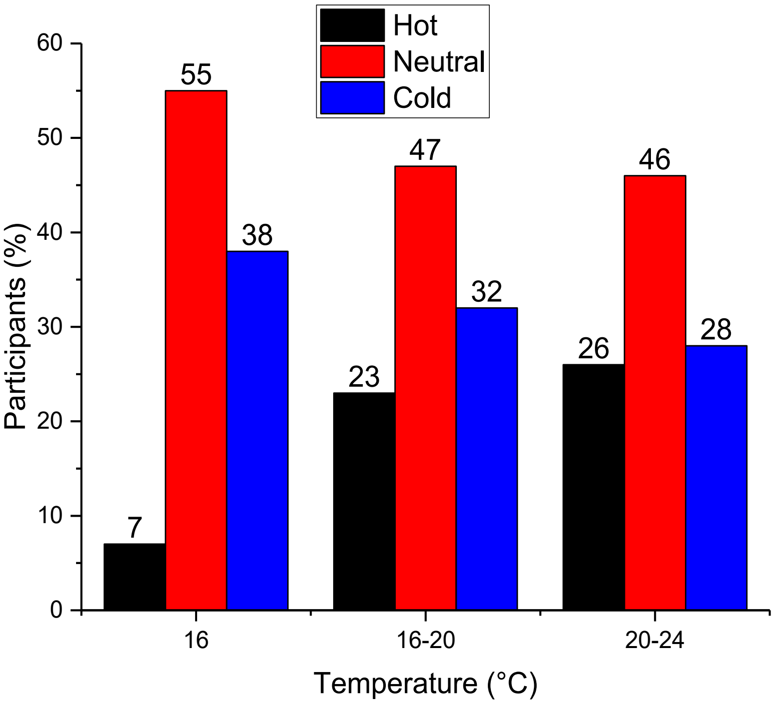


Fig 12. Thermal perception of Yoruba and Igbo participants

10.48047/jocaaa.2025.34.07.13

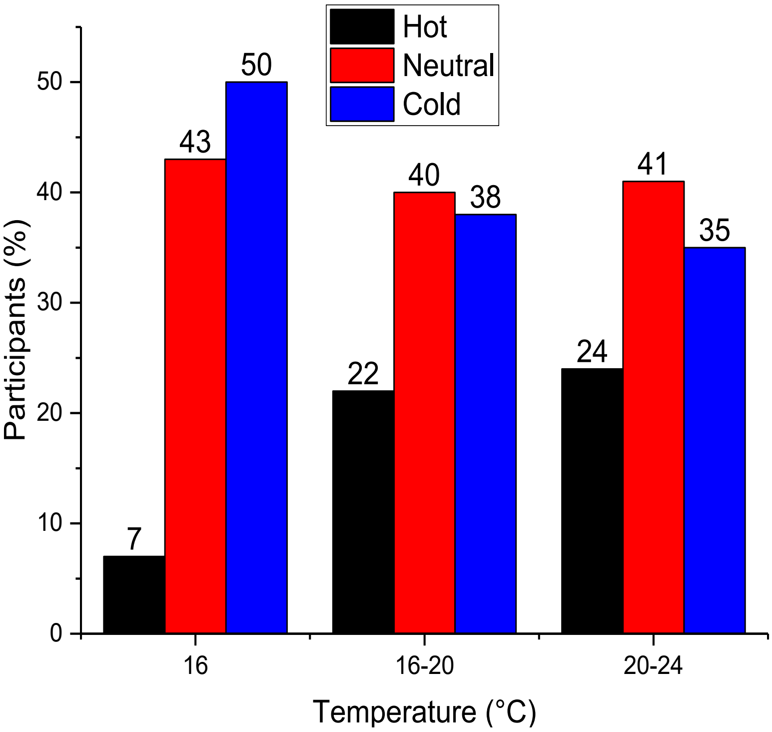


Fig 13. Thermal perception of Hausa and Fulani participants

Additionally, the results of the logistic ordinal regression, which showed the range correlation of the participants' distinct thermal sensations, were used to calibrate the PET index. This finding indicates that PET measurements become "Very Cold" and cannot be attributed to any nationality after temperatures drop significantly. For all nations, the upper bound of the "Cold" range is the same. But compared to Hausa, Fulani have a smaller "Cool" range, but Yoruba and Igbo have a wider "Neutral" range. It was challenging to determine the "Warm" range; nonetheless, all subjects reported experiencing the same occurrence for Igbo, which suggests that they feel "hot" and "very hot" as the temperature rises.

5.3 Energy consumption

The room temperature is the primary component taken into consideration for the controller setting, which is necessary to keep the thermostat running in order to retain the zone comfort at least somewhat closer to the appropriate comfort level, because the cost of power does not signal much influence on energy use in the setting. This also held true during periods of high energy consumption; the compressor continued to run until the zone temperature reached the thermostat's targeted level. The compressor turns off once it reaches the proper temperature. As long as there are people in the room, this process will keep happening. This procedure establishes the air conditioner's cycle.

An air conditioner's cycle time is how long it needs to operate in order to keep the zone at the desired temperature. The compressor operates for a significant period of time to lower the zone temperature, as seen in Fig. 14, which lengthens the cycle duration. The suggested method uses 38.708 kW of energy, while the current method uses 39.159 kW. Under comparable temperature conditions in [30], the suggested controller's HVAC systems use less energy than those in the current study. The suggested method optimally modifies setpoint temperature through a feedback loop as the devices' energy consumption could not be delayed for long periods of time to let users to utilise them during peak hours. Figure 14 illustrates how the suggested method reduces energy use by almost 45% when compared to the prior method.

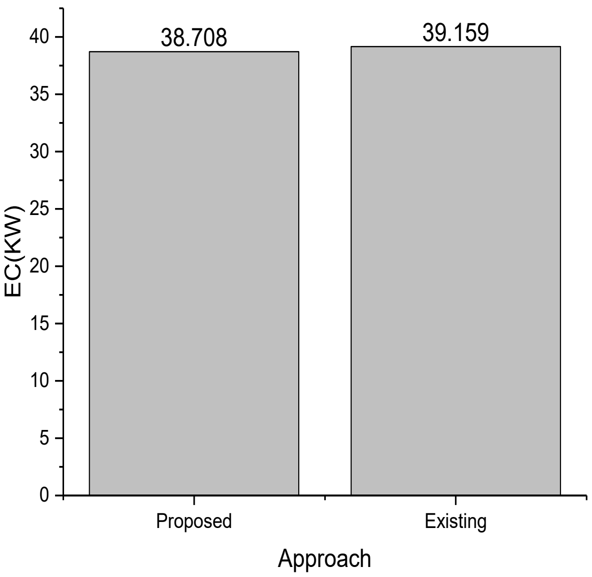


Fig. 14. Comparison of energy consumption

5.0 Discussion

Occupancy prediction in a smart building can assist with demand management ventilation strategies that balance thermal comfort and energy consumption. The significance of occupancy privacy, especially in residential buildings, makes it impractical to implement many of the proposed occupancy prediction techniques that use intrusive technology, such as wearables, cameras, and Wi-Fi, in residential settings. This led to a lot of interest in the recently developed environmental sensing technology. Unfortunately, as shown in the literature, the environmental sensing performs relatively low due to the poor quality of the training dataset, the prediction model's ML technique selection, and the weak feature correlation between predictors and predicting variables.

Occupancy prediction: Using information from many sensor streams that significantly correlate with building occupancy, this study proposed a method for forecasting interior occupancy. The suggested method was tested and taught using a performance evaluation prototype system. The random forest strategy is chosen because of its adaptability and effectiveness in addressing issues with both regression and classification techniques, even though the suggested prototype can make accurate predictions utilising a range of machine learning techniques. The evaluation's findings show that the suggested strategy performs well when compared to the total effectiveness of earlier strategiesAdditionally, the findings demonstrate that, in comparison to utilising a single variable parameter or sensor data directly, adding more variable factors with a strong correlation may significantly increase the accuracy of occupancy prediction. One of the main challenges of the suggested technique is user intervention during data collection to verify that data from environmental sensors matched or matching data captured by the camera. If not handled correctly, this might damage the quality of the dataset and the prediction's overall performance. In place of the traditional if-else classifier suggested in this work, this problem can be further addressed in a variety of methods, including camera fusion, validation classifiers using machine learning or deep learning to guarantee that only matched and legitimate related records are collected in training datasets.

One crucial component of the suggested prediction method is feature selection, which aims to find the ideal collection of characteristics that enables the construction of an optimised occupancy prediction model. The time (occupancy departure and arriving) is no longer appropriate to include in occupancy forecast because the data collection was conducted in residential buildings. Before the dataset sample was included to the model for evaluation, the study found a variable (time) with poor association at this point in the feature selection process, and it was removed. As a result, the suggested method is unable to forecast building occupancy using a timetable. Energy conservation requires accurate occupancy schedules for building energy models. In order to create accurate indoor occupancy predictions even when occupancy schedules are altered, a deep learning technique may be investigated to include reference occupancy schedules including sets that significantly differ from the actual occupancy changes into the establishment occupancy fluctuation pattern.

Similarly an appropriate thermal comfort is a crucial factor to take into account in order to control building energy consumption and provide an appropriate interior environment. Effectively determining appropriate thermal comfort is difficult, though, because it relies on personal characteristics and interior environments. Additionally, collecting datasets of indoor environmental and person characteristics may occasionally be unfeasible and challenging in terms of work and resources. Because of this, the study's design concentrated on setting up the system to allow for on-site interviews with residents about their experiences with thermal comfort in various temperature settings using a questionnaire created for residents of various nationalities. The findings show that, depending on nationality, there are comfort margins or ranges that are appropriate to maintain acceptable comfort. It is crucial to create a prediction model for personal thermal comfort using transfer learning in order to transfer information from datasets of people in different indoor and thermal conditions, even when the target subject's physiological and environmental data are insufficient. This will provide a more general approach to guarantee that the HVAC controller always maintains an acceptable comfort level regardless of the nationality. This may be accomplished by using wearable wristbands and sensors to collect data on the interior environment and each subject's physiological condition. Then, by combining machine learning and deep learning techniques, the datasets may be utilised to produce a pre-trained model. Based on the pre-trained model, the transfer learning technique might be utilised to improve the poor generalisation performance caused by insufficient datasets for each target subject.

7.0 Conclusion

Over the past ten years, one of the main emphases of building energy efficiency research has been building occupancy modelling and prediction. Despite the fact that invasive technologies, such as camera-based occupancy, have demonstrated strong effectiveness in forecasting building occupancy, their use in residential buildings has lately decreased because of privacy concerns. One of the major research challenges now facing the majority of earlier techniques is occupancy overlapping, which impairs prediction model accuracy and, therefore, HVAC operating performance, causing discomfort. This study thus concentrated on a non-intrusive environmental sensing method for estimating room occupancy. Depending on the occupancy thermal comfort needs, the smart controller uses the prediction's result as one of its input parameters to manage HVAC operation under three distinct settings. A prototype was created and connected to gather occupancy-related data in the building for model training in order to evaluate the effectiveness of the suggested model. Utilising a random forest regressor, the building occupancy was predicted. A questionnaire was created to ascertain the optimum thermal perception of participants from various nations in order to test and evaluate the suggested HVAC controller against energy usage and total occupant thermal satisfaction.

Despite this, psychological and social variables can affect how people react to heat perception, and these elements can change or fluctuate depending on thermal adaptation. Differences in PET calibration between participants in various nations show how a person's social characteristics impact their thermal perception. The PET neutral comfort value range for Yoruba and Igbo is 16°C, which is frigid to Hausa and Fulani. When comparing thermal perception, this illustrates how the comfort zone varies by about 4°C. This indicates that the calibration needs to be adjusted for the particular climatic zone and that a perfect PET calibration is not feasible. Therefore, in order to mitigate the consequences of uncomfortable conditions, it is essential to understand the sociological and psychological factors that influence how people perceive their environment when developing HVAC controllers, determining comfort levels, and making decisions pertaining to urban development. Lastly, the testing findings demonstrate that the suggested controller can run the HVAC system while maintaining a comfortable level of comfort and saving up to 45% of the energy used with the prior method. To improve building occupancy detection, future research might look at integrating a number of deep learning techniques with the suggested framework.

Acknowledgments: The authors wish to thank Waziru Umar Federal Polytechnic Birnin Kebbi, Nigiera for providing a valuable resource to carry out this study.

Ethics Approval and Consent to Participate: Graduate Research, Waziru Umar Federal Polytechic obtains ethical approval.

Competing Interests: The authors declare that they have no competing interests.

The availability of Data

Available upon request

Funding: The authors are thankful to the TETFUND Nigeria for funding this work under the local resesearch program grant.

Author Contributions:
All authors contributed to the research as follows: “Conceptualization — Shahidu Ibrahim, Methodology— Ibrahim A. Bello.; Formal analysis— Shahidu Ibrahim & Ibrahim A. Bello; Resources— Shahidu Ibrahim Writing— Abubakar Bashir, Original draft preparation— Ibrahim A. Bello & Abubakar Bashir.

References

[1] M. S. Aliero, K. N. Qureshi, M. F. Pasha, and G. Jeon, "Smart Home Energy Management Systems in Internet of Things networks for green cities demands and services," Environmental Technology & Innovation, vol. 22, 2021.
[2] M. S. Aliero et al., "Non-Intrusive Room Occupancy Prediction Performance Analysis Using Different Machine Learning Techniques," Energies, vol. 15, no. 23, 2022.
[3] N. Cao, J. Ting, S. Sen, and A. Raychowdhury, "Smart Sensing for HVAC Control: Collaborative Intelligence in Optical and IR Cameras," IEEE Transactions on Industrial Electronics, vol. 65, no. 12, pp. 9785-9794, 2018.
[4] X. Jin, G. Wang, Y. Song, and C. Sun, "Smart building energy management based on network occupancy sensing," Journal of International Council on Electrical Engineering, vol. 8, no. 1, pp. 30-36, 2018.
[5] H. Sadikoglu-Asan, "‘User-Home relationship’ regarding user experience of smart home products," Intelligent Buildings International, pp. 1-17, 2020.
[6] E. d. L. Oliveira, R. D. Alfaia, A. V. F. Souto, M. S. Silva, C. R. L. Francês, and N. L. Vijaykumar, "SmartCoM: Smart Consumption Management Architecture for Providing a User-Friendly Smart Home based on Metering and Computational Intelligence," Journal of Microwaves, Optoelectronics and Electromagnetic Applications, vol. 16, no. 3, pp. 736-755, 2017.
[7] F. Baig, A. Mahmood, N. Javaid, S. Razzaq, N. Khan, and Z. Saleem, "Smart home energy management system for monitoring and scheduling of home appliances using zigbee," J. Basic. Appl. Sci. Res, vol. 3, no. 5, pp. 880-891, 2013.
[8] J. Shin, Y. Park, and D. Lee, "Who will be smart home users? An analysis of adoption and diffusion of smart homes," Technological Forecasting and Social Change, vol. 134, pp. 246-253, 2018.
[9] M. V. Moreno, M. A. Zamora, and A. F. Skarmeta, "User-centric smart buildings for energy sustainable smart cities," Transactions on Emerging Telecommunications Technologies, vol. 25, no. 1, pp. 41-55, 2014.
[10] H. Na, J.-H. Choi, H. Kim, and T. Kim, "Development of a human metabolic rate prediction model based on the use of Kinect-camera generated visual data-driven approaches," Building and Environment, vol. 160, 2019.
[11] Zeynep Duygu Tekler, Raymond Low, Chau Yuen, and L. Blessing, "Plug-Mate: An IoT-based Occupancy-Driven Plug Load Management System in Smart Buildings," Building and Environment, vol. 223, 2022.
[12] Avgoustinos Filippoupolitis, William Oliff, and G. Loukas, "Bluetooth Low Energy based Occupancy Detection for Emergency Management," 2016 15th International Conference on Ubiquitous Computing and Communications 2016.
[13] Z. D. Tekler and A. Chong, "Occupancy prediction using deep learning approaches across multiple space types: A minimum sensing strategy," Building and Environment, vol. 226, 2022.
[14] M. S. Aliero, M. F. Pasha, A. N. Toosi, and I. Ghani, "The COVID-19 impact on air condition usage: a shift towards residential energy saving," Environ Sci Pollut Res Int, vol. 29, no. 57, pp. 85727-85741, Dec 2022.
[15] B. Abade, D. Perez Abreu, and M. Curado, "A Non-Intrusive Approach for Indoor Occupancy Detection in Smart Environments," Sensors (Basel), vol. 18, no. 11, Nov 15 2018.
[16] C. Brennan, G. W. T. , and P. S. , "Designing Learned CO2-based Occupancy Estimation in Smart Buildings," IET Research Journals, 2015.
[17] O. Hänninen et al., "Analysis of CO2 monitoring data demonstrates poor ventilation rates in Albanian schools during the cold season," Air Quality, Atmosphere & Health, vol. 10, no. 6, pp. 773-782, 2017.

[18] S. Zhai, Z. Wang, X. Yan, and G. He, "Appliance Flexibility Analysis Considering User Behavior in Home Energy Management System Using Smart Plugs," IEEE Transactions on Industrial Electronics, vol. 66, no. 2, pp. 1391-1401, 2019.

[19] A. M. Thomas et al., "Smart care spaces: needs for intelligent at-home care," International Journal of Space-Based and Situated Computing, vol. 3, no. 1, pp. 35-44, 2013.

[20] M. S. Zuraimi, A. Pantazaras, K. A. Chaturvedi, J. J. Yang, K. W. Tham, and S. E. Lee, "Predicting occupancy counts using physical and statistical Co2-based modeling methodologies," Building and Environment, vol. 123, pp. 517-528, 2017.

[21] P. Wei et al., "Impact Analysis of Temperature and Humidity Conditions on Electrochemical Sensor Response in Ambient Air Quality Monitoring," Sensors (Basel), vol. 18, no. 2, Jan 23 2018.

[22] M. Yazici, O. Ceylan, A. Shafique, S. Abbasi, A. Galioglu, and Y. Gurbuz, "A new high dynamic range ROIC with smart light intensity control unit," Infrared Physics & Technology, vol. 82, pp. 161-169, 2017.

[23] M. S. Aliero, M. Asif, I. Ghani, M. F. Pasha, and S. R. Jeong, "Systematic Review Analysis on Smart Building: Challenges and Opportunities," Sustainability, vol. 14, no. 5, 2022.

[24] H. L. J. Ahmad, R. Emmanuel and M. Mannion, "Occupancy detection in non-residential buildings – A survey and novel privacy preserved occupancy monitoring solution," Applied Computing and Informatics, vol. Vol. 17 No. 2
pp. pp. 279-295, 2021 2021.

[25] Bharathan Balaji, J. X. , A. N. , R. G. , and Y. A. , "Sentinel: Occupancy Based HVAC Actuation using Existing WiFi Infrastructure within Commercial Buildings," SenSys '13: Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, 2013.

[26] "Wearable-Based Human Activity Recognition Using an IoT Approach," Journal of Sensor and Actuator Networks, vol. 6, no. 4, 2017.

[27] R. Low, Z. D. Tekler, and L. Cheah, "An End-to-End Point of Interest (POI) Conflation Framework," ISPRS International Journal of Geo-Information, vol. 10, no. 11, 2021.

[28] S. P. Manar Amayri , Nizar Bouguila , Frederic Wurtz "Database quality assessment for interactive learning: Application to occupancy estimation," Energy & Buildings, vol. 209 (2020) 109578, 2020.

[29] M. A. Abuhussain, B. S. Alotaibi, M. S. Aliero, M. Asif, M. A. Alshenaifi, and Y. A. Dodo, "Adaptive HVAC System Based on Fuzzy Controller Approach," Applied Sciences, vol. 13, no. 20, 2023.

[30] A. Talebi and A. Hatami, "Online fuzzy control of HVAC systems considering demand response and users' comfort," Energy Sources, Part B: Economics, Planning, and Policy, vol. 15, no. 7-9, pp. 403-422, 2020.

[31] L. M. Candanedo and V. Feldheim, "Accurate occupancy detection of an office room from light, temperature, humidity and CO 2 measurements using statistical learning models," Energy and Buildings, vol. 112, pp. 28-39, 2016.

[32] M. Aftab, C. Chen, C.-K. Chau, and T. Rahwan, "Automatic HVAC control with real-time occupancy recognition and simulation-guided model predictive control in low-cost embedded system," Energy and Buildings, vol. 154, pp. 141-156, 2017.

[33] S. Lu, W. Wang, S. Wang, and E. Cochran Hameen, "Thermal Comfort-Based Personalized Models with Non-Intrusive Sensing Technique in Office Buildings," Applied Sciences, vol. 9, no. 9, 2019.

[34] J. Guo, M. Amayri, F. Najar, W. Fan, and N. Bouguila, "Occupancy estimation in smart buildings using predictive modeling in imbalanced domains," Journal of Ambient Intelligence and Humanized Computing, vol. 14, no. 8, pp. 10917-10929, 2022.

[35] C.-C. Huang, R. Yang, and M. W. Newman, "The potential and challenges of inferring thermal comfort at home using commodity sensors," presented at the Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp '15, 2015.

[36] Q. Huang, "Occupancy-Driven Energy-Efficient Buildings Using Audio Processing with Background Sound Cancellation," Buildings, vol. 8, no. 6, 2018.

[37] D. Yang, B. Xu, K. Rao, and W. Sheng, "Passive Infrared (PIR)-Based Indoor Position Tracking for Smart Homes Using Accessibility Maps and A-Star Algorithm," Sensors (Basel), vol. 18, no. 2, Jan 24 2018.

[38] Z. D. O'Neill, Y. Li, H. C. Cheng, X. Zhou, and S. T. Taylor, "Energy savings and ventilation performance from CO2-based demand controlled ventilation: Simulation results from ASHRAE RP-1747 (ASHRAE RP-1747)," Science and Technology for the Built Environment, vol. 26, no. 2, pp. 257-281, 2019.

10.48047/jocaaa.2025.34.07.13

[39] M. R. Sisco, V. Bosetti, and E. U. Weber, "When do extreme weather events generate attention to climate change?," Climatic Change, vol. 143, no. 1-2, pp. 227-241, 2017.
[40] H. Elkhoukhi et al., "A platform architecture for occupancy detection using stream processing and machine learning approaches," Concurrency and Computation: Practice and Experience, 2019.
[41] K. Sun, I. Qaisar, M. A. Khan, T. Xing, and Q. Zhao, "Building occupancy number prediction: A Transformer approach," Building and Environment, vol. 244, 2023.
[42] F. Wang et al., "Predictive control of indoor environment using occupant number detected by video data and CO 2 concentration," Energy and Buildings, vol. 145, pp. 155-162, 2017.
[43] W. Jung and F. Jazizadeh, "Human-in-the-loop HVAC operations: A quantitative review on occupancy, comfort, and energy-efficiency dimensions," Applied Energy, vol. 239, pp. 1471-1508, 2019.
[44] N. S. Fayed, M. Abu-Elkheir, E. M. El-Daydamony, and A. Atwan, "Sensor-based occupancy detection using neutrosophic features fusion," Heliyon, vol. 5, no. 9, p. e02450, Sep 2019.
[45] L. F. C. Maschi, A. S. R. Pinto, R. I. Meneguette, and A. Baldassin, "Data Summarization in the Node by Parameters (DSNP): Local Data Fusion in an IoT Environment," Sensors (Basel), vol. 18, no. 3, Mar 7 2018.
[46] J. P. Roselyn, R. A. Uthra, A. Raj, D. Devaraj, P. Bharadwaj, and S. V. D. Krishna Kaki, "Development and implementation of novel sensor fusion algorithm for occupancy detection and automation in energy efficient buildings," Sustainable Cities and Society, vol. 44, pp. 85-98, 2019.
[47] Z. Zhang, J. Wang, H. Zhong, and H. Ma, "Optimal scheduling model for smart home energy management system based on the fusion algorithm of harmony search algorithm and particle swarm optimization algorithm," Science and Technology for the Built Environment, vol. 26, no. 1, pp. 42-51, 2019.
[48] Y.-Y. Choo, Y.-J. Ha, Y.-B. Kim, S.-J. Lee, and H.-D. Choi, "Development of CoAP-based IoT Communication System for Smart Energy Storage System," presented at the Proceedings of the 2nd International Symposium on Computer Science and Intelligent Control - ISCSIC '18, 2018.
[49] G. Fan et al., "Investigation of indoor thermal environment in the homes with elderly people during heating season in Beijing, China," Building and Environment, vol. 126, pp. 288-303, 2017.
[50] Z. Chen, M. K. Masood, and Y. C. Soh, "A fusion framework for occupancy estimation in office buildings based on environmental sensor data," Energy and Buildings, vol. 133, pp. 790-798, 2016.
[51] "The Internet Information and Technology Research Directions based on the Fourth Industrial Revolution," KSII Transactions on Internet and Information Systems, vol. 10, no. 3, 2016.
[52] "Internet of Things based Smart Energy Management for Smart Home," KSII Transactions on Internet and Information Systems, vol. 13, no. 6, 2019.
[53] B. Gregorutti, B. Michel, and P. Saint-Pierre, "Correlation and variable importance in random forests," Statistics and Computing, vol. 27, no. 3, pp. 659-678, 2016.
[54] G. Zittis, "Observed rainfall trends and precipitation uncertainty in the vicinity of the Mediterranean, Middle East and North Africa," Theoretical and Applied Climatology, vol. 134, no. 3-4, pp. 1207-1230, 2017.
[55] R. Zhu, D. Zeng, and M. R. Kosorok, "Reinforcement Learning Trees," J Am Stat Assoc, vol. 110, no. 512, pp. 1770-1784, 2015.
[56] M. S. Aliero, M. F. Pasha, A. N. Toosi, and I. Ghani, "The COVID-19 impact on air condition usage: a shift towards residential energy saving," Environ Sci Pollut Res Int, Jan 10 2022.
[57] R. Adhikari, M. Pipattanasomporn, and S. Rahman, "An algorithm for optimal management of aggregated HVAC power demand using smart thermostats," Applied Energy, vol. 217, pp. 166-177, 2018.
[58] "The Python Standard Library," 2021.