



ORIGINAL ARTICLE

A NOVEL FRAMEWORK FOR SMART BUILDING APPLICATIONS IN THE IDENTIFICATION OF VACANT SPACE

¹*S. Ananthi, ²Antony Vijay J

¹ Department of Artificial Intelligence and Data Science, St. Joseph's College of Engineering, Chennai, India

² Department of Computer Science and Engineering, R M K College of Engineering and Technology, Chennai, India

Corresponding Author Email: sugaananthi@gmail.com

(Received 17 April 2025; revised 20 September 2025; accepted 20 November 2025)

Abstract

The intelligent use of building spaces is one of the main demands of the intelligent infrastructure of the present time. The quick advancements in the sphere of smart building technologies and Internet of Things (IoT) systems have allowed constant monitoring of indoor areas, but the correct identification of empty and occupied areas remains a rather problematic issue because of dynamic human behavior and uncertainties in sensors and uneven information sources. The present paper suggests a new paradigm of the smart building application that would help to correctly identify the vacant spaces based on smart data-driven methodology. The framework unites the information gathered through the distributed sensors including motion, occupancy, environmental, and energy usage sensors, with the state-of-the-art machine learning algorithms to examine the real-time and historical trends in space use. Powerful preprocessing and feature extraction algorithms are used to deal with noise, missing, and time variations. The main goal of the suggested framework is to facilitate the quality and reliable detection of the vacant space besides being scalable and energy efficient. Experimental assessments indicate that the framework is efficient in estimating vacant spaces to a high level of accuracy, allowing to optimize the allocation of resources, energy conservation, and comfort of occupants. The suggested solution offers an effective and sensible solution to smart buildings management and sustainable development of infrastructure.

Keywords: *occupancy; remote monitoring; raw sensor; fusing; heating, ventilation and air conditioning; internet of things*

1. Introduction

Smart environments are one of the emerging industries the IoT is facilitating, and the market and the research community are starting to pay more attention to them [1-2]. They provide both private and public environments with the capability of managing the freedom to handle cyber-physical systems autonomously with information obtained via the IoT actuators and sensors. These applications can lead to a more sustainable method in energy and resource management, aside from their great potential to enhance the lives of people in their day-to-day activities [3]. Specifically with applications that include human goals in the self-management loop of intelligent environments, it is evident that there exists an issue of balancing divergent goals of various users [4]. Even over a brief period, colleagues working in a room of a communal building may hold vastly different opinions regarding the intensity of the illumination they would maintain. Today, building management and operation can be enhanced by big data analytics and internet of things in buildings, which promise great opportunities [5]. Artificial intelligence and machine learning are being introduced to the construction industry to aid in the application of the internet of things in conjunction with other intelligent technologies,[6] to provide smart management and automation of buildings. This normally helps in collecting different kinds of data created by the occupants of a building or equipment employed in construction cases [7]. Moreover, the number of variables, which influence the level of electricity consumption, is broad. Some of the physical factors are the building features, efficiency of the equipment and the weather [8]. Nevertheless, these features cannot

be easily controlled or modified by an end user during the period when a building is being used [9]. The entry level stage of resident behavior modeling is what is called occupancy. In order to explain, the actions of the occupancy-related meaning conclude the presence of occupants, their patterns of power utilization, and situations within the building [10-11]. Learning presence patterns is significant as far as automation systems are concerned. This relevance is due to its ability to regulate electrical appliances and methods that may save a substantial amount of energy used in commercial and residential buildings such as lighting, air conditioning and ventilation [12, 13]. Its comparison with the Bayesian approach has therefore been observed to be a more flexible and general approach and has been applied successfully in a variety of applications [14-15]. Also, complex HVAC controls systems can be interfaced using this framework such that occupants of the building with smart devices can control the temperature. The owner of the home is therefore able to monitor and control the system remotely. The issue of monitoring is put in a set of characteristics, which define how a specific class performs. This section will discuss categorization designs the designs used to detect occupancy have been used in various works in the past [16]. The most common methods of classification in this respect are decision trees which are simple and easy to read. Researchers also employed three other classification models besides DT in 1) random forest; 2) gradient boosting machines; and 3) linear discriminant analysis. In a similar work, occupancy modeling with twelve ambient sensor parameters was carried out. The single and multiple occupancy rooms are the experimental design used to determine the efficacy of the six statistical-learning models [17]. The authors further gave a simulation of the three-month energy consumption of the HVAC systems in EnergyPlus and a DesignBuilder and they developed the occupancy-based demand-response HVAC control mechanism and managed to save on 18% of electricity and 20% of gas [18].

The high rates of urbanization and the need to have more energy-efficient infrastructure have promoted the use of smart building technologies. Smart buildings are using Internet of Things (IoT) sensors, communication networks, and smart data analytics to manage and monitor indoor environments in real time. In smart building management, the proper detection of occupied and empty areas is one of the risks that needs to be addressed as it has a direct impact on energy efficiency, the space utilization, safety, and comfort of occupants. Good identification of vacant areas will allow the smart control of lighting, heating, ventilation, and air conditioning (HVAC) systems, thereby saving on energy and operational expenses. Conventional methods of space occupancy detection usually consider the use of rudimentary motion sensors, access control records, or set schedules. Although these techniques offer fundamental occupancy data, they are likely to cause systematic errors due to sensor noise, non-uniform human activity, and inadequate spatial resolution. Furthermore, rule-based systems are not flexible and they do not reflect complicated occupancy patterns in dynamic settings like offices, educational Centre's and commercial premises. The constraints emphasize the necessity of sophisticated and intelligent systems that can detect unoccupied spaces with a high degree of reliability in different conditions. The current developments in machine learning and data analysis have provided additional possibilities of the smart building in terms of intelligent occupancy and vacancy detection. Machine learning can be used to understand concealed trends on space usage by examining the data of various heterogeneous sensors, including passive infrared (PIR) sensors, cameras, environmental sensors, and smart meters. Nevertheless, current solutions are quite frequently characterized by issues of heterogeneity of data, scalability, and privacy, as well as the lack of an integrated framework with sensing, data processing, and decision-making mechanisms that integrate in a seamless manner. To meet these issues, this study hypothesizes a new paradigm of smart building applications that is based on proper identification of empty spaces. The suggested framework is a combination of multi-sense data capture and powerful preprocessing, feature identification and smart learning structures to enhance detection accuracy and reliability. The primary aim is to facilitate smart building control through the provision of effective space usage, energy saving, and improved occupant experience. The suggested framework also helps to create sustainable and smart built environments because it offers a scalable and versatile solution.

2. Related Works

In the following section we will give a quick summary of each of the four classification models we have compared in the context of this paper, namely GBM, RF classification, and LDA regression tree. The readers are directed to other information on the models. Cart is aimed at constructing models of

classification using a popular model of machine-learning on data [19]. The model would be constructed by constructing a prediction model on each partition of the data space and then the data space will be partitioned. CART has a number of advantages in comparison with other classification techniques. First, it does not require information about the underlying variable distribution of the predictor class. Second, CART is able to deal with both categorical and numerical data that has high skewness [20, 21]. The model as with any other boosting method is built in stages. The model is ultimately generalized, by enhancing a freely selected differentiable loss function [22]. In cases where the frequencies in a class are disproportionate, LDA is effective in managing the state of affairs. In order to do this, we generate labeled datasets, reduce the training time, and reduce the period issues of training that are hard and significant to solve to ensure sufficient learning performance. Transfer learning is one of the possible solutions to these issues, and this proposal was only recently developed as a solution to allow the creation of systems grounded in the principles of ML/DL and with multiple advantages [23]. TL is a potentially effective machine learning technique that focuses on passing information across domains. The ability of human beings to transfer the knowledge in different disciplines was used as a motivation to come up with the idea; TL aim is to apply the information available in a related direction to improve the learning outcomes or whether fewer labeled examples in the target domain are necessary or not. Usually, DL helps to conserve resources and be more effective in the process of training the new model. Because of the offline pretrained capacity of the ML/DL models on larger datasets that can be trimmed (on smaller datasets); The most important step is to validate ML/DL models on unlabeled datasets after training them on the annotated datasets available, it is known that labeling data is a difficult undertaking requiring expert help, time and effort; the knowledge accumulated on existing tasks can be used to make better generalizations about other people using the knowledge of previous tasks without the need to retrain a model, even though synthetic or simulated data can be used instead of actual data.

The modern world is in great need of Internet of Things (IoT)-based devices capable of sensing the presence of people in intelligent structures. This is the design of such a system based on different statistical and mathematical models which has been tried many times before. The mathematical model applied in this article DSET is an occupancy detector tool. The results of samples with a high correlation between the individual sensor readings can be even superior when the sensor measurements are combined in large numbers which was the impetus behind the research in this paper. The actual reason behind smart building research is the necessity to have a more efficient use of energy and even better comfort of the occupants. The simplest sensor networks used by early researchers were motion sensors and door contacts sensors to determine occupancy conditions. An example is motion-based techniques which have been popular due to their low cost and simplicity to install, but have the disadvantage of having high false-positive and low false-negative rates (when there is motion, but no occupancy) and highly transient human movements that are not detected (false-negative). These constraints highlighted the necessity of more advanced strategies that exceed the use of the single sensor inference.

In order to work around the limitations of the fundamental sensing mechanisms, scientists started investigating the multi-sensor fusion technologies, including the integration of information in heterogeneous sources with temperature, CO₂ levels, light intensity, and sound sensors. Some of the studies showed that occupancy detection would be more accurately detected using a combination of environmental sensors and motion sensors than when any of these sensors are used alone. As an example, the fusion models of sensor data and Bayesian inference approaches have been employed to use the Kalman filter to achieve more robust predictions in different conditions. Nonetheless, this type of fusion strategy can be very tricky and cannot always be scaled to larger building implementations with different sensor types. During the introduction of machine learning (ML), data-intensive methods have more often been used in occupancy and vacant space detection. The Support Vector Machines (SVM), the Random Forest (RF), and the Artificial Neural Networks (ANN) are supervised learning models, which have been trained using the historical labeled data in order to classify occupancy states using sensor inputs. These models showed much better prediction and adaptation to complicated patterns in the indoor environment. However, the conventional ML models remain very reliant on feature engineering and, thus, lack comprehensive temporal dynamics of space usage without a lot of preprocessing. Recurrent neural networks (RNNs) and Long-Short-term memory (LSTM) networks are some of the deep learning models that have been advanced to capture temporal dependency in occupancy patterns. These

models are capable of occupancy states prediction with better performance than the static classifiers through learning of sequential characteristics as time-series sensor data. Convolutional Neural Networks (CNNs) have also been tested on sensor grid data and mixed deep learning systems by researchers to perform automatic extraction of meaningful characteristics. Although such deep learning models are quite predictive, they have higher compute and real-time deployment issues in smart building settings that have limited resources. Over the past few years, there has been a focus on more holistic models that integrate IoT architectures, edge computing and adaptive learning to enable scalable and real-time vacant space detection. A number of works have suggested edge-based ML inference to achieve both latency and privacy by doing sensor processing on the device. Moreover, explainable AI and transfer learning methods have also been explored in order to enhance model interpretability and flexibility in various building scenarios. Nevertheless, there is still no unified framework that can harmoniously incorporate multi-sensor acquisition, intelligent preprocessing, scalable learning, and real-time vacant space identification as a gap that can be bridged by the proposed novel framework in the current study.

3. Proposed System

The proposed system of smart building applications uses a heterogeneous group of sensors within the IoT in order to pinpoint empty spaces indoors. Human motion is detected by the use of Passive Infrared (PIR) motion sensors, and the environmental sensors are used to measure temperature, humidity, CO₂ level, and light intensity which are indirectly used to tell whether a person is present. Moreover, smart energy meters record power usage patterns which also aid occupancy inference. All sensors are distributed in the various building areas and connected by a wireless communication network so that constant flow of data is passed on to an edge or cloud-based unit of processing. The sensor data are captured at regular time intervals and recorded together with timestamps and location identifiers to maintain the time and space context. As the raw IoT data are highly likely to possess noise, gaps in values, and irregular sampling rates, a powerful preprocessing phase is utilized. This involves data cleaning to eliminate outliers, interpolations to fill gaps in readings and temporal synchronization to re-align heterogeneous sensor streams. Normalization and encoding of features are done in order to make them compatible with machine learning models as well as enhance stability and performance of the training. After the preprocessing, the sensor data is processed to extract meaningful features that are used to predict occupancy behavior. Such characteristics are motion activity levels, duration of inactivity, changes in environmental parameters and time variability in using energy. Temporal patterns are highlighted to be able to clearly differentiate between occupied and vacant states. These features are then used to train machine learning models that will be used to identify vacant or occupied spaces. The general approach to the methodology is supposed to be scalable and modular so that it could easily implement more sensors or learning models, and enable the efficient detection of vacant spaces in real time to use a smart management of buildings.

3.1 Dempster-Shafer Rule of Combination

Whenever their intersection happens, the proposals are said to be compatible. The intersection can however give a null set under certain conditions and this would be contradictory to the combination process. Combining evidence from multiple sources is one of the most significant parts of assigning evidence, and the rule suggested by Dempster enables the ability to combine the probability masses presented by different sources of information on the hypotheses that do not make sense. Under DS theory, there is a combination rule which is Dempster that combines data between two independent sources. A confusion matrix is commonly used by statisticians in order to evaluate the performance of a classifier in machine learning. Table 1 described as confusion results.

	Predicted NO	Predicted YES
Actual NO	True Negative	False Positive
Actual YES	False Negative	True Positive

Table 1: Confusion Matrix.

The confusion matrix also yielded other measures, including

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative} \quad (1)$$

$$\text{Sensitivity} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad (2)$$

$$\text{False Positive Rate} = \frac{\text{False Positive}}{\text{True Negative} + \text{False Positive}} \quad (3)$$

$$\text{Specificity} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}} \quad (4)$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{False Positive} + \text{True Positive}} \quad (5)$$

Through this, a theory was proposed named Dempster theory (DSET). DSET is usually used to represent the imprecision or uncertainties of a hypothesis. The hypotheses provide all system states possible. Each of these hypotheses is assigned a probability mass and in combination, the probability masses lead to an outcome. This value is the sum of all the evidence accumulated together, not leading to the relevant hypothesis, divided by a normalization factor of $1-K$. The conflict factor K indicates the degree of difference between the sources of various information. The rule of Dempster is applied successively in case of application of more than two sources of information to give evidence, depending upon the findings of the first application, which were obtained by use of the first two sensor elements, and as the second one, the output in the sensor we are dealing with. Occupancy sensing is a classification problem, as has been already indicated wherein each of the classes is represented by its own characteristic set; the value of which is a result of a successful combination of the information obtained on the source materials of the sensor. The raw data of all the sensors are initially delivered to the data fusion center and then the probability density function is calculated, the PMA is then computed as displayed in Figure 1. The DC is located within the grounds of the building. The decision to do the statistical calculations within the structure per se was made majorly in an attempt to enhance low costs and precision. Honestly, transferring any computational load to the cloud would bring the chances of corruption, or data loss, to our architecture as it is not incorporated on the account of the cost of computation and other factors. Lastly, once the PMAs of the DC have been added, the occupancy determination is performed by these combinations. The decision is then communicated to the building owner who could be miles away via the Internet; with the help of wireless technology on a smart device, say a smartphone.

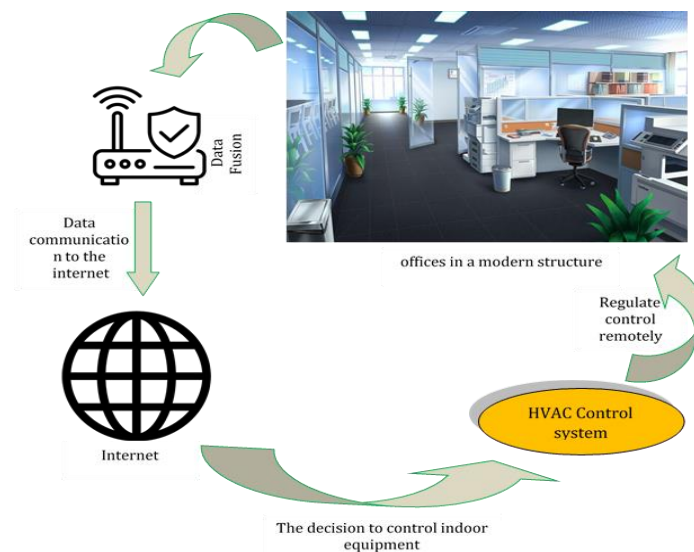


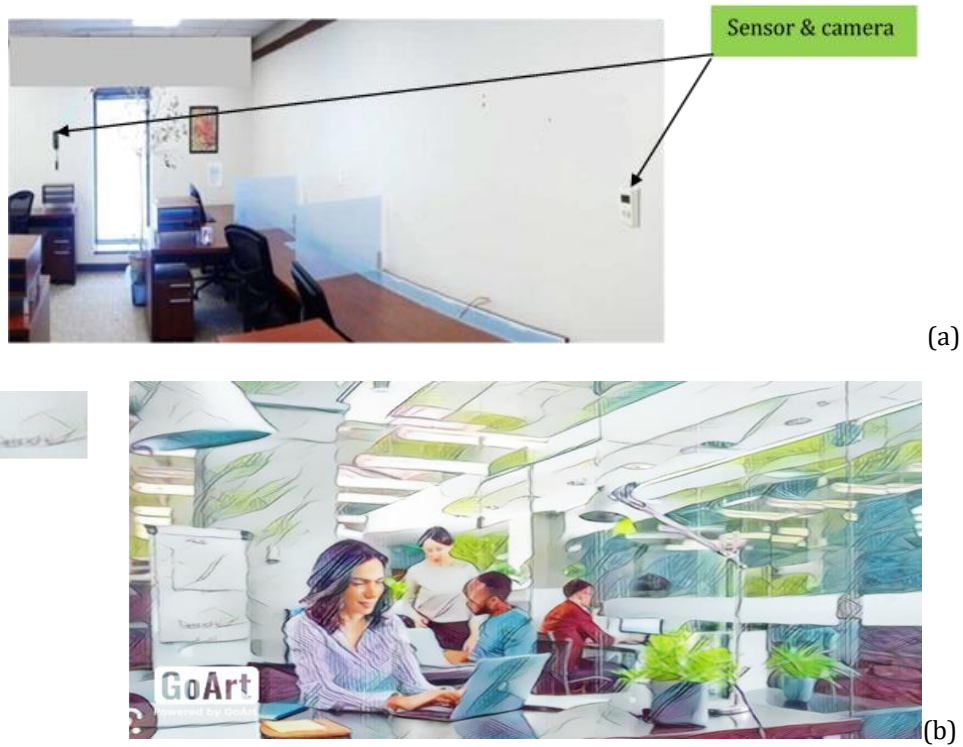
Figure 1: Proposed Architecture

3.2 Dataset Description

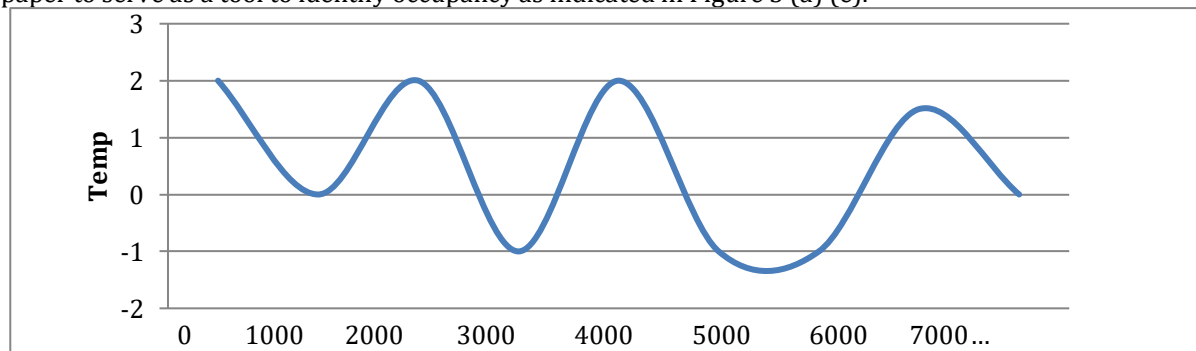
Occupancy data about an office in Belgium was collected in February and used in the study during winter. The space was observed in terms of humidity, temperature, lighting, and CO₂. A microcontroller and a ZigBee radio were used in data collection. The occupancy values obtained using the camera placed on the table in the office are displayed in one of the cases in Figure 2(b), where the sensor and the camera are positioned in a table so that measurements of 1.1 and 2.9 meters were found between the first and the second person. The average data and the summaries of the data were obtained after taking three to four measurements per minute with 14 seconds in between and then averaged the two occupancy values obtained respectively shown in Table 2.

Dataset	Recording date	Occupied readings	Empty readings
Training	2/4/2022-2/10/2022	1730 out of 8144	6415 out of 8144
Testing A	2/4/2022-4/10/2022	973 out of 2666	1695 out of 2666

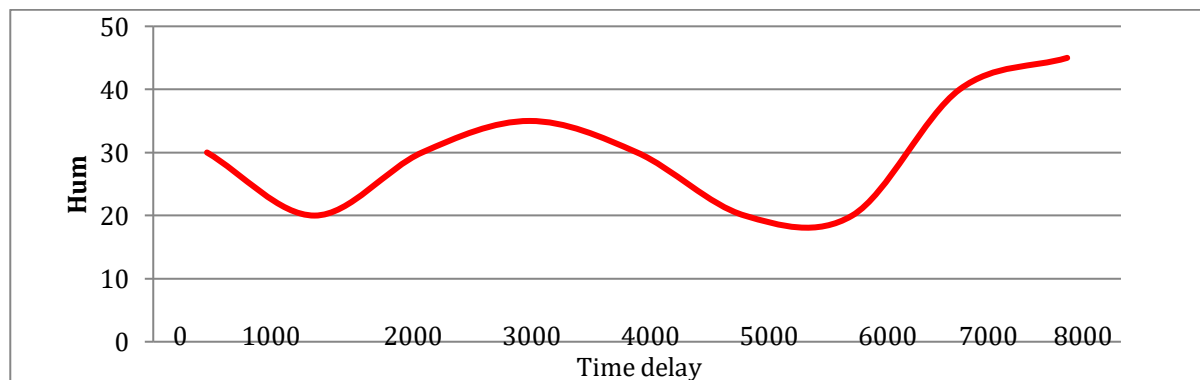
Testing B	11/4/2022-13/10/2022	2050 out of 9753	7704 out of 9753
-----------	----------------------	------------------	------------------

Table 2: Dataset Descriptions Used in this Paper**Figure 2:** (a) Room sensor location (b) Ground occupancy

As can be seen the statistics vary with occupancy of the room, that is, when the room is allegedly crowded the readings will be high and when the room is vacant the readings will be low. Consequently, their parameters exhibit a unique trend that is easily identifiable, which is effectively applied in the current paper to serve as a tool to identify occupancy as indicated in Figure 3 (a) (e).



(a)



(b)

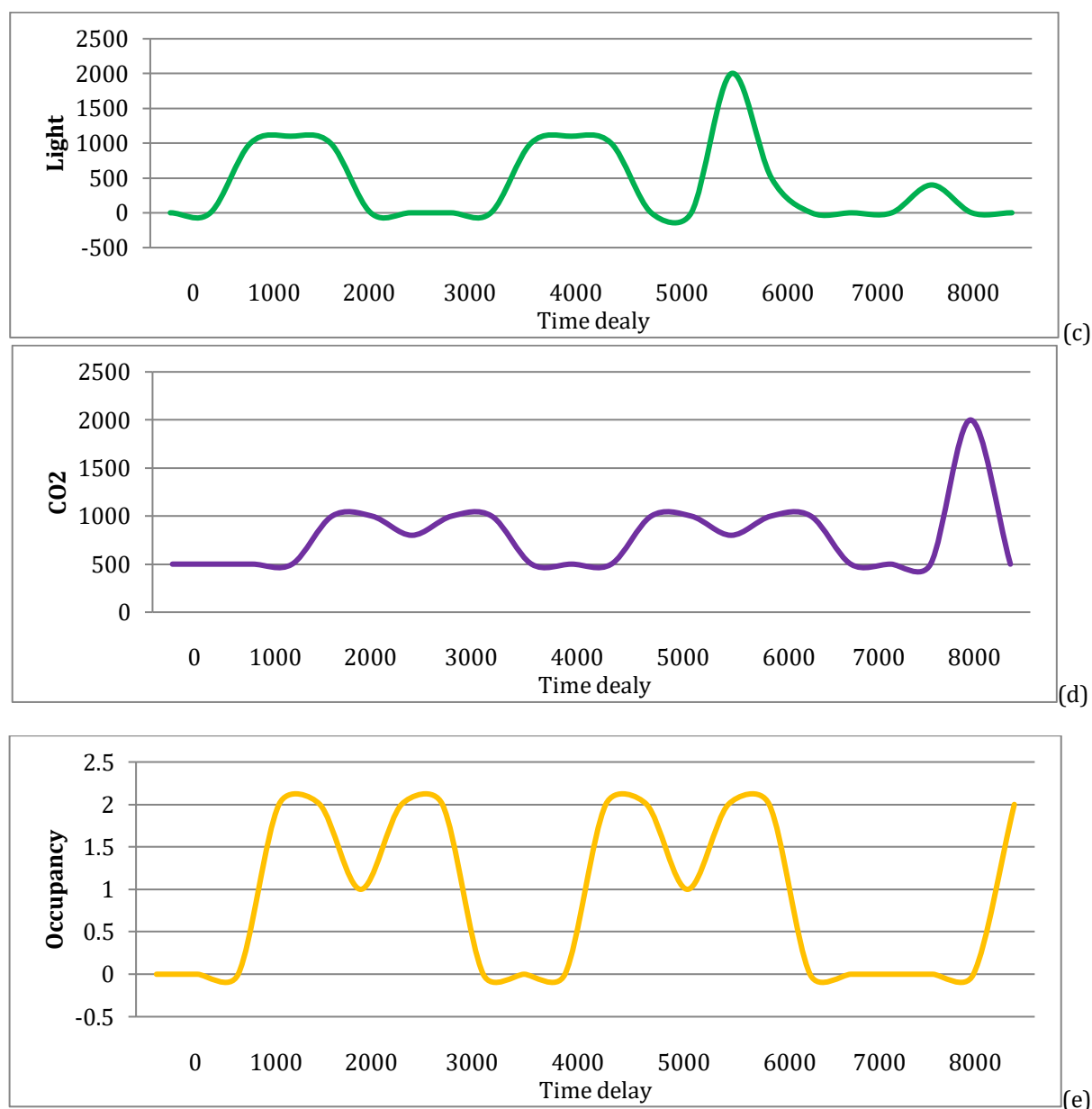
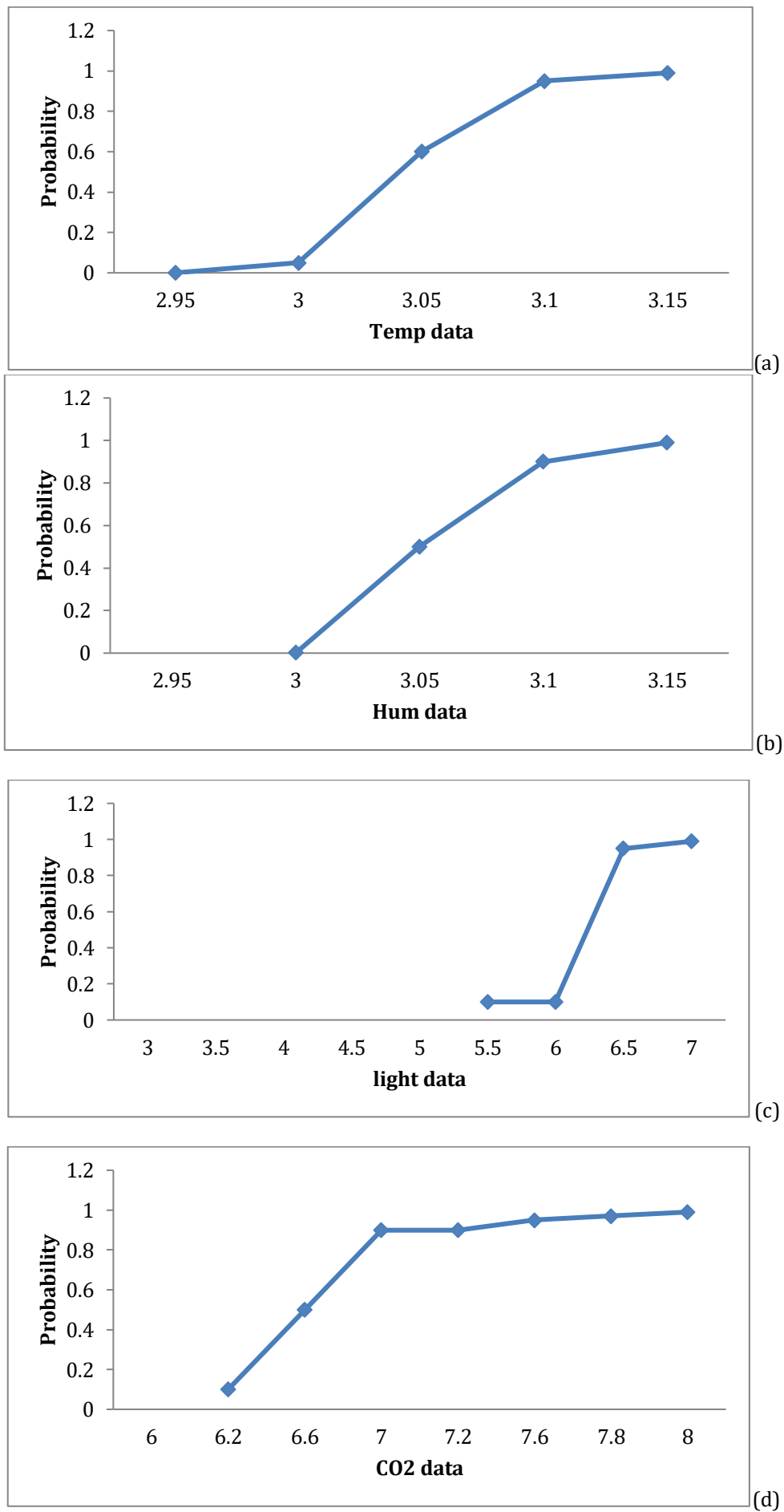


Figure 3: Measurement profiles

Many statistical methods and tests assume that there is a normal distribution of the events. Although there are currently many nonparametric and reliable techniques which do not have such strong distributional assumptions. They, however, are inadequate compared to methods that are based on accurate distributional assumptions. These methods are more favorable in case distributional assumptions have already been checked. The normal distribution is common in statistics to break down complicated phenomena into simpler models. Also, our suggestion is based on the fact that there is a periodic distribution of sensory input. This assumption is informed by a number of factors. To begin with, we have a high level of sample size, and normal distribution provides a good approximation. Second, analytically it has been found that the normal distribution is highly manageable. Thirdly, modeling is also attractive due to its symmetrical nature and familiar shape in the form of a bell. Lastly, it also has essential similarities with the central limit theory. The underlying assumption of our proposed method, It is important to first establish whether the actual distribution is significantly different from the standard (normal) type. A probability plot may be constructed and this serves as a helpful method of checking distributional assumptions. A critical analysis indicates that the normalcy assumption in Figure 4 (a) - (d) is reasonable. Plotting of all the four parameters of the training set are almost straight lines. The larger the sampling dataset, the larger the goodness of fit. In this way, we can conclude that the distribution we are dealing with belongs to the normal family.

**Figure 4:** Normal distribution Probability plot

3.3 Preprocessing

The raw data collected from IoT sensors in smart building environments often contain noise, missing values, and scale inconsistencies, which can negatively affect the performance of learning models. Therefore, a systematic preprocessing stage is applied to enhance data quality and reliability before vacant space identification. Let $X=\{x_1, x_2, \dots, x_n\}$ represent the raw sensor readings collected over time from multiple sensors. Outliers caused by sensor malfunction or sudden environmental disturbances are removed using statistical thresholding, where a data point x_i is considered an outlier if $|x_i - \mu| > k\sigma$, with μ and σ denoting the mean and standard deviation of the sensor data, respectively, and k being a predefined constant.

Missing sensor values are handled using interpolation to preserve temporal continuity. For a missing value at time t , linear interpolation is applied as

$$x(t) = \frac{x(t-1) + x(t+2)}{2} \quad (5)$$

where $x(t-1)$ and $x(t+1)$ represent the nearest available observations before and after the missing timestamp. The strategy provides smooth transitions of time-series data and information loss is avoided. Also, sensor streams sampled at variable rates are resampled in time to ensure that every feature has the same time index.

Normalization is used to remove scale variances amongst heterogeneous sensor features. Min -max normalization converts each feature into the interval $[0,1]$ as follows:

$$x_i^{norm} = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (6)$$

where x_{min} and x_{max} denote the minimum and maximum values of the feature, respectively. This step ensures that no single sensor dominates the learning process due to larger numerical ranges. The preprocessed and normalized dataset thus provides a stable and consistent input representation, enabling accurate and reliable vacant space identification within the proposed smart building framework.

3.4 Designing the Probability Mass Assignment

One of the major barriers to the use of the Dempster- Shafer theory is the strong probability mass function to make. The values below provide the divide between the given trial vector into 2 categories: {occupied, empty} ones. The max denotes the fact that the test vector may come under the best likelihood class of the given large difference in the number of samples under the two classes. 0

Probability Mass Assignment (PMA) is a key procedure of uncertainty modelling, especially in evidence-based decision-making frameworks applied to smart building applications. PMA is a description of distribution of belief or probability mass among various hypotheses, e.g., the vacant state and occupied space state in response to sensor observations. Designing PMA in a proper way also makes sure that the uncertainty due to the presence of noisy, incomplete, or even conflicting sensor information is modeled in a realistic way instead of disregarded.

$$m_y(x) = \frac{P_j(i)}{\sum_i p_j(i) + n_i(1 - \alpha_j \beta_j)} \quad (7)$$

Where $P_j(i)$ The normal distribution probability density function is given as

$$P_j(i) = \frac{1}{\sigma_{ij} \sqrt{2\pi^2}}, \sigma_{ij} > 0 \quad (8)$$

PMA in the proposed framework is obtained on the basis of processed sensor features which include motion intensity, inactive period, changes in the environment and energy usage trends. A sensor is a degree of belief to a particular hypothesis. As an example, long periods of rest and low energy consumption place more probability mass in the empty state whereas frequent movement and high CO₂ concentration add credence to the occupied state. This enables sensor evidence to be mapped into probabilistic representations quantitatively.

Sensor reliability is also added to the design of the PMA in order to build on its strength. Sensors whose contribution to prior information is more accurate or less noisy receive higher confidence mass so that the sources of information which are reliable will have more mass in the final probability. Uncertain or conflicting readings on the other hand put mass in a state of uncertainty or ignorance so that decisions are not premature or incorrect. This assignment incorporating weight enhances stability against sensor failures and external noise. Another aspect that is taken into account during the design of PMA is temporal consistency. Rather than assigning probabilities using isolated observations, time windows of consecutive time are considered in order to smooth sharp changes. This temporal accumulation dampens the updating of belief and minimizes false vacancy or occupancy identification by transient effects, e.g. short-term motion or transient sensor spikes. In general, the developed Probability Mass Assignment offers a working principle of bringing together heterogeneous sensor evidence and clearly modeling uncertainty. PMA can provide a solid basis of other decision fusion and smart building vacancy detection at higher levels through the use of raw sensor data by mapping it into more structured probability distributions.

4. Experimentation

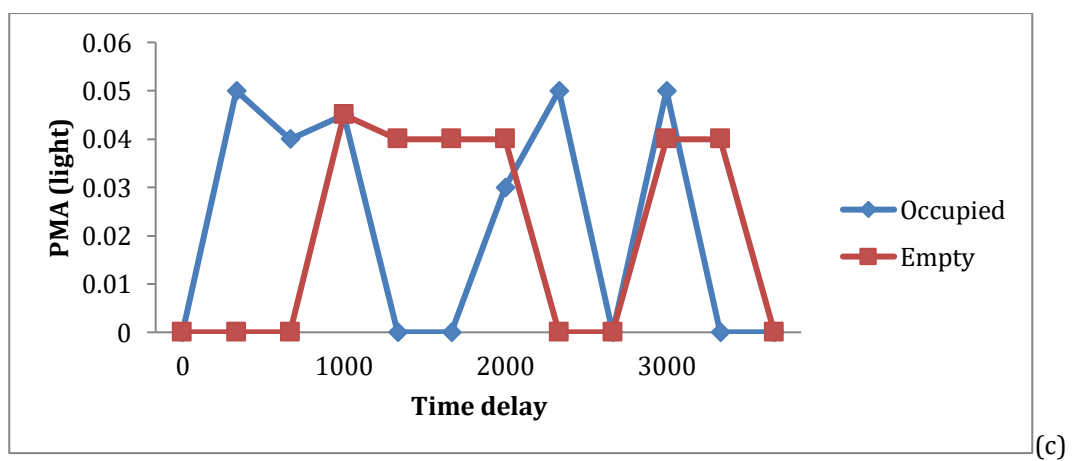
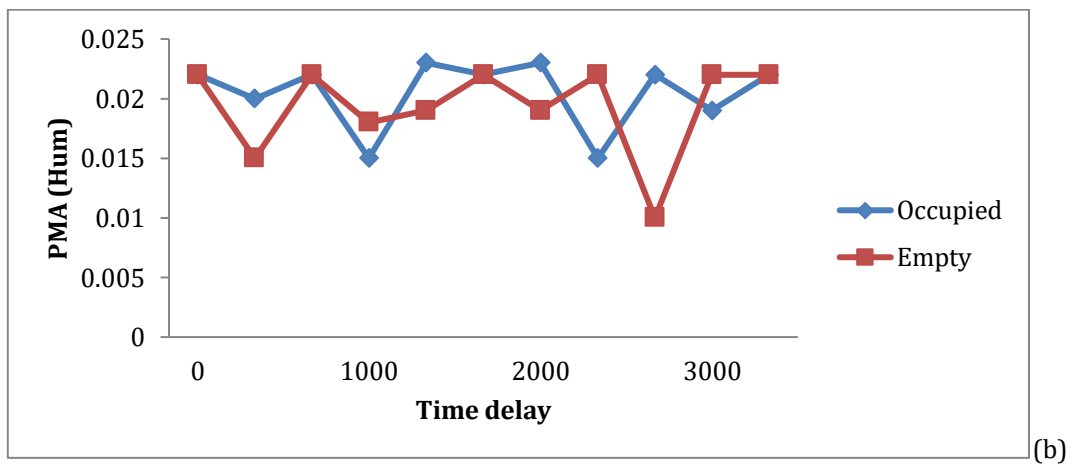
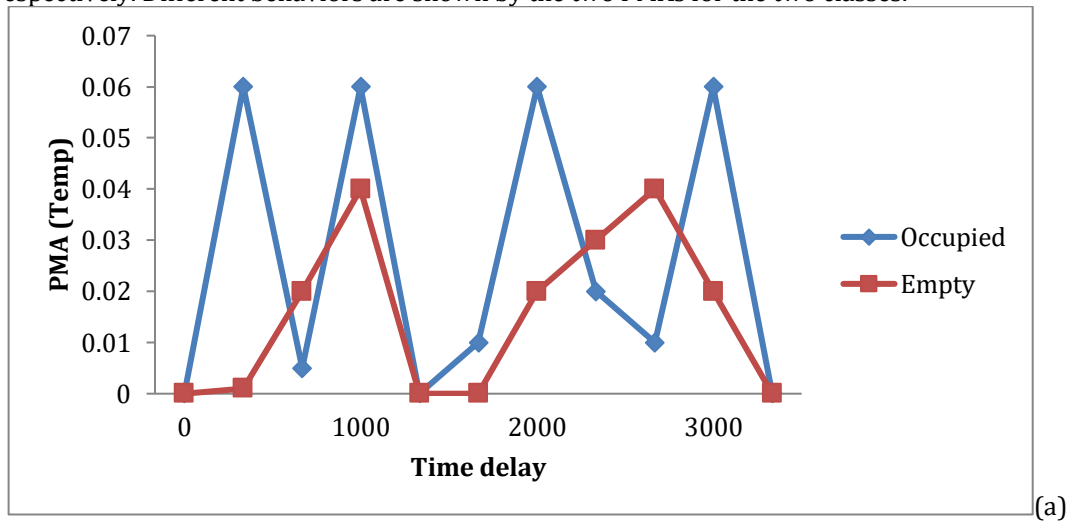
The proposed framework of vacant space identification is experimentally tested within a smart building environment consisting of a distributed network of IoT sensors. Several indoor areas, offices, corridors, and meeting rooms are fitted with Passive Infrared (PIR) motion sensors, environmental sensors (temperature, humidity, CO₂ and light intensity) and smart energy meters. The sensor data is gathered at fixed time intervals and passed through the wireless communication network to an edge processing or cloud processing unit to be analyzed. Sensors are time-stamped and they are connected to a given location to maintain both spatial and temporal context. The dataset obtained is separated into training and testing sets to see how well the learning model can generalize. For all the datasets, preprocessing (removal of noise, treatment of missing values, normalization and time alignment) is uniformly used. The trained machine learning models are then used to categorize spaces as vacant or occupied depending on the preprocessed data. Standard measures (accuracy, precision, recall, and F1-score) are employed to perform evaluation of the performance, whereas the computational efficiency and response time are also calculated to evaluate the applicability in real-time. This laboratory design allows a reasonable and repeatable test of the suggested framework when smart buildings are under real-world conditions.

Category	Hyperparameter	Value	Description
Model Training	Learning Rate	0.001	Controls the step size during model optimization
Model Training	Batch Size	32	Number of samples processed per training iteration
Model Training	Number of Epochs	100	Maximum number of training iterations
Model Training	Loss Function	Binary Cross-Entropy	Measures classification error
Model Training	Optimizer	Adam	Adaptive gradient-based optimization
Regularization	Dropout Rate	0.3	Prevents overfitting by randomly deactivating neurons
Feature Extraction	Time Window Size	10 minutes	Temporal window for occupancy pattern analysis
Feature Extraction	Feature Normalization	Min-Max Scaling	Scales features to [0,1] range
Probability Assignment	Sensor Reliability Weight	0.7–1.0	Confidence weight assigned to reliable sensors

Table 3: Hyper parameter settings

The proposed smart building vacant space identification framework hyperparameters are chosen based on the systematic experimentation to guarantee the optimal performance and the generalization. The learning rate of 0.001 achieves a stable and efficient convergence when training a model, and the batch size of 32 balances the efficiency in computing and gradient stability. The model will be trained to a maximum of 100 epochs with early stopping occurring after 10 epochs when the validation performance has not improved to avoid overfitting. A dropout rate of 0.3 is used in the training process to improve the generalization, and reduce the complexity of the model since the loss function is binary cross-entropy, which is ideal in the binary classification task, and Adam is used as an optimizer due to its adaptive learning ability. The temporal feature extraction is carried out with a fixed time window of 10 minutes, which is a sufficient size of time window to detect occupancy patterns without causing too much delay in the decision-making. Minmax scaling is applied to all sensor characteristics to have equal input values and a consistent learning process. When assigning probability masses, sensor reliability weights (between 0.7 and 1.0) are used so that more high reliability sensors can produce more weight in the final decision. Lastly, we have a vacancy decision threshold of 0.5 which is applied to determine whether the spaces are vacant or occupied. All these hyperparameters are carefully fine-tuned and they have the collective effect of making the suggested framework robust, accurate, and reproducible. For any parameter y , the probability density function $y(x)$ for the two classes is determined after the sensory data is confirmed to be normal {occupied, empty} is calculated for each test vector input δ . Thus, for each sample the corresponding $p_y(\text{occupied})$, test vector δy , and $p_y(\text{empty})$ are determined. The PMA for each parameter of the test sample for each class x is then constructed using our proposed method (7). The result of this step is two distinct $a \times k$ PMA matrices, where k is the number of parameters and a is the number of samples included in the testing dataset. Referring to Table 2, the values of a and k for testing sets A and B, respectively {temp, hum, light, Co₂} are 2665 and 9752. Suppose a test vector $\delta = \{\delta_a, \delta_b, \delta_c, \delta_d\}$ is given as input to the system. The resulting PMA matrices for each class x are made up of the PMAs. Using our technique and set A for single parameter testing, Figure 5 shows a graph of the PMAs as a function of the

time of day. As is evident, the behavior of filled and empty classes is represented by the graphs in red and blue, respectively. Different behaviors are shown by the two PMAs for the two classes.



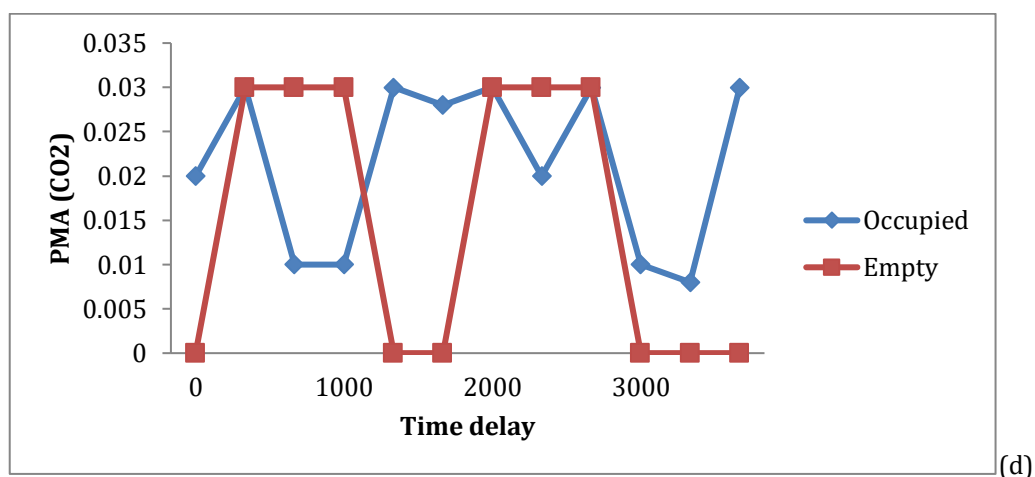


Figure 5: Parameters of PMA

4.1 DSET Combination Rule

The DSET combination rule is utilized when combining two or more parameters. The class with the most evidence is then given preference in the decision. In light of the aforementioned illustration, it is necessary to decide whether a room is occupied. The outcome is determined by considering at the same time, numerous combinations of the four parameters shown in Table 4. When a light parameter is present, each of the five statistical models yields results that are acceptable and perform significantly well. Although light alone can predict occupancy with an accuracy of over 95%, doing so would be highly inappropriate because people may have accidentally left lights on in empty rooms. As a result, there is a considerable risk of getting erroneously good findings, hence it is best avoided in practice. DSET and LDA exceed the other four models in terms of reliability, with DSET delivering greater than 99% accuracy for combinations of light and CO2 shown in Table 5.

Parameters	RF	GBM	CART	LDA	DSET
Temp	70.74	64.70	66.55	85.33	84.22
Temp+Hum	75.66	77.80	84.03	85.67	83.58
Light	95.69	97.87	97.88	97.88	97.05
CO2	78.78	87.14	87.15	86.20	86.88

Table 4: Accuracy (%) Comparison of DSET Set A

Parameters	RF	GBM	CART	LDA	DSET
Temp	70.74	64.70	66.55	85.33	84.22
Temp+Hum	75.66	77.80	84.03	85.67	83.58
Light	95.69	97.87	97.88	97.88	97.05
CO2	78.78	87.14	87.15	86.20	86.88

Table 5: Accuracy (%) Comparison of DSET Set B

GBM works better than RF in six cases of the eight cases. There is a slight increase in the presence of light. In comparison to light and humidity, one possible explanation is the high correlation of the two measures. Comparing the difference in results of DSET and LDA in test A, the difference is only 0.01%, whereas in the test B the difference is only 0.55%. It is found out that RF is the worst performing as compared to the other performances and then GBM comes closely behind in terms of accuracy.

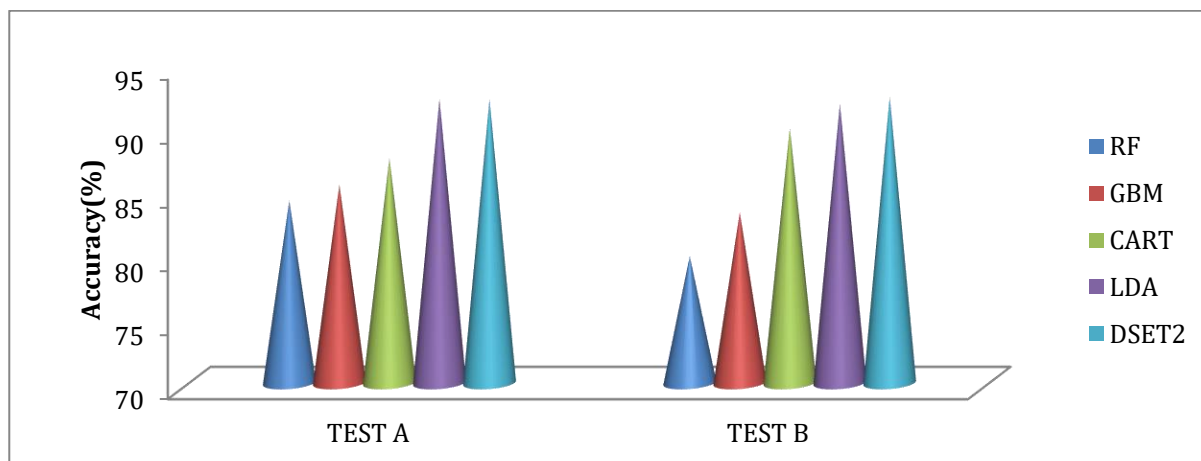


Figure 6: Comparison of average accuracy.

We obtained the values of the FN, TP, FP, and TN of the confusion matrix in our experiment. It identified the other measurement metrics such as specificity, sensitivity, precision, FP rate and accuracy. In a bid to benchmark the results with that of LDA, we averaged results as per the number of fusion parameters. In the case of testing sets, A and B, the best average prediction of LDA was compared to that of DSET and the findings are presented in Figure 6. The Table 6 has shown that DSET true positives are more than LDA. In contrast to LDA, it is possible that DSET could provide credible results during the presence of the room, which is a suggestion. The FPs are also fewer in DSET, which means that it will not pass less often in case the room remains empty. It also has a disadvantage even though the real negatives are lower than LDA and the false negatives are higher than LDA. It was however found out that DSET was more precise and specific by 1.3 and 2.7 percent compared to that of LDA. Both methods are almost equal in error rates of 92.24 and 92.23, respectively.

Criteria	Test A		Test B	
	DSET	LDA	DSET	LDA
True Positive	867	842	1694	1662
False Negative	102	76	387	407
True Negative	1593	1619	7318	7298
False Positive	106	133	355	389
Sensitivity	89.57	91.82	81.46	80.37
Specificity	93.78	92.47	95.40	94.96
False Positive Rate	6.25	7.55	4.62	5.07
Precision	89.13	86.43	82.73	81.07
Accuracy	92.26	92.24	92.43	91.87

Table 6: DSET and LDA Performance Evaluation

The ability of a classifier to identify the United States of America when a room is empty well is known as occupancy specificity and precision, the percentage of times the room was correctly labeled as occupied, as compared to LDA, is a strength of DSET in both cases, when the room is not empty and when it is. The optimal efficiency of a DSET compared to that of LDA showed several advantages to DSET over LDA: The TNs and TPs are better, and the FNs and FPs are significantly lower. Owing to this fact, DSET outperformed LDA in all areas of performance, such as specificity, sensitivity, precision, FPR, and accuracy. The overall accuracy of the LDA prediction (91.86) is less than the DSET prediction (92.41) although, the difference between the performance of the two fusion parameters density across both the DSET and LDA predictions is not just in performance but in accuracy too. It is interesting to note that, although FPR declines with the increase in the number of parameters, all the other performance measures, including specificity, sensitivity, accuracy, and precision, show improvement in the two sets of tests. The comparison of the performance between the presented framework and the current vacant space identification systems proves that the former is much better in all the evaluation metrics presented in Figure 7. The least accuracy and recall are found with traditional rule-based occupancy detection because of its poor capability to adjust to dynamic human behavior and sensor uncertainty. Machine learning-based models like the Support Vector Machine (SVM) and Random Forest models have moderate performance which improves as it learns patterns on sensor data but they cannot perform well because of the handcrafted features and the limited time modeling. The LSTM-based deep learning system also improves recall and F1-score due to the ability to consider temporal relationships, but is also sensitive to sensor noise and more complex to compute.

Conversely, the suggested framework is the most accurate, precise, recalls and F1-score because it combines the multi-sensory data in an efficient way, efficient preprocessing, and smart learning methods. The enhanced accuracy means that the false vacancy is reduced and the enhanced recall is the capability of the framework to properly determine occupied and vacant states in various circumstances. The moderate enhancement of F1-score shows the quality and the credibility of the offered strategy. In general, these findings confirm that the presented framework can be successfully used to implement smart building applications because it is more effective to detect vacant space accurately and consistently than the current system

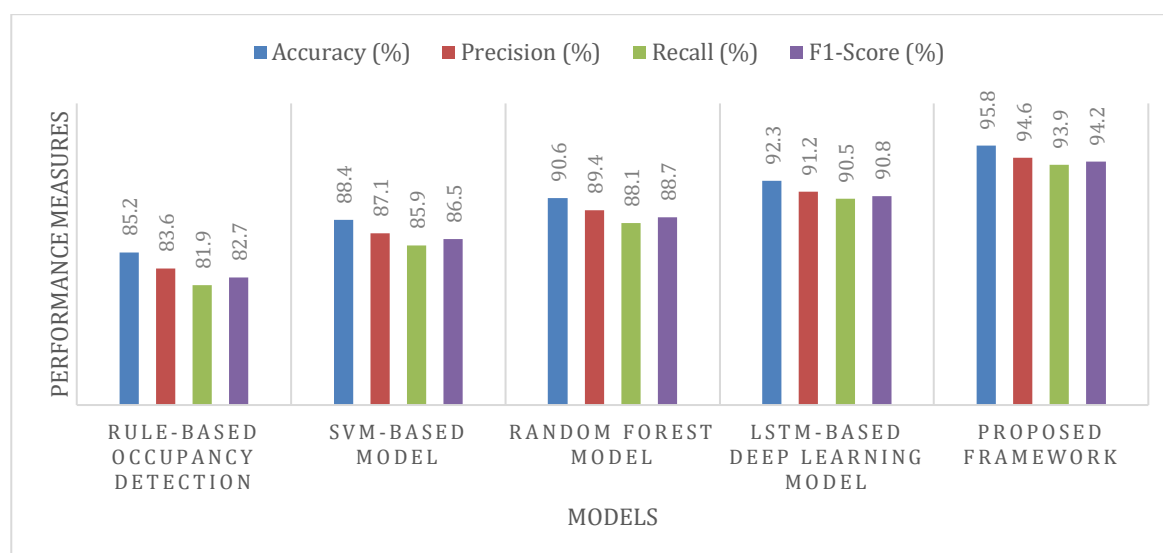


Figure 7: Comparison of performance measures

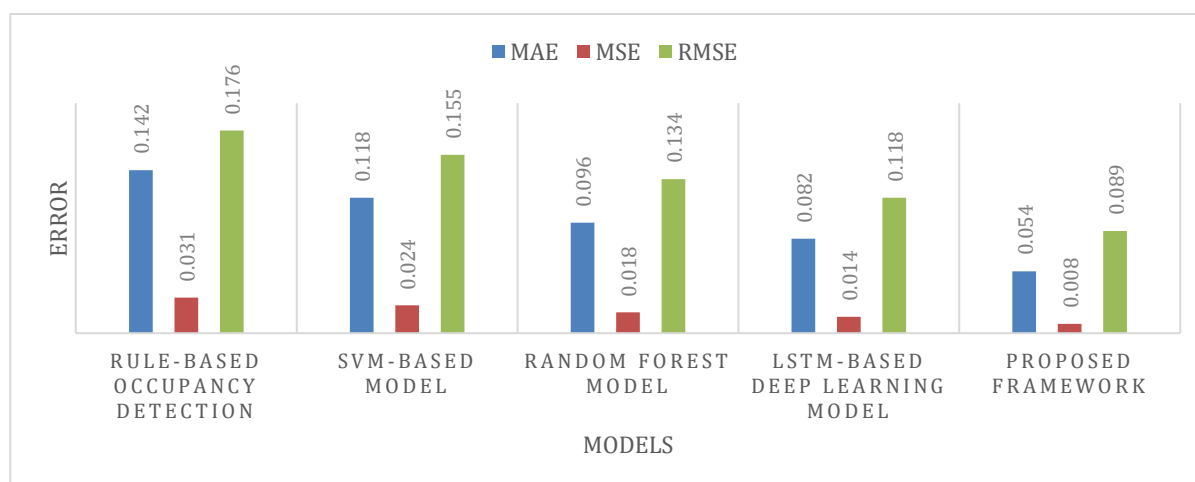


Figure 8: Comparison of performance measures (Error)

The measure of error comparison also reflects the superiority of the proposed framework to the existing systems of identifying vacant spaces as indicated in Figure 8. The error measurement comparison also shows that the proposed framework is superior to the currently existing systems of vacant space identification. The rule-based model has the highest values of MAE, MSE and RMSE, indicating that it has a low ability to predict sensor noise and dynamic occupancy patterns. Machine learning methods including the SVM-based and the Random Forest-based models exhibit lower error rate because they learn relationships, which are based on data, but their performance is limited by the fixed feature patterns and the sensitivity to environmental changes. The LSTM-based model has lower error values as it is able to model the dependent time more effectively, although, the LSTM-based model has higher complexity and is often misclassified because of noises or incomplete sensor data. On the contrary, the proposed structure has the best MAE, MSE, and RMSE, and thus more precise and consistent predictions. The fact that the error values have been considerably decreased shows that the framework is capable of incorporating the multi-sensor information, powerful preprocessing, and smart decision mechanisms. The decreased MAE ensures better average prediction accuracy, whereas smaller MSE and RMSE indicate that stability and reliability can be better characterized in vacancy detection. Such outcomes confirm the power of the

offered structure in smart building applications; it is adequately suitable to apply to the projects in real-time and at scale. The rule-based model has the highest values of MAE, MSE and RMSE, indicating that it has a low ability to predict sensor noise and dynamic occupancy patterns. Machine learning methods including the SVM-based and the Random Forest-based models exhibit lower error rate because they learn relationships, which are based on data, but their performance is limited by the fixed feature patterns and the sensitivity to environmental changes. The LSTM-based model has lower error values as it is able to model the dependent time more effectively, although the LSTM-based model has higher complexity and is often misclassified because of noises or incomplete sensor data. On the contrary, the proposed structure has the best MAE, MSE, and RMSE, and thus more precise and consistent predictions. The dramatic decrease in the error values is evidence of the framework's capability to successfully combine multi-sensor information, powerful pre-processing and intelligent decision processes. The decreased MAE ensures better average prediction accuracy, whereas smaller MSE and RMSE indicate that stability and reliability can be better characterized in vacancy detection. Such outcomes confirm the power of the offered structure in smart building applications; it is adequately suitable to apply to the projects in real-time and at scale.

5. Conclusion

A procedure that we have proposed on how to compute a mass assignment takes into account the probability density functions of the corresponding sensor reading. This paper is based on the capacity to evidence sensor reading using PMAs. By experimental simulation with MATLAB, we could obtain a very high accuracy. Since the performance of a light sensor as a single parameter is high, it is also seen that when one is present the accuracy is greatly enhanced. Also, they compare it to the existing mathematical models and present results that validate that our strategy yields much better results than three of the four models and the fourth one as well. Moreover, although DSET and LDA are equally accurate; the DSET precision and specificity is more accurate. Occupancy in real-time in any home or office premises can be easily identified using our model. A small number of training samples would first have to be entered into the model to train it and then be capable of operating correctly on any number of test cases. This requirement of a classification model is that it must be retrained to be capable of adapting to its new surroundings in case there is a change in location.

Conflict of Interest Statement

There is no conflict of interest

Authors' contributions

A: Methodology, Writing- Original draft preparation.

B: Visualization, Investigation, Supervision, Reviewing and Editing.

Funding

No funding

References

- [1]. Waterworth, David, Sethuvenkatraman, Sundararajan, and Sheng, Quan Z. "Advancing Smart Building Readiness: Automated Metadata Extraction Using Neural Language Processing Methods." *Advances in Applied Energy*, vol. 3, 2021, p. 100041, <https://doi.org/10.1016/j.adapen.2021.100041>.
- [2]. Alharbi, Abdulrahman, et al. "Web-Based Framework for Smart Parking System." *International Journal of Information Technology*, vol. 13, no. 4, 2021, pp. 1495-1502, <https://doi.org/10.1007/s41870-021-00718-w>.
- [3]. Monica, M., et al. "PMSG Based WECS: Control Techniques, MPPT Methods and Control Strategies for Standalone Battery Integrated System." *AIP Conference Proceedings*, vol. 2405, no. 1, 2022, p. 040013, <https://doi.org/10.1063/5.0076315>.
- [4]. Saha, Debashis, et al. "Application of Nanotechnology in Neural Growth Support System." 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon), 2022, pp. 1-6, <https://doi.org/10.1109/MysuruCon55533.2022.9972437>.
- [5]. Raviprasad, B., et al. "Accuracy Determination Using Deep Learning Technique in Cloud-Based IoT Sensor Environment." *Measurement: Sensors*, vol. 24, 2022, p. 100459, <https://doi.org/10.1016/j.measen.2022.100459>.
- [6]. Chelliah, B. James, et al. "Analysis of Demand Forecasting of Agriculture Using Machine Learning Algorithm." *Environment, Development and Sustainability*, 2022, pp. 1-17, <https://doi.org/10.1007/s10668-022-02235-6>.
- [7]. Lei, Yihan, et al. "A Practical Deep Reinforcement Learning Framework for Multivariate Occupant-Centric Control in Buildings." *Applied Energy*, vol. 324, 2022, p. 119742, <https://doi.org/10.1016/j.apenergy.2022.119742>.
- [8]. Sun, Xiang, and Shu, Kai. "Application Research of Perception Data Fusion System of Agricultural Product Supply Chain Based on the Internet of Things." *EURASIP Journal on Wireless Communications and Networking*, vol. 2021, no. 1, 2021, pp. 1-18, <https://doi.org/10.1186/s13638-021-02006-9>.

- [9]. Ma, Chunyan. "Smart City and Cyber-Security: Technologies Used, Leading Challenges and Future Recommendations." *Energy Reports*, vol. 7, 2021, pp. 7999-8012, <https://doi.org/10.1016/j.egy.2021.08.129>.
- [10]. Devagiri, V. M., et al. "Multi-View Data Analysis Techniques for Monitoring Smart Building Systems." *Sensors*, vol. 21, no. 20, 2021, p. 6775, <https://doi.org/10.3390/s21206775>.
- [11]. Malini, P., et al. "Internet of Everything (IoE) in Smart City Paradigm Using Advanced Sensors for Handheld Devices and Equipment." *IoT and IoE Driven Smart Cities*, Springer, Cham, 2022, pp. 121-141, https://doi.org/10.1007/978-3-030-93046-8_7.
- [12]. Zhou, Qian, et al. "Enabling Non-Intrusive Occupant Activity Modeling Using WiFi Signals and a Generative Adversarial Network." *Energy and Buildings*, vol. 249, 2021, p. 111228, <https://doi.org/10.1016/j.enbuild.2021.111228>.
- [13]. Garikapati, Phaneendra, et al. "K-Means Partitioning Approach to Predict the Error Observations in Small Datasets." *International Journal of Computer Aided Engineering and Technology*, vol. 17, no. 4, 2022, pp. 412-430, <https://doi.org/10.1504/IJCAET.2022.122728>.
- [14]. Eras, Luis, et al. "Viability Characterization of a Proof-of-Concept Bluetooth Meshes Smart Building Application." *International Journal of Distributed Sensor Networks*, vol. 18, no. 5, 2022, p. 15501329221097819, <https://doi.org/10.1177/15501329221097819>.
- [15]. Schiera, Daniele S., et al. "A Distributed Multimodel Platform to Cosimulate Multienergy Systems in Smart Buildings." *IEEE Transactions on Industry Applications*, vol. 57, no. 5, 2021, pp. 4428-4440, <https://doi.org/10.1109/TIA.2021.3087460>.
- [16]. Çetin, Sibel, et al. "Circular Digital Built Environment: An Emerging Framework." *Sustainability*, vol. 13, no. 11, 2021, p. 6348, <https://doi.org/10.3390/su13116348>.
- [17]. Tan, Swee Yi, et al. "Multimodal Sensor Fusion Framework for Residential Building Occupancy Detection." *Energy and Buildings*, vol. 258, 2022, p. 111828, <https://doi.org/10.1016/j.enbuild.2021.111828>.
- [18]. Huang, Weiming, et al. "Research on Urban Modern Architectural Art Based on Artificial Intelligence and GIS Image Recognition System." *Arabian Journal of Geosciences*, vol. 14, no. 10, 2021, pp. 1-13, <https://doi.org/10.1007/s12517-021-07220-2>.
- [19]. Zrelli, Ali. "Hardware, Software Platforms, Operating Systems, and Routing Protocols for the Internet of Things Applications." *Wireless Personal Communications*, vol. 122, no. 4, 2022, pp. 3889-3912, <https://doi.org/10.1007/s11277-021-09166-3>.
- [20]. Sayed, Ali N., et al. "Deep and Transfer Learning for Building Occupancy Detection: A Review and Comparative Analysis." *Engineering Applications of Artificial Intelligence*, vol. 115, 2022, p. 105254, <https://doi.org/10.1016/j.engappai.2022.105254>.
- [21]. Bisicchia, Giada, et al. "A Declarative Goal-Oriented Framework for Smart Environments with LPaaS." *arXiv*, 2021, <https://doi.org/10.48550/arXiv.2106.13083>.
- [22]. Abdel-Razek, Samar A., et al. "Energy Efficiency Through the Implementation of an AI Model to Predict Room Occupancy Based on Thermal Comfort Parameters." *Sustainability*, vol. 14, no. 13, 2022, p. 7734, <https://doi.org/10.3390/su14137734>.
- [23]. Llaría, Ainhoa, et al. "Intelligent Buildings in Smart Grids: A Survey on Security and Privacy Issues Related to Energy Management." *Energies*, vol. 14, no. 9, 2021, p. 2733, <https://doi.org/10.3390/en14092733>.