Building and Environment 171 (2020) 106681

Contents lists available at ScienceDirect

Building and Environment

journal homepage: http://www.elsevier.com/locate/buildenv

A scalable Bluetooth Low Energy approach to identify occupancy patterns and profiles in office spaces

Zeynep Duygu Tekler [a,*], Raymond Low [a], Burak Gunay [b], Rune Korsholm Andersen [c], Lucienne Blessing [a]

[a] Singapore University of Technology and Design, 8 Somapah Road, 487372, Singapore
[b] Department of Civil and Environmental Engineering, Carleton University, 1125 Colonel by Drive, Ottawa, ON, K1S 5B6, Canada
[c] Department of Civil Engineering, Technical University of Denmark, Nils Koppels Allé, 2800, Kgs. Lyngby, Denmark

ARTICLE INFO

Keywords:
Occupancy detection
Bluetooth low energy
Machine learning
K-means clustering
Office spaces
Occupancy profiles

ABSTRACT

Building occupants are often assumed to follow deterministic schedules in building performance simulation programs. Therefore, to accurately capture the dynamic nature of the occupants' movement patterns, researchers have proposed various indoor localisation technologies to infer occupancy information with varying degrees of accuracy and resolution. Among these technologies, the Bluetooth Low Energy (BLE) technology emerged as a popular alternative due to its availability in smartphone devices, as well as its low cost and power demand. In this study, we proposed a scalable and less intrusive occupancy detection method that leverages existing BLE technologies found in smartphone devices to perform zone-level occupancy localisation, without the need for a mobile application. The proposed method uses a network of BLE beacons for data collection before passing the pre-processed data into a machine learning model to infer the occupants' zone-level location. A supervised ensemble model and a semi-supervised clustering model were proposed and evaluated to identify the best performing model. The feasibility of the proposed method is demonstrated during a five-week case study involving two office spaces in an academic building in Singapore. While the supervised ensemble model produced the best performance in terms of accuracy and macro-average f1-score, the semi-supervised model was able to produce a reasonable performance while using a fraction of the training data (<4%) and time needed by the supervised model. By analysing the occupancy information obtained through the best performing model, we further identified a set of occupancy profiles to represent the diverse occupancy patterns observed in the study area.

1. Introduction

1.1. Background

Occupant behaviour (OB) can be defined to encompass both occupancy and the occupants' interaction with different building systems such as HVAC systems, lighting systems, windows, blinds, shades, and plug load appliances [1]. While many researchers have agreed on the influence of OB on the building's energy performance, the occupants' dynamic behaviours are still inadequately represented in most building performance simulations (BPS); where they are often assumed to exhibit deterministic behaviours and follow regular schedules. In reality, occupants exhibit stochastic and diverse behaviours that contribute to a wide discrepancy between the simulated and actual building energy consumption. This discrepancy is also known as the energy performance gap. Therefore, researchers must first gain a better understanding of OB so that it can be properly represented and integrated into existing energy simulations in the form of OB-centric models. These models would serve as a useful decision-making tool for building planners and engineers when designing high-performance, low-energy office buildings.

1.2. Occupancy in the built environment

Given the definition provided above, OB can be more appropriately defined to be dependent on occupancy as the occupants' interaction with different building components is first conditioned on their presence and movement patterns within the building. Thus, the first step towards understanding OB naturally involves the collection of detailed occupancy information. This information is not only valuable in the development of OB-driven models but also serves as a reliable indicator of

* Corresponding author.
E-mail address: duygutekler_zeynep@mymail.sutd.edu.sg (Z.D. Tekler).

https://doi.org/10.1016/j.buildenv.2020.106681
Received 8 November 2019; Received in revised form 30 December 2019; Accepted 16 January 2020
Available online 18 January 2020
0360-1323/

future demand which can be used in building control systems to improve energy efficiency and performance of commercial buildings [2]. Multiple studies reported substantial energy savings when implementing such occupancy-driven control systems to desktops, HVAC, and lighting systems [3–5]. Given these potential use cases, researchers have continually proposed increasingly advanced occupancy detection systems that are aimed at obtaining high-resolution occupancy information.

The resolution of the occupancy information can be categorised based on its spatial (i.e. building, floor, room), temporal (i.e. hours, minutes, seconds), and occupancy (i.e. presence, count, identity, and activities conducted) aspects [2,6]. Unsurprisingly, the complexity involved in data collection increases with the resolution of the data. As a result, these challenges have led to an alarming number of commercial buildings relying on simplistic occupancy schedules and profiles during operation, resulting in energy wastage [7].

While many occupancy studies have explored a wide range of sensing technologies, each technology comes with its own set of limitations including accuracy, implementation cost, inability for multiple occupant detection, privacy concerns, scalability, social acceptability, and time latency [8]. A detailed literature review of the various occupancy detection technologies has been provided in Section 2. Among the different technologies proposed, the BLE technology has recently emerged as a popular alternative due to its availability in most smartphone devices, as well as having a low cost, power demand, and data latency [9]. However, as recent studies continue to propose increasingly advanced occupancy detection systems that allow the collection of more fine-grained occupancy information, usually through the installation of a mobile application, this improvement often comes with an increased level of intrusiveness and power demand on the occupants' devices.

1.3. Study objective and contribution

Given these trends, the main objective of this study is to propose a scalable and less intrusive occupancy detection method that leverages existing BLE technologies found in smartphone devices to perform zone-level occupant localisation, without the need for a mobile application. The proposed methodology involves the deployment of BLE beacons to collect the smartphone devices' received signal strength indicator (RSSI) values before passing the pre-processed data into a machine learning model to infer the occupants' zone-level location. Two different machine learning models were also proposed, based on a supervised ensemble approach as well as a semi-supervised clustering approach, and evaluated against the ground truth to identify the best performing model. The feasibility of the proposed method is demonstrated during a five-week case study involving two office spaces in an academic building in Singapore. By using the best performing model to infer the occupancy information based on the RSSI values recorded during the study period, we further showed the value of the proposed method by conducting different types of analysis, which includes the identification of a set of occupancy profiles. These profiles were used to generalise the diverse occupancy patterns observed in the study areas.

2. Related work

2.1. Past occupancy detection systems

Occupancy detection systems can be classified based on a terminal or non-terminal-based method [10]. In terminal-based methods, occupants are required to carry a wearable sensor or a smartphone device to track their movement patterns within the built environment. Some examples of such methods include the use of wireless technologies such as Radio Frequency Identification (RFID) [11,12], WIFI [13–17], and BLE technology [18–21]. While the implementation of occupancy detection systems based on these wireless technologies have allowed researchers to track the location of individual occupants at relatively high accuracies, the same implementation might intrude upon the occupants' routine and impact their natural behaviours. This phenomenon is known as the Hawthorne effect as it describes a change in the occupants' behaviour towards social acceptability due to their awareness of being studied [22].

On the other hand, non-terminal-based methods involve passively monitoring a particular area or space for the occupant's presence without requiring the occupant to carry around or wear any sensing devices. Some examples of occupancy detection systems that adopted such methods include CO_2-based detection systems [23–25], Passive Infrared (PIR) detection systems [26–29], Ultrasonic detection systems [26,30], sound detection systems [31], cameras [32,33], and smart power meters [34,35]. While some of these methods are less intrusive, the level of resolution and accuracy of the occupancy information obtained are very limited. Some studies have attempted to address these limitations by combining non-terminal-based methods with terminal-based methods, resulting in a rise in infrastructure cost and data processing effort needed for sensor fusion.

Table 1 provides a summary of the different methods used in past occupancy detection systems, categorised based on their terminal and non-terminal nature, as well as a brief description of their various benefits and limitations.

2.2. Occupancy detection systems based on Bluetooth Low Energy technology

Given the benefits and limitations that come with each method as summarised in Table 1, Ahmad et al. [8] concluded that the WIFI and BLE methods provided an optimal ratio between the cost of implementation and the quality of data obtained. While the WIFI method has been shown to be successful in detecting occupancy at the building or zone level [36], a higher resolution of occupancy information would require a significant amount of calibration to reduce the false-positive rates caused by a large number of WIFI-enabled devices that can be found in a densely populated area, such as an office space. Therefore, BLE technology recently emerged as a popular choice among researchers due to its ability to obtain high-resolution occupancy information and is widely embedded in most smartphone devices. Conte et al. [37] proposed an accurate and power-efficient occupancy detection system, named BLUE-SENTINEL, which uses a mobile application that modifies the iBeacon protocol in iOS devices to collect the RSSI values of the nearest iBeacon at regular intervals. These RSSI values were passed through the K-nearest Neighbour and Decision Tree classification algorithms to infer the occupants' room-level location. The use of unique combinations of RSSI values to perform indoor localisation is commonly known as the Bluetooth fingerprinting approach.

This work is followed up by an Android version of the application [18] where the localisation performance of the system was improved by passing the RSSI values of multiple neighbouring iBeacons into a Support Vector Machine. The authors also attempted to reduce the energy consumption of the mobile application by sending the RSSI values to a nearby BLE beacon via Bluetooth connection, which retransmits the information to an online server for further processing. Other than the use of mobile applications, Shen and Newsham [39] developed a computer program to detect the occupant's presence by scanning for his Bluetooth-enabled smartphone device when he is within range of his office computer.

Occupancy detection systems have also been proposed in the context of emergency management where first responders will be able to utilise such systems to quickly pinpoint the location of victims in the building [40]. The proposed localisation system uses a combination of BLE beacons, a mobile application, and a server, to collect the RSSI values for location inference using various machine learning approaches. Another application area is in the field of energy management where researchers [41,42] proposed a system of BLE beacons, smart energy plugs, and a mobile application to control the operation of certain appliances based on the occupant's presence within the room. Some recent studies have

Table 1
Summary of past occupancy detection methods, categorised based on their terminal and non-terminal nature, a brief description of their benefits and limitations [8,38], as well as their highest spatial and occupant (Identity > Occupant Count > Binary Presence) resolution.

Terminal-based Methods			
	Benefits	Limitations	Highest Resolution
RFID	• Cost effective • Commercially available • High detection accuracy and precision	• Privacy concerns • Possible behaviour change due to the Hawthorne effect • Sensitive to external radio frequency devices and environmental conditions • Additional deployment cost • Occupants have to carry the sensor with them all the time	Identity at zone level
WIFI	• Power-efficient • Utilise existing communication infrastructure within building • High detection accuracy and precision	• Inaccurate estimation of occupancy level due to ownership of multiple WIFI-enabled devices • Existence of different locations with same WIFI signatures • High false-positive detection rates in densely-populated areas • Occupants have to carry the sensor with them all the time	Identity at zone level
BLE	• Power-efficient • Cost effective • Commercially available • Widely available in many modern smartphone devices • High detection accuracy and precision	• Privacy concerns • Possible behaviour change due to the Hawthorne effect • Occupants forgetting to turn on their smartphone's Bluetooth mode • Additional deployment cost • Occupants have to carry the sensor with them all the time	Identity at zone level
Non-Terminal Methods			
	Benefits	Limitations	Highest Resolution
CO_2-based detection systems	• Part of the existing building infrastructure	• Slow response times due to gradual gas dispersion • Sensitive to environmental and building conditions (e.g. wind speed, pressure difference, airflow rate, interzonal air transfer, envelope airtightness, and CO_2 concentration of air supply) • Unable to obtain spatial distribution of occupancy at the zone level • Relies on assumptions regarding occupants' CO_2 emissions • Need to be combined with other environmental sensors to improve accuracy	Occupant count at zone level
PIR detection systems	• Ease of implementation • Cost effective • Power-efficient	• Low reliability as it requires a clear line of sight and continuous motion for occupancy detection • Study area layout and sensor positioning also affects reliability • Only provides binary status of occupancy • Can be triggered by heat currents from HVAC systems • Need to be combined with other environmental sensors to improve accuracy	Binary presence at zone level
Ultrasonic detection systems	• Ease of implementation • Cost effective • Do not require line of sight or continuous movement	• Accuracy decreases for large spaces due to cross-talk between transducers • Slow response time • Prone to false presence detection due to environment (i.e. air turbulence from HVAC systems) and motion of inanimate objects • Only provides binary status of occupancy	Binary presence at zone level
Sound detection systems	• Provide occupancy information on location and presence	• Prone to false negatives when the occupant is not generating any sound • Prone to false positives due to sounds from non-human sources • Need to be combined with other environmental sensors to improve accuracy	Binary presence at zone level
Image detection systems (Cameras)	• High resolution data (occupant count, location, and movement) • Readily available in security networks	• Privacy concerns • Requires line of sight with minimal obstructions • Advanced signal processing and expensive hardware required • Data from security networks are often stored in standalone databases and rarely integrated into HVAC and lighting controls	Occupant count at zone level
Smart power meters	• Commercially available • Low privacy intrusion	• Only provides binary status of occupancy • Unable to detect occupants if they are not using an appliance • Need to be combined with other environmental sensors to improve accuracy	Binary presence at zone level
Terminal + Non-terminal Methods			
	Benefits	Limitations	Highest Resolution
Sensor Fusion	• Improves overall performance of occupancy estimation • Compensate for the limitations of each sensor	• Challenges in processing the sensor data efficiently in real-time, leading to slow response • Often require additional infrastructure • Heterogeneous nature of data requires advanced data processing	Identity at zone level

also proposed the integration of BLE-based occupancy detection systems with HVAC systems to enable occupancy-driven system controls based on high-resolution occupancy distribution information [43,44]. Finally, Tekler et al. [45] have proposed a sensor fusion approach to monitor the occupancy and plug load interaction patterns of occupants in an office environment by combining the use of BLE beacons and smart energy meters.

However, for several of the BLE occupancy detection systems highlighted above, occupants are either required to install a mobile application on their smartphone devices or download a custom program onto their personal plug loads in order to obtain individual-specific occupancy information. This requirement negatively impacts the occupants' natural behaviour due to the Hawthorne effect. Furthermore, the increased power burden on the occupant's smartphone device due to the mobile application also reduces the likelihood of widespread adoption. Due to these limitations, past occupancy studies are often limited in their participation rates, thereby reducing the generality and validity of the data collected.

In an attempt to reduce the intrusiveness of such studies, recent papers have started exploring occupancy detection systems that adopted less intrusive methods. Park et al. [20] proposed a non-invasive Bluetooth-based occupancy detection system that detects the number of discoverable Bluetooth devices in an area before mapping them to the room's occupancy level using a carefully calibrated ratio While the method does not require the occupants to install any specific software on their devices, recalibration of the ratio must be performed between different buildings and over time due to the changing ownership patterns of Bluetooth devices. This limitation reduces the scalability of the method. The method proposed by Tekler et al. [21] leverages the existing BLE technologies found in the occupants' smartphone devices to track the occupants' movement patterns using BLE beacons. Unlike many existing methods, the proposed method does not require the installation of a mobile application but only requires the occupants to provide their devices' Bluetooth media access control (MAC) address. This step is necessary as the Bluetooth MAC address is used as a unique identifier to track the occupant's movement patterns and cannot be sniffed unless the device is on "discoverable" mode. An additional consideration when collecting the occupant's Bluetooth MAC address is to obtain the true MAC address directly from the occupant's smartphone device, under the "Phone Settings" page. This step is necessary as the device's advertised address might differ from the true address due to the MAC randomisation features implemented by the phone manufacturers as a security measure. By using the BLE beacons to scan the vicinity for the occupants' Bluetooth-enabled devices, the RSSI values obtained are passed through a gradient boosting algorithm to infer the occupants' location, thereby allowing the authors to not only monitor the occupancy levels in each zone but also track the movement patterns of individual occupants between different zones.

Given that many of the occupancy detection methods that leverage on RSSI values to perform occupant localisation requires the system to be trained based on a supervised approach, the scalability of the method is limited as a large amount of labelled data needs to be collected for the training process every time the detection system is implemented in a new study area. Therefore, by proposing a semi-supervised approach when developing the occupancy detection method, we can improve upon the method proposed by Tekler et al. [21] by reducing the amount of labelled data needed, thereby increasing its scalability.

3. Methodology

3.1. Data collection

The implementation of the proposed occupancy detection method takes advantage of the existing BLE technology found in modern smartphone devices, combined with the usage of BLE beacons, to perform indoor localisation without the need for a mobile application. The BLE beacons, used in this study, are built using Raspberry Pi 3 Model B's which are relatively inexpensive and widely available. By programming the BLE beacons to scan the vicinity for the occupants' Bluetooth-enabled smartphone devices based on their MAC addresses, the presence of any neighbouring devices was recorded by storing the device's Bluetooth MAC address, RSSI value, as well as the current timestamp, on the BLE beacon. The recorded information can also be sent wirelessly via a WIFI connection to a centralised server to be stored for further processing.

The RSSI value can be interpreted as a unit of measurement that represents the relative quality of a received signal from a device. While a single RSSI value may not provide reliable estimations of the distance between two devices, the combination of multiple RSSI values can provide a reasonable indicator, especially when combined with advanced machine learning approaches.

3.2. Data pre-processing

Once the information had been recorded and stored on the BLE beacon, it was extracted and consolidated with the information from other BLE beacons deployed in the study area. Using the MAC address as a unique identifier, the RSSI values corresponding to a particular occupant were filtered out and sorted based on the timestamp information. Each entry, which contains a single RSSI value, was converted into an RSSI tuple $(x_1, x_2, x_3, \ldots, x_n)$ where each x_i, $i = 1, \ldots, n$, corresponds to the RSSI value recorded by a particular BLE beacon deployed in the study area for a particular occupant and unique timestamp. Since the scanning patterns of the BLE beacons were not synchronised, the formation of the RSSI tuple resulted in the presence of missing values which were imputed based on the last RSSI value recorded by each corresponding BLE beacon. However, if the last recorded RSSI value occurs beyond a user-defined time window, the missing values were instead imputed by an arbitrarily large RSSI value. This imputation strategy is based on our observations that occupants tend to remain stationary for the majority of the time while they were in the office. Therefore, an RSSI value received by a particular BLE beacon should remain valid for a short time after it has been recorded. An opposite argument holds when a particular BLE beacon is unable to detect an occupant's device for an extended period. This event is a strong indication that the occupant has left the scanning region of the BLE beacon and, therefore, is represented using a large RSSI value. While this user-defined time window can be adjusted based on the mobility patterns of the occupants, we observe that a window between 1 and 2 min is a reasonable value for the cases we studied. By performing the series of data pre-processing and imputation steps described above, the resulting data follows the dimensions $D \times M \times N$, where D represents the number of occupants monitored, M represents the number of RSSI tuples that were obtained using the BLE beacons, and N represents the number of BLE beacons deployed.

3.3. Model development

In the following subsections, we will propose a supervised and a semi-supervised machine learning model to infer the occupants' zone-level location using the pre-processed data.

3.3.1. Supervised bluetooth ensemble model

The supervised Bluetooth ensemble model follows an ensemble 1-vs-all architecture by developing a binary classifier for each zone in the study area to perform zone-level occupant localisation. By comparing the probabilistic outputs of each binary classifier, the classifier with the highest probability is identified, and its corresponding location is selected as the occupant's final predicted location. Given that there might be a considerable imbalance between the number of positive classes (i.e. occupant is present in the zone) and negative classes (i.e. occupant is not present in the zone) when training each binary classifier, additional steps were taken to overcome the class imbalance issue when developing the ensemble model. By splitting the data in the negative class into multiple batches, where the size of each batch equal to the size of the positive class, a binary classifier is trained on the resulting dataset made up of data from the positive class and data from each batch containing the negative class. The combination of these binary classifiers results in an ensemble sub-model of binary classifiers for each specific zone. Finally, by combining the ensemble sub-models from each zone, the final model was made up of multiple ensemble sub-models containing binary classifiers.

The model's training process involves the collection of a set of labelled data which contained both the RSSI tuples and the corresponding locations of the occupants when the RSSI tuples were recorded. This data was obtained by placing several Bluetooth-enabled devices at different zones within the study area to collect their corresponding RSSI tuples using BLE beacons. The data collected was processed using the steps described in subsection 3.1 and 3.2 and labelled based on the corresponding zone locations of the training devices at those timestamps. While it is common practice to randomly split the labelled data into a training set and a validation set, we observed during the study that a BLE beacon can record slightly different RSSI values for different smartphone devices even when they are placed in the same exact location. Since the model and the operating system of the device can contribute to these slight variations, the labelled data is split based on the device to result in a train-test ratio of 70:30. This approach is taken to accurately represent a real-world implementation of the proposed detection method where it is very likely that the devices used to train the Bluetooth fingerprinting model differs from the occupants' actual devices.

The classification algorithm used in this study was the gradient boosting algorithm [46] due to its superior predictive performance over other classification algorithms. The algorithm follows an iterative functional gradient descent method that minimises its loss function by iteratively introducing base learners, defined based on the errors made by the current model, to boost subsequent model performance. The different models proposed in this paper were optimised via a hyperparameter tuning and five-fold cross-validation approach. Given that the gradient boosting algorithm also allows us to assign an importance score to each input feature, the input features with the lowest importance scores were dropped from the final model to prevent model overfitting and decrease training time.

3.3.2. Semi-supervised bluetooth clustering model

The second machine learning model proposed in this study is modified based on an unsupervised clustering algorithm, guided by a small amount of labelled data, to perform zone-level occupant localisation on a set of unlabelled RSSI tuples. The clustering algorithm used in this study was the K-means algorithm due to its popularity and ease of implementation. The K-means algorithm begins by specifying the number of clusters K that can be found in the dataset. In this case, K was user-defined and based on the number of zones demarcated within the study area. By randomly selecting K data points as the initial centroids, the sum of squared distance between each data point and the centroids were calculated. In the next step, each data point was assigned to a particular cluster based on the nearest centroid and a new set of centroids were obtained by calculating the average of all of the data points that belong to the same cluster. This step was repeated until there is no change in cluster assignments for each data point.

However, given that the K-means algorithm was only able to provide a numerical label to the data points that belong to the same cluster, an additional step was needed to map each cluster label to a particular zone in the study area. This mapping step was achieved by collecting a small sample of labelled data from each zone and calculating the centroid of these labelled samples. Next, by calculating the Euclidean distance between a labelled centroid and the cluster centroids, we identified the nearest cluster centroid-labelled centroid pair and assigned the corresponding zone label to the data points in that cluster. Additional steps were also taken to ensure that each zone was uniquely mapped to a particular cluster. Another complication of this approach comes from a limitation of the K-means algorithm. As the initial K centroids were randomly selected from the dataset, different initialisation of the K centroids might result in different cluster labels for the same data point, leading to large fluctuations in model performance. This issue has been overcome by initialising the K-means algorithm with the labelled centroids described earlier to stabilise the algorithm and ensure that a similar prediction is produced every time the clustering algorithm was rerun. Furthermore, by initialising the clustering algorithm with the labelled centroids, convergence time decreases accordingly.

3.4. Model evaluation

Since we are interested in the models' ability to accurately perform zone-level localisation, the resulting models will be evaluated based on two evaluation metrics: zone-level accuracy and macro-average f1-score. The formulations of both evaluation metrics are provided below.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Macro - ave\, f1 = \frac{2Precision_{macro}Recall_{macro}}{Precision_{macro} + Recall_{macro}}$$

where

$$Precision_{macro} = \frac{\sum_{i=1}^{l} \frac{TP_i}{TP_i + FP_i}}{l}$$

$$Recall_{macro} = \frac{\sum_{i=1}^{l} \frac{TP_i}{TP_i + FN_i}}{l}$$

TP, TN, FP, and FN refer to the true positive, true negative, false positive and false negative zone-level prediction values, respectively, while refers to the number of prediction classes. The macro-average precision score can be interpreted as the fraction of true positives among the predicted positive instances in a class, averaged over all prediction classes. On the other hand, the macro-average recall score is interpreted as the fraction of true positives among all correctly predicted instances in a class, averaged over all prediction classes. While accuracy is a standard evaluation metric used by many researchers, the macro-average f1-score, which is the harmonic mean of both macro-average precision and recall scores, was also used in this study to provide a balanced evaluation of the model's performance for each class. A flowchart summarising the proposed methodology was provided in Fig. 1.

4. Case study

4.1. Study area description

The feasibility of the proposed method was demonstrated by conducting a five-week data collection effort in two office spaces, situated at an academic building in Singapore. The rationale behind conducting this study in two different offices was to ensure the generality of the method when encountering different office layouts and to capture a wider range of occupancy behaviours.

Each study area is divided into zones demarcated by a combination of physical walls and partition lines for open spaces. Study area 1 spans approximately 650 m^2 and houses a meeting area, a pantry, two open office areas, and several rooms with research equipment. The study area is divided into a total of 12 zones, as depicted in Fig. 2, with an additional zone named "Out" to indicate the absence of the occupant. Study area 2 has a smaller area of 248 m^2 and contains 8 zones, as depicted in Fig. 3, with an additional zone named "Out" to indicate the occupant's absence. The remaining zones include a pantry, a meeting room, a printer area, and three open office areas.

4.2. Study implementation

During the deployment of BLE beacons, the location of each beacon is restricted based on the availability of power sockets in the study area. In addition, several general guidelines for beacon deployment were also observed to produce good results during the study. The first guideline relates to larger-sized zones where the location and number of beacons deployed should ensure a complete signal coverage of the area. In addition, another guideline states that more beacons should be deployed in smaller-sized zones since it is harder for the model to infer the occupant's location due to the higher resolution. With the deployment of more beacons, the increased number of features can help to increase the model's accuracy, while being limited by the number of power sockets available in the area. In this case, study area 1 is made up of a mixture of larger-sized zones and smaller-sized zones. Hence the number of beacons that are deployed in the larger-sized zones are mostly to ensure signal coverage while more beacons are concentrated in the pantry area to improve the model's accuracy. Study area 2, on the other hand, is mostly made up of intermediate-sized zones with a few smaller-sized zones. Therefore, more beacons are deployed mainly to improve the model's accuracy. As a result, a total of 21 BLE beacons are deployed in study area 1, while 18 beacons are deployed in study area 2.

During the 5-week data collection period, the BLE beacons were configured to scan the vicinity of the occupants' Bluetooth-enabled smartphone device during the weekdays to obtain a total of 25 days of occupancy data between both study areas. In addition, a total of 60 occupants from both study areas have agreed to participate in this study, resulting in the collection of more than 2 million data points by the end of the study period with an average scanning frequency between 1 and 2 min for each study participant.

Two researchers were also placed in each study area to manually track the zone-level locations of each study participant over the duration of 5 working days to obtain the ground truth. The first researcher is placed at a strategic location that overlooks the study area and various exits while the second researcher is placed at a different location to cover any potential blind spots.

4.3. Descriptive statistics

Out of the 60 occupants who have agreed to participate in this study, we assigned 46 to be permanent occupants, and the remaining 14 as temporary occupants. Permanent occupants are defined as occupants who have an assigned workspace in the study area while temporary occupants are occupants who visit the study area for work-related purposes but do not have an assigned workspace. Table 2 provides a detailed breakdown of the different occupant types, arranged based on their designations and study area.

4.4. Model comparison results

Table 3 describes the overall accuracy and macro-average f1-score of both supervised and semi-supervised models for both study areas. It can be observed that the supervised ensemble model was able to outperform the semi-supervised clustering model both in terms of overall accuracy and macro-average f1-score for both study areas. The performance of the supervised ensemble model is further validated by processing its output to generate the predicted occupancy levels in both study areas and plotting it against the ground truth over five working days. It can be observed from Fig. 4 that the predicted occupancy level closely matches the true occupancy level recorded over the observation period.

As the ensemble model is made up of a combination of multiple ensemble sub-models, each of these sub-models was optimised to improve their performance via hyperparameter tuning and feature selection. While this required us to search over a much broader feature space to identify the optimal set of hyperparameters and input features to include in the final model, this also allowed us to achieve a much lower error rate. Another reasoning behind this performance could also be attributed to the removal of noise during the feature selection step. By dropping the least significant set of input features from each zone-level sub-model, this step decreases the likelihood of overfitting by reducing model complexity. Finally, it should be noted that the layout of the study area has an influence on model performance as the supervised ensemble model reported different performances in both study areas.

Another notable observation from Table 3 is the poor performance of the proposed semi-supervised approach when applied to the same 12 zones in study area 1. The prediction results were visualised using a confusion matrix in Fig. 5, where the columns represent the instances in each predicted class and the rows represent the true classes of each instance. Given that the correct predictions will be located in the diagonal of the matrix, it was observed that the clustering model commonly misclassifies zones that are adjacent to each other. This observation is particularly true for the three zones found in the pantry area as well as the area covering the office areas and the printer. Due to this observation, we improved the semi-supervised model's accuracy by merging these adjacent zones together to form a single zone for the pantry area, as well as a combined zone for the office area, to result in a total of 9 zones in study area 1.

While the resulting performance of the semi-supervised model is still lower than the supervised ensemble model (refer to Table 3), the practical advantages of the semi-supervised approach are clear as it only uses a small fraction of the labelled data and training time when compared to the supervised ensemble model. Furthermore, by experimenting with different number of labelled samples used when calculating the labelled centroids, we were able to achieve consistency in model performance when the number of labelled data reaches a value of 15 for each zone. This is equivalent to less than 4% of the total number of labelled data

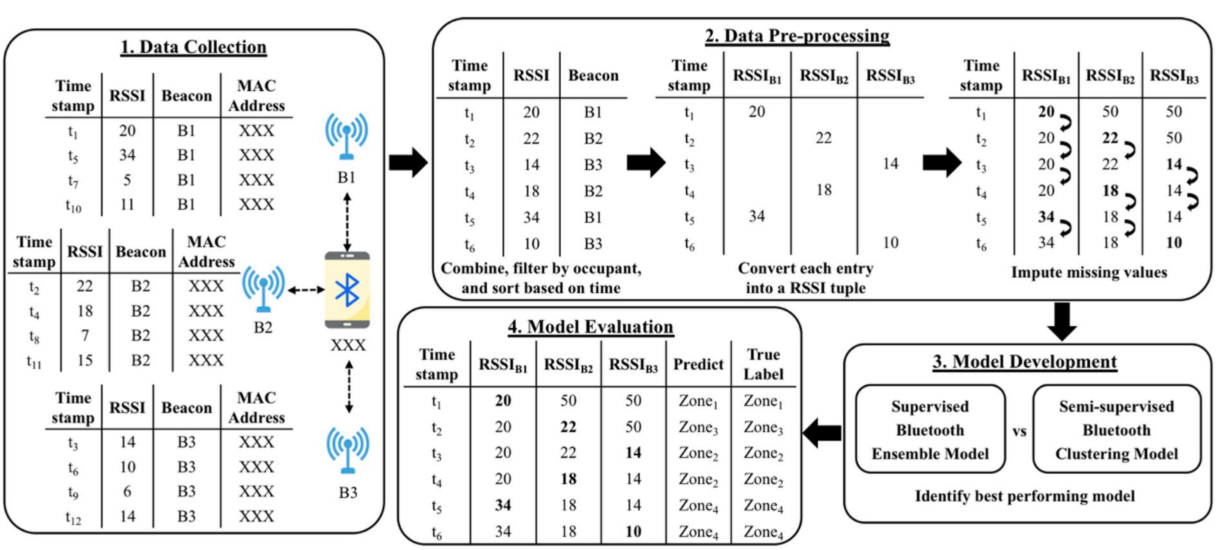

Fig. 1. A flowchart of the proposed methodology.

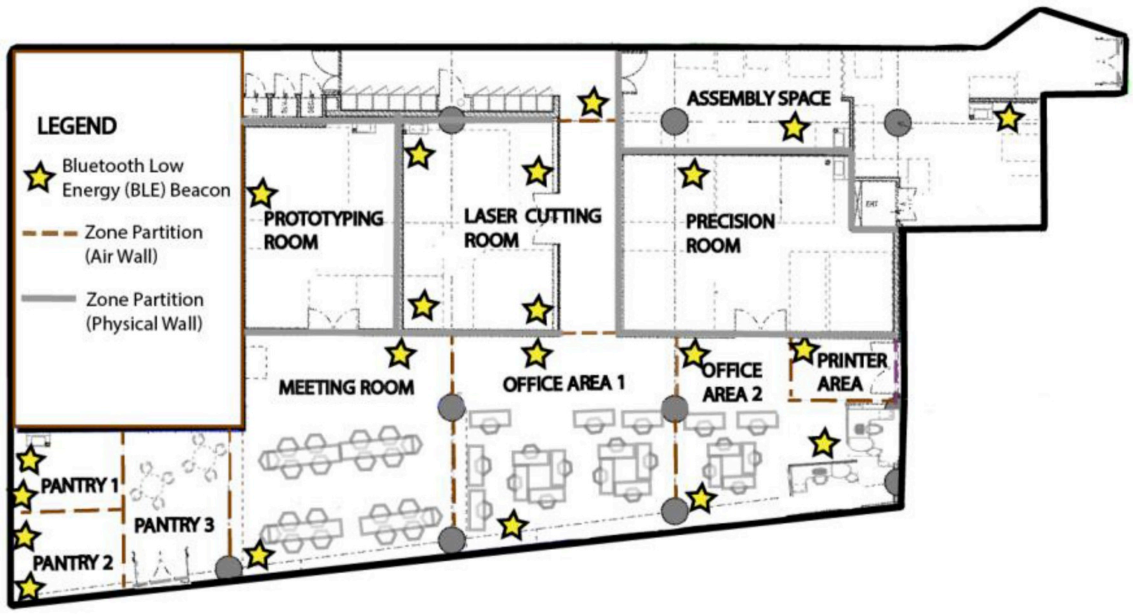

Fig. 2. Layout of study area 1 (n = 28).

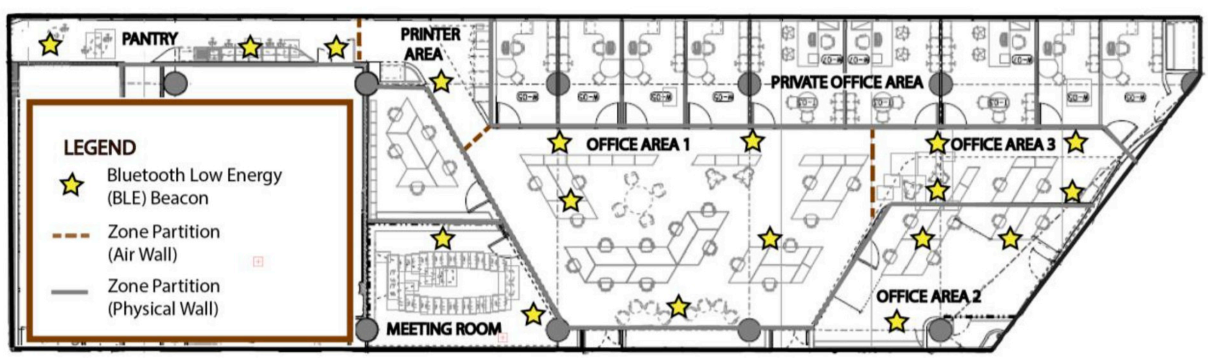

Fig. 3. Layout of study area 2 (n = 32).

Table 2
Detailed breakdown of study participants based on occupant type, designations, and study area.

	Occupant Type (n)	Designation Type (n)					
		Grad Student	Post-Doc	Design Engineer	Researcher	Staff	Faculty
Study Area1 (28)	Permanent Occupants (18)	3	1	5	7	2	0
	Temporary Occupants (10)	4	2	3	1	0	0
Study Area 2 (32)	Permanent Occupants (28)	11	4	0	2	6	5
	Temporary Occupants (4)	0	0	1	0	2	1

Table 3
Model performance for the supervised and semi-supervised models when compared to the ground truth (Study Area 1 and 2).

Model	Supervised Bluetooth Ensemble Model			Semi-supervised Bluetooth Clustering Model		
	Study Area 1 (12 Zones)	Study Area 1 (9 Zones)	Study Area 2 (8 Zones)	Study Area 1 (12 Zones)	Study Area 1 (9 Zones)	Study Area 2 (8 Zones)
Accuracy	0.81	**0.91**	**0.85**	0.60	0.79	0.82
Macro-average f1-score	0.73	**0.85**	**0.85**	0.56	0.71	0.82

used in training the supervised ensemble model.

The choice between the supervised ensemble model and the semi-supervised clustering model can be viewed as a trade-off between accuracy and scalability. Therefore, while the semi-supervised clustering model has clear practical advantages in real-world applications, we will apply the supervised ensemble model in both study areas to obtain the occupancy information needed for subsequent analysis due to its higher accuracy. The purpose of performing the subsequent analysis using the inferred occupancy information instead of the ground truth is to demonstrate the value of the proposed method by showing the different types of analysis that could be conducted based on the data collected.

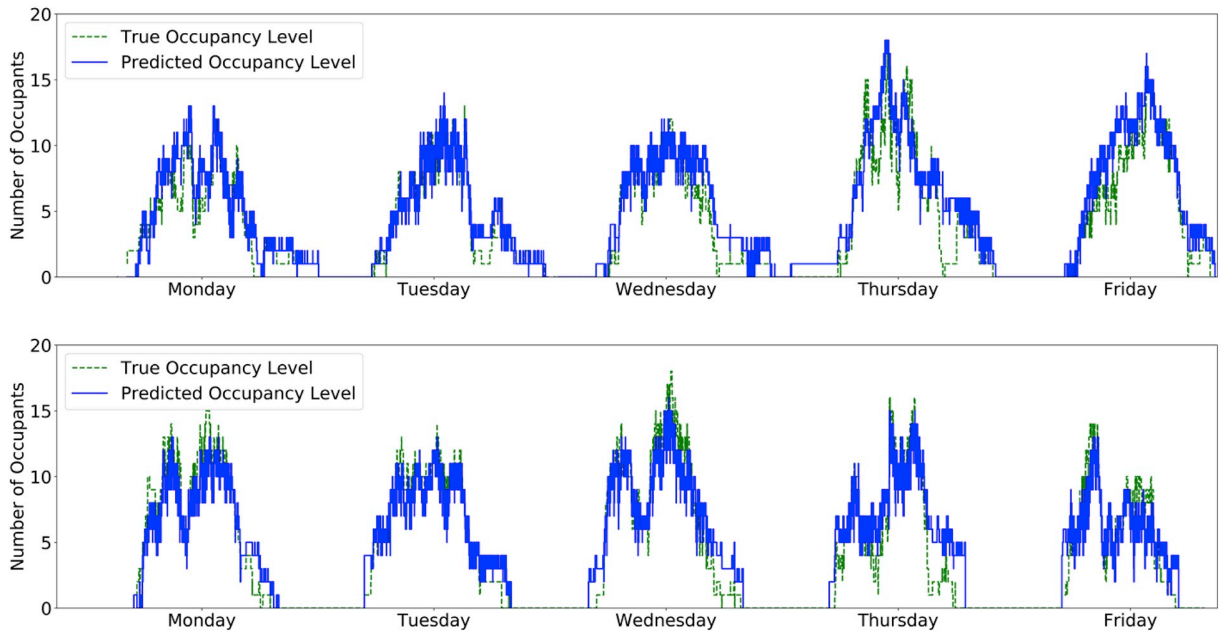

Fig. 4. The predicted occupancy levels in study area 1 (top) and study area 2 (bottom), plotted against the ground truth over five working days. The supervised ensemble model was used to infer the occupancy levels.

4.5. Occupancy patterns analysis

Based on the occupancy information obtained at the end of the study period, the occupancy levels for the permanent and temporary occupants in both study areas are visualised in Fig. 6, in the form of an occupancy density graph. The graph indicates the average percentage of study participants who are present in their respective study areas with time.

A comparison of the occupancy density graphs of the permanent occupants in both study areas showed some notable differences. It is observed that most permanent occupants in study area 2 follow regular office hours (9 a.m.–6 p.m.) while some of the permanent occupants in study area 1 follow flexible working schedules, with a significant minority staying after office hours. Furthermore, the occupancy patterns for the temporary occupants in both study areas also differ slightly in terms of the time of peak occupancy but generally follows the same triangle-shaped distribution. Based on the detailed breakdown of the occupants' designation type provided in Table 2, a significant portion of the occupants in study area 2 are administrative staff and faculty members who might prefer to follow more regular working hours while the occupants in study area 1 are mostly comprised of researchers and design engineers. As the differences in occupancy patterns between both study areas may be attributed to the disproportion in the occupants' designation types, it would be beneficial to merge the occupancy information from both study areas to obtain a more balanced distribution, thereby capturing a comprehensive set of occupancy profiles.

The merging process is conducted by relabeling the occupants' inferred zone-level location into one of the following five zones: Desk Area, Pantry, Printer, Meeting Room, and Other Work Areas. These new zone labels allow us to take into account different office layouts as well

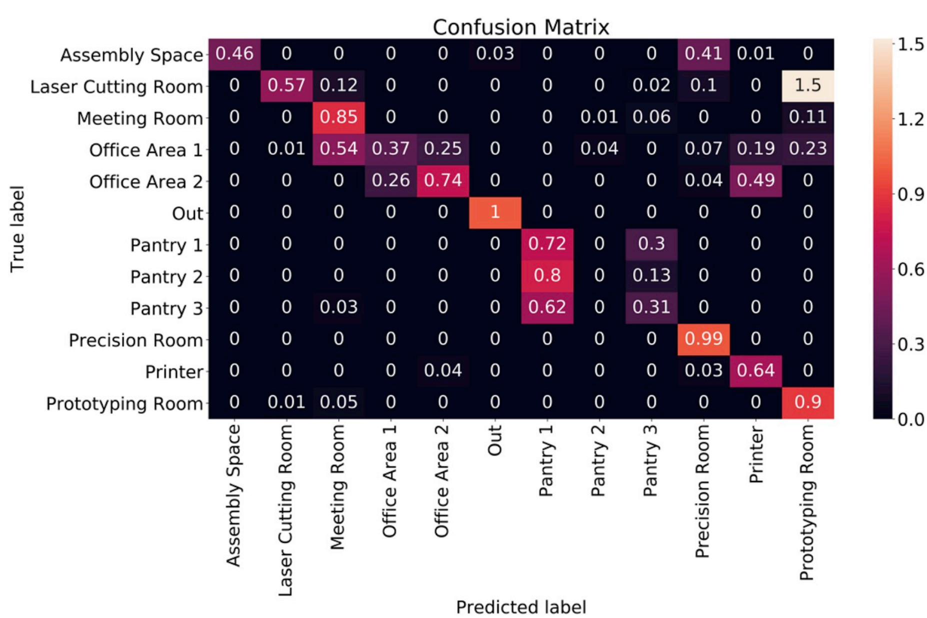

Fig. 5. Confusion matrix for semi-supervised clustering model for Study Area 1 (12 zones).

as the relative location of the occupant based on their assigned workstations. Therefore, when comparing between occupants who are seated in different zones of the office, their occupancy information, especially the fraction of time spent in each zone, remains comparable despite their different seating arrangements.

By combining the occupancy information of all permanent and temporary occupants from both study areas, the resulting occupancy density graph, depicted in Fig. 7, showed many notable differences from the ASHRAE Standard 90.1–2004. It is observed that the resulting occupancy density graph reflected the dynamic schedules of the occupants by recording a lower maximum occupancy density value of 0.6, while the ASHRAE standard assumes that the office will be mostly occupied (95%) for most of the office hours. The recorded change in occupancy levels during the start and end of each day are also less steep when compared to the ASHRAE standard, thus supporting the claim that most occupants exhibit diverse behaviours from each other. Finally, the recorded occupancy patterns showed more frequent instances where occupants stayed longer or returned to the office after regular office hours to resume their work. This observation is not adequately represented in the ASHRAE standard.

When comparing the occupancy patterns between the permanent and temporary occupants, there are many differences, in terms of their first arrival time, last departure time, as well as their average visiting frequencies based on a 5-day work week. It is observed from Fig. 8 that the permanent occupant's first arrival and last departure times can be represented by two distinct peaks, while most temporary occupants tend to arrive early afternoon and depart uniformly throughout the rest of the day. In terms of the occupants' weekly visiting frequency, a small group of occupants are observed to have surprisingly low visiting frequencies which could be attributed to flexible working schedules and the rise in employees opting to work remotely from their assigned workstations. On the other hand, a small minority of temporary occupants are observed to have surprisingly high visiting frequencies which could indicate a particular dependency on certain facilities or occupants to perform their daily roles.

Finally, Fig. 9 reflects the fraction of time both permanent and temporary occupants would spend at different zones in the study area. As expected, most permanent occupants will spend the majority of their time working at their desk area and the remainder of their time evenly divided between the pantry, meeting room, and other work areas. On the other hand, temporary occupants tend to spend most of their time either at the pantry, meeting room or other work areas. Based on these occupancy patterns, we can assume several reasons behind the temporary occupants' visit, which include attending a meeting, joining a friend for lunch at the pantry, approaching other occupants at their desks for work-related discussions, or using certain research facilities.

4.6. Occupancy profiles

By taking advantage of the detailed occupancy information obtained through the proposed method, the final output of this study is to identify a set of occupancy profiles that generalizes the different types of occupancy behaviours observed in a university office environment.

In this study, occupancy profiles are developed at two different levels. The first level expresses the occupancy information in terms of the presence of the occupant in the office. Therefore, occupancy information is represented in a binary format where a value of 0 indicates that the occupant is out of the office, and a value of 1 indicates that the occupant is present. The occupants' presence information for each day will form a single data point consisting of 1440 input features (due to 1440 min in a day) and be fed into a clustering algorithm to identify similar presence profiles.

The second level expresses the occupancy profiles at the occupant level, whereby the occupant's movement patterns over the study period are summarised based on several descriptive statistics and passed into the clustering algorithm as input features.

The clustering algorithm that is used to generate the different occupancy profiles is the K-means algorithm, and the optimal number of clusters K is determined based on the Davies Bouldin Index (DBI) [47].

4.6.1. Presence profiles

Presence profiles are most useful and informative when the occupants' movement patterns do not differ significantly between each day. Thus, the permanent occupants' daily presence information is ideal for developing such profiles and will be the focus of this subsection. By passing the permanent occupants' daily presence information through the K-means algorithm and calculating the corresponding average DBI score when K is varied between 2 and 10, it can be observed from Fig. 10 that the DBI score is lowest when K equals to 3. Therefore, the optimal number of clusters K is chosen to be 3 for the permanent occupants.

By clustering the permanent occupants' daily presence information into three distinct profiles, we can compare the occupancy density

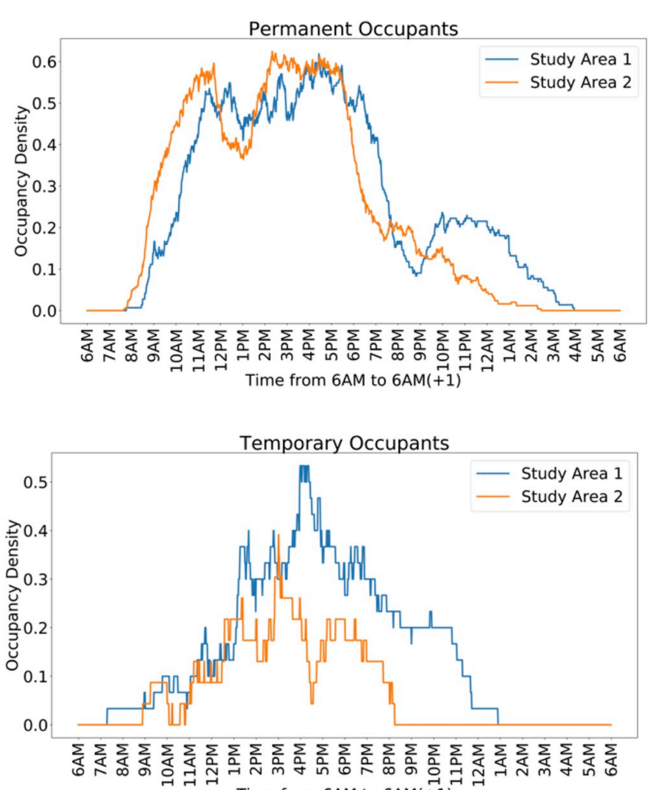

Fig. 6. Occupancy density graphs of permanent and temporary occupants in both study areas.

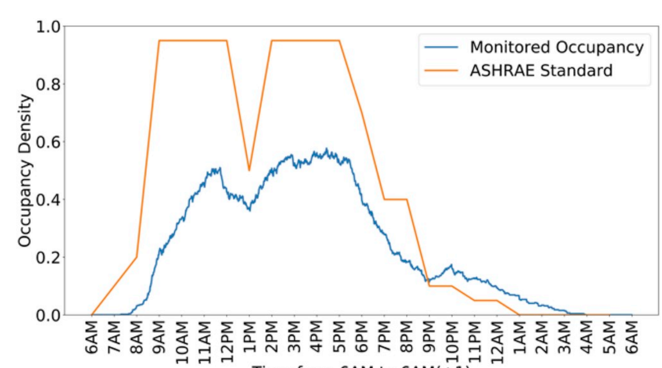

Fig. 7. Combined occupancy density graph for all occupant types compared with the ASHRAE standard for medium offices.

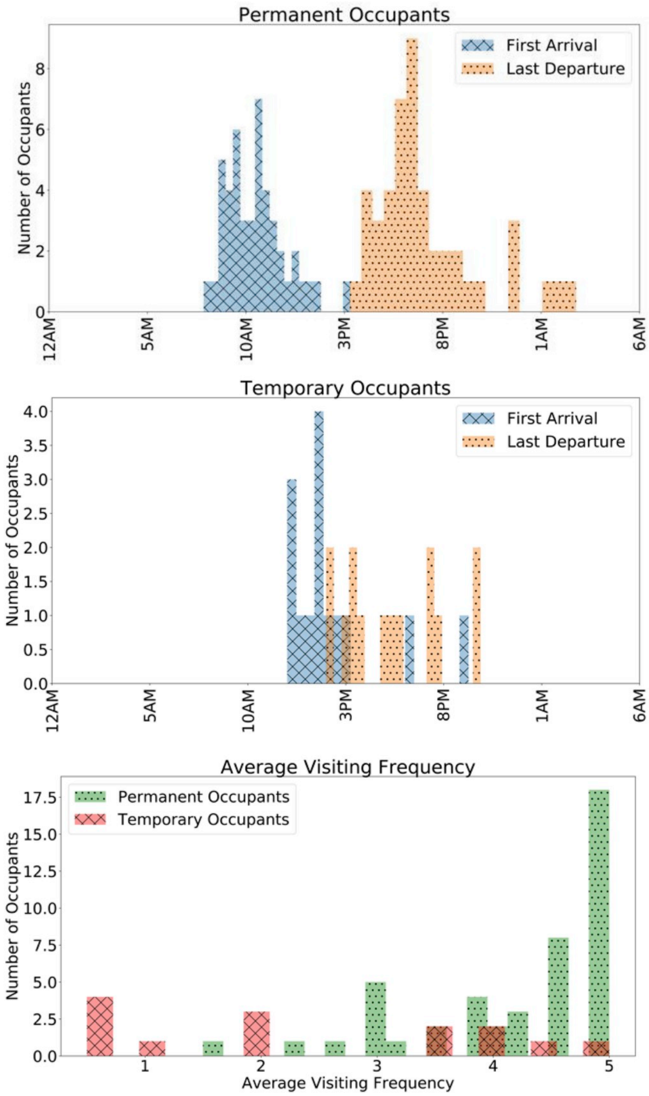

Fig. 8. First arrival, last departure and visiting frequencies of permanent and temporary occupants.

graphs of each profile in Fig. 11. Profile 2 (46.2%) follows the occupancy patterns of a regular office worker who comes to work in the morning at around 9 a.m.–10 a.m., takes a lunch break between 12 p.m. and 1:30 p.m., and leaves the office for the day between 6 p.m. and 10 p.m. Profile 1 (37.3%) seems to follow a similar occupancy pattern to Profile 2 in terms of the first arrival and last departure time but reports an average occupancy density between 0.2 and 0.4. This result is significantly lower as compared to Profile 2, which reported an average occupancy density between 0.6 and 0.8 during office hours. This observation could be attributed to the occupants' flexible work schedules or occupants who follow regular office hours but are continually moving in and out of the office due to external commitments. Finally, Profile 3 (16.5%) reflects an occupancy density graph of an occupant who has a first arrival time between 9 a.m. and 2 p.m., and last departure time between 11 p.m. and 3 a.m. in the next morning. Given the wide range of the first arrival time and last departure time, these office workers tend to follow more flexible work schedules that deviate from the regular office hours. Therefore, based on these observations, we will label the occupants that fall under Profile 1 as *Mobile Workers*, the occupants in Profile 2 as *Regulars*, and the occupants from Profile 3 as *Flexi-Timers*.

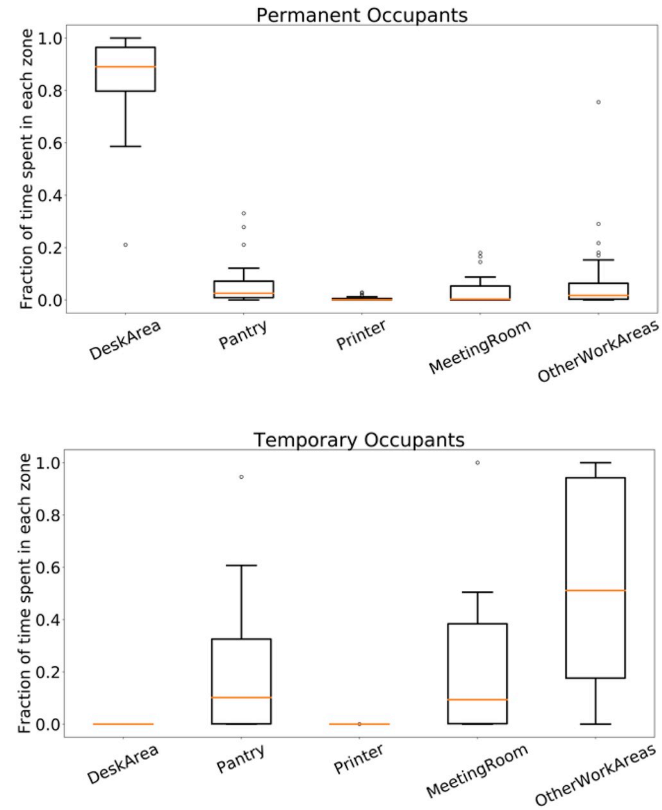

Fig. 9. Fraction of time spent in each zone for permanent and temporary occupants.

4.6.2. Occupant profiles

As we have concluded in subsection 4.5 that the temporary occupants' visits tend to be purpose-driven and are not dependent on the time of day, their movement patterns are best represented using different descriptive statistics such as visiting frequency, and the fraction of time spent in different zones in the office. In the case of temporary occupants, they are assumed not to have access to the printer, nor do they have an assigned desk area, as defined in subsection 4.3. Therefore, the fraction of time spent in these areas are assumed to be zero and do not need to be considered. By passing these descriptive statistics into the K-means algorithm as input features, the corresponding average DBI score is computed for each K value as K is varied between 2 and 10. Since the DBI score was the lowest when K equals to 5, the temporary occupants will be grouped based on five distinct clusters, as depicted in Fig. 12.

Based on a comparison of the different occupant profiles obtained through the clustering algorithm, we observed that temporary occupants

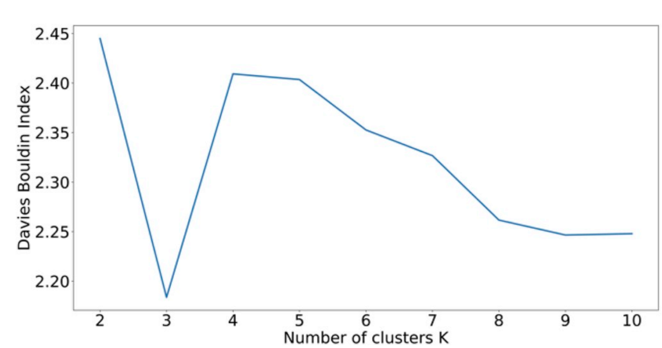

Fig. 10. A plot of the corresponding average DBI score (over 10 iterations) when the number of clusters K is varied between 2 and 10.

in Profile 1 (28.5%) and Profile 2 (21.4%) tend to visit the study areas more frequently than the other occupants over a 5-day work week, with the occupants in Profile 3 (14.3%) and Profile 4 (14.3%) spending most of their time in the pantry area and meeting room, respectively. While the occupants in Profile 2 and Profile 5 (21.4%) differ in terms of their visiting frequencies, they are similar in terms of the fraction of the time spent in the other work areas. The occupants in Profile 1 seem to visit the study area for multiple reasons as they spend their time evenly between the pantry, meeting room, as well as other work areas. Therefore, based on all of these insights, we will label the temporary occupants in Profile 1 to Profile 5 as *Mixed Visitors*, *Frequent Visitors*, *Lunch Buddies*, *Meeting Attendees*, and *Rare Visitors*, respectively.

5. Conclusion

In this study, we proposed a scalable and less intrusive occupancy detection method that leverages existing Bluetooth Low Energy (BLE) technologies found in smartphone devices to perform zone-level occupant localisation, without the need for a mobile application. The proposed method uses a network of BLE beacons to record the received signal strength indicator (RSSI) values of neighbouring devices which were consolidated and pre-processed to obtain a set of RSSI tuples. These RSSI tuples were passed into a machine learning model to infer the occupant's zone-level location. A supervised ensemble model and a semi-supervised clustering model were proposed and evaluated against the ground truth to identify the best performing model. The feasibility of the proposed method is demonstrated during a five-week case study involving two office spaces in an academic building in Singapore. While the supervised ensemble model produced the best performance in terms of overall accuracy and macro-average f1-score, the semi-supervised clustering model demonstrated practical advantages as it was able to produce a reasonable performance while using a fraction of the training data (<4%) and training time needed when compared to the supervised model.

Based on the occupancy information obtained using the supervised ensemble model, we further demonstrated the value of the proposed method by performing a series of analysis to identify a set of occupancy profiles to generalise the types of behaviours observed in the study area. The occupancy profiles are generated based on the occupants' presence information (presence profiles) as well as a summary of the occupant's movement patterns (occupant profiles). The K-means algorithm is used to cluster the occupants' movement information, and the optimal number of clusters K is determined based on the Davis-Bouldin Index (DBI). As a result, we identified three distinct presence profiles for permanent occupants and five unique occupant profiles for temporary occupants. Based on the outputs of this study, facility managers will have a better understanding of the other types of occupancy information that can be obtained through the implementation of the system, allowing them to manage the operations of their buildings better.

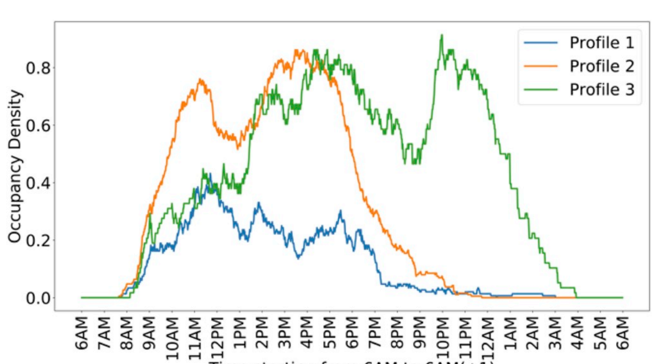

Fig. 11. Occupancy density graphs of three distinct presence profiles obtained based on the permanent occupants' movement patterns.

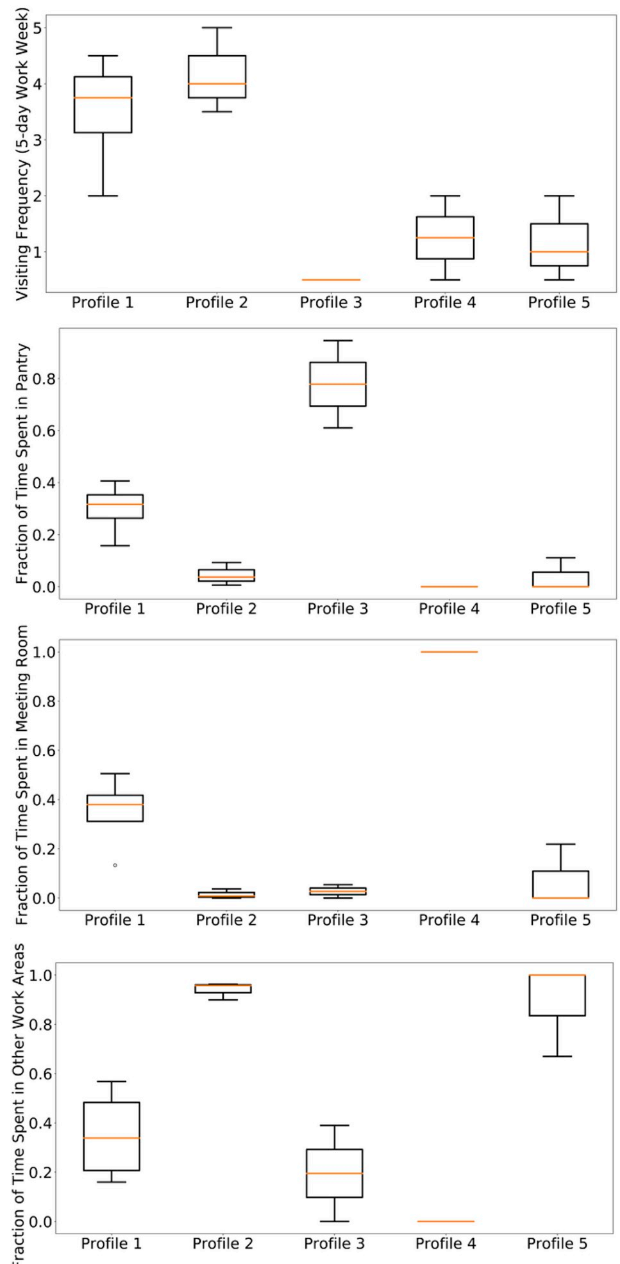

Fig. 12. Comparison between different temporary occupants grouped according to their occupant profiles.

Furthermore, while the occupancy profiles are obtained based on a small sample size of 60 occupants, the scalability of the proposed method allows the study to be easily implemented to other office spaces, thereby allowing us to extract realistic profiles that can be used as inputs into OB-centric models, given a sufficiently large sample size.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] T. Hong, D. Yan, S. D'oca, C.-F. Chen, Ten questions concerning occupant behavior in buildings: the big picture, Build. Environ. 114 (2017) 518–530.

[2] D. Yan, W. O'brien, T. Hong, X. Feng, H.B. Gunay, F. Tahmasebi, A. Mahdavi, Occupant behavior modeling for building performance simulation: current state and future challenges, Energy Build. 107 (2015) 264–278.
[3] M.A. Ul Haq, M.Y. Hassan, H. Abdullah, H.A. Rahman, M.P. Abdullah, F. Hussin, D. M. Said, A review on lighting control technologies in commercial buildings, their performance and affecting factors, Renew. Sustain. Energy Rev. 33 (2014) 268–279.
[4] M. Mysen, J. Rydock, P. Tjelflaat, Demand controlled ventilation for office cubicles—can it be profitable? Energy Build. 35 (2003) 657–662.
[5] J. Brooks, S. Kumar, S. Goyal, R. Subramany, P. Barooah, Energy-efficient control of under-actuated HVAC zones in commercial buildings, Energy Build. 93 (2015) 160–168.
[6] R. Melfi, B. Rosenblum, B. Nordman, K. Christensen, Measuring building occupancy using existing network infrastructure, in: 2011 International Green Computing Conference and Workshops, IEEE, 2011.
[7] T. Labeodan, W. Zeiler, G. Boxem, Y. Zhao, Occupancy measurement in commercial office buildings for demand-driven control applications—a survey and detection system evaluation, Energy Build. 93 (2015) 303–314.
[8] J. Ahmad, H. Larijani, R. Emmanuel, M. Mannion, A. Javed, Occupancy Detection in Non-residential Buildings–A Survey and Novel Privacy Preserved Occupancy Monitoring Solution, Applied Computing and Informatics, 2018.
[9] F. Topak, M.K. Pekeriçli, A.M. Tanyer, Technological viability assessment of bluetooth low energy technology for indoor localization, J. Comput. Civ. Eng. 32 (2018), 04018034.
[10] S. Lee, K.N. Ha, K.C. Lee, A pyroelectric infrared sensor-based indoor location-aware system for the smart home, IEEE Trans. Consum. Electron. 52 (2006) 1311–1317.
[11] D. Hahnel, W. Burgard, D. Fox, K. Fishkin, M. Philipose, Mapping and localization with RFID technology, in: IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA'04. 2004, IEEE, 2004.
[12] N. Li, B. Becerik-Gerber, Performance-based evaluation of RFID-based indoor location sensing solutions for the built environment, Adv. Eng. Inf. 25 (2011) 535–546.
[13] S. Depatla, A. Muralidharan, Y. Mostofi, Occupancy estimation using only WiFi power measurements, IEEE J. Sel. Area. Commun. 33 (2015) 1381–1393.
[14] B. Balaji, J. Xu, A. Nwokafor, R. Gupta, Y. Agarwal, Sentinel: occupancy based HVAC actuation using existing WiFi infrastructure within commercial buildings, in: Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems, ACM, 2013.
[15] W. Wang, J. Chen, T. Hong, N. Zhu, Occupancy prediction through Markov based feedback recurrent neural network (M-FRNN) algorithm with WiFi probe technology, Build. Environ. 138 (2018) 160–170.
[16] W. Wang, J. Chen, X. Song, Modeling and predicting occupancy profile in office space with a Wi-Fi probe-based Dynamic Markov Time-Window Inference approach, Build. Environ. 124 (2017) 130–142.
[17] W. Wang, T. Hong, N. Li, R.Q. Wang, J. Chen, Linking energy-cyber-physical systems with occupancy prediction and interpretation through WiFi probe-based ensemble classification, Appl. Energy 236 (2019) 55–69.
[18] A. Corna, L. Fontana, A.A. Nacci, D. Sciuto, Occupancy detection via iBeacon on Android devices for smart building management, in: Proceedings of the 2015 Design, Automation & Test in Europe Conference & Exhibition, EDA Consortium, 2015.
[19] A. Filippoupolitis, W. Oliff, G. Loukas, Occupancy detection for building emergency management using BLE beacons, in: International Symposium on Computer and Information Sciences, Springer, 2016.
[20] J.Y. Park, T. Dougherty, Z.N. Intelligent, A Bluetooth Based Occupancy Detection For Buildings, 2018.
[21] Z.D. Tekler, R. Low, L. Blessing, An alternative approach to monitor occupancy using bluetooth low energy technology in an office environment, J. Phys. Conf. Ser. 1343 (2019), 1. https://doi.org/10.1088/1742-6596/1343/1/012116.
[22] G. Diaper, The Hawthorne effect: a fresh examination, Educ. Stud. 16 (1990) 261–267.
[23] S.J. Emmerich, A.K. Persily, State-of-the-art Review of CO2 Demand Controlled Ventilation Technology and Application, Citeseer, 2001.
[24] S. Meyn, A. Surana, Y. Lin, S.M. Oggianu, S. Narayanan, T.A. Frewen, A sensor-utility-network method for estimation of occupancy in buildings, in: Proceedings of the 48h IEEE Conference on Decision and Control (CDC) Held Jointly with 2009 28th Chinese Control Conference, IEEE, 2009.
[25] N. Nassif, A robust CO2-based demand-controlled ventilation control strategy for multi-zone HVAC systems, Energy Build. 45 (2012) 72–81.
[26] X. Guo, D. Tiller, G. Henze, C. Waters, The performance of occupancy-based lighting control systems: a review, Light. Res. Technol. 42 (2010) 415–431.
[27] J. Lu, T. Sookoor, V. Srinivasan, G. Gao, B. Holben, J. Stankovic, E. Field, K. Whitehouse, The smart thermostat: using occupancy sensors to save energy in homes, in: Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems, ACM, 2010.
[28] Y.P. Raykov, E. Ozer, G. Dasika, A. Boukouvalas, M.A. Little, Predicting room occupancy with a single passive infrared (PIR) sensor through behavior extraction, in: Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ACM, 2016.
[29] T. Teixeira, G. Dublon, A. Savvides, A survey of human-sensing: methods for detecting presence, count, location, track, and identity, ACM Comput. Surv. 5 (2010) 59–69.
[30] O. Shih, A. Rowe, Occupancy estimation using ultrasonic chirps, in: Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems, ACM, 2015.
[31] S. Uziel, T. Elste, W. Kattanek, D. Hollosi, S. Gerlach, S. Goetze, Networked embedded acoustic processing system for smart building applications, in: 2013 Conference on Design and Architectures for Signal and Image Processing, IEEE, 2013.
[32] V.L. Erickson, S. Achleitner, A.E. Cerpa, POEM: power-efficient occupancy-based energy management system, in: Proceedings of the 12th International Conference on Information Processing in Sensor Networks, ACM, 2013.
[33] D. Liu, X. Guan, Y. Du, Q. Zhao, Measuring indoor occupancy in intelligent buildings using the fusion of vision sensors, Meas. Sci. Technol. 24 (2013), 074023.
[34] D. Chen, S. Barker, A. Subbaswamy, D. Irwin, P. Shenoy, Non-intrusive occupancy monitoring using smart meters, in: Proceedings of the 5th ACM Workshop on Embedded Systems for Energy-Efficient Buildings, ACM, 2013.
[35] M. Milenkovic, O. Amft, Recognizing energy-related activities using sensors commonly installed in office buildings, Procedia Comput. Sci. 19 (2013) 669–677.
[36] B.W. Hobson, D. Lowcay, H.B. Gunay, A. Ashouri, G.R. Newsham, Opportunistic occupancy-count estimation using sensor fusion: a case study, Build. Environ. 159 (2019) 106154.
[37] G. Conte, M. De Marchi, A.A. Nacci, V. Rana, D. Sciuto, BlueSentinel: a first approach using iBeacon for an energy efficient occupancy detection system, in: BuildSys@ SenSys, Citeseer, 2014.
[38] W. Shen, G. Newsham, B. Gunay, Leveraging existing occupancy-related data for optimal control of commercial office buildings: a review, Adv. Eng. Inf. 33 (2017) 230–242.
[39] W. Shen, G. Newsham, Smart phone based occupancy detection in office buildings, in: 2016 IEEE 20th International Conference on Computer Supported Cooperative Work in Design (CSCWD), IEEE, 2016.
[40] A. Filippoupolitis, W. Oliff, G. Loukas, Bluetooth low energy based occupancy detection for emergency management, in: 2016 15th International Conference on Ubiquitous Computing and Communications and 2016 International Symposium on Cyberspace and Security (IUCC-CSS), IEEE, 2016.
[41] M. Choi, W.-K. Park, I. Lee, Smart office energy management system using bluetooth low energy based beacons and a mobile app, in: 2015 IEEE International Conference on Consumer Electronics (ICCE), IEEE, 2015.
[42] A. Pratama, W. Widyawan, A. Lazovik, M. Aiello, Multi-user low intrusive occupancy detection, Sensors 18 (2018) 796.
[43] W. Wang, J. Chen, T. Hong, Modeling occupancy distribution in large spaces with multi-feature classification algorithm, Build. Environ. 137 (2018) 108–117.
[44] W. Wang, J. Chen, G. Huang, Y. Lu, Energy efficient HVAC control for an IPS-enabled large space in commercial buildings through dynamic spatial occupancy distribution, Appl. Energy 207 (2017) 305–323.
[45] Z.D. Tekler, R. Low, L. Blessing, Using smart technologies to identify occupancy and plug-in appliance interaction patterns in an office environment, IOP Conf. Ser. Mater. Sci. Eng. 609 (2019), https://doi.org/10.1088/1757-899x/609/6/062010, 062010.
[46] J.H. Friedman, Greedy function approximation: a gradient boosting machine, Ann. Stat. (2001) 1189–1232.
[47] D.L. Davies, D.W. Bouldin, A cluster separation measure, IEEE Trans. Pattern Anal. Mach. Intell. (1979) 224–227.