Building and Environment 200 (2021) 107936

Contents lists available at ScienceDirect

Building and Environment

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

A framework to identify key occupancy indicators for optimizing building operation using WiFi connection count data

Nastaran Alishahi, Mazdak Nik-Bakht *, Mohamed M. Ouf

Department of Building, Civil and Environmental Engineering, Concordia University, Montreal, QC, Canada

ARTICLE INFO

Keywords:
Building occupancy patterns
Peak time
Arrival and departure time
WiFi connection count
Energy efficiency
Machine learning
Poisson regression

ABSTRACT

Adapting building systems' operation to occupancy variations can provide significant energy savings, but this is typically constrained by the unavailability of occupancy information. Although several technologies have been introduced to integrate real-time occupancy information in heating, ventilation and air-conditioning (HVAC) systems operations (e.g., CO_2 monitoring), the logistical and cost issues associated with deploying these technologies remained a key issue. Previous studies proposed using WiFi counts as a proxy for occupancy in building operations and showed a strong correlation between occupancy and WiFi counts in several building types. However, the difficulty of integrating real-time WiFi traffic data in building automation systems hinders wide-scale deployment of this approach. To this end, this study proposes a framework for extracting occupancy indicators from WiFi traffic data. The proposed framework utilizes several machine learning algorithms and statistical analysis methods to predict patterns of building occupancy as well as to identify peak occupancy time and earliest/latest arrival and departure times. To validate the proposed framework, it was implemented in a case-study using data collected from an academic building in Montreal, Canada between January and March 2020. Results revealed that the proposed models could successfully predict weekly building occupancy patterns, with an average accuracy (R^2_D) of 0.98 for weekdays and 0.81 for weekends. Furthermore, the analysis identified peak occupancy timing, as well as arrival and departure times variations between different zones. These findings provided a proof-of-concept for the proposed framework and demonstrated its potential to provide actionable information to modify the sequences of operation of building systems based on buildings' unique occupancy patterns.

1. Introduction

Occupants are among the key stakeholders in the operation phase of any building's life cycle and have a great impact on its energy consumption [1]. Studies have stressed the importance of understanding occupancy patterns for implementing energy-saving strategies in buildings, especially for controlling Heating, Ventilation, and Air Conditioning (HVAC) systems, which are responsible for about 40–50% of the energy consumption in buildings [2,3]. Based on previous studies, the integration of occupancy information into the HVAC system operation can result in between 10% and 40% reduction in building energy consumption [4]. This shows the importance of accurate information about occupancy patterns for adjusting HVAC operation schedules.

Despite the previous findings, most buildings are still being operated under the assumption of full or nearly full occupancy; which can lead to significant energy waste and occupant discomfort [5]. Previous studies have proved considerable temporal and spatial deviations in building occupancy patterns, especially in large commercial and institutional buildings [6,7]. In such buildings, occupants arrive and/or leave the spaces at different times of the day, while peak occupancy can be influenced by different events. However, many studies suggest that although significant variations exist in occupancy patterns of different buildings (even those of similar types), they tend to follow relatively repetitive patterns for individual buildings [8–10]. In other words, once the occupancy patterns of a specific building are thoroughly investigated and identified, no significant changes should be expected in those patterns, unless exceptional events take place.

In recent years, researchers investigated various technologies for obtaining occupancy information, such as carbon dioxide (CO_2) or passive infrared (PIR) sensors [11,12]. These studies focused on developing HVAC control strategies to utilize occupancy information obtained from the adoption of these methods/technologies [5]. However,

* Corresponding author. 1455 De Maisonneuve Blvd. W., Montreal, QC, H3G 1M8, Canada.
E-mail address: mazdak.nikbakht@concordia.ca (M. Nik-Bakht).

https://doi.org/10.1016/j.buildenv.2021.107936
Received 10 December 2020; Received in revised form 27 April 2021; Accepted 1 May 2021
Available online 13 May 2021
0360-1323/

many of these technologies have limitations in terms of accuracy, cost, intrusiveness, and privacy. For example, while demand-controlled ventilation using CO_2 sensors has been successfully commercialized, its deployment in large buildings is still relatively limited due to the significant cost and maintenance requirements [13]. On the other hand, using WiFi connection counts as proxy data for occupancy has attracted increasing attention. The wide availability of WiFi networks in buildings and the strong correlation between WiFi connection counts and occupancy counts [13,14] are among the reasons for the popularity of this technology in occupancy sensing research.

Previous studies focused on investigating the correlation between WiFi connection counts and the number of occupants [13–18]; predicting occupancy patterns using historical data [8,15,19,20]; or using sensor fusion techniques with other data streams to increase the accuracy of occupancy patterns estimation [21,22]. At the core of these studies, automated data exchange between WiFi networks and building automation systems (BAS) is assumed as the path forward for utilizing WiFi connection counts data to optimize the HVAC system's operations. However, since WiFi networks and the BAS are typically managed by two different teams within most organizations, the widespread adoption of this approach in action has been hindered due to practical and liability issues, including data exchange, communication, and operators' reluctance to automation. Furthermore, potential issues due to sudden WiFi interruptions or temporary system shutdowns may have significant negative consequences for the operation of HVAC systems and compromise occupants' health and safety (e.g., due to under ventilation).

To this end, the goal of this paper is to introduce a framework for extracting occupancy pattern information using WiFi connection count data in existing buildings. To help with bridging the implementation gaps, the required inputs for this framework are limited to authenticated and associated counts of WiFi connections, which are readily available and accessible in most buildings. The framework focuses on identifying key occupancy metrics relevant to HVAC operation offline. To validate the capability of the proposed framework in identifying these key metrics, a case study is presented, focusing on (i) learning and predicting weekly occupancy patterns; (ii) predicting peak occupancy as well as identifying its occurrence at the building-level; and (iii) identifying earliest/latest arrival and departure times at the zone-level. All this information regarding the building's unique occupancy patterns can later be used to adjust the building operation schedules.

The rest of this paper is organized as follows. Section 2 provides a review of relevant studies; Section 3 presents the methodology; Section 4 introduces the case study; Section 5 presents the results of the experiment; Section 6 discusses the major findings; and finally, Section 7 provides the conclusion by explaining the study contribution as well as limitations and proposing the future works.

2. Literature review

Previous studies focused on obtaining accurate occupancy information at different temporal and spatial resolutions [23,24] to optimize the operation of HVAC systems and to maintain acceptable thermal comfort in buildings [25]. These studies utilized different technologies and methods to investigate occupancy count information with three main objectives, including (i) real-time occupancy estimation; (ii) future occupancy prediction; and (iii) occupancy pattern identification. In the following sub-sections, details on different technologies and methods adopted in previous studies are provided and then studies that have used WiFi technology for occupancy counting are discussed.

2.1. Occupancy counting technologies

Currently, the level of CO_2 concentration is one of the most commonly used indicators of occupancy levels in HVAC operations. Accordingly, many studies have focused on improving the accuracy of occupancy estimation using CO_2 sensors [26–32]. However, CO_2 concentration can be misleading since it is sensitive to situational factors including the location of sensors, occupants' activity, occupancy density, and open/closed state of the doors and windows. Moreover, there is always a delay between the change in the occupancy count in space and the variation in the level of CO_2. Other environmental sensors, measuring temperature, humidity, and pressure, can improve the accuracy of occupancy estimation, when combined to form a sensor fusion approach [33–37]. The most important step in using environmental sensors is feature engineering since they are an indirect proxy for occupancy [12,38].

Using cameras is another technology that measures occupancy counts more directly, through image and/or video processing [39–43]. The cameras are also used for obtaining occupancy information at other occupancy resolutions such as presence detection, localization, tracking, and identification based on their types (i.e. depth or non-depth) and their installed locations (i.e. entrance overhead or room interior) [11]. Compared to other technologies that are more applicable for occupancy counting at a small scale (i.e. rooms with low occupancy rate), several studies used some types of cameras for real-time crowd counting at a large scale (i.e. large areas with a dense occupancy) in complex scenes [41]. In general, cameras can be used to estimate occupancy with comparatively high accuracy. However, the major issues associated with using cameras for occupancy estimation are high computational complexity and privacy concerns [12]. Hence, they are mostly used as a source of providing ground truth data for validating results from other technologies [44].

Two other common technologies proposed in previous studies are radio frequency identification (RFID) tags [45,46] and Bluetooth Low Energy (BLE) beacons [47,48]. Despite the high level of accuracy, both technologies have deployment limitations. They both need users to carry tags (or other forms of wearables), while the BLE can take advantage of the Bluetooth on users' devices (such as smartphones or smartwatches); the users must always keep the Bluetooth of their devices on, which is not necessarily common. Therefore, they are not suitable, particularly for occupancy estimation in large areas with a high level of occupancy variation, dynamic and ad-hoc occupants, such as institutional or academic buildings. Using Passive infrared (PIR) sensors is another common direct approach, mostly applicable to detect occupants' presence [49,50]. However, there are also studies using PIR sensors in more complex set-ups to detect moving directions of occupants [51] or identify their motion patterns [52] and finally infer occupancy counts. In general, all these technologies require installing sensors and an additional infrastructure, allocated to the occupancy counting, which will require extra installation and maintenance costs.

2.2. Occupancy counting models

In recent years, the application of machine learning algorithms to extract occupancy information has been increased [4]. Artificial Neural Networks (ANN) and Support Vector Machines (SVM) are among the most frequently used data-driven methods for real-time occupancy estimation [53]. Zuraimi et al. (2017) developed ANN and SVM models to estimate occupancy in a theater hall with a maximum occupancy of 200 people, from CO_2 concentration level data. The models reached an average accuracy of 70% and 76%, respectively [54].

For the prediction of future occupancy, Markov models are typically being proposed. Ryu and Moon (2016) developed a Hidden Markov Model (HMM) for occupancy prediction, achieving an accuracy ranging from 85% to 93.2% in a small-scale experiment [55]. Li and Dong (2018) developed an inhomogeneous Markov model to predict occupancy in office buildings and evaluated its performance, in various horizons from 15-min ahead to 24-h ahead [56]. They showed the proposed model performs better in short-term horizons (e.g. 15-min) with Root Mean Squared Error (RMSE) of 0.510, achieved in an experiment with a maximum of 6 occupants. In another study on residential buildings by

Huchuk et al. (2019), it was shown that random forest outperforms logistic regression, Markov model, HMM, and Recurrent Neural Networks (RNN) in predicting occupancy over various horizons from 30 min to 3 h, with a median accuracy of approximately 75% [57]. In general, few studies have developed prediction models for future occupancy prediction, and they have mostly focused on short-horizon (at the most day-ahead) with a limited number of occupants.

2.3. Occupancy counting using WiFi

WiFi networks are widely installed in modern buildings, especially in offices and educational buildings. Furthermore, using WiFi-enabled devices, including smartphones, tablets, and laptops, is very common among occupants of such buildings. These factors make WiFi connection count data a cost-efficient and reliable approach to obtain building occupancy information non-intrusively, and use it for different purposes, including HVAC operation optimization. For example, Wang et al. (2018) used WiFi probe technology to estimate real-time occupancy for an occupancy-based ventilation strategy. They compared the energy consumption share of ventilation in their proposed strategy, to the fixed-rate ventilation, in an experiment on a graduate students' office. The proposed strategy showed a saving of about 44.26% and 55.5% on weekdays and weekends, respectively [58]. In another study on a commercial building, almost 26.4% of energy consumption in cooling and ventilation demands was saved through inferring occupancy information from WiFi connections [59]. Balaji et al. (2013) also used WiFi data for HVAC operation in a commercial building and reached an electricity saving of 17.8% [60].

A statistically significant strong positive correlation between WiFi connection counts and actual occupancy counts has been confirmed by previous studies [13,14]. For example, Ouf et al. (2017) showed a correlation coefficient (R) of 0.84 between the two, which was stronger than the relationship of CO_2 concentration level and occupancy counts [13]. In another study, a stronger correlation was found between occupancy counts and WiFi connection counts (with an R of 0.84) compared to beam counters' log [14]. Although these experiments were conducted at a room-level, they successfully identified the potentials of using WiFi connection count data as a proxy to estimate occupancy counts at the building-level. Simma et al. (2019) investigated the correlation of WiFi connection counts and occupancy counts using the linear regression algorithm resulting in a coefficient of determination (R^2) value ranged between 0.86 and 0.96 [18]. Wang and Shao (2018) calculated the ratio of WiFi connection counts and occupancy counts in a typical study room to be around 1.16 [16]. This ratio is a function of time and space type [15,47]. All these studies revealed that WiFi connection counts can be a strong representative of the occupancy pattern. Following this strong correlation, there are some recent studies that directly used WiFi connection count instead of occupancy counts for different purposes including space utilization [61], emergency evacuation [62], and evaluating building energy performance [63,64].

The main focus of WiFi-based occupancy sensing studies has been on real-time occupancy estimation. Among them, many have also compared WiFi data with other data streams, using different machine learning algorithms. Hobson et al. (2019) developed ANN and Multiple Linear Regression (MLR) models using data from WiFi APs, CO_2 sensors, PIR motion detectors, and plug and light electricity load meters, to estimate the occupancy counts [21]. The developed models using sensor fusion containing WiFi data showed significantly higher accuracy than other datasets. The maximum and mean R^2 values resulted from models developed using WiFi data alone were 0.97 and 0.71, respectively in ANN models and 0.96 and 0.74, in MLR models. Wang et al. (2018) employed ANN, k-Nearest Neighbors (kNN), and SVM to estimate occupancy counts using three datasets: an environmental dataset, a WiFi signal dataset, and a fused dataset of both [22]. They showed that SVM and kNN trained on the WiFi data alone, provide fewer counting errors.

There are studies that benefited from other potentials of WiFi technologies, such as Received Signal Strength Indicator (RSSI) to estimate the position of occupants, in these studies, some included the occupancy counting as another occupancy resolution. For example, Zou et al. (2017) proposed a system for providing fine-grained occupancy information at different resolutions including occupancy detection, counting, and tracking [65]. In an experiment conducted on an office with almost 20 occupants, their system obtained a Normalized Root Mean Square Deviation (NRMSD) of 0.096 in estimating the number of occupants. Yoo et al. (2020) proposed a stations-oriented indoor localization system [66]. In an experiment on multiple offices with a maximum of 12 occupants, they obtained the overall Normalized Root Mean Square Error (NRMSE) of 0.0309 in estimating occupancy counts. Although using RSSI for occupancy counting at the building-level increases the complexity, it might be a better option than using the WiFi connection counts for estimating the number of occupants inside a room or zone. Since the connections that are out of the desired boundary can be filtered out based on a threshold defined for RSSI [47,67]. Longo et al. (2019) took this approach to estimate the number of occupants and achieved an RMSE ranging between 1.42 and 5.12 while experimenting on multiple academic spaces with a maximum of 132 occupants [47]. However, this approach might not be a reasonable candidate for occupancy estimation at large scale since it might invade occupants' privacy due to the need for MAC addresses and also the environment including building components can affect the signal strength.

On the other hand, Channel State Information (CSI), another potential of WiFi technologies, describes the details of WiFi signal propagation from the transmitter to the receiver. Studies using CSI measured the impact of a certain number of occupants on signal propagation and estimated occupancy counts through classification techniques [68]. Studies that used CSI for occupancy counting could achieve an accuracy ranging between 81% and 96% [69,70]. However, this approach is mostly applicable in small spaces with a limited number of occupants since it is highly dependent on the environment and movements.

Among the studies using WiFi data, only a few focused on future occupancy prediction. Wang et al. (2017) used WiFi probe to predict occupancy in the following time window based on previous time windows, through a Dynamic Markov Time-Window Inference (DMTWI) model [19]. They conducted an experiment with a time window length of 20 min on a research student office room with less than 20 occupants. They achieved a prediction accuracy of 80% with a tolerance of four (4) occupants on weekdays; three (3) on holidays, and two (2) on weekends. Their proposed model was also compared versus Auto-Regressive Moving Average (ARMA) model and Support Vector Regression (SVR), and showed slightly higher accuracy. Ashouri et al. (2019) developed MLR and ANN models to predict day-ahead occupancy [15]. In an experiment on an office building, they achieved an R^2 of 0.96 for ANN and 0.88 for MLR, indicating a superior performance for ANN. However, the authors proposed MLR as a more reasonable candidate for future occupancy prediction due to its lower computational complexity. They also developed a linear regression model with limited ground truth data to translate WiFi connection counts to occupancy counts and achieved an R^2 of 0.9. In another study, Hobson et al. (2020) used WiFi data to develop a classification tree for predicting day-ahead occupancy day types [8]. During an experiment on an academic office building, they fed the lighting and plug load profiles (which follow the same patterns of WiFi data in 84.5% of days) into the classification tree and reached a successful classification rate of 70.4%. Apostolo et al. (2021) used WiFi connection data at AP-level in order to predict the Wi-Fi network demand for energy-efficient smart buildings [20]. They developed multiple classification and regression models based on a combination of different parameters such as features, machine learning algorithms and etc. to find the best model. In an experiment on a classroom building, they reached the best Root Mean Squared Percentage Error (RMSPE), Mean Absolute Percentage Error (MAPE), and RMSE for connection counts prediction, with the values of 0.29, 0.41, and 8.41, respectively. A summary of major research works focusing on WiFi-based occupancy counting is presented in Table 1.

Although several previous studies investigated using WiFi data as a reliable proxy for occupancy estimation, relatively few studies focused on using this data for extracting occupancy indicators relevant to HVAC operation. Most of these studies achieved comparatively high accuracy, however, they mostly focused on a room-level/zone-level occupancy estimation or validation, which is less nuanced than the entire building's occupancy. Furthermore, the developed models were tested over relatively short periods of time. The present study aims to bridge these identified gaps in the literature and develop a framework for extracting practical occupancy indicators from WiFi connection count data, that can be later used by building operators to adjust HVAC systems' schedules. This approach can overcome various challenges related to the automation of data exchange between WiFi networks and BAS.

3. Methodology

As illustrated in Fig. 1, the proposed framework consists of three main components to identify key occupancy indicators using WiFi connection count data. The first component focuses on learning and predicting weekly occupancy patterns at the building-level. Next, the analysis focuses on predicting peak occupancy and identifying its occurrence interval on different days of the week. Finally, earliest/latest arrival and departure times are identified at a zone-level. Extracting these metrics can provide practical information for building operators, especially for adjusting ventilation schedules based on weekly occupancy patterns and the peak occupancy occurrence, as well as adjusting temperature setpoints and scheduling setbacks for different building zones.

It is worth noting that the proposed framework focuses on identifying indicators that characterize occupancy patterns (e.g., arrival/departure times, or occurrence of peak occupancy) rather than the exact number of occupants by considering that there is a strong correlation between WiFi connection counts and occupancy counts. Accordingly, the times at which the first and last WiFi connections are observed are considered as the earliest arrival and latest departure times of

Table 1
Summary of major WiFi-based occupancy counting studies.

Problem	Reference	Data	Spatial Resolution	Case	Experiment Duration	Method (Technique)	Performance
Occupancy estimation using WiFi data	Di Domenico et al. (2016) [69]	CSI	Room-level	Office room (max 7 occupants)	Not mentioned	Linear discriminant classifier	2-Accuracy: 81–91%
	Zou et al. (2017) [65]	RSSI	Zone/Room level	Multi-functional office and lab (max 20 occupants)	4 weeks	WiFi-based non-intrusive Occupancy Sensing System (WinOSS)	NRMSD: 0.096 Mean error: 0.196 STD error: 0.578
	Ouf et al. (2017) [13]	Device counts, CO_2	Room-level	Classroom (max 80 occupants)	1 week	MLR; Pearson's product-moment correlation	R^2: 0.703 R = 0.839
	Mohottige et al. (2018) [14]	Device counts	Room-level	Classroom (max 200 occupants)	1 week	Pearson's product-moment correlation	R = 0.85 SMAPE: 12.1%
	Zou et al. (2018) [70]	CSI	Room-level	Meeting room (max 11 occupants)	2 days	Transfer kernel learning	Accuracy: 90.2–96%
	Wang et al. (2018) [22]	RSSI	Room-level	Graduate student office (max 20 occupants)	3 days	kNN; SVM; and ANN	MAE: 2.1–2.5 MAPE: 34.3–37 RMSE: 2.8–3.3
	Longo et al. (2019) [47]	RSSI	Room-level	Multiple academic spaces (max 132 occupants)	25 h	Regularized linear regression; Regularized multinomial logistic regression	RMSE: 1.42–5.12 MAE: 1.05–4.25 Accuracy: 68–95%
	Azam et al. (2019) [71]	RSSI	Room-level	Open office (max 20 occupants)	9 weeks	Decision Tree; Random Forest; Gradient Boosting; Extremely Randomized Trees	Accuracy: 95%
	Wang et al. (2019) [17]	Device counts	Zone-level	Private and cubicle offices (max 74 occupants)	5 weeks	Long Term Short Term Memory (LSTM) networks; Random Forest; ANN	2-Accuracy: 70–72% RMSE: 3.95–4.62
	Hobson et al. (2019) [21]	Device counts	Floor-level	Academic office (max 72 occupants)	208 h	MLR; ANN	Max R^2: 0.96–0.97 Mean R^2: 0.71–0.74
	Jagadeesh Simma et al. (2019) [18]	Device counts	Room-level	Lecture room (max 100 occupants)	6 weeks	Linear regression	R^2: 0.86–0.96
	Ashouri et al. (2019) [15]	Device counts	Building-level	Office (max 80 occupants)	19 h	Linear model	R^2: 0.9
Occupancy pattern clustering	Wang and Shao (2018) [16]	Connection data	Room-level	Study room (max 20 occupants)	14 h	K-means clustering	Not reported
	Hobson et al. (2020) [8]	Device counts	Building-level	Academic office (max 570 occupants)	7 months	K-means clustering	Not reported
Future occupancy prediction using WiFi data	Wang et al. (2017) [19]	Connection data	Room-level	Research students' office rooms (max 68 occupants)	3 days	DMTWI; ARMA; SVR	Accuracy (when up to 5 occupants error is allowed): 85%
	Ashouri et al. (2019) [15]	Device counts	Building-level	Office (max 80 occupants)	9 weeks	MLR; ANN	R^2: 0.88–0.96 RMSE: 2.9–5.0 Accuracy: 83.1–90.1%
	Hobson et al. (2020) [8]	Device counts, plug and light load	Building-level	Academic office (max 570 occupants)	7 months	Decision Tree	Classification rate: 70.4%Error: 47 ± 69 occupants
	Apostolo et al. (2021) [20]	Connection data	AP-level	A classroom building	6 months	Random Forest; Decision Tree; kNN; XGBoost	Min RMSE: 8.4130 Min RMSPE: 0.29 Min MAPE: 0.41

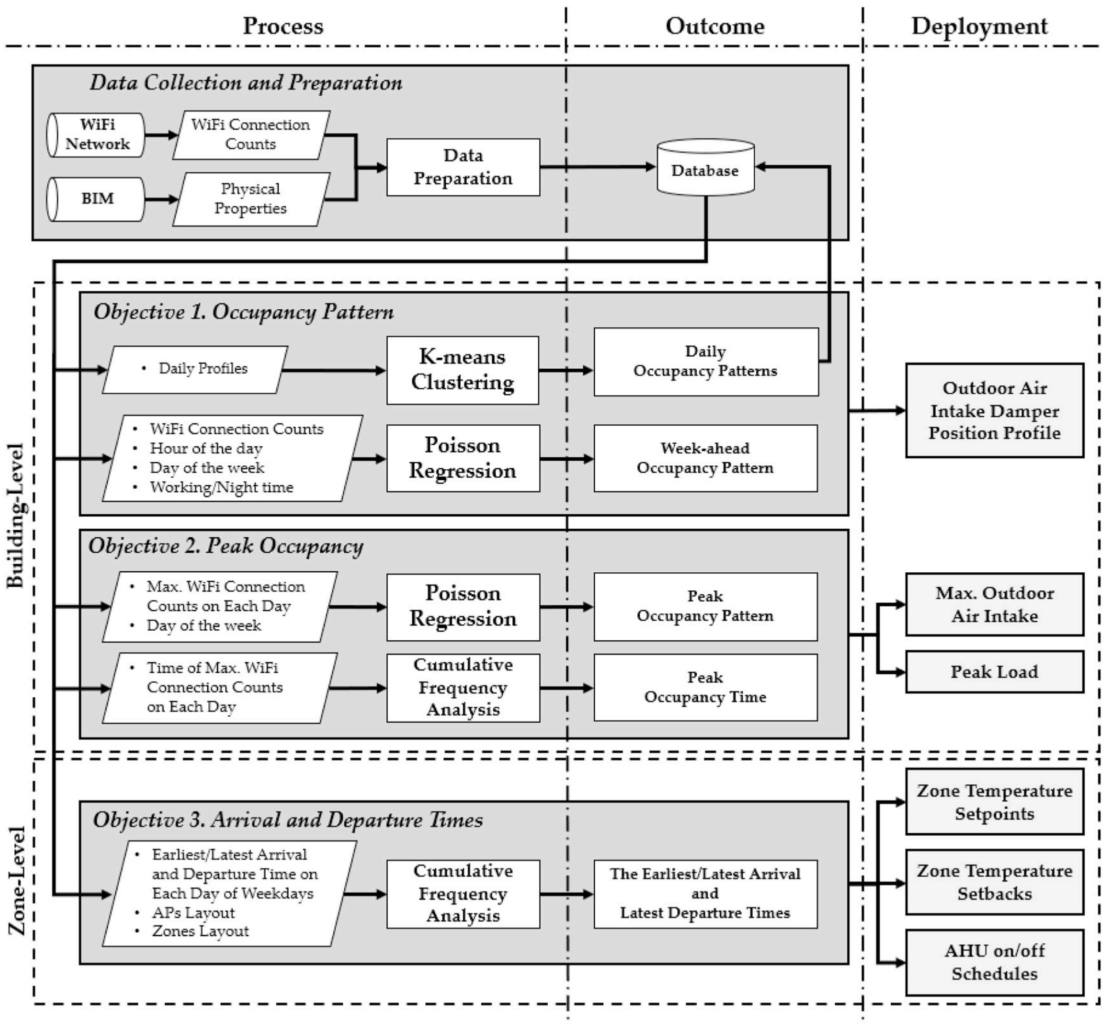


Fig. 1. The high-level methodology of the study.

occupants, respectively. Moreover, the time at which the maximum WiFi connection counts is observed is considered as the peak occupancy time.

3.1. Data collection and preparation

The first step entails collecting the number of devices connected to each WiFi Access Point (AP) in the building. Typically, most WiFi network administration platforms can provide hourly or sub-hourly reports on the number of 'authenticated' and 'associated' device counts. While the former refers to the connected devices, the latter provides the total number of devices trying to connect to the network (successfully and unsuccessfully) [72]. In the proposed framework, the number of associated device counts is used due to its additional coverage of occupants (e.g., visitors who are not authenticated in the WiFi network). Privacy concerns are limited in this data collection approach since it only relies on the aggregated device counts (i.e., no personally identifying information is being collected or used). In addition to WiFi data, the building's contextual information including physical properties and floor areas of the building zones, as well as the HVAC system's layout are retrieved from the 3D BIM (Building Information Model). By integrating the location of APs in the BIM, APs' coverage can be estimated based on the floor layouts and building components' materials, if available, then it can be attributed to zones served by individual Air Handling Units (AHU). The last step entails cleaning and transforming the collected WiFi data by handling missing values, duplications, and outliers. To identify outliers among daily profiles and investigate the possibility of grouping various days based on the similarity of their occupancy patterns, the Kruskal-Wallis H test is used, which is a non-parametric statistical test used to determine whether statistically significant differences exist between different subgroups.

3.2. Occupancy pattern prediction

Occupancy pattern prediction entails two steps, i.e. identifying day type groups that follow similar patterns; and developing prediction models for each group. For the first step, 'day type' can be decided extrinsically (e.g. weekdays vs weekends) or intrinsically (i.e. through clustering of observed patterns on different days). For the latter, the centroid-based clustering algorithm (k-means clustering) is selected after evaluating k-shape and k-means clustering techniques. The clustering is applied to normalized daily occupancy profiles and the similarity between data-points is measured through Euclidian distance. To find the optimal *k* value in k-means clustering, Davies Bouldin Index (DBI) is used as the mathematical performance measure besides two other factors, namely the number of clusters and the cluster size (i.e. the number of items in each cluster).

The next step entails developing prediction models, considering the results of clustering daily occupancy patterns to improve the prediction's performance. The features include the values of WiFi connection counts, the hour of the day, and the day of the week, all of which are extracted from the collected time-series WiFi connection count data. In addition, comparing the performance of models developed for predicting the entire day, versus models predicting the working time period only (i.e. 8:00 to 21:00) suggested that introducing an additional Boolean feature to distinguish the working time (i.e. 8:00 to 21:00) against nighttime (i.e. before 8:00 and after 21:00), can improve the prediction performance despite its correlation with the 'the hour of the day' attributes. All these features are used for training prediction models. Evaluating prediction models developed on different datasets shows slightly better performance in predicting typical days when using separate models for specific day types, rather than having one model for all.

To develop the prediction model, Poisson regression, which is a type of Generalized Linear Models (GLM), is selected. Unlike MLR, which is the common modeling technique for occupancy estimation, Poisson regression uses the Maximum Likelihood Estimation (MLE) technique to identify regression coefficients. Poisson regression assumes the response variable follows a Poisson distribution; hence it is ideal for modeling event-based and discrete WiFi connection count data. The performance of the developed models is then evaluated based on the error, using RMSE and MAPE, and Pseudo R-squared (R^2_D) – which is also called 'percentage of deviance'. R^2_D is a generalization of regular R^2 in the Ordinary Least Squares (OLS) method and is calculated through the following equation [73]:

$$\mathrm{n}\boldsymbol{R}_D^2 = 1 - \frac{\sum_i \left[y_i \log\left(\frac{y_i}{\widehat{\mu}_i}\right) - \left(y_i - \widehat{\mu}_i\right)\right]}{\sum_i \left[y_i \log\left(\frac{y_i}{\overline{y}}\right) - \left(y_i - \overline{y}\right)\right]}, \tag{1}$$

n#

where $\boldsymbol{y} = (y_1, \ldots, y_n)$ and $\widehat{\boldsymbol{\mu}} = (\widehat{\mu}_1, \ldots, \widehat{\mu}_n)$are the actual values of the dependent variable and the corresponding predicted values, respectively; and

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n} \tag{2}$$

3.3. Peak occupancy analysis

Peak occupancy is investigated in two levels including the maximum WiFi connection counts and the time of its occurrence. Firstly, the maximum WiFi connection counts on each day and day of the week are identified as features for training prediction models through Poisson regression. R^2_D is also used as the measure to evaluate the impact of extending the training period, on improving the performance of peak occupancy prediction. The process starts by predicting the peak occupancy of each week from the previous week's data, then introducing data from additional past weeks until reaching an acceptable accuracy of prediction. Through this process, the R^2_D value is calculated by averaging the performance of prediction models trained with the same number of training weeks.

Secondly, to investigate the occurrence time of peak occupancy, cumulative relative frequency distributions of the time when the maximum WiFi connection counts occur on each day are used. These plots can identify the probability of peak occupancy occurrence at different times of the day, for each day of the week. The identified occupancy profile on different days, as well as the most likely time of peak occupancy occurrence, can then be used by building operators to adjust ventilation schedules accordingly to mimic these profiles.

3.4. Arrival and departure times analysis

Arrival/departure times analysis needs zone-level WiFi connection counts which firstly entails mapping APs to AHUs' zones; and secondly involves aggregating WiFi connection counts of all APs assigned to each zone. Then, on each day, the time at which the first WiFi connection is detected is classified as arrival time, while the time at which the number of WiFi counts drops to zero is classified as departure time. Cumulative relative frequency distributions of arrival and departure times in each zone are plotted to identify the probabilities of earliest/latest arrival and departure times in different zones. These indicators can then be used by building operators to schedule temperature setpoints and setbacks.

4. Case study

The proposed framework was tested on a 17-story academic building with a gross floor area of 37,000-m^2, located in Montreal, Canada. The building is LEED silver certified, and its floors vary in layout and space type. Classrooms are mostly located on the lower eight floors, starting at sub-basement two through the sixth floor. Therefore, these eight floors were selected for this case study. The specifications of each floor retrieved from the BIM are summarized in Table 2.

4.1. Building-level data collection and preparation

A total of 89 Cisco Aironet AIR-CAP3702I and AIR-CAP3602I APs are located throughout the first eight floors of the case study building; mostly concentrated in classrooms and auditoriums. According to the Instructional and Information Technology Services (IITS) team's verification, these APs cover the entire eight floors of the building and there are no dead zones on these floors. For this study, the anonymized reports of WiFi connection counts (generated by the APs) were provided in 3-min intervals for the period of nine weeks, from January 13, 2020, to March 12, 2020, including one week of winter break at which no classes took place. The collected data from the APs of the entire eight floors were aggregated into hourly counts and were flattened into one dataset for the building-level analysis.

4.2. Zone-level data collection and preparation

Four zones with two different space types, i.e. classroom and office spaces, were selected for investigating early and late arrival/departure times. The number of the APs assigned to these zones, as well as the space type that they support are presented in Table 3. Considering the position of these APs in the selected zones, and their coverage which is about 10–12 m [74], these APs reported the WiFi connection counts that were within the desired zones with minimum error. The collected data from these APs were aggregated into 30-min intervals.

5. Results

Descriptive data analytics on the collected dataset are presented at the beginning of this section, followed by the case study results according to this paper's three main objectives.

5.1. Descriptive analytics

Fig. 2 shows the WiFi connection count data, over the entire eight floors of the building, within the study period. The time-series contained almost nine weeks with similar weekly patterns, except for week no. 7, which coincided with the university's winter break. The total number of WiFi connection counts during weekdays of this week was significantly lower than other weekdays, therefore it was considered an outlier and was removed from the analysis.

As expected, similar patterns were observed among weekdays, which were different from the weekends. This difference suggested developing separate models for weekdays and weekends. Despite variations in the patterns of Mondays through Fridays, based on the results of the Kruskal-Wallis H test, no statistically significant differences were observed among them, indicating that it is acceptable to group weekdays as one dataset for the next step. Moreover, Saturdays and Sundays can be grouped, except for week no. 2, in which some special events may have

Table 2
Specifications of the first eight floors of the case study building.

	Classroom/Auditorium (m^2)	Study/Meeting room (m^2)	Corridor/Lobby (m^2)	Service (m^2)	Office (m^2)	Number of APs
Sub-basement 2	1050	80	870	450	0	14
Sub-basement 1	800	100	1100	400	0	10
Ground floor	600	0	1600	250	0	14
Floor 2	1100	260	840	200	0	11
Floor 3	1000	260	940	200	0	12
Floor 4	150	50	920	200	680	8
Floor 5	700	170	850	180	100	9
Floor 6	380	160	940	180	340	11
Total	5780	1080	8060	2060	1120	89

Table 3
Description of selected APs and corresponding spaces.

Zone name	Space type	Number of APs in each zone
Class1	Classroom	2
Class2	Classroom	2
Office1	Office	1
Office2	Office	1

taken place. Descriptive statistics of weekdays and weekends datasets are presented in Table 4.

The range, quantiles, median, and mean of the building's peak WiFi connection counts for each day of the week during the experiment are plotted in Fig. 3. Three distinct levels of peak values were observed, which were correlated with the days of the week as follows. (i) Monday through Thursday; (ii) Friday; and (ii) Saturday, Sunday; showing an average of about 2455, 1650, and 610 WiFi connections, respectively.

5.2. *Occupancy pattern prediction*

In order to identify the governing behavioral regimes, representative day types were extracted using *k*-means clustering. The range of values tested for *k* was limited to 3–10 to avoid the relatively small or large number of clusters. On the one hand, a high value for *k* may lead to over-fitting and might not help to improve the accuracy of prediction. On the other hand, a low value might overshadow some frequent patterns. Moreover, a cluster with a low number of items is not acceptable as a representative of frequently repeated daily occupancy patterns. According to the plot in Fig. 4, *k* equals both 3 and 5 showed low DBI values. However, setting *k* equal to 5 resulted in having one cluster with only two data-points, which was not acceptable. Therefore, *k* equal to 3 was chosen as the optimal number of clusters for k-means clustering.

Fig. 5 shows the three clusters of daily occupancy patterns, along with their centroids. In all three patterns, the counts start to increase at 07:00 and continue with a sharp growth between 08:00 and 10:00, with a rapid decrease after 19:00. The major difference between the three clusters is the average level of peak WiFi connection counts, and the time of its occurrence, which can be summarized as follows.

- Pattern 1 showed an average peak value of about 1,700, by around 11:00.
- Pattern 2 showed an average peak value of about 2,200, by around 16:00.
- Pattern 3 showed an average peak value of about 2,700, by around 16:00.

The membership of days of the week to each cluster is shown in Fig. 6. Pattern 1, which represented the lowest average of WiFi connection counts with a distinct shape, was the only fully homogeneous cluster, entirely allocated to Fridays. This can be associated with Fridays' characteristics; not only fewer classes are usually scheduled during Friday evenings, but it is also less common to organize events (such as presentations, talks, or workshops) on Friday evenings. The other two patterns happened on the other four days, with an almost similar rate on Mondays, Wednesdays, and Thursdays. However, the dominant pattern for Tuesdays was pattern 3, which represented the highest average of WiFi connection counts, suggesting that more classes and events happen during Tuesdays in this building.

These findings, including (i) the significant difference between weekdays' and weekends' occupancy; and (ii) having Fridays as a separate cluster, suggested dividing the complete dataset into three independent subsets, i.e. (i) Weekdays (Monday through Thursday); (ii) Fridays; and (iii) Weekends. Three separate prediction models were then developed for each subset, to improve the performance of week-ahead occupancy pattern prediction. In these three models, the hours around peak times had higher coefficients which was expected, given that the WiFi connection count was of a considerably larger order of magnitude

Table 4
Descriptive statistics of weekdays and weekends datasets.

	Weekdays	Weekends
Min	7	3
Max	3076	823
Median	862	122
Mode	16	8
Mean	1011	195
Standard deviation	974	211

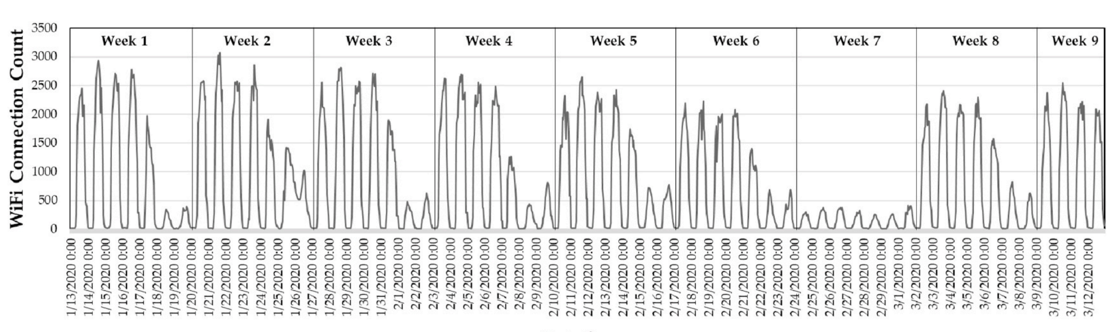


Fig. 2. Hourly WiFi connection counts in eight floors of the case study building, reported from January 13, 2020, to March 12, 2020.

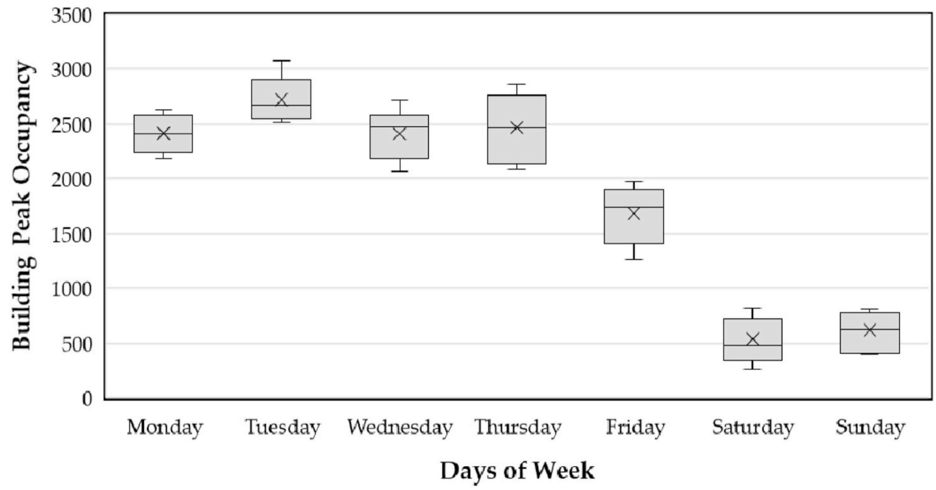


Fig. 3. Building peak WiFi connection counts (occupancy) for different days of the week within the study period.

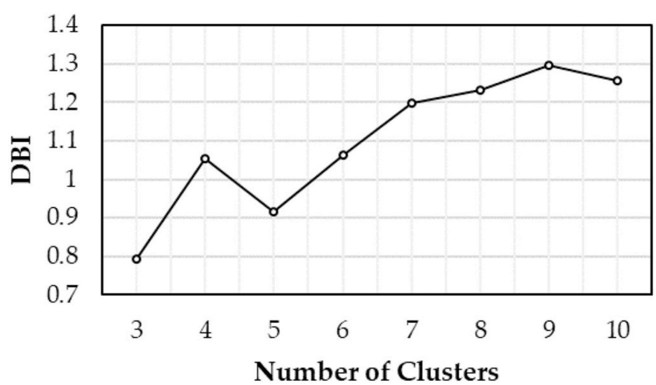


Fig. 4. DBI graph of different values of k in k-means clustering of daily occupancy patterns.

during these hours.

The performance of the three prediction models was evaluated through time-based n-fold cross-validation employed based on the available data of the three subsets, with n equals 8, 7, and 7 for 'Weekdays', 'Fridays', and 'Weekends', respectively. As an example, Fig. 7 shows the combined results of all three models, for one fold of the n-fold validation. In this fold, week no. 8 was considered as the test dataset, while all three models were trained with the rest of the weeks.

The performance of each of the three prediction models in predicting week-ahead occupancy patterns was evaluated by averaging the score of $R^2{}_D$, RMSE, and MAPE in all folds of cross-validation. The results are reported in Table 5. The statistical characteristics of deviance residuals show a mean near to zero in all three prediction models, indicating that the models have not been biased. Moreover, since the dependent variable ranged from 3 to around 3000 counts, the resulted RMSE (ranging between 110 and 178) is deemed acceptable. According to the table, 'Weekdays' and 'Fridays' models achieved significantly higher $R^2{}_D$ values (0.98 and 0.97, respectively), compared to the 'Weekends' model (with an $R^2{}_D$ of 0.81). Furthermore, the results of MAPEs showed a high level of error in the 'Weekends' model (83.99%) compared to 'Weekdays' and 'Fridays' models (with 42.79% and 38.33% error, respectively). The lower performance of the 'Weekends' model can be justified due to the considerably higher frequency of temporary events, such as short-term workshops and special meetings, during the weekends. Such events make the occupancy patterns far less predictable, while the university schedule during weekdays can help to regularize the occupancy patterns to some extent. On the other hand, comparing the MAPEs of working time and nighttime prediction, revealed that due to the small number of WiFi connections during nighttime as well as the significant discrepancy between working time and nighttime WiFi connection counts, the nighttime prediction has been the main contributor to the overall error. Since the present study was mainly aiming at the prediction of occupancy patterns during working time, the MAPEs calculated for this period will be of the main concern. Those errors are of a

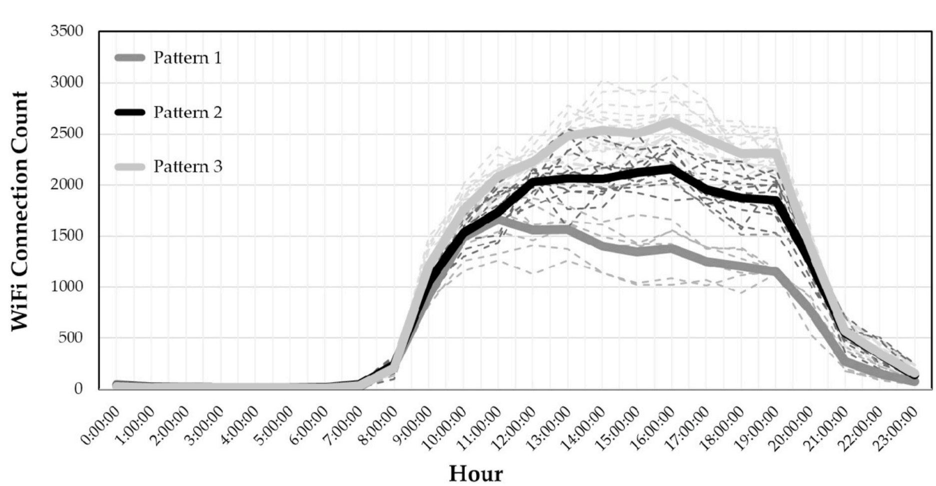


Fig. 5. Three clusters of daily occupancy patterns of weekdays with their centroids.

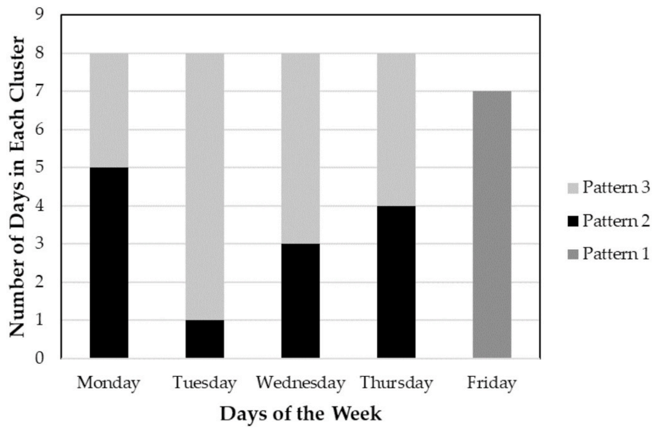


Fig. 6. Membership of days of the week to each cluster.

considerably lower magnitude, and range between 12.49% and 16.61%. Fig. 8 presents the breakdown of errors for days of the week and hours of the day. A higher level of errors was observed for nighttime, which was resulted from the lower number of connection counts during this time and likely more sporadic occupancy outside the school's operation hours. During the working time when occupants use the space more actively, the level of error was lower and more stable (around 20%). Comparing the prediction errors during different days of the week also showed lower performance of the 'Weekends' model which was mainly due to the lower number of counts as well as irregular events such as workshops.

5.3. Peak occupancy prediction

Multiple prediction models were developed and tested to predict week-ahead peak WiFi connection counts which is a proxy for peak occupancy. This process relied on using a range of training sets, starting from one week of training data up to seven weeks of training. As plotted in Fig. 9, the resulting average R^2_D value from cross-validation for each certain number of weeks of training was slightly improved by increasing the number of training weeks from one to two, where it reached its highest value (of 0.94). Extending the training window to beyond two weeks caused the R^2_D to decline. This can be due to introducing more variation in counts to the model which causes more noise. Therefore, two weeks of training prior to the week of target was determined as the optimum number of weeks required for peak occupancy prediction in this building for the winter semester.

To further investigate the occurrence of peak building occupancy, the cumulative relative frequency distributions of the time at which the WiFi connection counts reach their daily maximum were plotted for each day of the week (see Fig. 10). Based on the plot, the most apparent finding was the significant difference between the time at which peak occupancy occurs on different days. In 90% of the time, peak occupancy occurs on or before 16:00 on Mondays, Tuesdays, and Thursdays; while Fridays and Wednesdays experience peak occupancy on or before 13:00 and 18:00, respectively. These variations in the probable peak occupancy time were also observed on weekends, which can help the building operators with adjusting ventilation schedules and ensuring maximum outdoor air intake coincides with the timing of peak occupancy on different days.

5.4. Arrival and departure times analysis

A similar approach was followed to identify arrival and departure times at the four selected zones. The cumulative relative frequency distributions in Fig. 11 showed considerable variation in arrival/departure times of zones, regardless of their space types. For example, the earliest arrival time observed in Office 1 was 5:30 while it was 7:00 in Office 2. Based on these observations, the heating setback schedule in Office 2 could be extended relative to Office 1.

In 99% of the time, it was observed that the latest arrival times for Offices 1 and 2 were 7:30 and 13:00, respectively, which were also different for classrooms. Building operators could use this information to schedule earlier temperature set-backs in these rooms if no occupancy is detected after the observed latest arrival time, which would avoid heating these empty rooms for several hours when standard operational schedules are used.

The same analysis was extended to departure times in the selected

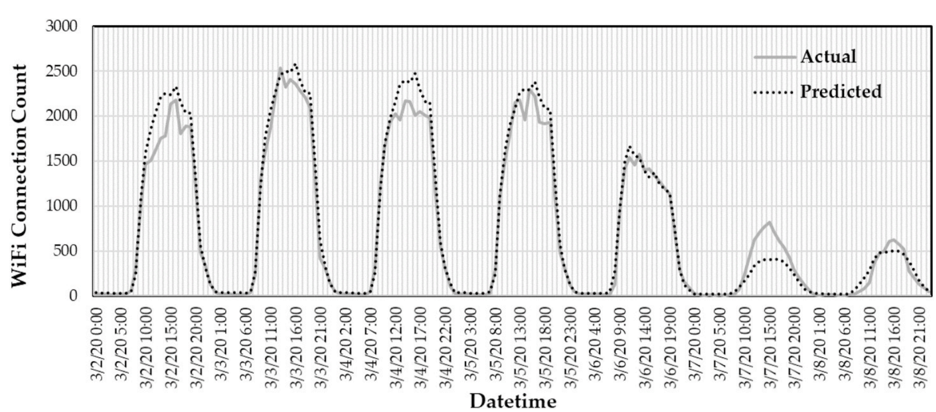


Fig. 7. The combination of all three prediction models' results for predicting test dataset, week no.8, versus actual values, starting from Monday, ending on Sunday.

Table 5
Performance of three prediction models developed on data subsets (i) Weekdays, (ii) Fridays, and (iii) Weekends.

Prediction model	R^2_D	RMSE	MAPE			Deviance Residual[a]
			All predictions	*working time prediction*	*nighttime prediction*	
Weekdays	0.98	178	42.79%	12.49%	85.22%	−0.104±4.077
Fridays	0.97	143	38.33%	16.61%	68.74%	- 0.106±3.956
Weekends	0.81	110	83.99%	60.50%	116.87%	- 0.289±5.233

[a] Mean±Standard deviation.

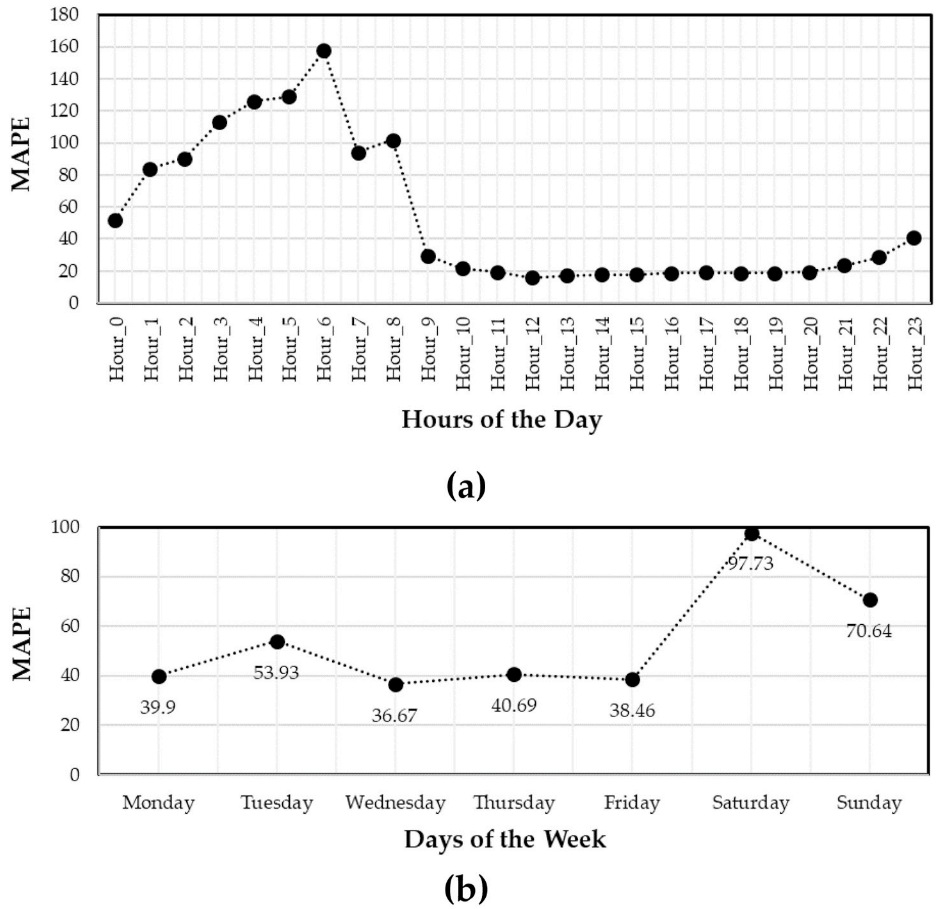


Fig. 8. The breakdown of errors over (a) hours of the day, (b) days of the week.

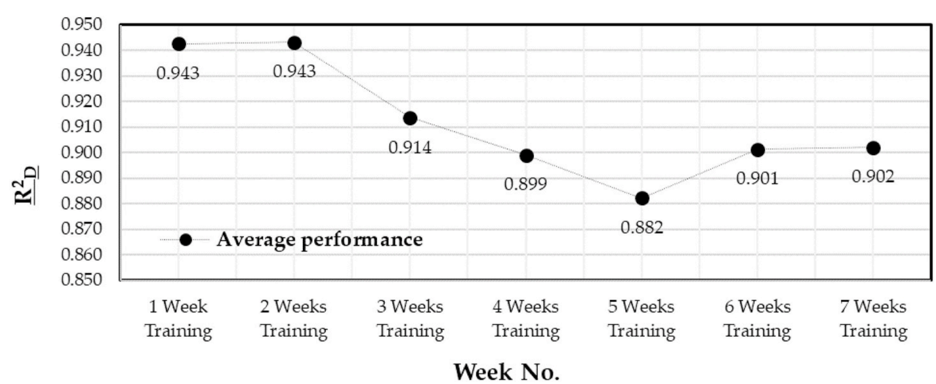


Fig. 9. The average performance of peak occupancy prediction with prediction models trained on a different number of weeks.

four rooms which also showed large variations. In 99% of the time, it was observed that the latest departure times were 22:00 and 1:00 in Offices 1 and 2, respectively. This information could also be used to adjust the timing of initiating temperature setbacks in different rooms. In many cases, building operators resort to keeping the temperature setpoint for 24 h if building occupancy is irregular to avoid occupant complaints. However, this practice can be avoided by gaining insights into the occupancy patterns of specific rooms and zones using the proposed approach.

It must be noted that the results of this analysis are only meant to show a proof of concept of the proposed methodology and the observed occupancy pattern variations in different rooms. However, data collection would have to be extended beyond the duration of this case-study for at least one year to capture seasonal variations and ensure results accuracy, before making any changes to the control sequences of building systems.

6. Discussion

The proposed framework aims to identify practical occupancy indicators using WiFi connection count data. Applying it to the case study provided a proof-of-concept and indicated that occupancy indicators derived from the analytics can provide a simple, yet more accurate understanding of the building's occupancy schedule, compared to standard schedules. The wide differences between occupancy attributes, such as the peak time, arrival, and departure times, even in spaces of the

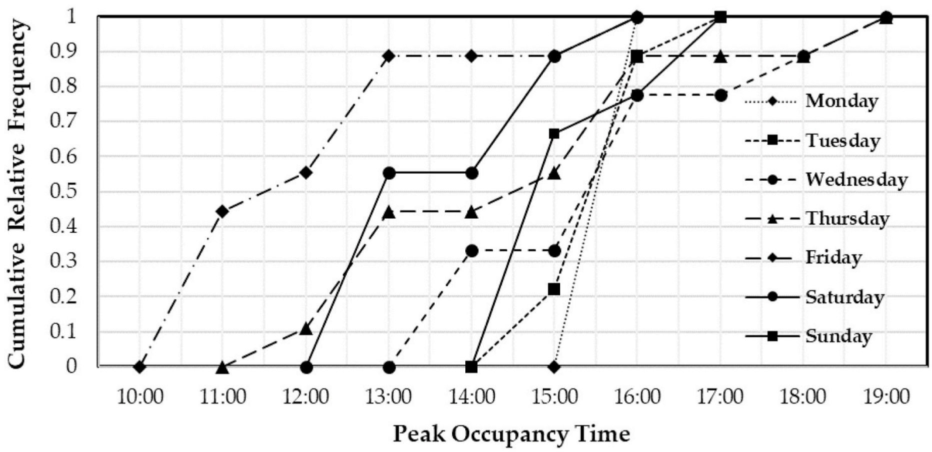
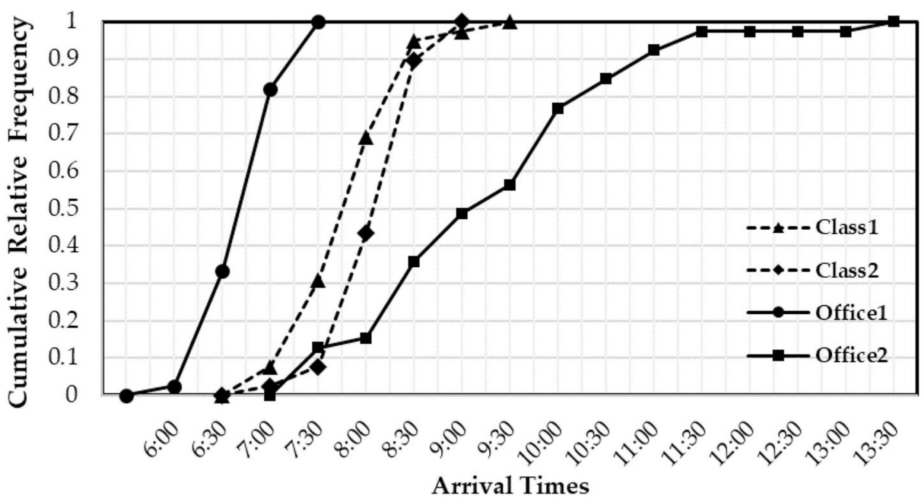


Fig. 10. Cumulative relative frequency distribution of building peak occupancy time.



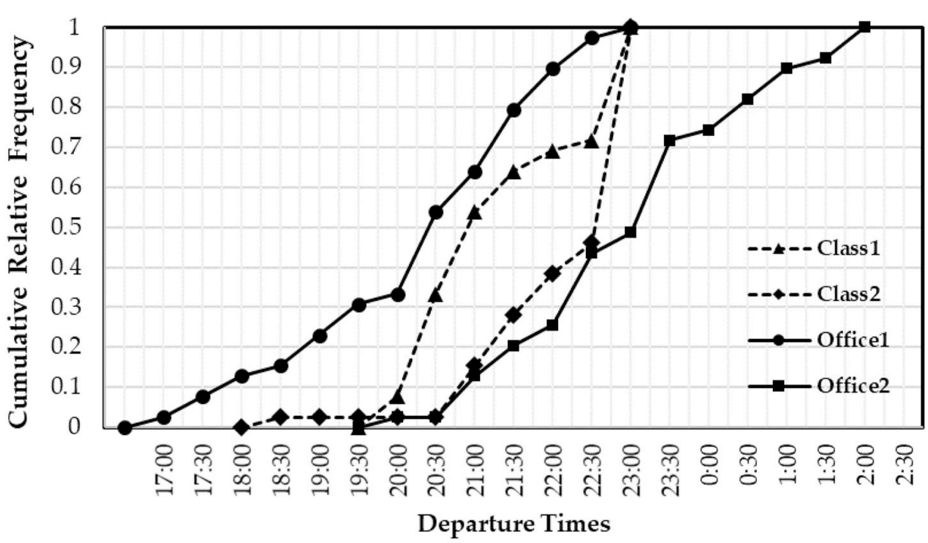


Fig. 11. Cumulative relative frequency distributions of (a) arrival times of occupants in four zones (b) departure times of occupants in four zones.

same type, are clear indicators of the need for building-specific sequences of operations.

Although the case-study demonstrated that WiFi can be a low-cost approach to analyze building-specific occupancy patterns with an acceptable level of accuracy, it must be noted that the results are only meant to show a proof of concept of the proposed methodology. Data collection would have to be extended beyond the duration of this case-study for at least one year to capture seasonal variations and ensure results accuracy, before making any changes to the control sequences of building systems. Key insights drawn from the case-study are summarized as follows.

- First, clustering daily WiFi connection count profiles extracted representative day types of occupancy patterns that were significantly different from standard operational schedules. Although the presented results may not identify or predict the exact occupancy counts (due to the duality of WiFi traffic and occupant numbers), they clearly show the deviation between the standard and actual schedules, especially in terms of peak occupancy, arrival and departure times.
- Second, the high accuracy of the developed prediction models showed that Poisson regression, which was rarely used in this field, improved model performance while still relying only on simple input data (i.e., aggregate WiFi connection counts). This algorithm was capable of learning and predicting the variations in WiFi counts and obtained higher prediction accuracy than the typically used MLR models.
- Third, although 'the hour of the day' was considered as a feature for developing prediction models, the high discrepancy between WiFi connection counts of working time and nighttime necessitated introducing a new Boolean feature to the model to represent these two states. Using this approach in developing the prediction model, considerably improved the accuracy of prediction, compared to prediction models developed for working time and nighttime separately.
- Fourth, the comparison between training a single prediction model to cover all days of the week and having separate prediction models for each subset derived from day types (e.g., weekends/weekdays) showed that although the latter slightly outperforms in predicting typical weeks; the former approach can better capture irregularities of different days in some weeks. Hence, using a single model can be helpful to predict occupancy in atypical weeks (such as exam times in academic buildings or holidays in retail and commercial buildings) which results from being trained by more diverse daily profiles.
- Fifth, a notable variation was demonstrated in the peak occupancy level, as well as the timing of its occurrence, for different days of the week. This can significantly influence the outdoor air ventilation demand on different days.
- Sixth, the study of arrival/departure times showed that the diversity of zone-level occupancy patterns can be considerable, which may result in inefficient operation of building systems in unoccupied rooms. Analyzing this diversity can be practical in adjusting zone-level sequences of operation with a high level of confidence, particularly in buildings with more variable occupancy patterns.
- Finally, unlike most of the technologies proposed for occupancy counting, simple data of aggregate WiFi connection counts provided the opportunity of extracting high-level occupancy indicators, such as building occupancy patterns and building peak occupancy time. However, analysis of WiFi data can also result in more detailed insights regarding occupants' behavioral patterns, which requires access to individuals' identifier attributes such as the MAC address.

7. Conclusions

Although several approaches have been proposed to adapt HVAC systems operation to occupancy variations, logistical, cost and integration challenges remain a key issue. The proposed framework addresses this gap by analyzing WiFi traffic data to identify key information about buildings' unique occupancy patterns, which can then be used to adjust sequences of operations. The paper contributes to the literature by (i) offering prediction models that predict occupancy patterns in a longer horizon (i.e. week-ahead) and larger scale (i.e. building-level) with a higher level of accuracy; and (ii) identifying and capturing dynamisms of peak occupancy occurrence as well as arrival and departure times at the room-level. The results of implementing the framework in a case study showed that Poisson regression can predict occupancy patterns with a higher level of accuracy, outperforming MLR. Furthermore, introducing more features to differentiate between daytime and nighttime WiFi connection counts notably improved the performance of prediction models. The case study results also demonstrated the potential for extracting practical occupancy indicators to adjust zone-level sequences of operations based on typical arrival and departure times, which varied significantly even between rooms of similar functions.

Despite the benefits of the proposed approach, some limitations should also be acknowledged and addressed in future work. Although several previous studies showed a strong statistically significant correlation between WiFi connection counts and actual occupancy counts, the proposed models should be calibrated with ground-truth data to account for stationary devices and occupants carrying more or less than one WiFi-connected device. The importance of having ground-truth data increases when the proposed methodology for occupancy pattern prediction is applied at a smaller scale such as room-level or zone-level instead of building-level. At these levels, it is more challenging to recognize the number of connections attributed to each zone without having access to the details of each connection, such as the RSSI. Furthermore, the data collection period should be extended to cover the impact of seasonality on occupancy variations before adjusting building operational sequences. The transferability of the developed models to other buildings with different space types, as well as their scalability to other zones should also be investigated.

Future work will focus on investigating the transferability of the prediction models to other buildings and using different techniques to calibrate the developed models with available ground-truth data on actual occupancy counts. Furthermore, quantifying energy savings after adjusting building operations based on the identified occupancy metrics will be investigated. These findings will further highlight the benefits of using the proposed approach to provide actionable information to modify sequences of operation and reduce energy consumption in large commercial and institutional buildings.

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.

Acknowledgment

This research is supported by the research funding provided by the Natural Sciences and Engineering Research Council (NSERC) of Canada. The authors acknowledge the support of the Instructional and Information Technology Services (IITS) team of Concordia University by providing the WiFi connection count data; and the Security and Facilities Management departments of Concordia University by providing the drawings of the building.

References

[1] S. D'Oca, T. Hong, J. Langevin, The human dimensions of energy use in buildings: a review, Renew. Sustain. Energy Rev. 81 (2018) 731–742, https://doi.org/10.1016/j.rser.2017.08.019.

[2] L. Pérez-Lombard, J. Ortiz, C. Pout, A review on buildings energy consumption information, Energy Build. 40 (2008) 394–398, https://doi.org/10.1016/j.enbuild.2007.03.007.

[3] M. Trčka, J.L.M. Hensen, Overview of HVAC system simulation, Autom. ConStruct. 19 (2010) 93–99, https://doi.org/10.1016/j.autcon.2009.11.019.

[4] X. Dai, J. Liu, X. Zhang, A review of studies applying machine learning models to predict occupancy and window-opening behaviours in smart buildings, Energy Build. 223 (2020) 110159, https://doi.org/10.1016/j.enbuild.2020.110159.

[5] M.M. Ouf, W. O'Brien, B. Gunay, On quantifying building performance adaptability to variable occupancy, Build. Environ. 155 (2019) 257–267, https://doi.org/10.1016/j.buildenv.2019.03.048.
[6] J. Yang, M. Santamouris, S.E. Lee, Review of occupancy sensing systems and occupancy modeling methodologies for the application in institutional buildings, Energy Build. 121 (2016) 344–349, https://doi.org/10.1016/j.enbuild.2015.12.019.
[7] M.S. Gul, S. Patidar, Understanding the energy consumption and occupancy of a multi-purpose academic building, Energy Build. 87 (2015) 155–165, https://doi.org/10.1016/j.enbuild.2014.11.027.
[8] B.W. Hobson, H.B. Gunay, A. Ashouri, G.R. Newsham, Clustering and motif identification for occupancy-centric control of an air handling unit, Energy Build. 223 (2020) 110179, https://doi.org/10.1016/j.enbuild.2020.110179.
[9] X. Liang, T. Hong, G.Q. Shen, Occupancy data analytics and prediction: a case study, Build. Environ. 102 (2016) 179–192, https://doi.org/10.1016/j.buildenv.2016.03.027.
[10] Y. Ding, W. Chen, S. Wei, F. Yang, An occupancy prediction model for campus buildings based on the diversity of occupancy patterns, Sustainable Cities and Society 64 (2021) 102533, https://doi.org/10.1016/j.scs.2020.102533.
[11] K. Sun, Q. Zhao, J. Zou, A review of building occupancy measurement systems, Energy Build. 216 (2020) 109965, https://doi.org/10.1016/j.enbuild.2020.109965.
[12] Z. Chen, C. Jiang, L. Xie, Building occupancy estimation and detection: a review, Energy Build. 169 (2018) 260–270, https://doi.org/10.1016/j.enbuild.2018.03.084.
[13] M.M. Ouf, M.H. Issa, A. Azzouz, A.-M. Sadick, Effectiveness of using WiFi technologies to detect and predict building occupancy, Sustain. Build 2 (2017) 7, https://doi.org/10.1051/sbuild/2017005.
[14] I.P. Mohottige, T. Sutjarittham, N. Raju, H.H. Gharakheili, V. Sivaraman, Role of campus WiFi infrastructure for occupancy monitoring in a large university, in: IEEE International Conference on Information and Automation for Sustainability (ICIAfS), IEEE, Colombo, Sri Lanka, 2018, pp. 1–5, https://doi.org/10.1109/ICIAFS.2018.8913341.
[15] A. Ashouri, G.R. Newsham, Z. Shi, H.B. Gunay, Day-ahead prediction of building occupancy using WiFi signals, in: 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE), IEEE, Vancouver, , BC, Canada, 2019, pp. 1237–1242, https://doi.org/10.1109/COASE.2019.8843224.
[16] Y. Wang, L. Shao, Understanding occupancy and user behaviour through Wi-Fi-based indoor positioning, Build. Res. Inf. 46 (2018) 725–737, https://doi.org/10.1080/09613218.2018.1378498.
[17] Z. Wang, T. Hong, M.A. Piette, M. Pritoni, Inferring occupant counts from Wi-Fi data in buildings through machine learning, Build. Environ. 158 (2019) 281–294, https://doi.org/10.1016/j.buildenv.2019.05.015.
[18] K.C. Jagadeesh Simma, A. Mammoli, S.M. Bogus, Real-time occupancy estimation using WiFi network to optimize HVAC operation, Procedia Computer Science 155 (2019) 495–502, https://doi.org/10.1016/j.procs.2019.08.069.
[19] 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, https://doi.org/10.1016/j.buildenv.2017.08.003.
[20] G.H. Apostolo, F. Bernardini, L.C.S. Magalhaes, D.C. Muchaluat-Saade, A unified methodology to predict wi-fi network usage in smart buildings, IEEE Access 9 (2021) 11455–11469, https://doi.org/10.1109/ACCESS.2020.3048891.
[21] 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, https://doi.org/10.1016/j.buildenv.2019.05.032.
[22] W. Wang, J. Chen, T. Hong, Occupancy prediction through machine learning and data fusion of environmental sensing and Wi-Fi sensing in buildings, Autom. ConStruct. 94 (2018) 233–243, https://doi.org/10.1016/j.autcon.2018.07.007.
[23] K. Christensen, R. Melfi, B. Nordman, B. Rosenblum, R. Viera, Using existing network infrastructure to estimate building occupancy and control plugged-in devices in user workspaces, Int. J. Commun. Network. Distr. Syst. 12 (2014) 4, https://doi.org/10.1504/IJCNDS.2014.057985.
[24] Thiago Teixeira, Gershon Dublon, Andreas Savvides, A survey of human-sensing: methods for detecting presence, Count, Location, Track, and Identity (2010) 5.
[25] M. Esrafilian-Najafabadi, F. Haghighat, Occupancy-based HVAC control systems in buildings: a state-of-the-art review, Build. Environ. 197 (2021) 107810, https://doi.org/10.1016/j.buildenv.2021.107810.
[26] C. Jiang, M.K. Masood, Y.C. Soh, H. Li, Indoor occupancy estimation from carbon dioxide concentration, Energy Build. 131 (2016) 132–141, https://doi.org/10.1016/j.enbuild.2016.09.002.
[27] M. Gruber, A. Trüschel, J.-O. Dalenbäck, CO 2 sensors for occupancy estimations: potential in building automation applications, Energy Build. 84 (2014) 548–556, https://doi.org/10.1016/j.enbuild.2014.09.002.
[28] N. Nassif, A robust CO2-based demand-controlled ventilation control strategy for multi-zone HVAC systems, Energy Build. 45 (2012) 72–81, https://doi.org/10.1016/j.enbuild.2011.10.018.
[29] I.B. Arief-Ang, M. Hamilton, F.D. Salim, A scalable room occupancy prediction with transferable time series decomposition of CO_2 sensor data, ACM Trans. Sens. Netw. 14 (2018) 1–28, https://doi.org/10.1145/3217214.
[30] G. Ansanay-Alex, Estimating Occupancy Using Indoor Carbon Dioxide Concen-Trations Only in an Office Building: A Method and Qualitative Assessment, 2013, pp. 1–8.
[31] C. Jiang, Z. Chen, R. Su, M.K. Masood, Y.C. Soh, Bayesian filtering for building occupancy estimation from carbon dioxide concentration, Energy Build. 206 (2020) 109566, https://doi.org/10.1016/j.enbuild.2019.109566.
[32] A. Franco, F. Leccese, Measurement of CO2 concentration for occupancy estimation in educational buildings with energy efficiency purposes, Journal of Building Engineering 32 (2020) 101714, https://doi.org/10.1016/j.jobe.2020.101714.
[33] M.K. Masood, Y. Chai Soh, V.W.-C. Chang, Real-time occupancy estimation using environmental parameters, in: International Joint Conference on Neural Networks (IJCNN), IEEE, Killarney, Ireland, 2015, pp. 1–8, https://doi.org/10.1109/IJCNN.2015.7280781.
[34] K. Weekly, M. Jin, H. Zou, C. Hsu, C. Soyza, A. Bayen, C. Spanos, Building-in-Briefcase: a rapidly-deployable environmental sensor suite for the smart building, Sensors 18 (2018) 1381, https://doi.org/10.3390/s18051381.
[35] Z. Chen, Q. Zhu, M.K. Masood, Y.C. Soh, Environmental sensors-based occupancy estimation in buildings via IHMM-MLR, IEEE Trans. Ind. Inf. 13 (2017) 2184–2193, https://doi.org/10.1109/TII.2017.2668444.
[36] A. Szczurek, M. Maciejewska, T. Pietrucha, Occupancy determination based on time series of CO2 concentration, temperature and relative humidity, Energy Build. 147 (2017) 142–154, https://doi.org/10.1016/j.enbuild.2017.04.080.
[37] M. Jin, N. Bekiaris-Liberis, K. Weekly, C.J. Spanos, A.M. Bayen, Occupancy detection via environmental sensing, IEEE Trans. Autom. Sci. Eng. 15 (2018) 443–455, https://doi.org/10.1109/TASE.2016.2619720.
[38] M.K. Masood, Y.C. Soh, C. Jiang, Occupancy estimation from environmental parameters using wrapper and hybrid feature selection, Appl. Soft Comput. 60 (2017) 482–494, https://doi.org/10.1016/j.asoc.2017.07.003.
[39] 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, https://doi.org/10.1088/0957-0233/24/7/074023.
[40] F. Wang, Q. Feng, Z. Chen, Q. Zhao, Z. Cheng, J. Zou, Y. Zhang, J. Mai, Y. Li, H. Reeve, Predictive control of indoor environment using occupant number detected by video data and CO 2 concentration, Energy Build. 145 (2017) 155–162, https://doi.org/10.1016/j.enbuild.2017.04.014.
[41] Y. Meng, T. Li, G. Liu, S. Xu, T. Ji, Real-time dynamic estimation of occupancy load and an air-conditioning predictive control method based on image information fusion, Build. Environ. 173 (2020) 106741, https://doi.org/10.1016/j.buildenv.2020.106741.
[42] V. Chidurala, X. Li, Occupancy estimation using thermal imaging sensors and machine learning algorithms, IEEE Sensor. J. 21 (2021) 8627–8638, https://doi.org/10.1109/JSEN.2021.3049311.
[43] E. Seghezzi, M. Locatelli, L. Pellegrini, G. Pattini, G.M. Di Giuda, L.C. Tagliabue, G. Boella, Towards an occupancy-oriented digital twin for facility management: test campaign and sensors assessment, Appl. Sci. 11 (2021) 3108, https://doi.org/10.3390/app11073108.
[44] S. Petersen, T.H. Pedersen, K.U. Nielsen, M.D. Knudsen, Establishing an image-based ground truth for validation of sensor data-based room occupancy detection, Energy Build. 130 (2016) 787–793, https://doi.org/10.1016/j.enbuild.2016.09.009.
[45] N. Li, G. Calis, B. Becerik-Gerber, Measuring and monitoring occupancy with an RFID based system for demand-driven HVAC operations, Autom. ConStruct. 24 (2012) 89–99, https://doi.org/10.1016/j.autcon.2012.02.013.
[46] H.-T. Wang, Q.-S. Jia, C. Song, R. Yuan, X. Guan, Building occupant level estimation based on heterogeneous information fusion, Inf. Sci. 272 (2014) 145–157, https://doi.org/10.1016/j.ins.2014.02.080.
[47] E. Longo, A.E.C. Redondi, M. Cesana, Accurate occupancy estimation with WiFi and bluetooth/BLE packet capture, Comput. Network. 163 (2019) 106876, https://doi.org/10.1016/j.comnet.2019.106876.
[48] Z.D. Tekler, R. Low, B. Gunay, R.K. Andersen, L. Blessing, A scalable Bluetooth Low Energy approach to identify occupancy patterns and profiles in office spaces, Build. Environ. 171 (2020) 106681, https://doi.org/10.1016/j.buildenv.2020.106681.
[49] L. Wu, Y. Wang, A low-power electric-mechanical driving approach for true occupancy detection using a shuttered passive infrared sensor, IEEE Sensor. J. 19 (2019) 47–57, https://doi.org/10.1109/JSEN.2018.2875659.
[50] S.S. Shetty, H.D. Chinh, M. Gupta, S.K. Panda, User presence estimation in multi-occupancy rooms using plug-load meters and PIR sensors, in: GLOBECOM 2017 - 2017 IEEE Global Communications Conference, IEEE, Singapore, 2017, pp. 1–6, https://doi.org/10.1109/GLOCOM.2017.8255036.
[51] F. Wahl, M. Milenkovic, O. Amft, A distributed PIR-based approach for estimating people count in office environments, in: 2012 IEEE 15th International Conference on Computational Science and Engineering, IEEE, Paphos, Cyprus, 2012, pp. 640–647, https://doi.org/10.1109/ICCSE.2012.92.
[52] 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, Heidelberg Germany, 2016, pp. 1016–1027, https://doi.org/10.1145/2971648.2971746.
[53] L. Rueda, K. Agbossou, A. Cardenas, N. Henao, S. Kelouwani, A comprehensive review of approaches to building occupancy detection, Build. Environ. 180 (2020) 106966, https://doi.org/10.1016/j.buildenv.2020.106966.
[54] M.S. Zuraimi, A. Pantazaras, K.A. Chaturvedi, J.J. Yang, K.W. Tham, S.E. Lee, Predicting occupancy counts using physical and statistical Co2-based modeling methodologies, Build. Environ. 123 (2017) 517–528, https://doi.org/10.1016/j.buildenv.2017.07.027.
[55] S.H. Ryu, H.J. Moon, Development of an occupancy prediction model using indoor environmental data based on machine learning techniques, Build. Environ. 107 (2016) 1–9, https://doi.org/10.1016/j.buildenv.2016.06.039.
[56] Z. Li, B. Dong, Short term predictions of occupancy in commercial buildings—performance analysis for stochastic models and machine learning approaches, Energy Build. 158 (2018) 268–281, https://doi.org/10.1016/j.enbuild.2017.09.052.

[57] B. Huchuk, S. Sanner, W. O'Brien, Comparison of machine learning models for occupancy prediction in residential buildings using connected thermostat data, Build. Environ. 160 (2019) 106177, https://doi.org/10.1016/j.buildenv.2019.106177.

[58] W. Wang, J. Wang, J. Chen, G. Huang, X. Guo, Multi-zone outdoor air coordination through Wi-Fi probe-based occupancy sensing, Energy Build. 159 (2018) 495–507, https://doi.org/10.1016/j.enbuild.2017.11.041.

[59] 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, https://doi.org/10.1016/j.apenergy.2018.11.079.

[60] B. Balaji, J. Xu, A. Nwokafor, R. Gupta, Y. Agarwal, Sentinel: occupancy based HVAC actuation using existing WiFi infrastructure within commercial buildings. Proceedings of the 11th ACM Conference on Embedded Networked Sensor Systems - SenSys '13, ACM Press, Roma, Italy, 2013, pp. 1–14, https://doi.org/10.1145/2517351.2517370.

[61] M. Oppermann, T. Munzner, Ocupado: visualizing location-based counts over time across buildings, Comput. Graph. Forum 39 (2020) 127–138, https://doi.org/10.1111/cgf.13968.

[62] I. Pasquel Mohottige, H.H. Gharakheili, A. Vishwanath, S.S. Kanhere, V. Sivaraman, Evaluating emergency evacuation events using building WiFi data, in: 2020 IEEE/ACM Fifth International Conference on Internet-Of-Things Design and Implementation (IoTDI), IEEE, Sydney, Australia, 2020, pp. 116–127, https://doi.org/10.1109/IoTDI49375.2020.00018.

[63] A. Chong, G. Augenbroe, D. Yan, Occupancy data at different spatial resolutions: building energy performance and model calibration, Appl. Energy 286 (2021) 116492, https://doi.org/10.1016/j.apenergy.2021.116492.

[64] H. Hou, J. Pawlak, A. Sivakumar, B. Howard, J. Polak, An approach for building occupancy modelling considering the urban context, Build. Environ. 183 (2020) 107126, https://doi.org/10.1016/j.buildenv.2020.107126.

[65] H. Zou, H. Jiang, J. Yang, L. Xie, C. Spanos, Non-intrusive occupancy sensing in commercial buildings, Energy Build. 154 (2017) 633–643, https://doi.org/10.1016/j.enbuild.2017.08.045.

[66] W. Yoo, H. Kim, M. Shin, Stations-oriented indoor localization (SOIL): a BIM-Based occupancy schedule modeling system, Build. Environ. 168 (2020) 106520, https://doi.org/10.1016/j.buildenv.2019.106520.

[67] A. Ravi, A. Misra, Practical server-side WiFi-based indoor localization: addressing cardinality & outlier challenges for improved occupancy estimation, Ad Hoc Netw. 115 (2021) 102443, https://doi.org/10.1016/j.adhoc.2021.102443.

[68] I. Sobron, J. Del Ser, I. Eizmendi, M. Velez, Device-Free people counting in IoT environments: new insights, results, and open challenges, IEEE Internet Things J 5 (2018) 4396–4408, https://doi.org/10.1109/JIOT.2018.2806990.

[69] S. Di Domenico, M. De Sanctis, E. Cianca, G. Bianchi, A trained-once crowd counting method using differential WiFi Channel State information, in: Proceedings of the 3rd International on Workshop on Physical Analytics - WPA '16, ACM Press, Singapore, Singapore, 2016, pp. 37–42, https://doi.org/10.1145/2935651.2935657.

[70] H. Zou, Y. Zhou, J. Yang, C.J. Spanos, Device-free occupancy detection and crowd counting in smart buildings with WiFi-enabled IoT, Energy Build. 174 (2018) 309–322, https://doi.org/10.1016/j.enbuild.2018.06.040.

[71] M. Azam, M. Blayo, J.-S. Venne, M. Allegue-Martinez, Occupancy estimation using wifi motion detection via supervised machine learning algorithms, in: 2019 IEEE Global Conference on Signal and Information Processing (GlobalSIP), IEEE, Ottawa, ON, Canada, 2019, pp. 1–5, https://doi.org/10.1109/GlobalSIP45357.2019.8969297.

[72] Cisco, Cisco Prime Infrastructure 3.4 User Guide, 2017. http://www.cisco.com.

[73] H. Heinzl, M. Mittlböck, Pseudo R-squared measures for Poisson regression models with over- or underdispersion, Comput. Stat. Data Anal. 44 (2003) 253–271, https://doi.org/10.1016/S0167-9473(03)00062-8.

[74] E. Cisco, Cisco Aironet Series 1700/2700/3700 Access Point Deployment Guide, 2014. http://www.cisco.com.