Proceedings of the Canadian Society for Mechanical Engineering International Congress
Computational Fluid Dynamics Canada International Congress
CSME - CFD-SC2023
May 28-31, 2023, Sherbrooke, QC, Canada

Occupancy detection in a residential building using sensor fusion data and machine learning algorithms

Hamed Heidarifar [1*], Mahdi Shahbakhti [1]
[1] Mechanical Engineering Department, University of Alberta, Edmonton, Canada
* heidarif@ualberta.ca

Abstract**—** This research implemented sensor fusion to predict the occupancy status of a residential building. A room in a residential building located in Edmonton, Alberta was used as the testbed for this study. To predict the occupancy, data including room temperature, relative humidity, CO_2 concentration, and day of the week were collected from March to April 2022. The actual occupancy status of the room was collected using a Passive Infrared (PIR) motion sensor and the data sheets filled by the occupants. The study considered four different machine learning algorithms including K-nearest neighbors (KNN), Gaussian Support Vector Machine (SVM), Artificial Neural Network (ANN), and Decision Tree (DT) to predict the state of occupancy and compared the accuracy of these methods. The results showed KNN method outperforms the other methods by reaching the Geometric Mean (GM) accuracy of 94% for occupancy prediction. In addition, it investigated the sufficiency of temperature and humidity sensors for occupancy detection and studied the importance of using recent data for occupancy prediction in residential buildings.

Keywords-component; Sensor fusion; Occupancy detection; KNN; ANN; DT; SVM

I. INTRODUCTION

Residential and commercial buildings are the largest energy-consuming sector in the world. They are responsible for 40% of all primary energy usage in the US and EU [1]. In Canada, residential and commercial/institutional sectors consume around 30% of the total energy consumption in the country. Among all the energy users in the buildings, space heating and cooling account for 66% and 59% of the energy usage in Canadian residential, and commercial buildings respectively [2, 3].

Despite all the efforts in the building industry to reduce energy consumption and emission production in the building sector, the building industry is still responsible for one-third of the greenhouse gas emissions worldwide [4]. Although new projects such as energy-zero buildings and green buildings have increased energy efficiency, but due to the rapid growth in the building sector, and world population the energy consumption in the building sector is expected to increase globally for the next 30 years by 1.3% [5]. To this end, methods and tools to decrease the energy consumption of buildings are highly needed.

One of the effective methods in a Building Energy Management (BEM) system to reduce the heating/cooling load of a building is detecting occupancy behavior in a building and adjusting Heating, Ventilation, and Air Conditioning (HVAC) operation accordingly [6]. The temperature range based on human comfort in occupied spaces depends on different parameters such as relative humidity, activity level, clothing, etc. The comfortable temperature for an occupied space varies from 19.5 °C to 27 °C based on ASHRAE 55 standard [7]. In the winter time, when a space in a building is unoccupied, the temperature can be lower depending on the minimum safe temperature to avoid freezing in a building and the minimum ventilation requirement to avoid forming mold. Although it is difficult to find a unique thermostat temperature for unoccupied buildings during the cold season, the temperature range between 12 °C to 16 °C is commonly used to avoid any damage to the building and save energy [8, 9]. This can reduce the heating load while the building is not occupied; thus, it reduces the building's energy consumption.

Occupancy-based control systems in buildings can reduce energy consumption significantly. The study in [10] shows the possibility of up to 80% savings in energy consumption by using an occupancy-based feedback control system. Another study reported around 30% energy saving by using an occupancy pattern in a conference room of a commercial building [11]. To this end, effective and low-cost methods for occupancy detection are of high interest for buildings.

Despite numerous studies conducted for commercial/educational buildings to develop occupancy models, residential buildings have not received the same attention in the literature. Fig. 1 compares the studies that use residential and non-residential buildings as their testbed.

The fact that residential buildings consume more energy than commercial and institutional buildings [12] along with the previously stated statistics, determines the need for more research to detect and predict occupancy in residential dwellings.

There are different methods which can be used to detect the occupancy in buildings without human intervention including:

- Video camera
- Passive Infra-Red (PIR) Motion detection sensor
- Radio frequency (RF) sensor
- WLAN, Wi-Fi and Bluetooth usage
- Sensor fusion

Among the above-mentioned methods, video cameras, motion detection sensors, sensor fusion, and Wi-Fi usage provide the highest to the lowest accuracy of occupancy detection respectively [13].

Despite the accuracy of cameras, due to strict privacy policies and costly equipment, using these methods is not feasible all the time, specifically for residential buildings [6, 14, 15]. Devices, such as laptops or smartphones, connected to the Wi-Fi network have been used to determine/count occupancy mostly in commercial buildings [16, 17]; however, for residential buildings, its accuracy will decrease significantly since all devices normally connect to one access point despite being in different zones in the house [16]. PIR accuracy can be used at a low cost; however, sometimes the lag in the sensor response can cause inaccuracy in the result [14]. RF signal technology is another tool for occupancy detection that has been studied in the building industry [18-21], but it requires a large number of reference points and its accuracy drops when reference points are not sufficient [14].

Using sensor fusion can provide accurate occupancy information for the building control system without creating any privacy issues; however, implementing this method usually needs data from multiple sensors such as CO2, temperature, humidity, light, etc. Some of these sensors could be expensive for a residential building, but in recent years some of these sensors are already installed in residential buildings for other reasons e.g., safety or being integrated into other smart devices such as smart appliances.

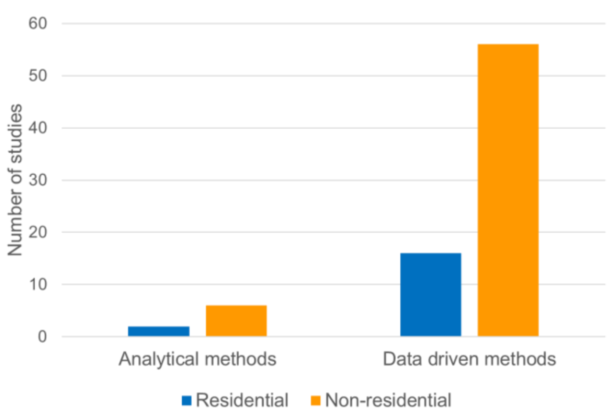


Figure 1: Comparison of the number of studies using occupancy detection methods in residential and non-residential buildings, using the data from [22].

The accuracy of occupancy prediction in the studies that combined sensor fusion and data-driven classification models for occupancy prediction varies between 60% to 98% [22]. A wide range of classification algorithms including Artificial neural network (ANN) [23], Support Vector Machine (SVM) [24], K-Nearest Neighbors (KNN) [24, 25], hidden Markov model (HMM)[6, 24] and Decision Tree [6] have been implemented in the literature for occupancy prediction.

In this study data including CO2, temperature, relative humidity (RH), and day of the week is collected to (1) compare the performance of different classifiers for occupancy prediction, (2) investigate the sufficiency of temperature and humidity sensors for occupancy detection as these sensors are common and affordable even in residential buildings and can easily be installed in different rooms, (3) study the importance of using recent data for occupancy prediction in residential buildings.

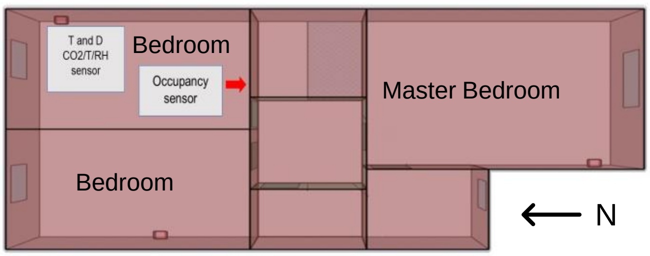
The structure of the paper is as follows. Section II describes the experimental testbed in this study, the data collected for this study and the sensors that have been used to collect data. Section III explains machine learning algorithms used to predict the room's occupancy state and compare their performance. Furthermore, it discusses the possibility of using temperature and humidity data for occupancy detection. Finally, Section IV discusses the summary and conclusion of the study.

II. TEST-BED AND DATA COLLECTION

The testbed in this study is a bedroom on the second floor of a half-duplex residential house located Southwest of Edmonton, Alberta. A CO2 sensor, model RTR-576 with an RTR500BW data collector made by T&D, was installed in the bedroom which collects and saves CO2 concentration, temperature, and humidity in the room. Fig. 2 shows a schematic of the house and bedrooms on the second floor. In addition, sensor specifications used in this study are presented in Table 1.

The time interval for the data collection has been defined as one minute. To collect the ground occupancy data, a PIR motion sensor was added to this room. The PIR data were cross-referenced with the occupancy data sheet filled by the residents of the building to validate the occupancy information generated by the machine learning model based on the environmental data.

a)



b)

Figure 2: a) North view of the house, b) schematic of the second floor and sensor's location.

The building is equipped only with a heating system which provides heat during the cold season (i.e., September to May). The heating system is fitted with a TRANE Upflow left-induced draft gas furnace (Model: TUE1B080A9361A) that runs with natural gas.

Data including CO2 concentration, temperature, and humidity were collected from March 4th, 2022, 12:00 a.m., till May 7th, 2022, 08:55 a.m. Considering the recording interval of one minute, 278,088 data samples were collected in total, for 92,696 minutes. Based on the data for the actual occupancy, the room was occupied for 21,006 minutes out of 92,696 minutes.

In addition, a Monnit wireless PIR motion detection sensor along with a log sheet filled by the occupants was used to record the actual state of occupancy in the room.

III. MACHINE LEARNING CLASSIFIERS

For this study, four different classifiers that reported high accuracy for occupancy prediction in the literature [26-31] were used to predict the state of occupancy in the room. These classifiers include (1) k-nearest neighbors (KNN), (2) artificial neural network (ANN), (3) Gaussian support vector machine (SVM), and (4) decision tree (DT). Fig. 3 depicts the inputs and output of these classifiers. CO2 concentration, room

Table 1: Sensor's specifications. a) T&D RTR-576 Sensor. b) Monnit wireless PIR motion detection sensor

a)

T&D RTR-576 Sensor 			
Channels	CO_2 concentration	Temperature	Relative humidity
Accuracy	$\pm(50\ ppm + 5\%\ of\ reading)at\ 5000\ ppm\ or\ less$	± 0.5°C	$5\%\ RH\ at$ 25°C
Measurement resolution	Minimum of 1 ppm	0.1°C	$1\%\ RH$
Measurement range	0 to 9,999 ppm	$0\ to\ 55$ °C	$10\ to\ 95\%\ RH$
Recording interval	1 minutes		

b)

Monnit wireless PIR motion detection sensor 	
Data logging	On Wi-Fi disruption, the unit will log the first 50 readings and transmit them when Wi-Fi connection is re-established
Wireless range	Up to 30 m
Sensor warmup time	30 Seconds
Sensing Technology	Passive infrared
Sensing range	5 m

temperature, relative humidity, and day of the week were defined as predictor features, and the state of occupancy was considered to be predicted.

KNN finds the k nearest training points based on a selected distance metric such as Euclidean, and Chebyshev and then classifies the test point according to the label of most of those k points [32]. A neural network feeds the input layer to the first hidden layer of the network and performs the activation calculation which can be chosen from different functions e.g., sigmoid, ReLU, etc. The output of each layer will be calculated by summing the input and applying the transfer function to it. After calculating the output of the network, the

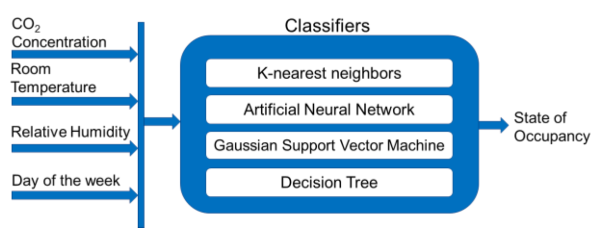


Figure 3: Structure of the classifiers and inputs and output of the model

weight of each neuron will be modified by using a backpropagation algorithm to minimize the error [33]. SVM classifiers label the data by forming hyperplanes to separate the data and then try to maximize the margin between the hyperplane and the data to increase the confidence of classification [34]. Finally, the decision tree splits parent nodes into child nodes by using the input variables till getting into the pure nodes or reaches the stopping rule [35]. The design parameters for the four classifiers used in this study are shown in Table 2.

As the state of occupancy in the room is not balanced (23% occupied and 77% unoccupied), Geometric Mean (GM), is used instead of accuracy to measure the performance of the classifiers. In the case of imbalanced data, accuracy can be highly affected by the dominant class, but GM considers the performance of both major and minor groups; thus, it won't overfit or underfit the imbalance classes [36].

$$GM = ((TP / (TP+FN)) * (TN / (TN +FP)))^{0.5} \quad (1)$$

Where; TP: True Positive, FP: False Positive, TN: True Negative, and FN: False Negative

Data collected from March 4th to April 3rd, 2022, were used to train models. 70% of the data were randomly selected to train and validate the model and the rest was used to test the models. In addition, five-fold cross-validation was implemented for training and validation of the model to avoid any overfitting.

MATLAB Classification Learner Toolbox was used to train and test all four algorithms. A comparison between their GMs is presented in Fig. 4.

In addition, the confusion matrix for all four algorithms while using four features including room temperature, relative humidity, CO_2 concentration, and day of the week for predicting the occupancy can be seen in Fig. 5. Both GM and confusion matrix show that the KNN model provides the best performance by predicting the occupancy with more than 90% accuracy.

Table 2: Design parameters of the four classifiers used in this study.

Machine learning Algorithm	Design parameters
K-nearest neighbors	Weighted KNN with 10 neighbors Distance metric: Euclidean Distance weight: Squared inverse
Artificial Neural Network	2 layers First layer size: 15 Second layer size: 15 Activation: ReLU Iteration limit: 1000 Lambda: 0
Gaussian Support Vector Machine	Kernel function: Gaussian Kernel scale: 8 Box constraint level: 1 Multiclass method: One-vs-One
Decision Tree	Maximum number of splits: 20 Split criterion: Gini's diversity index Surrogate decision splits: Pilot

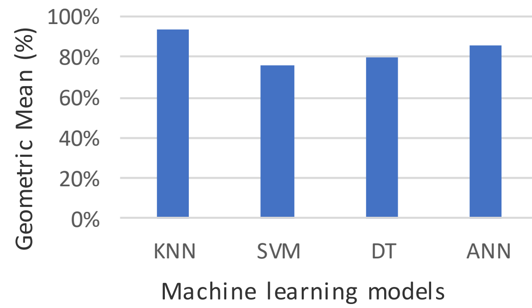


Figure 4: Geometric Mean for four occupancy prediction models. All models used the same standardized data.

If all four features are used for the occupancy prediction, all algorithms provide high-quality occupancy data which can be used by the building control system to optimize energy consumption. Unfortunately, CO_2 sensors are expensive and cannot be used widely in residential buildings. To investigate the possibility of using the most available data in residential buildings such as room temperature and relative humidity, all the algorithms were trained and tested using all features except CO_2 concentration.

Fig. 6 depicts the performance of different algorithms while removing CO_2 concentration data. Despite the decrease in the accuracy of the models after removing CO_2 as one of the features, the KNN method can predict the state of occupancy in the room accurately by having the room temperature, relative humidity, and day of the week.

To study the effect of having a lag between the data used for training the model and the data used for making the

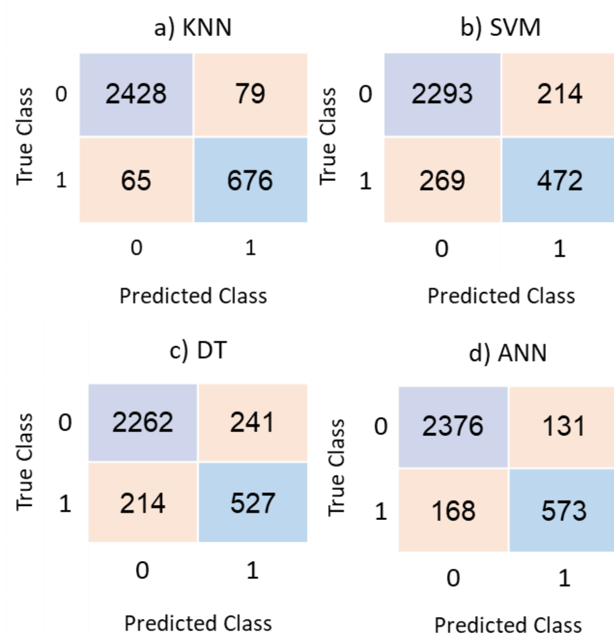


Figure 5: Confusion matrices for occupancy detection test using different ML algorithms with four predictor features (i.e., T_{room}, CO_2, RH, and Day of the week).

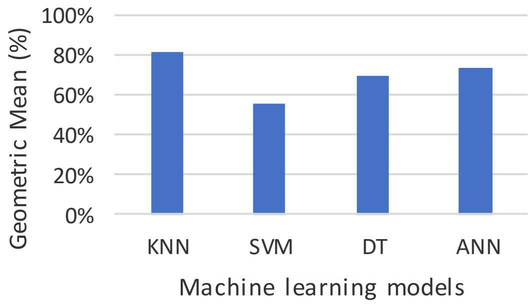


Figure 6: Geometric Mean, while CO_2 concentration has been removed from predictor features. All models used the same standardized data.

prediction, the KNN, ANN, SVM, and DT models were used to predict the state of occupancy for four different weeks. As can be seen in Fig. 7 accuracy of using a model that has been trained by data obtained a month ago drops dramatically. This can be explained as occupancy in residential buildings changes during the year. The outside activity during spring and summer could be different from activities during fall and winter, and these changes in the occupant behavior should be captured by the model, otherwise, a model that has been trained based on the data collected during summer, cannot predict the occupancy in winter accurately.

To make sure a model performs accurately during the year, the times that can change the occupant behavior should be identified and the model should be trained again with the data collected close to those occasions.

IV. SUMMARY AND CONCLUSION

This study investigated KNN, ANN, SVM, and DT algorithms to predict occupancy in one bedroom of a 3-bedroom residential house. The input data to these algorithms include room temperature, relative humidity, CO_2 concentration, and day of the week. Among all four algorithms which have been used in this study, KNN provides the best performance in both cases of adding or removing CO_2 concentration by achieving 94% and 81% GM respectively. Despite the decrease in the performance of ANN model (from 86% to 74%) after removing CO_2 concentration from predictor features it still can be implemented for occupancy detection even with limited sensor data.

Considering the reasonable performance of occupancy prediction models by using room temperature, day of the week, and relative humidity, further research can be conducted to use environmental data from different zones in the house to determine the occupancy in each zone. The state of occupancy along with other available data can be fed into an advanced control system e.g., a model predictive control system to minimize the energy consumption in the house in real-time.

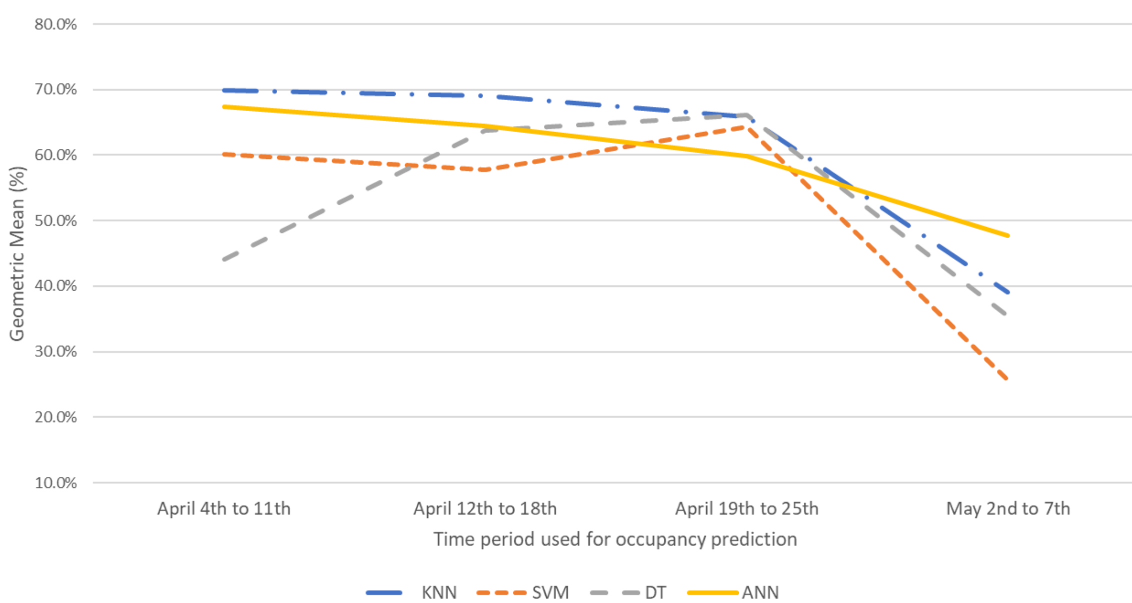


Figure 7: The performance of models trained using data collected between March 3rd and April 4th for predicting occupancy for the unseen data.

REFERENCES

[1] X. Cao, X. Dai, and J. Liu, "Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade," Energy and buildings, vol. 128, pp. 198–213, 2016.

[2] N. R. Canada. "Commercial/institutional sector – canada." (2018), [Online]. Available: https://oee.nrcan.gc.ca/corporate/statistics/neud/dpa/menus/trends/comprehensive/trends com ca.cfm (visited on 09/27/2022).

[3] N. R. Canada. "Residential sector – canada." (2018), [Online]. Available at https: //oee.nrcan.gc.ca/corporate/statistics/neud/dpa/menus/trends/comprehensive/trends res ca.cfm (visited on 09/27/2022).

[4] A. Mirakhorli and B. Dong, "Occupancy behavior based model predictive control for building indoor climate—a critical review," Energy and Buildings, vol. 129, pp. 499–513, 2016

[5] .B. Hojjati. "Global energy consumption driven by more electricity in residential, commercial buildings." (2019), [Online]. Available: https://www.eia.gov/todayinenergy/detail.php?id=41753. (visited on 09/27/2022).

[6] S. H. Ryu and H. J. Moon, "Development of an occupancy prediction model using indoor environmental data based on machine learning techniques," Building and Environment, vol. 107, pp. 1–9, 2016.

[7] A. Standard, "Standard 55-2020 thermal environmental conditions for human occupancy," Ashrae: Atlanta, GA, USA, 2020.

[8] D. Energy. "What should my thermostat be set to in winter?" (2022), [Online]. Available: https : / / www . directenergy. ca / learn / recommended - thermostat -settings-winter (visited on 11/16/2022).

[9] M. Barnes. "What temperature should you keep a vacant house at?" (2022), [Online]. Available: https : / / apollocover . com / magazine / what - temperature -should-you-keep-a-vacant-house-at/ (visited on 11/16/2022).

[10] J. Brooks, S. Kumar, S. Goyal, R. Subramany, and P. Barooah, "Energy-efficient control of under-actuated hvac zones in commercial buildings," Energy and Buildings, vol. 93, pp. 160–168, 2015. 32

[11] B. Dong and B. Andrews, "Sensor-based occupancy behavioral pattern recognition for energy and comfort management in intelligent buildings," in Proceedings of building simulation, International Building Performance Simulation Association Vancouver, 2009, pp. 1444–1451

[12] G. of Canada. "Energy fact book 2021-2022." (2021), [Online]. Available: https://www.nrcan.gc.ca/science- and- data/data- and-analysis/energy- data- and-analysis/energy-facts/20061 (visited on 11/18/2022).

[13] J. Ahmad, H. Larijani, R. Emmanuel, M. Mannion, and A. Javed, (2021). Occupancy detection in non-residential buildings–A survey and novel privacy preserved occupancy monitoring solution. Applied Computing and Informatics, 17(2), 279-295.

[14] Y. Ding, S. Han, Z. Tian, J. Yao, W. Chen, and Q. Zhang, (2021, August). "Review on occupancy detection and prediction in building simulation." Building Simulation (pp. 1-24). Tsinghua University Press.

[15] D. Liu, X. Guan, Y. Du, and Q. Zhao, (2013). Measuring indoor occupancy in intelligent buildings using the fusion of vision sensors. Measurement Science and Technology, 24(7), 074023.

[16] S. Zhan, and A. Chong, (2021). Building occupancy and energy consumption: Case studies across building types. Energy and Built Environment, 2(2), 167-174.

[17] K. C. J. Simma, A. Mammoli, and S. M. Bogus, (2019). Real-time occupancy estimation using WiFi network to optimize HVAC operation. Procedia Computer Science, 155, 495-502.

[18] J. Ahmad, H. Larijani, R. Emmanuel, M. Mannion, and A. Javed, (2020). Occupancy detection in non-residential buildings–A survey and novel privacy preserved occupancy monitoring solution. Applied Computing and Informatics, 17(2), 279-295.

[19] J. Yang, M. Santamouris, and S. E. Lee, (2016). Review of occupancy sensing systems and occupancy modeling methodologies for the application in institutional buildings. Energy and Buildings, 121, 344-349.

[20] X. Li, Y. Zhang, I. Marsic, A. Sarcevic, and R. S. Burd, (2016, November). Deep learning for rfid-based activity recognition. In Proceedings of the 14th ACM Conference on Embedded Network Sensor Systems CD-ROM (pp. 164-175).

[21] L. M. Ni, Y. Liu, Y. C. Lau, and A. P. Patil, (2003, March). LANDMARC: Indoor location sensing using active RFID. In Proceedings of the First IEEE International Conference on Pervasive Computing and Communications, 2003.(PerCom 2003). (pp. 407-415). IEEE.

[22] L. Rueda, K. Agbossou, A. Cardenas, N. Henao, and S. Kelouwani, "A comprehensive review of approaches to building occupancy detection," Building and Environment, vol. 180, p. 106 966, 2020.

[23] K. Tutuncu, O. Cataltas, and M. Koklu, (2016). Occupancy detection through light, temperature, humidity and CO2 sensors using ANN. Int. J. Ind. Electron. Electr. Eng, 5, 63-67.

[24] W. Kleiminger, C. Beckel, T. Staake, and S. Santini, (2013). Occupancy detection from electricity consumption data. In Proceedings of the 5th ACM workshop on embedded systems for energy-efficient buildings (pp. 1-8).

[25] A. Szczurek, M. Maciejewska, and T. Pietrucha, (2017). Occupancy determination based on time series of CO2 concentration, temperature and relative humidity. Energy and Buildings, 147, 142-154.

[26] S. H. Ryu and H. J. Moon, "Development of an occupancy prediction model using indoor environmental data based on machine learning techniques," Building and Environment, pp. 1-9, 2016

[27] W. Kleiminger, C. Beckel, T. Staake and S. Santini, "Occupancy detection from electricity consumption data," BuildSys, 2013.

[28] L. Candanedo and V. Feldheim, "Accurate occupancy detection of an office room from light, temperature, humidity and CO2 measurements using statistical learning models," Energy Build, vol. 112, pp. 28-39, 2016.

[29] K. Tutuncu, O. Cataltas and M. Kokl, "Occupancy detection through light, temperature, humidity and co2 sensors using ANN," in roceedings of ISER 45th International Conference, Rabat, Morocco, 2016.

[30] A. Szczurek, M. Maciejewska and T. Pietrucha, "Occupancy determination based on time series of CO2 concentration, temperature and relative humidity.," Energy and Buildings, pp. 142-154, 2017.

[31] A. Javed, H. Larijani, A. Ahmadinia and D. and Gibson, "Smart random neural network controller for HVAC using cloud computing technology," IEEE Transactions on Industrial Informatics, pp. 351-360, 2016.

[32] L. Kozma, "k Nearest Neighbors algorithm (kNN)," Slides for Special Course in Computer and Information Science, Helsinki University of Technology, 2008.

[33] K. P. Sudheer, A. K. Gosain, D. Mohana Rangan and S. M. Saheb, "Modelling evaporation using an artificial neural network algorithm.," Hydrological Processes, vol. 16, pp. 3189-3202, 2002.

[34] W. S. Noble, "What is a support vector machine?," Nature biotechnology, vol. 24, no. 12, pp. 1565-1567, 2006.

[35] Y. Y. Song and L. U. Ying, "Decision tree methods: applications for classification and prediction," Shanghai Archives of Psychiatry, vol. 27, no. 2, p. 130, 2015..

[36] A. Tharwat, "Classification assessment methods," Applied computing and informatics, vol. 17, no. 1, pp. 168–192, 2021