The 10th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications
18-21 September, 2019, Metz, France

Assessment of Occupancy Estimators for Smart Buildings

Claudia Chitu[1] , Grigore Stamatescu[1, 2] , Iulia Stamatescu [1] , Valentin Sgârciu[1]
[1] Department of Automatic Control and Industrial Informatics, University Politehnica of Bucharest, Romania,
email: *claudia@imtt.pub.ro; iulia.stamatescu@aii.pub.ro; vsgarciu@aii.pub.ro*
[2] Institute of Technical Informatics, Technical University of Graz, Austria, e-mail: *gstamatescu@tugraz.at*

***Abstract* — Dense distributed sensor systems deployed to monitor and control the built environment open up wide application areas for improving the quality of life through smart spaces. By implementing these systems in residential or commercial buildings, more efficient, pervasive, real-time and finally more robust decision support is achieved. As one of the key tasks, high accuracy occupancy detection and estimation within buildings offers the potential to improve HVAC utilization while reducing maintenance needs, enhancing energy savings for building operators as well as a more suitable thermal comfort for occupants. In this context the main contribution of the paper consists of an assessment of supervised machine learning techniques namely random forests and extreme machine learning neural networks for indirect occupancy profile estimation. The methods are applied on a benchmarking dataset of indoor carbon dioxide measurements and ventilation damper positions as collected from the building information system. The impact of data preprocessing through filtering and smoothing as well as hardware constraints and cloud distributed processing for algorithm deployment are also discussed. We highlight key challenges and discuss how the learned models can be integrated for online operation to improve energy efficiency.**

Keywords* — *statistical learning; occupancy estimation; sensor data processing; smart buildings

I. INTRODUCTION

Increasing constraints on the energy efficiency of buildings, including the ones imposed through environmental regulations in Europe, has led to more advanced control systems to optimise building operation. The buildings are thus gradually transformed into smart spaces with new generation of networked sensors and actuators, multimodal user interfaces and analytics, alerting and reporting for the owner or operator. These are usually integrated under the umbrella of next generation Building Management Systems (BMS). Interdisciplinary research groups have started to collaborate more intense within this field through teams of data scientists, civil engineers, system engineers, facility managers, for a better understanding of the occupant patterns, which can be recognized as one critical driving element of overall energy consumption. The example of interest for this contribution is related to the indoor ventilation requirements, that is the need to satisfy the need of reducing CO_2 levels by gradually replacing indoor air with fresh one in the economizer with minimal energy waste [1].

In terms of sensor data processing, addressing data fusion techniques, inference learning, or historical data analysis for thousands of non-residential buildings [2], [3], [4] current open source statistical learning libraries can be used. Computing aspects are critical for such algorithms in on-line operation through parallel computing or distributed nodes using cloud platforms [5], [6].

The challenges of handling large quantities of heterogeneous building data concern noisy, structured, semi-structured and unstructured data and the ability to run the algorithms on big data platforms on cloud. Measurements collected by the heating, ventilation and air conditioning systems (HVAC) can subsequently be used to indirectly estimate room usage with the purpose of substituting dedicated additional sensors for people counting, thereby significantly reducing monitoring system cost. We thus present a use case of estimating occupancy patterns in a benchmark non-residential building, using already deployed sensors [7].

Our contributions presented in this paper are argued to be:

- implementation and benchmarking of two machine learning techniques, namely Random Forest and Extreme Learning Machine, on real data collected in a reference academic building for occupancy level estimation, including a data preprocessing phase;
- assessment of technical aspects regarding algorithm implementation for local machine and for cloud instances using Spark technology as analytics engine for large-scale data processing and discussion of potential real world validation.

The rest of the paper is organized as follows. Section II presents a summary of relevant works applied for buildings data. Section III focuses on the methods to describe from a statistical point of view the mathematical modeling and the data processing frameworks. The presentation and discussion of the experimental results is included in Section IV, emphasizing the importance and existing limitations of the work within the context of smart spaces. The paper concludes with future work in Section V.

II. Related work

Occupancy detection and forecasting in non-residential and office buildings is enabled by a large range of physical sensor systems [8] where the information is extracted with statistical [9] and machine learning techniques [10], [11]. This has led to enhancing the accuracy of estimation and prediction for occupancy related tasks. Through a data fusion approach by consideration of sources already present in the building infrastructure, indirect estimation of room occupancy levels can be achieved. This can subsequently be integrated into a predictive control algorithm at the HVAC level for occupant aware optimised control.

A research of user behavior using several classification techniques including Random Forests (RF) was presented by [12]. This stresses the importance or proper feature selection and model structure for high accuracy in the problem of user detection. Though the topic is intensively studied and several machine learning algorithms are implemented and well exploited, one salient limitation is that the occupancy is binary, in terms of occupied or unoccupied rooms. It is very important to know a good approximation of the total person count for minimizing the electricity consumption of the HVAC system as it is presented in [13], while accounting for the ventilation requirements and net-energy contribution to the space of each user.

Prediction of the occupant pattern in buildings is modeled in [14] using Extreme Learning Machine neural networks (ELM) for CO_2 data. The authors present a variation of ELM algorithms with high enhancement of occupancy detection for a zone with 24 occupants, but testing only for one zone. The ELM technique is exploited in several other studies as in [15] where a reference discussion and implementation are carried out for efficient implementation on low resource edge devices while arguing a 3x speed-up in a particular context. [16] introduce a two-stage approach for occupancy detection learning algorithms using ELM for the first stage of fast non-linear classification followed by inserting the results in a Support Vector Machine (SVM) algorithm for final outcomes. In [17] the authors focus on feature selection to improve the effectiveness of ELM for occupancy estimation. This relies on a filter component and a wrapper component which use raw measurements of carbon dioxide, relative humidity (RH), temperature and ambient air pressure. The method provides good results for indoor human detection accuracy, from 75-85% up to 90-95%.

The indirect room usage estimation based on carbon dioxide sensor measurements, has thus gathered intensively researcher attention in the last decade [18], being still an actual interesting topic [19] because of its high potential to improve energy efficiency and occupant comfort in modern building automation applications.

III. Methods

In this section, we explain the theoretical models used to estimate occupancy based on CO_2 concentration levels and duct airflow provided in terms of damper openness while having the ground truth in terms of number of occupants.

Often it is argued that CO_2 concentration levels are not accurate, being affected by error accumulation. To address this aspect, especially that the data providers [20] suggest addressing known issues with CO_2 concentration levels as drifts and offsets, we applied 2 different steps for this measurement in order to pre-process it: CO_2 data smoothing technique presented in [14] and a Kalman filter.

To deal with the offsets induced by the air movement of the persons getting closer to the sensors and measurements noise represented as spikes, we estimate the current value using a smoothing technique first. Let us consider our measurements of concentration levels as a vector $c = [c_1, c_2, ..., c_n]^T$. Then, the smoothed data denoted here c_s can be found, as presented in [14], by minimizing the function:

$$J(c_s) = ||c - c_s||_2^2 + \lambda||\nabla c_s||_2^2 \tag{1}$$

where ∇ is the gradient, $||.||_2^2$ is the square of Norm 2 (Euclidian norm) and λ is a weighting factor. A larger value of λ leads to increased smoothing effect. After solving the problem according to the same source, the c_s is implemented with the formula:

$$c_s = (I + \lambda)^{-1} c \tag{2}$$

In addition, we used a Kalman filter to compare the previous values of the concentration levels. This type of filter is suitable for real time problems, being light on memory and very fast while dealing effectively with the uncertainty induced by noisy sensor readings.

The problem formulation states that $x \in R^n$ is the signal value to be estimated of a discrete control process. The signal is governed by the equation:

$$x_k = Ax_{k-1} + Bu_{k-1} + w_k \tag{3}$$

and the measurement is given by $z \in R^m$:

$$z_k = Hx_k + v_k \tag{4}$$

with w_k and v_k random variables and representing the process noise and the measurement noise. This could be interpreted that every value of the signal is a linear combination of its previous value, a control signal and process noise. Equation (4) says that a measurement value is represented as a linear combination of the signal value and measurement noise. For $A = 1$, the state does not change from step to step, and no control input meaning $u = 0$, replacing in (3), we have:

$$x_k = x_{k-1} + w_k \tag{5}$$

and if we consider a measurement $z \in R^1$ and $H = 1$, as the signal directly is measured with the noisy measurement, in (4):

$$z_k = x_k + v_k \tag{6}$$

In many of the signal filtering situations, the variables A, B and H are not matrices, but numeric values and for simplified models, assume they do not change from one state to another, remaining constant. Following the mathematical explanation of the Kalman model from [21], we will use the next equations, considered at the state k:

$$K_k = \frac{P_k^-}{P_k^- + R} \tag{7}$$

$$\hat{x}_k = \hat{x}_k^- + K_k(z_k - \hat{x}_k^-) \tag{8}$$

$$P_k = (1 - K_k)P_k^- \tag{9}$$

with $\hat{x}_k$ the a priori estimate, K_k the Kalman Gain, R the covariance of the observation noise and P_k^- the a priori estimate covariance. These are known as the measurement correction equations. They are coupled with what is called the time update equations, where Q is the covariance of the process noise:

$$\hat{x}_k^- = \hat{x}_{k-1} \tag{10}$$

$$P_k^- = P_{k-1} + Q \tag{11}$$

The goal is to find the $\hat{x}_k$ which is the estimate of x. The current estimates will be input for the next states, this meaning that the initial value of P is needed (P_0), and in practice it is assigned, as well as x_0.

Once having the data pre-processed with these filters, we input them into a processing pipeline using a Random Forest (RF) model. Let us consider the input dataset $X = \{x_1, x_2, ..., x_n\}$ (features) and an output dataset $Y = \{y_1, y_2, ..., y_n\}$ (labels). The RF model assumes a training of a classification tree T_b on a sample with replacement dataset from X_b and Y_b with $b = \{1, ..., B\}$. The output of the algorithm is the ensemble of trees $\{T_b\}_1^B$. For the classification task, the prediction for a new unseen sample, x, is described by:

$$\hat{T}_{rf}^B(x) = MajorityVote\{\hat{T}_b(x)\}_1^B \tag{12}$$

The split node within the tree is chosen using the Gini impurity given by:

$$1 - \sum_{i=1}^{c} p(i)^2 \tag{13}$$

where $p(i)$ is the probability of a certain classification per dataset. In other words, it tells us how often, by choosing a random sample, it would be incorrectly labeled considering the distribution of labels and doing a random labeling.

In parallel of this technique, we used Extreme Learning Machine for tackle the same problem. This type of feed-forward neural network is designed usually with one non-linear hidden layer of neurons. The number of occupants is given by:

$$\sum_{i=1}^{L} = \beta_i h(w_i^T x_k + b_i) \tag{14}$$

where $w_i \in R^n$ represent the random generated weights from the input layer to the $i^t h$ hidden neuron, b_i is the random bias of the input for the $i^t h$ hidden neuron, $h(.)$ is called the activation function and $\beta = [\beta_1, \beta_2, ..., \beta_L]^T$ contains the weights from the hidden layer to output and is found by solving a least square problem. Because we address the estimation problem for city-scale energy footprint estimation, we approach an implementation of ELM optimized for Big Data applications as presented in [15] to reduce the implementation complexity. Each parameter w for the input to hidden layer and from the hidden layer to the output layer is mapped using the transformation:

$$w_q = round(w(-1 + 2^{n_b} - 1) \tag{15}$$

where n_b is the number of bits for the neurons in the hidden layer.

IV. Results

The data sources are presented in [20], summing up a number of 21600 records per room, with a total of 4 rooms, for 15 days collection period across March and April 2017, at 1 minute frequency, from an academic space (study zones and lecture rooms). We aim to estimate the occupancy in each room based on the CO_2 concentration level and air flow given in terms of damper openness. This is important because these data sources are usually present in any building and once the occupancy level is estimated, it could be included in a control strategy to condition the room for the actual necessity and not for full occupancy as it happens in BMS. In the same time, this trade off between user comfort and energy savings is compliant with the European guidelines for ventilation requirements in buildings, with respect to CO_2 concentration.

Occupancy at maximum level recorded for each room is presented in Table I. The first room and the last one are dedicated to teaching and the second and third rooms are used as study zones.

Table I. Room occupancy

	Room1	Room2	Room3	Room4
Max. Occupants:	67	28	35	39

For the CO_2 smoothing technique that we presented in the previous section, we used $\lambda = -0.5$, because this value was suitable to our dataset, after trying other ranges. The implementation that we did for Kalman filter for the same variable, based on the equations (7), (8) and (9) is the following:

Algorithm 1 Kalman filter algorithm for CO_2 data

for $i \leftarrow 1$ to length(c) **do**
 $k \leftarrow \frac{p}{p+r}$
 $x \leftarrow x + k(c(i) - x)$
 $p \leftarrow p + q$
 current_estimate $\leftarrow$ x
end for

As we mentioned previously, x_0 should be initialized. For every room, we set this value to the first value recorded on the corresponding dataset. The other parameters were set to: $p_0 = 0.1$, $q = 1$, $r = 50$. Because the feature vectors, here the CO_2 concentration level and the air flow, are less numerous in diversity, we designed the smoothing and filtering techniques as soft filters, being relatively very close to the real data. This is justified by the motivation of staying close to the ground truth, to diminish the chances of misclassification. This was done based on data analysis: we noticed that at the concentration level of 411 ppm (parts per million) one person was in the room according to the ground truth; at the concentration level of 464 ppm, the records have again one person assigned. But, continuing the inspection, we found that for the value of 626 ppm we could have again one person in the room and at 629 ppm 2 persons. So, we could not drift with the filtering too much from the initial value. In Figure 1 are represented the first 100 values for the dataset corresponding to room 1, for a zoomed in image. We have now a new dataset which will name 'smooth' and it is very similar to the inital one, and a second new dataset which will call 'kalman' and represents a more dramatic filter comparing with 'smooth'.

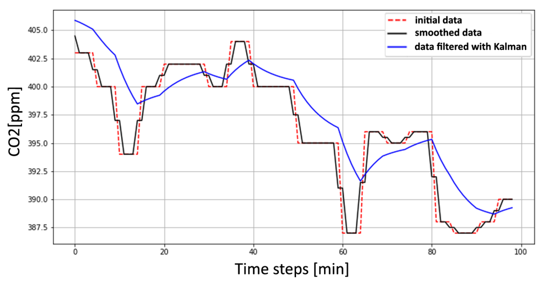

Figure 1. CO_2 filters comparison

In Table II we summarize the performance of the RF model for each room, where we refer to each dataset containing the CO_2 pre-processed as we mentioned above, and by 'Rn' we mean Room number n, with $n = 1, 2, 3, 4$. RMSE stands for root mean square error; the split was 80% for training data, with a random state of 42 and 100 trees in the forest. From this table we see that the

Table II. RANDOM FOREST RMSE

	RMSE		
	initial data	'smooth'	'kalman'
R1:	**8.39**	8.56	10.44
R2:	3.06	3.12	**2.75**
R3:	**3.36**	3.38	3.39
R4:	4.08	**3.82**	4.2

Table III. RANDOM FOREST TRAINING TIME

	Time[sec]		
	initial data	'smooth'	'kalman'
R1:	**0.16**	0.17	0.19
R2:	0.06	**0.05**	0.07
R3:	**0.09**	0.11	0.12
R4:	**0.12**	0.15	0.22

smoothed dataset was not performing better than the initial one, though was very similar, excepting one case, for the room 4. The same conclusion could be draw for the dataset obtained with Kalman filter. This could be interpreted that the initial dataset was clean since it is a relatively recent collected, the CO_2 sensors were well calibrated and the accumulated error was not impacting that much the performance of the classification model. On the other hand, we could not be sure about this hypothesis, and assume that the less performing behavior could be associated with a poor feature data vector. Our features were the concentration level for CO_2 and the airflow given by the damper position. The highest value for RMSE is accounted for the Room1, but this is the room with the largest maximum number of occupants (67) and highest occupancy every daily as could be observed from the Figure 2. The associated training time is presented in Table III and could be seen that again, the initial dataset had the best performance, here the least training time, in 3 out of 4 cases. In Figure 3 is a plot of the first 5 hours of training data and predicted occupancy values for room 1.

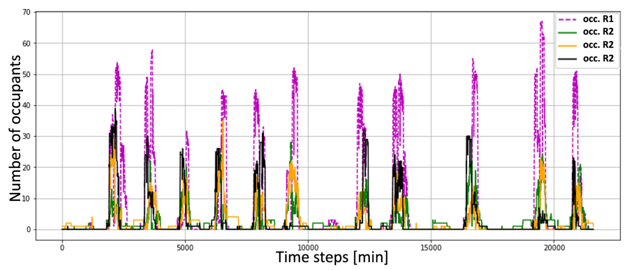

Figure 2. Occupancy distribution for room R1, R2, R3, R4

The RF model was implemented in Python, on a local machine. Despite the fact that at the moment we do not have access to large amount of dataset, the research community is preparing to assist to a city scale image of building performance, in terms of occupancy patterns, usage, electricity consumption which are all related and correlated with transportation and other urban services. For such platform to support these objectives, we tested

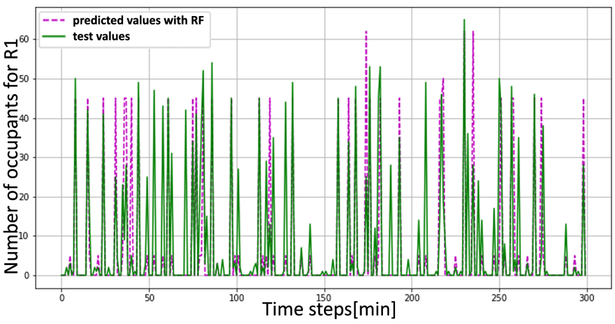

Figure 3. Occupancy distribution for room R1 using RF

Table IV. ELM ACCURACY FOR ROOM1 DATASET

Accuracy	with tanh	
initial data	'smooth'	'kalman'
69.17%	69.08%	67.95%
Accuracy	with abs	
initial data	'smooth'	'kalman'
68.11%	68.11%	67.92%

the RF model on a Spark machine, in cloud, provided by Databricks, using Pyspark, the Spark Python API that exposes the Spark model to Python programming language. We used Spark 2.4.0 version, 6GB memory and 1 Driver. The implementation is using the MLlib which is the Apache Spark machine learning library. The advantages of using Spark are multiple from using tens of data type sources to running of different environments such as EC2, Hadoop etc. For the first room dataset, using RF on Spark, we obtained an accuracy of 67.65%, with a training time of 4.94 seconds, a higher time than in the case when the model ran on the local machine. It is known that big data engines return a poor running time for small datasets comparing with the time needed for processing GB of data.

A second model is tested for the same data as the RF model, ELM with 2 types of activation functions: abs - absolute valued and tanh - hyperbolic tangent function given by the equation (15). We tried also the ReLU (Rectfied Liniar Unit), multi-quadratic and sigmoid, but these 2 presented the best results.

$$tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}} \tag{16}$$

In Table IV and Table V are summarized these results for the first room, as this one presented the highest RMSE when evaluating the RF model, i.e. the worst-case scenario. Here, the accuracy is given by the sum of true positive and true negative cases divided by the total values representing how often is the classifier correct. We keep the previous notations for 'smooth' and 'Kalman'. For this model described in the previous section, the parameters are set to: 2000 number of neurons, coefficient c is 0.1, and n_b for input is 4 and for output is set to 0.

What is to be noticed here is that the ELM model is

Table V. ELM TRAINING TIME

Training time[s]	with tanh	
initial data	'smooth'	'kalman'
2.14	1.85	1.99
Training time[s]	with abs	
initial data	'smooth'	'kalman'
1.95	1.82	1.98

more time demanding for the training phase comparing to the RF model, aspect specific to the neural networks. In this section we have explored and assessed the RF and ELM models as occupancy counters, analyzing the accuracy of each one and preparing the path for following up with more data to test on a big data platform.

V. CONCLUSION

In this study we have investigated occupancy estimator's performance in terms of accuracy and training time for nonresidential spaces, such as lecture rooms. We carried out a comparative assessment of two learning models, Random Forest and Extreme Learning Machine, for datasets from 4 rooms, considering measurements that usually are collected by the HVAC systems. With application on several important services for smart cities, building occupancy level in real time would potentially demand for big data processing frameworks, as already researchers have started to consider tens and hundreds of buildings to be analyzed [4]. In this sense, we tested Spark framework for Random Forest model, as an example to be followed up with massive datasets. A limitation of this study is that the algorithms were implemented without being tested on a real building, but this work could be followed by other researchers to extend the horizon of it. The highest importance or this research is that we forayed into data already present in buildings usually, which means that it is less pervasive and cost effective for being explored in many buildings. We believe that with an additional feature - the room temperature-, the models would have a very high accuracy in detection, but even as it is, included in a predictive control strategy, would enable more efficient ways to perform HVAC control. We found that by addressing CO_2 filtering methods the accuracy did not improve which could be explained by the fact that a good sensor calibration is reducing the offsets, so technological advancements decreased the need of pre-processing data. Because of the actual uncertainty of the estimation due to poor dataset in terms of features, this type of usage detection is recommended for large office spaces or academic lecture rooms where many persons are using the space and a band of few persons in detection could be added to reduce the estimation error and still save energy by not conditioning at maximum level. This is an initial step to understand patterns usage and unlock energy savings in the building sector with a good balance for user thermal comfort.

References

[1] F. C. Sangogboye, K. Arendt, M. Jradi, C. Veje, M. B. Kjærgaard, and B. N. Jørgensen, "The impact of occupancy resolution on the accuracy of building energy performance simulation," in *Proceedings of the 5th Conference on Systems for Built Environments*. ACM, 2018, pp. 103–106.

[2] C. Miller, "Predicting success of energy savings interventions and industry type using smart meter and retrofit data from thousands of non-residential buildings," in *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*. ACM, 2017, p. 17.

[3] C. Miller and F. Meggers, "The building data genome project: An open, public data set from non-residential building electrical meters," *Energy Procedia*, vol. 122, pp. 439–444, 2017.

[4] G. Fierro, M. Pritoni, M. AbdelBaky, P. Raftery, T. Peffer, G. Thomson, and D. E. Culler, "Mortar: an open testbed for portable building analytics," in *Proceedings of the 5th Conference on Systems for Built Environments*. ACM, 2018, pp. 172–181.

[5] M. Bărbulescu, R.-O. Grigoriu, G. Neculoiu, I. Halcu, V. C. Săndulescu, O. Niculescu-Faida, M. Marinescu, and V. Marinescu, "Energy efficiency in cloud computing and distributed systems," in *2013 RoEduNet International Conference 12th Edition: Networking in Education and Research*. IEEE, 2013, pp. 1–5.

[6] I. Stamatescu, V. Bolboacă, and G. Stamatescu, "Distributed monitoring of smart buildings with cloud backend infrastructure," in *2018 5th International Conference on Control, Decision and Information Technologies (CoDIT)*. IEEE, 2018, pp. 1035–1039.

[7] V. L. Erickson and A. E. Cerpa, "Occupancy based demand response hvac control strategy," in *Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building*. ACM, 2010, pp. 7–12.

[8] C. Chiţu, G. Stamatescu, I. Stamatescu, and V. Sgârciu, "Wireless system for occupancy modelling and prediction in smart buildings," in *2017 25th Mediterranean Conference on Control and Automation (MED)*. IEEE, 2017, pp. 1094–1099.

[9] C. Chitu, M. I. Khadem, and V. Sgarciu, "Predictive modeling of occupancy patterns in smart buildings," in *2017 9th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, vol. 2. IEEE, 2017, pp. 622–627.

[10] C. Nichiforov, I. Stamatescu, I. Făgărăşan, and G. Stamatescu, "Energy consumption forecasting using arima and neural network models," in *2017 5th International Symposium on Electrical and Electronics Engineering (ISEEE)*, Oct 2017, pp. 1–4.

[11] C. Nichiforov, G. Stamatescu, I. Stamatescu, I. Făgărăşan, and S. S. Iliescu, "Intelligent load forecasting for building energy management systems," in *2018 IEEE 14th International Conference on Control and Automation (ICCA)*. IEEE, 2018, pp. 896–901.

[12] L. M. Candanedo and V. Feldheim, "Accurate occupancy detection of an office room from light, temperature, humidity and co2 measurements using statistical learning models," *Energy and Buildings*, vol. 112, pp. 28–39, 2016.

[13] Y. Wei, L. Xia, S. Pan, J. Wu, X. Zhang, M. Han, W. Zhang, J. Xie, and Q. Li, "Prediction of occupancy level and energy consumption in office building using blind system identification and neural networks," *Applied Energy*, vol. 240, pp. 276 – 294, 2019.

[14] C. Jiang, M. K. Masood, Y. C. Soh, and H. Li, "Indoor occupancy estimation from carbon dioxide concentration," *Energy and Buildings*, vol. 131, pp. 132–141, 2016.

[15] R. Dogaru and I. Dogaru, "Optimized extreme learning machine for big data applications using python," in *2018 International Conference on Communications (COMM)*. IEEE, 2018, pp. 189–192.

[16] T. Liu, Y. Li, Z. Bai, J. De, C. V. Le, Z. Lin, S.-H. Lin, G.-B. Huang, and D. Cui, "Two-stage structured learning approach for stable occupancy detection," in *2016 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2016, pp. 2306–2312.

[17] M. K. Masood, C. Jiang, and Y. C. Soh, "A novel feature selection framework with hybrid feature-scaled extreme learning machine (hfs-elm) for indoor occupancy estimation," *Energy and Buildings*, vol. 158, pp. 1139–1151, 2018.

[18] M. Gruber, A. Trüschel, and J.-O. Dalenbäck, "Co2 sensors for occupancy estimations: Potential in building automation applications," *Energy and Buildings*, vol. 84, pp. 548–556, 2014.

[19] I. B. Arief-Ang, F. D. Salim, and M. Hamilton, "Da-hoc: Semi-supervised domain adaptation for room occupancy prediction using co 2 sensor data," in *Proceedings of the 4th ACM International Conference on Systems for Energy-Efficient Built Environments*. ACM, 2017, p. 1.

[20] K. Arendt, A. Johansen, B. N. Jørgensen, M. B. Kjærgaard, F. C. Sangogboye, J. H. Schwee, and C. Veje, "Room-level occupant counts, airflow and co2 data from an office building," in *Proceedings of the First Workshop on Data Acquisition To Analysis*, 2018, pp. 13–14.

[21] G. Bishop, G. Welch *et al.*, "An introduction to the kalman filter," *Proc of SIGGRAPH, Course*, vol. 8, no. 27599-23175, p. 41, 2001.