Scalable Room Occupancy Prediction with Deep Transfer Learning Using Indoor Climate Sensor

Davor Stjelja 1,2,*, Juha Jokisalo 1 and Risto Kosonen 1,3

1 Department of Mechanical Engineering and Automation, Aalto University, 02150 Espoo, Finland; juha.jokisalo@aalto.fi (J.J); risto.kosonen@aalto.fi (R.K.)
2 Innovations, Granlund Oy, 00701 Helsinki, Finland
3 College of Urban Construction, Nanjing Tech University, Nanjing 211800, China
* Correspondence: davor.stjelja@granlund.fi

Abstract: An important instrument for achieving smart and high-performance buildings is Machine Learning (ML). A lot of research has been done in exploring the ML models for various applications in the built environment such as occupancy prediction. Nevertheless, the research focused mostly on analyzing the feasibility and performance of different supervised ML models but has rarely focused on practical applications and the scalability of those models. In this study, a transfer learning method is proposed as a solution to typical problems in the practical application of ML in buildings. Such problems are scaling a model to a different building, collecting ground truth data necessary for training the supervised model, and assuring the model is robust when conditions change. The practical application examined in this work is a deep learning model used for predicting room occupancy using indoor climate IoT sensors. This work proved that it is possible to significantly reduce the length of ground truth data collection to only two days. The robustness of the transferred model was tested as well, where performance stayed on a similar level if a suitable normalization technique was used. In addition, the proposed methodology was tested with room occupancy level prediction, showing slightly lower performance. Finally, the importance of understanding the performance metrics is crucial for market adoption of ML-based solutions in the built environment. Therefore, in this study, additional analysis was done by presenting the occupancy prediction model performance in understandable ways from the practical perspective.

Keywords: transfer learning; occupancy prediction; indoor climate; IoT; model robustness; space efficiency; energy efficiency

1. Introduction

The CO₂ emissions of buildings in the European Union are 36% and 28% on the global scale [1,2], while HVAC systems in developed countries are responsible for 50% of building energy consumption alone [3]. To battle this problem, the EU has set the goal of developing a sustainable, competitive, secure and decarbonised energy system by the year 2050 [1].

One of the tools to achieve this goal is through the digitalization of energy systems and buildings and the EU has introduced a smart readiness indicator (SRI). The purpose of this indicator is to determine the capability of buildings in using information and communication technologies to adapt the building operation to the needs of the occupants and the grid while improving the overall performance of the buildings [1]. Recently, many research attempts have been made using advanced technologies in the field of computer science such as Artificial Intelligence (AI), Machine Learning (ML) and the Internet of Things (IoT) in building operations. These technologies are enabling use cases such as model predictive control, system fault detection and diagnosis, occupancy estimation and detection, demand response and more in general they are system integrators of different building subsystems and occupants [4,5]. The mentioned technologies can
help to significantly improve building performance, from better indoor climate and lower energy consumption to better space efficiency. However, it is important to minimize the interference with privacy and with work activities in order for office occupants to accept smart technologies, as was found in a recent survey [6].

Office building occupancy efficiency has significant potential for improvement; one British study from 2013 concluded that regular offices had an average occupancy level between 60 and 70% [7]. The recent COVID-19 pandemic has fast tracked remote work, making occupancy levels lower, where in one Israeli study 82% of interviewees were working full-time from the office before the pandemic and during the pandemic 41% of them claim to not go to the office at all. While after the pandemic more than 60% expect to come to the office only two to three times a week [8]. Furthermore in a report by McKinsey & Company, the authors presume that organisations will use their space for predominantly collaboration rather than individual work, which might require that most office space is turned into collaboration rooms [9]. If these claims come true, existing buildings and their HVAC systems could become even more inefficient, since the standard static occupancy profile, according to which they were designed, is not valid anymore.

This all leads to the expectation that information on building occupancy is going to become more valuable and needed than ever before as it is important for optimizing the space and the building’s system operation. Fortunately, a considerable amount of literature has been published on the matter of building occupancy estimation and detection. One of the extensive reviews in the literature on the topic was done by Chen et al. [10], where reviewed technologies range from presence and environmental sensors to WiFi and cameras, each of them with their own set of opportunities and challenges.

For example, Passive Infrared (PIR) sensors show good accuracy in occupancy detection if occupants are moving. On the other hand, static occupants or occupants outside the field of view are not going to be detected, which can be seen from [11], where PIR data does not show uniform occupancy during any of the occupied sessions. Most of the work based on PIR occupancy prediction sensors was about detection, but some, such as [12], worked on occupancy prediction aided with ML.

Another interesting data source which can be used to extract information on building occupancy is smart meters, since they record and communicate energy consumption near real time. Using the metered consumption of electricity or water with ML-based methods, it is possible to infer information on occupancy. The assumption is that appliances, lighting or water consumption are highly correlated with building occupancy. Most of the work done in this field was done for occupancy detection [13–15], but there have been some studies working on the prediction of the number of people or at least an occupancy range [16,17]. Typically, smart meters are installed to measure the consumption in the wider spatial scale, such as a whole building, a floor or a zone. Consequently, occupancy can be estimated only for an area of that scale. If the goal is to estimate the occupancy on a room or even a sub-room scale, then the more helpful sensors are those which are usually installed on that level such as PIR, air quality sensors, cameras, etc. Since the focus of this work is inferring occupancy information using indoor climate sensors, the literature review focuses on them.

To infer and analyze the information on occupancy in a larger scale, the commonly available sensors in buildings should be utilised. Indoor climate sensors commonly available in buildings are CO₂, room air temperature, relative humidity and in some cases Total Volatile Organ Compound (TVOC), of which CO₂ has the highest correlation with occupants in a room [18,19]. CO₂ as a sensor for occupancy estimation has been used for a longer time. Physical methods to predict the number of persons in the space are based on the mass balance equation and are either steady state or dynamic. In general physical methods can only approximate the number of people and with significant delays because of the air volume of the space [20,21]. The challenge with physical methods is their low scalability to other spaces, which decreases their potential to be applied widely. For example, the following input is needed for CO₂ mass balance model: Air volume of space,
adjacent room information such as CO\textsubscript{2} concentration or exchanged air between rooms, outdoor CO\textsubscript{2} concentration, exchanged air with outdoors, and ventilation schedule [22].

Recently, there have been studies attempting to solve the scalability problems of physical models by combining them with stochastic differential equations (SDEs) into the so-called grey box models, such as in [23]. The strength of this model is that it is generic, in the sense that it only uses measured CO\textsubscript{2} and the information is the ventilation system on or off as inputs. Other information about the space and ventilation system is not necessary. On the other hand, this model requires a training period and needs to go through different phases (occupied while ventilation is off, occupied when ventilation on, etc.) for the model to work properly. Wolf et al. [23] concluded that this model is accurate in terms of estimating binary occupancy (occupied or not) and that errors occur more frequently for higher numbers of occupants.

Regarding the data-driven methods for occupancy prediction with room CO\textsubscript{2} and/or other indoor climate sensor data, they can be divided into statistical, machine learning or deep learning methods. From statistical methods, models based on Hidden Markov Model (HMM) have been the most popular, which have a good understanding of temporal dependency in occupancy data, but can be easily affected by sensor noise and do not consider different occupancy dynamics in different time periods [18,19,23–26].

Common machine learning methods have been used as well, such as the random forest [27], support vector machine (SVM) [18,28–30], k-nearest neighbours (k-NN) [31] or different artificial neural network methods [18,30,32–34]. Machine learning methods have shown good accuracy with occupancy prediction, but require a process which is called feature engineering. Feature engineering is used for extracting more representative information from raw sensor data. First, raw sensor data is noisy, which can increase the prediction inaccuracy. Second, the effect of occupancy on the Indoor Air Quality (IAQ) is not momentary, but extends over a period of time and this information needs to be brought to the model. The downside is that the process of feature engineering is time consuming and labour intensive while the list and ranking of features useful for prediction can differ from case to case, making it difficult to scale.

In the work by Chen et al., these challenges were solved using a deep learning based approach [35]. The authors have used the Convolutional Deep Bidirectional Long Short-Term Memory (CDBLSTM) method, which automatically extracts significant features from raw data and understands temporal dependency of sequential data such as IAQ data. The proposed method has outperformed other statistical and machine learning based methods in both occupancy detection and estimation. Furthermore, the authors have tested the method with noisy data and with the data from a different room, all of it showing good performance. However, for applying the method to another room, the authors have needed to collect new ground truth data and train the model from scratch. The collection of occupancy ground truth is challenging and has been identified in several works as costly and time consuming [5,36,37] where it requires installation of devices such as: Cameras, PIR sensors or even manual counting by a person. This is an issue if this method would want to be used for large-scale practical purposes.

Most of the earlier introduced ML models [27,29,30,33–35] for occupancy prediction utilized data from one or two rooms only, where the output was the model accuracy compared to the ground truth and/or models from previous research. Wang et al. [5] emphasised that research in the field focuses highly on ML algorithm development instead of focusing on the overall ML model implementation process. The authors, in their literature review of machine learning in building energy management field, identified practical issues in the ML model implementation, such as inadequate model adaptability, limited data collection techniques, lack of user confidence, etc. This makes practical applications of ML in the field of building operations rare.

One way to solve the scalability issue of the ML methods in building operations is to use an unsupervised method, for which ground truth occupancy data and model training are not needed. Recently two studies were published with the purpose of inferring
occupancy information for commonly measured variables in buildings [17,38]. In a study by Mora et al. hierarchical clustering was used for identification of the occupancy patterns of a single-person office [38]. The clustering was applied on data such as CO\textsubscript{2}, electric power, room air temperature, relative humidity, air conditioning operation mode and window opening status. With their methodology, the authors were able to detect office room occupancy presence in different time periods, especially using the CO\textsubscript{2} and power as the predictors. In another study, Stjelja et al. [17] used data of office equipment and lighting electricity consumption as well as water consumption to predict the building floor occupancy with a supervised and unsupervised ML approach. With the unsupervised approach, the authors succeeded in categorising the days by their occupancy into three activity levels using only consumption data. Both of these studies were able to avoid training of the model, but ended up with lower resolution of either the inferred occupancy information (presence) or timescale (day).

A possible step in the right direction is to use some of the previously mentioned supervised machine learning methods which have shown good accuracy, such as a deep learning method by Chen et al. [35] with transfer learning. Transfer learning is a technique for extracting knowledge from extensive and known source datasets and using it to improve the learning of a model on the lesser-known target dataset [39]. In a case of deep learning this is usually done by transferring the pretrained weights of different neural network layers to the new (target) model. Once transferred, the target model can be fine-tuned (retrained) with a smaller dataset, therefore requiring significantly less amount of training data. The transfer learning research domain has been most extensively focused on image recognition, but lately, with the rise of wearables and activity recognition, there has been development of it for time-series data [40–43]. In a work by Hoelzemann et al. [42] the authors used a convolution-Long Short-Term Memory (LSTM) deep learning model and transfer learning of human activity recognition. The authors experimented with freezing and retraining of different layers of the model with different datasets, measured by different sensors. Some of the combinations were unsuccessful and some successful, such as the recommendation to keep the convolutional layer frozen, while retraining the LSTM part of the model.

Weber et al. [44] in their work used transfer learning in order to minimise training data needed for room occupancy detection. The authors used the same CDBLSTM algorithm as in [35], which was trained on a synthetic dataset produced by simulating occupancy and room CO\textsubscript{2} concentration. Model weights from the source model were used for a target model, where all layers were then retrained with a much smaller real-world dataset. The results show that with transfer learning of synthetic data it is possible to get good accuracy of occupancy detection even when only little ground-truth data for training is available.

A literature review has led to the formation of this work to provide one of the solutions in closing the following research gap:

- **Model scalability**
  - Occupancy ground truth: Collecting the ground truth for model training is problematic; can it be minimized if not removed?
  - Model robustness: Assess the model robustness if conditions in a room change, comparing to the condition during model training phase. How can the model robustness be increased?
  - Scalability without retraining: Is training always needed? If rooms and their conditions are similar, what is the performance concerning using a previously trained model?

- **Understanding results**
  - Model performance: Give insight concerning the occupancy prediction performance in practical terms.
To close the research gap, this paper proposes a deep transfer learning methodology for room occupancy prediction. The novelty of this work is the use of the transfer learning, where the knowledge on CO₂-based occupancy prediction from a model trained on a room with a larger available training dataset is transferred to a different target room from a different building using small available ground truth data for retraining. Furthermore, the performance of the target prediction model and its robustness are assessed using more understandable performance metrics developed from the perspective of building operations and real-estate domain use cases. Additionally, the transferred model is as well tested on two similar rooms in the target building, without additional retraining.

2. Methodology

2.1. Data Acquisition

The source and target dataset for this work were collected with the same type of the IoT sensor, measuring room air temperature and CO₂ in two meeting rooms located in different buildings, both located in Finland.

The source room is located in an office building (shown on the left in Figure 1). The source room is a typical small meeting room with seats for four persons and with an approximate size of 6 m². It is located in the middle of the floor and surrounded by an open office area. The ventilation system has constant airflow of 15 L/s consisting of 100% fresh outdoor air. The indoor climate IoT sensor was placed on the desk in the room. The camera for collecting ground truth occupancy was installed on the ceiling, just above the door. The ground truth for the source model was acquired using a video camera recording (the image was blurred for privacy purposes), which was used for manual counting of the number of people in the source room for the period 13 February–9 March 2020.

The target room for the transfer of the occupancy prediction model in this work is a large meeting room located in a hospital building. The target room (shown on the right in Figure 1) has a capacity for 12 people with an area of 21 m² and its ventilation system is designed to be of variable airflow rate. Room air temperature and CO₂ were collected with a IoT sensor which was placed on the conference desk. The ground truth for the transfer dataset was acquired using an infrared (IR) time-of-flight based people counting camera made by Irisys [45].

The measurement in the target room was done in two periods. The first period was from 22 March to 19 April 2021. During this period, the room ventilation system operated with a constant airflow rate, even though the system operated with a variable airflow rate. After this period, because of certain changes in the building, the ventilation operation was reset to perform with a variable airflow rate, as intended. The variable airflow rate in the room is controlled by the signal from three BMS sensors, PIR, CO₂ and room air temperature in order to keep both variables inside at wanted levels. During the measurements a problem with the variable airflow rate operation was noticed, where the source of the problem was too high air supply temperature. This manifested in airflow varying between 50 and 90 L/s, while the lowest airflow of 30 L/s was never reached, even when the room was unoccupied for longer periods. By increasing the airflow, the ventilation system tried to cool down the room and therefore there was no clear connection between the airflow and occupancy as such, making it more difficult to use the occupancy prediction model based on indoor air quality data. Nevertheless, this issue was seen as an opportunity to test the model robustness, which is addressed in Section 3.4, and the measurement continued for the second period from 10 May to 12 July 2021. It is important to note that the target room operated regularly during the COVID-19 pandemic, since it was used by hospital staff.
2.2. Overview of CDLSTM Model

The deep learning model used for predicting the occupancy from the indoor climate variables in this work is Convolutional Deep Long Short-Term Memory (CDLSTM). This model was inspired from the work by Chen et al. [35] since it outperformed other state-of-the-art machine learning models; it proved reliable even with the noisy data, while removing the need for a feature engineering process. The particular CDLSTM model uses a combination of different layers’ specific properties to classify occupancy states from raw data. Mainly the convolutional operation (C) is used for extracting features from raw data, while a DLSTM is used to understand the temporal dependence of the data.

In Figure 2, the architecture of the CDLSTM model used in this work is presented. The neural network model functions in the following way. Windows of room air temperature and CO$_2$ data enter first convolutional layer, which slides a filter window to extract the characteristics. Extracted features are then passed through the pooling operation, which compresses the features by removing less important local features. Following are the LSTM layers, commonly used with sequential data processing, which are good in learning long-term dependencies of features over a time sequence. A Deep LSTM (DLSTM) model, consisting of more than one LSTM layer is used in this work to get better representation of dependencies. After each LSTM layer, there is a dropout layer used for regularisation to solve the overfitting problem. Following the LSTM layers are fully connected dense layers, with their dropout layers as well. The fully connected part of the model is used to learn more abstract features in the data such as learning non-linear combinations of input features and preparing the features for their final representation. The final representation of the learned features from raw time-series data is then fed finally to the softmax classification layer. The softmax classification layer translates given features into classes, which are in the case of this work occupancy states.

![Figure 2. Architecture of the CDLSTM model used in this work with respective layer size.](image-url)
2.3. Proposed Transfer Method

Training a deep learning model such as the proposed CDLSTM requires significant amount of training data. Labelled training data is difficult to collect when it comes to room occupancy, especially the occupancy count or level. Therefore, in this work the transfer method of the CDLSTM model is studied.

The principle behind transfer learning is that a model is trained on a labelled and extensive dataset, called the source dataset. The acquired knowledge from training the source model can then be transferred to solve a similar problem on a target dataset. The weights of the trained CDLSTM model were then transferred to a target CDLSTM model which was retrained on a target dataset as shown in Figure 3. During the transfer some layers of the CDLSTM model were frozen (locked) and other layers were fine-tuned, using a smaller target dataset. Depending on the difference between the source and the target dataset, different model layers give different results, which is analysed in this work as well.

![Figure 3. Diagram of proposed transfer learning method.](image)

2.4. Experimental Setup

Data was gathered from IoT temperature and CO$_2$ sensors with a sampling rate of one minute, while the ground truth for the source dataset was acquired in three-minute intervals. Therefore, all other measurements were resampled to three minutes.

Before they were used for the deep learning model, the raw sensor data were pre-processed. Missing values, caused by connection problems or other IoT sensor problems, were interpolated using polynomial interpolation of second order since these missing periods were very short. Raw sensor data were noisy and through the experimentation we noticed that, using data as such, the accuracy of the CDLSTM model was lower and with longer computational time. Therefore, data were smoothed before the model input using the Kalman filtering with the Python package tsmoothie [46]. Kalman smoothing was performed on time-series level components.

During the initial experimentation, it was noticed that partly balancing the classes in the dataset also improved the results. Imbalanced classes in the dataset mean that certain classes are overrepresented compared to the other classes. In this case, the class of zero or empty room makes up a large majority of the dataset, while the occupied state makes up a smaller portion of the dataset. Training the deep learning model on the imbalanced dataset can lead to a model with lower sensitivity to classes which are underrepresented. In this
study, this issue was reduced by removing the periods when occupancy is expected to be zero, such as during the nighttime or the weekend.

Before using the data with neural model, it was normalized. Z-score normalization was used, where the values are normalized according to their mean and standard deviation. Data was normalized using the mean and standard deviation from the training data, while for the later use of the model with variable airflow, a sliding method was used as described in Section 3.4.

Normalized and preprocessed data were then ready to be used with the CDLSTM model. The model was created using the deep learning Python package Tensorflow/Keras [47] and was designed as follows and as presented in Figure 2. Room air temperature and CO\textsubscript{2} data enter the convolutional layer which was created with 100 output filters, a kernel size of three which specifies the 1D convolutional window and a pooling size of two, with ReLU activation function. Following is the deep LSTM model, consisting of three LSTM layers, with hidden layer sizes of 100, 150 and 200, respectively. After each layer there is a dropout layer served for regularization with masking probability of 0.5. Following the LSTM layers, there are two fully connected dense layers with hidden size 200 and 300. Another dropout layer is placed between the fully connected layers with a probability of 0.3. At the end, there is a softmax layer classifying the model outputs to classes, in our case occupancy states. The loss function used for the training of the CDLSTM model was cross-entropy, binary for occupancy detection and categorical for occupancy level prediction.

Another important setting for the CDLSTM model is the batch size and window sequence length and both vary depending on the sampling rate of the data on which the model was trained. For example, when using the dataset with three-minute sampling rate, a batch size of 16 and a sequence length of four were used. This means that one window was a size of four timesteps or twelve minutes which were fed to the model in batches of 16 windows at a time. The exact sizes of the batches and sequences were found with the trial and error method.

2.5. Performance Metric

To evaluate the performance of the model, the main metric used was the Matthews correlation coefficient (MCC) [48]. Similar work on occupancy estimation used a more traditional accuracy score or an F1 score, which happens to show wrongly optimistic results on overly imbalanced datasets. Since the rooms in question are the majority of the time vacant, the number of times a vacant class (zero) is present in the dataset is much higher compared to the number of times the room is occupied, which means that if a model would always predict the room is vacant, it would be awarded with a high accuracy score. The MCC score, on the other hand, gives more expected accuracy metrics for those cases: To achieve a high quality score, the classifier (such as the occupancy prediction model) has to make a prediction in the majority of positive and negative cases independently of their ratios in the overall dataset [48]. The MCC score lies in the range \([-1, +1\)], where \(-1\) and \(+1\) are reached in perfect misclassification and classification, respectively. On the other hand, an MCC score of zero indicating no relationship between the prediction and the truth. The main equation for calculating MCC is shown in Equation (1), where \(tp\) indicates true positives, \(fp\) false positives, \(tn\) true negatives and \(fn\) false negatives. MCC was implemented using the Python module Sci-\textit{kit learn} [49].

\[
\text{MCC} = \frac{tp \times tn - fp \times fn}{\sqrt{(tp + fp)(tp + fn)(tn + fp)(tn + fn)}}
\]  

(1)

On the other hand, the MCC score gives a very general metric, which is not straightforward for the potential use cases of this work. To make the results more understandable additional performance metrics were developed. Aside from the room occupancy in each timestep, from the perspective of optimising the ventilation system and space utilisation, it is useful to know the following about the meeting room occupancy: The time of the first
and last occupancy in the day and the duration of room occupancy. Therefore, the corresponding metrics were developed. The difference in the first or last occupancy of the day between the measured and predicted occupancy is shown as averaged throughout the period and in minutes. Duration of room occupancy is shown as a percentage of the average day being occupied.

3. Results

3.1. Analysis Workflow

The analysis of the proposed methodology is divided into six parts, which can be followed from the flowchart in Figure 4 and are described in the following sections. In Section 3.2 the source CDLSTM occupancy detection model was trained using data with long occupancy ground truth. In Section 3.3, a pretrained source model was transferred to a target room in a different environment with minimal occupancy ground truth data for model retraining. In Section 3.4, the operation of the ventilation system has experienced a change and the target detection model created in the previous part is used with the input data normalized using a sliding window to retain the performance. In Section 3.5, a target model’s performance is checked on the data from two similar rooms from the same building. In Section 3.6, the room efficiency is analysed to present the importance of occupancy information and at the same time the performance of the target model is interpreted. Finally, in Section 3.7, the transfer methodology is attempted predicting the room occupancy level.

![Figure 4. Workflow of the paper: Training of source occupancy prediction model, transferring the model, change of ventilation system operation (robustness check), using target model with similar rooms, room efficiency analysis and transfer of occupancy level prediction.](image-url)
3.2. Source Model Training

The source model for occupancy estimation was trained with data from a small meeting room located in an office building from the period between 13 February and 9 March 2020. Training of the model was done using the first 22 days (14 days occupied) and validation of the training was done with the following three days. The model was first trained using the raw sensor data in 30 epochs, which were enough for the model to converge and then in the second iteration smoothed data were used. Using raw and smoothed versions of the same data showed better performance during training and it served as an additional regularisation technique, making the model less prone to overfitting.

The trained source model showed an MCC score of 85% and the resulting comparison of predicted and measured occupancy detection can be seen in Figure 5.

![Figure 5. Three-day showcase of the source model occupancy detection performance and comparison with measured items.](image)

3.3. Transfer to the Target Room

In this section, the transfer of the source model to be used for occupancy detection with the data of a target room in a different building is assessed. The transfer learning of the source model for successful occupancy detection is analysed from the point of the CDLSTM specific process for retraining and minimum training data needed from the target room.

The process for retraining of the CDLSTM model considers which layers of the model are frozen and which are retrained with the new data and in which way. If a particular layer is not frozen, it can be trained from scratch (weights are reinitialised) or the layer weights can be transferred and then retrained.

Models used in this work and their characteristics and differences are explained in Table 1, where different layers of the CDLSTM model are visualised in Figure 2. The model that was trained using training data from the only target dataset is called Baseline model, while the model that was not retrained at all is Pure source model. The Transfer Method 1 is a model with pre-trained weights for all layers, which were retrained. Transfer Method 2 had a convolutional part of the model frozen, while LSTM and rest of the final layers were retrained. The Transfer Method 3 had the convolutional part and the first two LSTM layers of the model frozen, while the last (third) LSTM and the rest of the final layers were retrained. The Transfer Method 4 had all the other layers frozen during retraining, while the softmax classification layer was retrained.

The previously trained occupancy detection model (source model) is transferred to IAQ IoT sensor data from the hospital’s meeting room during the time that ventilation system was set to operate with a constant airflow rate setting, as explained in Section 2.1. Data for this analysis were collected in the period from 22 March until 2 May 2021. For (re)training of the model, different training durations from this period were used. Training period lengths used here were one, two, three and five working days. Additionally, in order to have more reliable results, two different training, validation and test sets were selected, serving as a cross-validation step. These periods can be seen in Table 2.
Table 1. Models used in this study and their characteristics. The baseline model is a model trained from scratch, where all layers are initialized, pure transfer is a source model used without retraining on a target dataset. Transfer methods 1, 2, 3 and 4 are a combination of frozen and trainable pre-trained transferred layers.

| Parameter                  | Baseline Model | Pure Transfer | Transfer Method 1 | Transfer Method 2 | Transfer Method 3 | Transfer Method 4 |
|----------------------------|----------------|---------------|-------------------|-------------------|-------------------|-------------------|
| Learning rate              | 0.001          | -             | 0.001             | 0.001             | 0.001             | 0.001             |
| Convolutional layer        | Initialized    | pre-trained (frozen) | pre-trained (trainable) | pre-trained (frozen) | pre-trained (frozen) | pre-trained (frozen) |
| First two LSTM layers      | Initialized    | pre-trained (frozen) | pre-trained (trainable) | pre-trained (trainable) | pre-trained (frozen) | pre-trained (frozen) |
| Third LSTM layer           | Initialized    | pre-trained (frozen) | pre-trained (trainable) | pre-trained (trainable) | pre-trained (trainable) | pre-trained (trainable) |
| Classification layer       | Initialized    | pre-trained (frozen) | pre-trained (trainable) | pre-trained (trainable) | pre-trained (trainable) | pre-trained (trainable) |

Table 2. The selected dates for training, validation and testing set.

| Period   | First Set | Second Set                        |
|----------|-----------|-----------------------------------|
| Train    | 22 March + (1, 2, 3, or 5 working days) | 12 April + (1, 2, 3, or 5 working days) |
| Validation | 29 March | 9 April |
| Test     | 30 March–18 April | 22 March–8 April |

The goodness of the models is first assessed with the MCC score and then the best cases are further analysed using the additional performance metrics. In Table 3, the MCC score of different models with different training data length is presented. Analysing the results from the training length perspective, it can be seen that, with already two days of ground-truth data with transfer learning, it is possible to get nearly as high the MCC score, as with a longer training period. Regarding the best transfer method to choose, it is either transfer method 1, 2 or 3. Which means that at least final LSTM, fully connected and classification layers should be retrained. From the table, it can be seen that Baseline model needs five days’ training data to perform and the proposed transfer method is performed with one day. Additionally, all methods show similar performance on both training and testing datasets, which is a sign of a good robustness of the proposed method.

Table 3. Different model MCC score of occupancy detection with regards to the length of training data.

| Detection | First Set | Second Set |
|-----------|-----------|------------|
| Baseline model | 0.67 | 0.80 | 0.70 | 0.75 | 0.78 |
| Pure source model | 0.71 | 0.71 | 0.71 | 0.70 | 0.70 | 0.70 |
| Transfer Method 1 | 0.77 | 0.82 | 0.82 | 0.76 | 0.80 | 0.79 | 0.80 |
| Transfer Method 2 | 0.78 | 0.81 | 0.82 | 0.78 | 0.79 | 0.78 | 0.79 |
| Transfer Method 3 | 0.77 | 0.78 | 0.78 | 0.74 | 0.81 | 0.80 | 0.81 |
| Transfer Method 4 | 0.75 | 0.75 | 0.75 | 0.72 | 0.75 | 0.75 | 0.81 |

For further analysis, five well-functioning cases were chosen for each set, which are highlighted in Table 3. The chosen cases were selected taking the highest MCC score for the shortest training length (1–2 days) and the longest period for comparison. Moreover, the baseline model, which is trained from scratch, is used for comparison. Selected cases are then further examined with additional performance metrics in order to understand what MCC score might mean for practical use and to select a final occupancy detection transfer model. This analysis can be seen in Table 4.

From the first set, the best result shows the case using five day data and Transfer Method 2, second to it is two day Transfer Method 1. Both of them have the same MCC.
score, but differ slightly in daily average occupied time compared to the ground truth and the first and the last detected occupancy of a day being similar. With the goal of minimizing the training data needed for transfer learning, two day Transfer Method 1 is selected as the optimal case. From the second set, using the same logic, the optimal case seems to be the two day Transfer Method 3. Which leads to the overall optimal case in this work Transfer Method 1 and the two day retraining from the first set, highlighted in Table 4. This case will be called the target model in the next sections. The target model’s performance is visualized in Figure 6, where it is compared to the measured ground truth occupancy.

### Table 4. Additional performance metrics for selected models of different training length and method.

| Training Days—Training Method | MCC  | Daily Average Occupied Time [%] | First Occupancy in a Day Average Difference [min] | Last Occupancy in a Day Average Difference [min] |
|-------------------------------|------|--------------------------------|-----------------------------------------------|-----------------------------------------------|
| Ground truth                  | -    | 7.5                            | -                                             | -                                             |
| First set                     | 1 day—Transfer Method 2 | 0.78                            | 9.7                                           | 10                                           |
|                               | 2 day—Baseline model    | 0.73                            | 11                                            | 5                                            |
|                               | **2 day—Transfer Method 1** | **0.82**                      | **8.9**                                       | **7**                                        |
|                               | 5 day—Baseline model    | 0.80                            | 9.4                                           | 5                                            |
|                               | 5 day—Transfer Method 2 | 0.82                            | 8.3                                           | 6                                            |

| Second set                    | Ground truth            | -                               | 12.1                                          | -                                            |
|                               | 1 day—Transfer Method 2 | 0.78                            | 14                                            | 8                                            |
|                               | 2 day—Baseline model    | 0.70                            | 14.5                                          | 7                                            |
|                               | **2 day—Transfer Method 3** | **0.81**                      | **12**                                        | **9**                                        |
|                               | 5 day—Baseline model    | 0.78                            | 12                                            | 4                                            |
|                               | 5 day—Transfer Method 3 | 0.81                            | 11.9                                          | 8                                            |

**Figure 6.** Three-day showcase of the selected target model performance and comparison with measured occupancy.

### 3.4. Robustness Check—Change in Ventilation System Operation Mode

The robustness of ML models for predicting the room occupancy using indoor climate variables is important for using them commercially, as certain changes might happen with HVAC system operation. In this section the performance of the transferred detection model is checked after a change in ventilation system operation occurs. In the target room, during the model transfer, ventilation operated with a constant airflow rate of about 90 L/s. Change in the operation mode was the demand-controlled variable air flow operation activation, where airflow was modulated between about 55 and 90 L/s, as explained in Section 2.1.

Robustness check on data from variable airflow rate operation was done using data measured in two parts, during spring (10–30 May 2021) and summer (20 June–11 July 2021). During the process, it became clear that the way z-score normalization is performed is important. For stationary time series, the z-score normalization statistics derived from the training dataset can also be used for the test data. However, for non-stationary time series this is not suitable [52]. In this case introducing the variable airflow makes the mean and standard deviation of CO₂ and temperature time series more susceptible to change over time. Therefore, a sliding window approach was used, where normalization statistics were calculated on an arbitrarily chosen 14 day window prior to the day being predicted.
Table 5 presents results from the robustness check of the transferred detection model from spring and summer periods and with two normalization approaches. Results are presented using MCC for model accuracy and the additional performance metrics. In both periods, the sliding window normalization produced significantly better results than using the normalization from the training phase. Using the sliding window normalization, the detection performance of the model is close to the original performance of the detection model during ventilation operation with the constant airflow rate. The largest difference in performance is in detection of the first daily occupancy. The difference in performance between spring and summer times probably comes from more variability in airflow and slightly higher room temperature during the summer period.

Table 5. Performance of target model after the room ventilation changed from constant to variable airflow rate operation mode.

| Test Period          | Case                      | MCC | Daily Average Occupied Time [%] | First Occupancy in a Day Average Difference [min] | Last Occupancy in a Day Average Difference [min] |
|----------------------|---------------------------|-----|-------------------------------|-----------------------------------------------|-----------------------------------------------|
|                      | Ground truth              | -   | 5.7                           | -                                             | -                                             |
| 10–30 May 2021       | Target model Normalized on training | 0.71 | 8.4 | -74 | 21 |
|                      | Target model Sliding normalization | 0.78 | 6.9 | -27 | 2 |
|                      | Ground truth              | -   | 11.0                          | -                                             | -                                             |
| 20 June–11 July 2021 | Target model Normalized on training | 0.5 | 25.0 | -103 | 91 |
|                      | Target model Sliding normalization | 0.72 | 12.0 | -28 | 7 |

3.5. Usage in Similar Rooms

While the proposed transfer method eases the scalability of the machine learning based occupancy prediction method by minimising the need for acquiring occupancy ground truth, it does not solve the problem if this process is needed even with similar rooms. Therefore, the target occupancy detection model has been tested in two similar rooms from the same hospital building. Room 2 is a meeting room of the same size as the target room, served by the same air handling unit, while room 3 is a slightly smaller (16.5 m$^2$) break room for hospital staff, ventilated by a different air handling unit. During the two-week test period, both rooms operated with constant and with maximum designed airflow (85 and 90 L/s). The same type of IoT sensor was placed centrally on the desk as in the target room.

Ground truth occupancy detection data were collected during two weeks in May 2021 in both rooms. Using the target model, the performance of the model to these rooms was tested and results can be seen in Table 6 and in Figure 7. For both rooms, sliding window normalization was used as discussed in the Section 3.4. Presented results indicate that the performance of the detection model is on par with the performance on the target room to which it was transferred, the MCC score being higher for room 2 and slightly lower with room 3. One potential reason why the results are better than with the target room might be because of regularity of room occupancy schedules. The target room has a more irregular occupancy schedule, while rooms 2 and 3 are usually occupied with more regular schedules. Moreover, additional rooms have had longer vacant time between the occupied times, which leaves time for the ventilation system to clean the room air, making it easier for the model to make a correct prediction.
Table 6. Performance of target model on the two similar rooms (to the target room) in the same building.

| Room          | Case     | MCC Occupied Time [%] | Daily Average Occupied Time | First Occupancy in a Day | Last Occupancy in a Day | Average Difference [min] |
|---------------|----------|-----------------------|-----------------------------|--------------------------|-------------------------|--------------------------|
| Room 2        | Ground   | 0.85                  | 5.3                         | -                        | -                       | -                        |
| (meeting room)| Model    | 0.85                  | 5.2                         | -7                       | -28                     |                          |
| Room 3        | Ground   | 0.83                  | 10.1                        | -                        | -                       | -                        |
| (break room)  | Model    | 0.83                  | 8.2                         | 7                        | -21                     |                          |

Figure 7. Three-day showcase of the target occupancy detection model performance on room 2 (meeting room).

3.6. Room Efficiency Analysis

To present one potential use of the occupancy prediction model and its value regarding energy and space inefficiency, an additional analysis was made using data from the target room. First, looking at the space usage efficiency, it was possible to see already from Tables 4 and 5 that average daily occupied time in those rooms for the tested periods is not more than 11% of the total ventilated time (24 h).

Regarding the ventilation energy usage, a simplistic analysis of it is done for the current case and for the ideal control scenario. In the ideal control scenario, it was assumed that during the present occupancy the ventilation airflow rate is as it was in reality, while during the unoccupied phase it works at the minimal designed airflow rate of 30 L/s. Additionally, after the room has become vacant, the ventilation continues as normal until the measured room CO₂ concentration is not under an arbitrarily chosen baseline of 430 ppm. The ideal control scenario is calculated based on the measured ground truth occupancy and on the predicted occupancy using the target occupancy detection model.

The calculation of energy uses two simplified methods based on the airflow ratio. The first one is for calculating the heat transfer between outdoor air temperature and supply air temperature and another is for calculating the power needed to transfer the required air volume. Both of them produce a relative result compared to the current case. The first equation is based on the heat transfer, while the fan power equation comes from fan laws [53] where we assume that the fan efficiency remains constant and therefore it is possible to use Equation (3), where \( \phi \) signifies heat flow, \( q \) airflow and \( P \) fan power for the ideal and for the current case.

\[
\phi_{\text{ideal}} = \frac{q_{\text{ideal}}}{q_{\text{current}}} \cdot \phi_{\text{current}} \tag{2}
\]

\[
P_{\text{ideal}} = \left( \frac{q_{\text{ideal}}}{q_{\text{current}}} \right)^3 \cdot P_{\text{current}} \tag{3}
\]

Simplified energy analysis for the ideal control scenario results are presented in Table 7. From the table, there are three things possible to conclude. First, results between the ideal control scenario based on the ground truth occupancy and a control based on predicted occupancy using the target model show small differences, confirming the good accuracy of the prediction model. Second, it can be seen that the performance of the control scenario based on the prediction model is similar in cases with constant and variable airflow. Finally,
this simplified energy analysis shows how inefficient the current room ventilation operation mode is compared to the actual room occupancy, in both airflow settings. It showed that the fan power needed could be one tenth compared to the constant airflow period or one fifth during the current implementation of variable airflow. On the other hand, the heat flow could be halved.

### Table 7. Simplified energy analysis for ideal control scenario based on the measured and predicted occupancy detection, presented for the two periods, one with constant and another with variable airflow.

|                              | Ground Truth Occupancy | Predicted Occupancy |
|------------------------------|------------------------|---------------------|
|                              | Fan Power (%P_{current}) | Heat Flow (%φ_{current}) | Fan Power (%P_{current}) | Heat Flow (%φ_{current}) |
| Constant Airflow Rate        |                         |                      |                         |                      |
| (30 March–18 April 2021)     | 10.7                    | 38.7                | 11                      | 38.9                 |
| Variable Airflow Rate        |                         |                      |                         |                      |
| (10–30 May 2021)             | 18                      | 50.2                | 18.2                    | 50.3                 |

### 3.7. Transfer of Occupancy Level Prediction

In this section, the same transfer learning methodology was applied in order to acquire more details about room occupancy using the same data. Ideally, it would be possible to provide a small amount of ground truth for model transfer and to get an accurate room occupancy count. For a successful classification, the machine learning model needs to see enough examples of each class, which could be a problem concerning using the model to estimate the number of persons in the room. In practice, this would mean that, for every possible number of persons in a room, there would need to be recording of the CO₂ and room air temperature. This is why the model was used for predicting the occupancy level, instead of counting. To do that, a source model from the source room was trained to predict the number of people and another model to predict the level of occupancy.

The same model architecture was used as with the detection, but with the difference that here there is a multi-class problem, instead of binary classification. This was solved by using different size CDLSTM output layer (maximum number of occupancy levels) and using categorical cross-entropy for loss function. Since the source room has a maximum occupancy of four persons in terms of occupancy level, it was divided into three levels: Zero (0 persons), low (1–2 persons) and high (3–4 persons). Data from the same period as in Section 3.2 were used to train the source model for occupancy level prediction and its MCC score was 0.82.

Following the same principle as with the occupancy detection, transfer learning is applied to the occupancy level prediction of the target meeting room in the hospital building. Transfer learning for occupancy level prediction was tested in the second period or during the time when the ventilation system operated with variable airflow. The reason for this is that, during this period, accurate ground truth data for the number of persons in the room were available.

Occupancy of the target room was divided into three levels—zero, low and high occupancy. Low occupancy meant one to three occupants and above was considered as high occupancy. This is because, during the measured period, most of the occupancy that occurred was between one and three occupants.

The MCC score for the baseline (model trained without transfer learning) and the transfer model is presented in Table 8, for different lengths of training with target room data. The transfer model score is already highest with one day retraining, while two and three days have lower score. Only after retraining with five days data, performance comes close to the performance of model with one day retraining. The possible reasons for this are discussed in the discussion section. In general the MCC score is lower for the occupancy level than comparing it with detection during the same period in variable airflow conditions (Table 5). However, comparing the transfer model scores to the baseline, it is noticeable
how transfer learning increases the performance of the occupancy level prediction model as well.

In Figure 8 occupancy level prediction using the transfer model with one day retraining data is presented. The figure confirms what the MCC score has showed; a limited accuracy with predicting occupancy level. Comparing this figure with Figures 5 and 6 it is visible that the occupancy detection performance is slightly lower as well.

Table 8. Occupancy level prediction performance analysis as the MCC score with different amounts of data for training of baseline and transfer model.

| Method          | 1 Day | 2 Day | 3 Day | 5 Day |
|-----------------|-------|-------|-------|-------|
| Baseline model  | 0.37  | 0.42  | 0.5   | 0.63  |
| Transfer model  | 0.67  | 0.62  | 0.64  | 0.65  |

Figure 8. Showcase of the occupancy level prediction using the transfer model with one day retraining data.

4. Discussion

As mentioned in the literature review, plenty of studies were done exploring the accuracy of the occupancy inferring methods from indoor climate variables. On the other hand, very little was done on the scalability of those methods and therefore they have not found their way in the industry. Research gaps, which are aimed to be closed with this work are obstacles preventing the wider use of ML-based occupancy prediction methods.

Having available room occupancy information would be beneficial in several built environment domains, consequently reducing carbon footprint while increasing the occupant’s comfort. A method based on commonly available data in buildings, such as room air temperature and CO$_2$ could be a potential solution. Optimising and controlling the HVAC system schedule and setpoints could improve the indoor comfort of occupied spaces while reducing the energy waste in unoccupied spaces. Real-estate management would have better insight into how their spaces are used and be better able to make decisions that would reduce inefficiency and increase usability and satisfaction. Consequently better space usage of existing buildings would decrease the need for new construction and in that way have a positive impact on climate change.

The main obstacle in the wider adoption of such a method was found to be the collection of ground truth data for the training of ML models. In this work, we sought to avoid or at least minimize the main obstacle by using the transfer learning method. Having a previously trained model on a large dataset and having a small amount of ground truth data (two days) from a different room, it was possible to get good occupancy detection accuracy.

The second obstacle considers the model robustness, or making sure the model performance stays at the same level if conditions in the room change. In this work, a target model prepared with the indoor climate data measured during constant airflow conditions was tested after the conditions in the room changed to variable airflow. The test showed that it is
possible to keep the performance on the same level, but the selection of an appropriate data normalization process is important. Here, a simple sliding window z-score normalization was used, but more advanced approaches exist such as adaptive normalization [52,54]. Since advanced normalization approaches were out of the scope of this work, they were not explored here, but are an important issue for future research.

While proving that collecting just two days’ room occupancy ground truth is enough for a detection model of good accuracy, carrying that out for every room is not convenient. Therefore, the data from two similar rooms of the same hospital building as the target room were used to check the target model performance. The results of this analysis have shown that performance of the model in similar rooms was on the same level as in the target room. Consequently, this means transfer learning could be performed on representative rooms in a building and the resulting models could be put to use in similar rooms directly. Eventually, a library of pretrained models could be created for different room sizes, types and ventilation systems, removing even the need for any transfer learning.

Knowing whether rooms are occupied or not is important, but there is additional value in knowing the level of room occupancy at the time. Following this work’s methodology, a transfer learning of room occupancy level prediction was performed as well. The results here were not as successful as with detection, where the most successful model was with one days’ data and providing more days for retraining just decreased the model performance. Since occupancy level prediction is actually a multi-class classification, the model needs to “see” enough examples of each class during (re)training to make an accurate prediction. Therefore, with the current methodology, even a small target dataset needs to contain enough samples of each class. Unfortunately, in this work the analysis of occupancy level prediction was tested only during the time when the room’s ventilation was working with the variable airflow rate. The variable airflow rate brings additional dynamics to indoor climate and occupancy and to perceive it requires more training data. Future work on this matter should try using the variable airflow rate as a signal to the prediction model as well as performing this methodology with constant airflow data.

One of the obstacles of larger adoption of similar methods is the understanding of the results, or what they mean for potential use cases. Therefore in this work, in addition to the MCC score, additional performance metrics was developed. Comparing the percentage of the actual occupied time to the predicted time gives an idea of the suitability of the occupancy prediction model output for optimizing the ventilation system. Similarly, knowing the time difference between the actual and the predicted first or last occupancy of the day is also important. Results of the selected target model have shown that the difference of the first and last occupancy in a day is about seven minutes between actual and predicted occupancy. The actual daily average occupied time was about 7.5%, while the target model predicted 8.9%. These results can be considered good, especially compared to the current state where occupancy information is not available at all, leading to the case such as in the target room where the system is operating non-stop. As previously mentioned, the same metrics can be useful for space usage optimisation purposes.

Furthermore, in this work a simplified energy analysis was done as if the room occupancy detection would be used for controlling ventilation in the room. Besides showing how wasteful the current situation is, this analysis has proved that a target model exhibits good performance in an understandable way. Estimated energy savings if the ventilation would be controlled by the occupancy prediction of the proposed target model or by measured ground truth occupancy were similar.

The most important limitation of this study lies in the occupancy ground truth collection. The main tool for this was the use of manual camera counts for the source dataset and the IR camera as the source for the target dataset. Based on Kuutti et al. [55], IR cameras’ true positive rate ranges between 0.67 and 1.00, with the average being 0.89. To verify that this issue would not be a source of large errors, data from IR cameras was manually checked from the presence and noise sensor data. The limitation where it is difficult to get an accurate ground truth shows the importance of this methodology, where the training
data is to be minimized. Another limitation is the work on the occupancy level prediction, as previously mentioned. During the period of constant airflow rate in the target room there were problems with the collection of the ground truth for the number of persons in the room. Therefore, a clear picture of the proposed methodology regarding the occupancy level prediction is not available.

Even using the proposed methodology of this work, a wider usage of the ML-based occupancy prediction from indoor climate variables is not going to happen. To achieve that, further work should continue with removing the need for ground truth collection completely, if possible. Additionally, methods for monitoring the drift of such models should be explored, where the biggest issue is not having ground truth at all. Finally, occupancy detection is necessary, but it is more valuable to know the number of people in the room or at least the level of room occupancy. Minimizing the training data needed for this purpose is difficult, since machine learning models need to see enough examples of every class. This is an important issue for future research and state-of-the-art methods from the computer science field should be explored.

5. Conclusions

The main goal of the presented study was to explore the scalability of a deep learning-based method for inferring room occupancy information from indoor climate measurements. Previous research has focused on testing different ML-based methods in order to get as high accuracy as possible, but since it is difficult to collect labelled training data, the methods were not used widely. In this study, a transfer learning method has been applied and it showed that it is possible to create an occupancy detection model of good accuracy with only two days of ground truth data, instead of several weeks.

Model robustness is another issue, which might prevent the ML-based model from wider use. To check the robustness, the transferred target model was applied in conditions different than during the transfer process. The target room, which was initially operated with constant airflow rate ventilation, changed to operate with variable airflow rate. The model robustness analysis showed that the occupancy detection model performance remained nearly the same, as long as sliding window normalization was applied. Furthermore, to make sure this methodology is scalable, the performance of the transferred target model was checked (without retraining) with two additional but similar rooms from the same building. In this case, the model performance remained on the same level as well.

On the other hand, using transfer learning to predict the level of room occupancy was slightly less successful since multi-class classification requires enough data samples of each class for good performance. It is important to note that the occupancy level prediction done here was with variable air flow, while using the same approach in constant airflow conditions could have given better results. Therefore, this is one limitation of this work as it has not fully explored the ability of scalable occupancy level or count prediction.

Finally, we tried to explain the performance of the target model from the perspective of the use case, such as ventilation system optimization. This was done by comparing the actual daily average occupied time and the first or last occupancy of the day between actual and predicted data. Additionally, a simplified energy analysis of potential savings was done in cases where predicted and actual detected occupancy was used for controlling the ventilation system. A simplified analysis proved the performance of the transfer deep learning method for occupancy detection and justified the need for room occupancy information by showing the existing room’s ventilation system inefficiency.

Exploring the scalability and robustness of the room occupancy prediction method in this work will hopefully contribute to better availability of occupancy information. Continuing this work and creating a library of pretrained models for different room conditions could make this possible on a larger scale, enabling a better HVAC system and space utilization and finally resulting in better performing buildings with lower carbon footprint.
Author Contributions: Conceptualization, D.S.; Data curation, D.S.; Formal analysis, D.S.; Investigation, D.S.; Supervision, J.J. and R.K.; Writing—original draft, D.S.; Writing—review & editing, J.J. and R.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Business Finland through the project Mad@Work.

Acknowledgments: The authors would like to thank Aalto University and Granlund Oy for making this research possible.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations
The following abbreviations are used in this manuscript:

- AI: Artificial Intelligence
- BMS: Building Management System
- CAV: Constant Air Volume
- CDBLSTM: Convolutional Deep Bidirectional Long Short-Term Memory
- CDLSTM: Convolutional Deep Long Short-Term Memory
- DLSTM: Deep Long Short-Term Memory
- HMM: Hidden Markov Model
- HVAC: Heating Ventilation Air Conditioning
- IAQ: Indoor Air Quality
- IoT: Internet of Things
- IR: InfraRed
- k-NN: k-Nearest Neighbours
- LSTM: Long Short-Term Memory
- MCC: Matthews correlation coefficient
- ML: Machine Learning
- PIR: Passive Infrared
- ReLU: Rectified Linear Unit
- RNN: Recurrent Neural Network
- SDE: Stochastic Differential Equations
- SVM: Support Vector Machines
- TVOC: Total Volatile Organ Compound
- VAV: Variable Air Volume

References
1. European Union. Directive 2018/844 of the European Parliament and of the Council of 30 May 2018 Amending Directive 2010/31/EU on the Energy Performance of Buildings and Directive 2012/27/EU on Energy Efficiency; Official Journal of the European Union: Strasbourg, France, 2018; pp. L 156/75–L 156/91.
2. UN Environment and International Energy Agency. Towards a Zero-Emission, Efficient, and Resilient Buildings and Construction Sector; Technical Report; UN Environment and International Energy Agency: London, UK, 2017.
3. Pérez-Lombard, L.; Ortiz, J.; Pout, C. A Review on Buildings Energy Consumption Information. Energy Build. 2008, 40, 394–398. [CrossRef]
4. Choi, J.H.; Yang, B.; Yu, C.W. Artificial Intelligence as an Agent to Transform Research Paradigms in Building Science and Technology. Indoor Built Environ. 2021, 1420326X2110176. [CrossRef]
5. Wang, Z.; Liu, J.; Zhang, Y.; Yuan, H.; Zhang, R.; Srinivasan, R.S. Practical Issues in Implementing Machine-Learning Models for Building Energy Efficiency: Moving beyond Obstacles. Renew. Sustain. Energy Rev. 2021, 143, 110929. [CrossRef]
6. Vigna, I.; Balest, J.; Pasut, W.; Pernetti, R. Office Occupants’ Perspective Dealing with Energy Flexibility: A Large-Scale Survey in the Province of Bolzano. Energies 2020, 13, 4312. [CrossRef]
7. Bedford, M.; Harris, R.; King, A.; Hawkeswood, A. Occupier Density Study; British Council for Offices (BCO): London, UK, 2013.
8. Naor, M.; Pinto, G.D.; Hakakian, A.I.; Jacobs, A. The Impact of COVID-19 on Office Space Utilization and Real-Estate: A Case Study about Teleworking in Israel as New Normal. J. Facil. Manag. 2021. [CrossRef]
9. Boland, B.; Smet, A.D.; Palter, R.; Sanghvi, A. Reimagining the Office and Work Life after COVID-19; McKinsey & Company: Chicago, IL, USA, 2020; pp. 1–5.
10. Chen, Z.; Jiang, C.; Xie, L. Building Occupancy Estimation and Detection: A Review. Energy Build. 2018, 169, 260–270. [CrossRef]
11. Dodier, R.H.; Henze, G.P.; Tiller, D.K.; Guo, X. Building Occupancy Detection through Sensor Belief Networks. Energy Build. 2006, 38, 1033–1043. [CrossRef]
12. Raykov, Y.P.; Ozer, E.; Dasika, G.; Boukouvalas, A.; Little, M.A. 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, Heidelberg, Germany, 12–16 September 2016. [CrossRef]

13. Kleiminger, W.; Beckel, C.; Staake, T.; Santini, S. Occupancy Detection from Electricity Consumption Data. In Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings, Roma, Italy, 11–15 November 2013; pp. 1–8. [CrossRef]

14. Kleiminger, W.; Beckel, C.; Santini, S. Household Occupancy Monitoring Using Electricity Meters. In Proceedings of the UbiComp 2015: 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan, 7–11 September 2015; pp. 975–986. [CrossRef]

15. Razavi, R.; Gharipour, A.; Fleury, M.; Justice, I. Energy & Buildings Occupancy Detection of Residential Buildings Using Smart Meter Data: A Large-Scale Study. Energy Build. 2019, 183, 195–208. [CrossRef]

16. Yang, L.; Ting, K.; Srivastava, M.B. Inferring Occupancy from Opportunistically Available Sensor Data. In Proceedings of the 2014 IEEE International Conference on Pervasive Computing and Communications, PerCom 2014, Budapest, Hungary, 24–28 March 2014; pp. 60–68. [CrossRef]

17. Stjelja, D.; Jokisalo, J.; Kosonen, R. From Electricity and Water Consumption Data to Information on Office Occupancy: A Supervised and Unsupervised Data Mining Approach. Appl. Sci. 2020, 10, 9089. [CrossRef]

18. Dong, B.; Andrews, B.; Lam, K.P.; Höynck, M.; Zhang, R.; Chiou, Y.S.; Benitez, D. An Information Technology Enabled Sustainability Test-Bed (ITEST) for Occupancy Detection through an Environmental Sensing Network. Energy Build. 2010, 42, 1038–1046. [CrossRef]

19. Candanedo, L.M.; Feldheim, V.; Deramaix, D. A Methodology Based on Hidden Markov Models for Occupancy Detection and a Case Study in a Low Energy Residential Building. Energy Build. 2017, 148, 327–341. [CrossRef]

20. Wang, S.; Jin, X. CO2-Based Occupancy Detection for on-Line Outdoor Air Flow Control. Indoor Built Environ. 1998, 7, 165–181. [CrossRef]

21. Sun, Z.; Wang, S.; Ma, Z. In-Situ Implementation and Validation of a CO2-based Adaptive Demand-Controlled Ventilation Strategy in a Multi-Zone Office Building. Build. Environ. 2011, 46, 124–133. [CrossRef]

22. Cali, D.; Matthes, P.; Huchtemann, K.; Streblow, R.; Müller, D. CO2 Based Occupancy Detection Algorithm: Experimental Analysis and Validation for Office and Residential Buildings. Build. Environ. 2015, 86, 39–49. [CrossRef]

23. Wolf, S.; Cali, D.; Krogtie, J.; Madsen, H. Carbon Dioxide-Based Occupancy Estimation Using Stochastic Differential Equations. Appl. Energy 2019, 236, 32–41. [CrossRef]

24. Lam, K.P.; Höynck, M.; Dong, B.; Andrews, B.; Chiou, Y.S.; Zhang, R.; Benitez, D.; Choi, J. Occupancy Detection through an Extensive Environmental Sensor Network in an Open-Plan Office Building. IBPSA Build. Simul. 2009, 145, 1452–1459.

25. Dong, B.; Lam, K.P. A Real-Time Model Predictive Control for Building Heating and Cooling Systems Based on the Occupancy Behavior Pattern Detection and Local Weather Forecasting. Build. Simul. 2014, 7, 89–106. [CrossRef]

26. Chen, Z.; Zhu, Q.; Masood, M.K.; Soh, Y.C. Environmental Sensors-Based Occupancy Estimation in Buildings via IIMM-MLR. IEEE Trans. Ind. Inform. 2017, 13, 2184–2193. [CrossRef]

27. Amayri, M.; Arora, A.; Ploix, S.; Bandhyopadhyay, S.; Ngo, Q.D.; Badarla, V.R. Estimating Occupancy in Heterogeneous Sensor Environment. Energy Build. 2016, 129, 46–58. [CrossRef]

28. Mamidi, S.; Chang, Y.; Maheswaran, R. Improving Building Energy Efficiency with a Network of Sensing, Learning and Prediction Agents. In Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems—Innovative Applications Track (AAMAS 2012), Valencia, Spain, 4–8 June 2012; pp. 45–52. [CrossRef]

29. Chen, Z.; Masood, M.K.; Soh, Y.C. A Fusion Framework for Occupancy Estimation in Office Buildings Based on Environmental Sensor Data. Energy Build. 2016, 133, 790–798. [CrossRef]

30. Zuraimi, M.S.; Pantazaras, A.; Chaturvedi, K.A.; Yang, J.J.; Tham, K.W.; Lee, S.E. Predicting Occupancy Counts Using Physical and Statistical Co2-based Modeling Methodologies. Build. Environ. 2017, 123, 517–528. [CrossRef]

31. Szczeruck, A.; Maciejewska, M.; Pietrucha, T. Occupancy Determination Based on Time Series of CO2 Concentration, Temperature and Relative Humidity. Energy Build. 2017, 147, 142–154. [CrossRef]

32. Alam, A.G.; Rahman, H.; Kim, J.K.; Han, H. Uncertainties in Neural Network Model Based on Carbon Dioxide Concentration for Occupancy Estimation. J. Mech. Sci. Technol. 2017, 31, 2573–2580. [CrossRef]

33. Masood, M.K.; Soh, Y.C.; Chang, V.W. Real-Time Occupancy Estimation Using Environmental Parameters. In Proceedings of the International Joint Conference on Neural Networks, Killarney, Ireland, 12–17 July 2015; Volume 2015, pp. 1–8. [CrossRef]

34. Masood, M.K.; Jiang, C.; Soh, Y.C. A Novel Feature Selection Framework with Hybrid Feature-Scaled Extreme Learning Machine (HFS-ELM) for Indoor Occupancy Estimation. Energy Build. 2018, 158, 1139–1151. [CrossRef]

35. Chen, Z.; Zhao, R.; Zhu, Q.; Masood, M.K.; Soh, Y.C.; Mao, K. Building Occupancy Estimation with Environmental Sensors via CDBLSTM. IEEE Trans. Ind. Electron. 2017, 64, 9549–9559. [CrossRef]

36. Dong, B.; Kjærgaard, M.B.; De Simone, M.; Gunay, H.B.; O’Brien, W.; Mora, D.; Dziedzic, J.; Zhao, J. Sensing and Data Acquisition. In Exploring Occupant Behavior in Buildings: Methods and Challenges; Springer International Publishing: New York, NY, USA, 2017; pp. 77–105. [CrossRef]

37. Dong, B.; Yan, D.; Li, Z.; Jin, Y.; Feng, X.; Fontenot, H. Modeling Occupancy and Behavior for Better Building Design and Operation—A Critical Review. Build. Simul. 2018, 11, 899–921. [CrossRef]
38. Mora, D.; Fajilla, G.; Austin, M.C.; De Simone, M. Occupancy Patterns Obtained by Heuristic Approaches: Cluster Analysis and Logical Flowcharts. A Case Study in a University Office. *Energy Build.* 2019, 186, 147–168. [CrossRef]
39. Pan, S.J.; Yang, Q. A Survey on Transfer Learning. *IEEE Trans. Knowl. Data Eng.* 2010, 22, 1345–1359. [CrossRef]
40. Cook, D.; Feuz, K.D.; Krishnan, N.C. Transfer Learning for Activity Recognition: A Survey. *Knowl. Inf. Syst.* 2013, 36, 537–556. [CrossRef]
41. Khan, M.A.A.H.; Roy, N.; Misra, A. Scaling Human Activity Recognition via Deep Learning-Based Domain Adaptation. In Proceedings of the 2018 IEEE International Conference on Pervasive Computing and Communications, PerCom 2018, Athens, Greece, 21–23 March 2018. [CrossRef]
42. Hoelzemann, A.; Van Laerhoven, K. Digging Deeper: Towards a Better Understanding of Transfer Learning for Human Activity Recognition. In Proceedings of the International Symposium on Wearable Computers, ISWC, Online, 12–17 September 2020; pp. 50–54. [CrossRef]
43. Li, F.; Shirahama, K.; Nisar, M.A.; Huang, X.; Grzegorzek, M. Deep Transfer Learning for Time Series Data Based on Sensor Modality Classification. *Sensors* 2020, 20, 4271. [CrossRef]
44. Weber, M.; Doblander, C.; Mandl, P. Towards the Detection of Building Occupancy with Synthetic Environmental Data. *arXiv* 2020, arXiv:2010.04209.
45. Irisys. Vector 4D Footfall People Counter System | Irisys. Available online: https://www.irisys.net/products/vector4d-people-counter (accessed on 24 January 2022).
46. Cerliani, M. Tsmoothie. 2020. Available online: https://pypi.org/project/tsmoothie/ (accessed on 12 July 2021).
47. Chollet, F. Keras. 2015. Available online: https://keras.io (accessed on 12 July 2021).
48. Chicco, D.; Jurman, G. The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation. *BMC Genom.* 2020, 21, 6. [CrossRef] [PubMed]
49. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* 2011, 12, 2825–2830.
50. Metrics and Scoring: Quantifying the Quality of Predictions. Available online: https://scikit-learn/stable/modules/model_evaluation.html (accessed on 29 December 2021).
51. Saha, H.; Florita, A.R.; Henze, G.P.; Sarkar, S. Occupancy Sensing in Buildings: A Review of Data Analytics Approaches. *Energy Build.* 2019, 188–189, 278–285. [CrossRef]
52. Ogasawara, E.; Martinez, L.C.; De Oliveira, D.; Zimbrão, G.; Pappa, G.L.; Mattoso, M. Adaptive Normalization: A Novel Data Normalization Approach for Non-Stationary Time Series. In Proceedings of the 2010 International Joint Conference on Neural Networks (IJCNN), Barcelona, Spain, 18–23 July 2010; pp. 1–8.
53. Papar, R.; Szady, A.; Huffer, W.D.; Martin, V.; McKane, A. Increasing Energy Efficiency of Mine Ventilation Systems. In *8th U.S. Mine Ventilation Symposium*; University of Missouri-Rolla: Rolla, MO, USA, 1999; p. 7.
54. Passalis, N.; Tefas, A.; Kannaiinen, J.; Gabbouj, M.; Iosifidis, A. Deep Adaptive Input Normalization for Time Series Forecasting. *IEEE Trans. Neural Netw. Learn. Syst.* 2020, 31, 3760–3765. [CrossRef]
55. Kuutti, J.; Blomqvist, K.H.; Sepponen, R.E. Performance of Commercial Over-Head Camera Sensors in Recognizing Patterns of Two and Three Persons: A Case Study. In Proceedings of the 2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), Amman, Jordan, 3–5 December 2013.