Transfer learning for smart buildings: A critical review of algorithms, applications, and future perspectives

Advances in Applied Energy 5 (2022) 100084

Contents lists available at ScienceDirect

Advances in Applied Energy

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

Transfer learning for smart buildings: A critical review of algorithms, applications, and future perspectives

Giuseppe Pinto [a], Zhe Wang [b,d], Abhishek Roy [c], Tianzhen Hong [b], Alfonso Capozzoli [a,*]

[a] *Department of Energy, TEBE Research Group, BAEDA lab, Politecnico di Torino*
[b] *Building Technology and Urban Systems Division, Lawrence Berkeley National Laboratory, One Cyclotron Road, Berkeley, CA 94720, United States*
[c] *University of California Berkeley, USA*
[d] *Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology*

A R T I C L E I N F O

Keywords:
Transfer learning
Deep learning
Machine learning
Smart buildings
Building energy management

A B S T R A C T

Smart buildings play a crucial role toward decarbonizing society, as globally buildings emit about one-third of greenhouse gases. In the last few years, machine learning has achieved a notable momentum that, if properly harnessed, may unleash its potential for advanced analytics and control of smart buildings, enabling the technique to scale up for supporting the decarbonization of the building sector. In this perspective, transfer learning aims to improve the performance of a target learner exploiting knowledge in related environments. The present work provides a comprehensive overview of transfer learning applications in smart buildings, classifying and analyzing 77 papers according to their applications, algorithms, and adopted metrics. The study identified four main application areas of transfer learning: (1) building load prediction, (2) occupancy detection and activity recognition, (3) building dynamics modeling, and (4) energy systems control. Furthermore, the review highlighted the role of deep learning in transfer learning applications that has been used in more than half of the analyzed studies. The paper also discusses how to integrate transfer learning in a smart building's ecosystem, identifying, for each application area, the research gaps and guidelines for future research directions.

1. Introduction

With the introduction of carbon neutral targets, the energy system is undergoing profound changes. The centralized architecture of the electrical grid, in which fossil-fuel plants generate electricity, are being shifted towards a distributed architecture that leverages renewable energy sources, energy storage, and optimal management [1]. As a result, the shift from fossil fuel to renewable energy sources will be accelerated even more [2]. Buildings account for about 40% of total energy use, and Grid-interactive Efficient Buildings (GEB) [3] play a key role in the energy transition benefiting building owners, occupants, and the electric grid [4]. Smart meters and grid automation technology account for the majority of digital grid investment, and digitalisation offers enormous potential to improve the efficiency, flexibility, and resilience of energy systems. In this perspective, coordinated optimization and collaborative management of various smart grid actors will become a trend [5], paving the way for power systems to fully enter the digital era, leveraging new technologies such as the Internet of Things (IoT), real-time monitoring and control, peer-to-peer energy, and smart contracts [6] to ensure more efficient, reliable, and sustainable electricity dispatch. To support the penetration of GEBs in the smart grid, several countries plan a mass deployment of advanced metering infrastructure (AMI) [7] that can provide useful insights on user's consumption patterns and distributed energy resources (DER) production. In this context, smart meters, artificial intelligence (AI), and connectivity can be used in different phases of the building cycle to improve power demand and generation forecasts, and to extract energy usage patterns. Data-driven models can be used for building operation and control, automating decision-making and easing the deployment of energy management at scale.

1.1. Motivation and scope of the review

The growing adoption of automation and control systems, information, and communication technologies (ICT) and IoT sensors in smart buildings has contributed recently to an unprecedented availability of long-term monitoring data related to the energy performance and indoor quality of the built environment. As a consequence, complex building-related databases are more available than in the past, and their exploration provides the opportunity to effectively characterise the actual building energy behaviour and to optimise the performance of its energy systems during operation. The size, complexity, and heterogeneity of building-related databases make it increasingly necessary for the introduction of frameworks based on an effective coupling of machine

* Corresponding author.
E-mail address: alfonso.capozzoli@polito.it (A. Capozzoli).

https://doi.org/10.1016/j.adapen.2022.100084
Received 8 November 2021; Received in revised form 4 January 2022; Accepted 4 January 2022
Available online 5 January 2022

Nomenclature

Acronyms

AMI	Advanced Metering Infrastructure
AI	Artificial Intelligence
ANN	Artificial Neural Network
ASO	Automated System Optimization
BNN	Bayesian Neural Network
BAS	Building Automation System
CNN	Convolutional Neural Network
DTL	Decoder Transfer Learning
DL	Deep Learning
DNN	Deep Neural Network
DRL	Deep Reinforcement Learning
DR	Demand Response
DSM	Demand Side Management
DER	Distributed Energy Resources
EIS	Energy Information System
FDD	Fault Detection and Diagnosis
GRU	Gated Recurrent Units
GEB	Grid-interactive Efficient Building
HVAC	Heating Ventilation and Air Conditioning
IEEE	Institute of Electrical and Electronics Engineers
IoT	Internet of Things
LSTM	Long Short-Term Memory
ML	Machine Learning
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
MPC	Model Predictive Control
MLP	Multi-Layer Perceptron
NILM	Non-Intrusive Load Monitoring
PCC	Pearson Correlation Coefficient
PV	Photovoltaic
PCA	Principal Component Analysis
RNN	Recurrent Neural Network
RL	Reinforcement Learning
RMSE	Root Mean Square Error
RBC	Rule Based Controller
SSS	Sub-keyword Synonym Searching
SMAPE	Symmetric Mean Absolute Percentage Error
TL	Transfer Learning

learning and energy domain knowledge to extract ready-to-implement strategies for optimizing the building energy performance exploiting this massive amount of building-related data [8]. Machine learning (ML) methods proved to be effective tools to valorize the knowledge that can be extracted from data to support the optimization of building energy performance [9,10] and have been applied in various applications across the building life cycle to improve building performance and occupant comfort and health [11]. The most promising applications for building energy management are: the prediction of energy demand required for the efficient operation of a building [12], the optimization of building operation [13–15], the detection and commissioning of operational failures of building equipment [16,17], the energy benchmarking analysis [18,19], the characterisation of energy demand profiles [20–22], and the assessment of the impact of user behaviour [23]. Currently the building industry is exploiting ML with the progressive introduction of energy management and information systems (EMIS), which enhance and integrate the functionalities of traditional building automation system (BAS) to analyse and control building energy use and system performance.

The EMIS includes the energy information systems (EIS) and fault detection and diagnostic (FDD) systems, which are aimed to support the decisions by means of informative solutions (one-way communication with the BAS), and the automated system optimization (ASO) tools, which optimize the control settings (two-way communication paradigm with the BAS) [24]. EIS include both predictive and descriptive analytics for performing tasks such as load prediction, anomaly detection, advanced benchmarking, load profiling, and schedule optimisation of building energy systems. FDD systems help to detect abnormal system states whose identification and diagnosis can lead to significant energy savings. The 2016–2020 Smart Energy Analytics Campaign [24] assessed the costs and benefits of EMIS installations for a number of different building types and sizes, including 104 commercial organizations across the United States and more than 6500 buildings. By the second year of installation, a median annual energy savings of three percent with EIS, and of nine percent with FDD tools, was evaluated, supporting the use of such technologies in buildings. ASO includes predictive and adaptive control solutions (e.g., model predictive control or reinforcement learning-based control) to optimise the settings of building energy systems considering the trade-off between multiple and contrasting objectives for enhancing energy flexibility, renewable energy integration, and building performance, achieving an annual cost reduction that ranges from 11% to 16% with respect to conventional building energy management systems (BEMS) [25].

One area demonstrating large potential benefits in the ASO field is the development and application of reinforcement learning-based building controls to optimize energy efficiency and energy flexibility [26,27]. To fully exploit the flexibility associated with buildings, the scale of analysis was shifted from individual devices or buildings to the so-called cluster of buildings [28], communities [29], districts [30], or integrated microgrid [31], in which multi-agent control techniques have recently proven to be effective. In this context, to overcome the computational complexity associated with the control of such environments, machine learning and deep learning techniques [32,33] have been used to lighten large-scale building simulations.

However, collecting and preparing a large amount of high quality data to train machine learning algorithms is time consuming and not always feasible, as most buildings lack reliable sensing or metering systems or lack the IT infrastructure to collect and store the data. Therefore, machine learning techniques have not yet been widely adopted by the industry, and real applications are often limited to research or early stage demonstration projects. To address this gap, one key technique needed is to transfer machine learning models trained and validated for buildings with rich data to buildings with limited or poor data. With this motivation in mind and given that existing efforts such as [34] and [35] reviewed TL for specific topics only, activity recognition and demand response respectively, and no in-depth literature review on transfer learning in smart buildings exists, we aimed to conduct a comprehensive and structured review on transfer learning. This study focused on how transfer learning is used for modeling, prediction, performance diagnosis, and performance optimization of commercial and residential buildings.

1.2. Structure of the review

This review effort aimed to provide insights into significant questions on transfer learning for buildings research and applications. In particular, Fig. 1 unfolds the structure of the review and the associated research questions that were considered for the work. The first part of the paper covers motivation behind the use of transfer learning in buildings, trying to understand how TL can integrate into the buildings research ecosystem and when to use it. Then attention is shifted towards background and applications, studying the different methods to transfer knowledge and which are the most common applications. The review also covers the questions of which algorithms, tools, and common platforms have been used and how it is possible to assess TL performance. Lastly, challenges and future directions are identified, providing useful guidelines for researchers.

The paper is structured as follows. First, Section 2 provides the fundamental background of the review, by introducing notations, defini-

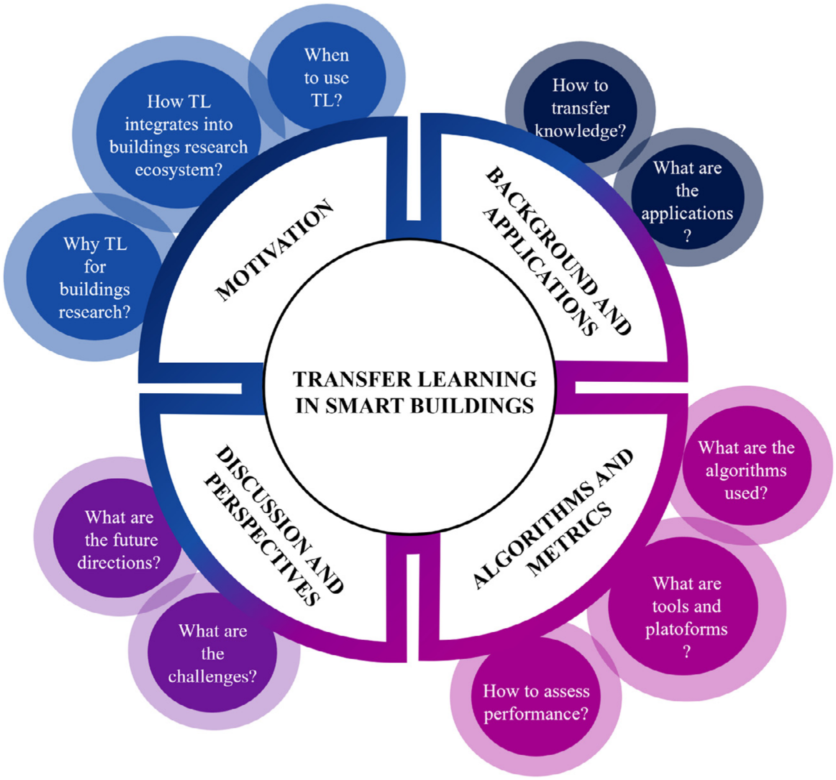

Fig. 1. Research questions and paper overview

Table 1
Transfer learning notations.

Notation	Definition
$\mathcal{X}$	Input feature space
$\mathcal{Y}$	Label space
$\mathcal{T}$	Predictive learning task
Subscript S	Denotes source
Subscript T	Denotes target
$P(X)$	Marginal distribution
$P(Y\|X)$	Conditional probability
$P(Y)$	Label distribution
$\mathcal{D}_S = f\{\mathcal{X}_S, P(X_S)\}$	Source domain data
$\mathcal{D}_T = f\{\mathcal{X}_T, P(X_T)\}$	Target domain data

tions, and terminology of transfer learning. Then, Section 3 describes the search method applied in the review process. Section 4 categorizes the reviewed papers based on the TL applications previously cited, also classifying papers according to the ML techniques and the type of TL used. Section 5 discuss the main findings of the review, along with potential future directions and applications. Finally, conclusions are given in Section 6, covering the research questions in Fig. 1, with condensed answers.

2. Background on transfer learning

2.1. Transfer learning

In this subsection, the notations used within the paper are reported in Table 1, and transfer learning related definitions are described for convenience. In addition, transfer learning categorization, examples, and reviews in the built environment are provided.

The starting point for the definition of transfer learning is the description of the concepts of "domain" and "task," reported below according to Pan and Yang [36].

Definition 1. Domain: a domain $\mathcal{D}$ consists of two components, a feature space $\mathcal{X}$ and a marginal probability distribution $P(X)$ where $X = \{x_1, \dots, x_n\} \in \mathcal{X}$.

For example, if the learning task is the electrical load prediction of a building, modelled as a regression problem, then $\mathcal{X}$ is the space of all influencing variables, (e.g., external temperature, occupancy, historical load), while x_i represents the i^{th} influencing variables and X a specific learning sample.

Definition 2. Task: a task consists of two components, a label space Y and an objective predictive function $f(\cdot)$ (denoted by $\mathcal{T} = \{Y, f(\cdot)\}$), which is not observed but can be learned from the training data, represented by a pair $\{x_i, y_i\}$, where $x_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$. The function $f(\cdot)$ is used to approximate the conditional probability $P(y|x)$ and predict the corresponding label of a new instance x.

Considering the same application of building load prediction, Y is a continuous space with the possible values of the building load.

Lastly, transfer learning definition is provided and, to ease the comprehension, the definition only considers the case of one source domain $\mathcal{D}_S$ and one target domain $\mathcal{D}_T$, since it represents the most common research problem. In particular, we denote the source domain data as $\mathcal{D}_S = \{(x_{S1}, y_{S1}), \dots, (x_{Sn_S}, y_{Sn_s})\}$, where $x_{Si} \in X_S$ is the data instance and $y_{Si} \in Y_S$ is the corresponding output. Similarly, the target domain data are denoted as $\mathcal{D}_T = \{(x_{T1}, y_{T1}), \dots, (x_{Tn_T}, y_{Tn_T})\}$, where $x_{Ti} \in X_T$ and $y_{Ti} \in Y_T$ are the corresponding outputs. In many cases, transfer learning provides advantages where $0 \le n_T \ll n_S$.

Definition 3. Transfer Learning: Given a source domain $\mathcal{D}_S$ and learning task $\mathcal{T}_S$, a target domain $\mathcal{D}_T$, and a learning task $\mathcal{T}_T$, transfer learning aims to help improve the learning of the target predictive function $f(\cdot)$ in $\mathcal{D}_T$ using the knowledge in $\mathcal{D}_S$ and $\mathcal{T}_S$, where $\mathcal{D}_S \neq \mathcal{D}_T$, or $\mathcal{T}_S \neq \mathcal{T}_T$.

A schematic representation of the application of transfer learning in buildings is shown in Fig. 2, highlighting the differences with respect to a classical machine learning problem, while below examples

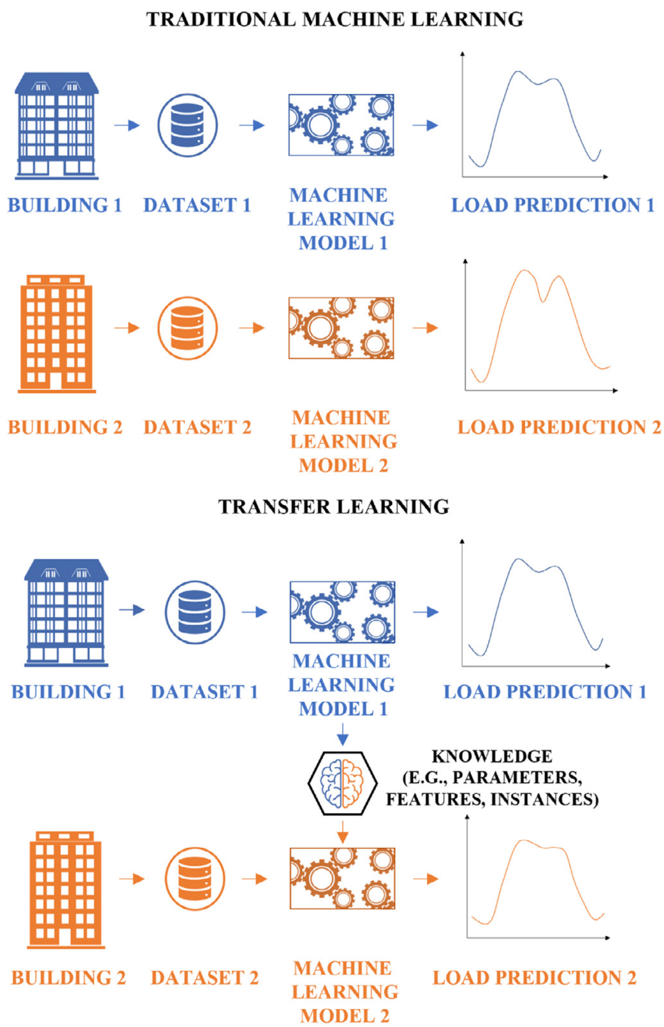

Fig. 2. Schematic representation of machine learning and transfer learning problem in buildings.

related to the built environment are discussed to ease the comprehension of transfer learning. Using Definition 1, a domain is a pair $\mathcal{D} = \{\mathcal{X}, P(X)\}$. Thus, two domains are different if they have (1) different feature spaces $\mathcal{X}_S \neq \mathcal{X}_T$ or (2) different marginal probability distributions $P(X_S) \neq P(X_T)$. Considering forecasting of building load, case (1) corresponds to when the two buildings have different energy systems, such as the presence/or absence of PV systems. Case (2) can be found in buildings with two different occupancy schedules or located in different climates.

Like the domain, learning tasks also can be different in two ways. They can either have (1) different label spaces, $\mathcal{Y}_S \neq \mathcal{Y}_T$ or (2) different conditional probability distributions

$P(Y_S|X_S) \neq P(Y_T|X_T)$. Recalling the building example, case (1) represents the situation where the aim of the source domain is to predict building electrical load, while the target domain focuses only on thermal related load prediction. Case (2) corresponds to the situation where the source and the target building are very unbalanced in terms of power usage.

2.2. *Categorization of transfer learning*

Transfer learning problems can be categorized based on different combinations among source and target domains, tasks, and solutions adopted.

2.2.1. *Label classification*

The first classification is based on the task similarity and label availability, and categorizes transfer learning in three subsettings: inductive transfer learning, transductive transfer learning, and unsupervised transfer learning, depending on the label availability, hereafter called *label classification.*

- In the inductive transfer learning setting, the target task is different from the source task ($\mathcal{T}_S \neq \mathcal{T}_T$), no matter whether the source and target domains are the same or not. In that case, inductive transfer learning aims to help improve the learning of the target predictive function $f_T(\cdot)$ in $\mathcal{D}_T$, using the knowledge in $\mathcal{D}_S$ and $\mathcal{T}_S$, exploiting some labelled data in the target domain to induce the objective predictive model $f_T(\cdot)$.
- In the transductive transfer learning setting, the source and target tasks are the same ($\mathcal{T}_S = \mathcal{T}_T$), while the source and target domains are different ($\mathcal{D}_S \neq \mathcal{D}_T$). In this case, transductive transfer learning aims to improve the learning of the target predictive function $f_T(\cdot)$ in $\mathcal{D}_T$, using the knowledge in $\mathcal{D}_S$ and $\mathcal{T}_S$. In the transductive transfer learning, no labelled data in the target domain are available, but labelled data in the source domain are available.
- Lastly, in the unsupervised transfer learning setting, the target task is different from but related to the source task ($\mathcal{T}_S \neq \mathcal{T}_T$). However, the unsupervised transfer learning focuses on solving unsupervised learning tasks in the target domain. In that case, no labelled data are available in both source and target domains during the training process.

2.2.2. *Space classification*

A further categorization is based on the similarity between source and target spaces (feature and label). Hereafter, this will be called *space classification.* In the space classification, if $\mathcal{X}_S = \mathcal{X}_T$ and $\mathcal{Y}_S = \mathcal{Y}_T$, the scenario is classified as homogeneous transfer learning. Otherwise, if $\mathcal{X}_S \neq \mathcal{X}_T$ and/or $\mathcal{Y}_S \neq \mathcal{Y}_T$, the scenario is classified as heterogeneous transfer learning.

It is important to note that label and space classifications can coexist, since the domain and tasks are characterized by feature and label space, but also by their conditional distribution. To introduce the problem of transfer learning in the context of buildings, Table 2 summarizes how the two classifications can coexist with examples taken from the reviewed works, highlighting different methods and applications analysed in smart buildings.

2.2.3. *Solution classification*

Lastly, transfer learning also can be categorized based on the strategy adopted to share the knowledge, i.e., data instance-based, model parameter-based, feature representation-based, and relational knowledge-based strategies; this classification hereafter will be called *solution classification.* To support easy understanding of those concepts, definitions are illustrated for the application of building load prediction, as reported in Fig. 3. The load prediction is a multi-variate time series problem that can highly benefits from transfer learning, working either with data (time series), reweighting or extracting features (instance-based and feature representation-based) or directly adapting model parameters (parameter-transfer or relational knowledge-based).

- The instance-based TL approach assumes that certain parts of the data in the source domain $\mathcal{D}_S, \mathcal{Y}_S$ can be reused for learning in the target domain $\mathcal{D}_T$ when some historical target task data $\mathcal{Y}_T$ are available. Instance-based TL select and reweight data in the source domain to facilitate the data-driven task in the target domain. This technique is typically used when the data variables are the same across different domains, and it increases the amount of data available for training without substantially changing the algorithm itself.
- The feature representation-based TL extracts and exploits features to map instances from the source and target domains to improve

Table 2
Joint categorization of label and space classification according to domain and task with practical examples related to machine learning applications in energy and buildings.

	Domain	Task	Example
Homogeneous Inductive Learning	$\mathcal{X}_S = \mathcal{X}_T$	$\mathcal{T}_S \neq \mathcal{T}_T$	Transfer learning is used to enhance building monthly electric load prediction leveraging information from similar buildings in different districts, that exhibits a different conditional probability [37].
Heterogeneous Inductive Learning	$\mathcal{X}_S \neq \mathcal{X}_T$	$\mathcal{T}_S \neq \mathcal{T}_T$	Transfer learning is used to fine-tune a pretrained neural network initially built to perform multi-class classification, to increase the accuracy of a prediction model for building temperature setback identification [38].
Homogeneous Transductive Learning	$\mathcal{X}_S = \mathcal{X}_T, P(X_S) \neq P(X_T)$	$\mathcal{T}_S = \mathcal{T}_T$	Transfer learning is used for improving the accuracy of home activity estimation by exploiting the data of a source house applied to a target house with no labelled data [39].
Heterogeneous Transductive Learning	$\mathcal{X}_S \neq \mathcal{X}_T$	$\mathcal{T}_S = \mathcal{T}_T$	Transfer learning is used to predict building dynamics by extracting features from multiple households in an online fashion, without having access to labelled data [40].

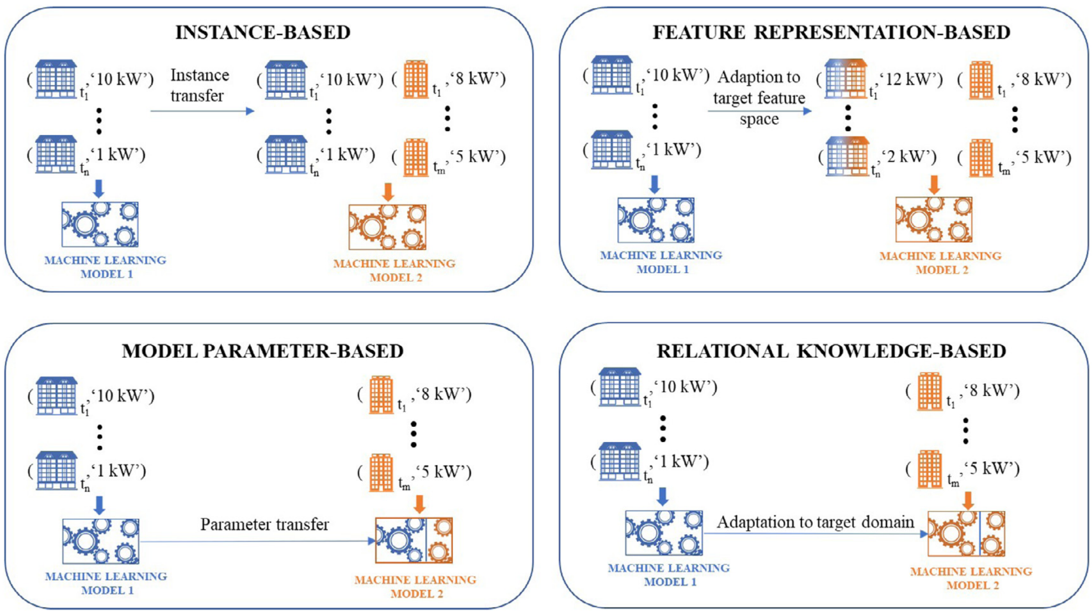

Fig. 3. Illustration of transfer learning according to the solution classification using the example of building load prediction (adapted from [41])

training on the target task. Feature representation-based TL is further classified in two approaches: the first approach transforms the features of the source through reweighting, to match the target domain (asymmetric feature transformation) more closely. The second approach discovers underlying meaningful structures between the domains to find a common latent feature space (symmetric feature transformation).

- The model parameter-based TL assumes that the source tasks and the target tasks share some parameters or prior distributions of the hyper-parameters of the models (e.g., neural networks). The latter is based on the assumption that models developed for similar tasks will have similar model parameters or hyper-parameters. The knowledge gained from the source task is transferred to another task as shared model weights in this case. The recent success of deep learning has spawned a new type of transfer learning, network-based transfer learning [42], which belongs to the parameter-based transfer learning category and can be further classified based on the strategy used to share model parameters.
  - The first way is to use the pretrained model for feature extraction. In this case, the weights of some layers are fixed, except for the input/output layer, which are domain dependent and need to be fine-tuned using target data. The main advantage is represented by the reduced amount of data needed to train the model, as well as the possibility to exploit data from different domains.
  - The second way is to use the pretrained model for weight initialization and fine-tuning. In such a case, the weights of the pretrained model are used for initialization purposes only and can be adjusted through a fine-tuning process.

  Fig. 4 displays the two strategies adopted to perform parameter-based TL, henceforth called *feature-extraction* and *weight-initialization.*
- Relational knowledge-based TL is generally used with multi-relational datasets. The underlying assumption is that some relationship among the data in the source and target domains are similar. Thus, the knowledge to be transferred is the relationship among the data.

Table 3
Different solutions used in different label settings.

	Inductive	Transductive	Unsupervised
Instance-based	✓	✓	
Feature-based	✓	✓	✓
Parameter-based	✓		
Relation-based	✓		

Table 3 reports the cases where different solutions are used for each label setting. It can be observed that instance-based TL can be used within both inductive and transductive settings, while feature-based TL is the only solution that can be employed within all the settings. On the other hand, parameter-based and relation-based TL can only be used in an inductive settings, showing the necessity of labelled data to apply TL. The literature review revealed that inductive TL is the most popu-

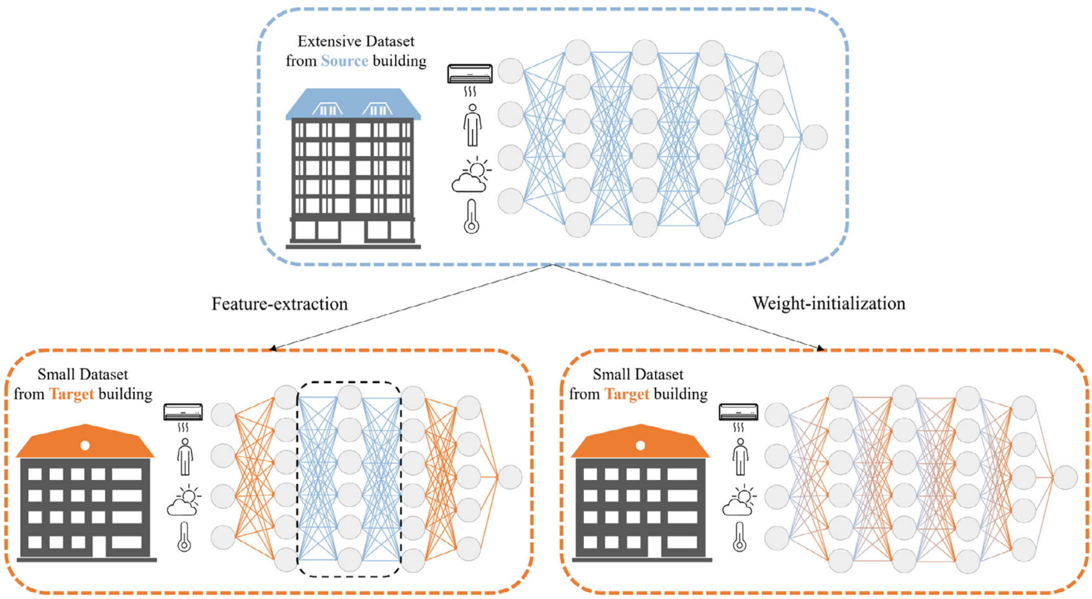

Fig. 4. Parameter-based transfer learning further classified in feature-extraction (left) and weight-initialization (right)

lar approach, while unsupervised TL is only adopted in the context of feature-based TL.

Although the first review of transfer learning of Pan and Yang [36] is recent, the evolution of deep learning rapidly increased and broadened the application in smart buildings of such methods, leading to the necessity of other reviews. In particular, Weiss et al. [43] provided an overview of TL with recent applications (up to 2016), while Day and Khoshgoftaar [44] provided a deep analysis of heterogeneous TL. More recently, Zhuang et al. [45] provided a comprehensive survey about TL from the perspectives of data and model, and Liang et al. [46] described the recent advances in TL. Tan et al. [42] analysed the role of deep TL, proposing a specific categorization. Lastly [47,48] reviewed the phenomena of negative transfer, together with the main factors leading to negative transfer, and which algorithms can help to mitigate it.

2.2.4. *Transfer learning in reinforcement learning*

Despite providing a complete classification for transfer learning in supervised and unsupervised settings, Pan and Yang [36] mentioned that their review excluded reinforcement learning, due to the inherent difference from supervised and unsupervised learning. Recently, several reviews, including Lazaric et al. [49] and Taylor and Stone [50], analysed the applications of transfer learning in the context of reinforcement learning. Furthermore, Zhu et al. [51] provided an overview on the application of TL in the context of deep reinforcement learning (DRL). Zhu et al. [51] classified TL methods based on the type of knowledge transferred, introducing five types of TL for DRL: i) reward shaping, ii) learning from demonstrations, iii) policy transfer, iv) intertask mapping, and v) representation transfer. A detailed introduction to reinforcement learning can be found in [52].

To provide a complete overview of the applications of transfer learning in smart buildings, the present work employed the same classification of Pan and Yang [36], adapting it for the case of RL.

The main differences with respect to the previously introduced definition lie in the fact that in a RL problem the conditional probability of the task is influenced by both the transition function and the reward function, which affect the agent's behaviour. Moreover, with respect to a traditional machine learning problem, in a reinforcement learning setting the input feature space ($\mathcal{X}$) and label space ($\mathcal{Y}$) commonly refers to as state space ($\mathcal{S}$) and action space ($\mathcal{A}$).

In the context of a smart building, different parts of the environment can alter the underlying transition function, thus affecting the conditional probability of the control policy (e.g., different weather, different energy systems, different user behaviour). Furthermore, in RL the reward function is computed after an interaction with the environment, no longer making it necessary to use labels. According to these differences, a new definition of task is provided.

Definition 4. Task: in a reinforcement learning setting, a task consists of three components, an action space A, a reward function r and a transition $f(\cdot)$, denoted by $\mathcal{T} = \{A, r, f(\cdot)\}$. The function $f(\cdot)$ is used to approximate the conditional probability $P(s_{t+1}, r_t | s_t, a_t)$ and select the optimal action. The goal of the agent is to find a control policy able to maximize the expected reward.

It should be noticed that since the transition function $f(\cdot)$ depends on states, actions and reward, in a transfer learning setting it will necessarily be different. Therefore, in a RL problem, dealing with the same task ($\mathcal{T}_S = \mathcal{T}_T$) implies dealing with the same action space A and reward function r, thus allowing to reuse the label classification previously introduced (inductive and transductive). Recalling the space classifications in Section 2.2.1 and the newly introduced label classification, Table 4 describes the different cases that can be encountered during the transfer of reinforcement learning in smart buildings. The present work refers to this classification when encountering application of TL for RL in building energy systems. Lastly, looking at solution classifications, policy transfer in DRL falls in the category of parameter-based TL.

3. Method

The novelty of the topic required the application of a specific methodology, described below, to include as much as possible all relevant works in the field of transfer learning applied to buildings. The first step was to use the Scopus search engine to identify and select relevant papers using the keywords:

$TS1$ = Transfer learning & Building & Energy

Furthermore, sub-keyword synonym searching (SSS) [53] was performed to cover the most common terminology of ML in buildings, coupling it with TL. As described in [53], the aim of the SSS methodology is to exhaust relevant papers by effectively searching literature using effective keywords, synonyms, and their combinations. As an example, "building load prediction" can often be referred to as "energy forecast in buildings," despite being the same research field. Therefore, this paper

Table 4
Joint categorization of label and space classification according to domain and task with practical examples related to reinforcement learning in buildings.

	Domain	Task	Example
Homogeneous Inductive Learning	$\mathcal{S}_S = \mathcal{S}_T \wedge \mathcal{A}_S = \mathcal{A}_T$	$r_S \neq r_T$	Reward function of the control problem changes from building energy cost minimization to peak load shaving.
Heterogeneous Inductive Learning	$\mathcal{S}_S \neq \mathcal{S}_T \vee \mathcal{A}_S \neq \mathcal{A}_T$	$r_S \neq r_T \vee \mathcal{A}_S \neq \mathcal{A}_T$	The energy system changes (i.e., actions or states change accordingly) but the reward function is the same, or the reward function changes together with the energy system and the action space.
Homogeneous Transductive Learning	$\mathcal{S}_S = \mathcal{S}_T \wedge \mathcal{A}_S = \mathcal{A}_T \wedge P(\mathcal{S}_S) \neq P(\mathcal{S}_T)$	$r_S = r_T$	The same controller including states, actions and reward is deployed with a different time-schedule of electricity price tariff.
Heterogeneous Transductive Learning	$\mathcal{S}_s \neq \mathcal{S}_T$	$r_S = r_T \wedge \mathcal{A}_s = \mathcal{A}_T$	The controller is deployed with the same actions and reward, but with a different state-space to consider different energy systems in the building.

Table 5
Summary of studies related to energy systems control.

ID	Objectives	ML problem	Dataset	Methods	Label	Space	Solution	NN
[64]	The paper aims to optimize the scheduling of a microgrid exploiting knowledge of another microgrid	Regression	Measurement	DDPG	Transductive	Homogeneous	Parameter	Weights
[66]	The work aims to optimize the operational planning of a battery using the knowledge of another battery in a different building	Classification	Measurement	Fitted Q-iteration,K-Shape, MILP	Transductive	Homogeneous	Parameter	Weights
[68]	The work uses transfer learning to share the knowledge of the optimal action of an HVAC from one building to another, to reduce energy consumption, costs and increase comfort	Classification	Measurement	DQN	Transductive	Heterogeneous	Parameter	Weights
[69]	The paper has the aim to increase the convergence rate of RL applied in new home by using the knowledge of a similar home with transfer learning	Regression	Both	PPO	Transductive	Homogeneous	Parameter	Weights
[70]	The work deals with potential scalability issues associated to RL considering grouping dependent electrical devices and applying TL, speeding up the process	Classification	Simulation	DPG,DQN	Transductive	Homogeneous	Parameter	Weights
[67]	The work examines the effects of TL on varying the spatial dimensions or geographical locations when learning policies for HVAC control	Classification	Measurement	Tabular Q-Learning	Transductive	Homogeneous	Parameter	Weights
[65]	The work studies different ways to apply TL in multiple homes, to reduce energy costs and increase PV self-consumption in a microgrid	Classification	Simulation	DQN	Transductive	Homogeneous	Parameter	Weights

exploited domain expertise to identify the main applications of ML in buildings (Load Prediction, Occupancy Detection, Building Dynamics, Building Systems Control), further described in Section 2, while not losing generality thanks to the previous search (TS1). The second search combined for each of the four applications, the sub-keywords C1 and C2 (if present). The full list of sub-keywords C1 and C2 is shown in Fig. 5 and an example is reported below.

For the application of the “load prediction,” the SSS methodology combines different keywords, exploring all the possible combinations among the four subsets: {Transfer Learning, Domain Adaptation}, {Building, Home, District, City}, {Consumption, Electricity, Energy, Load} and {Forecasting, Prediction}. As a result, the iterative procedure search for 2x4x4x2 = 64 keywords for load prediction, 24 keywords for occupancy detection, 32 keywords for building dynamics, and 32 keywords for building systems control.

$$TS2 = \{A \ \& \ B \ \& \ (C1 \ \& \ C2)\}$$

Lastly, papers were manually filtered according to their relevance, without including citations thresholds due to the novelty of the topic, while preferring journal articles over conference articles when dealing with similar topics. The extracted papers were manually reviewed, categorized, and organized into five categories, henceforth called: “Load Prediction”, “Occupancy & Activities”, “Building Dynamics”, “Systems Control”, and “Other”, with the latter including all the possible applications of TL in buildings not included in the previous categories. A graphical representation is reported in Fig. 6.

4. Results of the review

This section describes in detail the results of the review. First, a metadata analysis is performed, to assess trends related to the applications of transfer learning in smart buildings. Then, the main application areas are identified and described, followed by an overview of the approaches used. Lastly, the most used tools and metrics are discussed.

4.1. Metadata analysis

Fig. 7 shows a summary of the reviewed literature. Fig. 7 (a) displays the keywords cloud of the reviewed articles, where the size of the text indicates the frequency of a keyword being used. ”Transfer learning,” ”building,” and ”energy” were among the most popular words representing the main topic of the analysis. Other words such as ”forecasting,” ”deep learning,” ”neural network,” and ”occupancy” highlight both topics and methods used within the reviewed articles. Fig. 7 (b) shows the geographical distribution of researchers studying TL. A total of 31 unique countries were found, with top contributors represented by China (20.3%), the United States (15.1%), and Australia, Canada, and Japan (5.4% each). Lastly, Fig. 7 (c) displays the journal article (conference papers, posters, and book chapters were excluded) distribution over the year and per journal. Among the 77 reviewed papers, 47 were published in 26 different journals. The main journals include *Energy and Buildings* (12.7%), *Energies* (10.6%), *Association for Computing Machinery (ACM)* (10.6%), and *Institute of Electrical and Electronics Engineers (IEEE)*

Table 6
Summary of studies related to occupancy detection and activity recognition.

ID	Objectives	ML problem	Dataset	Methods	Label	Space	Solution	NN
[86]	Transfer learning is used to increase the accuracy of activity recognition in buildings with respect to unsupervised learning models	Classification	Measurement	PCA	Transductive	Heterogeneous	Feature	
[78]	transfer learning is used to develop an intelligent human counting system,further utilized for energy optimization in smart building management	Regression	Measurement	CNN	Inductive	Heterogeneous	Parameter	Weights
[103]	The work aims at collecting data from multiple houses and use TL to perform activity recognition in houses with unlabelled data, increasing the accuracy of the models	Classification	Measurement	1 Nearest Neighbor (1NN)	Transductive	Heterogeneous	Feature	
[102]	The paper exploit activity recognition to evalute the heat gains and further use a LSTM to simulate building dynamics. Lastly, it uses these information to perform control reducing the energy consumption of a classroom	Classification	Measurement	CNN, LSTM	Inductive	Heterogeneous	Parameter	Weights
[77]	Transfer learning is used to overcome data unavailability for occupancy prediction task in a room exploiting data of other rooms	Classification	Measurement	MLP, LSTM, RF,SVM	Inductive	Homogeneous	Parameter	Weights
[88]	The work uses TL to improve activity recognition in smart homes, mapping sensors from multiple houses applying LSTM trained on the entire dataset, outperforming standard models in case of no labelled data and avoiding negative transfer	Classification	Measurement	Word2Vec	Transductive	Heterogeneous	Feature	
[106]	The paper evaluates similarities between two houses and uses using unsupervised TL to improve the accuracy of activity recognition	Measurement	LSTM	Unsupervised	Heterogeneous	Feature		
[87]	The paper exploits TL to solve the problem of fully retrain occupancy forecasting models on fresh data, speeding up their convergence	Classification	Measurement	LSTM	Inductive	Homogeneous	Parameter	Weights
[80]	The work has the aim to predict occupancy using synthetic data and CO_2 sensors, applying TL to overcome data unavailability	Classification	Simulation	CDBLSTM (Convolu-tional deep bidirectional LSTM)	Inductive	Homogeneous	Parameter	Features
[84]	The paper proposes an approach for improving the accuracy of home activity estimation using both transductive and unsupervised TL	Classification	Measurement	DT,RF	Unsupervised	Heterogeneous	Feature	
[81]	The paper explores the use of TL for human occupancy counting using CO_2 levels in a room, overcoming data unavailability problem	Classification	Measurement	DA-HOC, SD-HOC, SVR	Transductive	Homogeneous	Feature	
[83]	The paper aims to increase activity recognition accuracy in different homes, mapping data into a common feature space and relating it to a semantic space that describes the sensors	Classification	Measurement	binary sensor semantic and time information method (BSST)	Inductive	Heterogeneous	Feature	
[85]	The paper performs TL for visual activity recognition, exploiting already pretrained models to reduce the amount of data needed to create a model from scratch	Classification	Measurement	CNN, DT, SVM	Inductive	Heterogeneous	Parameter	Weights
[39]	The paper aims to tackle the data unavailability problem for activity recognition, transferring features from one house to another to increase the accuracy	Classification	Measurement	Feature Space Remapping (FSR), PCA	Transductive	Heterogeneous	Feature	
[82]	The paper exploited the Brick schema for representing metadata in IoT-enabled environments, exploiting transfer learning to ease the occupancy prediction task	Regression	Measurement	Knowledge Graph	Inductive	Homogeneous	Parameter	Weights
[79]	The paper uses a pretrained AlexNet to extract features from images captured by thermal cameras to identify the number of occupants	Regression	Measurement	CNN	Inductive	Heterogeneous	Feature	

(10.6%). In general Fig. 7 (c) confirms the growing interest for the topic in the recent years, with a greater presence in energy and building related journals, followed by journals in the IT field.

4.2. Application areas of transfer learning in smart buildings

Fig. 8 (a) and 8 (b) show the application distribution over recent years and by topic of TL in smart buildings. The main application of TL in smart buildings is related to load prediction, which represents about 34% of the surveyed works. Another common application is related to occupancy detection and activity recognition, while during recent years an increasing trend of works focused on building dynamics and systems control works was observed. The following subsections describe the papers collected for the present review according to their different application fields.

4.2.1. Load prediction

Building load prediction is an essential part of many building control and analytics activities, as well as grid-interactive and energy-efficient building operations. It may be found in a variety of applications across the built environment, from single components to multiple buildings. Among the 77 articles reviewed, 25 used transfer learning to facilitate more accurate, data-efficient, and robust load prediction. The literature shows that transfer learning can be used at different scales, from the prediction of appliance consumption through non-intrusive load monitoring (NILM) [54]; to specific equipment such as heating, ventilation, and air conditioning (HVAC) systems [55]; and wastewater treatment [56]. Moreover, it has been applied at the whole building scale [57,58] or district level [59], with hourly [55] or monthly resolution [37].

Among the techniques used to predict appliance consumption, the most popular is NILM. It refers to the process of analyzing changes in the voltage and current of a building to deduce which appliances are used in the house and their individual energy consumption. In D'Incecco et al. [60], the features extracted by using CNN are transferred across different appliances (e.g., kettle, microwave, washing machine) and across households in different regions (e.g., the UK and United States), and the regression layer is then fine-tuned. D'Incecco et al. [60] confirmed that it is possible to train a universal model for residential appliances that

Table 7
Summary of studies related to building dynamics.

ID	Objectives	ML problem	Dataset	Methods	Label	Space	Solution	NN
[117]	The work uses ANN and TL to speed the dataset creation process when dealing with multiple elements, obtaining the energy use intensity of the represented buildings	Regression	Simulation	MLP	Inductive	Homogeneous	Parameter	Weights
[75]	This paper has presented results from using different type of transfer learning to accelerate the real world performance of black-box systems, used to represent the dynamics of hot water systems	Regression	Measurement	MLP	Inductive	Heterogeneous	Parameter	Weights
[73]	The work analyse the effect of DNN architecture on TL for the prediciton of building thermal responses for HVAC controls, favoring the use of DNN rather than shallow NN	Regression	Measurement	MLP	Inductive	Homogeneous	Parameter	Weights
[40]	The presented paper proposes different methods to predict temperature evolution inside a buildings in an online way, coupling the prediction with an MPC controller	Regression	Simulation	TCA	Transductive	Heterogeneous	Feature	
[72]	The works discuss pros and cons of transferring the tail or the head of the DNN for building dynamics prediction (T and RH), claiming that transferring the head is equal to transferring building property, while transferring the tail may be more related to physical laws	Regression	Simulation	MLP	Inductive	Homogeneous	Parameter	Both
[116]	The paper employs DL and TL to speed building performance simulation, reusing the heating prediction layer to increase the performance of the cooling prediciton layer, reducing prediction gap for a high number of buildings.	Regression	Simulation	MLP,LSTM	Inductive	Homogeneous	Parameter	Features
[76]	The paper aims to predict and evaluate thermal comfort of occupants using learning-based approach for thermal comfort modeling, overcoming the problems of data and parameter inadequacy.	Classification	Measurement	MLP	Inductive	Heterogeneous	Parameter	Features
[74]	The paper develops a Resistance-Capacitance model for temperature dynamics, coupling clustering techniques and TL for buildings with insufficient data about building properties, increasing prediction accuracy	Regression	Measurement	BNN	Inductive	Homogeneous	Parameter	Weights
[138]	The authors proposed an online transfer learning framework to predict building dynamics, in which the online prediction are evaluated as a weight are combined from an offline source domain and the online target domain	Regression	Simulation	GOTL	Transductive	Homogeneous	Instance	
[71]	The works employs TL to adapt pretrained models for building dynamics from one building to another, increasing the predictive performance of target model with only a limited amount of data	Regression	Measurement	LSTM,RNN	Inductive	Homogeneous	Parameter	Weights

TARGET	DOMAIN	APPLICATION	SUB-KEYWORD1	SUB-KEYWORD2
A	B		C1	C2
Transfer Learning	Building*	BUILDING LOAD PREDICTION	Consumption	Forecasting
Domain Adaptation	Home*		Electricity	Prediction
	District*		Energy	
	City*		Load	
		OCCUPANCY DETECTION	Occupancy	Detection
				Estimation
				Prediction
		BUILDING DYNAMIC	Dynamic*	
			Surrogate Model*	
			Comfort	
			Temperature	
		BUILDING SYSTEMS CONTROL	Optimization	
			Control	
			Policy	
			Reinforcement Learning	

Fig. 5. Input for the second topic search (TS2) that explored the combination of target, domain, and application using sub-keywords, where * indicates both the singular and plural form

can be then adapted by using reduced training data, thus achieving the oracle NILM, remarkable computational savings as well as a decreased dependency on specific appliance-labelled data. In Liu et al. [61], the voltage-current (V-I) trajectory is first visually represented as images through color encoding, to enhance the load signature's uniqueness and to enable TL from image recognition. Then a deep learning model pre-trained on a visual recognition dataset is transferred to train the NILM classifier.

Moving at higher scale, building or city load prediction can be seen as typical time-series data and can be predicted using recurrent neural network (RNN) and its variants, such as long short-term memory (LSTM). For those neural networks, transfer learning can be applied using weight initialization or feature extraction, as explained in Section 2.

Whole building load prediction represents the most analysed topic in smart buildings, due to the data availability, especially at the meter level. Fang et al. [57] used transfer learning to enhance energy predic-

Table 8
Summary of studies related to load prediction.

ID	Objectives	ML problem	Dataset	Methods	Label	Space	Solution	NN
[37]	The work uses TL to enhance building monthly electric load prediction in different districts, with a DNN-based forecasting model on the source data, fine-tuned on the target data	Regression	Measurement	MLP, MLR, RF, XGB	Inductive	Homogeneous	Parameter	Weights
[111]	The paper exploits clustering and TL to enhance the performance of a prediction model for energy consumption, speeding up the process with respect to traditional TL	Regression	Measurement	K-means, LSTM	Inductive	Homogeneous	Parameter	Weights
[108]	The paper employs TL to overcome the problem of lack of training data BAS energy models to perform energy prediction, providing insights on how to select appropriate source buildings	Regression	Measurement	MLP, MLR, RF, XGB	Inductive	Heterogeneous	Parameter	Weights
[54]	The paper study the relation of feature selection among several electrical appliances, exploiting transfer learning to increase their classification in NILM application	Classification	Measurement	KNN, LDA, MLP, PCA, RF	Inductive	Heterogeneous	Feature	
[56]	The paper compared different TL approaches to improve the performance of wastewater treatment plant prediction	Regression	Measurement	CNN, GRU, LSTM	Inductive	Homogeneous	Parameter	Both
[105]	The paper addresses generalizability problems in NILM for load disaggregation using GANs and two TL solutions, feature-based and parameter-based	Classification	Measurement	MLP, GAN	Inductive	Heterogeneous	Feature	
[57]	The paper uses transfer learning to enhance energy prediction in buildings with few labelled data exploiting feature extraction and domain adaption, studying the effects of different time horizon, architectures and buildings		Measurement	CNN, FC, LSTM-DANN, RF	Inductive	Homogeneous	Parameter	Features
[128]	The paper employs transfer learning to perform short-term load prediction, proposing an effective method on how to find the single building most useful to perform weight-initialization TL	Regression	Measurement	LGBM, LSTM, RF, XCORR, XGB	Inductive	Homogeneous	Parameter	Weight-initialization
[109]	The paper propose a nove approach to perform load prediction with no data at all or augmenting data in case of a small dataset	Regression	Measurement	BIGAN, FC, LSTM, SVR	Inductive	Homogeneous	Parameter	Features
[62]	The work studies the application of transfer learning for building forecast prediction analysing how data availability and duration period availability influences parameter-based TL	Regression	Measurement	CNN, LSTM	Inductive	Homogeneous	Parameter	Both
[55]	The objective of the work is to exploit information of a building with a detailed sensor systems to perform a medium-term energy prediction on another building with few available data	Regression	Both	MLP, SVR, TrAdaBoost	Transductive	Homogeneous	Instance	
[59]	The objective of the work is to predict thermal load using data from other buildings based on an introduced similarity index and transfer learning	Regression	Measurement	LSTM	Inductive	Homogeneous	Parameter	Weights
[129]	The main objective of the work is to use RL, together with TL to perform energy prediction of multiple buildings extracting common features from other residential or commercial buildings	Regression	Measurement	DBN, SARSA, Q-Learning	Transductive	Heterogeneous	Feature	
[121]	This paper proposes Hephaestus, a novel cross-building energy prediction method based on transfer learning with seasonal and trend adjustment to improve prediction for a target buildings	Regression	Measurement	Hephestus, MLP, SVR	Inductive	Homogeneous	Parameter	Features
[119]	The paper proposes 2 TL models (seq2seq LSTM and CNN+attention) to increase accuracy prediction in target building with low data availability, comparing their effectiveness	Regression	Measurement	CNN, seq2seq LSTM	Inductive	Homogeneous	Parameter	Both
[118]	The paper explores the use of ML models to predict energy consumption of a leisure center, proposing a TL approach to enhance the performance of energy prediction in other leisure center, analysing the possibility to exploit information of an office building as additional source	Regression	Measurement	DT, EET, KNN, LightGBM, RF	Inductive	Homogeneous	Parameter	Weights
[130]	This paper proposes Similarity-based Chained Transfer Learning (SBCTL), a novel solution for building neural network-based forecasting models for a large number of smart meters, using previously fine-tuned network to transfer to the most similar meter, with an iterative process	Regression	Both	Seq2seq RNN, Similarity-Based Chained Transfer Learning (SBCTL)	Inductive	Homogeneous	Parameter	Weights
[127]	The paper compares transfer learning and meta learning with statistical methods and DL methods, showing the superiority of TL and Metalearning with few or no data at all, representing a suitable solution for short-term load prediction	Regression	Measurement	ARIMA, EGB, LSTM, ResNet LSTM, Seq2seq RNN	Inductive	Homogeneous	Parameter	Weights
[63]	The paper proposes a TL approach to increase the prediction performance of a model that forecast the customers' response to incentives, where the forecasting accuracy of a certain customer is improved using information of related customer	Regression	Measurement	MLP	Inductive	Homogeneous	Instance	
[58]	The paper proposes a TL approach for energy forecasting in absence of data, using CNN to extract feature reducing the amount of data needed to effectively perform the prediction	Regression	Measurement	CNN	Inductive	Homogeneous	Parameter	Features
[61]	The paper exploits TL to increase the accuracy of a NILM model, encoding Voltage-Current trajectory as images t and using a pretrained dataset to train a NILM classifier	Classification	Measurement	CNN	Inductive	Heterogeneous	Parameter	Features
[110]	The paper exploits weight-initialization transfer learning to increase the performance of residential short-term forecasting	Regression	Measurement	CNN	Inductive	Heterogeneous	Parameter	Weights
[60]	The paper exploits TL to increase the performance of a regression model to predict appliances consumption, achieving two benefits: reduce the number of sensors for each appliance to be installed and offer computational savings	Regression	Measurement	CNN	Inductive	Heterogeneous	Parameter	Features
[104]	The work uses a model trained on the most similar building to enhance the load prediction on the target building, using several metrics to evaluate similarities among time series	Regression	Measurement	DT, MLR, MRF, RF	Inductive	Homogeneous	Instance	
[115]	The paper propose domain adaptation for energy disaggregation models, improving the performance and reducing labelled data requirement of the model	Classification	Measurement	CNN	Inductive	Homogeneous	Parameter	Features

Table 9
Summary of studies related to other building and energy related applications.

ID	Objectives	ML problem	Dataset	Methods	Label	Space	Solution	NN
[94]	The paper aims to solve the problem of insufficient samples in training data for weather forecasting in ML using TL, transferring the data from related cities and building a forecast model based on the extended dataset	Regression	Measurement		Transductive	Homogeneous	Instance	
[96]	The paper uses TL to overcome data unavailability to perform wind power forecasting in different zones, increasing prediction performance	Regression	Measurement	GBDT	Transductive	Homogeneous	Instance	
[97]	The objective of the study is to exploit transfer learning to increase classification accuracy of earthquake damage detection in building, using a pretrained neural network adopted for image classification, increasing the performance of the classifier	Classification	Measurement	CNN, CVA, RF, SVM	Inductive	Heterogeneous	Parameter	Features
[112]	The paper addresses the need to speed up inspection work after construction in buildings, using a transfer learning approach that lead to higher accuracy and better efficiency, paving the way for the integration of fully autonomous mobile robot systems	Classification	Measurement	R-CNN	Inductive	Heterogeneous	Parameter	Features
[113]	The paper uses TL to overcome labelled data unavailability to detect historical buildings, using pretrained ResNet50 as backbone and image augmentation to increase dataset dimensions	Classification	Measurement	R-CNN	Inductive	Heterogeneous	Parameter	Features
[107]	The paper presents an architecture that exploits deep features selection and TL to overcome limited dataset when dealing with intelligent decision support for power transformers	Classification	Measurement	DBN, MLP, BLOCK HSIC Lasso	Inductive	Heterogeneous	Feature	
[100]	The work aims at increasing the accuracy of a prediction model for the building identification using pretrained CNN and TL	Classification	Simulation	CNN	Inductive	Heterogeneous	Parameter	Features
[89]	The paper uses TL to build a binary classifier for HVAC component degradation to overcome lack of labelled data, increasing classifier accuracy and reducing computational cost	Classification	Measurement		Inductive	Homogeneous	Parameter	Weights
[95]	The work aims at predicting thermal comfort in different buildings and climatic zone using transfer learning, overcoming traditional ML methods with the highest performance in different buildings with the same climate	Classification	Measurement	Adaboost, DT, GAN, KNN, MLP, RF	Inductive	Heterogeneous	Parameter	Features
[120]	The objective of the work is to exploit TL to predict FDD in chillers, performing domain adaption to account for different size	Classification	Measurement	BN, DT, KNN, SVM	Inductive	Heterogeneous	Parameter	Both
[92]	The main objective of the work is to exploit TL to reconstruct missing data in buildings, combining transfer learning with DL techniques	Regression	Measurement	FCNN, KNN, LSTM, LSTM-BIT (LSTM with bidirectional input transfer), RNN, RF, SVM	Inductive	Homogeneous	Parameter	Weights
[38]	The paper exploits TL to identify night setback in District Heating systems. This is done transforming time series into images (heatmaps) and using transferred knowledge from the Imagenet dataset to properly classify the presence or not of night setbacks	Classification	Measurement	CNN	Inductive	Heterogeneous	Parameter	Features
[93]	The paper exploits RL and TL to extend the comfort model from one building to another building	Regression	Measurement	MLP, Q-learning	Inductive	Heterogeneous	Parameter	Weights
[98]	The paper proposes a method that employs TL to detect and classify seven classes of old building damage in Medina of Fez and Meknes in Morocco	Classification	Measurement	Logistic Regression, RF,SVM	Inductive	Heterogeneous	Parameter	Weights
[91]	The paper proposes a method to automatically map building's sensoring and control points according to several features, able to identify common characteristics in different buildings even with different metadata conventions	Classification	Measurement	CNN	Transductive	Homogeneous	Feature	
[114]	The paper employs parameter-based TL with feature-extraction to increase the accuracy of a building image recognition model, comparing several pretrained architectures	Classification	Measurement	CNN, CRNN, DTL, GAN, LSTM	Inductive	Heterogeneous	Parameter	Features
[101]	The paper introduces a deep decoder transfer-learning (DTL) framework to address personal air quality prediction problem, using a pretrained DNN and the Wasserstein distance to match the heterogeneous distribution between the source and target domains	Regression	Measurement	CNN	Inductive	Heterogeneous	Parameter	Features
[90]	The paper performs an in-depth analysis of the application of TL for fault detection and diagnosis of chillers, exploring several architectures and boundary conditions	Classification	Measurement	CNN	Inductive	Heterogeneous	parameter-based	Both
[99]	The paper employs TL for PV panel defects detection, using a pretrained AlexNet to increase the effectiveness of the proposed classifier, addressing the problem of labelled data unavailability	Classification	Measurement		Inductive	Heterogeneous	Parameter	Features

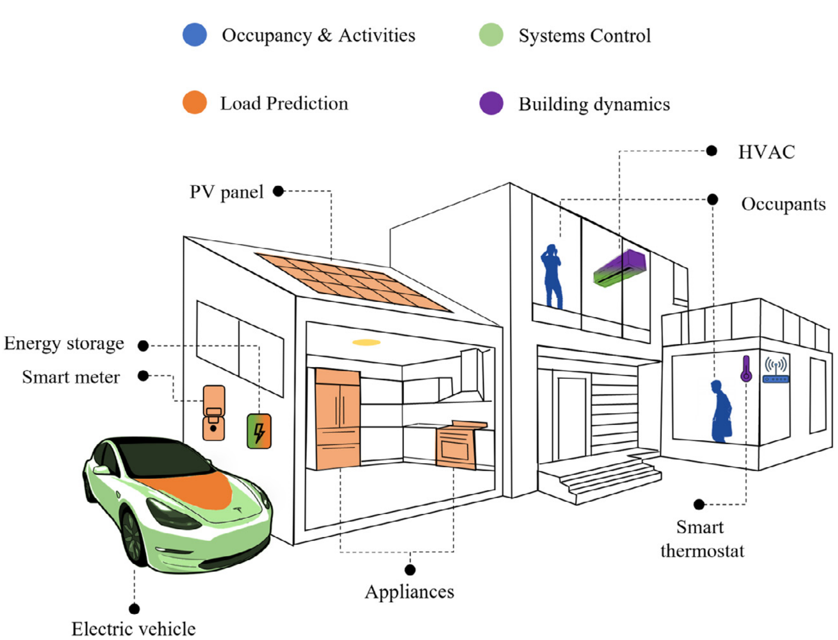

Fig. 6. Machine learning applications in a smart building

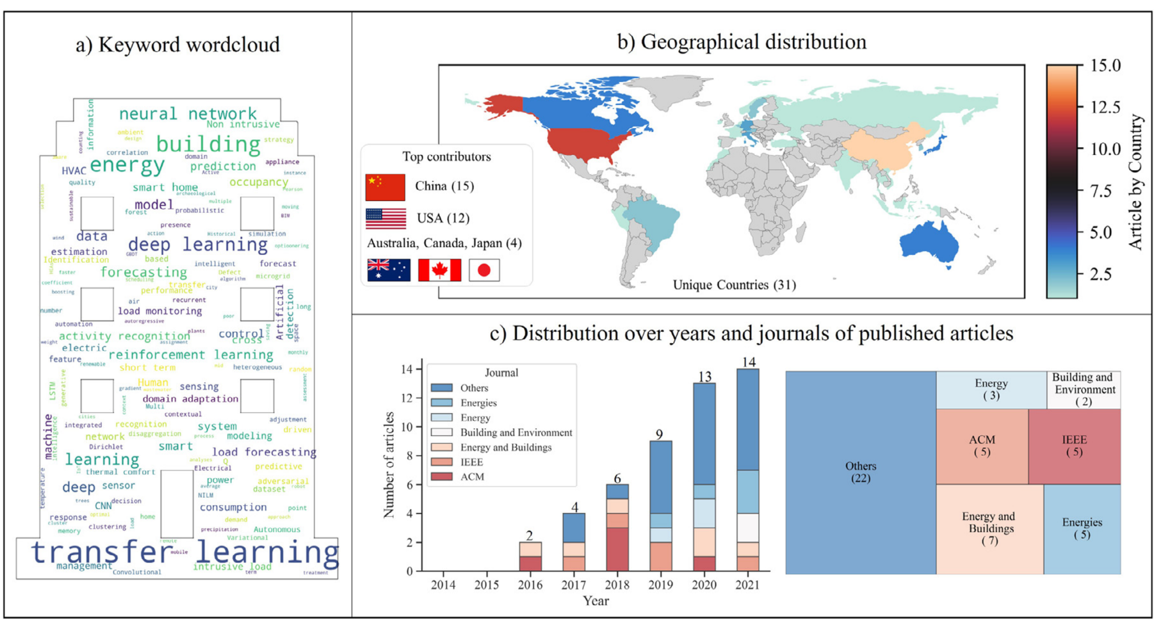

Fig. 7. Overview of the reviewed literature: (a) keyword wordcloud of the selected literature; (b) geographical distribution of the researchers; (c) publication by journal of recent years

tion in buildings with few labelled data, employing an LSTM as feature extraction and further fine-tuning a regression layer for domain adaptation, studying the effects of different time horizons, architectures, and buildings. Fan et al. [62] compared several parameter-based architectures to enhance building forecasting prediction, analysing how data availability and duration period affects performance. Lastly, Cai et al. [63] exploited TL to increase the accuracy of incentive-based Demand Response (DR), characterized by stochastic and sporadic events, using data from similar customers.

4.2.2. Systems control

BAS are computer-based automated systems that monitor and regulate all energy-related systems in buildings, including mechanical and electrical equipment. BAS are frequently used to automate all services and operations within a building in order to optimize its performance, efficiency, and energy usage. With a significant role in distributed energy resources exploitation and energy transition, this technology enables the execution of essential energy management activities such as automating demand response techniques and supervising energy prices. Among the 77 papers reviewed, 7 used transfer learning to enhance building systems control. The papers exploited a policy-transfer approach [51] in combination with RL to optimize control at different scales: microgrid [64,65], batteries [66], HVAC systems [67,68], and appliances [69,70]. A key pain point of applying RL controllers in buildings is the training process that is time- and data-demanding before it can converge. To address this problem, Zhang et al. [69] first identified several homes similar to the target home that have the same number and type of appliances. Then the RL controller was trained on the source home and fine-tuned for the target homes. The results showed that TL can effectively reduce the training time of a new policy if the target home is similar to the source homes. Tsang et al. [70] used transfer learning to train a DRL controller of Household Energy Management.

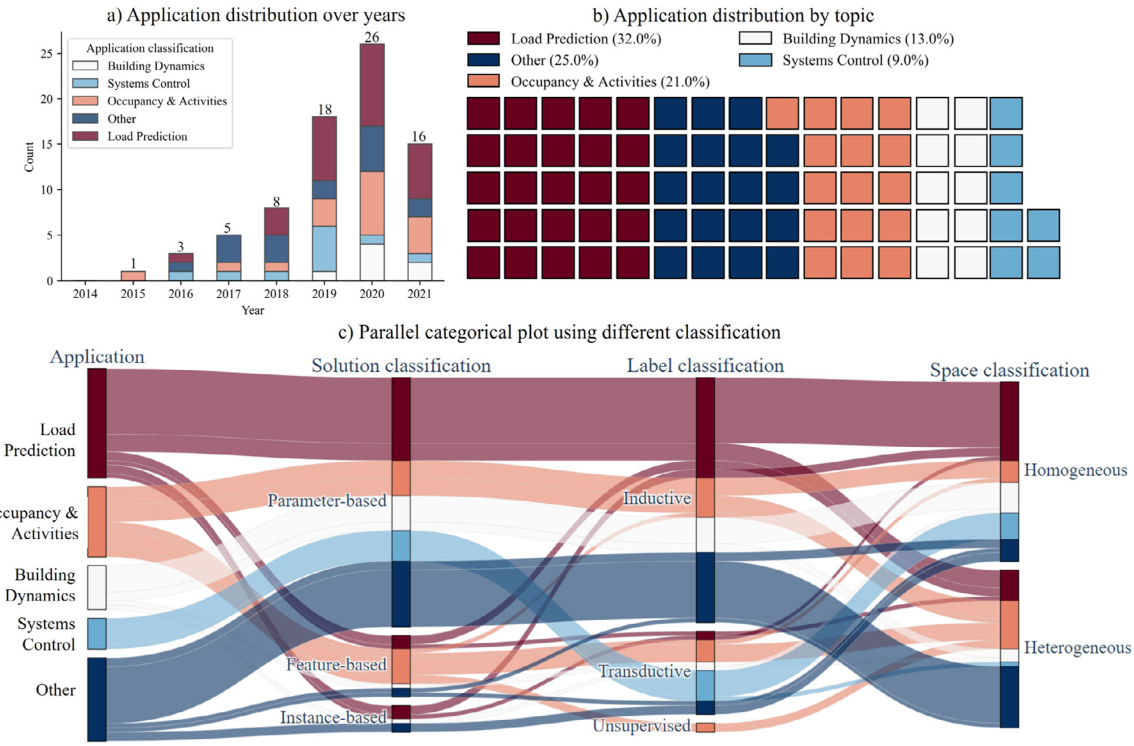

Fig. 8. Overview of the application of TL in smart buildings (a) publication by application over recent years; (b) distribution of the applications; and (c) parallel categorical plot using different classification within the review

The agents in the target domain are advised by the suggested actions of the existing model pretrained in the source domain.

Xu et al. [68] applied the same process, shifting the domain from appliances to HVAC systems, transferring the policy of DRL-based HVAC controllers from source buildings to target buildings with different materials and layouts, HVAC equipment, and weather conditions. Furthermore, they analysed a case with a different number of thermal zones, being the only work that used heterogeneous TL for control application, thanks to its ability to generalize over thermal zones. Similarly, Lissa et al. [67] studied the effect of transferring the policy of an HVAC controller from one room to another in the same building, performing several experiments to test the robustness of the controller and assessing the impact on occupant discomfort time, showing reductions using a TL approach. Looking at microgrid scale, Fan et al. [64] evaluated similarity between the production and generation of two different microgrids to find the optimal way to transfer knowledge, sharing the weight of the DRL neural networks and speeding the controller performance, paving the way for possible application at a large scale. Lissa et al. [65] proposed an inter-agent transfer, in which knowledge is shared with another agent with similar characteristics, and this agent is able to merge the transferred knowledge with its own experience. This concept is called *parallel transfer learning,* where the knowledge to be shared between agents does not need to wait until the end of the process to be available. Lastly, Mbuwir et al. [66] applied transfer learning to speed the convergence of the learning algorithm to optimize thermal and battery storage planning, improving also its scalability. The results show that reinforcement learning coupled with transfer learning can represent a suitable alternative when few data are available, despite further studies are needed to demonstrate the ability of transfer learning to generalize across multiple buildings, especially when controlling different energy systems.

4.2.3. Building dynamics

Building thermal dynamic models (predicting how the building thermal state will evolve under different weather, disturbances, and other factors) have many applications, including but not limited to advanced controls such as Model Predictive Control (MPC) and DRL or the increased accuracy of load prediction models. Conventional building thermal models are developed through a physics-based approach, such as in EnergyPlus. The shortcomings of physics-based building modelling are the high time and expertise demand needed to develop such a model and the need for a great deal of information about the building and system features. An alternative approach to model building thermal dynamics is data driven modelling. However, a large amount of historical data may be needed to train such data-driven building thermal dynamics models, which is challenging, especially for buildings that are brand new or not yet commissioned [71]. This highlights how transfer learning can be leveraged for this application. Among the 77 papers reviewed, 10 focus on using transfer learning to develop building dynamics models. Jiang and Lee [71] pretrained an LSTM S2S model using a large amount of data from source buildings to study building temperature evolution. Then weight initialization was used to enhance the performance of the target building. In that case, the whole model was fine-tuned without freezing any hidden layers. Similarly, Chen et al. [72] applied transfer learning to predict not only internal temperature but also relative humidity. In other studies (such as [73]), the hidden layers have been frozen while only the last fully connected layers were fine-tuned. It was found that the deep supervised domain adaptation is effective to adapt the pretrained model from one building to another, and has better predictive performance than learning from scratch with only a limited amount of data [71].

Hossain et al. [74] trained a Bayesian neural network (BNN) to directly learn an RC model rather than estimating parameters. The work proved that at least several weeks of data are necessary to obtain good performance. The paper proposes a methodology on how to transfer these models in new buildings with only one day of data, identifying and selecting the best RC model according to consumption patterns and outperforming time-series methods directly constructed on available data.

Additionally, data-driven models have been used to represent specific temperature evolution, as in Kazmi et al. [75], which applied TL to train a model to predict the thermal behaviors of hot water storage systems; or Hu et al. [76], which applied transfer learning to predict the thermal comfort state in buildings. Lastly, Grubinger et al. [40] present an interesting approach of online transfer learning coupling the resulting prediction with an MPC controller, paving the way for possible application of this technique.

4.2.4. Occupancy & activities

Building occupancy data are useful for improving the effectiveness of energy management systems so that energy consumption may be reduced while occupant comfort is maintained. Occupancy prediction aims to predict the occupant counts/states using the historical occupant counts/states (for instance, in [77]).

Occupancy detection aims to predict the number of occupants in a space from images ([78], [79]) or environmental sensors (primarily CO_2 sensors, such as in the study [80], [81], [82]), while activity recognition finds wide applications in remote elderly care and the healthcare industry [83], [84], [85], [86],[39]. Among the 77 papers reviewed, 16 are about occupancy detection or activity recognition. Regarding the first topic, CNN have been widely used to deal with camera images, and transfer learning can help to avoid having to train a complicated CNN from scratch, speeding the process or dealing with insufficient data. For instance, Mosaico et al. [79] transferred a pretrained AlexNet to extract features from images captured by thermal cameras to identify the number of occupants, and confirmed that the occupant detection approach built upon transfer learning can achieve higher performance with respect to standard models. Another way to detect the occupant counts uses environmental sensing data; Arief-Ang et al. [81] applied semi-supervised domain adaptation to a human occupancy counting model so it could be implemented in any room without adequate labelled data, while Weber et al. [80] applied TL to pretrain and transfer a deep neural network to reduce the amount of data needed for training. Lastly, occupant counts can be predicted from historical occupant data, and using this approach RNN is the most widely used algorithm. Similar to applying TL to predict building load, TL was applied to pretrain either RNN or LSTM ([87] [77]).

On the other hand, most of the existing activity recognition methods are based on supervised classification algorithms, which are limited by the shortcoming that the classification model learned in one smart home environment usually cannot be used in another. For a new smart home environment, sufficient sensor readings have to be collected and labelled to learn the needed classification model. This process is time consuming and expensive. Transfer learning can help to address this challenge. Niu et al. [88] used TL to improve activity recognition in smart homes, comparing the approach with standard unsupervised ML models and outperforming them, avoiding negative transfer. Inoue and Pan [84] proposed an approach for improving the accuracy of home activity estimation by transferring existing data to a new household with a novel approach that exploited unsupervised learning.

4.2.5. Others

Our literature review shows that transfer learning has been applied extensively in other domains of the building field as well.

For example, Dowling and Zhang [89] applied transfer learning to develop a log-likelihood classifier to detect faults, errors, and degradation of HVAC systems. The detector was pretrained on a building with a larger amount of data and then used on a building with less data. Similarly, [90] performed an in-depth analysis of the application of TL for the classification of fault detection and diagnosis of chillers, exploring several architectures and analysing the effect of data availability on TL performances. Hong et al. [91] used transfer learning to develop a metadata model for buildings. The model can learn a set of statistic classifiers of the metadata from a labelled source building and adaptively integrate those classifiers to another unlabelled target building, even if the two buildings have very different metadata conventions. This approach can automatically label more than 36% percent of the labels in a new building with at least 85% accuracy, and for some cases up to 81% with more than 96% accuracy. TL was also used to tackle the missing value problem in building energy related data using a bidirectional LSTM [92], or combined with RL to represent the occupant behavior on set point adjustment and thermal comfort [93]. Some other applications include applying TL to predict weather data [94] or comfort conditions in different cities according to weather data [95]. Cai et al. [96] utilized instance-based transfer learning to improve the accuracy of probabilistic wind power forecasting, in which different weights are assigned to different auxiliary training sets according to their relatedness to the target problem to reflect the real relatedness between source domains and the target domain.

Another common application is related to the use of pretrained ImageNet for classification tasks in building systems. Abdi and Jabari [97] and Masrour et al. [98] transferred a ResNet-18 CNN model to detect the building damage resulting from an earthquake and age-related deterioration. Zyout and Oatawneh [99] transferred a pretrained AlexNet to classify the surface of photovoltaic (PV) panel images as either normal or defective. Mao et al. [100] used the same approach for building identification with a pretrained Recog-Net, increasing the accuracy by 10%. Lastly, [38] converted time-series into images and used pretrained CNN to identify night setback in district heating substations, and retrained classification layers using target domain data.

Additionally, to predict the personal exposure to air pollution, Zhao and Zettsu [101] designed a transfer learning framework based on an encoder-decoder structure, in which the Wasserstein Distance was used to match the heterogeneous distribution of the source domain (the data from the atmospheric monitoring stations) and the target domain (the personal air quality), which is referred to as*decoder transfer learning* (DTL). DTL matched the feature distributions by reducing the Wasserstein distance between the source feature distributions and the target feature distributions.

Lastly, Paudel et al. [102] exploited TL for activity recognition, using such modelling to increase the performance of an LSTM model used to simulate building dynamics, and in turn using this model to optimize energy consumption. It is worth mentioning that the paper combines almost every application identified, highlighting integration of TL in the built environment ecosystem.

4.3. Transfer learning approaches in smart buildings

After analysing the main application of TL in smart buildings, this study analyzes the relation among applications and the classification introduced in Section 2, providing insights on the type of algorithms used for each of them. As Fig. 8 (c) shows, we analysed TL in smart buildings considering the following classifications: (1) applications, (2) solutions, (3) labels, (4) space. The TL parallel categorical plot in Fig. 8 (c) relates the four types of classification, in which each chunk indicates the number of studies, color-coded by applications (Load Prediction, Occupancy & Activities, Systems Control, Building Dynamics, and Other). Analysing the solution classifications, it can be noticed the absence of reviewed work that adopted relation-based TL, since multi-relational dataset applications are not so common in buildings and emerging techniques like graph neural networks still need to be properly explored in this field. Moreover, looking at label classification, it can be seen that unsupervised TL is used only in a few works for occupancy detection and activity recognition, due to the intrinsic nature of these problems, while other applications have not explored this type of transfer learning. Lastly, despite being used in different applications, the number of homogeneous and heterogeneous works is almost the same, suggesting that their use mainly depends on data availability and task. The subsection unfolds over the three main solution approaches used in transfer learning, assessing their relation with the other classifications.

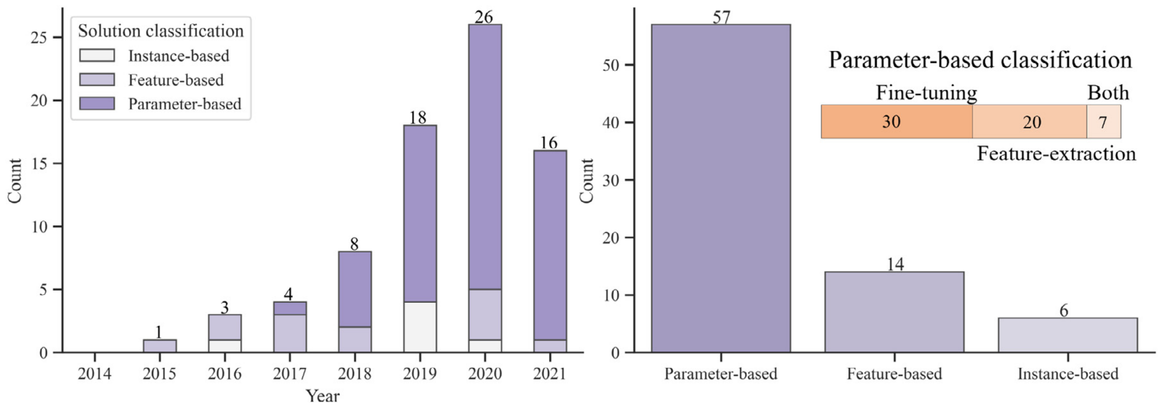

Fig. 9. Solution categorization: (a) over the years; (b) by solution type, with further division for parameter-based classification

4.3.1. *Instance-based transfer learning*

As shown in Fig. 9(b), instance-based TL has been used in 6 papers out of 77, with sparse application across the years. This approach has been used mainly in occupancy detection and activity recognition, thanks to the ability to directly reuse source data to improve performance in the target domain. Among the papers that exploited instance-based for occupancy detection and activity recognition, [103] used data from multiple source spaces to recognize activities using no labelled data from the target space, improving the results despite the different apartment layouts and resident schedules. This also can be seen looking at Fig. 8, which shows that all the instance-based approaches used for activity recognition are transductive transfer learning applications, highlighting how this approach is particularly useful in absence of labelled data. Moreover, few applications of load prediction used instance-based TL to improve their performance. Among them, Qian et al. [55] used instance-based TL with the aim of improving medium-term energy prediction of a building with few data. To achieve that goal, the paper exploited the TrAdaBoost algorithm using real data and simulated data from a source building, to improve load prediction of the target building, solving the data availability problem and comparing the performance with Artificial Neural Network (ANN) methods. Moon et al. [104] used instance-based TL, studying the similarity among the target building and source buildings, using different distance metrics creating a robust methodology called SPROUT, and improving prediction with respect to other ML methods. Lastly, other applications that involve the use of instance-based TL can be found in probabilistic wind power forecasting, in which the goodness of source data are quantified using maximum likelihood, consequently weighting the source data to increase prediction accuracy in the target domain.

4.3.2. *Feature-based transfer learning*

Fig. 9 (b) shows that 14 of the 77 papers reviewed used feature-based TL. Feature-based TL represents the second most used method; it is applied mainly in load prediction and occupancy detection, and is used together with other applications in the context of smart buildings. Among the load prediction applications, feature-based TL was used in [54] and [105] for NILM. Looking at occupancy detection application, feature-extraction has been used in [81] to determine the occupancy of rooms based on CO_2 sensor data, extracting features that allow the proposed algorithm DA-HOC, to improve the binary classification. Sonia and Baruah [106] exploited feature-based TL for activity recognition in several smart homes. In particular, the analysis dealt with different numbers and types of sensors, finding relations among location, number, and type of sensors between the source and the target domain. The main advantage of the proposed methodology relies on its ability to find similarities without labelled data in both domains, representing one of the few applications of unsupervised transfer learning. Moreover, feature-based TL has been used in [91] for metadata representation in buildings, and Chernov et al. [107] used it for intelligent decision support for power systems characterized by a high-dimensional space and limited data availability.

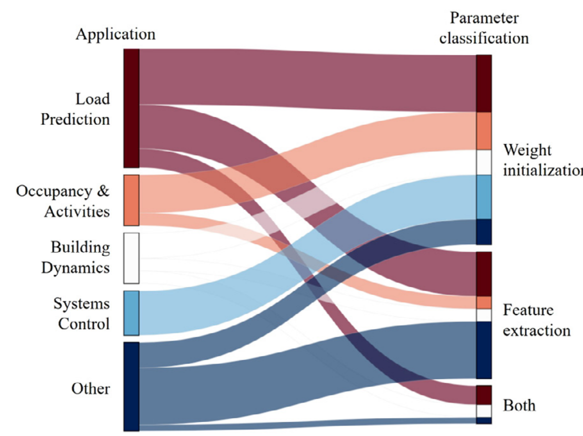

Fig. 10. Sankey diagram of application and parameter-based classification of TL in smart buildings.

4.3.3. *Parameter-based TL*

Parameter-based TL represents the most used method in smart buildings, with the largest adoption occurring during the last few years. Among the 77 papers reviewed, 57 used a parameter-based approach, representing the vast majority of the papers reviewed. Fig. 10 shows that parameter-based methods are mostly used in load prediction, with several architectures including: MLP ([108]) LSTM ([109]), GRU (gated recurrent units, [56]), CNN (convolutional neural network, [58,110]), CNN + LSTM ([62]), and LSTM-TLL ([111]). Other common applications are building dynamics and systems control, linked by the nonlinear nature of these topics, affected by stochastic variables or driven by physical laws that justify the wide adoption of neural networks. Fig. 4 provides insights on the relation between the application and the parameter classification, weight-initialization, or feature-extraction.

Feature-extraction

In particular, classification tasks such as activity recognition or image processing applications often rely on feature-extraction; among the others, the tackled problems include building damage detection [97], building quality assessment [112], and historical building detection [113,114]. In these cases, feature-extraction enables users to exploit commonly used pretrained neural networks such as ResNet-18 and ResNet50 to overcome data scarcity, adapting input and output layers with a fine-tuning process. Moreover, [115] proposed a novel method to perform energy disaggregation, exploiting CNN to extract features and consequently adapting the domain to account for different appliances, trying to minimize the distribution discrepancy using semi-supervised learning. Singaravel et al. [116] exploited feature-extraction to speed the prediction of heating and cooling demand by a factor of 1,000. The method used LSTM and TL to speed building performance simulation, reducing prediction gaps for a high number of buildings. Looking at building load prediction applications, Fang et al. [57] exploited several Deep Neural Network (DNN) architectures as feature-extractors, including fully connected layers (FC), convolutional neural networks (CNN), and long short-term memory (LSTM) neural networks to improve short-term energy prediction among buildings.

Weight-initialization

A large part of regression task problems, such as load prediction, building dynamics, and system control, are often associated with weight-initialization, followed by a fine-tuning process that exploits target data. Demianenko and De Gaetani [117] used ANN and TL to speed the creation of parametric datasets when performing building simulations. The proposed approach is able to significantly reduce the simulation period, considering multiple design factors and evaluating the impact on the final value of energy use intensity (EUI). Similarly, Banda et al. [118] proposed a similarity-based chained transfer learning to increase the accuracy of load prediction exploiting data from smart meters in an iterative fashion. Lastly, parameter-based TL has been used in system control, where weight-initialization can help data-driven controllers to jumpstart their performance while trying to achieve an optimal control policy. Fan et al. [64] used TL to exploit accumulated knowledge of an RL controller used to schedule the optimal strategy of a microgrid, fine-tuning the learned policy (encoded in the DNN) and deploying it in another environment. The same approach was used in Lissa et al. [65] and Lissa et al. [67] to speed the training of a DRL controller for heat pump and HVAC management.

Hybrid method and comparison

A last approach analyses complex architectures, in which both feature-extraction and weight-initialization are performed for different parts of the neural networks, or the two approaches are compared. Gao et al. [119] proposed two neural-network architectures (seq2seq and CNN) deployed in five models, using different TL approaches. In particular, they compared the necessity to adopt feature-extraction of the context vector for the seq2seq LSTM and the CNN, comparing it with a simple TL of a dense-layer in the LSTM. Zhu et al. [120] used different architectures of a DNN to detect faults in building chillers. The first layers were used to extract features and frozen, while the other hidden layers were fine-tuned. Finally, Fan et al. [62] and Chen et al. [72] proposed a detailed comparison among the two approaches for the application of load prediction and building dynamics, respectively, providing useful insights on the difference between the two methods. In particular, Fan et al. [62] performed a statistical investigation on the lack of data and its distribution among time periods when performing TL for load prediction, while Chen et al. [72] selected several MLP architectures analysing the role of the first and last layers of the neural network when performing TL for building dynamics. Another interesting application combined parameter-based and instance-based TL [121] for load prediction, accounting for seasonal and trend adjustment using data from similar buildings. Kazmi et al. [75] compared feature-based TL with a parameter-based TL that exploited weight-initialization, observing that when approaching a heterogeneous TL problem, weight-initialization and fine-tuning represent the best option; on the other hand, feature-based TL was found to better perform in a homogeneous context.

4.4. Tools and metrics

The subsection focuses on the description of adopted tools, data, and metrics to assess the performance of TL. Particular focus was devoted on the approaches adopted to evaluate similarities among source and target datasets.

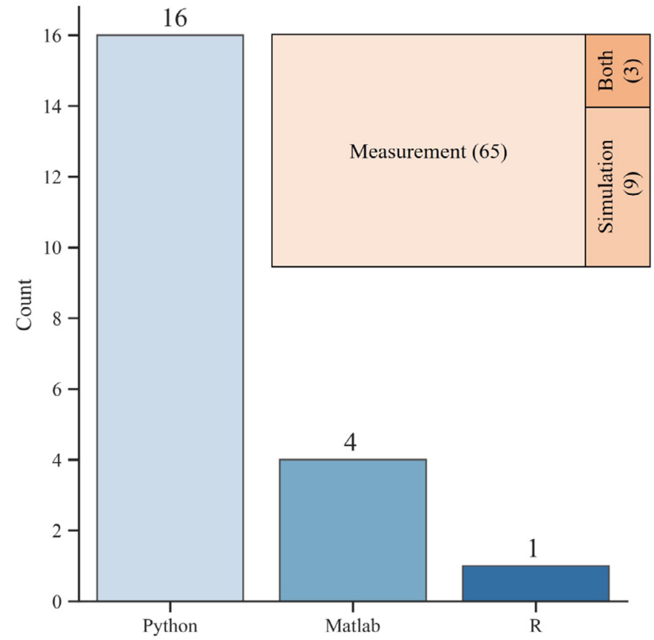

Fig. 11. Countplot of most popular tools and treemap of data sources used for TL works in smart buildings.

4.4.1. Tools and data

Fig. 11 shows the most commonly used tools: Python, MATLAB, and R and the data source (i.e., measured, simulated, or both). Fig. 11 displays how only about 20 papers out of the 77 surveyed clearly stated the tools adopted for the analysis. The most popular tool is Python, which allows an easy implementation of parameter-based TL thanks to the packages TensorFlow[122], Keras[123], and PyTorch[124]. On the other hand, MATLAB uses Deep Learning Toolbox [125] to ease the implementation of neural networks, followed by R [126]. Looking at data availability, it can be seen that more than 80% of the analysed papers relied on monitoring data, highlighting the practical role of TL in the built environment. Few cases exploited both real data and simulated data to increase the accuracy of predictive models, while control application with RL exploited only simulated data.

4.4.2. Metrics

From the analysis it was observed that a large part of surveyed works rely on measured data, and that the most common metrics are the ones typically used to measure performance in other machine learning tasks. There are few new metrics being proposed specifically for transfer learning, to evaluate similarities between the datasets and to specifically quantify TL performance.

Fig. 12 displays three common measures used to quantify learning improvement. The first metric used is the ”jumpstart,” which describes the increase in the initial performance achievable in the target task using the transferred knowledge, before any further learning. The second metric is the ”time to threshold,” used to quantify the amount of time it takes to achieve certain performance in the target task given the transferred knowledge, compared to the amount of time necessary to learn it from scratch. The last metric is the ”asymptotic performance” level achievable in the target task compared to the one without transfer. Independently from the specific ML task, jumpstart and asymptotic performances are evaluated, often using the same metrics according to the problem, while the time to threshold is measured as a reduction in computational cost. The following subsections describe the principal metrics encountered in classification and regression tasks, together with other metrics used to quantify the similarities between buildings or tasks.

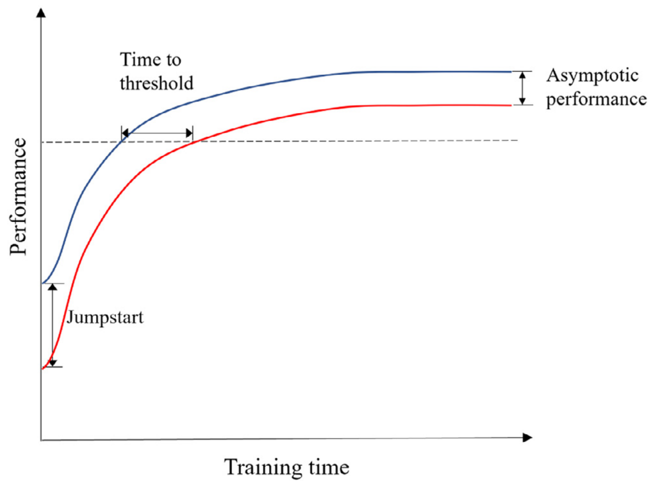

Fig. 12. Common measures used to quantify the performance of transfer learning.

4.4.3. Classification

This subsection analyses common metrics used for classification problems to quantify the jumpstart or asymptotic performance due to the application of TL with respect to ML. Results show that for classification problems the most widely used metric is accuracy, found in 24 of the 77 papers analysed, followed by F1-score (or Macro F1), precision, and recall. An example of accuracy application can be found in [88], which used the metric to quantify the improvement of activity recognition models using transfer learning with respect to unsupervised machine learning models. Precision, recall, and F1-score have often been coupled with accuracy to get a better understanding of the performance of the classification model. In fact, if there is significant class imbalance, then it might be beneficial to check the F1-score rather than just accuracy. For instance, in [105], the main aim was to attain high accuracy in load disaggregation. Since the data were heavily imbalanced, accuracy was complemented with metrics like Macro-F1 score and Transfer Gain (the difference between the F1-score of the transfer learning model and the F1-score of a model trained on the target data from scratch).

Another metric used (such as in Rashidi, Parisa and Cook [103]) is the recognition rate, defined as the "percentage of sensor events predicted with the correct label." This metric provides an overall understanding of how well the model is recognizing different activities. Inoue and Pan [84] delve deeper by using metrics like true positive rate (TP-rate), true negative rate (TN-rate), and balanced classification rate. These, if used in combination, provide an imbalance independent overview on the model.

4.4.4. Regression

As previously done for classification problems, common metrics used in regression problems are analysed to quantify jumpstart or asymptotic performance. The most widely used metrics among the surveyed papers in the regression task is root mean squared error (RMSE) or mean squared error (MSE), found in 29 works. An example of using MSE is Pardamean et al. [78], which uses transfer learning in the context of computer vision to count the number of occupants in a room through video data, employing the RMSE to evaluate the performance increase due to the transfer learning application.

Other popular metrics are mean absolute error (MAE), mean absolute percentage error (MAPE), and R^2. MAE was used in about 19 different papers and MAPE in 15 papers. Additionally, some variations of MAPE are found in the reviewed studies, such as mean MAPE (MMAPE, in [59]) and symmetric MAPE (SMAPE, in [127] and [101]).

Additionally, quantile forecasting score (QS) has been used to measure the performance of quantile regression used to estimate wind power quantiles in Cai et al. [96].

Mosaico et al. [79] used mean bias error and calculated the standard deviation of this error by defining another metric called *error standard deviation.*

In papers focusing on energy forecasting or disaggregation, like D'Incecco et al. [60], some unique metrics have been defined. For example, D'Incecco et al. [60] uses normalized signal aggregate error (relative error of the total energy predicted), energy per day (absolute error in predicted energy used in a day), and normalized disaggregate error (normalized error of the squared difference between the prediction and the ground truth of the appliances).

4.4.5. Other metrics

In transfer learning, a necessary step to avoid performance decrease, also called *negative transfer,* is to use it when source and target datasets have a certain degree of similarity. It is therefore important to propose a robust methodology to quantify similarities. Among the various applications that attempt to quantify building similarities, the greatest efforts have been made in the field of load prediction. For example, Jung et al. [37] evaluated the Pearson correlation coefficient (PCC) to select the most similar time-series and used it to initialize the weights of a DNN. Ozer et al. [128] studied the correlation of newly constructed building characterized by low data availability with other buildings, to employ weight-initialization, outperforming DNN directly trained on the target building. Mocanu et al. [129] used the Kolmogorov-Smirnov test, with the maximum difference between an empirical and a hypothetical cumulative distribution. Moreover,Lu et al. [59] studied the analogies among the datasets using similarity measurement index, while Tian et al. [130] employed Euclidean, Cosign, and Manhattan distances between the distributions. Looking at feature-based applications, Chen et al. [86] used principal component analysis (PCA) in the source and target domain to a space with a higher divergence that contains more independent information, adopting Gale-Shapley similarity measurement and Jensen-Shannon divergence to estimate similarity between each feature as the reference for feature mapping. Lastly, clustering is an unsupervised learning technique that can help group domains and aid the process of choosing an appropriate source domain for a corresponding target, enhancing the process of transfer learning. An important metric used for clustering is silhouette coefficient, used in Le et al. [131], which indicates the consistency of data within clusters and for each point, provides information how similar that point is to the other data points in its cluster compared to other clusters.

5. Discussion

In this section we present and discuss the reviewed works, as well as key challenges and opportunities of applying TL in smart buildings.

5.1. Research trends and open questions

Applying transfer learning in smart buildings is an emerging research topic that has attracted increasing research attention. The idea of transfer learning was originated from machine learning, which was accelerated as more data and computing power became available in the past decade. As shown in Fig. 7, the number of publications on this topic has increased since 2015. Also, due to the multidisciplinary research area, the papers on this topic have been published in both computer science-oriented journals and at conferences (e.g., IEEE, ACM), as well as in building science-oriented journals (e.g., *Energy and Buildings, Building and Environment,* and *Applied Energy*).

However, research on this topic is still at the very early stage or, in the case of relation-based TL, still needs to be explored in smart buildings. Moreover, despite using real data, existing literature used such data in an offline fashion, without deploying them in real world applications. This approach tends to be simplified and may not reflect real world problems in real buildings. More in-field studies are needed to validate TL performance in real buildings. Collaboration and coordination between academia, industry, and policymakers are needed to apply TL to solve real-world problems and make true impacts.

Despite the emerging interest for transfer learning in smart buildings, a number of research gaps still need to be addressed. Below are reported considerations and insights for a number of open questions based on the knowledge extracted from the present review:

Why Transfer Learning for energy and buildings? Higher data availability in buildings is leading more and more to a data-centric energy management with the opportunity of exploiting complex AI-based models. In this context, TL can support the penetration of ML for energy management in buildings by contributing to reduced implementation costs (i.e., pipeline of the machine learning frameworks) and time. The natural use of TL can be found in existing buildings recently equipped with monitoring infrastructure (i.e., no historical data) or new buildings (with limited historical data). However, guidance on the type and number of sensors needed to fully exploit the benefits of TL are heavily dependent on applications, and are still not clear. The present review highlights different use cases according to the applications previously discussed. For example, while aggregated building load can be estimated analysing similar buildings, hourly and sub-hourly resolution can be difficult to estimate. In this context, transfer learning can help to increase the accuracy of building load predicition even at sub-meter level, such as in NILM applications. Moreover, while building dynamics often requires physics-based approaches or a lot of data, limiting their adoption, transfer learning can speed-up and overcome data availability, allowing for an effective coupling with advanced control strategy. The same considerations can be performed for occupancy detection, that can benefit from the application of this technique. Looking at energy systems control, the application of transfer learning is crucial to broaden the adoption of advanced control strategies, which have a bottleneck of high effort to train and tune models. In fact, these approaches are too data intensive to be applied at scale. Lastly, there are other applications in the building energy field that can benefit from transfer learning, especially classification application that can highly benefit from the comprehensive and large-volume datasets. In general, the main advantages of TL use in smart buildings are the increase of performance and the potential to scale-up and speed-up ML processes. However, compared to computer vision applications, deep learning models used for buildings are not computationally demanding, therefore further analysis is needed to assess computational advantages when the scale of the analysis is larger (e.g.,communities, districts, or cities).

When to use Transfer Learning in the built environment?

As previously said, TL find its natural application when trying to apply ML models in existing buildings with few, poor, or no historical data, as well as new buildings without historical data. However, its application is further complicated by the type of task to be completed. A prerequisite for TL applications is a certain degree of similarity among the two domains; however, except for a few studies in building load prediction that tried to quantify time-series analogies, no studies have quantified the specific features importance on the similarity, and those studies are needed.In particular, looking at building load prediction and building dynamics estimation, similarity plays a key role, since buildings can have similar (or different) shape, use, climatic condition, and equipment that, depending on the considered task, may have more or less influence. Therefore, to fully understand the advantages of transfer learning applications in building load prediction and dynamics estimation, a proper definition of similarity must be defined, to contour the range of application of transfer learning. Looking at occupancy detection and activity recognition, transfer learning can be used to overcome data unavailability for specific room in a building, leveraging information from other rooms. In this context, another research question to be explored is related to the opportunity of setting minimum requirements in terms of data availability that can effectively enable a parameter-based TL. Lastly, control application may benefit from transfer learning when buildings are subject to a retrofit of energy systems and the optimal control strategy may obtain a significant jumpstart using the initial control policy from a similar building.

What are the challenges?

Summarizing the previous questions, the review highlighted the main challenges related to the application of TL in buildings. Some challenges are common to the different tasks and related to the models, while others are related to specific applications. In particular, further studies are necessary to propose robust methods on how to select the right source building, quantifying the similarity between buildings, thus avoiding negative transfer. Some solutions have been proposed and explored [132,133]; for instance, Ahmed et al. [105] proposed to use an auxiliary discriminator to determine whether the source domain and the target domain follow the same distribution. TL will be applied only when the statistical distance between the source and the target domains in the feature space are small enough. Despite these attempts, there are no well recognized standards or principles, and guidance is needed in this regard.

In particular, looking at parameter-based TL, it is not yet clear which of the feature-extraction and weight-initialization brings the greatest benefits in smart buildings applications. In particular, the analysis shows that the feature-extraction is much more used for classification problems, while for regression problems there is not enough evidence, representing one of the challenges to be overcome to increase the effectiveness of transfer learning.

Another common question that still needs to be addressed is related to the amount of data necessary in the source building and the amount of data necessary to properly transfer knowledge in the target buildings. This becomes even more true when considering applications that can be highly dependent on seasonality, such as building dynamics, systems control or load prediction. Regarding specific application, in RL there are a lot of unexplored TL settings, that in principle can drastically increase the application of such control strategies in the built environment. Despite the potential of TL for energy systems control, one of the main challenges is related to ensuring a safe control policy that does not preclude user comfort or increase costs associated with control. Furthermore, as pointed out by [35], the effectiveness of transferring high-level knowledge in energy systems still need to be assessed.

5.2. Future opportunities

Based on this literature review, transfer learning has demonstrated its potential to enhance data efficiency, accelerate training speed, and increase model accuracy. To promote the application of TL in the smart building field, comprehensive, large-volume, well-recognized datasets can be very helpful. A good example is how ImageNet advanced the field of computer vision (CV) through transfer learning [134]. ImageNet was used to train CV models, which can be transferred to other tasks with some fine tuning. In the building field, the Building Data Genome Project [135] can be really helpful in promoting a more robust TL process from two perspectives. First, models learned from more buildings are more likely to be generalizable. Second, the availability of data from a very high number of buildings can be used to help perform a benchmarking analysis to compare different TL methods. In contrast to the Building Data Genome Project, which focuses on hundreds of different buildings, [136] contains about 1400 simulations of a medium office considering several climatic conditions, energy efficiency levels of the building and systems, and different occupant behaviour, which can be used to analyse the effects of different features on building dynamics or control applications.

The second opportunity for applying TL in the building field is the potential to scale-out advanced controls in single buildings and ease the scale-up of multiple buildings' controllers with multi-agent architectures towards data-driven energy communities. Indeed, the vast majority of

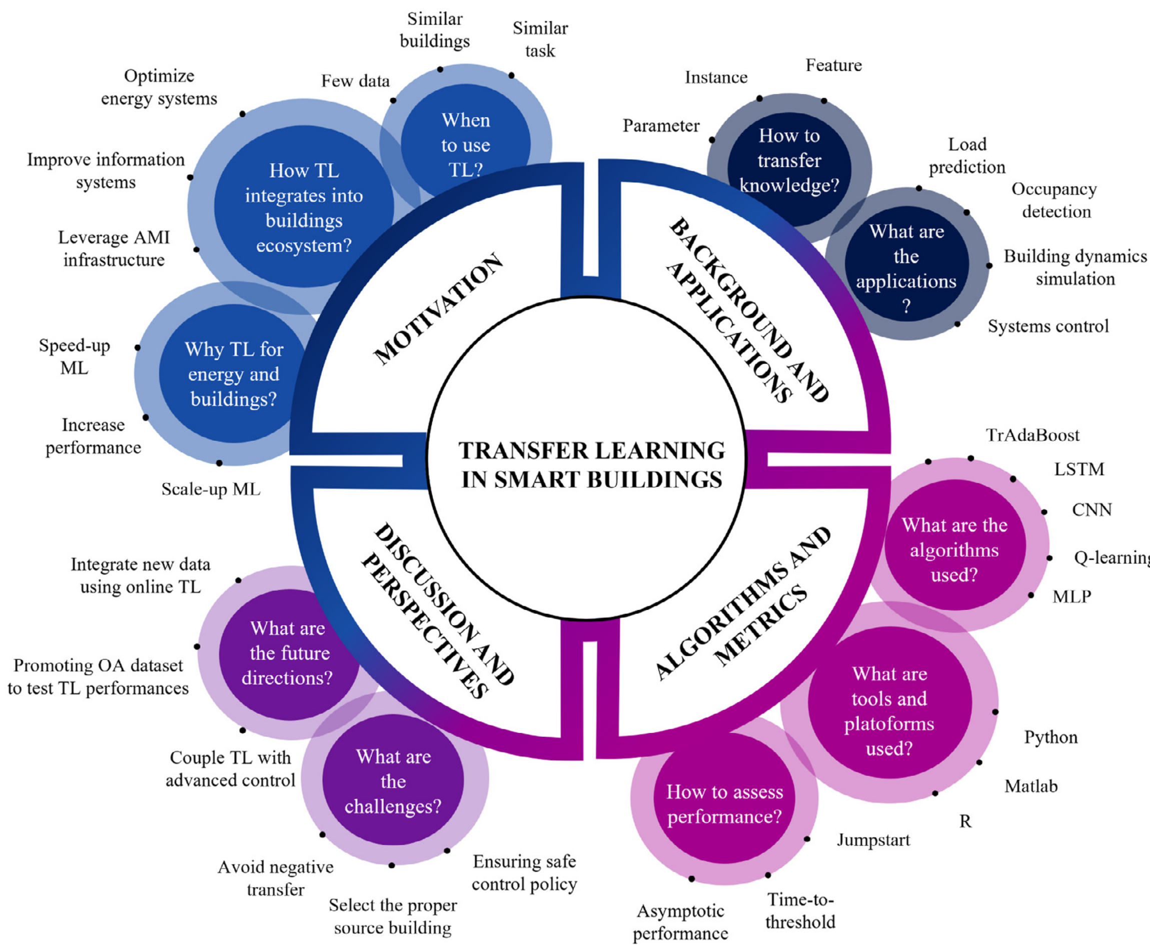

Fig. 13. Research questions and short answers

existing studies are mainly focused on supervised learning tasks (such as prediction), overlooking the potential of TL for control. The major pain-point of applying data-driven controllers (such as reinforcement learning) to buildings is that it takes too much time and data to train the controller; however, controllers in buildings have a certain degree of similarity, which provide an opportunity that can be leveraged by using TL. In this context, TL can help to scale data-driven control by transferring the knowledge of the controller or transferring the knowledge of the building dynamics (e.g., differentiable predictive control [137]), as done in Grubinger et al. [138], which integrated online TL to model building dynamics for control purposes.

The third research opportunity is about benchmarking the performance of TL. Researchers applied different TL techniques in different tasks, all claiming TL can help to improve accuracy or accelerate training. However, a standardized evaluation framework to benchmark and compare different TL approaches is still needed. The standardized evaluation framework needs to include (a) a couple of well defined building-related tasks, (b) prepared datasets for the source and task domain, and (c) well-recognized performance metrics to quantify the benefits of TL. Many research questions still need to be answered to develop this evaluation framework. Taking the metrics, for example, it might be easy to quantify the accuracy improvement, but consensus about how to quantify the training acceleration is still lacking.

5.3. Contribution and limitations

This paper presents a comprehensive review on applying transfer learning in smart buildings, which has not been done before. This review focuses on various categorizations of TL and how TL is used to solve different tasks in the smart building field. By doing so, it provides a clear picture of which categorization of TL is the most widely used in the smart building field, and which categorization might demand more research attention; which tasks use TL more frequently; how TL performs in different tasks; and more. Based on this review, we identified the current research trends and future research opportunities on this topic.

One limitation of this study is that the smart building related application is a vague term. For instance [114] applied transfer learning to recognize a type of historical building in Peru. This is a typical computer vision task, rather than a smart building task. But to make sure this review is as inclusive as possible, we included this study in this review under the category of ”others.” Therefore in the ”others” category, there are some nontypical smart building related applications, which readers might need to pay attention to.

Another limitation is the algorithm level of details present in the reviewed papers, which does not always include clear information about neural network architectures and hyper-parameter settings, thus limiting capability to reproduce results and to conduct critical analysis on each study.

6. Conclusions

This review focuses on applications and algorithms of transfer learning in smart buildings, which has been identified as a promising technique to scale up the adoption of machine learning models in real-world applications.

The study first presented the main applications of machine learning in buildings, divided into energy information systems and automated system optimization, followed by a theoretical background on transfer learning and its classifications. These two concepts were used to perform a systematic review with the aim of identifying the most common applications and relating them with the type of transfer learning and the techniques adopted.

The review analyzed 77 papers, leading to the conclusion that the main applications can be categorized in four groups: (1) building load prediction, (2) occupancy detection and activity recognition, (3) building dynamics prediction, and (4) energy systems control.

The most adopted application was found to be in building load prediction; however, despite using real data, few studies were deployed in the real world, highlighting the necessity to fully integrate transfer learning in an end-to-end machine learning pipeline, ensuring a certain quality even in the presence of few data. To this end, techniques such as data augmentation and online learning seem to represent the best ways to deal with the problem.

Moreover, the review highlighted the use of different categorizations of transfer learning, with the most and least used solutions; respectively, parameter-based and relation-based transfer learning. Given the prevalence of the contribution of parameter-based transfer learning (over 50% of the papers), a specific analysis on deep learning models was carried out, assessing the recent use of LSTM and CNN in regression and classification tasks.

Finally, Fig. 13 summarizes the research questions, along with condensed answers based on the results of the review. Research gaps and future directions have been identified, as follows:

- There is still no clear way to identify the right ”source building” to perform an effective transfer learning. Further studies on how different monitoring infrastructure, use destination, climate condition, building features, and data availability affects transfer learning results are required, specifically in building dynamics and energy systems control.
- Further analysis on the application of transfer learning in building dynamics simulation and energy systems control can unveil transfer learning potential to scale-up data-driven energy management in buildings. However, the safety of the transferred control policy or building dynamic models should be better investigated.
- Workflow standardization and open source codes are fundamental to increase the reproducibility of the works. It is needed to define common guidelines to evaluate transfer learning performance across different applications (e.g., regression, classification) and create high-quality datasets to benchmark different transfer learning approaches.

Declaration of Competing Interest

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

Acknowledgment

Dr. Zhe Wang and Dr. Tianzhen Hong’s work was supported by the Laboratory Directed Research and Development (LDRD) Program of Lawrence Berkeley National Laboratory, provided by the Director, Office of Science, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Appendix A

The following tables describe the analysed literature based on their application. To provide a more readable table, the following abbreviations are used: in the column NN, that describes the two subcategories of parameter-based TL, Weights is used to represent weight initialization, while Features is used to represent feature-extraction (note that ”Features” in the NN column should not be confused with ”Feature” in the previous column, which represents feature-based TL).

References

[1] European Commission. The European Green Deal. 2019. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52019DC0640&from=EN.
[2] Victoria M, Zhu K, Brown T, Andresen GB, Greiner M, et al. Early decarbonisation of the european energy system pays off. Nat Commun 2020;11(1):1–9. arXiv:2004.11009. doi:10.1038/s41467-020-20015-4.
[3] Satchwell A., Piette M.A., Khandekar A., Granderson J., Frick N.M., Hledik R., Faruqui A., Lam L., Ross S., Cohen J., et al. A national roadmap for grid-interactive efficient buildings2021;.
[4] Aggarwal S, Orvis R. Grid flexibility: methods for modernizing the power grid. Energy Innovation San Francisco, California March 2016.
[5] Kerscher S, Arboleya P. The key role of aggregators in the energy transition under the latest european regulatory framework. International Journal of Electrical Power & Energy Systems 2022;134:107361. doi:10.1016/j.ijepes.2021.107361. URL: https://www.sciencedirect.com/science/article/pii/S0142061521006001
[6] Andoni M, Robu V, Flynn D, Abram S, Geach D, Jenkins D, McCallum P, Peacock A. Blockchain technology in the energy sector: asystematic review of challenges and opportunities. Renewable Sustainable Energy Rev 2019;100:143–74. doi:10.1016/j.rser.2018.10.014. URL: https://www.sciencedirect.com/science/article/pii/S1364032118307184
[7] Zhou S, Brown MA. Smart meter deployment in europe: a comparative case study on the impacts of national policy schemes. J Clean Prod 2017;144:22–32. doi:10.1016/j.jclepro.2016.12.031. URL: https://www.sciencedirect.com/science/article/pii/S0959652616320868
[8] Capozzoli A, Cerquitelli T, Piscitelli M. Enhancing energy efficiency in buildings through innovative data analytics technologies. In: Dobre C, Xhafa F, editors. Next Generation Platforms for Intelligent Data Collection; 2016. p. 353–89.
[9] Molina-Solana M, Ros M, Ruiz MD, Gómez-Romero J, Martin-Bautista MJ. Data science for building energy management: areview. Renewable Sustainable Energy Rev 2017;70:598–609. doi:10.1016/j.rser.2016.11.132. URL: https://www.sciencedirect.com/science/article/pii/S1364032116308814
[10] Antonopoulos I, Robu V, Couraud B, Kirli D, Norbu S, Kiprakis A, Flynn D, Elizondo-Gonzalez S, Wattam S. Artificial intelligence and machine learning approaches to energy demand-side response: a systematic review. Renewable Sustainable Energy Rev 2020;130:109899. doi:10.1016/j.rser.2020.109899. URL: https://www.sciencedirect.com/science/article/pii/S136403212030191X
[11] Hong T, Wang Z, Luo X, Zhang W. State-of-the-art on research and applications of machine learning in the building life cycle. Energy Build 2020. doi:10.1016/j.enbuild.2020.109831.
[12] Amasyali K, El-Gohary NM. A review of data-driven building energy consumption prediction studies. Renewable Sustainable Energy Rev 2018;81:1192–205. doi:10.1016/j.rser.2017.04.095. URL: https://www.sciencedirect.com/science/article/pii/S1364032117306093
[13] Kathirgamanathan A, De Rosa M, Mangina E, Finn DP. Data-driven predictive control for unlocking building energy flexibility: a review. Renewable Sustainable Energy Rev 2021;135:110120. doi:10.1016/j.rser.2020.110120. URL: https://www.sciencedirect.com/science/article/pii/S1364032120304111
[14] Drgoňa J, Arroyo J, Cupeiro Figueroa I, Blum D, Arendt K, Kim D, Ollé EP, Oravec J, Wetter M, Vrabie DL, Helsen L. All you need to know about model predictive control for buildings. Annu Rev Control 2020;50:190–232. doi:10.1016/j.arcontrol.2020.09.001. URL: https://www.sciencedirect.com/science/article/pii/S1367578820300584
[15] Serale G, Fiorentini M, Capozzoli A, Bernardini D, Bemporad A. Model predictive control (MPC) for enhancing building and HVAC system energy efficiency: problem formulation, applications and opportunities. Energies 2018;11(3). doi:10.3390/en11030631.
[16] Kim W, Katipamula S. A review of fault detection and diagnostics methods for building systems. Science and Technology for the Built Environment 2018;24(1):3–21. doi:10.1080/23744731.2017.1318008.
[17] Himeur Y, Ghanem K, Alsalemi A, Bensaali F, Amira A. Artificial intelligence based anomaly detection of energy consumption in buildings: a review, current trends and new perspectives. Appl Energy 2021;287:116601. doi:10.1016/j.apenergy.2021.116601. URL: https://www.sciencedirect.com/science/article/pii/S0306261921001409
[18] Li Z, Han Y, Xu P. Methods for benchmarking building energy consumption against its past or intended performance: an overview. Appl Energy 2014;124:325–34. doi:10.1016/j.apenergy.2014.03.020. URL: https://www.sciencedirect.com/science/article/pii/S0306261914002505
[19] Capozzoli A, Piscitelli MS, Neri F, Grassi D, Serale G. A novel methodology for energy performance benchmarking of buildings by means of linear mixed effect model: the case of space and DHW heating of out-patient healthcare centres. Appl Energy 2016;171:592–607. doi:10.1016/j.apenergy.2016.03.083. URL: https://www.sciencedirect.com/science/article/pii/S0306261916304093
[20] Chicco G. Overview and performance assessment of the clustering methods for electrical load pattern grouping. Energy 2012;42(1):68–80. doi:10.1016/j.energy.2011.12.031. 8th World Energy System Conference, WESC 2010, URL: https://www.sciencedirect.com/science/article/pii/S0360544211008565
[21] Wang Y, Chen Q, Kang C, Zhang M, Wang K, Zhao Y. Load profiling and its application to demand response: a review. Tsinghua Sci Technol 2015;20(2):117–29. doi:10.1109/TST.2015.7085625.
[22] Piscitelli MS, Brandi S, Capozzoli A. Recognition and classification of typical load profiles in buildings with non-intrusive learning approach. Appl Energy 2019;255:113727. doi:10.1016/j.apenergy.2019.113727. URL: https://www.sciencedirect.com/science/article/pii/S030626191931414X

[23] He Z, Hong T, Chou SK. A framework for estimating the energy-saving potential of occupant behaviour improvement. Appl Energy 2021;287:116591. doi:10.1016/j.apenergy.2021.116591. URL: https://www.sciencedirect.com/science/article/pii/S0306261921001343
[24] Kramer H., Curtin C., Lin G., Crowe E., Granderson J.. Proving the business case for building analytics. 2020.. 10.20357/B7G022
[25] Lee D, Cheng C-C. Energy savings by energy management systems: areview. Renewable Sustainable Energy Rev 2016;56:760–77. doi:10.1016/j.rser.2015.11.067. URL: https://www.sciencedirect.com/science/article/pii/S1364032115013349
[26] Wang Z, Hong T. Reinforcement learning for building controls: the opportunities and challenges. Appl Energy 2020. doi:10.1016/j.apenergy.2020.115036.
[27] Vazquez-Canteli J, Dey S, Henze G, Nagy Z. Citylearn: standardizing research in multi-Agent reinforcement learning for demand response and urban energy management. tbd 2021. arXiv:2012.10504.
[28] Pinto G, Piscitelli MS, Vázquez-Canteli JR, Nagy Z, Capozzoli A. Coordinated energy management for a cluster of buildings through deep reinforcement learning. Energy 2021;229:120725. doi:10.1016/j.energy.2021.120725. URL: https://www.sciencedirect.com/science/article/pii/S0360544221009737
[29] Xiong L, Li P, Wang Z, Wang J. Multi-agent based multi objective renewable energy management for diversified community power consumers. Appl Energy 2020;259:114140. doi:10.1016/j.apenergy.2019.114140. URL: https://www.sciencedirect.com/science/article/pii/S0306261919318276
[30] Stoyanova I, Jendoubi I, Monti A. Model predictive control for cooperative energy management at city-district level. In: 2018 Power Systems Computation Conference (PSCC); 2018. p. 1–7. doi:10.23919/PSCC.2018.8442619.
[31] Parisio A, Wiezorek C, Kyntäjä T, Elo J, Strunz K, Johansson KH. Cooperative MPC-based energy management for networked microgrids. IEEE Trans Smart Grid 2017;8(6):3066–74. doi:10.1109/TSG.2017.2726941.
[32] Drgoňa J, Picard D, Kvasnica M, Helsen L. Approximate model predictive building control via machine learning. Appl Energy 2018;218:199–216. doi:10.1016/j.apenergy.2018.02.156. URL: https://www.sciencedirect.com/science/article/pii/S0306261918302903
[33] Pinto G, Deltetto D, Capozzoli A. Data-driven district energy management with surrogate models and deep reinforcement learning. Appl Energy 2021;304:117642. doi:10.1016/j.apenergy.2021.117642. URL: https://www.sciencedirect.com/science/article/pii/S0306261921010096
[34] Ali SMM, Augusto JC, Windridge D. A survey of user-Centred approaches for smart home transfer learning and new user home automation adaptation. Applied Artificial Intelligence 2019;33(8):747–74. doi:10.1080/08839514.2019.1603784.
[35] Peirelinck T, Kazmi H, Mbuwir BV, Hermans C, Spiessens F, Suykens J, Deconinck G. Transfer learning in demand response: a review of algorithms for data-efficient modelling and control. Energy and AI 2022;7:100126. doi:10.1016/j.egyai.2021.100126. URL: https://www.sciencedirect.com/science/article/pii/S2666546821000732
[36] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2010;22(10):1345–59. doi:10.1109/TKDE.2009.191.
[37] Jung SM, Park S, Jung SW, Hwang E. Monthly electric load forecasting using transfer learning for smart cities. Sustainability (Switzerland) 2020;12(16). doi:10.3390/SU12166364.
[38] Zhang F, Bales C, Fleyeh H. From time series to image analysis: a transfer learning approach for night setback identification of district heating substations. Journal of Building Engineering 2021;43(April):102537. doi:10.1016/j.jobe.2021.102537.
[39] Chiang Y-T, Lu C-H, Hsu JY-j. A feature-based knowledge transfer framework for cross-environment activity recognition toward smart home applications. IEEE Trans Hum Mach Syst 2017;47(3):310–22.
[40] Grubinger T, Chasparis GC, Natschläger T. Generalized online transfer learning for climate control in residential buildings. Energy Build 2017;139:63–71. doi:10.1016/j.enbuild.2016.12.074.
[41] Maschler B, Weyrich M. Deep transfer learning for industrial automation: areview and discussion of new techniques for data-driven machine learning. IEEE Ind Electron Mag 2021;15(2):65–75. doi:10.1109/MIE.2020.3034884.
[42] Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. A survey on deep transfer learning. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 2018;11141 LNCS:270–9. arXiv:1808.01974. doi:10.1007/978-3-030-01424-7_27.
[43] Weiss K, Khoshgoftaar TM, Wang DD. A survey of transfer learning, vol 3. Springer International Publishing; 2016. ISBN 4053701600. doi:101186/s40537-016-0043-6.
[44] Day O, Khoshgoftaar TM. A survey on heterogeneous transfer learning. J Big Data 2017;4(1). doi:10.1186/s40537-017-0089-0.
[45] Zhuang F, Qi Z, Duan K, Xi D, Zhu Y, Zhu H, et al. A comprehensive survey on transfer learning. Proc IEEE 2021;109(1):43–76. arXiv:1911.02685. doi:10.1109/JPROC.2020.3004555.
[46] Liang H, Fu W, Yi F. A survey of recent advances in transfer learning. International Conference on Communication Technology Proceedings, ICCT 2019:1516–23. doi:10.1109/ICCT46805.2019.8947072.
[47] Zhang W., Deng L., Zhang L. , Wu D. A survey on negative transfer. arXiv:2009.00909.
[48] Wang Z., Dai Z., Póczos B., Carbonell J. Characterizing and avoiding negative transfer. arXiv:1811.09751
[49] Lazaric A, Lazaric A, Learning R, Marco S. Transfer in reinforcement learning : a framework and a survey; 2013.
[50] Taylor ME, Stone P. Transfer learning for reinforcement learning domains: a survey. Journal of Machine Learning Research 2009;10(56):1633–85. URL: http://jmlr.org/papers/v10/taylor09a.html
[51] Zhu Z., Lin K., Zhou J. arXiv:2009.07888. 1–23.
[52] Sutton and Barton RS, Andrew G. Reinforcement learning: an introduction. 2nd. Cambridge, Massachusetts: MIT Press; 2014.
[53] Zhang L, Wen J, Li Y, Chen J, Ye Y, Fu Y, Livingood W. A review of machine learning in building load prediction. Appl Energy 2021;285:116452. doi:10.1016/j.apenergy.2021.116452. URL: https://www.sciencedirect.com/science/article/pii/S0306261921000209
[54] Houidi S, Fourer D, Auger F, Sethom HBA, Miègeville L. Comparative evaluation of non-intrusive load monitoring methods using relevant features and transfer learning. Energies 2021;14(9):2726.
[55] Qian F, Gao W, Yang Y, Yu D, et al. Potential analysis of the transfer learning model in short and medium-term forecasting of building HVAC energy consumption. Energy 2020;193(C).
[56] Oliveira P, Fernandes B, Analide C, Novais P. Forecasting energy consumption of wastewater treatment plants with a transfer learning approach for sustainable cities. Electronics (Basel) 2021;10(10):1149.
[57] Fang X, Gong G, Li G, Chun L, Li W, Peng P. A hybrid deep transfer learning strategy for short term cross-building energy prediction. Energy 2021;215:119208. doi:10.1016/j.energy.2020.119208.
[58] Hooshmand A, Sharma R. Energy predictive models with limited data using transfer learning. In: Proceedings of the Tenth ACM International Conference on Future Energy Systems; 2019. p. 12–16.
[59] Lu Y, Tian Z, Zhou R, Liu W. A general transfer learning-based framework for thermal load prediction in regional energy system. Energy 2021;217:119322. doi:10.1016/j.energy.2020.119322.
[60] D'Incecco M, Squartini S, Zhong M. Transfer learning for non-Intrusive load monitoring. IEEE Trans Smart Grid 2020;11(2):1419–29. arXiv:1902.08835. doi:10.1109/TSG.2019.2938068.
[61] Liu Y, Wang X, You W. Non-intrusive load monitoring by voltage–current trajectory enabled transfer learning. IEEE Trans Smart Grid 2018;10(5):5609–19.
[62] Fan C, Sun Y, Xiao F, Ma J, Lee D, Wang J, et al. Statistical investigations of transfer learning-based methodology for short-term building energy predictions. Appl Energy 2020;262:114499.
[63] Cai L, Wen H, Gu J, Ma J, Jin Z. Forecasting customers' response to incentives during peak periods: a transfer learning approach. International Transactions on Electrical Energy Systems 2020;30(7):1–19. doi:10.1002/2050-7038.12251.
[64] Fan L, Zhang J, He Y, Liu Y, Hu T, Zhang H. Optimal scheduling of microgrid based on deep deterministic policy gradient and transfer learning. Energies 2021;14(3):1–15. doi:10.3390/en14030584.
[65] Lissa P, Schukat M, Keane M, Barrett E. Transfer learning applied to DRL-based heat pump control to leverage microgrid energy efficiency. Smart Energy 2021:100044. doi:10.1016/j.segy.2021.100044. URL: https://www.sciencedirect.com/science/article/pii/S2666955221000447
[66] Mbuwir BV, Paridari K, Spiessens F, Nordström L, Deconinck G. Transfer learning for operational planning of batteries in commercial buildings. In: 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids (SmartGridComm). IEEE; 2020. p. 1–6.
[67] Lissa P, Schukat M, Barrett E. Transfer learning applied to reinforcement learning-based HVAC control. SN Computer Science 2020;1. doi:10.1007/s42979-020-00146-7.
[68] Xu S, Wang Y, Wang Y, O'Neill Z, Zhu Q. One for many: Transfer learning for building hvac control. In: Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation; 2020. p. 230–9.
[69] Zhang X, Jin X, Tripp C, Biagioni DJ, Graf P, Jiang H. Transferable reinforcement learning for smart homes. In: Proceedings of the 1st International Workshop on Reinforcement Learning for Energy Management in Buildings & Cities; 2020. p. 43–47.
[70] Tsang N, Cao C, Wu S, Yan Z, Yousefi A, Fred-Ojala A, et al. Autonomous household energy management using deep reinforcement learning. In: 2019 IEEE International Conference on Engineering, Technology and Innovation (ICE/ITMC). IEEE; 2019. p. 1–7.
[71] Jiang Z, Lee YM. Deep transfer learning for thermal dynamics modeling in smart buildings. In: 2019 IEEE International Conference on Big Data (Big Data). IEEE; 2019. p. 2033–7.
[72] Chen Y, Tong Z, Zheng Y, Samuelson H, Norford L. Transfer learning with deep neural networks for model predictive control of HVAC and natural ventilation in smart buildings. J Clean Prod 2020;254:119866.
[73] Chen Y, Zheng Y, Samuelson H. Fast adaptation of thermal dynamics model for predictive control of HVAC and natural ventilation using transfer learning with deep neural networks. In: 2020 American Control Conference (ACC). IEEE; 2020. p. 2345–50.
[74] Hossain MM, Zhang T, Ardakanian O. Evaluating the feasibility of reusing pre-trained thermal models in the residential sector. UrbSys 2019 - Proceedings of the 1st ACM International Workshop on Urban Building Energy Sensing, Controls, Big Data Analysis, and Visualization, Part of BuildSys 2019 2019:23–32. doi:10.1145/3363459.3363529.
[75] Kazmi H, Suykens J, Driesen J. Large-scale transfer learning for data-driven modelling of hot water systems. Building Simulation Conference Proceedings 2019;4:2611–18. doi:10.26868/25222708.2019.210352.
[76] Hu W, Luo Y, Lu Z, Wen Y. Heterogeneous transfer learning for thermal comfort modeling. In: Proceedings of the 6th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation; 2019. p. 61–70.
[77] Khalil M, McGough S, Pourmirza Z, Pazhoohesh M, Walker S. Transfer Learning Approach for Occupancy Prediction in Smart Buildings. In: 2021 12th International Renewable Engineering Conference, IREC 2021; 2021. p. 0–5. doi:10.1109/IREC51415.2021.9427869.
[78] Pardamean B, Muljo HH, Cenggoro TW, Chandra BJ, Rahutomo R. Using trans-

fer learning for smart building management system. J Big Data 2019;6(1). doi:10.1186/s40537-019-0272-6.
[79] Mosaico G, Saviozzi M, Silvestro F, Bagnasco A, Vinci A. Simplified state space building energy model and transfer learning based occupancy estimation for HVAC optimal control. 5th International Forum on Research and Technologies for Society and Industry: Innovation to Shape the Future, RTSI 2019 - Proceedings 2019:353–8. doi:10.1109/RTSI.2019.8895544.
[80] Weber M, Doblander C, Mandl P. Detecting building occupancy with synthetic environmental data. In: Proceedings of the 7th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation; 2020. p. 324–5.
[81] Arief-Ang IB, Hamilton M, Salim FD. A scalable room occupancy prediction with transferable time series decomposition of CO2 sensor data. ACM Transactions on Sensor Networks (TOSN) 2018;14(3–4):1–28.
[82] Anjomshoaa A, Curry E. Transfer learning in smart environments. Machine Learning and Knowledge Extraction 2021;3(2):318–32.
[83] Wang W, Miao C. Activity recognition in new smart home environments. In: Proceedings of the 3rd International Workshop on Multimedia for Personal Health and Health Care; 2018. p. 29–37.
[84] Inoue S, Pan X. Supervised and unsupervised transfer learning for activity recognition from simple in-home sensors. ACM International Conference Proceeding Series 2016;28-November-2016:20–7. doi:10.1145/2994374.2994400.
[85] Kumar A, Yordanova K, Kirste T, Kumar M. Combining off-the-shelf image classifiers with transfer learning for activity recognition. ACM International Conference Proceeding Series 2018. doi:10.1145/3266157.3266219.
[86] Chen W-H, Cho P-C, Jiang Y-L. Activity recognition using transfer learning. Sens Mater 2017;29(7):897–904.
[87] Leeraksakiat P, Pora W. Occupancy Forecasting using LSTM Neural Network and Transfer Learning. In: 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON 2020; 2020. p. 470–3. doi:10.1109/ECTI-CON49241.2020.9158103.
[88] Niu H, Nguyen D, Yonekawa K, Kurokawa M, Wada S, Yoshihara K. Multi-source Transfer Learning for Human Activity Recognition in Smart Homes. In: Proceedings - 2020 IEEE International Conference on Smart Computing, SMARTCOMP 2020; 2020. p. 274–7. doi:10.1109/SMARTCOMP50058.2020.00063.
[89] Dowling CP, Zhang B. Transfer learning for HVAC system fault detection. In: 2020 American Control Conference (ACC). IEEE; 2020. p. 3879–85.
[90] Liu J, Zhang Q, Li X, Li G, Liu Z, Xie Y, Li K, Liu B. Transfer learning-based strategies for fault diagnosis in building energy systems. Energy Build 2021;250:111256. doi:10.1016/j.enbuild.2021.111256. URL: https://www.sciencedirect.com/science/article/pii/S0378778821005405
[91] Hong D, Wang H, Ortiz J, Whitehouse K. The building adapter: Towards quickly applying building analytics at scale. In: Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments; 2015. p. 123–32.
[92] Ma J, Cheng JCP, Jiang F, Chen W, Wang M, Zhai C. A bi-directional missing data imputation scheme based on LSTM and transfer learning for building energy data. Energy Build 2020;216. doi:10.1016/j.enbuild.2020.109941.
[93] Deng Z, Chen Q. Reinforcement learning of occupant behavior model for cross-building transfer learning to various HVAC control systems. Energy Build 2021;238:110860. doi:10.1016/j.enbuild.2021.110860.
[94] Huang T, Jiao F. Study on data transfer in meteorological forecast of small and medium-sized cities and its application in zhaoqing city. Comput J 2020;63(7):1076–83.
[95] Gao N, Shao W, Rahaman MS, Zhai J, David K, Salim FD. Transfer learning for thermal comfort prediction in multiple cities. Build Environ 2021;195(January):107725. arXiv:2004.14382. doi:10.1016/j.buildenv.2021.107725.
[96] Cai L, Gu J, Ma J, Jin Z. Probabilistic wind power forecasting approach via instance-based transfer learning embedded gradient boosting decision trees. Energies 2019;12(1):159.
[97] Abdi G, Jabari S. A multi-Feature fusion using deep transfer learning for earthquake building damage detection. Canadian Journal of Remote Sensing 2021;0(0):1–19. doi:10.1080/07038992.2021.1925530.
[98] Masrour T, El Hassani I, Bouchama MS. Deep convolutional neural networks with transfer learning for old buildings pathologies automatic detection. In: Ezziyyani M, editor. Advanced Intelligent Systems for Sustainable Development (AI2SD'2019). Cham: Springer International Publishing; 2020. p. 204–16. ISBN 978-3-030-36671-1.
[99] Zyout I, Oatawneh A. Detection of PV solar panel surface defects using transfer learning of the deep convolutional neural networks. In: 2020 Advances in Science and Engineering Technology International Conferences (ASET). IEEE; 2020. p. 1–4.
[100] Mao J, Yu L, Yu H, Hu Y, Sheng W. A Transfer Learning Method with Multi-feature Calibration for Building Identification. In: Proceedings of the International Joint Conference on Neural Networks; 2020.
[101] Zhao P, Zettsu K. Decoder Transfer Learning for Predicting Personal Exposure to Air Pollution. In: Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019; 2019. p. 5620–9. doi:10.1109/BigData47090.2019.9006604.
[102] Paudel P, Kim S, Park S, Choi KH. A context-aware iot and deep-learning-based smart classroom for controlling demand and supply of power load. Electronics (Switzerland) 2020;9(6):1–12. doi:10.3390/electronics9061039.
[103] RashidiParisa and Cook D. Multi Home Transfer Learning for Resident Activity Discovery and Recognition. In: Proceedings of the International Workshop on Knowledge Discovery from Sensor Data; 2010. p. 53–63.
[104] Moon J, Kim J, Kang P, Hwang E. Solving the cold-start problem in short-term load forecasting using tree-based methods. Energies 2020;13(4). doi:10.3390/en13040886.
[105] Ahmed AMA, Zhang Y, Eliassen F. Generative adversarial networks and transfer learning for non-intrusive load monitoring in smart grids. In: 2020 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids, SmartGridComm 2020; 2020.
[106] Sonia S, Baruah RD. Transfer Learning in Smart Home Scenario. In: Proceedings of the International Joint Conference on Neural Networks; 2020.
[107] Chernov A, Butakova M, Kostyukov A. Intelligent Decision Support for Power Grids Using Deep Learning on Small Datasets. In: Proceedings - 2020 2nd International Conference on Control Systems, Mathematical Modeling, Automation and Energy Efficiency, SUMMA 2020; 2020. p. 958–62. doi:10.1109/SUMMA50634.2020.9280654.
[108] Li A, Xiao F, Fan C, Hu M. Development of an ANN-based building energy model for information-poor buildings using transfer learning. Build Simul 2021;14(1):89–101. doi:10.1007/s12273-020-0711-5.
[109] Zhou D, Ma S, Hao J, Han D, Huang D, Yan S, et al. An electricity load forecasting model for integrated energy system based on biGAN and transfer learning. Energy Reports 2020;6:3446–61. doi:10.1016/j.egyr.2020.12.010.
[110] Vos M, Bender-Saebelkampf C, Albayrak S. Residential Short-Term Load Forecasting Using Convolutional Neural Networks. In: 2018 IEEE International Conference on Communications, Control, and Computing Technologies for Smart Grids, SmartGridComm 2018; 2018.
[111] Le T, Vo MT, Kieu T, Hwang E, Rho S, Baik SW. Multiple electric energy consumption forecasting using a cluster-based strategy for transfer learning in smart building. Sensors 2020;20(9). doi:10.3390/s20092668.
[112] Liu L, Yan RJ, Maruvanchery V, Kayacan E, Chen IM, Tiong LK. Transfer learning on convolutional activation feature as applied to a building quality assessment robot. Int J Adv Rob Syst 2017;14(3):1–12. doi:10.1177/1729881417712620.
[113] Xiong Y., Chen Q., Zhu M., Zhang Y., Huang K.. Accurate detection of historical buildings using aerial photographs and deep transfer learning. 2020. 1586–1589.
[114] Leon-Malpartida J, Farfan-Escobedo JD, Cutipa-Arapa GE. A new method of classification with rejection applied to building images recognition based on Transfer Learning. In: Proceedings of the 2018 IEEE 25th International Conference on Electronics, Electrical Engineering and Computing, INTERCON 2018; 2018. p. 15–18. doi:10.1109/INTERCON.2018.8526392.
[115] Wang S, Du L, Zhou Q. A semi-Supervised deep transfer learning architecture for energy disaggregation. IEEE Power and Energy Society General Meeting 2019;2019-August:6–10. doi:10.1109/PESGM40551.2019.8973556.
[116] Singaravel S, Suykens J, Geyer P. Deep-learning neural-network architectures and methods: using component-based models in building-design energy prediction. Adv Eng Inf 2018;38(June):81–90. doi:10.1016/j.aei.2018.06.004.
[117] Demianenko M, De Gaetani CI. A procedure for automating energy analyses in the BIM context exploiting artificial neural networks and transfer learning technique. Energies 2021;14(10). doi:10.3390/en14102956.
[118] Banda P, Bhuiyan MA, Zhang K, Song A. Transfer learning for leisure centre energy consumption prediction. In: Rodrigues JMF, Cardoso PJS, Monteiro J, Lam R, Krzhizhanovskaya VV, Lees MH, Dongarra JJ, Sloot PMA, editors. Computational Science – ICCS 2019. Cham: Springer International Publishing; 2019. p. 112–23. ISBN 978-3-030-22734-0.
[119] Gao Y, Ruan Y, Fang C, Yin S. Deep learning and transfer learning models of energy consumption forecasting for a building with poor information data. Energy Build 2020;223:110156. doi:10.1016/j.enbuild.2020.110156.
[120] Zhu X, Chen K, Anduv B, Jin X, Du Z. Transfer learning based methodology for migration and application of fault detection and diagnosis between building chillers for improving energy efficiency. Build Environ 2021;200(April):107957. doi:10.1016/j.buildenv.2021.107957.
[121] Ribeiro M, Grolinger K, ElYamany HF, Higashino WA, Capretz MAM. Transfer learning with seasonal and trend adjustment for cross-building energy forecasting. Energy Build 2018;165:352–63. doi:10.1016/j.enbuild.2018.01.034.
[122] Abadi M., Agarwal A., Barham P., Brevdo E., Chen Z., Citro C., Corrado G.S., Davis A., Dean J., Devin M., Ghemawat S., Goodfellow I., Harp A., Irving G., Isard M., Jia Y., Jozefowicz R., Kaiser L., Kudlur M., Levenberg J., Mané D., Monga R., Moore S., Murray D., Olah C., Schuster M., Shlens J., Steiner B., Sutskever I., Talwar K., Tucker P., Vanhoucke V., Vasudevan V., Viégas F., Vinyals O., Warden P., Wattenberg M., Wicke M., Yu Y., Zheng X.. TensorFlow: Large-scale machine learning on heterogeneous systems. 2015. Software available from tensorflow.org, URL: https://www.tensorflow.org/.
[123] Chollet F., et al. Keras. https://keras.io; 2015.
[124] Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S. Pytorch: an imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Beygelzimer A, Alché-Buc Fd, Fox E, Garnett R, editors. Advances in Neural Information Processing Systems 32. Curran Associates, Inc.; 2019. p. 8024–35. URL: http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf
[125] The MathWorks, Natick, MA, USA, Matlab deep learning toolbox. 2021b.
[126] R Core Team. R: A Language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2017. URL: https://www.R-project.org/
[127] Lee E, Rhee W. Individualized short-term electric load forecasting with deep neural network based transfer learning and meta learning. IEEE Access 2021;9:15413–25. doi:10.1109/ACCESS.2021.3053317.
[128] Ozer I, Efe SB, Ozbay H. A combined deep learning application for short term load forecasting. Alexandria Engineering Journal 2021;60(4):3807–18. doi:10.1016/j.aej.2021.02.050.
[129] Mocanu E, Nguyen PH, Kling WL, Gibescu M. Unsupervised energy prediction in

a smart grid context using reinforcement cross-building transfer learning. Energy Build 2016;116:646–55. doi:10.1016/j.enbuild.2016.01.030.

[130] Tian Y, Sehovac L, Grolinger K. Similarity-Based chained transfer learning for energy forecasting with big data. IEEE Access 2019;7:139895–908. doi:10.1109/ACCESS.2019.2943752.

[131] Le T., Vo M.T., Kieu T., Hwang E., Rho S.. Using a Cluster-Based Strategy for Transfer Learning in Smart Building

[132] Fang X, Gong G, Li G, Chun L, Peng P, Li W. A general multi-source ensemble transfer learning framework integrate of LSTM-DANN and similarity metric for building energy prediction. Energy Build 2021;252:111435. doi:10.1016/j.enbuild.2021.111435. URL: https://www.sciencedirect.com/science/article/pii/S0378778821007192

[133] Aggarwal CC. Educational and software resources for data classification. Data Classification: Algorithms and Applications 2014:657–65. doi:10.1201/b17320.

[134] Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. Ieee; 2009. p. 248–55.

[135] Miller C, Meggers F. The building data genome project: an open, public data set from non-residential building electrical meters. Energy Procedia 2017;122:439–44.

[136] Li H, Wang Z, Hong T. A synthetic building operation dataset. Sci Data 2021;8:213. doi:10.1038/s41597-021-00989-6.

[137] Drgoňa J, Tuor A, Skomski E, Vasisht S, Vrabie D. Deep learning explicit differentiable predictive control laws for buildings. IFAC-PapersOnLine 2021;54(6):14–19.

[138] Grubinger T, Chasparis GC, Natschläger T. Online transfer learning for climate control in residential buildings. 2016 European Control Conference, ECC 2016 2017(2):1183–8. doi:10.1109/ECC.2016.7810450.