Sustainable Cities and Society 85 (2022) 104059

Contents lists available at ScienceDirect

Sustainable Cities and Society

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

Next-generation energy systems for sustainable smart cities: Roles of transfer learning

Yassine Himeur [a,*], Mariam Elnour [a], Fodil Fadli [a], Nader Meskin [b], Ioan Petri [c], Yacine Rezgui [c], Faycal Bensaali [b], Abbes Amira [d,e]

[a] Department of Architecture & Urban Planning, Qatar University, Doha, Qatar
[b] Department of Electrical Engineering, Qatar University, Doha, Qatar
[c] School of Engineering, BRE Institute of Sustainable Engineering, Cardiff University, Wales, United Kingdom
[d] Department of Computer Science, University of Sharjah, United Arab Emirates
[e] Institute of Artificial Intelligence, De Montfort University, Leicester, United Kingdom

ARTICLE INFO

Keywords:
Artificial intelligence
Transfer learning
Energy systems for sustainable smart cities
Deep transfer learning
Domain adaptation
Computing platforms

ABSTRACT

Smart cities attempt to reach net-zero emissions goals by reducing wasted energy while improving grid stability and meeting service demand. This is possible by adopting next-generation energy systems, which leverage artificial intelligence, the Internet of things (IoT), and communication technologies to collect and analyze big data in real-time and effectively run city services. However, training machine learning algorithms to perform various energy-related tasks in sustainable smart cities is a challenging data science task. These algorithms might not perform as expected, take much time in training, or do not have enough input data to generalize well. To that end, transfer learning (TL) has been proposed as a promising solution to alleviate these issues. To the best of the authors' knowledge, this paper presents the first review of the applicability of TL for energy systems by adopting a well-defined taxonomy of existing TL frameworks. Next, an in-depth analysis is carried out to identify the pros and cons of current techniques and discuss unsolved issues. Moving on, two case studies illustrating the use of TL for (i) energy prediction with mobility data and (ii) load forecasting in sports facilities are presented. Lastly, the paper ends with a discussion of the future directions.

1. Introduction

1.1. Preliminary

Developing smart and sustainable cities represents a significant and challenging approach to curtail wasted energy and promote a transition towards clean energy using digitization (Carrera, Peyrard, & Kim, 2021). This has become possible using next-generation energy systems, requires new characteristics and functionalities to manage smart city services, satisfy consumers' needs, improve resilience and enhance energy efficiency. In this respect, increasing the reliability and resilience of energy systems has become a significant concern, where digitalization is a crucial player (Himeur, Ghanem, Alsalemi, Bensaali and Amira, 2021; Kathirgamanathan, De Rosa, Mangina, & Finn, 2021). Therefore, energy systems have recently undergone several changes due to the advance of the smart grids, adoption of artificial intelligence (AI) and the internet of things (IoT), integration of renewable energies (e.g., wind and solar photovoltaic), and electric vehicles (EV), and deployment of cyber-security measures (e.g., blockchain) (Elnour et al., 2022; Singh et al., 2020).

AI-based big data analytics have become essential in the energy sector as they play a crucial role in developing the next generation of energy systems (Himeur et al., 2021). The digitalization of the energy sector and the large amount of data produced by energy systems require powerful and intelligent tools to make the energy industry more efficient and secure (Petri, Rana, Rezgui, & Fadli, 2021). Thus, AI-based big data analytics tools found their way by effectively analyzing and evaluating large-scale datasets (Varlamis et al., 2022). Specifically, areas of application are numerous, among them building energy management systems (BEMSs) (Fan, Sun, Zhao, Song, & Wang, 2019; Qin, Ke, Wang, & Fedorovich, 2021), smart grid (SG) and energy trading (Wang, Li, Ho and Qiu, 2021), anomaly/fault detection and diagnosis (AFDD) (Himeur, Alsalemi, Bensaali, & Amira, 2020a), thermal comfort control (Gao, Li and Wen, 2020; Ngarambe, Yun,

* Corresponding author.
E-mail addresses: yassine.himeur@qu.edu.qa (Y. Himeur), me1003659@qu.edu.qa (M. Elnour), f.fadli@qu.edu.qa (F. Fadli), nader.meskin@qu.edu.qa (N. Meskin), PetriI@cardiff.ac.uk (I. Petri), RezguiY@cardiff.ac.uk (Y. Rezgui), f.bensaali@qu.edu.qa (F. Bensaali), aamira@sharjah.ac.ae (A. Amira).

https://doi.org/10.1016/j.scs.2022.104059
Received 5 February 2022; Received in revised form 11 June 2022; Accepted 10 July 2022
Available online 19 July 2022

Abbreviations

AFDD	Anomaly/fault detection and diagnosis;
AI	Artificial intelligence;
BAMS	Building energy management systems;
BDGP	Building data genome project;
BGEC	Baltimore gas and electric company;
BGRU	Bidirectional gated recurrent units;
BiLSTM	Bi-directional long-short term memory;
BPNN	Backpropagation neural network;
CAE	Convolutional autoencoder;
CAISO	California independent system operator;
CART	Classification and regression tree;
CBTL	Cross-building transfer learning;
CETL	Cross-equipment transfer learning;
CGAN	Conditional generative adversarial network;
CNN	Convolutional neural network;
CRTL	Cross-region TL;
CRL	Clustered reinforcement learning;
CSTL	Cross-site Cross-site;
CV-RMSE	Coefficient of variance of the root mean squared error;
CZTL	Cross-climate zones transfer learning;
DANN	Domain adversarial neural network;
DATCNN	Domain adversarial transfer convolutional neural networks;
DBN	Deep belief network;
DDNA	Deep domain network adaptation;
DEA	Denoising autoencoder;
DL	Deep learning;
DNN	Deep neural network;
DRL	Deep reinforcement learning;
DTL	Deep transfer learning;
EV	Electric vehicles;
GANT	Generative adversarial network transfer;
GEFC	Global energy forecasting competition;
GRU	Gated recurrent unit;
GIS	Gas-insulated switchgear;
HVAC	Heating, ventilation, and air conditioning;
HTL	Heterogeneous transfer learning;
HmTL	Homogeneous transfer learning;
IoT	Internet of things;
iTCM	Intelligent thermal comfort management;
ITCNN	Intelligent thermal comfort neural network;
ITL	Inductive transfer learning;
LSTM	Long-short term memory;
MAPE	Average absolute percent error;
MCC	Matthews correlation coefficient;
ML	Machine learning;
MMD	Maximum mean discrepancy;
MPC	Model predictive control;
MTL	Multi-task transfer learning;
NILM	Non-intrusive load monitoring;
NILMTK	Non-intrusive load monitoring toolkit (NILMTK);
NN	Neural networks;
OTL	Online transfer learning;
OfTL	Offline transfer learning;
PIR	Performance improvement ratio;
PMV	Predicted mean vote;
PRISMA	Preferred reporting items for systematic reviews and meta-analysis;
RL	Reinforcement learning;
RNN	Recurrent neural network;
REDD	Reference energy disaggregation dataset;
REFIT	Personalized retrofit decision support tools for UK homes;
S2P	Sequence-to-point;
S2S	Sequence-to-sequence;
SARSA	State–action–reward–state–action;
SEP	Sub-transmission expansion planning;
SG	Smart grid;
SMI	Similarity measurement index;
STL	Sequential transfer learning;
STVS	Short-term voltage stability;
SVM-AD	Support vector machines with adapting decision boundaries;
SVR	Support vector regression;
TCA	Transfer component analysis;
TGAN	Tabular generative adversarial networks;
TL	Transfer learning;
UAV	Unmanned aerial vehicle;
UTL	Unsupervised transfer learning.

& Santamouris, 2020), electric vehicle (Zhang et al., 2020), renewable energy and microgrids (Aslam et al., 2021; Cheng, Zhu, Gu, Yang and Mohammadi, 2021), non-intrusive load monitoring (NILM) or energy disaggregation (Himeur, Alsalemi, Bensaali, & Amira, 2020c; Himeur, Alsalemi, Bensaali and Amira, 2021), etc. Additionally, AI-based big data analytics have great potential for accelerating and supporting the energy sector's transformation from fossil-based to zero-carbon (Varlamis et al., 2022). They can serve as an intelligent layer across different energy systems to identify patterns, improve systems' performance, and predict outcomes of complex situations (Elnour et al., 2021; Himeur, Alsalemi, Al-Kababji, Bensaali and Amira, 2020). Moreover, for various energy tasks of the building energy sector, such as load forecasting and thermal comfort control, conventional physics-based models require thorough and case-specific building information (e.g., material, geometry, windows size, etc.) for creating energy efficiency models (Fadli et al., 2021; Himeur et al., 2022). Also, its development needs tremendous effort while achieving sufficient accuracy and efficiency is challenging for runtime building control, and ensuring scalability for field implementations is tough. By contrast, data-driven-based algorithms only need building operational data, e.g., thermal response, building operations, environmental patterns, to train their models (Himeur et al., 2022; Sardianos et al., 2020). Put simply, the model structures could remain invariant for distinct buildings (Chen, Tong, Zheng, Samuelson and Norford, 2020).

On the other hand, consumers can contribute to a stable and green energy grid intelligently connected to energy networks. Typically, smart building solutions and smart meters already exist, and vast amounts of data can be recorded, processed, and analyzed to polish end-users energy consumption behaviors (Wang, Xu, Wang and Li, 2019). This is possible by providing them with on-time, engaging, and explainable energy-saving recommendations (Alsalemi et al., 2020; Sardianos et al., 2021). Additionally; in smart networked buildings, for example, connected appliances can react to prices on the electricity market and adapt to the building energy consumption profiles (e.g., to boost their consumption when electricity is abundant and cheap and do the inverse is the price increases) for saving energy and reducing costs, thanks to the AI-based big data analytics (Más & Kuiken, 2020). In this context, integrating end-users information (such as energy consumption preferences, habits, time, windows, etc.) and building parameter settings into the data-driven models can help further improve energy optimization (Alsalemi et al., 2020; Sayed, Alsalemi, Himeur, Bensaali and Amira, 2021).

However, the broad utilization of AI-based big data analytics in energy systems can be prevented or delayed by various key barriers, among them (i) data scarcity, where historical data or real-time records may not be promptly available due to shortages of grid communication infrastructures (Li, Wu, Attar and Xu, 2022; Yang, Hou, Liu, Zhai and Niu, 2021), (ii) lack of labeled datasets for training machine learning (ML) models (Himeur, Alsalemi, Bensaali, & Amira, 2020b; Tariq et al., 2021), (iii) high computing resource requirements, especially when deep learning (DL) models are trained on a massive range of energy data (He et al., 2020; Liu, Yu, Liang, Griffith and Golmie, 2021); (iv) supervised learning can create highly accurate models by training ML models for completing a wide range of tasks using annotated datasets, however, its application on real-world scenario may encounter some issues if actual data deviates or strays from the training sets. To that end, reducing the volume of training datasets, creating labeled datasets, and decreasing the training time while maintaining adequate learning performance are challenging and crucial issues (Lv, Lou, Kumar Singh, & Wang, 2021).

To alleviate the impact of the above-mentioned barriers, transfer learning (TL) has been recently introduced as a solution that can bring numerous advantages to the development process of energy systems based on AI and big data analytics (Feng et al., 2022; Li, Li, Liu, Wang and Zhang, 2021). Typically, TL helps to (i) save computing resources and improve efficiency when training new models since the ML models can be pretrained offline on large-scale datasets and then fine-tuned on small datasets (Ahmed, Jeon, Chehri, & Hassan, 2021); (ii) train ML models on available annotated datasets before validating them on unlabeled datasets, which is of utmost importance, keeping in mind that labeling data is an arduous task that takes time and effort and requires the intervention of experts (Che, Deng, Lin, Hu, & Hu, 2021; Zheng, Qi, Zhuang, & Zhang, 2021); (iii) train ML models using simulated or synthetic data instead of real-world environments (Ko & Park, 2021), (iv) leverage knowledge from existing models instead of starting from scratch each time, and (v) exploit the knowledge acquired from previous tasks for improving generalization about others (Kim, Choi, Kim, & Choi, 2021; Prakash, Murugappan, Hemalakshmi, Jayalakshmi, & Mahmud, 2021).

1.2. Transfer learning conceptual background

TL consists of training a model on a specific domain (or task) and then transferring the acquired knowledge to a new, similar domain (or task). For example, if we consider building thermal comfort prediction, the data from a specific building can be used to train an ML model and learn the optimal model parameters before retraining a part of the model (i.e., fine-tuning) and performing the validation process on new target buildings (which can be from the same or different region/country) (Yang, Cheung, Ding and Tan, 2021). In this regard, TL has mainly been introduced to (i) overcome the problems of data scarcity encountered with small datasets, which hinders training a full-scale model from scratch, and (ii) reduce computational cost. We briefly summarize the definitions widely used in TL:

Def. 1 - Domain: Let us consider a specific dataset $X = \{x_1, \ldots, x_n\} \in \chi$, in which χ represents the feature space, and $P(X)$ refers to the marginal probability distribution of X. A domain is defined as $\mathbb{D} = \{X,\ P(X)\}$. In TL, the domain that contains the initial knowledge is defined as the source domain, where it is represented by $\mathbb{D}_S$. By contrast, the domain including the unknown knowledge to be learnt is named the target domain, it corresponds to $\mathbb{D}_T$ (Lu, Tian, Zhou, & Liu, 2021).

Def. 2 - Task: Considering the previously defined dataset $X = \{x_1, \ldots, x_n\} \in \chi$, which corresponds to a set of labels $Y = \{y_1, \ldots, y_n\} \in \gamma$, where γ represents the label space. A task can be defined as $\mathbb{T} = \{Y, F(X)\}$, where F is denotes the learning objective predictive function that could be represented as well as a conditional distribution $P(Y|X)$. Following the task definition, the label spaces of the source and target domains are represented as γ_s and γ_T, respectively (Ramirez, Tonioni, Salti, & Stefano, 2019).

Def. 3 - Transfer Learning (TL): if we consider a source domain $\mathbb{D}_S$ and its corresponding task $\mathbb{T}_S$, a learned function F_S can be interpreted as the knowledge obtained in $\mathbb{D}_S$ using $\mathbb{T}_S$. When there is a difference between domains or tasks, the goal of TL is to get the target predictive function F_T for target task $\mathbb{T}_T$ with target domain $\mathbb{D}_T$. Put differently, TL aims to help improve the performance of F_T by utilizing the knowledge F_S, where $\mathbb{D}_S \neq \mathbb{D}_T$ and $\mathbb{T}_S \neq \mathbb{T}_T$. In doing so, TL can be represented as follows (Lu et al., 2015):

$$\mathbb{D}_S = \{X_S, P(X_S)\},\ \mathbb{T}_S = \{Y_S, P(Y_S/X_S)\} \rightarrow \mathbb{D}_T = \{X_T, P(X_T)\},$$
$$\mathbb{T}_T = \{Y_T, P(Y_T/X_T)\} \tag{1}$$

Def. 4 - Domain Adaptation: considering the source domain $\mathbb{D}_S$ for the task $\mathbb{T}_S$ and the target domain $\mathbb{D}_T$ for task $\mathbb{T}_T$, where $\mathbb{D}_S \neq \mathbb{D}_T$. Domain adaptation aims at learning a predictive function F_T so that the knowledge obtained from $\mathbb{D}_S$ and $\mathbb{T}_S$ can be used for enhancing F_T. In other words, the domain divergence is adapted in F_T (Li, Gu, Zhang, & Chen, 2020).

Overall, classifying data where $\mathbb{D}_S \neq \mathbb{D}_T$ or $\mathbb{T}_S \neq \mathbb{T}_T$ is the main challenge that TL algorithms attempt to meet. One popular idea to do so is by reducing the difference between domains or tasks, which ensures certain similarities between the corresponding feature or label spaces (Tuia, Persello, & Bruzzone, 2016). Fig. 1 explains the difference between conventional ML and TL techniques.

1.3. Contribution of the paper

Although the significant effort recently put by the energy research community to develop TL-based energy applications, there is only one survey study that has recently been introduced by Pinto, Wang, Roy, Hong, and Capozzoli (2022). Typically, the authors have focused on investigating the use of TL for smart buildings, with a brief description of its utilization for load forecasting, systems control, and building dynamics. Also, open questions and future directions have superficially been discussed. However, we present in this paper a comprehensive survey of the research progress of TL in the energy sector. Specifically, this study attempts to intensely discuss the latest trends in using TL to perform different big energy data analytics, such as load forecasting, thermal comfort control, energy disaggregation (or NILM), AFDD, and non-intrusion detection in smart grids, etc. Moreover, a critical discussion is conducted to identify the open challenges that need to be addressed and facilitate the broad deployment of TL-based solutions in energy systems. This includes the negative transfer, overfitting, measurement of transfer gains, unification of TL, and reproducibility of TL results. Additionally, two case studies are presented to provide the reader with insight into real-world applications. They illustrate (i) the use of TL for load forecasting in sports facilities that suffer from a lack of public datasets and (ii) energy prediction during the COVID-19 pandemic using mobility data. Moving forward, future research directions that provide insights into where the actual research effort is concentrated and the challenges that will attract considerable research and development in the near future are identified. Lastly, the essential findings following this study are derived. Overall, the principal contributions of this paper are summarized as follows:

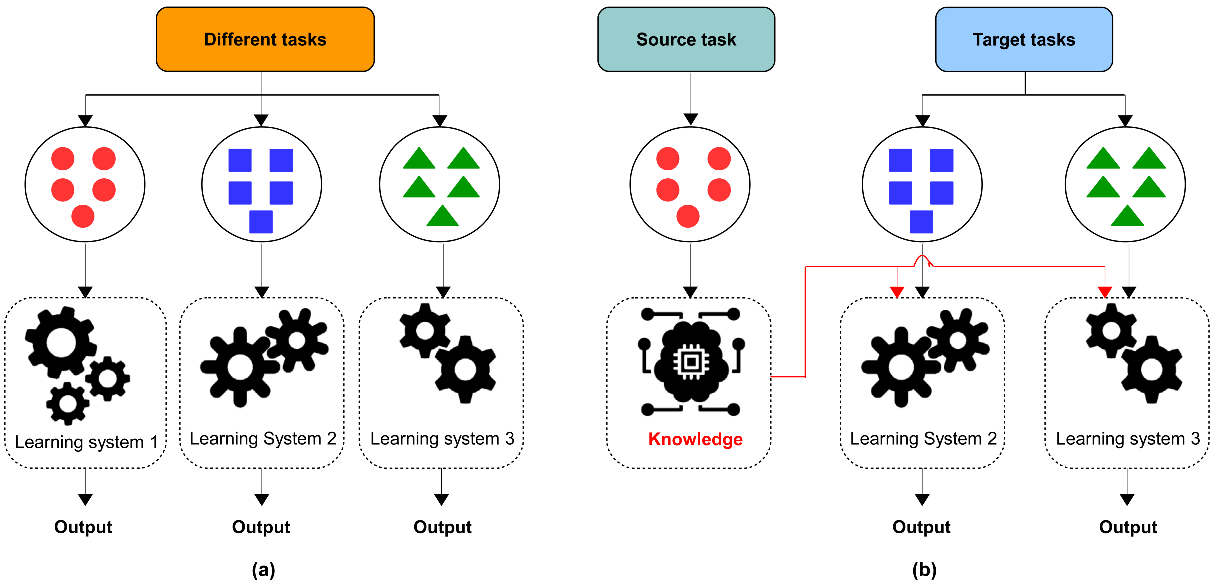


Fig. 1. Difference between conventional ML and TL techniques for multiple tasks: (a) conventional ML and (b) TL.

- A thorough overview of existing TL-based energy systems for sustainable smart city applications is introduced. It is based on a generic taxonomy for classifying them into different categories with reference to the similarity of the domain/task, learning process, feature space, and application scenario.
- Detailed analysis and in-depth discussion are performed to inform the state-of-the-art of current issues of TL-based energy systems. This includes negative transfer, overfitting, reproducibility of scientific results, knowledge gain quantification, and unification of TL.
- Introducing two case studies that explain the use of TL-based models to (i) forecast loads of sports facilities suffering from the lack of open datasets. Typically, a TL is adopted, where neural network (NN)-based model predictive control (MPC) is pretrained on a simulated dataset and then fine-tuned on a small target dataset. The latter has been collected from a sports facility at Qatar University; and (ii) forecast electrical loads using mobility data during the COVID-19 pandemic.
- Future research directions towards improving the performance of TL-based energy systems are described, which revolve around using (i) online TL, (ii) federated TL, and (iii) transfer reinforcement learning.

findings.

Fig. 2 portrays the structure of this review and summarizes the main. The rest of this paper is organized as follows. Section 2 explains the adopted research methodology. Section 3, presents an overview of TL-based energy contributions in sustainable smart cities. Moving on, TL potential and notable use cases are described in Section 4. Following, a critical discussion is conducted in Section 5 to identify key challenges. Next, two case studies are presented in Section 6 before deriving the future research directions in Section 7. Lastly, the paper is concluded in Section 8.

2. Review methodology

2.1. Objective of the study

Our review methodology is based on the approach described in Moher, Liberati, Tetzlaff, and Altman (2009). Specifically, identifying the necessity to conduct a survey is of utmost importance, and the results of review papers help clarify the state of knowledge and identify needed research. The necessity for the survey presented in this paper comes from the fact that a variety of research studies, solutions, and methodologies have been proposed recently in the energy sector to tackle different challenges. The latter includes (i) the lack of labeled datasets, data scarcity, and small real-world datasets, and (ii) the necessity to reduce the training time of DL models, etc. Moreover, our investigation reveals that no systematic survey has already been proposed to explain the use of TL in energy systems. Therefore, by conducting this survey, the following questions will be answered:

1. Why is TL attracting substantial interest in energy systems for smart and sustainable cities, and what are its challenging applications?
2. What are the methodologies used to achieve them and how can they be classified?
3. What are the principal issues encountered by the research community while developing TL-based energy systems?
4. What are the essential research directions that TL-based energy systems should follow in the near future to improve their performance and facilitate their applications in the real world and promote sustainable cities?

2.2. Literature search strategy

We conducted the bibliometric research under the perspective of a narrative review. Existing studies on using TL for energy systems have been searched. Our search instigation has been taken place on the "Scopus" and "Web of science", "ScienceDirect", "IEEE", "ACM", and "Taylor & Francis" databases from 2016 to 2021. In this way, the following terms have been searched in the titles, abstracts, and keywords: ("transfer learning" OR "domain adaptation") AND ("energy systems" OR "fault diagnosis" OR "anomaly detection" OR "load forecasting" OR "thermal comfort prediction" OR "energy disaggregation" OR "smart grid" OR "energy trading" OR "renewable energy") AND (publication date: 2016–2021).

In this respect, research studies introduced between January 2016 and February 2022 are discussed in this framework. This period has arbitrarily been selected to evaluate the recent and pertinent contributions. Typically, this framework discusses English-written peer-reviewed journal articles, conference proceedings papers, and book chapters. The selection process adopted in this review relies on adhering to the specifications of the "preferred reporting items for systematic

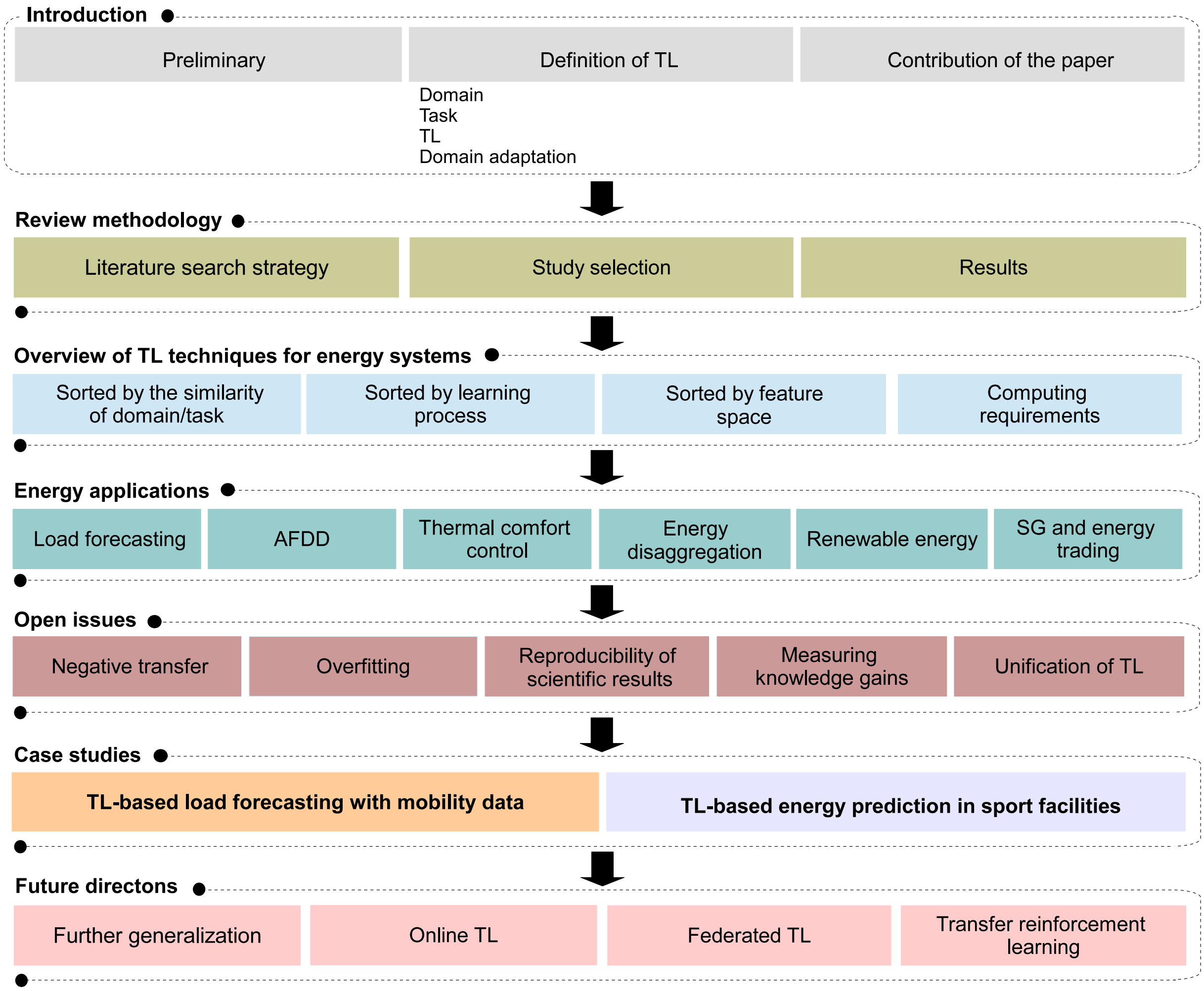


Fig. 2. Organization of the review paper and its main contributions.

reviews and meta-analysis (PRISMA)" (Moher et al., 2009), which is a practical and efficient approach for writing survey studies. Concretely, a search was performed for the last seven years (January 2016–February 2022), where 493 articles were initially identified. Moving on, 109 duplicates have been eliminated using reference manager software. The remaining frameworks have been filtered by their titles, keywords, and abstracts, where 296 references were reserved. Lastly, a full-text evaluation is performed, and in the end, 211 articles that fit the conditions are retained. In addition to reviewing existing TL-based energy systems for sustainable smart city applications, two case studies are included in this article to provide the reader with insights about using TL for (i) energy prediction with mobility data and (ii) load forecasting in sports facilities. Fig. 3 explains the main literature search steps considered in this systematic review.

3. Overview of TL-based energy systems for smart cities

The focus of this section is on reviewing TL-based energy systems introduced to improve smart energy city services and solve different research gaps. Typically, this paper sheds light on the latest trends in using TL in energy systems for smart cities. To that end, we attempted to cover most of the tasks related to the use of energy systems for sustainable and smart city applications, including building energy forecasting, AFDD, thermal comfort control, energy disaggregation, the proliferation of renewable energy technologies (e.g., EVs within smart cities), development of smart grid (SG), and energy trading. Specifically, these tasks are the most important for optimizing energy functions in smart cities and promoting sustainability and economic growth while at the same time enhancing the quality of life for citizens. Although there is no unified or standard taxonomy of TL techniques, we attempted in this section to categorize them based on different aspects, such as the similarity of domain/task, learning process, feature space, and application scenario. Those metrics have been adopted to help the readers to understand TL from different perspectives. Fig. 4 illustrates the categories of TL identified using the adopted taxonomy.

3.1. Sorted by the similarity of domain/task

3.1.1. Transductive TL

In this case, the source and target tasks are similar, but their domains are different. Moreover, annotated data is only available in the source domain, making this category quite similar to semi-supervised learning processes. Additionally, if the source feature space is similar to the target feature space ($X_S = X_T$), but the marginal probability distributions of input data are different ($P(X_S) \neq P(X_T)$), the transductive TL is related to a domain adaptation (Ribani & Marengoni, 2019). Another important example of transductive TL is the use of synthetic data in a source task, generated by energy simulation software (e.g., EnergyPlus, TRNSYS, eQUEST, etc.), for boosting the performance of a target task that has real-world data (Ko & Park, 2021; Ribani & Marengoni, 2019).

In energy systems, the domain can refer to different space features. Typically, this includes the electric equipment and devices that consume energy heating, ventilation, and air conditioning (HVAC)

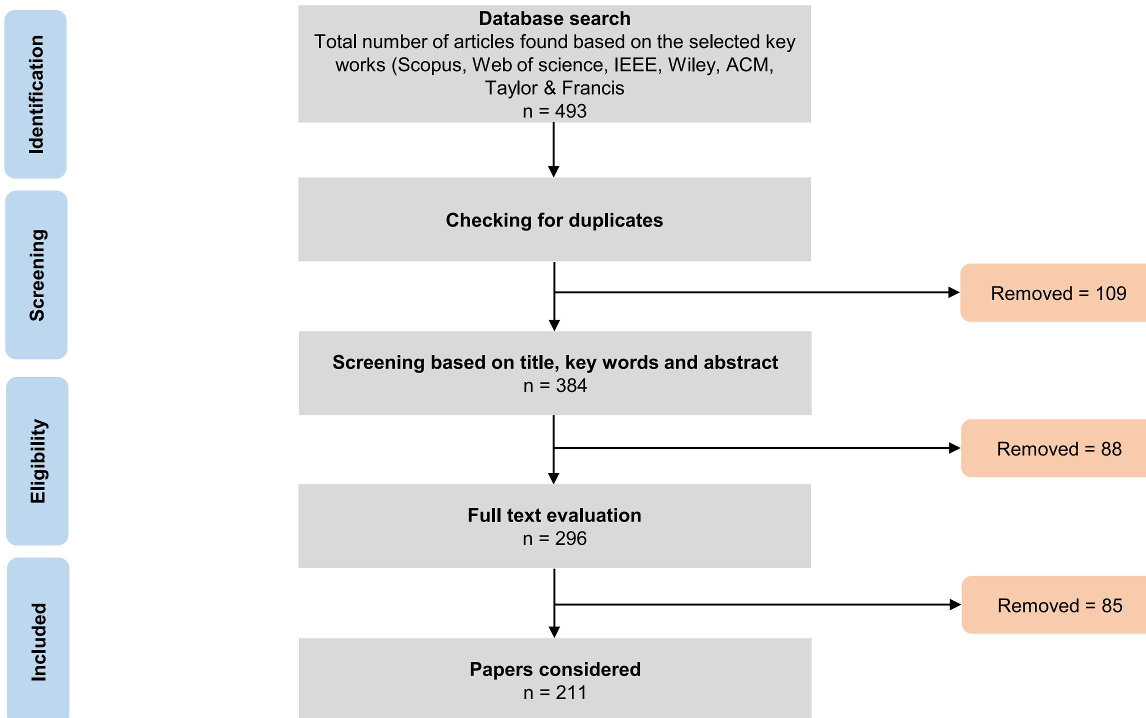


Fig. 3. The main literature search steps considered in this systematic review.

systems, chillers, domestic appliances, smart meters, etc.), buildings including electric equipment, and having different operation parameters, industrial infrastructures, etc. To that end, domain-adaptation-based TL techniques have been classified into four sub-categories:

(a) Cross-equipment TL (CETL): refers to transferring knowledge acquired with an equipment (or more) to another one that can be from the same category or a different one, but they have some similarities. For example, in energy disaggregation, an ML algorithm can be trained to segregate aggregated loads, and then the characteristics learned of an appliance can be transferred to other appliances. This is the case of D'Incecco, Squartini, and Zhong (2019), where a sequence-to-point (S2P) model has been used to transmit and adapt the implicit characteristics learned by a "complex" device (i.e., washing machine) to a "simple" device (i.e., a coffee machine). Moreover, to avoid overfitting due to data scarcity and lack of annotated datasets in AFDD systems, Liu et al. (2021) investigate a CETL using four CNN-based strategies to diagnose and detect faults in different building's chillers. In doing so, two TL tasks are conducted by splitting two chillers' data into the source and target domains, where limited data has been reserved for the latter. CETL has improved the diagnosis accuracy by 12.63% and 8.18% under two different scenarios with limited modeling data. Moving forward, a CETL-based AFDD scheme is introduced in Di, Shao, and Xiang (2021), which is based on transferring the knowledge acquired by a stacked autoencoder (SAE) of bevel-gear cross-operation conditions.

Cross-site TL (CSTL): also called cross-building TL (CBTL), aims at transferring the knowledge of ML algorithms by training and testing them on similar sites (e.g., buildings, industrial infrastructures, etc.) but with different parameters (e.g., thermal energy density profiles, the number of occupants, etc.), or entirely different sites (e.g., households vs. commercial buildings) but recorded from the same geographical region. In Park and Park (2021), individual thermal comfort prediction is conducted using a CSTL-based scheme that relies on transferring the knowledge of the combination between CNN and support vector machines (SVM) models (i.e., CNN–SVM) from different but similar indoor buildings and thermal environments. In doing so, environmental conditions and physiological data have been used to train a CNN–SVM on a data-rich building before testing it on a target subject with insufficient data. Specifically, a homogeneous ensemble TL mechanism is adopted, in which (i) 1D-CNN and DNN are first pretrained on the source building dataset, (ii) a fine-tuning is conducted based on the availability of data in the target subject, (iii) SVM and random forest (RF) are trained on randomly sampled data batches from the target building, and (iv) weighted soft voting is performed to estimate the final individual thermal comfort prediction. The experimental tests conducted on two target subjects have shown that the accuracy and F1 score varies from one subject to another. Typically, they both reached 95% for the first subject while they attained 85% and 83%, respectively, for the second subject.

Similarly, in Wang et al. (2021), a CSTL-based thermal load forecasting system is introduced, which helps in modeling a district heating station's energy consumption based on analyzing other heating stations' data. The main idea behind this work is to directly transfer some hidden layers in the pretrained model with source domain data and then fine-tune the remaining hidden layers with target domain data. Hence, only the parameters of the layers directly transferred are entirely trained with the source domain data, effectively avoiding overfitting that may occur during the transfer process. In this context, collected source data is first preprocessed before being used to train an RNN model. Next, the parameters of most of the hidden layers are frozen, while those of the rest are assigned and retrained with the target domain dataset.

(c) Cross-region TL (CRTL): in this case, data from different sites/equipment/devices located in other geographical regions are considered to train and transfer the knowledge of ML models. A typical example is proposed in Gao et al. (2021), in which a CRTL-based thermal comfort control approach is proposed, where data recorded from different cities with a similar climate zone is used for learning thermal comfort patterns. In this context, a TL-based multilayer perceptron (MLP) model has been designed and trained to predict thermal comfort accurately. Additionally, the prediction performance has been enhanced using a GAN-based resampling method, i.e., tabular generative adversarial networks (TGAN), which can handle the class

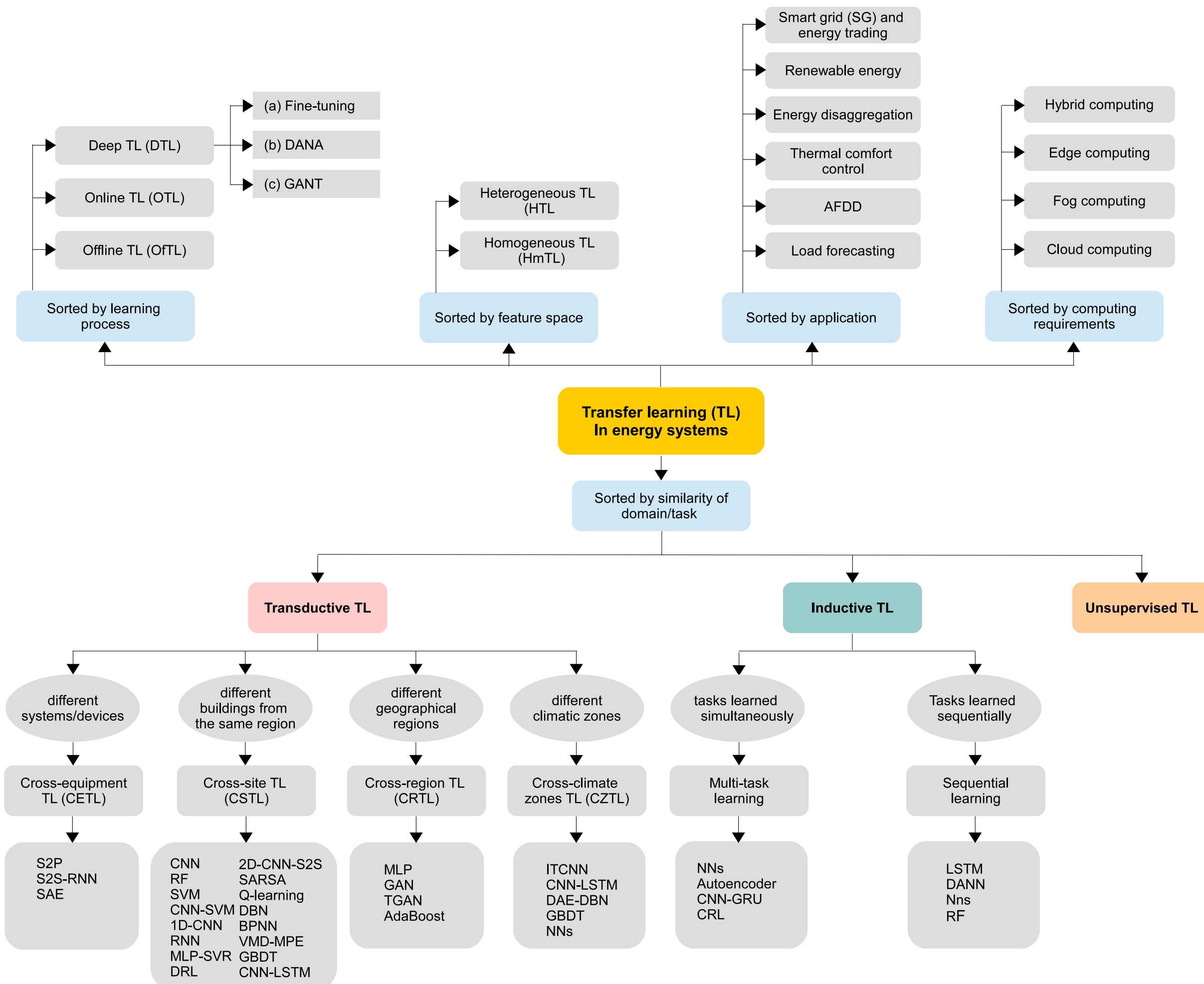


Fig. 4. The taxonomy adopted to review existing TL-based energy systems sorted with reference to different aspects, including (i) learning process, (ii) feature space, (iii) application, (iv) computing requirements, and (v) similarity of domain/task.

imbalance characteristics of occupants' thermal sensation data. Empirical evaluation has been conducted on the ASHRAE RP-884 dataset,[1] where the superiority of the TL-MLP approach versus state-of-the-art has been demonstrated in terms of the prediction accuracy and F1 score. In doing so, an accuracy of 54.50% with an F1 score of 55.12% have been achieved.

(d) Cross-climate zones TL (CZTL): it capitalizes data from different sites/equipment/devices located in different regions with distinct climate zones. Although predicted mean vote (PMV) has been investigated for decades to optimize the thermal comfort in buildings, its performance is low accurate. It has two major issues related to modeling data inadequacy and thermal comfort parameter inadequacy. To close this gap, a CZTL-based approach is introduced in Hu, Luo, Lu, and Wen (2019), which predicts the thermal comfort of target buildings from a different climate zone than the buildings used in the pre-training process. Therefore, data has been recorded for five months to build a new dataset, called intelligent thermal comfort management (iTCM), with reference to the intelligent thermal comfort management system used in the collection process. The pre-training has been conducted on the famous ASHRAE dataset, while iTCM has been used for the target domain. In this line, the pre-training has been performed using a DL model called intelligent thermal comfort neural network (ITCNN) before validating its performance on the iTCM data. The experimental results obtained in terms of the accuracy, macro-F1 score, and Matthews correlation coefficient (MCC) have revealed the superiority of the CZTL-ITCNN approach compared to the PMV model and other conventional ML algorithms, although the performance in general needs further improvement.

3.1.2. Inductive TL (ITL)

The ITL refers to the case where the target and source domains are the same (or extracted from the same site/equipment under the same conditions); however, the target and source tasks are different. In this context, ITL algorithms use the inductive biases of the source domains to help enhance the target tasks. Based on whether the source domains contain annotated datasets or not, ITL can additionally be split into two subgroups called multitask learning and Sequential (self-taught) learning respectively.

(a) Multi-task TL (MTL): involves training an ML model on different tasks jointly. Hence, it learns standard features among various tasks that share some layers. In energy systems, it has been shown that multitask TL schemes can undergo performance degradation in comparison with single-task learning (Chen, Zheng, Hu, Wang and Liu, 2019). Typically, in Chang, Chen, and Lin (2018), an autoencoder is utilized to develop MTL-based energy disaggregation. While in Xuan, Shouxiang, Qianyu, Shaomin, and Liwei (2021), an MTL-based load forecasting approach is introduced using CNN and gated recurrent unit (GRU) models for regional integrated energy systems. Moving on, a clustered

[1] https://www.sydney.edu.au/architecture/staff/homepage/richard_de_dear/ashrae_rp-884.shtml.

reinforcement learning (CRL) is adopted in Chen, Zheng et al. (2019) to implement an MTL-based solution, promoting energy efficiency in edge computing.

(b) Sequential TL (STL): in this case, the source and target tasks are not necessarily similar, and the knowledge is transferred with a sequence of steps. Additionally, STL comprises two main stages by contrast to MTL (i) the ML model is pretrained in the first stage on a source dataset (e.g., unsupervised), and (ii) it is adapted to another target task (e.g., supervised) in the second phase. A typical example is an STL-based scheme, introduced in Fang et al. (2021), to transfer the knowledge of long-term load forecasting tasks to a short-term prediction task with limited data. Typically, a long-short term memory (LSTM) has been utilized to capture the temporal characteristics across target and source–target buildings. In contrast, using domain adaptation, a domain adversarial neural network (DANN) model has been deployed to identify domain-invariant characteristics between the target and source buildings.

3.1.3. Unsupervised TL (UTL)

In unsupervised TL, both source and target domains are different ($\mathbb{D}_S \neq \mathbb{D}_T$), and their datasets are unlabeled. To transfer the knowledge and learn the target task, unsupervised TL relies on measuring the correlation between the source and target domains using unsupervised ML models. So far, there is little research work on this setting. It has been utilized in applications for fault and anomaly detection. For example, in Su, Yang, Xiang, Hu, and Xu (2022), a dilated convolution deep belief network (DCDBN) is developed to extract transferable characteristics from raw vibration data of rotary machinery's bearings under variable running conditions from the source and target domain simultaneously. Then the DCDBN model is used along with pseudo labeling technology to train an NN-based classifier for bearings fault detection. Similarly, in Michau and Fink (2021), an adversarial domain adaptation for feature extraction is combined with a one-class classifier for unsupervised anomaly detection in industrial applications. It aims to accommodate the changing operating conditions that can cause variations in the monitored data. That is, source and target data have different distributions, and UTL is applied mainly to extract transferable features between the two domains.

3.2. Sorted by learning process

Based on the learning process, TL can be categorized into offline TL (OfTL) and online TL (OTL).

3.2.1. Offline TL (OfTL)

In this case, both the source and target domains are established, where the learning process's goal is to perform a one-time knowledge transfer to adjust the model. Put simply, the source domain will not be updated with actual data during the training. For instance, various energy disaggregation frameworks have used real-world public datasets, such as the reference energy disaggregation dataset (REDD) (Kolter & Johnson, 2011), and personalized retrofit decision support tools for UK homes (REFIT) (Murray et al., 2015) to train their model in an offline manner. However, its performance may drop when applied to new unseen data because of the lack of online updates.

3.2.2. Online TL (OTL)

OTL refers to the learning process in which once the training is running; the model is updated with new data recorded from actual measurement, making the models constructed by this learning process more adaptable to the target domain (Kang, Yang, Yang, Fang, & Zhao, 2020; Zeng, Li, Jiang, & Song, 2021). Moreover, measuring the variation of data distribution is quite tricky as the data in the target domain is dynamically injected and processed (in real-time). This makes the OTL more challenging than OfTL (Bao, Mohammadpour Velni, & Shahbakhti, 2020). A typical example, in this case, is the work described in Grubinger, Chasparis, and Natschläger (2017), where an OTL for monitoring indoor environmental quality in residential buildings has been designed to be part of a model-predictive-control implementation. Specifically, weighted predictions of an offline ML model learned on a set of data-rich buildings (source domain) have been used to update an online ML model validated on new buildings with limited data (target domain).

3.2.3. Deep TL (DTL)

DL, which is a specific type of NNs that consist of an input layer, an output layer, and several hidden layers, has been widely used in energy systems for different purposes, e.g., load forecasting in buildings (Runge & Zmeureanu, 2021), renewable energy prediction (Kumari & Toshniwal, 2021; Wang, Lei, Zhang, Zhou and Peng, 2019), powerlines inspection (Jenssen, Roverso, et al., 2018), AFDD (Taheri, Ahmadi, Mohammadi-Ivatloo, & Asadi, 2021), etc. In doing so, various DL models have been used to process energy data, among them CNN, recurrent neural network (RNN), deep auto-encoder (DAE) (Fan et al., 2019). Every layer of a DL model comprises numerous nodes, which combine data inputs with several coefficients or weights that either amplify or dampen those inputs (they assign importance to input data based on the targeted learning task). The output of each node layer represents the input for the subsequent layer. To perform specific energy tasks, energy data is cleaned and preprocessed before being sent to the input layer to calculate the classification result. The output is then passed in a multistep process until satisfactory classification accuracy.

(a) Fine-tuning: the most popular DTL in energy systems relies on fine-tuning a pretrained DL model when the source and target domain are almost similar. Fig. 5(a) illustrates an example of DTL conducted using fine-tuning of a hybrid CNN-LSTM model for thermal comfort prediction in buildings. Typically, two scenarios are shown (i) when the whole network is fine-tuned, and (ii) some particular layers (learning layers) are fine-tuned. This TL architecture has been adopted in Somu, Sriram, Kowli, and Ramamritham (2021), where the parameter weights from the two source domains from the same climate zone (ASHRAE and Scales project datasets) have been utilized for improving the learning process of the hybrid CNN-LSTM architecture in the target domain (from a distinct climate zone). The first layers close to the input data have been retrained on the target domain data as they capture the pertinent features (climatic characteristics) of the dataset. In contrast, the deeper layers have been frozen since they are responsible for the classification task i.e., thermal comfort labeling.

Developing fine-tuning-based DL has recently received increasing interest in energy applications. In Chen, Tong et al. (2020), the authors use a fine-tuning-based MLP paradigm to perform an MPC of HVAC with natural ventilation in smart buildings. Typically, indoor air temperature and related humidity are predicted using the fine-tuning-based MLP model, in which the few layers have been frozen. Following, the remaining layers have been fine-tuned using a dataset of the target domain. In Hooshmand and Sharma (2019), the authors develop a TL-based energy predictive model for energy assets that have a limited amount of data. A CNN model (which is appropriate for capturing daily, weekly, and interday cyclostationary samples, seasonalities, and trends) has been designed and pretrained on public datasets (from the Open PV Project[2]) to forecast daily energy consumption before fine-tuning the weights of the last layer to train the model on a limited training dataset. Similarly in Jung, Park, Jung, and Hwang (2020), a TL-based DNN model is implemented to forecast monthly electricity consumption using data from different districts or cities (in South Korea). After training and testing the model on source data, all the layers are fine-tuned using the target domain. Also, in Jiang and Lee (2019), all the layers of the TL-based LSTM model used to predict temperature and

[2] https://energy.duke.edu/research/energy-data/resources.

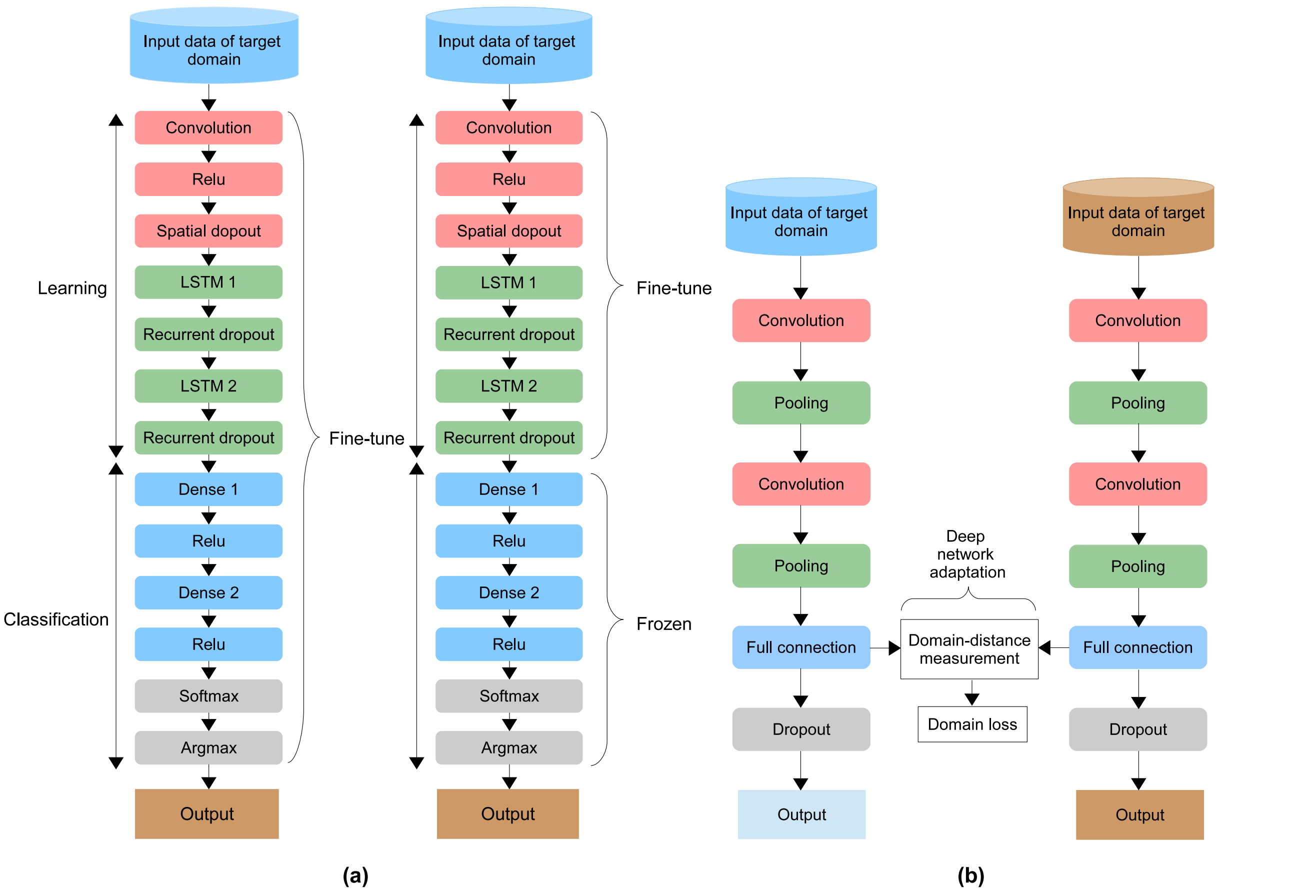


Fig. 5. An example explaining the principle of a deep TL-based on a hybrid CNN-LSTM model: (a) Structure of fine-tuning, and (b) Structure of deep domain network adaptation (DDNA).

energy demands in buildings have been fine-tuned (this is instead of freezing the parameters of the encoder/decoder and only fine-tuning the parameters of the dense layer). Moving on, a TL-based ANN model to predict one-hour-ahead energy demand is proposed in Li, Xiao, Fan and Hu (2021). Typically, the errors obtained by the first layers of the target model are back-propagated into the base (copied) features for fine-tuning them to fit the target task.

(b) Deep domain network adaptation (DDNA): Fine-tuning has been widely used because of its simplicity; also, it is easy to implement and be understood. However, fine-tuning becomes less effective when the distributions of source and domains are distinct. Seeking for other alternatives, many studies have investigated the consideration of distance measurements in TL into the original networks. This is named networks adaptation, which is based on adjusting the cost function of the initial network through the addition of a domain loss that measures the distribution of the source and target datasets (Wang et al., 2021). Fig. 5(b) portrays an example of a DDNA framework built on CNN to adjust the distribution in fully connected layers through domain distance measurement.

A typical example of DDNA-based frameworks is proposed in Fang et al. (2021), which is based on using and LSTM-based deep adversarial neural networks (LSTM-DANN) to extract domain invariant features from the source and target domains. Next, the distance between extracted source and target domains' components is measured using maximum mean discrepancy (MMD). Lastly, MMD results have been considered as the similarity metric indices for calculating regression weights and prediction values of the LSTM-DANN model. Moving on, a DDNA-based approach that helps detect intrusions against SG attacks is proposed in Zhang and Yan (2019). In doing so, domain-adversarial training is first introduced for creating a mapping between the labeled source data and the unlabeled target data. Specifically, the dissimilarity of the features mapped from the source and target domains is minimized to help classifiers learn better when applied to a new feature space against unknown threats. In Lin, Ma, Zhu, and Liang (2021) a DDNA-based energy disaggregation method is designed using temporal CNNs, which helps learn the dynamic characteristics of individual appliance loads. Distances between the source and target domain distributions have been measured to quantify domain adaptation losses. The latter has been combined to identify the NN's weights using backpropagation for training.

(c) Generative adversarial network transfer (GANT): GANs consisting of generative networks (generators) and discriminative networks (discriminators) excel in generating new datasets which have similar statistics as the training sets (Goodfellow et al., 2020). The synthetics patterns are generated by generative networks with reference to the input datasets and then estimated by the discriminative networks to distinguish them from the source. Although GANs have initially been proposed to perform unsupervised learning, they have also been effective in supervised, semi-supervised, and reinforcement learning processes. Moreover, the principles of GANs have successfully been applied to TL to develop the GANT models, where the characteristics of the source and target domains learned by the generators are transmitted to the discriminators. The latter identifies the source of the features and then sends back the results to the generators. This operation is repeated until the features become indistinguishable.

In Ahmed, Zhang, and Eliassen (2020), GANs are used to perform a GANT-based energy disaggregation. In doing so, the generator receives the aggregate energy measurements and generates appliance-level data, while the discriminator discriminates between real and synthetic data. Moving on, the knowledge learned from the source

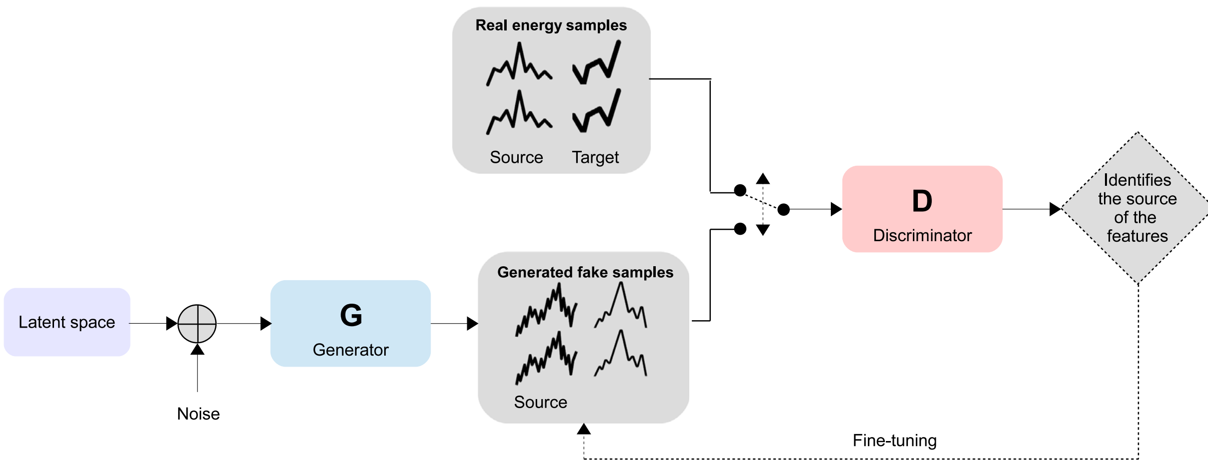


Fig. 6. A general structure of a GANT model used in Ahmed et al. (2020) for NILM. The discriminator **D** implicitly learns the loss function to detect the source of the appliance-level signature.

domain is transferred to the target domain based on addressing the model generalizability by (i) using a TL of parameter sharing and (ii) minimizing the distance between the features extracted from the source and target domains. Besides, in Han et al. (2021), Han et al. introduce GANT-based AFDD for actual power lines in SG. In this regard, a conditional GAN (CGAN) has been deployed to augment fault observations and hence increase the data amount. Then, the loss function of CNN is redesigned based on TL to develop an improved CNN-based fault classification framework. The improved CNN has been then trained using both simulated and adversarial data to make the distribution of both categories of feature patterns closer. This has resulted in better fault classification performance. Fig. 6 illustrates a general structure of a GANT model.

3.3. Sorted by feature space

Based on the nature of the features in the source and target domains, we can also be classified TL into homogeneous TL (HmTL) and heterogeneous TL (HTL). In HmTL, the dimensions and semantics of the feature sets in both the source and target domains are similar. By contrast, in HTL, the dimensions and semantics of the feature spaces are different in the source and target domains. In Hu et al. (2019), an HTL-based thermal comfort monitoring system is developed. It aims at utilizing source domain datasets for developing a related pretrained classification model based on a DNN architecture. Following, the resulting model is used to update the features of the original target domain, which are mapped to a higher dimension space. Additionally, in Gao, Shao et al. (2021) and Somu et al. (2021), HTL is used to develop an accurate NN-based forecasting model for thermal comfort for a building in a particular city with limited data, utilizing data from multiple cities in the same climate. The data of the two domains have different features with some common ones.

The deployment of HTL is proposed in Pardamean, Muljo, Cenggoro, Chandra, and Rahutomo (2019) to address the availability limitations of datasets for computer vision-based crowd-counting applications in terms of size and abundance when compared to image classification and object detection problems. It presents the development of a smart human-counting system using ImageNet dataset as the source, and a CCTV collected data as the target. The proposed system can be used for energy optimization in buildings. In Zhang, Bales and Fleyeh (2021), DL and HTL are used for night setback identification of district heating substations in which the time-series data are configured into images and used on pretrained DL models.

A HmTL-based control framework for HVAC systems is developed in Lissa, Schukat, and Barrett (2020) using RL. Lissa et al. investigated the learning transfer of the strategy of an HVAC controller from one room to another in the same building to speed up the learning time of the RL agent. While in Ahn and Kim (2022), HmTL is employed to develop a prediction model of the power consumption of office buildings towards improving efficiency. The simulation dataset of the reference building is used as the source dataset, while the target domain data are the short-term collected data. Similarly, in Demianenko and De Gaetani (2021), an ANN and HmTL are used to support automated building energy analyses in the context of building information modeling. The proposed approach helps in reducing simulation period while considering various design parameters and assessing their impact on the building energy use.

HmTL is applied extensively to tackle the big data requirement for DL models, such as in Chen, Tong et al. (2020), in which it is used to develop a deep neural network (DNN) of a building with limited available data for model predictive control of HVAC and natural ventilation. The DNN is first trained using the abundant data of another building before being retrained for the desired building. In Gao, Ruan, Fang and Yin (2020), HmTL is applied for energy consumption forecasting for a building with poor information data using a convolutional neural network (CNN) and an LSTM network, in Ma et al. (2020), for the reconstruction and imputation of missing building data before the application of DL techniques, and in Moon, Kim, Kang, and Hwang (2020) for an accurate and reliable building consumption short-term load forecasting modeling.

3.4. Sorted by computing strategy

Despite that the breakthroughs in DL over the last decade have revolutionized the energy sector (along with other research and development fields), the state-of-the-art DL models (such as deep convolutional neural networks (DCNN), AlexNet, GRU, DAE, LSTM) have a plethora of parameters, and they require to be pretrained on large-scale and comprehensive energy datasets. However, this task is not always doable, especially when dealing with a real-time application or using a constrained computing platform (Chen & Ran, 2019; Mahmoudi, Belarbi, Mahmoudi, Belalem, & Manneback, 2020). To that end, the idea of retraining existing DL models on new specific tasks or datasets has gained growing attention. To assist TL in retraining DL models, four computing methodologies have been investigated, including cloud computing (Alsalemi, Al-Kababji, Himeur, Bensaali and Amira, 2020), fog computing (Chen, Zheng et al., 2019), edge computing (Tan et al., 2018), and hybrid computing (Al Maruf, Singh, Azim, & Auluck, 2021).

3.4.1. Cloud computing

Cloud computing has first been adopted to ensure the computation requirements of TL-based energy systems due to its (i) storage capacities that allow unlimited and scalable storage space and help in integrating, aggregating, and sharing a massive amount of data and (ii) processing capabilities, where unlimited virtual processing on-demand is provided by cloud data centers (Khan, Tian, Ilager and Buyya, 2021). For instance, in Brik, Esseghir, Merghem-Boulahia, and Snoussi (2021), a cloud platform has been used for developing an IoT-based thermal comfort prediction of people with physical disabilities. Accordingly, the cloud has provided the end-users with various network services, such as statistical analytics, real-time monitoring, and graphical visualization.

3.4.2. Fog computing

Fog computing processes in the middle layer between edge devices and the cloud server for different purposes, including data filtering. With the widespread deployment of smart meters and connected devices, large-scale energy datasets are recorded (Farooq, Shabir, Javed, & Imran, 2021). However, data processing and analyzing in cloud centers have become impractical in many situations, especially for decentralized and data-driven energy applications, because of its high latency, downtime, and privacy preservation concerns. To alleviate these issues, fog computing can help build distributed, latency-aware, and privacy-preserving energy applications (Al Maruf et al., 2021). Moreover, it is a promising option for time-sensitive TL-based energy applications, such as energy disaggregation and AFDD, since it provides lower communication delays and improved network durability versus the cloud. Additionally, it can offer practical privacy preservation as data is processed by various nodes in a complex distributed system (Dey, Mukherjee, Pal, & Balamuralidhar, 2018), as it is demonstrated in Cao, Liu, Wu, Guan, and Du (2019) and Kelati, Dhaou, Kondoro, Rwegasira, and Tenhunen (2019).

3.4.3. Edge computing

The emergence of edge computing is due to its distributed computing architecture that adopts decentralized processing power, and its advantage of processing data (energy, environmental conditions, building parameter settings, etc.) and sensitive information closer to its source on IoT devices. The latter plays a significant role in promoting energy-saving and sustainability since they help make the building environments and energy-related infrastructures sophisticated and automated using ML- and DL-based solutions. Put simply, the intelligence and computing tasks are moved to the edge devices to reduce the latency (Pérez, Arroba, & Moya, 2021). Moreover, consumers are increasingly reluctant to see their personal information being transferred to cloud data centers. To that end, processing data on edge is much safer. Additionally, the commercialization of an energy-related solution (e.g., a NILM solution) on a substantial subscription basis, which requires important resources to be allocated to the cloud (e.g., in terms of maintenance), is still a challenge. Here again, going for an edge-based alternative can significantly facilitate the commercialization process (Ahmed & Bons, 2020).

However, as edge devices are equipped with low computing resources, running DL and other complex algorithms on them is challenging. To that end, TL has recently been proposed as an effective alternative, which can significantly reduce the computing requirements (Sufian, Ghosh, Sadiq, & Smarandache, 2020). In this regard, effective light DTL architectures for energy systems can be developed and implemented on edge computing environments as it has been done in other research fields (Gong, Lin, Gong, & Lu, 2020; Sufian, You, & Dong, 2021; Zhou et al., 2020). Moving forward, even though the number of works targeting the implementation of TL on edge devices for energy systems is still limited, this research direction has promising perspectives for many reasons. For instance, actual IoT devices are increasing in sophistication in addition to their vast proliferation, and the computing requirements of ML/DL algorithms request that future IoT devices should be comprised of more than simple sensors (having 8-bit microprocessors) (Pokhrel, Pan, Kumar, Doss, & Vu, 2021; Zhang, Si, Wang, Cao and Zhang, 2021).

3.4.4. Hybrid computing

Hybrid computing refers to using hybrid edge-cloud or fog-cloud computing platforms, where the information processing operations are distributed among different architectures (Leroy & Goedemé, 2021). In this respect, low-level data preprocessing operations (e.g., data filtering, data cleaning, etc.) are usually conducted either at the edge devices or fog nodes, while high-level data processing (e.g., data augmentation, classification, prediction, etc.) are performed at the cloud (Lee, Kim, & Youn, 2020; Tao, Qiu, & Lai, 2021).

4. TL potential and notable use cases in sustainable smart cities

As most of the development levels and datasets of smart energy cities are imbalanced, different problems are faced by the data scientists, including the cold-start issue, data scarcity, and lack of annotated datasets to train supervised ML models. To overcome these pitfalls, TL is leveraged to accelerate the development of smart cities. In this context, many studies have been investigated. For instance, in Dridi, Amayri, and Bouguila (2022) and Lu, Zhou, Liu, and Zhang (2022), the term urban TL is coined to explore the paradigms providing relevant practitioners and city planners with recommendations about using the TL technology. In Dridi et al. (2022), TL is used to estimate occupancy detection in smart buildings for better energy and safety management. However, the smart energy city that aims at optimizing urban energy systems and improving the quality of life for citizens is a bit vague. It is also realized through various tasks, including load forecasting, AFDD, thermal comfort control, energy disaggregation, renewable energy integration, SG development, energy trading, etc. (Fan et al., 2022; Lu et al., 2022). This study investigates the applications of TL for smart energy cities since it is a cross-disciplinary research topic that helps overcome some of the data-based challenges encountered with ML tools (Hu et al., 2022). Typically, this review focuses on analyzing the algorithmic contributions, which are the most critical part of building TL-based smart energy city systems. Moreover, this is also because TL is a series of ML strategies transferring knowledge from a source domain (with rich data) to a target domain (with scarce data) (Anjomshoaa & Curry, 2022; Maghdid et al., 2022).

4.1. Load forecasting

With the broad deployment of smart meters, the popularization of sensor-based data, data-driven energy prediction is booming. It analyzes historical sensor records to predict future energy usage using ML algorithms, particularly DL models (Jin, Acquah, Seo, & Han, 2022; Seyedzadeh, Rahimian, Rastogi, & Glesk, 2019). However, their main issue is the necessity of vast amounts of historical data (often over extended periods) for training DL algorithms and achieving robust predictions (Qian, Gao, Yang, Yu, et al., 2020; Yang, Peng, Ye, Lu and Zhong, 2021). That is not always possible, especially with the new buildings or buildings having newly installed smart meters that provide only small quantities of historical data, which cannot be sufficient for creating accurate predictions. Typically, TL approaches have been introduced as a practical solution to overcome the data scarcity issues via using cross-domain datasets for improving predictions (Khan et al., 2021; Wu, Wang, Precup, & Boulet, 2019; Ye & Dai, 2018).

In Mocanu, Nguyen, Kling, and Gibescu (2016), a CSTL method is proposed to perform unsupervised load forecasting of buildings with new behaviors and completely different buildings using the knowledge of existing buildings. Accordingly, the idea relies on learning a building model by integrating a generalization of the state space domain. In doing that, the RL state–action–reward–state–action (SARSA) and Q-learning algorithms are explored for modeling building power consumption. Then, a deep belief network (DBN) is integrated into the RL models, making them appropriate for continuous states. This approach has been validated on a comprehensive dataset (provided by Baltimore gas and electric company (BGEC)[3]), including various commercial and residential buildings, five buildings profiles, and gathered for seven years.

In Tian, Sehovac, and Grolinger (2019), a pretrained RNN model is used to develop a CETL approach for predicting the energy consumption recorded by a target meter based on knowledge acquired from different kinds of meters. Typically, a similarity-based chained TL has been utilized to train a sequence-to-sequence RNN (S2S-RNN) model and learn the similar energy consumption features of more than 400 meters in a chain manner. This has helped significantly reducing the computational cost when the CMTL has been validated on the global energy forecasting competition (GEFC) dataset. While in Hooshmand and Sharma (2019), the authors investigate the application of CSTL for short-term electricity load forecasting (24-h ahead) with limited data. In doing so, a CNN-based energy predictive model is first developed to capture inter-day, daily, and weekly energy consumption footprints, and trends and seasonalities of energy time-series. The TL is based on a pre-training process, which is introduced to learn the relevant characteristics common across the buildings having the same type as the target building. The acquired knowledge has been then adapted to the target prediction model and tested on 23 randomly selected commercial and industrial facilities (e.g., school, hospital, bank, grocery, light industry, etc.) with limited data.

In Ribeiro, Grolinger, ElYamany, Higashino, and Capretz (2018), a CBTL approach is developed to predict building power consumption using time-series multi-feature regression, and trend and seasonal regulation. Mixed data from buildings with different characteristics (i.e., distributions and seasonal profiles) gleaned over a long time frame have been used to predict the energy consumption of another building having one month of data. MLP and support vector regression (SVR) models have been adopted to adjust the source buildings' data. The effects of time over time-series adaptation have been removed, and time-independent features have been prepared via a non-temporal domain adaptation. Besides, in Gao, Ruan et al. (2020), Gao et al. introduce two DL models, a 2D-CNN and a sequence-to-sequence (S2S) model with an attention layer, (2D-CNN-S2S) to implement a CSTL scheme, which enhances the forecasting accuracy of a target building. Typically, data from three office buildings in China has been considered to forecast the energy consumption of the target building. In this context, compared to LSTM, S2S and 2D-CNN have enhanced (i) average absolute percent error (MAPE) by 19.69% and 20.54% on average, respectively; and (ii) coefficient of variance of the root mean squared error (CV-RMSE) by 31.18% and 30.32% on average, respectively. Accordingly, to predict one year of energy consumption, they have employed one month of data for fine-tuning the model pretrained on two years of data (2015–2016) from every building in the source domain. Similarly, a CSTL is investigated for short-term load forecasting, where 24-h ahead building energy demand is predicted in Fan et al. (2020). Specifically, the knowledge learned from buildings with high-quality data has been utilized to ease load forecasting in buildings with limited data. In doing so, data of 407 buildings (from the building data genome project (BDGP) Miller & Meggers, 2017) have been used as the source domain, while the data of the remaining 100 buildings have been utilized as the target domain. The model architecture encompasses three parts (i) 1D-CNN layers that have been deployed for automatically extracting local temporal characteristics from time-series, (ii) recurrent layers that have been used for capturing interactions of intra-temporal characteristics, leading to better accuracy; and (iii) bidirectional operations and dropout techniques, which have been integrated to improve the prediction accuracy and avoid overfitting. In this regard, by adopting this CSTL approach, the prediction error has been reduced from 15% to 78%. Additionally, the TL value has been quantified using the performance improvement ratio (PIR) and statistical tests. Accordingly, a mean PIR of 77.9% has been reached for the feature extraction, while a mean PIR of 0.752 has been attained for weight initialization.

In Li, Fu, Fung, Qu and Lau (2021), a CSTL scheme is proposed to predict short-term energy consumption (for one-hour ahead) of buildings with minimal data using a backpropagation neural network (BPNN). The knowledge of BPNN acquired from 404 different source buildings of the BDGP dataset has been transferred for the model design for target buildings. This study has demonstrated that the critical features influencing the performance of the CSTL scheme are the energy consumption behavior and building industry. On the other hand, most of the TL-based energy prediction schemes ignore discussing the process of selecting the source domain data. Therefore, the prediction accuracy depends mainly on the similarity between the source and target domains. To overcome this issue, Lu et al. (2021) propose a general STL-based load prediction approach that integrates a similarity measurement index (SMI) to select the source tasks' data having the most similar percentile with the target task data. An LSTM model has been pretrained on data-rich tasks before being tuned on the target data-scarce task. In this line, by integrating the SMI to the STL, prediction error has been reduced by 1.81–5.65% compared to the STL without SMI. Fig. 7 portrays the example of the STL based load forecasting framework proposed in Lu et al. (2021).

TL is applied in Le et al. (2020) to develop a framework for multiple electric energy consumption forecasting in a smart building utilizing LSTM networks while minimizing computation time. The consumption data is pre-processed to produce clean and complete daily consumption data. K-means clustering algorithm is used to identify the multiple daily load demand profiles in which Silhouette analysis is applied to determine the appropriate number of clusters. For each cluster, i.e., demand profile, an LSTM model is trained to develop the multiple profiles forecasting LSTM models using TL. This reduces the training time of the numerous computationally demanding LSTM networks. While in Jain, Gupta, Sathanur, Chandan, and Halappanavar (2021) TL is applied to tackle the effect of practical data sparsity/quality on the accuracy of building consumption forecasting models. ML-based consumption forecasting models, are first trained using the abundant simulation data and then, using TL, are retrained using the limited practical data. The use of TL in this context is investigated in two schemes: (i) feature extraction scheme in which only the output segments of the ML model are retrained, and (ii) parameters initialization scheme, in which the full ML model is retrained. Due to limited practical data availability, TL demonstrates its effectiveness in solving the load forecasting model's overfitting and poor performance issues. Similarly, in Xu and Meng (2020), a hybrid TL model based on time series decomposition is proposed in which the trend and seasonal components are handled by a standard ML model for better interpretation of the seasonal cycles of the load data. Then, a two-stage forecasting TL model is developed for the irregular component to improve forecasting accuracy. Table 1 presents a summary of the reviewed TL-based load forecasting frameworks, in terms of the adopted ML model, type of TL, dataset, limitations and best performance. The performance of reviewed methods have been compared based on the metrics used in each framework, e.g., the accuracy, F1 score, mean square error (MSE), root mean squared error (RMSE), mean absolute error (MAE), PIR, or other statistical inspired coefficients.

4.2. Anomaly/fault detection and diagnosis (AFDD)

Diagnosis of energy systems and monitoring their condition can efficiently alleviate the impact of failures. This operation can help identify the faults and anomalies due to energy components and equipment, or even those that occurred because of consumers' behaviors (Zhang & Cheng, 2021). Data-driven-based AFDD has recently received growing interest from the energy research community. It benefits from the actual and historical big data generated by the energy systems

[3] http://www.supplier.bge.com/.

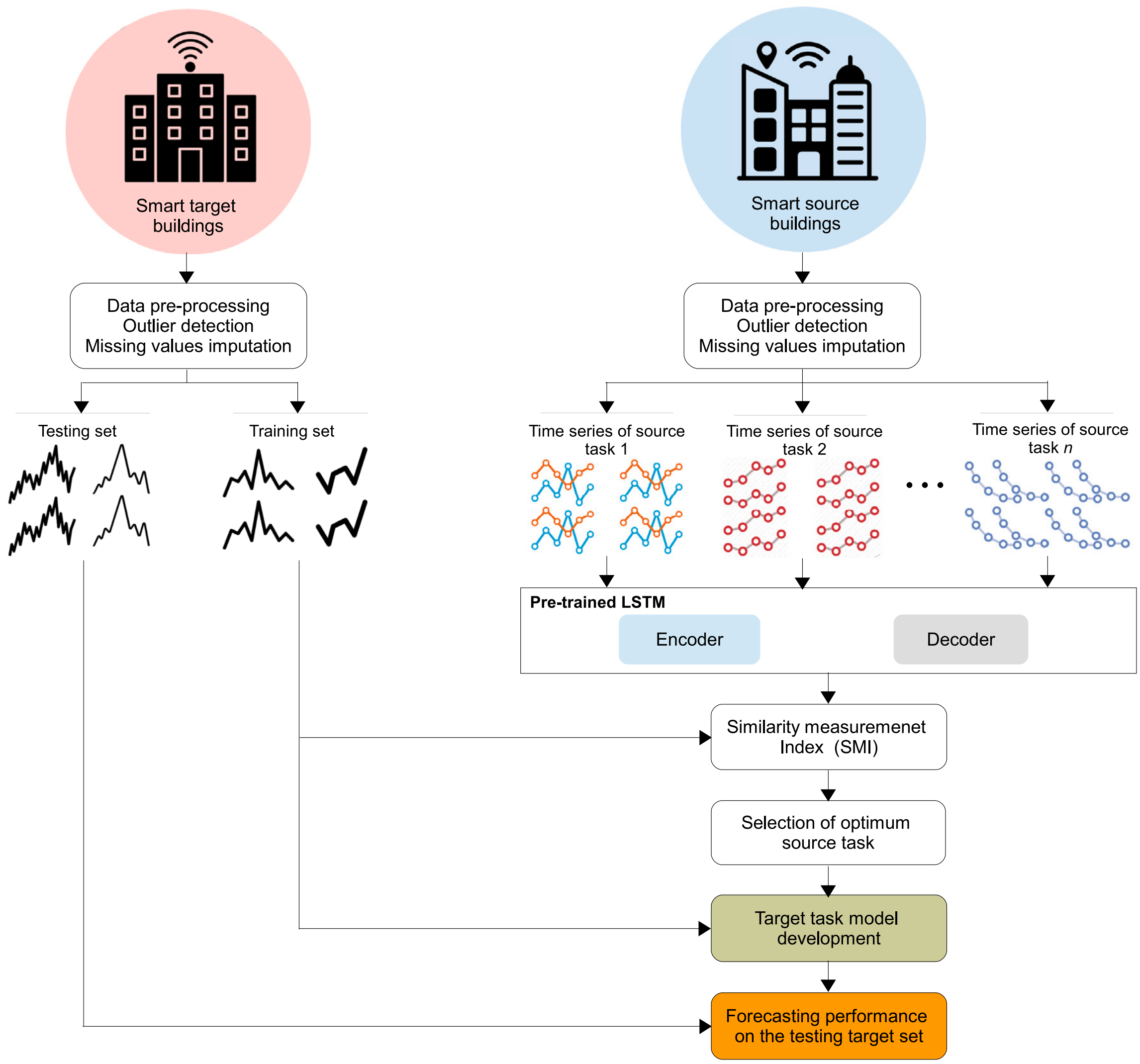


Fig. 7. Example of an STL-based load prediction framework used in Lu et al. (2021) to enhance the prediction accuracy of target task under the condition of limited data.

to identify anomalies and faults. Thanks to the advancement of ML, intelligent automatic AFDD systems have been developed at a low cost, and their reliability and accuracy are ever-increasing. However, the broad deployment of AFDD is still limited by the quality and quantity of training data and also the generalizability of the ML models (Tang, Tang, & Chen, 2019).

The building energy sector puts significant effort into developing AFDD systems to save energy for building operations as more than 20% can be wasted due to faults and anomalies during routine operations (Teke & Timur, 2014). However, it is still challenging to collect sufficient data from the faulty operation of building energy systems since not all the faults can be reproduced without causing damages to the buildings' energy systems. Additionally, the quality of data can also affect the diagnostic accuracy of AFDD models and produce high false alarm rates (Wang et al., 2022). To that end, TL has been actively used to transfer knowledge learned from data-rich building energy systems to AFDD tasks in data-sparse systems (Han et al., 2021; Wu, Jiang, Zhao, & Li, 2020). Moving on, since faulty HVAC chillers can lead to increased energy consumption, reduced thermal comfort, and increasing maintenance costs, developing robust AFDD strategies help significantly alleviate these effects. In this line, an AFDD system is proposed in Dowling and Zhang (2020) by conducting a CSTL using a Bayesian classifier that helps HVAC faulty operations from normal operations. Accordingly, the Bayesian classifier has been trained on a data-rich building before transferring its knowledge to another data-scarce building.

While most existing AFDD frameworks have focused on training fault diagnostic ML models specific chillers, very few studies have been reported to overcome the problem of transferring AFDD among highly different datasets. This is a promising but challenging task, which was addressed in Zhu, Chen, Anduv, Jin, and Du (2021). Typically, a generic model that can learn the fault diagnosis knowledge on information-rich source chillers and then transfer it to new target chillers is proposed. In this respect, after standardizing heterogeneous patterns of different chillers, a CETL based on domain adaptation is implemented to avoid domain shifts that emerge from the difference of feature distributions across chillers. Therefore, a DDNA is developed for generating the diagnostic model for the target chillers, where the information about normal operation profiles of the target chillers and the prior knowledge are considered to train the NL models. For the empirical evaluation, two screw chillers have been utilized as the target and source domains accordingly, where extensive fault simulations and experiments have been conducted. Results have indicated that the transferred diagnostic knowledge has enabled decent diagnostic performance on a small dataset. For instance, the accuracy for the first and second cycles are 81.27% and 74.93%, respectively. The block diagram of the CETL-based AFDD system proposed in Zhu et al. (2021) is portrayed in Fig. 8.

Table 1
Summary of the reviewed TL-based load forecasting frameworks.

Work	Year	Model	Type of TL	Dataset	Best performance	Limitations
Ribeiro et al. (2018)	2017	MLP, SVR	CSTL	Private data	MAPE = 0.0928, MSE = 66,509	• Source and target buildings were from the same region. No information about the performance of the model on buildings from different regions.
Mocanu et al. (2016)	2019	SARSA, Q-learning, DBN	CSTL	Private data (BGEC)	RMSE = 0.02, R = 0.99	• Performance varies from a scenario to another.
Tian et al. (2019)	2019	S2S-RNN	CMTL	CEFG	MAPE < 10%	• Validated on meters' records with similar duration and reading intervals, no idea about its performance on datasets with different duration and reading intervals.
Hooshmand and Sharma (2019)	2019	CNN	CSTL	UCI repository	Total normalized averaged MAE (TNA.MAE) = 0.7	• The performance varies significantly from a building to another.
Gao, Ruan et al. (2020)	2020	S2S, 2D-CNN	CSTL	Private data	MAPE = 0.0860, CV-RMSE = 0.1482	• Validation on three buildings of the same functional type and climate. The case of distinct kinds of source, target buildings, and/or different climate zones was not investigated.
Le et al. (2020)	2020	LSTM	STL	Private data	RMSE < 1.4, MAE < 0.9	• High computational cost for determining the optimized hyper-parameters.
Xu and Meng (2020)	2020	AdaBoost regression	CRTL	GEFCOM2012, Australia electricity dataset	MAPE < 20	• Moderate performance and privacy issues.
Fan et al. (2020)	2021	1D-CNNN, LSTM	CSTL	PIR (feature extraction) = 77.9, PIR (weight initialization) = 0.752	BDGP	• High computational cost especially in the training phase. Buildings in the empirical evaluation were from the same region.
Li, Fu et al. (2021)	2021	BPNN	CSTL	BDGP dataset	MAPE < 15	• The performance is moderate and needs further improvement.
Lu et al. (2021)	2021	LSTM	STL	Private data	MAPE < 11	• Have deviations in selecting the optimum source tasks the target data is very limited, i.e., less than 9% of the source tasks data.
Fang, Gong, Li, Chun, Li et al. (2021)	2021	LSTM, DANN	STL	BDGP	MAE = 5.3732, MAPE = 0.0635, CV-RMSE = 0.0811	• High computational cost for determining the optimized hyper-parameters (using trial and error).
Jain et al. (2021)	2021	NN, RF	STL	Private data	Accuracy = 80%	• Security and privacy issues.

In Wang, Yan et al. (2021), a CETL-based method is developed using domain adversarial transfer CNN (DATCNN) to diagnose gas-insulated switchgear (GIS) faults using small data samples. Typically, feature representations are first learned from the target and source domains using a residual CNN. Moving on, a domain adversarial training is performed to transfer features, in which a conditional adversarial process is used. Then, the joint distribution of labels and features is enhanced to a random linear combination that conducts a simultaneous adaptation of labels and features. Similarly, in Chen, Qiu, Feng, Li, and Kusiak (2021), Chen et al. proposed a CETL-based approach using Inception V3 and TrAdaBoost, which enables detection and diagnosis faults of wind turbines. In this context, two kinds of faults are detected using SCADA data, i.e., gear cog belt fracture and blade icing accretion. TrAdaBoost has been utilized to deal with imbalanced datasets better than conventional ML models. Besides, in Yang, Zhang, Lv and Wang (2021), an image recognition model based on DNN is introduced to accurately extract image features and automatically detect wind turbine blade damage. The Otsu threshold segmentation method has been employed to segment the blade image and eliminate the image background's impact on the detection process. To enhance this method's capability in extracting abstract features and accelerating the convergence efficiency, a TL is adopted using a CNN. Moreover, an RF-based ensemble learning classifier is used for improving the accuracy of detecting the blade defects. The performance of the overall model has been evaluated using unmanned aerial vehicle (UAV) images of the wind turbine blades. Additionally, although AFDD systems are critical for the operation of wind turbines since they reduce the impact of failures, enormous wind turbines cannot implement AFDD models because of insufficient data. In this respect, because operational data of corresponding wind turbines can have similar failure characteristics, the authors in Li, Jiang, Zhang and Shu (2021) try to employ this data to develop a CETL-based AFDD approach using convolutional autoencoder (CAE). This helps in (i) alleviating the problems due to data shortage of wind turbines and (ii) transferring knowledge from similar wind turbines to target wind turbines.

While most existing chiller AFDD frameworks require annotated datasets to train their models, van de Sand, Corasaniti, and Reiff-Stephan (2021) attempted to avoid this issue by proposing a TL-based technique to allow sharing the AFDD knowledge between heterogeneous chillers. In this respect, a domain adaptation that uses an SVM with adapting decision boundaries (SVM-AD) and transfer component analysis (TCA) for diagnosing faults is adopted. Thus, annotated source domain data and unlabeled target domain data are aggregated in the training phase. Moving on, since data cannot be sufficiently and representatively recorded for short periods when newly-built yarn-spinning workshops are studied. Moreover, anomalous energy profiles in new spinning workshops may significantly differ from old ones since, most likely, the anomaly rate is much lower in the new ones. This data variation can affect the performance of TL algorithms and make them less effective by underutilizing knowledge of the data-rich workshops. To address these challenges, a cluster-based DDAN

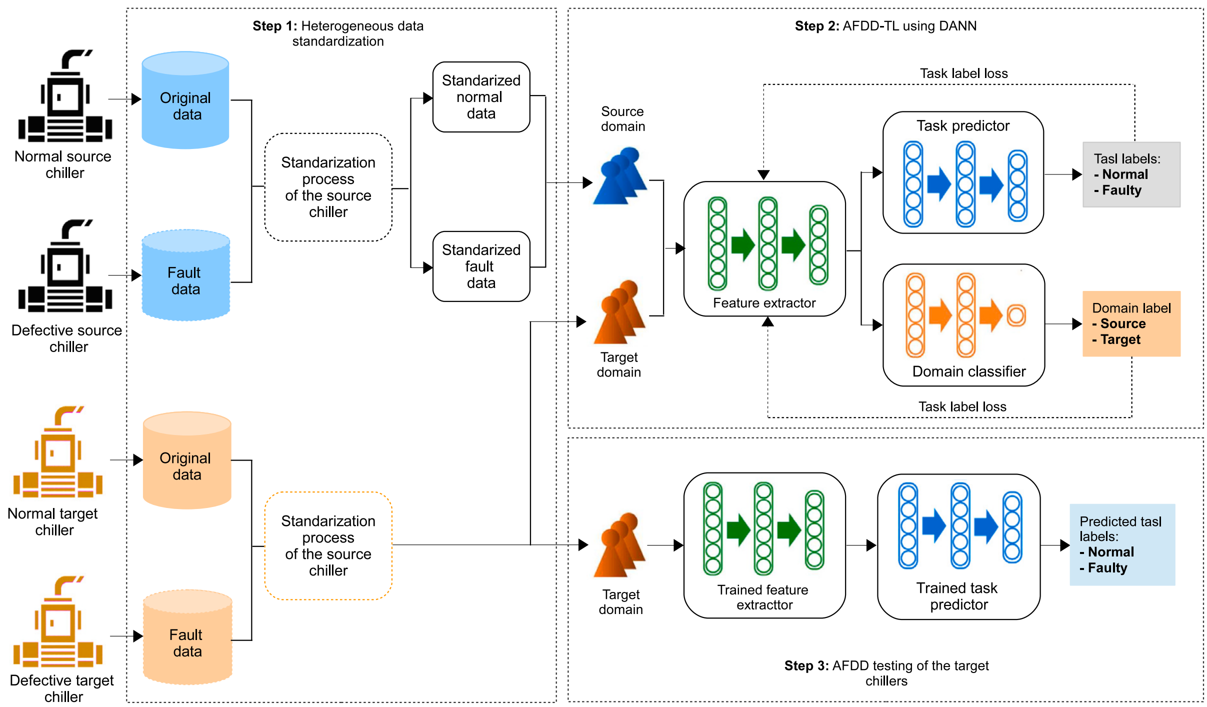


Fig. 8. The CETL-based AFDD system proposed in Zhu et al. (2021), which is applied in three steps: (i) heterogeneous data standardization, (ii) AFDD-TL using DANN and (iii) AFDD testing of the target chillers.

is introduced in Xu, Wang, Zhang, and Li (2021), which (i) moves the knowledge learned from old spinning workshops with rich energy consumption data to the target data-scarce spinning workshops, and (ii) improves the performance of TL. AFDD plays a critical role in the aluminum extrusion process since energy consumption anomalies frequently emerge because of improper extrusion parameters. To that end, a DNN-based AFDD approach is introduced in Liang, Yang, Chen, Xiao, and Lan (2018), where a large-scale datasets of extrusion machine is used to train the DNN model before transferring the acquired knowledge to data-insufficient extruding machines. Table 2 summarizes the TL-based AFDD frameworks discussed above with regard to various parameters.

4.3. Thermal comfort control

Collecting real-world datasets of indoor spaces and individual characteristics is delicate in terms of both the cost and time of the collection operation and is unrealistic in certain cases (Valinejadshoubi, Moselhi, Bagchi, & Salem, 2021). To that end, the quality of model-based control algorithms representing building systems and dynamics is essential to ensure the decent performance of smart buildings' control and automation. For instance, using DL to perform a predictive control of HVAC systems in buildings coupled with natural ventilation helps effectively predict the buildings' thermal responses considering several operational and environmental conditions. However, reliably training DL models to identify the dynamics of complicated physical processes (e.g., natural ventilation) requires the record of a vast amount of historical data. Unfortunately, this is not the case in most realistic scenarios, and sufficient building operational data may not be available for different reasons (Brik et al., 2021; Li, Fu et al., 2021).

To alleviate these issues, TL has been adopted in various studies. For instance, in Grubinger et al. (2017), a CSTL-based scheme is introduced to perform an online learning framework, which aims at enhancing temperature predictions in households using a generalized OTL (GOTL). Typically, a weighted combination of the source and target predictors has been employed, and the convergence to the best-weighted predictor has been guaranteed. Additionally, using transfer component analysis (TCA) has enabled the combination of multiple source domains (residential buildings), i.e., transferring knowledge from different source domains. Moving forward, in Mosaico, Saviozzi, Silvestro, Bagnasco, and Vinci (2019), a simplified state-space CSTL-based energy optimization scheme is introduced that relies on (i) estimating occupancy of HVAC systems using images collected from thermal cameras, and (ii) transferring the knowledge of the AlexNet+SVR. Moreover, considering various parameters, i.e., the occupancy, external temperature, temperature setpoints, solar irradiance, wind speed, and humidity, HVAC energy consumption has accurately been predicted.

In Chen, Tong et al. (2020) and Chen, Zheng and Samuelson (2020), a deep MLP is used to perform a CSTL for MPC of natural ventilation and HVAC systems in smart homes. The knowledge of the deep MLP model has been transferred by freezing most of the deep MLP layers (with 42,902 parameters) have been pretrained on a large-scale dataset (with multi-year data) from a source building (in Beijing) before being re-trained on another small dataset (including only 15 days data) from a different building (in Shanghai) having 200 trainable parameters. The two buildings have completely different window sizes, building materials, and floor areas.

Besides, deep reinforcement learning (DRL) attracts increasing interest due to its ability to perform an accurate control without analyzing physical models at runtime. However, the long training time needed to reach the desired performance still impedes its broad deployment. To close this gap, Xu, Wang, Wang, O'Neill, and Zhu (2020) introduce a BCTL-based scheme, which transfers the knowledge of a DRL-based HVAC controller trained using data from a source building to another controller applied on a data scare building. This was possible with less effort and enhanced performance than a conventional DRL-based approach without TL. While, in Bao et al. (2020), an STL-based MPC scheme is proposed by online transferring the knowledge of an ANN-based state-space linear parameter-varying (LPV-SS) model using closed-loop data. The LPV-SS model is initially identified offline based on inputs and outputs data, and an MPC approach was developed using this model. To enhance the performance, the model is also refined using collected closed-loop batch data and an online TL.

Table 2
Summary of the reviewed TL-based AFDD techniques.

Work	Year	Model	Type of TL	Dataset	Best performance	Limitations
Liang et al. (2018)	2018	DNN	CSTL	Private data	RMSE = 0.921, MAPE = 3.41%	• Further investigations are needed to optimize DNN parameters, e.g., the number of hidden layers, hidden units, and learning rate for covering different data complexity.
Dowling and Zhang (2020)	2020	Bayesian classifier	CSTL	Private data	MSE = 0.7	• Moderate fault detection performance.
Liu, Zhang et al. (2021)	2021	CNN	CETL	CEFG	Accuracy = 94.37%	• Requires more stable TL performance and stronger learning ability.
Zhu et al. (2021)	2021	CETL	DANN	Private data	Accuracy = 81.27%	• Need to be applied for more faulty scenarios occurred in the chillers. Did not be validated for different types of chillers (e.g., to transfer the knowledge from screw chiller to centrifugal chiller).
Wang, Yan et al. (2021)	2021	DATCNN	CETL	Private data	Accuracy = 98.9%	• Only working with labeled data, although it is quite difficult to promptly identify the type of insulation defect in GIS.
Chen et al. (2021)	2021	Inception V3, TrAdaBoost	CETL	Private data	Accuracy = 86.4%	• Validation on three buildings of the same functional type and climate. The case of distinct kinds of source, target buildings, and/or different climate zones was not investigated.
Yang, Zhang et al. (2021)	2021	CNN-RF	CETL	Private data	Accuracy = 98%	• High computation complexity and no classification of the type and severity of the defect.
van de Sand et al. (2021)	2021	TCA, SVM-AD	CETL	ASHRAE-1043	F1 = 98%	• Reproducing discriminative structures of the underlying target domains' distributions for faults available at low severity is challenging.
Xu et al. (2021)	2021	AlexNet	DDAN	Private data	Accuracy = 97.50%, F1 = 94.87%	• High false alarm rate and security and privacy issues.

In Lissa et al. (2020), Lissa et al. explore the spatial change of performance accuracy of a CSTL-based thermal comfort control system using an RL algorithm for buildings from the exact geographical location. Specifically, the knowledge learned by a Q-learning model for monitoring an HVAC system has been transferred by adjusting it following some spatial changes. Using the TL has enabled reducing the learning time required for training optimal (or near-optimal control) by a factor of 6 compared to the case without the TL. Moreover, when the spatial variation was less than 50%, similar performance for both static and dynamic HVAC control was reached, which presents an average time-out comfort (ToC) error of 3.83% and 2.55%, respectively (Mohammadi, Al-Fuqaha, Guizani, & Oh, 2017). Besides, in Somu et al. (2021), a CZTL-based thermal comfort prediction in buildings is proposed to overcome the data inadequacy issue. Typically, source buildings from a similar climate zone have been used to train a CNN-LSTM model and then transfer the acquired knowledge to a target building from a different climate zone. Accordingly, the hybrid CNN-LSTM model capitalizes the spatio-temporal characteristics of the thermal comfort patterns (TCPs) for efficiently modeling thermal comfort. Only the first layers of the hybrid model are re-trained using data of the target building. The experimental validation has been conducted using two source datasets (i.e., Scales Project[4] and ASHRAE RP-884) and one target dataset (from a US office). The synthetic minority oversampling technique (SMOTE) has been utilized to overcome the scarcity of samples across all thermal conditions in the considered datasets. Fig. 9 illustrates the typical flowchart of the CZTL-based thermal comfort prediction.

In Natarajan and Laftchiev (2019), a cross-users TL (CUTL)-based approach is developed for thermal comfort prediction in the same building. Typically, data from the target users varies from that of the source users concerning the thermal parameters used in each office space, such as a thermostat, space heater, humidity, and environmental stability, where an HVAC system has been used to control these parameters. Physiological data has also been recorded, including the metabolic rate, calories consumed, skin temperature, heart rate, barometer, altimeter, steps taken, and elevation. This data is then annotated by the users by providing thermal comfort ratings. Moving on, the knowledge of a ridge regression has been acquired and transferred to measure the performance within-users and between users and its within-user has been measured in both scenarios.

In Jiang and Lee (2019), a DTL-based thermal dynamics modeling approach is proposed to predict the indoor temperature and energy demand. In doing so, an LSTM-based S2S architecture has been adopted to perform a deep domain adaptation (DDA). It has been trained on a large-scale temperature and energy datasets temperature and then fine-tuned on other small datasets collected from other buildings. Specifically, SML (Zamora-Martínez, Romeu, Botella-Rocamora, & Pardo, 2014), and AHU[5] datasets have been used to model temperature variation, while two other energy data have been recorded from two commercial buildings. Table 3 outlines the pertinent TL-based thermal comfort control frameworks.

4.4. Energy disaggregation

Energy disaggregation, also called NILM, refers to the process of inferring individual appliance energy consumption fingerprints from an aggregated waveform (usually recorded by existing meters) using ML modes. This application is receiving increasing attention from the building energy community due to its ability to perform fine-grained load monitoring at a low cost, in which no extra fee is required (Xia, Ba, & Ahmadpour, 2021). However, although some labeled datasets exist, the overall limited amount of annotated data used to train ML models is a crucial challenge that hinders the deployment of this technology at a large scale. To fill that gap, recent studies have investigated the

[4] http://www.scales-project.net/.

[5] https://data.openei.org/submissions/636.

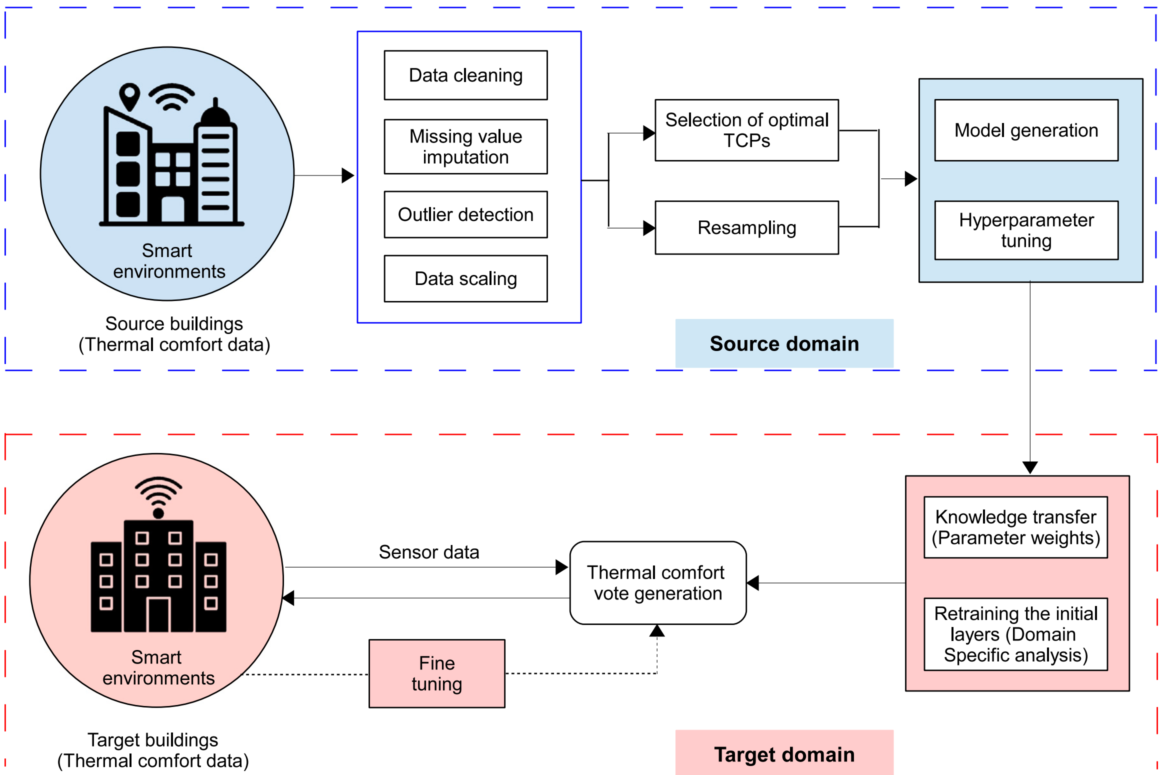


Fig. 9. A typical flowchart of TL-based thermal comfort prediction approach (Somu et al., 2021). Once the model is developed for the source building, it is refined for the actual target location (distinct zone/building, etc.) using TL.

possibility to train ML models on existing rich-labeled datasets and then generalize (or transfer their knowledge to different domains. That is, even the test data has been recorded in another country compared to the training labeled patterns (Katranji, Thuillier, Kraiem, Moalic, & Selem, 2016).

Data scarcity can be a severe issue when applying smart NILM solutions, especially those based on DL models. The latter requires vast datasets to learn individual (unidentifiable) loads accurately. Also, their generalizability to distinct domains is not guaranteed, e.g., when the test sequences are derived from a different region/country compared to the trained sequences (Murray, Stankovic, Stankovic, Lulic, & Sladojevic, 2019; Schirmer, 2021). To overcome these problems, TL has been introduced, which has significant benefits, among them (i) providing possibilities to perform oracle NILM, where a unique model can be deployed for all residential appliances everywhere; (ii) reducing the number of sensors to be installed as the learned characteristics of an appliance can be transferred to other appliances, i.e., by adopting CETL or CSTL, and hence reducing financial costs; (iii) offering considerable computational savings since pretrained models can be reutilized for other appliances or domains (D'Incecco et al., 2019).

In this context, the authors in D'Incecco et al. (2019) use a S2P model to investigate CETL and CSTL. In doing so, it has been demonstrated that when performing a CETL, the implicit characteristics learned by a "complex" device (e.g., washing machine) could be adapted to a "simple" device (e.g., a coffee machine). Regarding the CSTL, it was also possible to transfer the knowledge of S2P when the training and test datasets are from the same domain (similar buildings) without fine-tuning. By contrast, fine-tuning is required before applying the S2P learning to the test data if they are not in the same domain. Fig. 10 shows the concept of TL, portrayed by phase I, where all the orange convolutional (conv) layers indicate trainable layers using a comprehensive NILM dataset (Yang, Liu and Liu, 2021). such as the personalized retrofit decision support tools for UK homes (REFIT) (Murray, Stankovic, & Stankovic, 2017). In phase II, the first $n-1$ conv layers are frozen (layers are indicated by blue where the error does not get propagated to), while the last conv layer (or layers) and the fully connected (FC) are being adapted and trained to smaller NILM datasets like the REDD (Kolter & Johnson, 2011), and UK-DALE (Kelly & Knottenbelti, 2015).

A third scenario would be training a DNN model on data from another domain, such as famously investigated imaging datasets like ImageNet as in Deng et al. (2009), and then transferring the learning to the NILM domain; this is called cross different domains TL (ADD-TL). However, transitive TL should be applied to combat negative TL, which occurs if the original and destination domains have little to nothing in common (Li et al., 2022). It entails training the transferred DNN into an intermediate domain that shares common features between the source and destination domains (Liu, Zhang et al., 2021). In this regard, Liu, Wang, and You (2019) adopt an ADD-based TL to disaggregate energy footprints. Accordingly, the knowledge acquired by a pretrained AlexNet CNN model is used for classifying the color-encoded V-I trajectories of several device loads. In Murray et al. (2019), two different CTL-based models are proposed to improve the generalization of CNN and GRU models.

In Cavalca and Fernandes (2021), an ADD-TL scheme is implemented to segregate energy fingerprints from the REDD repository. Typically, the load time series is first transformed into 2D images. Next, features are extracted using DTL. Then, the classification and annotation of individual loads are performed. In Ahmed et al. (2020), two ML approaches are introduced based on GAN and TL-based GAN (TrGAN) models to perform NILM. Specifically, a GAN-NILM is first implemented using parameter sharing TL. However, as the results demonstrated that the parameter sharing is sensitive to the similarity between the source and target domains, TrGAN-NILM has been introduced to minimize the statistical distance between source and target domains in the feature space. Overall, TrGAN-NILM performs well and outperforms most of the existing approaches. Its only issue is that compact representations cannot be learned directly from the aggregated load without the target domain's devices' data.

Table 3
Summary of the reviewed TL-based thermal comfort control frameworks.

Work	Year	Model	Type of TL	Dataset	Best performance	Limitations
Grubinger et al. (2017)	2017	S2P	OTL	Private data	Moving average RMSE curves	• Training the GOTL over a period of more than five months is challenging.
Hu et al. (2019)	2019	ITCNN	CZTL	ASHRAE RP-884 , iTCM	Accuracy = 63.08%, macro-F1 = 53.06% 45.50%	• The performance needs further improvements. Data from one target domain (climate zone) was used in the validation.
Jiang and Lee (2019)	2019	LSTM (S2S)	DTL	SML, AHU, private data	CV-RMSE = 4.842%, MAPE = 3.402%, RMSE = 11.182%	• Not appropriate for multi-source TL or unsupervised domain adaptation (unlabeled target domain).
Mosaico et al. (2019)	2019	AlexNet + SVR	ADD	Private data	MAE = 0.7, RMSE = 1.2, NRMSE = 13%	• High complexity as occupancy is detected by analyzing images from thermal cameras.
Chen, Tong et al. (2020) and Chen, Zheng et al. (2020)	2020	Deep MLP	CSTL	Private data	MSE = 0.16	• Low transparency and interpretability.
Xu et al. (2020)	2020	DRL	CSTL	REDD	Violation rates $A\theta = 0$ and $M\theta = 0$	• Low transparency and interpretability. Did not consider the impact of building construction materials or weather conditions.
Lissa et al. (2020)	2020	Q-learning	CSTL	Private data	Average ToC error = 2.55%	• Consider only buildings from the same geographical location.
Bao et al. (2020)	2020	ANN + LPV-SS	OTL	Private data	MSE = 0.53	• Offline identified model and the online controlled system are similar.
Wang, Yuan et al. (2021)	2021	S2P, attention DNN	CTL	REFIT, REDD, UK-DALE	CV-RMSE = 17.49%	• Excessively introduced knowledge may cause the negative transfer phenomenon.
Gao, Shao et al. (2021)	2021	MLP, TGAN	CRTL	ASHRAE RP-884	Accuracy = 54.50%, F1 = 55.12%	• Low accuracy and F1 score performance. Only one target building was considered in the validation part. Did not consider buildings from different climate zones.
Somu et al. (2021)	2021	CNN-LSTM	CSTL	Scales Project, ASHRAE RP-884	Accuracy = 59.84%, PR = 56.68%, F1 = 56.54%	• Dependency to intrusive parameters and challenge in assessing its generalizability to different climate zones.
Park and Park (2021)	2021	CNN–SVM	CSTL	Private data	Accuracy = 95%, F1 = 95%	• The performance varies from a target subject to another.
Natarajan and Laftchiev (2019)	2021	S2P, attention DNN	CTL	REFIT, REDD, UK-DALE	RMSE = 0.82 ± 0.05	• Consider users' data from the same building. The number of users participated in the study is very low (five).

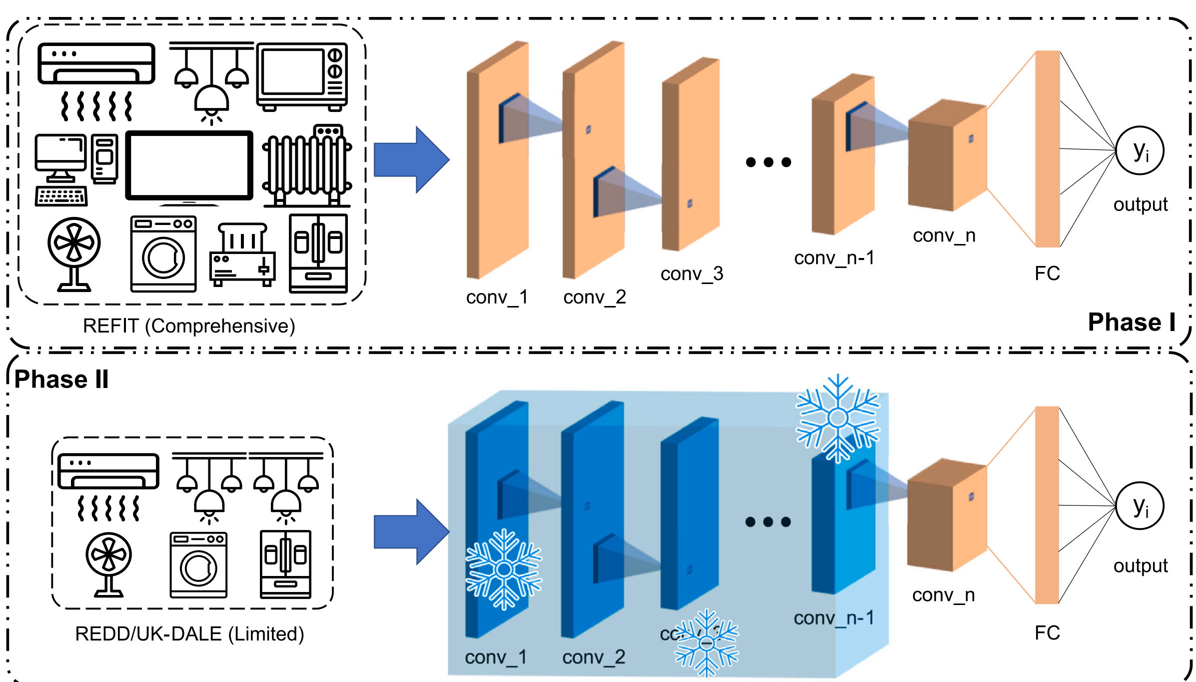


Fig. 10. A typical example of a TL-based energy disaggregation system (Yang, Liu et al., 2021), where (i) the orange convolutional (conv) layers indicate trainable layers and (ii) the first $n-1$ conv layers in blue are frozen (so the error does not get propagated to). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In Zhou et al. (2021), a smart online NILM solution is introduced, which has the ability of online inferring individual appliance loads using transferred knowledge with limited recorded data. In doing so, a DTL technique scheme is designed by using two DL networks in two steps. An LSTM model was first employed for extracting lower-level spatial and temporal characteristics from load gray-scale images generated after a 2D transformation. Secondly, a probabilistic neural network (PNN) was utilized to classify devices and transfer the knowledge between appliances. Also, the overall solution has been embedded into a smart plug for demonstrating its efficiency in real-world scenarios. Moving on, in Lin et al. (2021), Lin et al. propose a TL-based NILM using a temporal CNN model for learning the dynamic characteristics of individual devices' loads. Specifically, a domain adaption loss has been adopted for quantifying the domain distribution discrepancy between source and target domain representations. Also, joint optimization of domain adaptation and energy disaggregation has been used to learn invariant representations across domains of individual device states.

In Yang, Liu et al. (2021), a DNN algorithm using an attention mechanism is used to perform TL-based NILM, which can improve the conventional S2P model with an attention layer and a time-embedding layer. Thus, this scheme helped abandon the RNN structure and shorten the training time, and hence, it was adequate to be used in model pretraining with large datasets. Its validation on three distinct datasets: REFIT, UK-DALE, and REDD, has shown the superiority of the new attention model compared to a S2P model. Similarly, in Wang, Du and Zhou (2019), a deep transfer-learning-based NILM technique is introduced. A unique disaggregator is used for every appliance signature, which involves a feature extractor and a regressor. Then, a learning model based on deep domain adaptation has been utilized, which somehow helps align the distribution of energy patterns using both the labeled and the unlabeled data.

Table 4 summarizes and compares the frameworks mentioned above in terms of the appearance year, learning model, type of TL, the dataset used in the validation process, and limitations. Hence, it is worth noting that by adopting the TL strategy, a NILM algorithm can be deployed in any environment without re-training it with new data from scratch (from the new environment). In other words, TL improves learning in a new environment via transferring knowledge from an initial climate that has already been learned. Therefore, this means that an end-user without any technical expertise can easily install such NILM solutions. However, the burden falls over the developers' shoulders to implement a delicate transfer of knowledge from one domain to another. The burden entails finding the closest domain to the NILM domain or identifying the best intermediate domain common between them. For the latter case, it might require data generation and labeling steps.

Although the increasing interest from academia and industry to developing NILM systems, the latter still requires cross-building adaptation. Typically, a system designed for one building cannot be generalized to other buildings (applicability) and needs significant hand-tuning before its application to those buildings (scalability). To alleviate these problems, Humala et al. (2018) introduce a semi-supervised CSTL-based NILM approach, namely UniversalNILM. The latter is based on (i) modeling electrical devices in a few training buildings that have fine-grained device-level signatures using combinatorial optimization (CO); and (ii) transferring this knowledge to the test buildings (i.e., having only aggregated consumption fingerprints) to extract appliance-level load footprints.

In Houidi et al. (2021), a CSTL-based appliance identification approach is developed, where pertinent and understandable features are first selected to help efficiently discriminate distinct individual appliances. In this line, a new dataset named home electrical appliances recordings of NILM (HEAR-NILM)[6] is introduced, which includes appliance-level electrical characteristics recorded in steady-state conditions. Next, the performance of various feature selection approaches and classification models, e.g., DNN, k-nearest neighbors (KNN), linear discriminant analysis (LDA), principal component analysis (PCA), and mutual information (MI) have been assessed. Moving on, a TL is applied to move the knowledge learned on the selected features from the new dataset to a large-scale dataset, i.e., PLAID. Similarly, to overcome the problem of data shortage, a CSTL-based energy disaggregation scheme is proposed in Zhang, Watkins, and Kuenzel (2021). Typically, multi-quantile LSTM (MQ-LSTM) and multi-quantile GRU models are introduced to segregate the aggregated loads. They have been trained on a public dataset and then fined tuned on a local dataset (target domain) from a different area with distinct interval resolution and different portions of energy signatures.

4.5. Renewable energy

The energy transition is accelerating the transformation of the energy sector and, with the expansion of renewable energies, is bringing more and more new components into the system. As a result, there may be little or no information available for new or modified plants. The usual ML models cannot be applied here, and alternative physical models do not always provide accurate results (Ding, Zeng, Hu, & Yang, 2022). To that end, many studies have recently been proposed, researching new methods for automated knowledge transfer between individual system components in renewable energy systems (Schreiber, 2019).

The broad adoption of wind power systems has made the surge in developing reliable and accurate probabilistic wind power forecasting frameworks. In Qureshi, Khan, Zameer, and Usman (2017), a CZTL-based short-term wind power forecasting scheme is introduced by exploiting the knowledge of the DNN-based meta-regressor technique (DNN-MRT) model. In this regard, a DAE has been used as a base-regressor, while a deep belief network (DBN) has been utilized as a meta-regressor. This system has experimentally been trained and tested on a three-year dataset gathered from five different wind farms situated in distinct climate zones in Europe. This dataset includes power measurement (p) along with meteorological forecasts related to components of wind, i.e., the corresponding speed (S), meridional component of surface wind (MS), zonal Component of Surface Wind (ZS), and direction (D) of wind. Similarly, in Cai, Gu, Ma, and Jin (2019), another CZTL-based probabilistic wind power forecasting method is introduced. Typically, a wind power quantile regression scheme is developed by combining an instance-based TL model and gradient boosting decision tree (GBDT). This approach exploits the spatial cross-correlation characteristics of wind power generations in different zones to predict wind power accurately.

Because obtaining the signal data of wind turbine faulty gearbox is challenging along with diagnosing of the health condition under variable working conditions, Ren, Liu, Shan, and Wang (2019) introduce a fault diagnosis approach. It relies on variational mode decomposition (VMD), multi-scale permutation entropy (MPE), and feature-based TL (FTL). Typically, a CETL-based methodology is adopted, where the target domain data is different from that of the source domain with reference to the working conditions. Data representing a series of mode components are collected by transforming the signals under variable conditions, according to the vibration signal characteristics of wind turbines. Moving on, the covariance between the source domain and the target domain has been minimized using a linear transformation matrix. The difference in data distribution between the two domains has been reduced. Next, characteristics of the covariance-aligned source and target domains have been fed into an SVM classifier for training and testing. In Guariso, Nunnari, and Sangiorgio (2020), a DDNA-based solar irradiance forecasting approach is introduced using feed-forward LSTM (FF-LSTM). Typically, the latter has been developed and pretrained on the source domain data (collected from Como station,

[6] http://dx.doi.org/10.21227/ww76-d733.

Table 4
Summary of the reviewed TL-based energy disaggregation frameworks.

Work	Year	Model	Type of TL	Dataset	Best performance	Limitations
Humala, Nambi, and Prasad (2018)	2018	CO	CSTL	REDD, UK-DALE	F1 = 0.58	• The performance needs further improvements and it has been tested on only four appliances.
Liu et al. (2019)	2019	AlexNet CNN	ADD	PLAID, WHITED	Accuracy = 99.05	• Energy disaggregation of multi-state loads is still challenging, where the performance has dropped.
Murray et al. (2019)	2019	CNN, GRU	CTL	REFIT, REDD, UK-DALE	F1 = 72.33%	• Performance varies from a dataset to another and from an appliance to another, e.g., under REFIT, recognizing the load of a dishwasher was very challenging, while under DRED, the challenge was with the refrigerator.
D'Incecco et al. (2019)	2020	S2P	CETL, CTL	REFIT, REDD, UK-DALE	MAE = 11	• When the domains are similar, fine tuning may result in overfitting.
Cavalca and Fernandes (2021)	2021	CNN VGG16	ADD	REDD	Accuracy = 96.7%	• validation on one dataset, which is not sufficient. Moderate performance and privacy and security issues.
Ahmed et al. (2020)	2021	GAN, TrGAN	CTL	REFIT, REDD, UK-DALE	F1 = 80%	• Compact representations cannot be learned directly from the aggregated load.
Zhou, Xiang, Xu, Yi et al. (2021)	2021	LSTM, PNN	ATL	PLAID	Accuracy = 92.27%	• The accuracy of disaggregation drops when multi-state appliances are considered.
Lin et al. (2021)	2021	TCNN	ADD	Private data	MAE < 3	• The disaggregation of energy footprints including simultaneous appliance switching operations is the main issue.
Yang, Liu et al. (2021)	2021	MQ-LSTM, MQ-GRU	CTL	REFIT, REDD, UK-DALE	MAE < 16	• Performance varies from an appliance to another, and ATL has not been investigated due to space limitations. Only low-frequency was used, which results in inaccurate prediction of appliance-related events.
Houidi, Fourer, Auger, Sethom, and Miègeville (2021)	2021	DNN	CTL	PLAID, HEAR-NILM	Accuracy = 99%	• Require a supervised pre-training on the target dataset to achieve high performance and high complexity to select the features.
Houidi et al. (2021)	2021	DNN	CTL	Private data	RMSE = 0.58, $R^2 = 0.96$	• Require labeled training datasets and the performance needs to be improved.

in Italy) before being applied on the target domain with different characteristics (e.g., latitude and geographical settings).

Another critical research topic that gains polarities in grid energy management and renewable energy is related to developing lithium-based batteries, which offer high specific power/energy densities. Specifically, developing efficient and reliable lithium-based batteries (i) plays a critical role in the mass roll-out of EVs, facilitates the penetration of renewable energies globally (Shen, Sadoughi, Chen, Hong, & Hu, 2019), and provides explicit socio-economic benefits when transitioning towards the next-generation of energy systems. To ensure a reliable and safe operation of lithium-based batteries, estimating their lifetimes and online capacities is of utmost importance. This task is performed using data-driven methods, including kernel regression models or NNs, by learning the complicated dependency of cell characteristic features extracted from the current and voltage measurements on the capacity of a cell (Zhang et al., 2020). In this context, DL models have demonstrated promising potentials in battery capacity estimation. However, they have a plethora of parameters to be optimized and require a massive amount of battery aging data for training. Also, most of the ML-based algorithms suppose identical data distributions between testing and training batteries, although distribution discrepancies exist in real-world applications (Cheng et al., 2021).

To overcome the issues mentioned earlier, Li, Li et al. (2021) use TL and network pruning to develop compact CNN-based lithium-ion battery capacity estimation, which works on small datasets with enhanced estimation performance. Accordingly, the CNN model is first pretrained on a large source battery dataset, and the acquired knowledge is then moved to a small target dataset for improving the estimation accuracy. Moving forward, an approach for pruning the transferred model using a fast recursive algorithm is introduced. This has helped reducing the computational cost and size of the model while conserving good performance. Moving on, in Che et al. (2021), the remaining useful life (RUL) of a lithium-ion battery is predicted using a technique-based online model correction with TL. This was possible by (i) optimizing the threshold for health indicators using a Gaussian process regression, and hence determining the end-of-life, and (ii) introducing an evaluation approach for assessing the health indicators, and (iii) combining TL and gated RNN for predicting the RUL and promoting online applications. Similarly, in Kim et al. (2021), Kim et al. propose a DTL-based health prediction approach for Li-ion batteries, which is based on a variational LSTM (VarLSTM). In this regard, after training the VarLSTM on the source dataset, its knowledge has been transferred and to the target dataset. In doing so, the weights of LSTM cells have been frozen and moved to the target dataset network while the weights of the fully connected have been re-trained.

In Azkue, Lucu, Martinez-Laserna, and Aizpuru (2021), the authors introduce a calendar aging approach for lithium-ion batteries using a DTL model. In this line, the weights of the baseline DNN algorithm have been retrained using a small dataset, where the initial weights of DTL model have been adapted to the target dataset on account of the input–output patterns learned during the pretraining. Besides, in Liu et al. (2021), a DTL-based model is designed to predict the state-of-charge for lithium-ion batteries, which is based on LSTM. The LSTM five layers have been used to learn the state-of-charge dependency using a large-scale dataset. Next, a TL strategy based on fine-tuning has been utilized to regulate the parameters of the fully-connected layer and share other layers' knowledge.

4.6. Smart grid (SG) and energy trading

Cyber-security in SG: the SG has been among the significant breakthroughs of the energy sector; it showcases the best usage of computer intelligence to manage the energy network. Its benefits are enormous, among them (i) improving the reliability and efficiency of the electricity supply, (ii) integrating renewable energy into the existing networks, (iii) providing end-users with the tools for optimizing their electricity consumption, and (iv) assisting the development of EVs (Wang, Jiang, Wang, Lv and Nowak, 2021). Smart meters installed in the SG generate large-scale datasets, where AI and ML models are employed to analyze recorded data for various purposes, e.g., optimizing energy consumption, detecting faults and anomalies of the electrical network, and improving thermal comfort and safety of end-users. Moreover, the SG faces rising cyber–physical attack threats, aiming to hack the cyber–physical systems (CPS) or the communication systems, resulting in national security deficits. Again, ML and DL play a critical role in protecting the SG infrastructures against sophisticated attacks (Babar, Tariq, & Jan, 2020).

However, most ML/DL algorithms assume that the testing and training datasets follow similar data distributions, which is not the case in most dynamic time-varying SG systems. This is because operating points can shift drastically over time, and therefore resulting data distribution variations can degrade classification performance and drive delayed incidence responses. To close this gap, TL has recently been proposed (Zhang, Bao, Yu, Yang, & Han, 2017). For instance, in Zhang and Yan (2019), Zhang et al. introduce a domain-adversarial TL (DATL) scheme to robustly detect intrusion against SG attacks. This approach uses domain-adversarial training (DAT) to map the annotated source domain and the unlabeled target domain. In this regard, the classifier deployed learns and detects unknown threats in the new feature domain. An SG cyber-attack dataset.[7] has been used to evaluate this approach and other baseline classifiers. Similarly, in Zhang and Yan (2020), a semi-supervised DATL scheme is proposed, in which NNs have trained a dataset of known attack incidences before transferring its knowledge to another domain where data of the attack incidences is rare, and only the normal data is labeled (abnormal data is unknown). This approach has been then evaluated and compared to five baseline classifiers, including ANN, SVM, classification and regression tree (CART), and RF.

Moving forward, in Wang, Wu, Zhang, and Wang (2018), the knowledge acquired using the TrAdaBoost algorithm (which is an instance-based TL algorithm Dai, Yang, Xue, & Yu, 2007) is transferred from a source domain with a large amount of data to a data-scarce target domain for constructing the dynamic security defense strategy of vehicle-to-grid (V2G) systems. The evaluation of this scheme has been conducted on simulated training/testing datasets, which have been generated using Mininet.[8] Moving one, While using AI IoT devices can help better manage the SG and ensure reliable communication between the grid nodes, the enormous amounts of data recorded in geographically vast grids overload the communication infrastructures. To alleviate this issue, data compression at the source before communicating compressed patterns has been investigated in the literature (Huang, Zhang and Hsieh, 2021; Zhou, Xiang, Xu, Wang and Shi, 2021). In Das, Garg, and Srinivasan (2020), an autoencoder is used to extract pertinent structures from the SG data and then proceed with its compression. Moving on, the generalization of this approach to data from distinct geographical locations has been explored. Typically, the knowledge of the autoencoders that have been pretrained on data from a specific location has been transferred and adapted to the idiosyncrasies of a target domain.

SG management: detecting faults in the powerlines of SG is challenging due to the insufficiency of fault data used to train ML algorithms. Indeed, ML-based AFDD models are trained using simulated data generated using different software (EnergyPlus, Matlab/Simulink, etc.). This causes some issues in detecting the faults in real-world scenarios due to the difference between simulated and accurate data. Thus, a fault classification approach is developed using DATL in Han et al. (2021) to meet this challenge. Accordingly, a conditional generative adversarial network (CGAN) has been utilized to augment the actual fault data. Following, the loss function of CNN has been redesigned based on TL to build a new fault classification scheme based on improved CNN. This scheme has been then trained on both simulated and adversarial data before being validated on real-world powerlines data. Moving forward, because early warning mechanisms are significant to maintain the reliability and security of SG systems, a vision-based dataset with a TL-based approach that enables early warning classification is developed in Gao et al. (2021). The dataset includes a large number of instances pertaining to ten groups of high-risk objects, and power grid infrastructures have been labeled. The knowledge learned with CNNs to recognize power grid infrastructures and high-risk objects has been transferred to evaluate its generalization ability in local regions by loading the trained local patch responder with frozen weights.

On the other hand, with the integration of renewable energy into the smart grid and the rapid increase of dynamic loads, new challenges have arisen to the stable operation of sub-transmission networks. Typically, the short-term voltage stability (STVS) problem rises in sub-transmission expansion planning (SEP) and therefore threatens the stable operation of energy systems. While conventional methods of treating STVS based on time-domain simulation are computationally costly, Huang, Zhang and Zheng (2021) introduce a DTL-based method based on bi-directional LSTM (BiLSTM), which efficiently identifies resilient network structures at a low cost by only retraining some layers of the model. In this respect, the knowledge acquired by the BiLSTM on the original source system has been expanded to a similar target source system (i.e., the similarity is in terms of composite loads, transmission lines, and physical laws). Moreover, both the source and target tasks aim at evaluating the STVS performance of different network structures, and there is almost no difference in terms of the data size. Similarly, in Li, Zhao, Lee, and Kim (2019), an NN-based TL scheme to infer voltage stability margin is proposed, which requires a small amount of offline-computed voltage stability margin data. Typically, the NN predictor is first trained on a large binary stability-labeled dataset, and the pretrained model is transferred and fine-tuned on a small dataset of margins.

Energy trading: energy price prediction is becoming of significant importance for all the deregulated markets of the world. Recent research has proved the significance of accurately predicting day-ahead electricity prices, where different historical data from numerous markets could be utilized as inputs for the prediction models. An ensemble of ML and DL models has been adopted to perform this task. However, a principal issue is how to exploit available multi-market source data is still neglected effectively. Recent research has focused on using TL for electricity price prediction (EPP) to answer this question. For instance, in Gunduz, Ugurlu, and Oksuz (2020), a TL is adopted as a strategy to exploit information from a set electricity price source markets and then forecast the electricity price of a target market. Accordingly, a bidirectional gated recurrent units (BGRU) model has been pretrained using data from the source markets before performing a fine-tuning and validating it on the target market. Since forecasting real-time electricity price for wind power is crucial in operating energy markets and avoiding price risks, DNN has been applied as it captures the temporal relations of historical price time-series. Also, it automatically extracts the relevant characteristics of the massive amount of data. However, DNN-based models still need further improvements to deal with small datasets and improve the prediction accuracy. In this context, a TL-based electricity price forecasting approach is proposed in Yang and Schell (2021), to transfer the knowledge acquired with a GRU on different source domain data collected from different wind farms (in the same region). This TL-GRU model has outperformed a DNN-based model by 6.7% in terms of the mean absolute percent error (MAPE).

[7] https://sites.google.com/a/uah.edu/tommy-morris-uah/ics-data-sets.

[8] http://mininet.org/.

5. Discussion of key challenges

TL-based domain adaptation techniques have proved their efficiency in many application scenarios as they can outperform conventional ML models in terms of different evaluation metrics. That is because target domain data could seriously diverge from that used to train the ML models. Moreover, the operation characteristics of the energy systems such as HVAC systems, chillers, BAMSs, electrical devices, etc, have a significant impact. However, it is worth noting that there remain various problems. For instance, most TL-based frameworks ignore describing the procedure of selecting the source domain data to facilitate the learning of the target task. Indeed, in most situations, the success of TL models lies in the similarity between the source–target domains. Thus, the absence of similarity between the source and the target domains significantly damages the TL, and even worse results in negative transfer (Weiss, Khoshgoftaar, & Wang, 2016; Xu & Meng, 2020). Additionally, most DL approaches heavily depend on preprocessing, e.g., frequency-domain analysis, time-frequency domain transform, or time-domain feature calculation, where the similarity between the source and target domains should be high, and the dimensions between them need to be consistent. However, few studies have been devoted to overcoming the inconsistency of the source and target information dimensions. Among them, the study proposed in Hu et al. (2019), where a DTL has been explored to conduct an efficient knowledge transfer among heterogeneous domains. This section highlights the pressing challenges attracting considerable interest in the actual time.

5.1. The problem of negative transfer

When the TL ends up with a degradation of the classification of prediction performance (or accuracy) of the newly developed model, it is due to negative transfer. Indeed, TL perfectly works if the source and target domains are sufficiently similar. Put differently, when the data used to pretrain the TL model is different enough than the data used to re-train this model (or some of its parts), the performance might be worse than expected. Moreover, regardless that the source and target domains can appear similar to humans, algorithms may not always agree with them. For example, in the building energy sector, even the source and target data are collected from very similar buildings located in the same region; it may be very different as each building is personalized due to the variations of operation mode, occupant behavior, thermal performance, etc. Moreover, the prediction performance of the data-driven model in similar buildings is different due to the factors mentioned above. In this regard, to correctly transfer the knowledge of TL and avoid negative transfer, it is of utmost importance to appropriately select the source domain, which is highly similar to the target domain.

It is also worth mentioning that some studies, such as Dai et al. (2007), have demonstrated that the quality of the TL performance has a direct relation with the Kullback–Leibler divergence estimated between the source domains and target domain datasets (Sousa, Silva, Alexandre, Santos, & De Sá, 2014). Put simply, it may not be practical for some application scenarios to use a TL; or, the successful TL's architecture should be reliable for heterogeneous problems. Additionally, although these intuitive ideas have experimentally shown a relation between domain divergence and TL algorithms' performance, theoretical descriptions for these behaviors are still unknown.

Although there has been an increasing interest from the ML community to fix the problem of negative transfer in TL algorithms, e.g., Chen, Wang, Fu, Long and Wang (2019), Gui, Xu, Lu, Du, and Zhou (2018) and Wang, Dai, Póczos and Carbonell (2019), only one study has investigated this issue in the case of smart city or energy systems. Specifically, Niu et al. (2020) attempted to avoid the negative transfer problem by using a multi-source TL. This helps optimize energy management using occupancy detection patterns and a multi-source knowledge transfer. Typically, the negative transfer caused when using only one data source domain has been avoided. Moreover, the performance has been improved compared to the case of unsupervised ML. To that end, one straightforward solution to avoid the negative transfer in smart cities and energy systems is to transpose and adapt what has been done in other research fields (Zhang, Deng, Zhang and Wu, 2020), e.g., by filtering out unrelated source data as explained in Wang, Dai et al. (2019).

5.2. The problem of overfitting

One of the challenges in developing TL-based techniques for energy applications is overcoming overfitting, which is due to re-training complex models with insufficient data. The size of the training dataset is not the only influencing factor in this case. It is accompanied by the complexity of the data-driven models, that is mainly in terms of the set of model's parameters. Although this issue is familiar with all the data-driven models, overfitting in TL also occurs when the developed model learns details and noises from source domain data that negatively impact its outputs (Delfosse, Hebrail, & Zerroug, 2020). In TL, the network layers cannot be removed to identify with confidence the best classification/prediction parameters of the ML models. Typically, removing the first layers may negatively impact the dense layers since the number of trainable parameters will change. On the flip side, the number of dense layers can be reduced; however, the analysis of the number of layers to be removed while avoiding the overfitting of the model is time-consuming and challenging.

Hence, the first step in applying TL is the careful and systematic devising of the TL problem in terms of (i) the suitability of TL under the existing constraints on the available data, and (ii) the proper TL model selection according to the problem at hand. That is, TL should be applied wisely to make use of its advantages. Even though TL is a solution, it is not necessarily the solution for all applications. For that, it is essential to conduct an assessment of the problem considering the task objective, available data, current constraints, and possible ways to tackle the problem. Nevertheless, overfitting can be partially mitigated following two schemes; (i) by using regularization techniques, e.g., least absolute shrinkage and selection operator (LASSO) regularization for multiple linear regressions (Das, Nair, Reddy, & Venkatesh, 2018), and dropout techniques for DL models (Alghamdi et al., 2020; Fan et al., 2020); (ii) by elaborately designing a model development approach. This is possible by dividing the overall datasets into three ensembles for training, validation, and testing (using cross-validation) to optimize model parameters; by applying early-stopping strategy in model's re-training (Prechelt, 1998). Also, avoiding overfitting can be achieved by adopting data augmentation strategies for generating synthetic data when training the DL model (Jha et al., 2019; Zhao, 2017).

5.3. Reproducibility of scientific results

Although the increasing interest devoted to using TL in multiple energy applications, some factors are hindering the broad adoption of TL-based models and essentially affecting reproducibility, and thus empirical comparisons of TL-based solutions: (i) it is still challenging to evaluate the generality of TL models as most of the frameworks were evaluated on datasets that are collected from the similar subjects that are under the same climate conditions; (ii) there is a significant lack of using the same datasets and benchmarks to validate the new TL models. This is because of the limited number of existing open-source, benchmarked datasets; and (iii) different metrics and parameter settings have been utilized to quantify the distance between the source and target domains and assess the performance of TL-based solutions in different datasets. The challenges above make the comparison of TL techniques uniformly complicated even impossible.

Despite that energy disaggregation community has launched the open-source non-intrusive load monitoring toolkit (NILMTK) (Batra et al., 2019) to allow fair and accessible comparisons of energy disaggregation algorithms in a reproducible fashion (Batra et al., 2019). However, the effort put in this direction is insufficient and does not consider the challenges introduced by TL. Moreover, in Cook, Feuz, and Krishnan (2013), Cook has attempted to define a realistic assumption to quantify the similarity between two domains using a universal and domain-independent distance. The latter enables intelligently select appropriate algorithms and assess the performance of TL solutions; however, its application has been limited.

5.4. Measuring knowledge gains

Measuring the knowledge gained when a TL model is adopted to conduct specific tasks is of utmost importance. However, this challenge did not receive its merit, and a few research works have targeted it. Bengio et al. in Glorot, Bordes, and Bengio (2011) have attempted to analyze how to quantify the TL gain. In this regard, four measures have been introduced to quantify the gain knowledge, i.e., transfer error, transfer loss, transfer ratio, and in-domain ratio. Despite that these measures can overcome some interpretation issues related to the performance results occurring when dealing with various source domains, it is unknown how they will behave in other TL-based methods, especially for energy applications where class sets are different between problems. Further, they can result in non-definite performance if a perfect baseline model is obtained.

To that end, simpler measures, including accuracy, F1 score, MSE, RMSE, MAE, PIR, or other statistical inspired coefficients, which could provide further information, e.g., the class agreement, have been widely investigated for evaluating TL-based solutions in the energy sector.

5.5. Unification of TL

One principal challenge that may still impede the advance of TL-based energy applications is the wide range of formulations used to describe the mathematical background of developed TL algorithms. For instance, while (Hu et al., 2019) promotes the idea of Heterogeneous TL, Fan et al. (2020) opts for statistical investigations of TL-based methodologies, Lin et al. (2021), Zhang and Yan (2019, 2020) focus on domain-adaptation TL. Although these frameworks and others included in this review share the same TL idea, they differ in their definition and implementation based on the scenario under consideration. More importantly, different variant terminologies are used, leading to confusion. To alleviate this issue, a unification of TL definitions and background is becoming an emergency. Although the first tentative for unifying TL has been proposed in Patricia and Caputo (2014), this is still not enough to cover the energy sector.

6. Case studies

6.1. TL-based energy prediction with mobility data

Because of the COVID-19 global pandemic, enormous disruptions have been brought to the operations of the energy systems. To optimize energy consumption and maintain the lighting systems to work perfectly during this problematic period, accurately forecasting energy demand to correctly schedule electricity generation has been of utmost importance. However, the high variations of occupancy levels and mobility patterns and the quarantine and curfew measures applied globally have considerably reshaped energy usage. Thus, precise forecasts of future energy demand have become challenging. For instance, California independent system operator (CAISO)[9] has released a public dataset, which shows that in April 2020 energy demand was consistently overforecasted. This is mainly due to the impact of new factors, such as the mobility data, non-availability of sufficient data to train forecasting models, etc.

We present in this section an example of using TL to predict energy demand while considering mobility data for informing the forecast model of socioeconomic changes. Moreover, an MTL is adopted to transfer the knowledge of an NN model among different load regions. In this context, the dataset used to validate this MTL-based load forecasting scheme includes multiple data sources from 12 areas of different countries. Mobility data is integrated into the NN similarly to weather data, where all available features are concatenated and fed into the input layer. The innovation of this model is related to integrating the mobility data as day-ahead weather forecasting models have been widely used and are pretty accurate. In this regard, to perform a day-ahead forecast, day-ahead weather forecasts are concatenated with the current day's mobility data. For an n-layer NN, every hidden layer $H_j; j = 1, \ldots, n-1$ is parameterized as a fully-connected layer:

$$H_j = \sigma_j(W_j H_{j-1} + b_j) \tag{2}$$

$$\widehat{P}_{t+k} = W_n H_{n-1} + b_n \tag{3}$$

where b_j are the biases and W_j are the trainable weights at layer j, while σ_j represent the nonlinear activation functions for promoting nonlinearity in the NN model. The stochastic gradient descent has been used for minimizing MAPE in the training phase while the actual load P_{t+k} is collected.

$$P_{MAPE} = \frac{1}{M} \sum_{j=1}^{M} \frac{\left| P_{t+k} - \widehat{P}_{t+k} \right|}{P_{t+k}} \tag{4}$$

In the evaluation phase, four forecasting algorithms are considered:

- "Model 1: NN", which is a standard NN for day-ahead energy prediction (without mobility data);
- "Model 2: Retrain", which has the same architecture as model 1 but retrained when COVID-19 pandemic takes the impact on energy consumption (training data ranging from 15/02/2020 to 30/04/2020);
- "Model 3: Mobi", which is based on including the mobility data (training data from 15/02/2020 to 30/04/2020); and
- "Model 4: Mobi-MTL", which extends model 3 by using multi-task where similar tasks with similar-sized load regions are selected. The MTL energy prediction process is designed by sharing the features among various prediction tasks. In this respect, different NNs are collectively co-trained as portrayed in Fig. 11. Typically, the features, including the actual load, weather data, timing, and mobility data from the Apple mobility report and Apple (2021) Google COVID-19 mobility report (Google, 2021) are split (or shared) between different prediction tasks (Fig. 11(b)) instead of using all the features with one single task model (a conventional NN model), as depicted in Fig. 11(a). Moreover, it is worth noting that the features are randomly and equally split (shared) between the various prediction tasks.
  For an ensemble of prediction tasks $i = 1, 2, \ldots, R$ with the training datasets D_{tr} considered in this framework, the same weights of the first q layers have been shared by the energy prediction models. While, the last $q - n$ layers have mapped the embeddings H_j to distinct outputs $\widehat{P}_{t+k}$. In doing so, a NN is constructed as follows:

$$H_q = \sigma_q(W_q(\sigma_{k-1}(W_1 H_0 + b_1)\ \cdots)) + b_q \tag{5}$$

$$\widehat{P}^{(i)}_{t+k} = W^{(i)}_n(\sigma^{(i)}_{n-1}(\cdots \sigma^{(i)}_{q-1}(W^{(i)}_{q+1} H_q + b^{(i)}_{q+1} \cdots)) + b^{(i)}_n \tag{6}$$

  For the training of the MTL prediction NN (defined by Eqs. (5) and (6)), a batch of training patterns has been sampled from $D^{(i)}_{tr}$ for every task i and the weights have been updated for $W_j, b_j; j = 1, \cdots, q$ and $W^{(i)}_j$, $b^{(i)}_j$,$\sim j = q+1, \ldots, n$. Moving on, a fine-tuning step

[9] https://www.caiso.com/Pages/default.aspx.

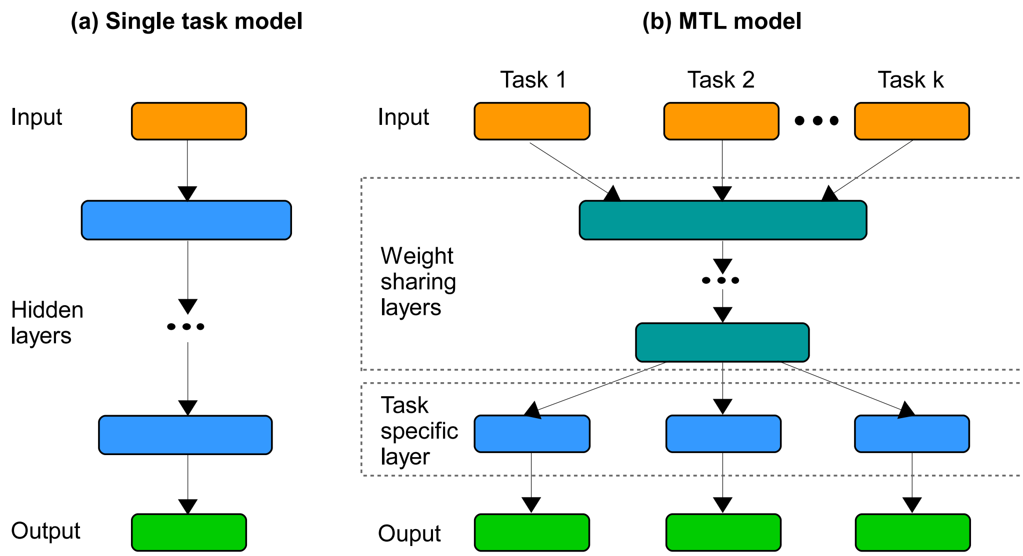


Fig. 11. The architectures of energy prediction models; (a) single prediction task, and (b) MTL.

has been employed to enhance the performance of this model, in which only a specific task i has was trained while the trained weights $W_j, j = 1, \ldots, q$ were fixed.

Two training datasets have been used to assess the performance of the aforementioned models. Mobility data, covering the period 01/01-15/05, 2020, has been excluded in the first dataset. By contrast, mobility data has been used in the second dataset, which spans for three months (15/02/–15/05, 2020). Both datasets include relatively small data for energy prediction. Moreover, the former represents the pre-lockdown period while the latter spans the after-lockdown duration. Fig. 12 portrays the day-ahead energy prediction results and error distributions obtained when the aforementioned models are considered. Typically, they represent two weeks of data that are related to the "Seattle City" light service data. Mobi-MTL has been trained using datasets recorded from "Boston", "Chicago" and "Mid-Atlantic" areas. The superiority of both "Mod. 3: Mobi" and "Mod. 4: Mobi-MTL" has been clearly seen compared to the other two models. Typically, it has been shown that the integration of mobility information helps the NN in better predicting the electricity usage profiles. Moreover, "Mod. 4: Mobi-MTL" has achieved the smallest energy prediction error. This validates the intuition that CZTL prediction knowledge could be beneficial when only limited data is available.

6.2. TL-based load forecasting in sport facilities

Buildings consumption modeling can be utilized to manage and optimize their performance and operation to enhance energy efficiency. Data-driven methods have been widely adopted for consumption prediction/forecasting due to the availability of data in the building automation and automation systems (BAMSs). Consumption prediction/forecasting for sports facilities was investigated and applied in several studies. For example, in Yuce et al. (2014), a NN was utilized for the prediction of energy consumption and thermal comfort level to manage and control an indoor swimming pool, in which simulation data was used to train the NN model, while real-world data was utilized to calibrate the model. Additionally, in Elnour et al. (2022, 2021), an NN-based dynamic forecasting model of a sports hall, developed using simulation data, was employed in a model predictive control (MPC) system to optimize the operation of its management system in terms of energy consumption and users thermal comfort. As shown in Fig. 13, the MPC system consists of an optimizer and a forecasting model of the building operation to determine the best HVAC system settings based on an objective function. It attempts to determine the control inputs $u(k+1)$, that are the zone's temperature setpoint, the AHU supply air's temperature setpoint, and the fresh air inlet flow rate. Numerical optimization is employed to find the proper inputs over a prediction horizon that provide the best-predicted performance according to the objective function. As indicted by Elnour et al. the accurate performance of the NN-based forecasting model is essential for a reliable operation of the proposed MPC-based system. These studies are developed based on simulation data which are likely to have a different distribution from the operation data collected from the actual facility. Hence, a sufficient amount of practical data of adequate quality is required for real-life applications.

Data-driven consumption modeling of new facilities and outdated existing ones may be challenging as they probably lack sufficient data to train the data-driven models. Moreover, BAMS operational data are likely to have quality issues related to missing data, outliers, and other anomalies. Hence, issues may be encountered for the real-life application of these methods as a result of insufficient performance of the data-driven forecasting model due to the limitation on the availability of data that cover diverse operating conditions. Consequently, TL is useful in these scenarios for information extraction with limited data. In this case study, we demonstrate the utilization of TL to boost the performance of an NN-based dynamic forecasting model of the sports hall in Elnour et al. (2021) when only limited practical data is available. Initially, the NN-based model is trained using data generated from a simulation model for 20 days that cover diverse operating conditions. Then, it is used to develop a calibrated version of the model using a limited practical dataset of around 2 days of the sports hall operation. The benchmark of the performance of the TL-based models is a standalone model that was trained, from scratch, using two days worth of practical data. Two TL-based models were developed, which are (i) Model 1 in which all the layers of the pretrained models are re-trained, and (ii) Model 2 where only the output layer of the pretrained model was retrained. In Model 1, the pretrained model is used for the weight initialization of the TL-based model, considering that the simulation data and the practical data differ in data distribution. While in Model 2, the fixed layer of the pretrained model is used for feature extraction, considering that the simulation data and the practical data have identical features.

Fig. 14 presents a comparison between the standalone model and the two TL-based models averaged over 30 repeats, given the stochastic nature of the NNs. The performance of the standalone model and the two TL-based models on the training dataset is adequate with an MSE of less than 0.01, as shown in Fig. 14(a). However, as demonstrated

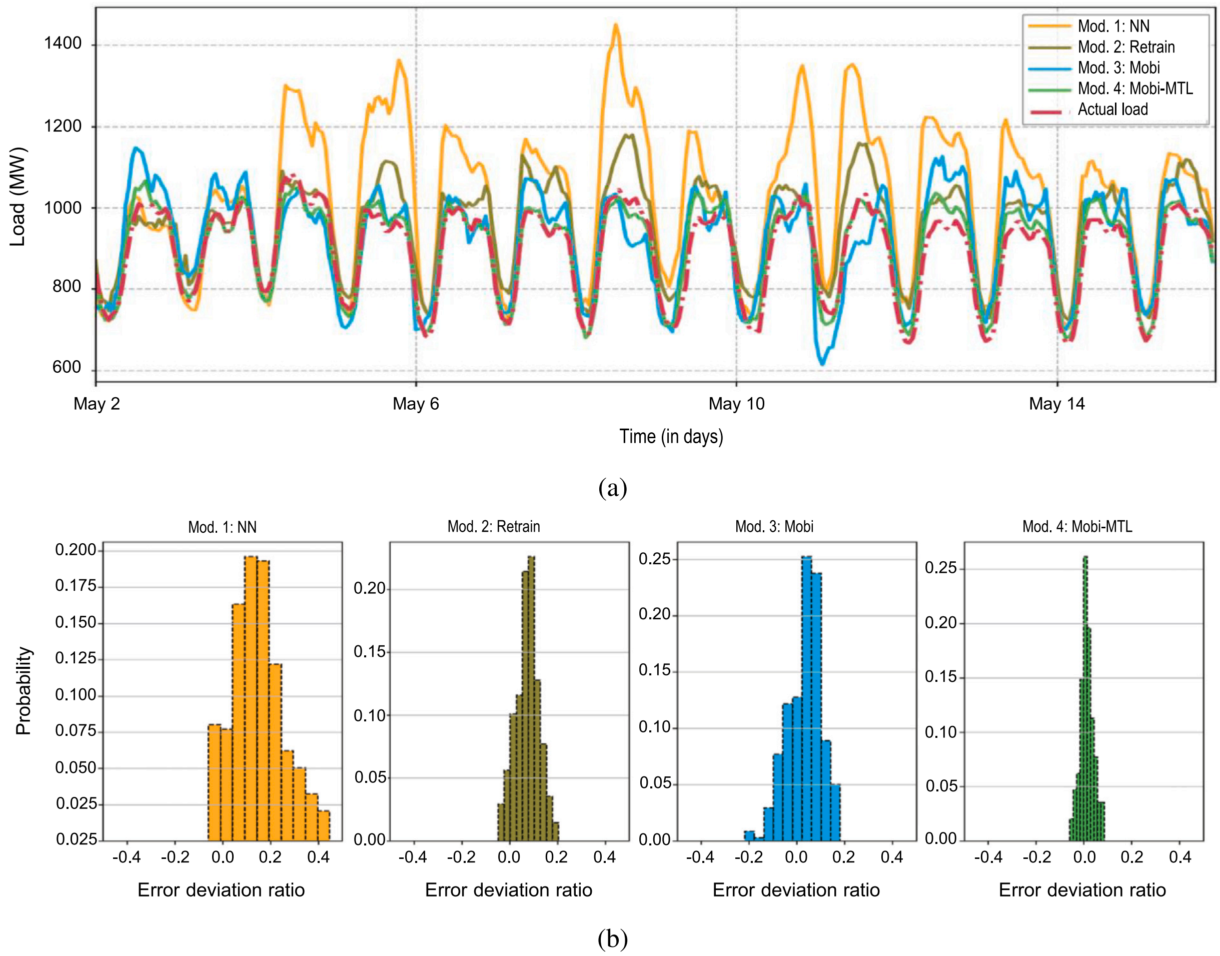


Fig. 12. Forecast performance of TL models validated on data collected during the COVID-19 pandemic using different model configurations: (a) forecast results, and (b) forecast error.

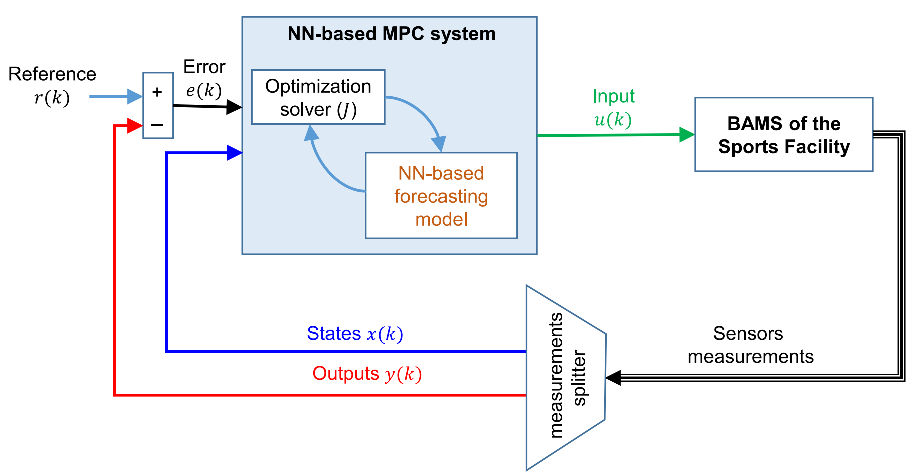
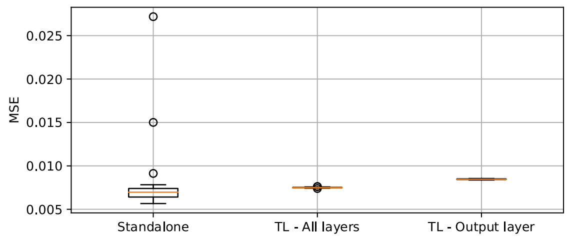
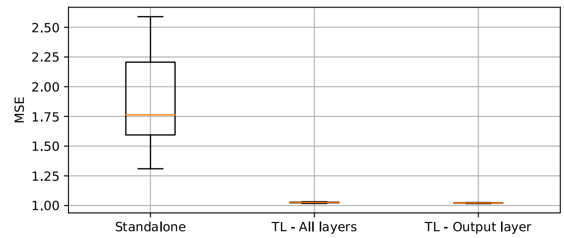


Fig. 13. The NN-based MPC system for sports facilities management and optimization. The MPC system consists of an optimizer and an NN-based forecasting model working collectively to decide the proper inputs $u(k)$. The BAMS's sensor measurements represents the states $x(k)$ which describe the system condition, and outputs $y(k)$ that are desired to reach certain reference values $r(k)$.

in Figs. 15 and 14(b), the standalone model evidently has poor performance with a median MSE on the test dataset is 1.75 due to the small size of the training dataset (2 days worth of data). However, even with the limited practical data for re-training the TL-based models, their test MSE is about 50% lower. Moreover, unlike the standalone model, the TL-based models have low variability since the models are developed based on a pretrained model with some determinism on the values of the weights. Nevertheless, the presented results are obtained when the training of the standalone model and TL-based models are conducted for 200 and 50 epochs, respectively. Hence, the computational training overhead is tremendously decreased while achieving favorable performance. In this case study, the application of



(a) Performance on the training dataset of the limited practical data.



(b) Performance on the test dataset of the limited practical data.

Fig. 14. Comparison between the standalone model and the two TL-based models of the NN-based forecasting model trained using the limited practical data. The box's upper and lower whiskers show the maximum and the minimum of each model's MSE for the 30 repeats, respectively and the box extends from the lower to the upper quartile values. The orange vertical line is the median and the dots represent the MSE outliers. The precision of the model is more when the whiskers are shorter, the box is narrower and overlapping with median line, and with minimal outliers.

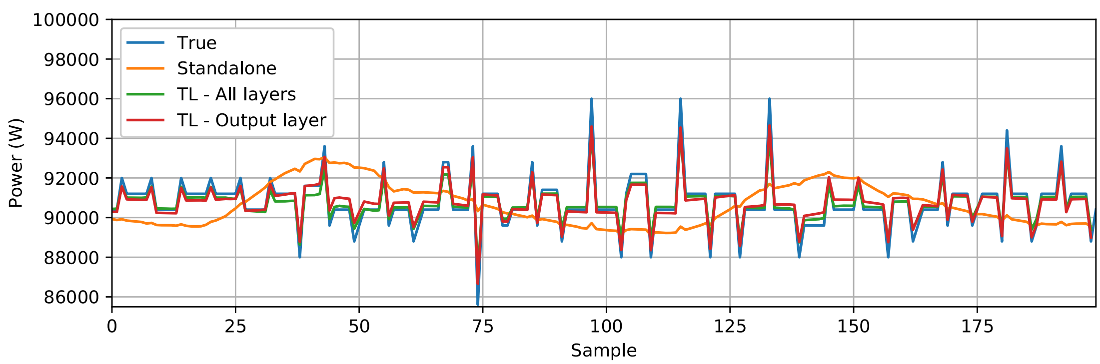


Fig. 15. Comparison between the consumption forecasting by the standalone model and the two TL-based models against the true power consumption on the training dataset. The blue plot depicts the actual data, the orange one represents the predicted consumption using the standalone model, and the green and red plots overlapping with the blue one are the outputs of the predicted consumption data using the TL models. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

TL is advantageous when the available historical data is insufficient to effectively train the NN-based model. TL demonstrates a remarkable capability in improving the performance of the forecasting model when the available data is scarce, given that the pretrained model used, which was trained on a dataset of a considerable size, is sufficiently familiarized with the type of features and/or tasks to be performed. By applying the TL, the pretrained model's parameters are tuned and adapted to the new data/problem.

7. Future directions

7.1. Further generalization

Although many TL-based techniques have excelled in transferring the knowledge of ML models from the source domains to different related target domains, still little information is available regarding what extent they can be generalized. To fill this gap, it is significant to carry out more investigations about the capability of TL not only with reference to the impact of spatial or geographical changes but also when different building environments (e.g., sports facilities, commercial buildings, office buildings, households, etc.), building construction materials, weather conditions (especially in larger countries or regions that experience higher temperature variations), occupants' behavior are considered simultaneously (Akhauri, Zheng, Goldstein, & Lin, 2021; Feng et al., 2021).

7.2. Real-time TL

In some applications of energy systems that require operate in real-time such as energy disaggregation, short-term load forecasting, and AFDD, online TL is more appropriate since data is available in sequential order and is utilized for updating the best predictor which predicts future operation data at every step (Grubinger, Chasparis, & Natschläger, 2016). This is by contrast to offline learning that generates the optimal predictor after being trained on the overall dataset at once (Wu et al., 2017). Moreover, online learning can be very helpful in learning computationally expensive energy tasks on edge, in which it is difficult, even infeasible, to train the ML models on the entire dataset while respecting the real-time condition (Mao, Ding, Tian, & Liang, 2020; Sayed, Himeur, Alsalemi, Bensaali and Amira, 2021).

Except (Bao et al., 2020; Grubinger et al., 2016, 2017), most of the included TL-based solutions for energy applications have concentrated on implementing offline TL, as we have been shown in this review. This could be justified because using TL in energy systems is still in its infancy and needs further investigation before being adopted in real-world scenarios. However, online learning using small datasets of real-time energy data is inevitable, especially for detecting faulty appliances and HVAC systems (Liao, Cai, Cheng, Dubey, & Rajesh, 2021), disaggregating energy data and identifying electrical devices (Krystalakos, Nalmpantis, & Vrakas, 2018), quantifying personalized thermal comfort (Ghahramani, Tang, & Becerik-Gerber, 2015), and detecting intrusions in smart grids (Yang, Zhai and Li, 2021). To that end, developing online TL techniques should be the target in the near future.

7.3. Federated TL

Developing efficient TL-based energy systems for load forecasting, energy disaggregation, thermal comfort control, AFDD, and many other tasks is inherently tied to the development of data (Alsalemi, Himeur, Bensaali, & Amira, 2021). In most energy systems, data may exist in the form of isolated domains (collected from different sources) with limited sharing perspectives between operators, consumers, and other stakeholders or between smart sensors and the central processing server

(CPS) that performs energy data analytics (Saputra et al., 2019). However, developing effective energy data analysis models that consider all the parameters impacting the energy systems requires frequent data sharing between the aforementioned nodes and contributors, thus raising privacy concerns and increasing the risks of attacks. Therefore, this hinders further development of TL and restricts its deployment (Fekri, Grolinger, & Mir, 2021; Ramu et al., 2022).

To alleviate those issues, federated learning has recently appeared as a potential solution that can manage the increasing data sharing demand without compromising users' privacy (Sharma, Park, & Cho, 2020). More particularly, federated TL has emerged as a variant of federated learning, enabling transferring knowledge among domains that do not have overlapping features and users. For instance, in Li, Ye, Song and Tse (2021), a distributed federated NILM system is introduced using federated learning and TL, which has achieved promising privacy preservation while maintaining good performance. Similarly, FedNILM, a practical federated learning-based energy disaggregation system on edge devices, is presented in Zhang et al. (2021). Typically, FedNILM has been developed to preserve users' privacy and provide energy disaggregation services to large-scale edge clients. This was possible by (i) personalizing edge model development using unsupervised TL, (ii) leveraging secure data aggregation through federated learning, and (iii) performing effective cloud model compression using multi-task learning and filter pruning. Moving on, the first framework adopting federated learning for energy prediction in residential buildings is introduced in Taïk and Cherkaoui (2020).

7.4. Transfer reinforcement learning

Reinforcement learning plays a significant role in various energy systems, especially for BAMSs, thermal comfort control, energy demand response monitoring, etc. Reinforcement learning excels at learning actions given a set of states without environment models and domain knowledge. This was doable by directly learning state–action value-functions or control policies. However, it is possible to wipe out the necessity of a domain expert if the environment model is learned based on the transitions experienced by the agent. Typically, reinforcement learning agents learn through interaction with the environment. Hence, the near-optimal policies can be learned based on previous experiences after several trials, while there is no need for prior knowledge.

However, many trials could be required to achieve optimal control performance. The learning agent faces a set of conditions and verifies the outcomes after carrying out various actions. This can result in high computational costs. To overcome this issue, TL can be integrated into reinforcement learning to reuse knowledge from another agent. In this regard, an agent that already has experience in a specific state can share that with another agent in a similar environment. This helps avoid the necessity to revisit the same state, as it has been demonstrated by Lissa et al. (2020), in which a considerable speed-up was achieved in the learning time on an HVAC control system. Similarly, in Lissa, Schukat, Keane, and Barrett (2021), TL is applied to a DRL-based heat pump monitoring system for better leveraging energy-saving in a microgrid. In this line, a DRL algorithm has been employed to monitor domestic hot water temperature and optimize PV self-consumption. Next, a TL has been applied to speed up the convergence process. Besides, in Fu et al. (2020), a building load forecasting scheme that combines DNN and TRL (DNN-TRL) is proposed. Specifically, a stack denoising autoencoder (DEA) is used for (i) extracting the deep characteristics of load forecasting and (ii) sharing the hidden layer structures for transferring the common information between various load forecasting problems. The output of the stack DEA network is utilized as the input of the Sarsa-based RL algorithm for improving the prediction performance of the building load forecasting. Moving on, a TRL-based rescheduling of differential power grids that considers the security challenges is presented in Wang and Tang (2022).

8. Conclusion

Cities are growing in size and resilience, but they require a more robust, smarter, and greener energy infrastructure to thrive fully. Saving energy is critical to this goal; however, implementing energy-saving programs that progressively shift away from fossil fuels and towards renewable energy resources is insufficient. In this regard, a transition to sustainable and renewable energy is pressing through developing the next-generation energy systems, especially by introducing the latest AI technologies. Training and testing AI models in practical applications face many problems because (i) data is gathered from distinct working environments or different energy systems (or devices), and the quantity of data is insufficient to train AI models, especially DL algorithms. Therefore, TL was proposed to overcome these issues by enhancing classification rates, avoiding overfitting, and improving energy systems' generalization ability.

This paper presented a comprehensive review of TL-based energy systems, which significantly impact developing the next-generation energy systems for smart city applications. To summarize, several achievements of TL have been reported for load forecasting, thermal comfort control, energy disaggregation, AFDD, smart grid and energy trading, etc. Although requiring a significant amount of training data, intelligent data-driven techniques for conducting the aforementioned energy tasks are attracting growing interest from academia and industry because of their performance superiority versus physics-based models and predicted mean vote models. TL has become a research hotspot in the energy sector since it helps in optimizing the performance of energy systems in different situations, e.g., when they cannot (i) provide sufficient data, (ii) generate labeled data, or (iii) meet the high computing resource requirements. We described in this framework the significant progress made by the TL community in energy systems via discussing existing frameworks, their learning methodologies, their pros, and cons. Moreover, the effectiveness of the TL models has been discussed under various application scenarios. Typically, we have highlighted the limitations of conventional ML algorithms and how TL can help improve their performance in energy systems. Also, in addition to crucial breakthrough works on this subject, many challenges still need to be overcome. Most of the principal challenges on TL, e.g., negative transfer, overfitting, measurement of transfer gains, unification of TL, and reproducibility of scientific results, have started to be explored. Furthermore, future directions that help better exploit TL, improve its generalizability, and expand its deployment in real-world energy applications have been identified and briefly described, including real-time TL, federated TL, and transfer reinforcement learning.

Overall, this review will be a comprehensive reference to guide the energy and smart city research communities in developing TL-based energy systems. This is mainly for applications that suffer from the unavailability of real-world datasets due to the difficulty of collecting multi-modal data, such as sports facilities.

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.

Data availability

Data will be made available on request.

Acknowledgments

This paper was made possible by National Priorities Research Program (NPRP) grant No. 12S-0222-190128 from the Qatar National Research Fund (a member of Qatar Foundation). The statements made herein are solely the responsibility of the authors. Open Access funding provided by the Qatar National Library.

References

Ahmed, S., & Bons, M. (2020). Edge computed NILM: a phone-based implementation using MobileNet compressed by tensorflow lite. In *Proceedings of the 5th international workshop on non-intrusive load monitoring* (pp. 44–48).

Ahmed, I., Jeon, G., Chehri, A., & Hassan, M. M. (2021). Adapting Gaussian YOLOv3 with transfer learning for overhead view human detection in smart cities and societies. *Sustainable Cities and Society*, *70*, Article 102908.

Ahmed, A. M., Zhang, Y., & Eliassen, F. (2020). 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)* (pp. 1–7). IEEE.

Ahn, Y., & Kim, B. S. (2022). Prediction of building power consumption using transfer learning-based reference building and simulation dataset. *Energy and Buildings*, *258*, Article 111717.

Akhauri, S., Zheng, L., Goldstein, T., & Lin, M. (2021). Improving generalization of transfer learning across domains using spatio-temporal features in autonomous driving. arXiv preprint arXiv:2103.08116.

Al Maruf, M., Singh, A., Azim, A., & Auluck, N. (2021). Faster fog computing based over-the-air vehicular updates: A transfer learning approach. *IEEE Transactions on Services Computing*, 1–10.

Alghamdi, A., Hammad, M., Ugail, H., Abdel-Raheem, A., Muhammad, K., Khalifa, H. S., et al. (2020). Detection of myocardial infarction based on novel deep transfer learning methods for urban healthcare in smart cities. *Multimedia Tools and Applications*, 1–22.

Alsalemi, A., Al-Kababji, A., Himeur, Y., Bensaali, F., & Amira, A. (2020). Cloud energy micro-moment data classification: a platform study. In *2020 IEEE/ACM 13th international conference on utility and cloud computing (UCC)* (pp. 420–425). IEEE.

Alsalemi, A., Himeur, Y., Bensaali, F., & Amira, A. (2021). Smart sensing and end-user behavioral change in residential buildings: An edge internet of energy perspective. *IEEE Sensors Journal*, *21*(24).

Alsalemi, A., Himeur, Y., Bensaali, F., Amira, A., Sardianos, C., Chronis, C., et al. (2020). A micro-moment system for domestic energy efficiency analysis. *IEEE Systems Journal*, *15*(1), 1256–1263.

Alsalemi, A., Himeur, Y., Bensaali, F., Amira, A., Sardianos, C., Varlamis, I., et al. (2020). Achieving domestic energy efficiency using micro-moments and intelligent recommendations. *IEEE Access*, *8*, 15047–15055.

Anjomshoaa, A., & Curry, E. (2022). Blockchain as an enabler for transfer learning in smart environments. arXiv preprint arXiv:2204.03959.

(2021). *Mobility trends reports*. Available online: https://covid19.apple.com/mobility. Accessed: 2021-11-05.

Aslam, S., Herodotou, H., Mohsin, S. M., Javaid, N., Ashraf, N., & Aslam, S. (2021). A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids. *Renewable and Sustainable Energy Reviews*, *144*, Article 110992.

Azkue, M., Lucu, M., Martinez-Laserna, E., & Aizpuru, I. (2021). Calendar ageing model for li-ion batteries using transfer learning methods. *World Electric Vehicle Journal*, *12*(3), 145.

Babar, M., Tariq, M. U., & Jan, M. A. (2020). Secure and resilient demand side management engine using machine learning for IoT-enabled smart grid. *Sustainable Cities and Society*, *62*, Article 102370.

Bao, Y., Mohammadpour Velni, J., & Shahbakhti, M. (2020). An online transfer learning approach for identification and predictive control design with application to RCCI engines. In *Dynamic systems and control conference, Vol. 84270*. American Society of Mechanical Engineers, V001T21A003.

Batra, N., Kukunuri, R., Pandey, A., Malakar, R., Kumar, R., Krystalakos, O., et al. (2019). A demonstration of reproducible state-of-the-art energy disaggregation using NILMTK. In *Proceedings of the 6th ACM international conference on systems for energy-efficient buildings, cities, and transportation* (pp. 358–359).

Batra, N., Kukunuri, R., Pandey, A., Malakar, R., Kumar, R., Krystalakos, O., et al. (2019). Towards reproducible state-of-the-art energy disaggregation. In *Proceedings of the 6th ACM international conference on systems for energy-efficient buildings, cities, and transportation* (pp. 193–202).

Brik, B., Esseghir, M., Merghem-Boulahia, L., & Snoussi, H. (2021). An IoT-based deep learning approach to analyse indoor thermal comfort of disabled people. *Building and Environment*, Article 108056.

Cai, L., Gu, J., Ma, J., & Jin, Z. (2019). Probabilistic wind power forecasting approach via instance-based transfer learning embedded gradient boosting decision trees. *Energies*, *12*(1), 159.

Cao, H., Liu, S., Wu, L., Guan, Z., & Du, X. (2019). Achieving differential privacy against non-intrusive load monitoring in smart grid: A fog computing approach. *Concurrency Computations: Practice and Experience*, *31*(22), Article e4528.

Carrera, B., Peyrard, S., & Kim, K. (2021). Meta-regression framework for energy consumption prediction in a smart city: A case study of Songdo in South Korea. *Sustainable Cities and Society*, *72*, Article 103025.

Cavalca, D. L., & Fernandes, R. A. (2021). Deep transfer learning-based feature extraction: An approach to improve nonintrusive load monitoring. *IEEE Access*.

Chang, F.-Y., Chen, C., & Lin, S.-D. (2018). An empirical study of ladder network and multitask learning on energy disaggregation in Taiwan. In *2018 conference on technologies and applications of artificial intelligence (TAAI)* (pp. 86–89). IEEE.

Che, Y., Deng, Z., Lin, X., Hu, L., & Hu, X. (2021). Predictive battery health management with transfer learning and online model correction. *IEEE Transactions on Vehicular Technology*, *70*(2), 1269–1277.

Chen, W., Qiu, Y., Feng, Y., Li, Y., & Kusiak, A. (2021). Diagnosis of wind turbine faults with transfer learning algorithms. *Renewable Energy*, *163*, 2053–2067.

Chen, J., & Ran, X. (2019). Deep learning with edge computing: A review. *Proceedings of the IEEE*, *107*(8), 1655–1674.

Chen, Y., Tong, Z., Zheng, Y., Samuelson, H., & Norford, L. (2020). Transfer learning with deep neural networks for model predictive control of HVAC and natural ventilation in smart buildings. *Journal of Cleaner Production*, *254*, Article 119866.

Chen, X., Wang, S., Fu, B., Long, M., & Wang, J. (2019). Catastrophic forgetting meets negative transfer: Batch spectral shrinkage for safe transfer learning. *Advances in Neural Information Processing Systems*, *32*.

Chen, Q., Zheng, Z., Hu, C., Wang, D., & Liu, F. (2019). Data-driven task allocation for multi-task transfer learning on the edge. In *2019 IEEE 39th international conference on distributed computing systems (ICDCS)* (pp. 1040–1050). IEEE.

Chen, Y., Zheng, Y., & Samuelson, H. (2020). 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)* (pp. 2345–2350). IEEE.

Cheng, D., Sha, W., Wang, L., Tang, S., Ma, A., Chen, Y., et al. (2021). Solid-state lithium battery cycle life prediction using machine learning. *Applied Sciences*, *11*(10), 4671.

Cheng, T., Zhu, X., Gu, X., Yang, F., & Mohammadi, M. (2021). Stochastic energy management and scheduling of microgrids in correlated environment: A deep learning-oriented approach. *Sustainable Cities and Society*, *69*, Article 102856.

Cook, D., Feuz, K. D., & Krishnan, N. C. (2013). Transfer learning for activity recognition: A survey. *Knowledge and Information Systems*, *36*(3), 537–556.

Dai, W., Yang, Q., Xue, G.-R., & Yu, Y. (2007). Boosting for transfer learning. In *ICML '07, Proceedings of the 24th international conference on machine learning* (pp. 193–200). New York, NY, USA: Association for Computing Machinery.

Das, L., Garg, D., & Srinivasan, B. (2020). NeuralCompression: A machine learning approach to compress high frequency measurements in smart grid. *Applied Energy*, *257*, Article 113966.

Das, B., Nair, B., Reddy, V. K., & Venkatesh, P. (2018). Evaluation of multiple linear, neural network and penalised regression models for prediction of rice yield based on weather parameters for west coast of India. *International Journal of Biometeorology*, *62*(10), 1809–1822.

Delfosse, A., Hebrail, G., & Zerroug, A. (2020). Deep learning applied to NILM: is data augmentation worth for energy disaggregation? In *ECAI 2020* (pp. 2972–2977). IOS Press.

Demianenko, M., & De Gaetani, C. I. (2021). A procedure for automating energy analyses in the BIM context exploiting artificial neural networks and transfer learning technique. *Energies*, *14*(10), 2956.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248–255). Ieee.

Dey, S., Mukherjee, A., Pal, A., & Balamuralidhar, P. (2018). Partitioning of cnn models for execution on fog devices. In *Proceedings of the 1st ACM international workshop on smart cities and fog computing* (pp. 19–24).

Di, Z., Shao, H., & Xiang, J. (2021). Ensemble deep transfer learning driven by multisensor signals for the fault diagnosis of bevel-gear cross-operation conditions. *Science China Technological Sciences*, *64*(3), 481–492.

D'Incecco, M., Squartini, S., & Zhong, M. (2019). Transfer learning for non-intrusive load monitoring. *IEEE Transactions on Smart Grid*, *11*(2), 1419–1429.

Ding, S., Zeng, J., Hu, Z., & Yang, Y. (2022). IOT-based social-economic management of distribution system with the high penetration of renewable energy sources. *Sustainable Cities and Society*, *76*, Article 103439.

Dowling, C. P., & Zhang, B. (2020). Transfer learning for HVAC system fault detection. In *2020 American control conference (ACC)* (pp. 3879–3885). IEEE.

Dridi, J., Amayri, M., & Bouguila, N. (2022). Transfer learning for estimating occupancy and recognizing activities in smart buildings. *Building and Environment*, *217*, Article 109057.

Elnour, M., Fadli, F., Himeur, Y., Petri, I., Rezgui, Y., Meskin, N., et al. (2022). Performance and energy optimization of building automation and management systems: Towards smart sustainable carbon-neutral sports facilities. *Renewable and Sustainable Energy Reviews*, *162*, Article 112401.

Elnour, M., Himeur, Y., Fadli, F., Mohammedsherif, H., Meskin, N., Ahmad, A. M., et al. (2022). Neural network-based model predictive control system for optimizing building automation and management systems of sports facilities. *Applied Energy*, *318*, Article 119153.

Elnour, M., Mohammedsherif, H., Fadli, F., Meskin, N., Ahmad, A. M., Rezgui, Y., et al. (2021). Neural network-based predictive control system for energy optimization in sports facilities: A case study. In *Proceedings of the 38th international conference of CIB, Vol. 2021* (pp. 734–743).

Fadli, F., Rezgui, Y., Petri, I., Ahmad, A. M., Hodorog, A., Elnour, M., et al. (2021). Building energy management systems for sports facilities in the gulf region: A focus on impacts and considerations. In *Proceedings of the 38th international conference of CIB W78* (pp. 3–15).

Fan, C., Lei, Y., Sun, Y., Piscitelli, M. S., Chiosa, R., & Capozzoli, A. (2022). Data-centric or algorithm-centric: Exploiting the performance of transfer learning for improving building energy predictions in data-scarce context. *Energy*, *240*, Article 122775.

Fan, C., Sun, Y., Xiao, F., Ma, J., Lee, D., Wang, J., et al. (2020). Statistical investigations of transfer learning-based methodology for short-term building energy predictions. *Applied Energy*, *262*, Article 114499.

Fan, C., Sun, Y., Zhao, Y., Song, M., & Wang, J. (2019). Deep learning-based feature engineering methods for improved building energy prediction. *Applied Energy*, *240*, 35–45.

Fang, X., Gong, G., Li, G., Chun, L., Li, W., & Peng, P. (2021). A hybrid deep transfer learning strategy for short term cross-building energy prediction. *Energy*, *215*, Article 119208.

Fang, X., Gong, G., Li, G., Chun, L., Peng, P., & Li, W. (2021). A general multi-source ensemble transfer learning framework integrate of LSTM-DANN and similarity metric for building energy prediction. *Energy and Buildings*, *252*, Article 111435.

Farooq, U., Shabir, M. W., Javed, M. A., & Imran, M. (2021). Intelligent energy prediction techniques for fog computing networks. *Applied Soft Computing*, *111*, Article 107682.

Fekri, M. N., Grolinger, K., & Mir, S. (2021). Distributed load forecasting using smart meter data: Federated learning with recurrent neural networks. *International Journal of Electrical Power & Energy Systems*, Article 107669.

Feng, S., Chen, J., Ye, Y., Wu, X., Cui, H., Tang, Y., et al. (2022). A two-stage deep transfer learning for localisation of forced oscillations disturbance source. *International Journal of Electrical Power & Energy Systems*, *135*, Article 107577.

Feng, S., Fu, H., Zhou, H., Wu, Y., Lu, Z., & Dong, H. (2021). A general and transferable deep learning framework for predicting phase formation in materials. *Npj Computational Materials*, *7*(1), 1–10.

Fu, Q., Liu, Q., Gao, Z., Wu, H., Fu, B., & Chen, J. (2020). A building energy consumption prediction method based on integration of a deep neural network and transfer reinforcement learning. *International Journal of Pattern Recognition and Artificial Intelligence*, *34*(10), Article 2052005.

Gao, F., Li, Q., Ji, Y., Ji, S., Guo, J., Sun, H., et al. (2021). EWNet: An early warning classification framework for smart grid based on local-to-global perception. *Neurocomputing*, *443*, 199–212.

Gao, G., Li, J., & Wen, Y. (2020). DeepComfort: Energy-efficient thermal comfort control in buildings via reinforcement learning. *IEEE Internet of Things Journal*, *7*(9), 8472–8484.

Gao, Y., Ruan, Y., Fang, C., & Yin, S. (2020). Deep learning and transfer learning models of energy consumption forecasting for a building with poor information data. *Energy and Buildings*, *223*, Article 110156.

Gao, N., Shao, W., Rahaman, M. S., Zhai, J., David, K., & Salim, F. D. (2021). Transfer learning for thermal comfort prediction in multiple cities. *Building and Environment*, *195*, Article 107725.

Ghahramani, A., Tang, C., & Becerik-Gerber, B. (2015). An online learning approach for quantifying personalized thermal comfort via adaptive stochastic modeling. *Building and Environment*, *92*, 86–96.

Glorot, X., Bordes, A., & Bengio, Y. (2011). Domain adaptation for large-scale sentiment classification: A deep learning approach. In *ICML*.

Gong, C., Lin, F., Gong, X., & Lu, Y. (2020). Intelligent cooperative edge computing in internet of things. *IEEE Internet of Things Journal*, *7*(10), 9372–9382.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., et al. (2020). Generative adversarial networks. *Communications of the ACM*, *63*(11), 139–144.

(2021). *Google COVID-19 mobility report*. Available online: https://www.google.com/covid19/mobility/. Accessed: 2021-11-05.

Grubinger, T., Chasparis, G. C., & Natschläger, T. (2016). Online transfer learning for climate control in residential buildings. In *2016 European control conference (ECC)* (pp. 1183–1188). IEEE.

Grubinger, T., Chasparis, G. C., & Natschläger, T. (2017). Generalized online transfer learning for climate control in residential buildings. *Energy and Buildings*, *139*, 63–71.

Guariso, G., Nunnari, G., & Sangiorgio, M. (2020). Multi-step solar irradiance forecasting and domain adaptation of deep neural networks. *Energies*, *13*(15), 3987.

Gui, L., Xu, R., Lu, Q., Du, J., & Zhou, Y. (2018). Negative transfer detection in transductive transfer learning. *International Journal of Machine Learning and Cybernetics*, *9*(2), 185–197.

Gunduz, S., Ugurlu, U., & Oksuz, I. (2020). Transfer learning for electricity price forecasting. arXiv preprint arXiv:2007.03762.

Han, J., Miao, S., Yin, H., Guo, S., Wang, Z., Yao, F., et al. (2021). Deep-adversarial-transfer learning based fault classification of power lines in smart grid. In *IOP conference series: Earth and environmental science, Vol. 701*. IOP Publishing, Article 012074.

He, X., Liu, Q., Tang, J., Ma, Q., Wang, G., & Liu, M. (2020). Power system frequency situation prediction method based on transfer learning. In *2020 12th IEEE PES Asia-Pacific power and energy engineering conference (APPEEC)* (pp. 1–5). http://dx.doi.org/10.1109/APPEEC48164.2020.9220684.

Himeur, Y., Alsalemi, A., Al-Kababji, A., Bensaali, F., & Amira, A. (2020). Data fusion strategies for energy efficiency in buildings: Overview, challenges and novel orientations. *Information Fusion*, *64*, 99–120.

Himeur, Y., Alsalemi, A., Al-Kababji, A., Bensaali, F., Amira, A., Sardianos, C., et al. (2021). A survey of recommender systems for energy efficiency in buildings: Principles, challenges and prospects. *Information Fusion*, *72*, 1–21.

Himeur, Y., Alsalemi, A., Bensaali, F., & Amira, A. (2020a). A novel approach for detecting anomalous energy consumption based on micro-moments and deep neural networks. *Cognitive Computation*, *12*(6), 1381–1401.

Himeur, Y., Alsalemi, A., Bensaali, F., & Amira, A. (2020b). Building power consumption datasets: Survey, taxonomy and future directions. *Energy and Buildings*, *227*, Article 110404.

Himeur, Y., Alsalemi, A., Bensaali, F., & Amira, A. (2020c). Robust event-based non-intrusive appliance recognition using multi-scale wavelet packet tree and ensemble bagging tree. *Applied Energy*, *267*, Article 114877.

Himeur, Y., Alsalemi, A., Bensaali, F., & Amira, A. (2021). Smart non-intrusive appliance identification using a novel local power histogramming descriptor with an improved k-nearest neighbors classifier. *Sustainable Cities and Society*, *67*, Article 102764.

Himeur, Y., Alsalemi, A., Bensaali, F., Amira, A., Varlamis, I., Bravos, G., et al. (2022). Techno-economic assessment of building energy efficiency systems using behavioral change: A case study of an edge-based micro-moments solution. *Journal of Cleaner Production*, *331*, Article 129786.

Himeur, y., Elnour, M., Fadli, F., Meskin, N., Petri, I., Rezgui, Y., et al. (2022). AI-big data analytics for building automation and management systems: A survey, actual challenges and future perspectives. *Artificial Intelligence Review*.

Himeur, Y., Ghanem, K., Alsalemi, A., Bensaali, F., & Amira, A. (2021). Artificial intelligence based anomaly detection of energy consumption in buildings: A review, current trends and new perspectives. *Applied Energy*, *287*, Article 116601.

Hooshmand, A., & Sharma, R. (2019). Energy predictive models with limited data using transfer learning. In *Proceedings of the tenth ACM international conference on future energy systems*,

Houidi, S., Fourer, D., Auger, F., Sethom, H. B. A., & Miègeville, L. (2021). Comparative evaluation of non-intrusive load monitoring methods using relevant features and transfer learning. *Energies*, *14*(9), 2726.

Hu, W., Luo, Y., Lu, Z., & Wen, Y. (2019). Heterogeneous transfer learning for thermal comfort modeling. In *Proceedings of the 6th ACM international conference on systems for energy-efficient buildings, cities, and transportation* (pp. 61–70).

Hu, C., Zhang, J., Yuan, H., Gao, T., Jiang, H., Yan, J., et al. (2022). Black swan event small-sample transfer learning (BEST-L) and its case study on electrical power prediction in COVID-19. *Applied Energy*, *309*, Article 118458.

Huang, J.-F., Zhang, G.-H., & Hsieh, S.-Y. (2021). Real-time energy data compression strategy for reducing data traffic based on smart grid AMI networks. *The Journal of Supercomputing*, 1–20.

Huang, W., Zhang, X., & Zheng, W. (2021). Resilient power network structure for stable operation of energy systems: A transfer learning approach. *Applied Energy*, *296*, Article 117065.

Humala, B., Nambi, A. S. U., & Prasad, V. R. (2018). UniversalNILM: A semi-supervised energy disaggregation framework using general appliance models. In *Proceedings of the ninth international conference on future energy systems* (pp. 223–229).

Jain, M., Gupta, K., Sathanur, A., Chandan, V., & Halappanavar, M. M. (2021). Transfer-learnt models for predicting electricity consumption in buildings with limited and sparse field data. In *2021 American control conference (ACC)* (pp. 2887–2894). IEEE.

Jenssen, R., Roverso, D., et al. (2018). Automatic autonomous vision-based power line inspection: A review of current status and the potential role of deep learning. *International Journal of Electrical Power & Energy Systems*, *99*, 107–120.

Jha, D., Choudhary, K., Tavazza, F., Liao, W.-k., Choudhary, A., Campbell, C., et al. (2019). Enhancing materials property prediction by leveraging computational and experimental data using deep transfer learning. *Nature Communications*, *10*(1), 1–12.

Jiang, Z., & Lee, Y. M. (2019). Deep transfer learning for thermal dynamics modeling in smart buildings. In *2019 IEEE international conference on big data (Big data)* (pp. 2033–2037). IEEE.

Jin, Y., Acquah, M. A., Seo, M., & Han, S. (2022). Short-term electric load prediction using transfer learning with interval estimate adjustment. *Energy and Buildings*, Article 111846.

Jung, S.-M., Park, S., Jung, S.-W., & Hwang, E. (2020). Monthly electric load forecasting using transfer learning for smart cities. *Sustainability*, *12*(16), 6364.

Kang, Z., Yang, B., Yang, S., Fang, X., & Zhao, C. (2020). Online transfer learning with multiple source domains for multi-class classification. *Knowledge-Based Systems*, *190*, Article 105149.

Kathirgamanathan, A., De Rosa, M., Mangina, E., & Finn, D. P. (2021). Data-driven predictive control for unlocking building energy flexibility: A review. *Renewable and Sustainable Energy Reviews*, *135*, Article 110120.

Katranji, M., Thuillier, E., Kraiem, S., Moalic, L., & Selem, F. H. (2016). Mobility data disaggregation: A transfer learning approach. In *2016 IEEE 19th international conference on intelligent transportation systems (ITSC)* (pp. 1672–1677). IEEE.

Kelati, A., Dhaou, I. B., Kondoro, A., Rwegasira, D., & Tenhunen, H. (2019). Iot based appliances identification techniques with fog computing for e-health. In *2019 IST-Africa week conference (IST-Africa)* (pp. 1–11). IEEE.

Kelly, J., & Knottenbelti, W. (2015). The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes. *Scientific Data*, *2*(150007), 1–14.

Khan, N., Haq, I. U., Khan, S. U., Rho, S., Lee, M. Y., & Baik, S. W. (2021). DB-Net: A novel dilated CNN based multi-step forecasting model for power consumption in integrated local energy systems. *International Journal of Electrical Power & Energy Systems*, *133*, Article 107023.

Khan, T., Tian, W., Ilager, S., & Buyya, R. (2021). Workload forecasting and energy state estimation in cloud data centers: ML-centric approach. *Future Generation Computer Systems*.

Kim, S., Choi, Y. Y., Kim, K. J., & Choi, J.-I. (2021). Forecasting state-of-health of lithium-ion batteries using variational long short-term memory with transfer learning. *Journal of Energy Storage*, *41*, Article 102893.

Ko, Y.-D., & Park, C.-S. (2021). Parameter estimation of unknown properties using transfer learning from virtual to existing buildings. *Journal of Building Performance Simulation*, *14*(5), 503–514.

Kolter, J. Z., & Johnson, M. J. (2011). REDD: A public data set for energy disaggregation research. In *Workshop on data mining applications in sustainability (SIGKDD), San Diego, CA, Vol. 25* (pp. 59–62).

Krystalakos, O., Nalmpantis, C., & Vrakas, D. (2018). Sliding window approach for online energy disaggregation using artificial neural networks. In *Proceedings of the 10th Hellenic conference on artificial intelligence* (pp. 1–6).

Kumari, P., & Toshniwal, D. (2021). Deep learning models for solar irradiance forecasting: A comprehensive review. *Journal of Cleaner Production*, Article 128566.

Le, T., Vo, M. T., Kieu, T., Hwang, E., Rho, S., & Baik, S. W. (2020). Multiple electric energy consumption forecasting using a cluster-based strategy for transfer learning in smart building. *Sensors*, *20*(9), 2668.

Lee, C., Kim, S.-H., & Youn, C.-H. (2020). Cooperating edge cloud-based hybrid online learning for accelerated energy data stream processing in load forecasting. *IEEE Access*, *8*, 199120–199132.

Leroy, P. A., & Goedemé, T. (2021). Optimal distribution of CNN computations on edge and cloud.. In *ICAART (2)* (pp. 604–613).

Li, L., Fu, Y., Fung, J. C., Qu, H., & Lau, A. K. (2021). Development of a back-propagation neural network and adaptive grey wolf optimizer algorithm for thermal comfort and energy consumption prediction and optimization. *Energy and Buildings*, *253*, Article 111439.

Li, W., Gu, S., Zhang, X., & Chen, T. (2020). Transfer learning for process fault diagnosis: Knowledge transfer from simulation to physical processes. *Computers & Chemical Engineering*, *139*, Article 106904.

Li, W., Huang, R., Li, J., Liao, Y., Chen, Z., He, G., et al. (2022). A perspective survey on deep transfer learning for fault diagnosis in industrial scenarios: Theories, applications and challenges. *Mechanical Systems and Signal Processing*, *167*, Article 108487.

Li, Y., Jiang, W., Zhang, G., & Shu, L. (2021). Wind turbine fault diagnosis based on transfer learning and convolutional autoencoder with small-scale data. *Renewable Energy*, *171*, 103–115.

Li, Y., Li, K., Liu, X., Wang, Y., & Zhang, L. (2021). Lithium-ion battery capacity estimation—A pruned convolutional neural network approach assisted with transfer learning. *Applied Energy*, *285*, Article 116410.

Li, J., Wu, D., Attar, H. M., & Xu, H. (2022). Geometric neuro-fuzzy transfer learning for in-cylinder pressure modelling of a diesel engine fuelled with raw microalgae oil. *Applied Energy*, *306*, Article 118014.

Li, A., Xiao, F., Fan, C., & Hu, M. (2021). Development of an ANN-based building energy model for information-poor buildings using transfer learning. In *Building simulation, Vol. 14* (pp. 89–101). Springer.

Li, Q., Ye, J., Song, W., & Tse, Z. (2021). Energy disaggregation with federated and transfer learning. In *IEEE 7th world forum on internet of things* (pp. 1–6).

Li, J., Zhao, Y., Lee, Y.-h., & Kim, S.-J. (2019). Learning to infer voltage stability margin using transfer learning. In *2019 IEEE data science workshop (DSW)* (pp. 270–274). IEEE.

Liang, P., Yang, H.-D., Chen, W.-S., Xiao, S.-Y., & Lan, Z.-Z. (2018). Transfer learning for aluminium extrusion electricity consumption anomaly detection via deep neural networks. *International Journal of Computer Integrated Manufacturing*, *31*(4–5), 396–405.

Liao, H., Cai, W., Cheng, F., Dubey, S., & Rajesh, P. B. (2021). An online data-driven fault diagnosis method for air handling units by rule and convolutional neural networks. *Sensors*, *21*(13), 4358.

Lin, J., Ma, J., Zhu, J., & Liang, H. (2021). Deep domain adaptation for non-intrusive load monitoring based on a knowledge transfer learning network. *IEEE Transactions on Smart Grid*.

Lissa, P., Schukat, M., & Barrett, E. (2020). Transfer learning applied to reinforcement learning-based hvac control. *SN Computer Science*, *1*(3), 1–12.

Lissa, P., Schukat, M., Keane, M., & Barrett, E. (2021). Transfer learning applied to DRL-based heat pump control to leverage microgrid energy efficiency. *Smart Energy*, *3*, Article 100044.

Liu, Y., Shu, X., Yu, H., Shen, J., Zhang, Y., Liu, Y., et al. (2021). State of charge prediction framework for lithium-ion batteries incorporating long short-term memory network and transfer learning. *Journal of Energy Storage*, *37*, Article 102494.

Liu, Y., Wang, X., & You, W. (2019). Non-intrusive load monitoring by voltage–current trajectory enabled transfer learning. *IEEE Transactions on Smart Grid*, *10*(5), 5609–5619. http://dx.doi.org/10.1109/TSG.2018.2888581.

Liu, X., Yu, W., Liang, F., Griffith, D., & Golmie, N. (2021). Towards deep transfer learning in industrial internet of things. *IEEE Internet of Things Journal*.

Liu, J., Zhang, Q., Li, X., Li, G., Liu, Z., Xie, Y., et al. (2021). Transfer learning-based strategies for fault diagnosis in building energy systems. *Energy and Buildings*, *250*, Article 111256.

Lu, J., Behbood, V., Hao, P., Zuo, H., Xue, S., & Zhang, G. (2015). Transfer learning using computational intelligence: A survey. *Knowledge-Based Systems*, *80*, 14–23.

Lu, Y., Tian, Z., Zhou, R., & Liu, W. (2021). A general transfer learning-based framework for thermal load prediction in regional energy system. *Energy*, *217*, Article 119322.

Lu, F., Zhou, G., Liu, Y., & Zhang, C. (2022). Ensemble transfer learning for cutting energy consumption prediction of aviation parts towards green manufacturing. *Journal of Cleaner Production*, *331*, Article 129920.

Lv, Z., Lou, R., Kumar Singh, A., & Wang, Q. (2021). Transfer learning-powered resource optimization for green computing in 5G-aided industrial internet of things. *ACM Transactions on Internet Technology (TOIT)*, *22*(2), 1–16.

Ma, J., Cheng, J. C., Jiang, F., Chen, W., Wang, M., & Zhai, C. (2020). A bi-directional missing data imputation scheme based on LSTM and transfer learning for building energy data. *Energy and Buildings*, *216*, Article 109941.

Maghdid, H. S., Ghafoor, K. Z., Al-Talabani, A., Sadiq, A. S., Singh, P. K., & Rawat, D. B. (2022). Enabling accurate indoor localization for different platforms for smart cities using a transfer learning algorithm. *Internet Technology Letters*, *5*(1), Article e200.

Mahmoudi, S. A., Belarbi, M. A., Mahmoudi, S., Belalem, G., & Manneback, P. (2020). Multimedia processing using deep learning technologies, high-performance computing cloud resources, and big data volumes. *Concurrency Computations: Practice and Experience*, *32*(17), Article e5699.

Mao, W., Ding, L., Tian, S., & Liang, X. (2020). Online detection for bearing incipient fault based on deep transfer learning. *Measurement*, *152*, Article 107278.

Más, H. F., & Kuiken, D. (2020). Beyond energy savings: The necessity of optimising smart electricity systems with resource efficiency and coherent waste policy in Europe. *Energy Research & Social Science*, *70*, Article 101658.

Michau, G., & Fink, O. (2021). Unsupervised transfer learning for anomaly detection: Application to complementary operating condition transfer. *Knowledge-Based Systems*, *216*, Article 106816.

Miller, C., & Meggers, F. (2017). The building data genome project: An open, public data set from non-residential building electrical meters. *Energy Procedia*, *122*, 439–444.

Mocanu, E., Nguyen, P. H., Kling, W. L., & Gibescu, M. (2016). Unsupervised energy prediction in a smart grid context using reinforcement cross-building transfer learning. *Energy and Buildings*, *116*, 646–655.

Mohammadi, M., Al-Fuqaha, A., Guizani, M., & Oh, J.-S. (2017). Semisupervised deep reinforcement learning in support of IoT and smart city services. *IEEE Internet of Things Journal*, *5*(2), 624–635.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Prisma Group (2009). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine*, *6*(7), Article e1000097.

Moon, J., Kim, J., Kang, P., & Hwang, E. (2020). Solving the cold-start problem in short-term load forecasting using tree-based methods. *Energies*, *13*(4), 886.

Mosaico, G., Saviozzi, M., Silvestro, F., Bagnasco, A., & Vinci, A. (2019). Simplified state space building energy model and transfer learning based occupancy estimation for HVAC optimal control. In *2019 IEEE 5th international forum on research and technology for society and industry (RTSI)* (pp. 353–358). IEEE.

Murray, D., Liao, J., Stankovic, L., Stankovic, V., Hauxwell-Baldwin, R., Wilson, C., et al. (2015). A data management platform for personalised real-time energy feedback. In *Procedings of the 8th international conference on energy efficiency in domestic appliances and lighting*.

Murray, D., Stankovic, L., & Stankovic, V. (2017). An electrical load measurements dataset of United Kingdom households from a two-year longitudinal study. *Scientific Data*, *4*, Article 160122. http://dx.doi.org/10.1038/sdata.2016.122.

Murray, D., Stankovic, L., Stankovic, V., Lulic, S., & Sladojevic, S. (2019). Transferability of neural network approaches for low-rate energy disaggregation. In *ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)* (pp. 8330–8334). IEEE.

Natarajan, A., & Laftchiev, E. (2019). A transfer active learning framework to predict thermal comfort. *International Journal of Prognostics and Health Management*, *10*(3).

Ngarambe, J., Yun, G. Y., & Santamouris, M. (2020). The use of artificial intelligence (AI) methods in the prediction of thermal comfort in buildings: energy implications of AI-based thermal comfort controls. *Energy and Buildings*, *211*, Article 109807.

Niu, H., Nguyen, D., Yonekawa, K., Kurokawa, M., Wada, S., & Yoshihara, K. (2020). Multi-source transfer learning for human activity recognition in smart homes. In *2020 IEEE international conference on smart computing (SMARTCOMP)* (pp. 274–277). IEEE.

Pardamean, B., Muljo, H. H., Cenggoro, T. W., Chandra, B. J., & Rahutomo, R. (2019). Using transfer learning for smart building management system. *Journal of Big Data*, *6*(1), 1–12.

Park, H., & Park, D. Y. (2021). Prediction of individual thermal comfort based on ensemble transfer learning method using wearable and environmental sensors. *Building and Environment*, Article 108492.

Patricia, N., & Caputo, B. (2014). Learning to learn, from transfer learning to domain adaptation: A unifying perspective. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1442–1449).

Pérez, S., Arroba, P., & Moya, J. M. (2021). Energy-conscious optimization of edge computing through deep reinforcement learning and two-phase immersion cooling. *Future Generation Computer Systems*, *125*, 891–907.

Petri, I., Rana, O., Rezgui, Y., & Fadli, F. (2021). Edge HVAC analytics. *Energies*, *14*(17), 5464. http://dx.doi.org/10.3390/en14175464.

Pinto, G., Wang, Z., Roy, A., Hong, T., & Capozzoli, A. (2022). Transfer learning for smart buildings: A critical review of algorithms, applications, and future perspectives. *Advances in Applied Energy*, Article 100084.

Pokhrel, S. R., Pan, L., Kumar, N., Doss, R., & Vu, H. L. (2021). Multipath TCP meets transfer learning: A novel edge-based learning for industrial IoT. *IEEE Internet of Things Journal*, *8*(13), 10299–10307. http://dx.doi.org/10.1109/JIOT.2021.3056466.

Prakash, N., Murugappan, M., Hemalakshmi, G., Jayalakshmi, M., & Mahmud, M. (2021). Deep transfer learning for COVID-19 detection and infection localization with superpixel based segmentation. *Sustainable Cities and Society*, *75*, Article 103252.

Prechelt, L. (1998). Early stopping - but when? In G. B. Orr, & K.-R. Müller (Eds.), *Neural networks: Tricks of the trade* (pp. 55–69). Berlin, Heidelberg: Springer Berlin Heidelberg, http://dx.doi.org/10.1007/3-540-49430-8_3.

Qian, F., Gao, W., Yang, Y., Yu, D., et al. (2020). Potential analysis of the transfer learning model in short and medium-term forecasting of building HVAC energy consumption. *Energy*, *193*(C).

Qin, Y., Ke, J., Wang, B., & Fedorovich, F. G. (2021). Energy optimization for regional buildings based on distributed reinforcement learning. *Sustainable Cities and Society*, Article 103625.

Qureshi, A. S., Khan, A., Zameer, A., & Usman, A. (2017). Wind power prediction using deep neural network based meta regression and transfer learning. *Applied Soft Computing*, *58*, 742–755.

Ramirez, P. Z., Tonioni, A., Salti, S., & Stefano, L. D. (2019). Learning across tasks and domains. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 8110–8119).

Ramu, S. P., Boopalan, P., Pham, Q.-V., Maddikunta, P. K. R., The, T.-H., Alazab, M., et al. (2022). Federated learning enabled digital twins for smart cities: Concepts, recent advances, and future directions. *Sustainable Cities and Society*, Article 103663.

Ren, H., Liu, W., Shan, M., & Wang, X. (2019). A new wind turbine health condition monitoring method based on VMD-MPE and feature-based transfer learning. *Measurement*, *148*, Article 106906.

Ribani, R., & Marengoni, M. (2019). A survey of transfer learning for convolutional neural networks. In *2019 32nd SIBGRAPI conference on graphics, patterns and images tutorials (SIBGRAPI-T)* (pp. 47–57). IEEE.

Ribeiro, M., Grolinger, K., ElYamany, H. F., Higashino, W. A., & Capretz, M. A. (2018). Transfer learning with seasonal and trend adjustment for cross-building energy forecasting. *Energy and Buildings*, *165*, 352–363.

Runge, J., & Zmeureanu, R. (2021). A review of deep learning techniques for forecasting energy use in buildings. *Energies*, *14*(3), 608.

Saputra, Y. M., Hoang, D. T., Nguyen, D. N., Dutkiewicz, E., Mueck, M. D., & Srikanteswara, S. (2019). Energy demand prediction with federated learning for electric vehicle networks. In *2019 IEEE global communications conference (GLOBECOM)* (pp. 1–6). IEEE.

Sardianos, C., Varlamis, I., Chronis, C., Dimitrakopoulos, G., Alsalemi, A., Himeur, Y., et al. (2021). The emergence of explainability of intelligent systems: Delivering explainable and personalized recommendations for energy efficiency. *International Journal of Intelligent Systems*, *36*(2), 656–680.

Sardianos, C., Varlamis, I., Chronis, C., Dimitrakopoulos, G., Himeur, Y., Alsalemi, A., et al. (2020). A model for predicting room occupancy based on motion sensor data. In *2020 IEEE international conference on informatics, IoT, and enabling technologies (ICIoT)* (pp. 394–399). IEEE.

Sayed, A., Alsalemi, A., Himeur, Y., Bensaali, F., & Amira, A. (2021). Endorsing energy efficiency through accurate appliance-level power monitoring, automation and data visualization. In *Networking, intelligent systems and security* (pp. 603–617). Singapore: Springer.

Sayed, A., Himeur, Y., Alsalemi, A., Bensaali, F., & Amira, A. (2021). Intelligent edge-based recommender system for internet of energy applications. *IEEE Systems Journal*.

Schirmer, P. A. (2021). *Modelling of electrical appliance signatures for energy disaggregation* (Ph.D. thesis), University of Hertfordshire.

Schreiber, J. (2019). Transfer learning in the field of renewable energies–a transfer learning framework providing power forecasts throughout the lifecycle of wind farms after initial connection to the electrical grid. arXiv preprint arXiv:1906.01168.

Seyedzadeh, S., Rahimian, F. P., Rastogi, P., & Glesk, I. (2019). Tuning machine learning models for prediction of building energy loads. *Sustainable Cities and Society*, *47*, Article 101484.

Sharma, P., Park, J., & Cho, K. (2020). Blockchain and federated learning-based distributed computing defence framework for sustainable society. *Sustainable Cities and Society*, *59*, Article 102220. http://dx.doi.org/10.1016/j.scs.2020.102220.

Shen, S., Sadoughi, M., Chen, X., Hong, M., & Hu, C. (2019). A deep learning method for online capacity estimation of lithium-ion batteries. *Journal of Energy Storage*, *25*, Article 100817.

Singh, S., Sharma, P. K., Yoon, B., Shojafar, M., Cho, G. H., & Ra, I.-H. (2020). Convergence of blockchain and artificial intelligence in IoT network for the sustainable smart city. *Sustainable Cities and Society*, *63*, Article 102364.

Somu, N., Sriram, A., Kowli, A., & Ramamritham, K. (2021). A hybrid deep transfer learning strategy for thermal comfort prediction in buildings. *Building and Environment*, *204*, Article 108133.

Sousa, R., Silva, L. M., Alexandre, L. A., Santos, J., & De Sá, J. M. (2014). Transfer learning: current status, trends and challenges. In *20th Portuguese conference on pattern recognition, RecPad* (pp. 57–58).

Su, H., Yang, X., Xiang, L., Hu, A., & Xu, Y. (2022). A novel method based on deep transfer unsupervised learning network for bearing fault diagnosis under variable working condition of unequal quantity. *Knowledge-Based Systems*, *242*, Article 108381.

Sufian, A., Ghosh, A., Sadiq, A. S., & Smarandache, F. (2020). A survey on deep transfer learning to edge computing for mitigating the COVID-19 pandemic. *Journal of Systems Architecture*, *108*, Article 101830.

Sufian, A., You, C., & Dong, M. (2021). A deep transfer learning-based edge computing method for home health monitoring. In *2021 55th annual conference on information sciences and systems (CISS)* (pp. 1–6). IEEE.

Taheri, S., Ahmadi, A., Mohammadi-Ivatloo, B., & Asadi, S. (2021). Fault detection diagnostic for HVAC systems via deep learning algorithms. *Energy and Buildings*, *250*, Article 111275.

Taïk, A., & Cherkaoui, S. (2020). Electrical load forecasting using edge computing and federated learning. In *ICC 2020-2020 IEEE international conference on communications (ICC)* (pp. 1–6). IEEE.

Tan, C., Sun, F., Kong, T., Zhang, W., Yang, C., & Liu, C. (2018). A survey on deep transfer learning. In *International conference on artificial neural networks* (pp. 270–279). Springer.

Tang, S., Tang, H., & Chen, M. (2019). Transfer-learning based gas path analysis method for gas turbines. *Applied Thermal Engineering*, *155*, 1–13.

Tao, Y., Qiu, J., & Lai, S. (2021). A hybrid cloud and edge control strategy for demand responses using deep reinforcement learning and transfer learning. *IEEE Transactions on Cloud Computing*.

Tariq, S., Loy-Benitez, J., Nam, K., Lee, G., Kim, M., Park, D., et al. (2021). Transfer learning driven sequential forecasting and ventilation control of PM2. 5 associated health risk levels in underground public facilities. *Journal of Hazardous Materials*, *406*, Article 124753.

Teke, A., & Timur, O. (2014). Assessing the energy efficiency improvement potentials of HVAC systems considering economic and environmental aspects at the hospitals. *Renewable and Sustainable Energy Reviews*, *33*, 224–235.

Tian, Y., Sehovac, L., & Grolinger, K. (2019). Similarity-based chained transfer learning for energy forecasting with big data. *IEEE Access*, *7*, 139895–139908.

Tuia, D., Persello, C., & Bruzzone, L. (2016). Domain adaptation for the classification of remote sensing data: An overview of recent advances. *IEEE Geoscience and Remote Sensing Magazine*, *4*(2), 41–57.

Valinejadshoubi, M., Moselhi, O., Bagchi, A., & Salem, A. (2021). Development of an IoT and BIM-based automated alert system for thermal comfort monitoring in buildings. *Sustainable Cities and Society*, *66*, Article 102602.

van de Sand, R., Corasaniti, S., & Reiff-Stephan, J. (2021). Data-driven fault diagnosis for heterogeneous chillers using domain adaptation techniques. *Control Engineering Practice*, *112*, Article 104815.

Varlamis, I., Sardianos, C., Chronis, C., Dimitrakopoulos, G., Himeur, Y., Alsalemi, A., et al. (2022). Smart fusion of sensor data and human feedback for personalized energy-saving recommendations. *Applied Energy*, *305*, Article 117775.

Wang, Z., Dai, Z., Póczos, B., & Carbonell, J. (2019). Characterizing and avoiding negative transfer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 11293–11302).

Wang, S., Du, L., & Zhou, Q. (2019). A semi-supervised deep transfer learning architecture for energy disaggregation. In *2019 IEEE power & energy society general meeting (PESGM)* (pp. 1–5). IEEE.

Wang, Z., Jiang, D., Wang, F., Lv, Z., & Nowak, R. (2021). A polymorphic heterogeneous security architecture for edge-enabled smart grids. *Sustainable Cities and Society*, *67*, Article 102661.

Wang, H., Lei, Z., Zhang, X., Zhou, B., & Peng, J. (2019). A review of deep learning for renewable energy forecasting. *Energy Conversion and Management*, *198*, Article 111799.

Wang, N., Li, J., Ho, S.-S., & Qiu, C. (2021). Distributed machine learning for energy trading in electric distribution system of the future. *The Electricity Journal*, *34*(1), Article 106883.

Wang, T., & Tang, Y. (2022). Transfer-reinforcement-learning-based rescheduling of differential power grids considering security constraints. *Applied Energy*, *306*, Article 118121.

Wang, W., Wang, C., Wang, Z., Yuan, M., Luo, X., Kurths, J., et al. (2022). Abnormal detection technology of industrial control system based on transfer learning. *Applied Mathematics and Computation*, *412*, Article 126539.

Wang, S., Wu, J., Zhang, S., & Wang, K. (2018). SSDS: A smart software-defined security mechanism for vehicle-to-grid using transfer learning. *IEEE Access*, *6*, 63967–63975.

Wang, Y., Xu, Z., Wang, Z., & Li, H. (2019). Analysis of thermal energy saving potentials through adjusting user behavior in hotel buildings of the Yangtze River region. *Sustainable Cities and Society*, *51*, Article 101724.

Wang, Y., Yan, J., Jing, Q., Qi, Z., Wang, J., & Geng, Y. (2021). A novel adversarial transfer learning in deep convolutional neural network for intelligent diagnosis of gas-insulated switchgear insulation defect: A DATCNN for GIS insulation defect diagnosis. *IET Generation, Transmission & Distribution*.

Wang, C., Yuan, J., Huang, K., Zhang, J., Zheng, L., Zhou, Z., et al. (2021). Research on thermal load prediction of district heating station based on transfer learning. *Energy*, Article 122309.

Weiss, K., Khoshgoftaar, T. M., & Wang, D. (2016). A survey of transfer learning. *Journal of Big Data*, *3*(1), 1–40.

Wu, Z., Jiang, H., Zhao, K., & Li, X. (2020). An adaptive deep transfer learning method for bearing fault diagnosis. *Measurement*, *151*, Article 107227.

Wu, D., Wang, B., Precup, D., & Boulet, B. (2019). Multiple kernel learning-based transfer regression for electric load forecasting. *IEEE Transactions on Smart Grid*, *11*(2), 1183–1192.

Wu, Q., Wu, H., Zhou, X., Tan, M., Xu, Y., Yan, Y., et al. (2017). Online transfer learning with multiple homogeneous or heterogeneous sources. *IEEE Transactions on Knowledge and Data Engineering*, *29*(7), 1494–1507.

Xia, D., Ba, S., & Ahmadpour, A. (2021). Non–intrusive load disaggregation of smart home appliances using the IPPO algorithm and FHM model. *Sustainable Cities and Society*, *67*, Article 102731.

Xu, X., & Meng, Z. (2020). A hybrid transfer learning model for short-term electric load forecasting. *Electrical Engineering*, *102*(3), 1371–1381.

Xu, S., Wang, Y., Wang, Y., O'Neill, Z., & Zhu, Q. (2020). 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* (pp. 230–239).

Xu, C., Wang, J., Zhang, J., & Li, X. (2021). Anomaly detection of power consumption in yarn spinning using transfer learning. *Computers & Industrial Engineering*, *152*, Article 107015.

Xuan, W., Shouxiang, W., Qianyu, Z., Shaomin, W., & Liwei, F. (2021). A multi-energy load prediction model based on deep multi-task learning and ensemble approach for regional integrated energy systems. *International Journal of Electrical Power & Energy Systems*, *126*, Article 106583.

Yang, C., Cheung, Y.-m., Ding, J., & Tan, K. C. (2021). Concept drift-tolerant transfer learning in dynamic environments. *IEEE Transactions on Neural Networks and Learning Systems*.

Yang, T., Hou, Y., Liu, Y., Zhai, F., & Niu, R. (2021). WPD-ResNeSt: Substation station level network anomaly traffic detection based on deep transfer learning. *CSEE Journal of Power and Energy Systems*.

Yang, M., Liu, Y., & Liu, Q. (2021). Nonintrusive residential electricity load decomposition based on transfer learning. *Sustainability*, *13*(12), 6546.

Yang, D., Peng, X., Ye, Z., Lu, Y., & Zhong, W. (2021). Domain adaptation network with uncertainty modeling and its application to the online energy consumption prediction of ethylene distillation processes. *Applied Energy*, *303*, Article 117610.

Yang, H., & Schell, K. R. (2021). Real-time electricity price forecasting of wind farms with deep neural network transfer learning and hybrid datasets. *Applied Energy*, *299*, Article 117242.

Yang, L., Zhai, Y., & Li, Z. (2021). Deep learning for online AC false data injection attack detection in smart grids: An approach using LSTM-autoencoder. *Journal of Network and Computer Applications*, *193*, Article 103178.

Yang, X., Zhang, Y., Lv, W., & Wang, D. (2021). Image recognition of wind turbine blade damage based on a deep learning model with transfer learning and an ensemble learning classifier. *Renewable Energy*, *163*, 386–397.

Ye, R., & Dai, Q. (2018). A novel transfer learning framework for time series forecasting. *Knowledge-Based Systems*, *156*, 74–99.

Yuce, B., Li, H., Rezgui, Y., Petri, I., Jayan, B., & Yang, C. (2014). Utilizing artificial neural network to predict energy consumption and thermal comfort level: An indoor swimming pool case study. *Energy and Buildings*, *80*, 45–56.

Zamora-Martínez, F., Romeu, P., Botella-Rocamora, P., & Pardo, J. (2014). On-line learning of indoor temperature forecasting models towards energy efficiency. *Energy and Buildings*, *83*, 162–172.

Zeng, F., Li, Y., Jiang, Y., & Song, G. (2021). An online transfer learning-based remaining useful life prediction method of ball bearings. *Measurement*, *176*, Article 109201.

Zhang, F., Bales, C., & Fleyeh, H. (2021). From time series to image analysis: A transfer learning approach for night setback identification of district heating substations. *Journal of Building Engineering*, *43*, Article 102537.

Zhang, X., Bao, T., Yu, T., Yang, B., & Han, C. (2017). Deep transfer Q-learning with virtual leader-follower for supply-demand stackelberg game of smart grid. *Energy*, *133*, 348–365.

Zhang, X., Chan, K. W., Li, H., Wang, H., Qiu, J., & Wang, G. (2020). Deep-learning-based probabilistic forecasting of electric vehicle charging load with a novel queuing model. *IEEE Transactions on Cybernetics*, *51*(6), 3157–3170.

Zhang, H., & Cheng, W. (2021). Fault line selection method based on transfer learning depthwise separable convolutional neural network. *Journal of Electrical and Computer Engineering*, *2021*.

Zhang, W., Deng, L., Zhang, L., & Wu, D. (2020). A survey on negative transfer. arXiv preprint arXiv:2009.00909.

Zhang, K., Si, D., Wang, W., Cao, J., & Zhang, Y. (2021). Transfer learning for distributed intelligence in aerial edge networks. *IEEE Wireless Communications*, *28*(5), 74–81. http://dx.doi.org/10.1109/MWC.011.2100061.

Zhang, Y., Tang, G., Huang, Q., Wang, Y., Wang, X., & Lou, J. (2021). FedNILM: Applying federated learning to NILM applications at the edge. arXiv preprint arXiv:2106.07751.

Zhang, Y., Tang, Q., Zhang, Y., Wang, J., Stimming, U., & Lee, A. A. (2020). Identifying degradation patterns of lithium ion batteries from impedance spectroscopy using machine learning. *Nature Communications*, *11*(1), 1–6.

Zhang, X.-Y., Watkins, C., & Kuenzel, S. (2021). Multi-quantile recurrent neural network for feeder-level probabilistic energy disaggregation considering roof-top solar energy.

Zhang, Y., & Yan, J. (2019). Domain-adversarial transfer learning for robust intrusion detection in the smart grid. In *2019 IEEE international conference on communications, control, and computing technologies for smart grids (SmartGridComm)* (pp. 1–6). http://dx.doi.org/10.1109/SmartGridComm.2019.8909793.

Zhang, Y., & Yan, J. (2020). Semi-supervised domain-adversarial training for intrusion detection against false data injection in the smart grid. In *2020 international joint conference on neural networks (IJCNN)* (pp. 1–7). IEEE.

Zhao, W. (2017). Research on the deep learning of the small sample data based on transfer learning. In *AIP conference proceedings, Vol. 1864*. AIP Publishing LLC, Article 020018.

Zheng, Z., Qi, H., Zhuang, L., & Zhang, Z. (2021). Automated rail surface crack analytics using deep data-driven models and transfer learning. *Sustainable Cities and Society*, *70*, Article 102898.

Zhou, Y., Han, M., He, J., Liu, L., Xu, X., & Gao, X. (2020). Abnormal activity detection in edge computing: A transfer learning approach. In *2020 international conference on computing, networking and communications (ICNC)* (pp. 107–111). IEEE.

Zhou, Z., Xiang, Y., Xu, H., Wang, Y., & Shi, D. (2021). Unsupervised learning for non-intrusive load monitoring in smart grid based on spiking deep neural network. *Journal of Modern Power Systems and Clean Energy*.

Zhou, Z., Xiang, Y., Xu, H., Yi, Z., Shi, D., & Wang, Z. (2021). A novel transfer learning-based intelligent nonintrusive load-monitoring with limited measurements. *IEEE Transactions on Instrumentation and Measurement*, *70*, 1–8. http://dx.doi.org/10.1109/TIM.2020.3011335.

Zhu, X., Chen, K., Anduv, B., Jin, X., & Du, Z. (2021). Transfer learning based methodology for migration and application of fault detection and diagnosis between building chillers for improving energy efficiency. *Building and Environment*, *200*, Article 107957.