



Article **Transfer Learning for Renewable Energy Systems: A Survey**

Rami Al-Hajj^{1,*}, Ali Assi², Bilel Neji¹, Raymond Ghandour¹ and Zaher Al Barakeh¹

- ¹ College of Engineering and Technology, American University of the Middle East, Egaila 54200, Kuwait; bilel.neji@aum.edu.kw (B.N.)
- ² Independent Researcher, Montreal, QC H1X1M4, Canada; assi@ieee.org
- Correspondence: rami.alhajj@aum.edu.kw

Abstract: Currently, numerous machine learning (ML) techniques are being applied in the field of renewable energy (RE). These techniques may not perform well if they do not have enough training data. Additionally, the main assumption in most of the ML algorithms is that the training and testing data are from the same feature space and have similar distributions. However, in many practical applications, this assumption is false. Recently, transfer learning (TL) has been introduced as a promising machine-learning framework to mitigate these issues by preparing extra-domain data so that knowledge may be transferred across domains. This learning technique improves performance and avoids the resource expensive collection and labeling of domain-centric datasets; furthermore, it saves computing resources that are needed for re-training new ML models from scratch. Lately, TL has drawn the attention of researchers in the field of RE in terms of forecasting and fault diagnosis tasks. Owing to the rapid progress of this technique, a comprehensive survey of the related advances in RE is needed to show the critical issues that have been solved and the challenges that remain unsolved. To the best of our knowledge, few or no comprehensive surveys have reviewed the applications of TL in the RE field, especially those pertaining to forecasting solar and wind power, load forecasting, and predicting failures in power systems. This survey fills this gap in RE classification and forecasting problems, and helps researchers and practitioners better understand the state of the art technology in the field while identifying areas for more focused study. In addition, this survey identifies the main issues and challenges of using TL for REs, and concludes with a discussion of future perspectives.

Keywords: transfer learning; knowledge transfer; renewable energy; energy forecasting; fault diagnosis; buildings load forecasting; reinforcement learning

1. Introduction

The numbers given by almost all energy bodies in the world show that RE is growing faster than all other traditional forms of energy. This fact is a result of the instability in the fossil fuels market and the numerous benefits of renewables. RE is clean energy that emits no or very little greenhouse gases and pollutants, which is good for the environment and better for the health of humans. RE also comes at low cost and is increasingly competitive, which keeps energy prices at affordable levels and minimizes fuel dependency. It creates jobs for local communities, makes the overall energy system more resilient, and prevents power shortages. RE is accessible to all, secure, and ensures the sustainability of energy systems.

As an important part of smart grids today, RE should support the stability of the overall energy system. Thus, it is important to successfully predict the amount of energy available from a given alternative source at a certain time so that energy needs will be secured. Fault diagnosis systems play a key role in supporting the stability and reliability of RE systems [1,2]. Faults can seriously reduce the efficiency of power generation and the lifespan of RE systems. Moreover, they may cause unexpected downtimes, affecting the production-consumption balance, which may also lead to an instability and inconsistency



Citation: Al-Hajj, R.; Assi, A.; Neji, B.; Ghandour, R.; Al Barakeh, Z. Transfer Learning for Renewable Energy Systems: A Survey. *Sustainability* **2023**, *15*, 9131. https:// doi.org/10.3390/su15119131

Academic Editors: Andrii Chub, Mohamed Trabelsi and Sertac Bayhan

Received: 10 March 2023 Revised: 6 May 2023 Accepted: 29 May 2023 Published: 5 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https:// creativecommons.org/licenses/by/ 4.0/). in the RE economy [2]. In this paper, we highlight four general goals for RE fault diagnosis: reduce the impact of failures in RE systems, improve their reliability, increase their lifespan, and reduce their maintenance costs.

Notably, RE resources are, in general, uncertain and variable [3–5]. To improve the stability of RE systems, numerous methods have been proposed to handle the variability in RE outputs, as well as the uncertainty in fault diagnoses [6-8]. In the past few decades, various ML techniques have been suggested to cope with these types of challenges. However, the performance of classical ML methods may be subject to some limitations due to several barriers, namely (i), the lack of sufficient training datasets, also named data scarcity [3,5]. This may happen particularly for newly implemented solar energy stations and wind turbines farms. (ii) The shortage or limitation of labeled data instances to train ML models, keeping in mind that the collection of labeled data records is a laborious and time-consuming task. (iii) The need for highly computing resources to train and validate ML models on newly collected training data, which becomes particularly challenging when deep learning (DL) models are trained on massive training datasets. Moreover, most applied statistical ML models are built on the assumption that training and testing data belong to the same feature space and follow the same data distribution, which is not true in most real applications. In addition, when the data distribution varies, the developed models must be retrained using newly gathered data, which is usually difficult and expensive in most real-world applications as data become outdated quickly [9]. Whenever an available training dataset is insufficient, the notion of augmenting the data with a similar dataset from a nearby domain has been tested many times. This happens to be the case with ML models applied to RE applications (e.g., solar photovoltaic (PV) and wind farm technologies) [10,11].

The purpose of TL, also known as knowledge transfer or domain adaptation, is to solve or alleviate all the above mentioned issues and limitations. In fact, this recently proposed ML framework can efficiently help in (i) transferring the learning from the already trained ML models with enough available annotated data to newly developed models, and then validate these models with limited available unlabeled data; (ii) save training time when training newly designed ML models by fine tuning the hyper-parameters of the already-trained models that belongs to similar domains, and therefore improve the overall efficiency. This is particularly important in the case of DL models; (iii) import the knowledge from one or more existing models instead of re-training every newly created model from scratch. This can be ensured by managing well the knowledge imported from previous similar domains and definitely improve incremental life-long learning. Recently, TL started to draw the attention of RE researchers, mainly to improve energy forecasting, buildings' power consumption, and fault diagnosis prediction tasks [10–15]. The growing interest of scholars and researches in applying TL techniques on RE systems is clearly noticed through the increasing number of relevant publications in recent years. Figure 1 depicts the distribution of the past six years' TL publications on RE in the three areas mentioned above. It is worth mentioning here that in the literature, we recently found one paper that surveyed the applications of TL on energy systems, and more precisely on smart cities and sustainability [16]. The authors in [16] focused on smart buildings and load forecasting-namely sports facilities, thermal comfort control, smart grids, and energy disaggregation—in particular types of cities.

In this paper, we present a comprehensive survey on the recent advances in TL applications to REs. After introducing TL and the main categories and types of knowledge transfer with particular examples of RE applications, the paper draws a qualitative analysis on the latest applications of TL on particular RE fields, namely RE forecasting that includes solar and wind energies, building load forecasting, and fault diagnosis in RE systems. The discussion in this paper helps researchers obtain a better understanding of the limitations and potentials of applying TL to the RE fields. The paper presents a survey covering research works reported between 2016 and 2022 in different popular online databases, including the Institute of Electrical and Electronics Engineers (IEEE) Xplore digital library, ScienceDirect, Scopus, the Association for Computing Machinery (ACM) digital

library, and Google Scholar. The paper attempts to discuss the challenging questions that face researchers when they tend to apply a TL framework to RE systems. The questions can be summarized by the following: When can we transfer? How do we transfer? And which part of knowledge is to be transferred? The paper also compares recent TL-based approaches in each of the three focus topics mentioned above in terms of TL subcategories, types of ML models, datasets, etc. Additionally, case studies for each of those topics will be presented showing the structure of the proposed TL-based approaches, the adopted ML models, source domains selection, and the improvement rates provided by TL. Furthermore, the paper concludes with the open challenges in applying TL to RE systems, such as the selection and evaluation of similar source domains, negative learning, and others. This paper will provide RE researchers with a useful survey to help in the identification of advantages and challenges when applying TL to the three main applications of RE mentioned above. It also identifies the research perspectives of applying TL to solve problems of REs in the near future.

Practically, TL consists of transferring data or interim model results from a source domain, D_S , to a target domain, D_T . In this context, a domain refers to the set of data records serving as an input to a model. In the literature, we distinguish between a source task, T_S , and a target task, T_T , where both could be related to either classification or regression tasks. TL is particularly important when the data in the target domain, D_T , are limited or not available. Therefore, when applied appropriately, TL improves the performance of T_T by transferring the right knowledge from T_S and D_S . The application of TL to RE is relatively new, and the reported performance and improvements in this field remain below the level of other TL applications, such as natural language processing, image classification, activity recognition, and others [17–22]. Figure 2 presents the distribution of the current TL publications. It is clear that more than 85% of TL publications were published in the past four years, indicating a growing interest.



Figure 1. Cont.



Figure 1. Transfer learning publications for the three surveyed renewable energy domains taken between 2016 and 2022. (a) Distribution of TL publications on surveyed topics in RE systems over the years between 2016 and 2022. (b) Number of TL publications on the surveyed topics in RE for the years between 2016 and 2022.

Figures 1 and 2 indicate the increasing number of reported research on TL for RE in the three recent years, particularly for energy forecasting and faults diagnosis in RE systems. In 2022, the number of publications has doubled compared to 2021. That is justified by the growing need to improve the fault diagnosis as well as the RE predictions.



(a)

Figure 2. Cont.



Figure 2. Number of TL publications on RE per publisher between 2016 and 2022. (a) Distribution of TL publications on surveyed topics in the top publishers over the years between 2016 and 2022. (b) Distribution of TL publications over the surveyed topics in RE by the top publishers for the years between 2016 and 2022.

Figure 3 depicts the general organization and the main contribution of the survey.



Figure 3. General organization and main contribution of the survey.

The TL and its main types and characteristics are discussed in Section 2. In Section 3, an overview of the recent TL applications on RE is provided, namely energy forecasting, cross-building energy consumption prediction, and fault diagnosis where we present and discuss the potential and limitations of the reported techniques. In Section 4, we discuss the main challenges and open problems facing researchers in applying TL for the prediction and classification in RE; in addition, we introduce some open challenges and potentials that might help in improving the generalizability of TL in order to expand its deployment in real world RE projects. Finally, we conclude this work with some perspectives for the near future.

2. Transfer Learning

Classical ML algorithms are data-driven methods to predict future data using statistical models that are designed and trained on historical data records [5,23,24]. Most of these methods assume that the amount of training data instances is sufficient and that the training and testing data belong to the same feature space and have the same distribution, which is not always true in real-world applications. TL addresses the cases in which domains, tasks, or data distributions in the training and testing phases are different. In fact, we encounter many examples of TL at the human learning level. For instance, the general ability to recognize animals in photos will help in the task of identifying cats in a separate set of photos. Likewise, the knowledge and experience of riding bicycles may help in learning how to ride a motorcycle. TL research is inspired by the fact that human beings exploit prior knowledge to learn new skills and solve new problems. In the field of ML, the main idea behind TL is to establish lifelong ML platforms that intake and reuse collected knowledge to solve new problems of classification, regression, and clustering [23,24]. TL is also referred to as life-long learning, domain adaptation, and incremental learning. The aim is to import knowledge from one or more source domains, and apply it to a target task in another different domain [23,25,26]. Figure 4 depicts the main differences between the learning process of classical ML algorithms and those that are adopted by TL techniques.



(a) Learning Process in Classical Machine Learning

(b) Learning Process in Transfer Learning

Figure 4. Learning processes in classical ML (**a**) and TL (**b**).

As shown in Figure 4, classical ML trains each model to perform a specific task from scratch, whereas TL exploits and transfers the knowledge accumulated from previous tasks to train for a new target task.

2.1. TL Definitions and Notations

In this section, we present the notations and definitions used in this survey. A domain, *D*, is defined by two components: a feature space, χ , and a marginal probability distribution, P(X), where $X = \{x_1, x_2, \ldots, x_n\} \in \chi$. For instance, if the learning task is a sentiment classification of a particular product given as Good, Neutral, or Bad, each term (i.e., comment) is encoded as a binary feature, and x_k is the kth term vector corresponding to the comment. Therefore, χ is the space of all possible term vectors, while x_k is a learning sample. Practically, the two learning domains are said to be different if they have different features' spaces or if the marginal probability distributions of those features' spaces are different. For a given domain, $D = \{\chi, P(X)\}$, a task, T(y, f(.)), related to this domain consists of a label space, y_i and a prediction function, f(.), which can be found by the training data sample, X. In general, a training sample consists of pairs in the form $\{x_k, y_k\}$, where $x_k \in X$ and $y_k \in \mathcal{Y}$. In our sentiment classification example, \mathcal{Y} is the set of all possible labels (i.e., $y = \{Good, Neutral, Bad\}$). The objective function, f(.), aims to associate a label, f(x), to a given instance in the learning sample, X. From the probability point of view, f(x) can be denoted as a conditional probability, P(y|x), that estimates the probability of having a label, y, given an input, x. Thus, the data of a source domain are denoted by the set of pairs, $D_s = \{(x_{S_1}, y_{S_1}), ..., (x_{S_k}, y_{S_k}), ..., (x_{S_n}, y_{S_n})\}$, where $x_{S_k} \in \chi_S$ is an input data instance in the source domain, and $y_{S_k} \in y_S$ is the associated label in the source set of labels for the instance, x_{S_k} . On the other hand, the data in a target domain are represented by $D_T = \{(x_{T_1}, y_{T_1}), ..., (x_{T_k}, y_{T_k}), ..., (x_{T_m}, y_{T_m})\}$, where $x_{T_k} \in \chi_T$ is an input data instance in the target domain, and $y_{T_k} \in y_T$ is the corresponding output/label in the target set of labels for the instance, x_{T_k} . In general, we assume that $0 \le n \ll m$.

The general goal of TL is to improve the learning of a target predictive function, $f_T(.)$, in a target domain, D_T , using the accumulated knowledge in a source domain, D_S , and a source learning task, T_S , with the assumption that $D_S \neq D_T$ or $T_S \neq T_T$. The assumption $D_S \neq D_T$ means that either the spaces of features χ_S and χ_T are not the same (i.e., $\chi_S \neq \chi_T$), or that the marginal distributions of the features' spaces are not equal (i.e., $P_S(X) \neq P_T(X)$). For instance, in our sentiment analysis example, either the set of term features in the source space differs from that of the target (e.g., comments are written for different categories of products or expressed in different languages), or the marginal distributions of the two sets are different.

On the other hand, there is the assumption that a source task defined as $T_S = \{y_S, P(Y_S|X_S)\}$, and a target task, $T_T = \{y_T, P(Y_T|X_T)\}$, are not equal if either (1), the label spaces, y_S and y_T , are not equal (i.e., $y_S \neq y_T$), or (2), the conditional probability distributions, $P(Y_S|X_S)$ and $P(Y_T|X_T)$, are different (i.e., $P(Y_S|X_S) \neq P(Y_T|X_T)$), where $Y_{S_k} \in y_S$ and $Y_{T_k} \in y_T$. In our example, the first case can be depicted as a situation in which the source task has two classes of evaluations (e.g., "liked" or "disliked" outcomes), whereas the target task has five classes of evaluations (e.g., number-of-stars score). The second case can be depicted as a situation in which the comments in the source and target domains belong to unbalanced user-predefined classes.

2.2. Categories of TL Techniques

The two main considerations of this survey when examining TL solutions are (i), which part of the knowledge is to be transferred, and (ii), how to make this knowledge transfer. Based on the different linkages between the source domain, D_S , and the target domain, D_T , the answers may lead to a common categorization of TL instances that consider three common sub-settings briefly described in the following subsections: inductive, transductive, and unsupervised TL [23–25].

- 1. The inductive TL has a source task, T_S , and a target task, T_T , that are different from each other. However, domains D_S and D_T may be the same or different. The inductive TL requires some labeled data in the target domain, D_T , to deduce an objective predictive function, $f_T(.)$ in the D_T . In addition, if there is a sufficient amount of labeled data in the source domain, D_S , in this case, inductive TL is considered similar to a multitasking learning case. However, if labeled data are not available in the source domain, D_S , inductive TL will be a self-taught learning case [21,23,26].
- 2. Transductive TL has tasks T_S and T_T , which are identical, whereas domains D_S and D_T are different (i.e. either the features spaces in the target and source domains are not the same, $\chi_S \neq \chi_T$, or the same feature spaces are defined in the two domains but with different marginal probability distributions, $\chi_S = \chi_T$ and $P_S(X) \neq P_T(X)$) [21,23]. In

this sub-setting, we assume that there are no labeled data in the target domain, D_T , while there are enough in the source domain, D_S .

3. Unsupervised TL assumes that the source task, T_S , and target task, T_T , differ, but are related [21,23]. In general, this TL setting is used to solve unsupervised learning problems in the target domain, D_T (e.g., dimensionality reduction and clustering). This type of sub-setting is not widely addressed in the RE literature.

The literature suggests that researchers and scholars consider additional TL categorizations: (i) homogeneous, where the input features are the same (i.e., $X_S = X_T$) and the space of labels is the same (i.e., $y_S = y_T$), or (ii) heterogeneous, where X_S and X_T do not have the same features (i.e., $X_S \neq X_T$), or the marginal distributions are not equal (i.e., $P_S(X) \neq P_T(X)$ and/or $y_S \neq y_T$ [23,25]). Table 1 depicts the common TL categorizations with real application examples.

	Transfer Type	Features Domain	Task	
Homogeneous	Inductive Learning	$\chi_S = \chi_T$	$T_S \neq T_T$	D_S and D_T have the same input features. For instance, temperature, humidity, and solar radiation. The T_S is tuned to forecast solar power instead of heat index or dew point [17].
	Transductive Learning	$\chi_S = \chi_T$ $P(X_S) \neq P(X_T)$	$T_S = T_T$	D_S and D_T have the same input features but different marginal distribution. For instance, temperature, humidity, and air density. The T_S and T_T are both used to forecast solar radiation [11].
Heterogeneous	Inductive Learning	$\chi_S eq \chi_T$	$T_S \neq T_T$	D_S and D_T have a different number of features. For instance, the air pressure is additionally used in the D_T to forecast solar power instead of heat index or dew point [25,26].
	Transductive Learning	$\begin{array}{c} \chi_S \neq \chi_T, \\ P(X_S) = P(X_T) \end{array}$	$T_S = T_T$	D_S and D_T have different but related features of one solar station location. The underlying marginal distributions are equal due to the similar location of solar station. T_S and T_T forecast the solar power [12,13].

Table 1. Transfer learning categories and data similarities.

Another distinction depends on which part of the knowledge is to be transferred from $D_{\rm S}$ to $D_{\rm T}$. In this context, four approaches are defined, which depend on the level of instances, features, and parameters. It may even offer a relational-based approach [21,25,26].

- Α. Instance-based approaches are data oriented and focus on transferring knowledge by adjusting data to reduce the distribution difference between the source and target domains. There are two main approaches for the instance-based category:
 - 1. Instance weighting: The instances are assigned weights to reduce the marginal distribution difference.
 - 2. Domain weighting: For multi-source domains, weights are assigned to each source to reduce the conditional distribution difference [27].
- B. Feature-based approaches are data-oriented and aim to find a new representation for each original feature. The main purpose of constructing a new feature representation is to minimize the conditional and marginal distribution differences and preserve potential data structures. This can be realized by feature augmentation, feature alignment, or feature reduction [21,23].

- C. Parameter-based approaches are model-based and transfer knowledge using model parameters that reflect the knowledge of those models. The two main approaches are as follows:
 - 1. Parameter sharing: Widely used to directly share parameters of the source learner with the target learner. A widely used technique is to freeze the first layers in a convolutional neural network and only fine-tune the last layers to produce a target model of the same type [10,12].
 - 2. Parameter restriction: The parameters of the target and source learners must be similar.
- D. The relational-based approaches transfer the logical relationships and rules accumulated in the source domain to the target.

Figure 5 summarizes the described TL categories. For a more comprehensive understanding of the topic, readers may refer to numerous reports [23,25,28–30]. Those interested in particular applications may refer to [19,31,32].



Figure 5. General TL Categories and Approaches.

3. TL for RE Systems

The number of TL applications for RE systems have recently increased, especially for solar and wind energy forecasting, building energy consumption prediction, and fault diagnosis. The main idea of TL is to cope with limited or absent historical datasets with which ML data-driven models (e.g., Support Vector Regressors, Random Forests (RFs), and Neural Networks (NN) [4,5]) must be trained. The scarcity of training data is a real challenge when building new PV plants and wind farms [10,11]. It is also prevalent in new and existing buildings with newly installed electricity meters. TL techniques for RE applications allow ML methods to be applied with limited or even no training data by inferring knowledge from existing energy forecasting models trained with sufficient data. Like the many TL domains discussed so far, the transfer of knowledge among RE forecasting models may occur in data or model domains. Therefore, the same conceptual questions arise about how the knowledge can be transferred from a source domain, D_S , where knowledge is readily available, to a target domain, D_T , where the knowledge is limited. In the following, we

introduce the TL techniques applied to RE forecasting and address the main challenges facing researchers when applying such techniques.

3.1. TL for Power Forecasting

For RE forecasting, domain *D* consists of a feature space, χ , that in general contains weather, wind speed, or solar radiation data alongside their marginal distribution, P (X) for $X \in \chi$. The task of the domain, T(y, f(.), assigns the expected power, Y, to an input, X, where f(.) is the rule that assigns the expected power, Y, to the input weather and solar data, where $Y \in y$.

The type of TL used for RE forecasting is chosen based on the types of challenges that arise. Table 1 summarizes the common TL categories applied to RE forecasting with respect to the similarity of data in D_S and D_T domains alongside the types of adopted learning.

In general, the use of TL for RE forecasting faces two main challenges [10]. The first challenge arises when researchers assume the absence of historical data in D_T . This is typically the case when building new PV plants or wind farms [11,12]. In this case, researchers may use numerical weather parameters and physical non-ML models [10] that provide small amounts of training data for D_T , which help in selecting similar solar PV or wind-farm configurations. The D_S and T_S of the selected source farms are therefore adapted to D_T to provide forecasting results using TL techniques. The main difficulty in this scenario is the likelihood of distorting the historical data, X_S and Y_S , owing to the specificity of the source PV or wind farm [10,11]. Another difficulty is the possible bias of the source forecasting model, T_S , toward a particular PV or wind farm.

The second challenge may take place when researchers assume the availability of limited data, Y_T , for new solar PV or wind farms. The available set of limited records may consist of measurements related to the first few months of establishing the new energy station. Among the issues that may arise from the available dataset, Y_T , is a potentially different statistical distribution related to the pre-selected stations in the D_{5} . This is likely to occur if the set of similar power stations and the one being planned are different. Other challenges may appear if the dataset related to the first few months of the new power station is biased by the seasonal weather conditions. Two common approaches have been suggested for dealing with these challenges. The first approach uses multi-task learning methods based on parameter sharing [10–12]. A model is chosen using the dataset of similar power stations. All models are assumed to be trained in parallel. The selected model is then trained on the historical data that are available for the source stations. Then, the trained model is fine-tuned to the tasks of the newly added power station. This method requires a small amount of new data to adapt the parameters of the newly modeled power station [10]. The second approach consists of combining fine-tuning with self-training. A generic model based on a set of pre-selected models for similar power stations is designed and trained. The generic model is then fine-tuned to the available historical data for a short period on the newly added power station. The updated version of this model is then used to generate artificial data from the available numerical weather data at the new station. Afterward, artificially generated data are used for the additional fine-tuning of the generic model. The described process of generating artificial data and then retraining provides a self-training technique that improves the prediction performance of the target. The finetuning of the two described techniques minimizes the overall design and training costs.

In the RE literature, numerous research works have focused on the application of TL to forecasting wind and solar power generation [10–14]. Most of the proposed TL approaches for RE prediction have considered short-term predictions and have applied feature and parameter transfer methods [11,12]. In [11], the authors applied a DL approach to extract a high-level representation of weather data to predict wind energy. The authors combined wind speed information from multiple sources to build a model with shared hidden layers. The shared layers ensure a universal feature transformation and help in extracting the hidden rules among wind-speed patterns. The output layers of the proposed model are farm-dependent, as the data distribution differs from one farm to another [11].

In [12], the authors assumed that the outputs of PV panels were mainly affected by the strength of solar radiation. Therefore, the time-series data records of the PV stations under study had similar probability distributions. The authors in this study considered the solar radiation data as D_S and the PV output data as D_T . They also implemented a feature-based TL by extracting common data between D_S and D_T through shared hidden layers. The output layer of the target model was fine-tuned using D_T data to improve the prediction performance. The authors claimed that increasing the amount of training data caused the advantages of TL to decrease gradually. Figure 6 shows the general structure of the model introduced in [12], which consists of a share-optimized-layer long short-term memory (LSTM). In this model, the general features are extracted through the hidden layer of the models pre-trained using D_S . In this work, authors claimed that the features had good generalizability.



Figure 6. The structure of share-optimized-layer long short-term memory with fine-tuned phase adapted from [12].

In [13], the authors introduced a cluster-based multi-source domain adaptation (MSDA) approach with inductive TL to transfer the wind-speed knowledge of multiple source stations to a newly built station in order to predict wind power. The proposed approach captures wind data distribution, and then computes the distributions similarity between the source and target domains. The authors claimed that similar weather parameters have similar influences on wind-power production, and they formulated the similarities between each source and target domain in terms of marginal and conditional probability distribution. These similarity measures were used to weigh the contribution of each source domain in estimating the predicted wind power. Their cluster-based MSDAweighted method provided a 16.88% better estimation than an MSDA approach that assigns equal weights to the source domains. The authors of [14] suggested a hybrid TL model consisting of a convolutional neural network (CNN) and a gated recurrent network (GRU) for short-term wind speed predictions. The CNN extracts complex features from meteorological data from surrounding cities to be used by the GRU model, which learns the relationships among time series data records to ensure the short-term prediction of wind speed. The source and target domains were the same with similar data distributions, and the authors fine-tuned the CNN target models accordingly. The authors claimed that their model improved the scores of statistical metrics by nearly 20%. In [33], the authors

introduced an adaptive TL (ATL) for a deep neural network (DNN) for the short-term prediction of wind power. The suggested ATL-DNN method continuously exploits the arriving information and handles the inductive transfer of knowledge between the task domains (i.e., from wind power to wind speed). The design dataset was comprised of wind directions and wind speed parameters collected from five wind farms in European regions. The suggested ATL-DNN method is ensemble-based, in which three base learners are trained adaptively every four months on the continuously generated data. The authors in [34] adopted a model-based TL strategy for multi-step-ahead wind-power prediction for newly established wind farms. The first layer of the proposed model consists of a serio-parallel feature extractor consisting of several CNNs separately connected to stacked LSTMs (CNN-LSTM). The CNNs extract meteorological features, whereas the LSTMs extract temporal information features from surrounding similar wind farms with sufficient training data. The second layer consists of a fully connected network for prediction. The first stage of the adopted parameter-based TL strategy transfers the partial parameters of the well-trained CNN-LSTM feature extractor of the source wind farm to the prediction model of the target. In the second phase, a personalized training of the fully connected network is assured using personalized training data of several weather features (e.g., temperature, relative humidity, wind direction, and air pressure). The authors in [35] proposed a multi-wind farm strategy for the prediction of wind speed using a bidirectional LSTM (Bi-LSTM) learning model. Four Bi-LSTM DL models were pre-trained with historical wind speed data collected from four wind farms to obtain an ensemble model. A transductive model-based TL was applied to transfer the knowledge of the four pre-trained models to a centralized control unit, and the wind speed at any newly created wind farm was predicted using this knowledge. A hybrid multi-objective optimization algorithm based on firefly and dragonfly algorithms [36,37] was used to weigh the predictions of a set of pre-trained models to obtain optimal wind-speed predictions.

The authors in [38] introduced a TL-based approach to extract features from sky images using a deep CNN to model the association between sky images and solar radiation. First, they trained a CNN classifier to extract high-level features to determine whether the sun is shaded by clouds. Using a parameter-based TL, part of the CNN classifier is used to construct a CNN regressor to extract information from sky images and quantitatively map the extracted information to continuous solar radiation. Practically, the parameters of the CNN classifier's convolutional layers are fixed, whereas the parameters of the subsequent fully connected layers are updated during the training phase.

In [39], the authors introduced a transductive instance-based TL method that uses gradient boosting decision trees (GBDTs) for wind power prediction. GBDT models were found effective for probabilistic forecasting [39]. The proposed regression model was trained using auxiliary wind power data from correlated zones as source problems. The incorporation of data from source problems during training helps in enhancing the prediction performance in the target zone. The authors in [40] proposed a parameters-based TL approach to train one-hour-ahead predictors of hourly average Global Solar Irradiance GHI at new locations with limited data. The authors tuned pre-trained recurrent neural network-based models to use them at other locations with limited data. The methodology is demonstrated by considering one source site in Egypt (rich in annotated data), and the target models were tuned using limited datasets at five other locations in Tunisia, Morocco, and Jordan. The approach in this work was found valid for all targeted locations in terms of mean absolute errors. Additionally, the authors reported the least and most influential parameters on the performance of their proposed TL models. Table 2 summarizes the recent publications that are relevant to the applications of TL in predicting wind power and solar power/radiation. The table compares the transfer types, adopted features, targets to be predicted, and the applied ML models.

Authors	TL Sub Setting	Transfer Type	Features	Target to Predict	ML Models
Hu, Q., et al. (2016) [11]	Homogeneous Transductive	Feature-based Transfer	Wind speed	Wind speed	Deep learning with shared hidden layers
Zhou, S., et al. (2020) [12]	Heterogeneous Inductive	Feature-based Transfer	Solar irradiance data, PV output data	Photovoltaic power forecasting	LSTM with sequential model-based global optimization SMBO
Tasnim, S., et al. (2018) [13]	Homogeneous inductive	Instance-based Transfer	Statistical features of wind speed (mean, std deviation, skewness,)	Wind power	Regression models with multi-Source data adaptation MSDA
Ji, L., et al. (2022) [14]	Homogeneous transductive	Instance-based Transfer	Meteorological features	Wind speed	CNN and GRU
Qureshi, A., et al. (2018) [32], (2017) [41]	Homogeneous inductive	Feature-based Transfer	Wind speed and wind direction	Wind speed	DNN and ensemble learning
Yin, H., et al. (2021) [33]	Homogeneous transductive	Model-based (parameters passed)	Weather features	Wind power	CNN, LSTM, and fully connected networks.
Liang, T., et al. (2021) [35]	Homogeneous transductive	Model-based	Wind speed and direction	Wind speed	Bi-LSTM with multi-objective optimization
Lin, Y., et al. (2019) [38]	Homogeneous transductive	Parameters-based	Sky images	Solar radiation	Deep CNN
Cai, L., et al. (2019) [39]	Homogeneous transductive	Instance-based	Power energy statistics, wind speed, wind directions	Wind power	GBDT
Abubak et al. (2022) [40]	Homogenous inductive	Parameter-based	Solar radiation	Solar radiation	Recurrent neural network
Cao, L., et al. (2018) [42]	Homogeneous transductive	Instance-based	Time series wind power	Wind power	Extreme gradient boosting
Liu, Y., et al. (2022) [43]	Homogeneous transductive	Model-based	Wind power	Wind power	MLP + particle swarm optimization
Abubak et al. (2022) [43]	Homogenous inductive	Parameter-based	Solar radiation	Solar radiation	Recurrent neural network
Khan et al. (2022) [44]	Homogeneous inductive	Parameter-based	Wind and energy measurements	Wind power	Deep neural network
Jifeng Song et al. (2022) [45]	Homogeneous inductive	Parameter-based	Weather, turbine, and rotor functions	Wind power	Temporal convolutional network + LSTM
JeongRim et al. (2022) [46]	Homogeneous inductive	Model-based	Weather data and wind power generation data	Wind power	Light gradient boosting machine
Schreiber et al. (2022) [47]	Homogeneous inductive	Parameter-based	Weather data	Wind/Solar power	Temporal convolutional network + auto encoders
Shuang Yu et al. (2022) [48]	Homogeneous inductive	Model-based	Weather and climate data	Wind speed	Random forest + XGBoost
Runhao et al. (2002) [49]	Homogeneous	Instance-based	Solar power data of PV power station	Solar power	Extreme learning machine + TrAdaBoost
Tajjour et al. [50]	Homogeneous inductive	Model-based	Satellite data of solar radiation	Photovoltaic power	Deep neural networks DNN

Table 2. Overview of the latest works on TL for predicting wind- and solar-power/radiation.

Table 2 shows that TL is widely applied with deep learning models to forecast solar power/radiation and wind power/speed. Most of the research works reported in the liter-

ature targeted wind prediction with homogeneous domains, which can be explained by the availability of benchmark data in the source domains and by the simplicity of applying transductive TL among wind power stations. Additionally, most of the reported research works applied the TL on DL models due to its ability to alleviate the issues of annotated data scarcity and learning costs we usually face when training DL models.

3.2. TL for Cross-Building Energy Load Forecasting

In the context of smart grids, the datasets of smart sensors-based data provided by the recently deployed smart meters make the sensor-based power consumption prediction more popular [51–55]. Short- and mid-term ML-based and particularly DL-based predictions require sufficient quantities of historical data to forecast future energy consumption in buildings and facilities. Recently, a few studies have applied TL to mitigate the scarcity of historical data for new and existing buildings with newly installed electricity meters where data were not sufficient to achieve accurate prediction [51–55]. In [51], the authors introduced a cross-building energy forecasting method using an inductive instance-based TL to improve the predictions for target buildings with limited training datasets. They trained a target model with a small dataset using measurements collected over an extended period from similar buildings (i.e., schools). Their approach enabled the use of standard ML algorithms with seasonal and trend adjustments to time-series data. The authors showed that their approach increased the forecasting accuracy by up to 11.2% compared to models that use a dataset of information from a target school gathered during only one month. In [52], the authors introduced a TL approach for the cross-building transfer of knowledge using two standard reinforcement learning algorithms coupled with a deep belief network (DBN) to model building energy consumption (i.e., building models) using unlabeled historical data. The proposed method handles new behaviors of existing buildings (e.g., variations in installation, changes in structure, and changes in energy consumption due to site renovation), as well as new types of buildings newly connected to the smart grid. The authors tested their approach using a real dataset collected over seven years, with a resolution of 1 h. The results showed that their approach significantly improved energy prediction accuracy in more than 91% of cases after using the DBN models to extract high-level features from the unlabeled data.

The authors in [53] proposed a short-term future energy consumption model for office buildings using transductive instance-based TL. The proposed approach uses a simulation dataset of a reference building with environmental features (e.g., temperature, humidity, and solar radiation) and time data. The approach also uses LSTM models for time-series regression. The authors trained their TL-LSTM model using an office-building power consumption dataset collected during 24 h, and made predictions for the next 24 h. They reported that the TL-LSTM model outperformed the stand-alone LSTM model. Moreover, they claimed that when the climate zone was the same for the source and target datasets, the TL-LSTM model showed high accuracy, even if the locations of the related buildings are different.

In [54], the authors introduced a TL-based framework with two DL models (i.e., an LSTM-based encoder–decoder and a 2D-CNN) to improve the prediction of power consumption in the offices of a target building. The proposed inductive TL approach used a seq2seq encode–decoder model with a simple LSTM model, and restricted data to improve prediction accuracy by more than 19% in terms of mean absolute percentage error. However, the 2D-CNN model provided a more than 20% improvement in prediction accuracy. The authors used numerical weather features (e.g., temperature, humidity, and dew points), categorical features (e.g., holidays, and months), and energy consumption labels.

To improve the accuracy of the short-term power-load prediction, the authors in [55] proposed a parameter-based TL framework using a CNN-GRU hybrid model. They estimated the bandwidth of the data distribution by optimizing the mean integrated square error, which can provide a load prediction interval and a fluctuating range curve of the future load at a particular confidence interval. The training process of the hybrid models

in D_S and D_T was realized using multi-source data that included time-series weather and load measurements. The authors indicated that their proposed method was not suitable for long-term predictions. Notably, as the prediction horizon increases, the interval prediction expands. Figure 7 depicts the CNN–GRU hybrid model with a parameter-based TL approach, as proposed in [55].



Figure 7. CNN gated recurrent unit model with parameter-based TL. The GRU model processes the outcomes of the CNN models. Adapted from [55].

Furthermore, recent studies demonstrated the potential benefits of applying reinforcement learning (RL) on learning-based building controls to optimize overall energy efficiency and management [56,57]. Reinforcement learning [58,59] is a field of ML that studies how an ML model can react with its environment to achieve a specific task. An RL model has to control a dynamic system by selecting actions in a sequential mode. The model moves to a new state after executing an action, and it receives a reward (numeric value) that shows how far the model is from the goal state.

Recently, numerous reviews including [60,61] analyzed the applications of TL with RL models. Actually, the TL techniques are applied with RL models to optimize their training time, which is one of the major difficulties in applying such types of data-driven models. This is possible due to a certain degree of similarity among buildings' activities [56]. Few research works have introduced the use of RL models with TL techniques for buildings' energy load prediction [52].

Table 3 summarizes the latest relevant publications that discuss TL applications for cross-building energy consumption forecasting. The table compares transfer type, adopted features, type of buildings, and applied ML model. Additionally, the table indicates that TL is widely used with DL to handle the energy consumption prediction for various types of buildings.

Authors	Authors TL Sub Setting Transfer Type Features		Features	Type-of Building	ML Models
Ribeiro, M. et al. (2018) [51]	Inductive	Instance-based	Time data + weather features + energy consumption	Schools	Regression models: MLP and SVR
Mocanu, E.et al. (2016) [52]	Unsupervised	Model-based	Loads profiles (time-of-use, lightening, existence of electrical heater,)	Cross-types residential + commercial	Reinforcement learning models + DBN
Ahn, Y. (2022) [53]	Transductive	Instance-based	Meteorological + time data	Offices	LSTM
Gao, Y. (2020) [54]	Inductive	Model-based	Numerical weather features + categorical (holiday, day of weak)	Offices	LSTM + 2D-CNN
Jin, Y. (2022) [55]	Transductive	Parameters-based	Time data + weather features + energy load	Residential + commercial	Hybrid CNN-GRU
Gonzalez-Vidal et al. (2022) [62]	Transductive	Instance-based	Weather features + building meta data	Non-residential + offices	LSTM + CNN

Table 3. Overview on the recent works on TL for cross-building energy load/consumption forecasting.

Among the drawbacks of cross-building sensor-based energy consumption forecasting, we mention the following: (1) the assumption that similarly categorized buildings (e.g., residents, schools, and commercial buildings) have the same power demands often leads to weak predictions. (2) The adopted methods for this type of forecasting must consider the effects of seasonality within the specific domain of usage, as well as activity trends in buildings and the changes that may happen on the activities due to renovations and extensions [51]. Little work has been focused on applying TL to buildings' load forecasting compared to the other surveyed fields of RE. This can be explained by numerous barriers such as confidentiality issues and the relatively limited number of smart sensors that have lately been deployed in real-world projects. In fact, most energy operators do not share enough information about the status of their systems. Furthermore, business owners do not reveal details about the daily power consumption of their residential buildings, facilities, and factories.

3.3. TL for Fault Diagnosis in Energy Systems

Fault diagnosis of RE systems plays a key role in ensuring consistency in the energy industry by supporting the stability and reliability of the relevant systems [1]. An efficient faults prediction and classification can certainly help in reducing the impact of failures in such systems. Various categories of faults may arise during the operation of RE systems. For instance, in wind turbines, faults may occur in one or more critical components (e.g., generator, blades, main bearing, and gearbox) [63]. Most faults occur in offshore wind farms or in remote areas that are difficult to access. On the other hand, in solar energy systems, PV panels are installed in complex climate conditions, which may cause anomalies such as fragmentation, short-circuits, shading, cracks, and dust accumulation [1].

Faults in energy systems reduce not only the efficiency of power generation, but also their life spans. Moreover, faults may cause unexpected downtimes, which consequently affects the production consumption balance, and may also lead to instability and inconsistency in the RE economy [63]. For energy systems, four main aspects of fault diagnosis techniques are prevalent (i.e., reduce the impact of failure, increase the reliability, increase the lifespan, and reduce the maintenance costs).

3.3.1. Classical Fault Diagnosis Approaches

The physical methods of fault diagnosis are the most conventional approaches, and they rely on mathematical models to explain anomalies in the collected data. In [64], Watson et al. proposed a new approach for the continuous calculation of damage accumulation using the parameters of turbine performance and failure physics. In [65], Qiu et al. presented a thermo-physics based method for fault diagnosis in wind turbines. Although physical methods can provide good explanatory models, they are difficult to develop because they require deep physical knowledge of the system and do not address all kinds of failures [1].

Apart from physical approaches, data-driven ML-based approaches have been introduced for fault diagnosis in RE systems. They are supposed to benefit from the historical data records collected from energy systems to automatically predict and classify anomalies and failures. Among the main challenges that face researchers and engineers in applying ML-based approaches are those occurring in the boundaries between normal and faulty data, as they are ambiguous and difficult to distinguish. Moreover, the data associated with failure are frequently less common than those associated with normal working conditions [1]. Notably, data-driven ML models do not require prior physical knowledge of energy systems, as they learn fault patterns from stochastic historical operational data.

In the literature, the operational data collected from supervisory control and data acquisition (SCADA) are widely used to develop fault diagnosis systems. Currently, these data are used to design and train numerous fault diagnosis systems [63]. For instance, Pramod et al. [66] applied an artificial neural network to diagnose gearbox bearing faults using real data from onshore wind turbines. The authors relied on SCADA data to identify gearbox-bearing damage. In [67], Xiao et al. proposed an optimized support vector machine for misalignment fault diagnosis in wind turbines. In [68], Chen et al. introduced a fault diagnosis for PV arrays based on RF methods.

However, the traditional ML methods reported in the literature require considerable effort when choosing the features and labels needed to build accurate models. DL that emerged in the last decade is currently demonstrating its ability to both learn complex relationships from data, and provide accurate solutions. DL extracts features through abstraction without the need for human interaction. For instance, Chang et al. [69] proposed an intelligent bearing-fault diagnosis system for wind turbines using a concurrent CNN. Their model extracts deep information from the vibration signals of generator bearings. In [70], Zhao et al. proposed an anomaly detection and fault analysis system for wind turbine components using SCADA data based on a deep auto-encoder. Their method provides early warnings of fault components and estimates possible fault locations.

3.3.2. TL Approaches to Fault Diagnosis

Despite the accuracy of DL methods in diagnosing faults in energy systems, they perform poorly with small-scale SCADA data. Moreover, when a new wind turbine is installed, the requirement for abundant training data cannot always be satisfied. In some cases, wind turbines and PV panels may be implemented in remote areas with poor communication conditions, which may limit data collection and transfer. Most wind farms use similar types of turbines, and some solar grids use the same types of PV plants. Thus, the operational data of these systems usually contain common failure information. Therefore, the TL technique is adequate for coping with data scarcity in cases where D_T has limited training data.

Researchers have introduced different TL approaches to build efficient fault diagnosis solutions for RE systems. In the following subsections, we overview the recently suggested TL-based fault diagnosis approaches in the literature where problems of small-scale target data are tackled.

A. TL for Fault Diagnosis in Wind Turbines

Recently, TL was widely used in predicting and diagnosing the faults that may occur in wind turbines, particularly in rolling bearings, which are considered one of the most fragile parts of wind turbines. In addition, several TL approaches have been introduced to detect the blade-icing issue as one of the frequent problems affecting wind turbines, since most wind farms are installed at high altitudes with low temperatures and high levels of humidity. In the literature, the TL is widely used with both DL and convolutional autoencoders (CAE) for the design of adequate approaches to deal with common issues in wind turbines. In this section, we brief some cases in the literature that show the efficiency of TL in diagnosing common faults in wind turbine systems. In [63], Li et al. proposed a new approach for wind-turbine fault diagnosis using parameter-based TL with small-scale data. The approach shares the model parameters that are trained using available data in D_S with the target learner. Then, the target learner freezes part of the parameter set and retrains the remaining part of the D_T . In their experiments, the authors used operational SCADA data from 15 identical wind turbines in Tianjin, China. The data consists of 52,560 records. Various features related to wind power, temperature, electrical, operating conditions, and statistical variables were considered. The authors used parameter-based TL and CAE for the proposed model. In that parameter-based TL, knowledge was transferred from D_S (14 wind turbines) to D_T (one wind turbine) by sharing the parameters and network structure trained by D_S . CAE is a general-purpose feature extractor that efficiently represents features extracted from a high-dimensional space to construct new input data records.

Because CAE is an unsupervised learning technique, the authors added a classifier to the trained CAE. The authors used a fine-tuning technique to freeze the parameters of the bottom layers of the trained network, and to retrain the top layer using D_T data. The CAE was used as a "carrier of knowledge" from D_S to D_T . In their experiments, the authors selected data from the 15th wind turbine as the target and those from the other 14 wind turbines as the source. The five-layer extractor from the CAE was combined with a three-layer classifier to produce the final CAE-TL network. The parameters were obtained from CAE, and then used to initialize the feature extractor layers while the parameters of the three-layer classifier were randomly initialized. In the fine-tuning step, only the parameters of the classifier network (i.e., the last three CAE-TL layers) were adjusted using the target data, while the parameters of the feature extractor (i.e., the first five layers) were frozen.

The results of [63] demonstrated that the CAE-TL has excellent accuracy and showed the high efficacy of TL in handling fault diagnosis problems.

In general, rolling bearings are principal components in wind turbines, as they are behind 30% of the mechanical faults. This is due to their nearly continuous movement [71]. These faults affect the efficiency of wind turbines in producing power, and may lead to catastrophic consequences. In [71], Zhang et al. proposed a new model based on CNN and deep TL for fault diagnosis in wind turbines, which performed well according to statistical results. The dataset in [65] included rolling bearing and gearbox data. Bearing data were collected at 1200 samples per second. The first dataset contained different types of rolling bearing fault data (e.g., normal operation, ball fault, inner ring fault, outer ring fault, and inner and outer ring combinations). The second dataset contained five types of gear fault data (e.g., missing teeth, broken teeth, wear, undercut, and normal operation).

The authors used the bearing dataset as the source domain D_s , whereas the gear dataset was used as the target domain, D_T . Although the two datasets were measured under the same operating conditions, they had different distribution characteristics. The authors tested two methods of parameter-based TL in their experiments. The first method consisted of training the network using the source domain, followed by fine-tuning the last layer of their model using the target data. The fine-tuning process included freezing feature-extraction parameters while training parameters of the last layers (i.e., fully connected and classified). The second parameter-based TL method tuned all network parameters using the target domain, which definitely required more computational costs. The results showed that the first method (i.e., freezing the first layers and fine-tuning the last layers) is not efficient and has low accuracy when the sample data size does not change significantly. Moreover, the results showed that the second method (i.e., training all parameters) was effective with higher accuracy. The experiments of [71] demonstrated that the suggested model had good TL performance and achieved 97.73% diagnostic accuracy.

Wind turbines are widely installed at high altitudes with high humidity and low temperatures. This fact renders the system vulnerable to blade-icing scenarios, which can result in power loss and electrical failures, in addition to mechanical failures. In general, ice detection methods fall into three categories: meteorological, external, and SCADA [72]. Conventional observation systems use meteorological data to detect icing, which is considered a challenging task. On the other hand, external monitoring systems require the installation of extra equipment for wind turbines. A SCADA system is considered a relatively low-cost method for ice detection, and it has been widely used for fault detection. SCADA data include the ambient temperature as well as the temperatures of mechanical components. In addition, the data contain electrical operating information and control variables, which offer comprehensive operating conditions.

Zhang et al. [72] proposed an ice detection model using neural networks and TL. The proposed model precisely detects ice in wind turbines by learning from the labeled SCADA data of the source domain. The authors collected SCADA data from two wind turbines, W1 and W2, from a large wind farm. The collected data contained 26 operational wind parameters. The proposed model used an inductive TL, which can deal with the problem of having different target classification models and limited labeled data items in the target domain. The model was trained using the large, labeled SCADA dataset from the source domain related to W1. Then, the knowledge was transferred to the target domain related to W2, which was trained using few SCADA data. The requirements for inductive TL were satisfied because the source and target domains shared the same feature space and similar ice detection methods. However, the target ice detection models were not identical. The authors tested different fully connected neural networks. The best accuracy, stability, and performance were compared to other models (e.g., AdaBoost, Quadratic Classifier, and RF). Accordingly, their model outperformed other models. The results also showed that inductive TL based on a fully connected neural network has higher detection accuracy (14%) and is more stable (13% better) than a fully connected neural network without TL.

The authors in [73] developed a TL-based approach for wind turbine fault diagnosis to address the problem of lacking data, unbalanced data, complex fault types, and weak generalizability of deep learning models. The authors proposed an improved residual network ResNet to implement deep TL in the form of pre-training and fine-tuning. They claimed that the proposed deep transfer model was able to be sufficiently trained on source and target domain vibrational data. The performance of the proposed model was verified by performing fault diagnosis on bearings and gears data from different sources. The authors reported that the deep residual network has high transferability, high accuracy, and high diagnostic speed. The reported results showed that the optimal diagnostic accuracy of the suggested deep TL model can be above 90%.

Most of the reviewed works in the literature showed that TL has proven an ability to improve RE fault diagnosis performance, particularly in the cases of wind turbines as depicted in Table 4.

Research	Energy	Base Model	Transfer Type	Source Data	Target Data
Li et al. (2021) [63]	Wind	Convolutional autoencoder	Homogenous, parameter-based	Operational data of 14 wind turbine	Operational data of 1 wind turbine
Zhang et al. (2022) [73]	Wind	Deep ResNet	Transductive model-based	Bearing datasets + gears datasets	Bearing datasets + gears datasets
Zyout et al. (2020) [74]	Solar	Alexnet CNN	Heterogeneous, parameter-based	Large object recognition dataset	RGB images of solar panels
Akram et al. (2020) [75]	Solar	CNN	Heterogeneous, parameter-based	Infrared images	Electroluminescence images
Yan Zhang et al. (2020) [71]	Wind	CNN	Heterogeneous, parameter-based	Bearing dataset	Gear dataset
Zhang et al. (2018) [72]	Wind	Fully-connected NN	Inductive, parameter-based	SCADA dataset	SCADA dataset
Korkmaz et al. (2022) [76]	Solar	Multi-scale CNN (Alexnet variant)	Heterogeneous, parameter-based	Large object recognition dataset	Infrared images

Table 4. Summary of fault diagnosis transfer-learning approaches discussed in Section 3.3.

Research	Energy	Base Model	Transfer Type	Source Data	Target Data
Yang et al. (2021) [77]	Wind	Alexnet CNN	Heterogeneous, parameter-based	Large object recognition dataset	UAV images
Demirci et al. (2019) [78]	Solar	CNN	Homogenous, parameter-based	Electroluminescent imaging	Electroluminescent Imaging
Zhu et al. (2022) [79]	Wind	LSTM	Feature-based	Monitoring data	Monitoring data
Hou et al. (2021) [80]	Solar	CNN	Heterogeneous, parameter-based	ImageNet	Electroluminescent Imaging
Chatterjee et al. (2020) [81]	Wind	LSTM, XGBoost	Transductive, feature-based	SCADA dataset	SCADA dataset
Korkmaz et al. (2022) [76]	Solar	Multi-scale CNN	Homogeneous parameters-based	Thermographic images	Thermographic images
Chen et al.(2021) [1]	Wind	CNN	Parameter-based	SCADA dataset	SCADA dataset
Liu et al. (2022) [82]	Wind	DLL generative adversarial Nets GAN	Transductive feature-based	Wind turbine transmission dataset (unbalanced)	Wind turbine transmission dataset

Table 4. Cont.

B. TL for the Defect Detection in PV Solar Panel Surfaces

The increased adoption of PV power in the energy portfolio worldwide makes performance monitoring a critical task. PV systems may suffer from various defects, such as cracks, shadows, bird droppings, and dust. Different image-based approaches are usually used to identify faulty PV panels, including electroluminescence (EL) images and infrared (IR) thermography. Deep CNN is widely used for defect classification in PV images thanks to its good performance in computer vision tasks. In the literature, we noticed that many researchers have adopted TL to enhance the performance of the developed models, especially when the available dataset is relatively small, or the computational resources are limited. A possible approach is to fine-tune a pre-trained DNN such as Alexnet [73], where the model is trained on a large object recognition dataset. Another approach is to train a CNN model using a source dataset that is relatively similar to the target one, followed by a fine-tuning [74].

In [68], Zyout et al. applied a parameter-based TL to detect PV panel defects. The authors implemented a pre-trained Alexnet CNN architecture composed of 25 layers. A dataset consisting of 599 images for normal and different defects was prepared by applying preprocessing techniques (e.g., cropping and image resizing). The efficiency of the proposed model was reportedly reduced by 2% after applying augmentation techniques (e.g., scaling, reflection, flipping, and rotation). Afterwards, the final three layers were replaced with a two-neuron layer, a SoftMax layer, and a classification layer. The new layers were trained using the PV panel image dataset. Moreover, fine-tuning was applied to the momentum parameter and the initial learning rate. The obtained results confirmed the importance of using TL with an existing pre-trained network for the detection of defects in PV panels.

In [75], Akram et al. introduced two ML-based models to detect PV panel defects from infrared images: (1) an isolated CNN model trained from scratch, and (2) another similar model that includes parameters-based TL. The same CNN architecture was used for both models to provide a fair comparison. The dataset consisted of 893 IR images of normal and defective modules taken before and after the defect induction. The isolated model was trained with IR images, whereas the TL model was first trained using a similar dataset of EL images. The model was then fine-tuned using the IR dataset. With TL, the selected layers were retrained using the experimental results, while other layers were not changed. In addition to using TL with a model trained by a dataset of a similar domain (i.e., EL images), the authors fine-tuned the Visual Geometry Group-16 CNN using the ImageNet dataset. The experimental results showed that the TL model achieved better performance than the isolated model in terms of accuracy and computational costs.

Table 4 summarizes the approaches surveyed in Section 3.3. It compares the knowledge transfer types, the base ML models, and the source and target data. As shown in this table, researchers applied various techniques to transfer knowledge across the domains. Notably, parameter-based (also named model-based) TL is a commonly used approach thanks to its simplicity compared to other types. A model is pre-trained using the source data and fine-tuned to the target data. Moreover, CNN is a widely used baseline model thanks to its ability to extract high-level features and its flexibility for architectural extensions and modifications.

Table 4 indicates the dominance of DL models with parameter-based TL in the fault diagnosis of RE systems. This fact can be justified by the capability of deep TL in resolving the conflict between limited data and ML models' generalization ability. Most of the reported approaches for wind turbines fault diagnosis implemented the pre-training fine-tuning process.

4. Discussions and Challenges

In this section, we overview the main challenges and open problems that need to be approached and the perspectives that can be addressed in the near future. We also depict the main limitations that faced the authors during the preparation of this survey.

4.1. Research Trends and Challenges

TL has proved its ability to help in reducing several data scarcity issues in ML-based prediction and the classification of RE systems. However, numerous challenges and limitations are still to be addressed in this field. For example, the procedure of evaluating and selecting the source domain with enough data to make an optimal transfer of knowledge to a target domain not having enough data is still an open challenge. In general, and based on the surveyed works on TL for RE, the success of knowledge transfer, also known as domain adaptation, depends on the similarity level of the source and target domains/tasks. This fact implies the necessity of defining a common process to assess how much the source and target domains are similar before starting the transfer of knowledge. That is particularly essential for the cases of homogeneous domains or when the source and target domains are of different feature spaces' dimensions. Practically, a weak similarity between the source domain(s) and target domain may not guarantee a better accuracy of the target model, and in the worst cases, may lead to what is known as negative transfer [20,22]. This scenario may happen when the data used to train the source models are not similar enough to those available in the target domain. In this case, the target models may not perform well, and their performance could be worse than those of the source domain(s). This is one of the common limitations of spreading TL in the area of RE applications. It usually occurs when the training dataset in the D_T is extensive but has low similarity to the D_S .

Moreover, the limitation in the number of open-source benchmarking datasets prevents the fair evaluation of the newly suggested TL-based approaches for RE systems, especially in the generality and mobility of the suggested approach. Therefore, the comparison of these TL-based solutions becomes complicated and even impossible. The problem of scarcity of open-source benchmarking datasets from real-world RE applications may be the result of two main reasons:

- (1) Most of the proposed TL frameworks for RE systems are trained and evaluated using source and target data records that are collected from particular areas and under common weather conditions.
- (2) Most of the collected datasets are private and cannot be shared due to security reasons, especially for faults diagnosis and buildings' load forecasting systems. Most of the operators do not reveal enough information about the status of their systems. In addition, owners do not disclose enough information about the daily power consumption activities of their residential buildings, offices, and factories.

Furthermore, a convenient and fair comparison among TL-based solutions requires the adoption of unified metrics to measure the similarity between source and target domains. Most of the surveyed works adopt different evaluation metrics, which makes the comparison of their performances a difficult task. In addition, quantifying the benefit of applying a TL-framework on a specific task, particularly on an RE application, is also an open challenge that has not been adequately addressed yet. However, we can still find some research that has been initiated to find standard algorithms and measures to quantify the benefit of applying knowledge transfer, including the transfer loss/gain and the transfer ratio [83]. To the best of our knowledge, it is not known how those measures will perform with the TL-based models for RE systems, and whether they are more suitable for homogeneous and/or heterogeneous domains.

The following points highlight the main challenging research gaps that still need to be addressed:

- Further studies are required to define a reliable method to evaluate and select the right source domain(s) to make an optimal transfer of knowledge to a target domain not having enough data. This necessitates a robust method to quantify the similarity between source and target domains, and therefore avoid the negative transfer issue.
- 2. Few or no shared comprehensive benchmark datasets are available to evaluate and compare the already existing and the newly suggested TL-based approaches for RE applications.
- There are no common evaluation metrics to assess the performance of the proposed TL-based approaches for RE applications. The absence of a standard evaluation framework prevents the quantifying of the benefits provided by a TL-based approach to a classical ML-based one.
- 4. There is no clear procedure to determine the amount of data in the source domain necessary for a successful transfer of knowledge. This limitation becomes critical in cases where applications depend on seasonal data; e.g., weather and solar radiation data for RE forecasting, daily activity data for buildings' load forecasting, and weather conditions data for turbine faults predictions

All the above-mentioned factors and others limit the broad adoption of TL-framework in the three RE applications that we mentioned at the beginning of this survey. It is worth mentioning that most of the approaches reported in the literature are still in a research and scholarly context. Therefore, most of the above-overviewed challenges are estimated to become more obvious when TL-based approaches are deployed in real-world RE projects.

4.2. Future Directions and Perspectives

The application of TL techniques to RE systems and smart grids is an emerging research field, and the number of reported works in the literature has been increasing since 2016. However, research and implementation achievements in this topic are still in their early stages, and a lot of perspectives need to be addressed. For instance, little work has been focusing on the application of TL to buildings' load forecasting compared to other fields of RE reviewed in this paper. Furthermore, very limited successful applications of TL to RL-based buildings' load directions have been reported in the literature [57]. However, TL is crucial when adopting RL for smart buildings' general controls, as it is computationally demanding, and therefore resource- and time-consuming to train RL-based load predictors for each new building [61]. Currently, most TL applications with RL-based models are proposed for smart building energy management [57,84–86]. This may help in deploying more RL-based methods within the TL frameworks for cross-building energy load forecasting.

Among the research opportunities that may help in mitigating the open challenging questions in applying TL to RE systems in the near future, we can highlight the following:

- The availability of shared comprehensive datasets for each of the applications in RE is very helpful to evaluate and compare the newly suggested TL-based approaches in RE systems. In addition, preparing a standard evaluation framework with common evaluation metrics to compare and benchmark different TL-based approaches in RE will surely make a true impact in this direction.
- 2. A potential opportunity to improve the TL-based approaches for RE systems consists of preparing robust guidance about the common input features required in the learn-

ing datasets that is available in a source domain. That will help in increasing the benefits of applying TL to RE systems. Therefore, studies on quantifying the importance of a particular feature or a specific set of features in a source domain are needed. This will also help to improve the overall performance of a TL-based approach for all three topics mentioned in this survey.

- 3. In addition, the use of standard setups and metrics to evaluate the performance of available TL-based solutions will certainly make the comparison between those solutions in terms of efficiency, stability, and mobility a possible and easy task.
- 4. Reinforcement learning is currently used to optimize energy usage in buildings, and there are potential opportunities to use it within the TL-frameworks to predict energy consumption in buildings. This will definitely allow the development of load prediction models that can be efficient and quickly adapted to different buildings and environments.
- 5. Not all the suggested TL-based approaches in the literature are easily implemented in real-world RE projects since most of the reported challenges and limitations are estimated to become more obvious when deployed in real applications. Therefore, a comprehensive survey about the deployed TL-based approaches in RE systems and smart grids is adequate and beneficial to helping RE researchers identify the challenges in real-world scenarios.

All the above-mentioned opportunities for research perspectives in the near future necessitate the coordination and collaboration among scholars, industries, and business owners to discuss the real-world problems and challenges in order to make true advances. Currently, planners of energy systems believe that TL will play a key role in improving the efficiency of ML-based models in the next generation of RE systems and sustainable smart grids. Figure 8 depicts the mind map of the most common motivations, open challenges, and future opportunities in TL for RE systems.



Figure 8. General motivations, open challenges, and future directions in TL research for renewable energy systems.

4.3. Contribution and Limitations

This paper presents a survey on applying transfer learning in selected applications in renewable energy systems. It shows the sub-categories of TL that are the most commonly used in solar and wind energies: forecasting frameworks, buildings' load prediction solutions, and faults diagnosis in RE systems. Additionally, it identifies current challenges and future research opportunities. The main limitations that we have encountered in preparing this survey can be summarized as follows:

- (1) Among the related research that we found in the literature, few works provide details about the improvement one can obtain by adopting TL for RE forecasting, load forecasting, and fault diagnosis and prediction. Furthermore, researchers who reported the benefit provided by TL have used different performance metrics. This fact prevented a fair comparison among the different suggested solutions.
- (2) The limitation and even the absence of comparative analysis in most of the related TL applications for RE systems prevented scholars from concluding the cons and pros of each particular approach.

The limitation of computing resources and open-source benchmarking datasets prevented the authors of this work from validating some particular results in most of the surveyed research found in the literature.

5. Conclusions

Smart grids are increasingly gaining particular attention from scholars and researchers. A smart grid aims to maintain the production/consumption balance among several available sources of energy, including traditional fuel-based energy and sustainable forms of energy. This fact requires the design and implementation of reliable algorithms for RE forecasting, short and long-term load prediction of buildings, and faults prediction of RE systems. Currently, data-driven ML models in RE systems face major challenges due to (i) the scarcity of sufficient historical data to adjust the hyper-parameters of classification and prediction models, (ii) the lack of labeled data for training, and (iii) the limitation of computing resources to train models from scratch. This last challenge becomes more serious with DL-based models that require a high number of training data records. In addition, most of the learning algorithms of ML models assume that training and testing data belong to the same feature space or follow the same data distribution, which is not true in most real applications. All of these challenges are common for newly built wind farms, solar stations, and newly implemented sensors for buildings' power consumption. TL is a modern ML framework that allows the transfer of knowledge from source domains with enough annotated data to target domains not having enough data. In this work, we investigated the applications of TL to RE systems as an efficient framework that helps in mitigating several data scarcity challenges in traditional ML tools for RE systems. In addition, we identified the main advantages and challenges of using TL techniques for RE systems. Energy forecasting, cross-building prediction of energy consumption, and fault diagnosis detection and prediction were the key areas highlighted in this work. We also focused on the applied sub-settings of TL in the three particular applications of RE.

First, we highlighted the increased interest in applying the TL framework to RE systems through the recent research works in the literature between 2016 and 2022 inclusive. Then, we introduced TL as a powerful ML framework that has a range of potential applications in the field of sustainable energy systems. We also outlined the types of TL used in energy forecasting, prediction of energy consumption in buildings, and fault diagnosis. The studies we have reviewed suggest that TL methods can significantly improve the accuracy of the abovementioned tasks. We have presented several works about each of the three application areas, and we compared them based on TL sub-settings, models, and adopted feature types. Most of the suggested approaches in this field rely on features- and parameter-based TL for their simplicity and data availability constraints. In these research works, the transfer of knowledge is performed using training data from the same or similar regions, and in relatively small geographical areas. Such approaches still require more operational and utility testing so that they can be applied to wider regions with different weather and/or topographical characteristics. Furthermore, most of the model-based techniques of TL for RE assume that the trained models are of the same type. For example, in most of the surveyed approaches, the source and destination models are often configurations of feed-forward neural networks, LSTMs, and statistical-based models, whereas in real-life applications, source and destination domain models may differ widely and use heterogeneous data. For instance, a newly built model for solar radiation could be an LSTM, while another could be an SVR or any other deep recurrent model. Therefore, further investigations of TL among different models are needed.

Finally, we discussed the main challenges and open problems that still face researchers and scholars in applying a TL-framework for prediction and classification in RE systems. Then, we concluded with some potential perspectives in the near future that might help in improving the generalization of TL in order to expand its deployment in real world energy projects. The summary tables containing information about publications regarding TL application to RE applications showed that more than 85% of the publications were generated in the past four years, which further indicate the importance of this research topic. However, this ramp-up in such research activity is correlated with an ever-increasing divergence of methods. Hence, this literature survey is beneficial to researchers working in the field.

Author Contributions: Conceptualization, R.A.-H. and A.A.; Methodology, R.A.-H., A.A., B.N. and Z.A.B.; Validation, R.A.-H.; Formal analysis, B.N. and R.G.; Investigation, R.A.-H.; Writing—original draft, R.A.-H. and A.A.; Writing—review & editing, R.A.-H., A.A., B.N., R.G. and Z.A.B.; Visualization, R.A.-H., A.A.; Supervision, R.A.-H. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Informed Consent Statement: Not applicable.

Data Availability Statement: This work does not use any specific experimental dataset.

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

References

- Chen, W.; Qiu, Y.; Feng, Y.; Li, Y.; Kusiak, A. Diagnosis of wind turbine faults with transfer learning algorithms. *Renew. Energy* 2021, 163, 2053–2067. [CrossRef]
- Ahmed, W.; Hanif, A.; Kallu, K.D.; Kouzani, A.Z.; Ali, M.U.; Zafar, A. Photovoltaic panels classification using isolated and transfer learned deep neural models using infrared thermo graphic images. *Sensors* 2021, 21, 5668. [CrossRef]
- Lai, J.-P.; Chang, Y.-M.; Chen, C.-H.; Pai, P.-F. A Survey of Machine Learning Models in Renewable Energy Predictions. *Appl. Sci.* 2020, 10, 5975. [CrossRef]
- Carneiro, T.C.; De Carvalho, P.C.M.; Dos Santos, H.A.; Lima, M.A.F.B.; Braga, A.P.D.S. Review on Photovoltaic Power and Solar Resource Forecasting: Current Status and Trends. J. Sol. Energy Eng. 2022, 144, 1–84. [CrossRef]
- Al-Hajj, R.; Assi, A.; Fouad, M. Short-Term Prediction of Global Solar Radiation Energy Using Weather Data and Machine Learning Ensembles: A Comparative Study. J. Sol. Energy Eng. 2021, 143, 051003. [CrossRef]
- 6. Kusiak, A.; Verma, A. Prediction of Status Patterns of Wind Turbines: A Data-Mining Approach. J. Sol. Energy Eng. 2011, 133, 011008. [CrossRef]
- 7. Al-Hajj, R.; Assi, A.; Fouad, M.; Mabrouk, E. A Hybrid LSTM-Based Genetic Programming Approach for Short-Term Prediction of Global Solar Radiation Using Weather Data. *Processes* **2021**, *9*, 1187. [CrossRef]
- Duchesne, L.; Karangelos, E.; Wehenkel, L. Recent developments in machine learning for energy systems reliability managment. Proc. IEEE 2020, 108, 1656–1676. [CrossRef]
- Blitzer, J.; Dredze, M.; Pereira, F. Biographies, Bollywood, Boom-Boxes and Blenders: Domain Adaptation for Sentiment Classification. In Proceedings of the 45th Ann. Meeting of the Assoc. Computational Linguistics, Prague, Czech Republic, 25–26 June 2007; pp. 432–439.
- 10. Schreiber, J. 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* **2019**, arXiv:1906.01168.
- Hu, Q.; Zhang, R.; Zhou, Y. Transfer learning for short-term wind speed prediction with deep neural networks. *Renew. Energy* 2016, *85*, 83–95. [CrossRef]

- Zhou, S.; Zhou, L.; Mao, M.; Xi, X. Transfer learning for photovoltaic power forecasting with long short-term memory neural network. In Proceedings of the IEEE International Conference on Big Data and Smart Computing (BigComp), Busan, Republic of Korea, 19–22 February 2020; pp. 125–132.
- 13. Tasnim, S.; Rahman, A.; Oo, A.M.T.; Haque, E. Wind power prediction in new stations based on knowledge of existing Stations: A cluster based multi source domain adaptation approach. *Knowl.-Based Syst.* **2018**, *145*, 15–24. [CrossRef]
- Ji, L.; Fu, C.; Ju, Z.; Shi, Y.; Wu, S.; Tao, L. Short-Term Canyon Wind Speed Prediction Based on CNN—GRU Transfer Learning. Atmosphere 2022, 13, 813. [CrossRef]
- Schreiber, J.; Vogt, S.; Sick, B. Task Embedding Temporal Convolution Networks for Transfer Learning Problems in Renewable Power Time Series Forecast. In Proceedings of the Joint European Conference on Machine Learning and Knowledge Discovery in Databases, Bilbao, Spain, 19–23 September 2021; pp. 118–134.
- 16. Himeur, Y.; Elnour, M.; Fadli, F.; Meskin, N.; Petri, I.; Rezgui, Y.; Bensaali, F.; Amira, A. Next-generation energy systems for sustainable smart cities: Roles of transfer learning. *Sustain. Cities Soc.* **2022**, *85*, 104059. [CrossRef]
- Blitzer, J.; McDonald, R.; Pereira, F. Domain Adaptation with Structural Correspondence Learning. In Proceedings of the Conference on Empirical Methods in Natural Language, Sydney, Australia, 22–23 July 2006; pp. 120–128.
- Wu, P.; Dietterich, T.G. Improving SVM Accuracy by Training on Auxiliary Data Sources. In Proceedings of the Twenty-First International Conference (ICML 2004), Banff, AB, Canada, 4–8 July 2004.
- Cook, D.; Feuz, K.D.; Krishnan, N.C. Transfer learning for activity recognition: A survey. *Knowl. Inf. Syst.* 2013, 36, 537–556. [CrossRef]
- Wang, J.; Zheng, H.; Huang, Y.; Ding, X. Vehicle Type Recognition in Surveillance Images From Labeled Web-Nature Data Using Deep Transfer Learning. *IEEE Trans. Intell. Transp. Syst.* 2017, 19, 2913–2922. [CrossRef]
- 21. Zhuang, F.; Qi, Z.; Duan, K.; Xi, D.; Zhu, Y.; Zhu, H.; Xiong, H.; He, Q. A Comprehensive Survey on Transfer Learning. *Proc. IEEE* 2020, 109, 43–76. [CrossRef]
- Gopalakrishnan, K.; Khaitan, S.K.; Choudhary, A.; Agrawal, A. Deep Convolutional Neural Networks with transfer learning for computer vision-based data-driven pavement distress detection. *Constr. Build. Mater.* 2017, 157, 322–330. [CrossRef]
- 23. Pan, S.J.; Yang, Q. A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* 2009, 22, 1345–1359. [CrossRef]
- Baralis, E.; Chiusano, S.; Garza, P. A lazy approach to associative classification. *IEEE Trans. Knowl. Data Eng.* 2007, 20, 156–171. [CrossRef]
- 25. Weiss, K.; Khoshgoftaar, T.M.; Wang, D.D. A survey of transfer learning. J. Big Data 2016, 3, 1345–1459. [CrossRef]
- Niu, S.; Liu, Y.; Wang, J.; Song, H. A decade survey of transfer learning (2010–2020). IEEE Trans. Artif. Intell. 2020, 1, 151–166. [CrossRef]
- 27. Belkin, M.; Niyogi, P.; Sindhwani, V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *J. Mach. Learn. Res.* **2006**, *7*, 2399–2434.
- 28. Day, O.; Khoshgoftaar, T.M. A survey on heterogeneous transfer learning. J. Big Data 2017, 4, 29. [CrossRef]
- Lu, J.; Behbood, V.; Hao, P.; Zuo, H.; Xue, S.; Zhang, G. Transfer learning using computational intelligence: A survey. *Knowl.-Based Syst.* 2015, 80, 14–23. [CrossRef]
- Tan, C.; Sun, F.; Kong, T.; Zhang, W.; Yang, C.; Liu, C. A Survey on Deep Transfer Learning. In Proceedings of the 27th International Conference on Artificial Neural Networks, Rhodes, Greece, 4–7 October 2018; pp. 270–279. [CrossRef]
- Shao, L.; Zhu, F.; Li, X. Transfer learning for visual categorization: A survey. *IEEE Trans. Neural Netw. Learn. Syst.* 2014, 26, 1019–1034. [CrossRef]
- Liu, R.; Shi, Y.; Ji, C.; Jia, M. A Survey of Sentiment Analysis Based on Transfer Learning. *IEEE Access* 2019, 7, 85401–85412. [CrossRef]
- 33. Qureshi, A.S.; Khan, A. Adaptive transfer learning in deep neural networks: Wind power prediction using knowledge transfer from region to region and between different task domains. *Comput. Intell.* **2019**, *35*, 1088–1112. [CrossRef]
- Yin, H.; Ou, Z.; Fu, J.; Cai, Y.; Chen, S.; Meng, A. A novel transfer learning approach for wind power prediction based on a serio-parallel deep learning architecture. *Energy* 2021, 234, 121271. [CrossRef]
- Liang, T.; Zhao, Q.; Lv, Q.; Sun, H. A novel wind speed prediction strategy based on Bi-LSTM, MOOFADA and transfer learning for centralized control centers. *Energy* 2021, 230, 120904. [CrossRef]
- 36. Yang, X.-S. Multiobjective firefly algorithm for continuous optimization. Eng. Comput. 2012, 29, 175–184. [CrossRef]
- 37. Bo, H.; Niu, X.; Wang, J. Wind Speed Forecasting System Based on the Variational Mode Decomposition Strategy and Immune Selection Multi-Objective Dragonfly Optimization Algorithm. *IEEE Access* **2019**, *7*, 178063–178081. [CrossRef]
- Lin, Y.; Duan, D.; Hong, X.; Han, X.; Cheng, X.; Yang, L.; Cui, S. Transfer Learning on the Feature Extractions of Sky Images for Solar Power Production. In Proceedings of the IEEE Power & Energy Society General Meeting (PESGM), Atlanta, GA, USA, 4–8 August 2019; pp. 1–5.
- Cai, L.; Gu, J.; Ma, J.; Jin, Z. Probabilistic Wind Power Forecasting Approach via Instance-Based Transfer Learning Embedded Gradient Boosting Decision Trees. *Energies* 2019, 12, 159. [CrossRef]
- Abubakr, M.; Akoush, B.; Khalil, A.; Hassan, M.A. Unleashing deep neural network full potential for solar radiation forecasting in a new geographic location with historical data scarcity: A transfer learning approach. *Eur. Phys. J. Plus* 2022, 137, 474. [CrossRef]

- 41. Qureshi, A.S.; Khan, A.; Zameer, A.; Usman, A. Wind power prediction using deep neural network based meta regression and transfer learning. *Appl. Soft Comput.* **2017**, *58*, 742–755. [CrossRef]
- Cao, L.; Wang, L.; Huang, C.; Luo, X.; Wang, J.-H. A transfer learning strategy for short-term wind power forecasting. In Proceedings of the Chinese Automation Congress (CAC), Xi'an, China, 30 November 2018; pp. 3070–3075.
- 43. Liu, Y.; Wang, J. Transfer learning based multi-layer extreme learning machine for probabilistic wind power forecasting. *Appl. Energy* **2022**, *312*, 118729. [CrossRef]
- 44. Khan, M.; Naeem, M.R.; Al-Ammar, E.A.; Ko, W.; Vettikalladi, H.; Ahmad, I. Power Forecasting of Regional Wind Farms via Variational Auto-Encoder and Deep Hybrid Transfer Learning. *Electronics* **2022**, *11*, 206. [CrossRef]
- Song, J.; Peng, X.; Yang, Z.; Wei, P.; Wang, B.; Wang, Z. A Novel Wind Power Prediction Approach for Extreme Wind Conditions Based on TCN-LSTM and Transfer Learning. In Proceedings of the IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia), Shanghai, China, 8–11 July 2022; pp. 1410–1415.
- Oh, J.; Park, J.; Ok, C.; Ha, C.; Jun, H.B. A Study on the Wind Power Forecasting Model Using Transfer Learning Approach. *Electronics* 2022, 11, 4125. [CrossRef]
- 47. Schreiber, J.; Sick, B. Multi-Task Auto-encoders and Transfer Learning for Day-Ahead Wind and Photovoltaic Power Forecasts. *Energies* **2022**, *15*, 8062. [CrossRef]
- Yu, S.; Vautard, R. A transfer method to estimate hub-height wind speed from 10 meters wind speed based on machine learning. *Renew. Sustain. Energy Rev.* 2022, 169, 112897. [CrossRef]
- Geng, R.; Long, H.; Gu, W. Small-Sample Interval Prediction Model based on Transfer Learning for Solar Power Prediction. In Proceedings of the IEEE 5th International Electrical and Energy Conference (CIEEC), Nanjing, China, 27–29 May 2022; pp. 4739–4743.
- Tajjour, S.; Chandel, S. Power Generation Forecasting of a Solar Photovoltaic Power Plant by a Novel Transfer Learning Technique with Small Solar Radiation and Power Generation Training Data Sets. 2022, p. 4024225. Available online: https://ssrn.com/ abstract=4024225 (accessed on 11 December 2022).
- 51. Ribeiro, M.; Grolinger, K.; ElYamany, H.F.; Higashino, W.A.; Capretz, M.A. Transfer learning with seasonal and trend adjustment for cross-building energy forecasting. *Energy Build*. **2018**, *165*, 352–363. [CrossRef]
- 52. Mocanu, E.; Nguyen, P.H.; Kling, W.L.; Gibescu, M. Unsupervised energy prediction in a Smart Grid context using reinforcement cross-building transfer learning. *Energy Build.* **2016**, *116*, 646–655. [CrossRef]
- Ahn, Y.; Kim, B.S. Prediction of building power consumption using transfer learning-based reference building and simulation dataset. *Energy Build.* 2022, 258, 111717. [CrossRef]
- 54. Gao, Y.; Ruan, Y.; Fang, C.; Yin, S. Deep learning and transfer learning models of energy consumption forecasting for a building with poor information data. *Energy Build.* 2020, 223, 110156. [CrossRef]
- Jin, Y.; Acquah, M.A.; Seo, M.; Han, S. Short-term electric load prediction using transfer learning with interval estimate adjustment. *Energy Build*. 2022, 258, 111846. [CrossRef]
- 56. Pinto, G.; Wang, Z.; Roy, A.; Hong, T.; Capozzoli, A. Transfer learning for smart buildings: A critical review of algorithms, applications, and future perspectives. *Adv. Appl. Energy* **2022**, *5*, 100084. [CrossRef]
- 57. Wang, Z.; Hong, T. Reinforcement learning for building controls: The opportunities and challenges. *Appl. Energy* 2020, 269, 115036. [CrossRef]
- 58. Wiering, M.A.; Van Otterlo, M. Reinforcement learning. Adapt. Learn. Optim. 2012, 12, 729.
- 59. Sutton, R.S.; Barto, A.G. Reinforcement Learning: An Introduction; The MIT Press: Cambridge, MA, USA, 2018.
- 60. Lazaric, A. Transfer in Reinforcement Learning: A Framework and a Survey. In *Reinforcement Learning: State-of-the-Art;* Springer: Berlin/Heidelberg, Germany, 2012; pp. 143–173.
- 61. Taylor, M.E.; Stone, P. Transfer learning for reinforcement learning domains: A survey. J. Mach. Learn. Res. 2009, 10, 1633–1685.
- Gonzalez-Vidal, A.; Mendoza-Bernal, J.; Niu, S.; Skarmeta, A.F.; Song, H. A Transfer Learning Framework for Predictive Energy-Related Scenarios in Smart Buildings. *IEEE Trans. Ind. Appl.* 2022, 59, 26–37. [CrossRef]
- 63. Li, Y.; Jiang, W.; Zhang, G.; Shu, L. Wind turbine fault diagnosis based on transfer learning and convolutional autoencoder with small-scale data. *Renew. Energy* **2021**, *171*, 103–115. [CrossRef]
- Gray, C.S.; Watson, S.J. Physics of Failure approach to wind turbine condition based maintenance. *Wind. Energy* 2010, 13, 395–405. [CrossRef]
- 65. Qiu, Y.; Feng, Y.; Sun, J.; Zhang, W.; Infield, D. Applying thermophysics for wind turbine drivetrain fault diagnosis using SCADA data. *IET Renew. Power Gener.* 2016, *10*, 661–668. [CrossRef]
- 66. Bangalore, P.; Tjernberg, L.B. An Artificial Neural Network Approach for Early Fault Detection of Gearbox Bearings. *IEEE Trans. Smart Grid* **2015**, *6*, 980–987. [CrossRef]
- 67. Xiao, Y.; Kang, N.; Hong, Y.; Zhang, G. Misalignment Fault Diagnosis of DFWT Based on IEMD Energy Entropy and PSO-SVM. *Entropy* **2017**, *19*, 6. [CrossRef]
- Chen, Z.; Han, F.; Wu, L.; Yu, J.; Cheng, S.; Lin, P.; Chen, H. Random forest based intelligent fault diagnosis for PV arrays using array voltage and string currents. *Energy Convers. Manag.* 2018, 178, 250–264. [CrossRef]
- 69. Chang, Y.; Chen, J.; Qu, C.; Pan, T. Intelligent fault diagnosis of Wind Turbines via a Deep Learning Network Using Parallel Convolution Layers with Multi-Scale Kernels. *Renew. Energy* **2020**, *153*, 205–213. [CrossRef]

- 70. Zhao, H.; Liu, H.; Hu, W.; Yan, X. Anomaly detection and fault analysis of wind turbine components based on deep learning network. *Renew. Energy* **2018**, 127, 825–834. [CrossRef]
- Zhang, Y.; Liu, W.; Wang, X.; Gu, H. A novel wind turbine fault diagnosis method based on compressed sensing and DTL-CNN. *Renew. Energy* 2018, 194, 249–258. [CrossRef]
- 72. Zhang, C.; Bin, J.; Liu, Z. Wind turbine ice assessment through inductive transfer learning. In Proceedings of the IEEE International Instrumentation and Measurement Technology Conference (i2mtc), Houston, TX, USA, 14–17 May 2018; pp. 1–6.
- 73. Zhang, Y.; Liu, W.; Gu, H.; Alexisa, A.; Jiang, X. A novel wind turbine fault diagnosis based on deep transfer learning of improved residual network and multi-target data. *Meas. Sci. Technol.* **2022**, *33*, 095007. [CrossRef]
- Zyout, I.; Oatawneh, A. Detection of PV Solar Panel Surface Defects using Transfer Learning of the Deep Convolutional Neural Networks. In Proceedings of the Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 9 April 2020; pp. 1–4.
- 75. Akram, M.W.; Li, G.; Jin, Y.; Chen, X.; Zhu, C.; Ahmad, A. Automatic detection of photovoltaic module defects in infrared images with isolated and develop-model transfer deep learning. *Sol. Energy* **2020**, *198*, 175–186. [CrossRef]
- 76. Korkmaz, D.; Acikgoz, H. An efficient fault classification method in solar photovoltaic modules using transfer learning and multi-scale convolutional neural network. *Eng. Appl. Artif. Intell.* **2022**, *113*, 104959. [CrossRef]
- 77. Yang, X.; Zhang, Y.; Lv, W.; Wang, D. Image recognition of wind turbine blade damage based on a deep learning model with transfer learning and an ensemble learning classifier. *Renew. Energy* **2021**, *163*, 386–397. [CrossRef]
- Demirci, M.Y.; Beşli, N.; Gümüşçü, A. Defective PV Cell Detection Using Deep Transfer Learning and EL Imaging. In *Proceedings Book*; Van Yüzüncü Yıl University: Van, Turkey, 2019; p. 311.
- Zhu, Y.; Zhu, C.; Tan, J.; Tan, Y.; Rao, L. Anomaly detection and condition monitoring of wind turbine gearbox based on LSTM-FS and transfer learning. *Renew. Energy* 2022, 189, 90–103. [CrossRef]
- Hou, D.; Ma, J.; Huang, S.; Zhang, J.; Zhu, X. Classification of Defective Photovoltaic Modules in ImageNet-Trained Networks Using Transfer Learning. In Proceedings of the IEEE 12th Energy Conversion Congress & Exposition-Asia (ECCE-Asia), Singapore, 24 May 2021; pp. 2127–2132.
- Chatterjee, J.; Dethlefs, N. Deep learning with knowledge transfer for explainable anomaly prediction in wind turbines. *Wind. Energy* 2020, 23, 1693–1710. [CrossRef]
- 82. Liu, X.; Ma, H.; Liu, Y. A Novel Transfer Learning Method Based on Conditional Variational Generative Adversarial Networks for Fault Diagnosis of Wind Turbine Gearboxes under Variable Working Conditions. *Sustainability* **2022**, *14*, 5441. [CrossRef]
- Glorot, X.; Bordes, A.; Bengio, Y. Domain adaptation for large-scale sentiment classification: A deep learning approach. In Proceedings of the 28th International Conference on Machine Learning (ICML-11), Washington, DC, USA, 28 June 2011; pp. 513–520.
- Yu, L.; Qin, S.; Zhang, M.; Shen, C.; Jiang, T.; Guan, X. A Review of Deep Reinforcement Learning for Smart Building Energy Management. *IEEE Internet Things J.* 2021, *8*, 12046–12063. [CrossRef]
- Zhang, X.; Jin, X.; Tripp, C.; Biagioni, D.J.; Graf, P.; Jiang, H. Transferable reinforcement learning for smart homes. In Proceedings of the 1st International Workshop on Reinforcement Learning for Energy Management in Buildings & Cities, Virtual, 17 November 2020; pp. 43–47.
- 86. Zhu, Z.; Lin, K.; Jain, A.K.; Zhou, J. Transfer learning in deep reinforcement learning: A survey. arXiv 2020, arXiv:2009.07888.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.