

Transfer learning in very-short-term solar forecasting: Bridging single site data to diverse geographical applications

Liwenbo Zhang, Robin Wilson, Mark Sumner, Yupeng Wu *

Faculty of Engineering, University of Nottingham, Nottingham, NG7 2RD, United Kingdom

ARTICLE INFO

Keywords:

Solar energy
Very-short-term solar forecasting
Computer vision
Deep learning
Vision transformer
Transfer learning
Sky images

ABSTRACT

Over the past decade, the rapid growth of solar energy penetration has posed significant challenges for grid balancing and scheduling, heightening the need for accurate and efficient short-term solar forecasting. While deep learning models have shown promise in improving forecasting accuracy, previous studies have often focused on data from specific sites, limiting their generalisability across different climatic and geographical conditions. This study addresses this limitation by employing a multimodal self-attention deep model, trained under the dry and clear climate conditions of Folsom, California, and integrating various transfer learning techniques. We examine the transferability of this model to a new dataset from Nottingham, UK, characterised by humid and rainy conditions. Specifically, we compare different transfer methods based on model architecture and validate performance with limited target site data (equivalent to two weeks of data). The model's expertise can be effectively transferred, reducing the required data for successful model training by 80% (from four months to two weeks). Simulations under realistic scenarios demonstrate that the model, trained with just two weeks of data from the deployment site, achieved performance surpassing the baseline. This work demonstrates the feasibility of transferring deep learning models for solar forecasting across diverse climatic conditions, significantly reducing the data and time needed for model adaptation and deployment. This has the potential to enhance the reliability and efficiency of solar energy integration into power grids globally.

1. Introduction

Very-short-term (VST) solar energy forecasting based on ground-based fisheye images has been extensively studied and proven effective in predicting the rapid intermittency of solar irradiance [1–3]. This forecasting approach typically offers a time resolution of less than 10 min and a spatial resolution covering an area with a radius of 1 kilometre from the ground observation point. By accurately forecasting solar intermittency, control systems in solar power generation facilities, such as solar power plants or Building-Integrated Photovoltaic (BIPV) systems, can preemptively respond to upcoming Ramp Events (RE), thereby achieving higher power generation efficiency and improved power quality [4,5]. Additionally, precise short-term solar irradiance predictions can provide valuable control assistance to urban architectural shading systems, such as Photovoltaic Shading Devices (PVSD) [6, 7], and daylight harvesting systems requiring high-frequency control mechanisms [8]. This control aids in optimising energy utilisation and enhancing the overall sustainability of urban structures.

Deep learning models, particularly those employing computer vision techniques, have demonstrated immense potential in VST solar forecasting [9]. While feature engineering-based image analysis

methods, such as pixel statistics [10], optical flow, graphical cross-correlation [11], cloud cover indices, and cloud height methods, can also yield impressive forecasting capabilities [12], they often require complex, researcher-defined feature engineering and domain-specific expertise. Moreover, these methods are typically developed for specific sites, limiting their generalisability [13]. In contrast, machine-learned features have been shown to outperform researcher-defined features across various domains [14,15]. Similarly, deep learning models for very-short-term solar forecasting by using ground-based sky images (VST-DL-GSI-SolarForecast) have exhibited superior performance, faster computational speed, and higher generalisation potential [16]. These models employ a diverse range of techniques. For instance, some use Convolutional Neural Networks (CNNs) that focus solely on sky images for feature extraction [17,18], while others leverage 3D CNNs [19, 20] or Vision Transformers (ViT) [21,22]. Methods like Long Short-Term Memory (LSTM) networks concentrate on the temporal continuity of sky images [23,24], and ConvLSTM methods integrate both spatial and temporal features [20]. Multimodal approaches focus on fusion strategies for multiple types of data [25]. Additionally, some models

* Corresponding author.

E-mail addresses: liwenbo.zhang@nottingham.ac.uk (L. Zhang), yupeng.wu@nottingham.ac.uk (Y. Wu).

are specifically designed for the VST-DL-GSI-SolarForecast domain, incorporating data preprocessing strategies like feature segmentation and recombination [16,26], as well as algorithms that combine deep models with physical constraints [27–29].

However, the black-box nature of these deep learning models raises concerns about their reliability and stability, especially in terms of generalisability [9]. Furthermore, solar energy datasets inherently differ from traditional computer vision datasets due to their susceptibility to climate and geographical conditions [30]. This unique spatial characteristic amplifies the challenges of generalisability in deep learning models. The situation is further complicated by the fact that most existing models are trained on datasets collected primarily from meteorological stations, which often high degree of spatial dependence [16, 31,32]. i.e., the developed model might not be suitable for other locations solar forecasting due to different climatic conditions. Moreover, the high cost of data collection poses another significant barrier [33–35]. Deep learning models require diverse samples to achieve better generalisability. However, the bi-periodicity inherent to solar irradiance necessitates that a truly generalised dataset should cover at least one full year [1,36]. This requirement imposes substantial time, personnel, and financial constraints, making it impractical to train models from scratch for every new location.

Given the limited spatial coverage of VST-DL-GSI-SolarForecast models and the burgeoning potential for their widespread application as solar energy technologies advance, the need for localised deployment becomes increasingly urgent. Yet, the spatial variations introduce significant disparities in datasets, further complicating the already challenging task of model training. Therefore, innovative solutions are needed to address these multi-faceted challenges in solar energy forecasting.

In scenarios with limited or incomplete datasets, transfer learning emerges as a promising solution [37,38]. It has been shown to improve model performance across various tasks by leveraging knowledge from related but not identical datasets [39,40]. Transfer learning has also been identified as having the potential to address the challenges of limited data and data collection constraints in the domain of solar irradiance forecasting [38,41,42]. Objective performance has been observed in intra-hour [43], intra-day and day-ahead forecasts [38, 42] that employ transfer learning strategies. However, for very-short-term solar irradiance forecasting, especially those incorporating sky images, the feasibility of transfer learning strategies for different modalities, such as image-based information, remains an area that warrants comprehensive research and investigation [44].

To enhance the reliability of solar forecasting models developed for specific sites and extend their applicability to broader regions or locations, this paper integrates innovative transfer learning techniques. We first developed our model based on the dataset from urban or suburban LA Folsom (USA) and then adapted it for use with a newly collected urban Nottingham dataset (UK). This adaptation successfully navigates the inherent geographical, climatic, and equipment-related disparities between the two datasets. The key contributions of this paper are as follows:(1) The introduction of advanced feature and label adaptation techniques, designed to facilitate effective transfer learning across diverse geographical and climatic conditions, thereby enhancing the model's generalisability. (2) The development of a robust validation framework that employs both quantitative and qualitative metrics to assess the efficacy of transfer learning in the specific case of very-short-term solar forecasting. (3) Validation of the model's feasibility for rapid localised deployment in real-world scenarios, demonstrating its practical applicability and efficiency. In the following sections, Section 2 describes the specific methodology used for transfer learning, including the model architecture, transfer methodology and evaluation criteria. Section 3 describes the experimental setup, including the dataset and experimental setup. Section 4 details the prediction results and performs sensitivity analyses. Section 5 discusses limitations, future work, challenges and opportunities. Section 6 concludes the paper.

2. Methodology

This section introduces the development of generic deep learning model used for local solar forecasting and also for transfer learners in Section 2.1, the definition of transfer learning used in this work, the methodology, and the strategy in Section 2.2, and the methodology for forecasting performance and evaluating the effectiveness of transfer learning in Section 2.3. The data collected and used for the project will be introduced in Section 3. Fig. 1 provides an overview of the methodology including the development of the generic deep learning model, transfer learning and forecasting performance employed in this study for very short-term solar energy forecasting using a fusion of image and numerical data. This frame work includes two phases. The first phase, often referred to as the pre-training phase, serves as the foundation for the entire process. In simpler terms, this is where we initially train our deep learning model using the Folsom dataset from the United States as the 'source domain' [37]. The goal is to create a model that can accurately predict solar energy output based on the given image and numerical inputs. Once the model performs well in this initial setting, we extract its 'weight parameters,' which are essentially the learned features that make the model effective. The second phase is known as the transfer learning or fine-tuning phase. Here, we take the weight parameters learned from the first phase and apply them to the Nottingham dataset in the UK, serving as the 'target domain.' The idea is to leverage the knowledge gained from the initial training to give our model a head start in this new setting. This allows the model to adapt more quickly and perform better in the new domain than it would if we started training from scratch.

2.1. Deep learning model in very-short term forecasting formulation

In this study, given the involvement of diverse solar forecasting tasks, we represent the very-short-term solar forecasting model as follows. For a specific forecast, a domain-task representation can be employed. A domain D consists of two parts: the feature space \mathcal{X} and a marginal distribution \mathbf{X} , where \mathbf{X} signifies a set of instances. This can be mathematically represented as:

$$D = \{\mathcal{X}, P(\mathbf{X})\} \quad (1)$$

$$\text{where } \mathbf{X} = \{\mathbf{x}|\mathbf{x}_i \in \mathcal{X}, i = 1, \dots, n\} \quad (2)$$

A task, \mathcal{T} , consists of a label space \mathcal{Y} and a decision function F as:

$$\mathcal{T} = \{\mathcal{Y}, F(\mathbf{x}; \mathbf{W})\} \quad (3)$$

$$\text{where } F(\mathbf{x}_j, \mathbf{W}) = \{p(\mathbf{y}_k|\mathbf{x}_j; \mathbf{W})|\mathbf{y}_k \in \mathcal{Y}, k = 1, \dots, |\mathcal{Y}|\} \quad (4)$$

W are the internal weights of the deep model F , indicating how the model maps the input space to the latent space. W^* represents a well-trained model. During the mapping process, $p(\mathbf{y}_k|\mathbf{x}_j; \mathbf{W})$ denotes the conditional probability of input \mathbf{x}_j to output \mathbf{y}_k under weight \mathbf{W} . In this research, we adopted the ViT-E model from Zhang et al. [25] as the standard model for transfer learning. In this model, the inputs at a particular time instance include sky image data and numerical observation data, represented as:

$$\mathbf{x}_t = \{\mathbf{x}_{img_t}; \mathbf{x}_I; \mathbf{x}_{Env_t}; \mathbf{x}_{I_c}; \mathbf{x}_{W_{ind}_t}; \mathbf{x}_{SA_t}\} \quad (5)$$

where \mathbf{x}_{img} , \mathbf{x}_I , \mathbf{x}_{Env} , \mathbf{x}_{I_c} , $\mathbf{x}_{W_{ind}}$, and \mathbf{x}_{SA} represent image input, irradiance input, environmental input, clear sky irradiance input, wind parameter input, and solar angle input, respectively.

Considering the model performance diminishes as forecast time increases, we selected a 2-minute lead time as the forecasting horizon. Thus, the model's forecasted result, or the only element in the output space, is $\hat{\mathbf{y}}_{t+2\min} \in \mathbf{y}_k$. That is, the model's forecasting process is represented as:

$$\hat{\mathbf{y}}_{t+2\min} = F_{\text{ViT-E}}(\mathbf{x}_t, \mathbf{W}) \quad (6)$$

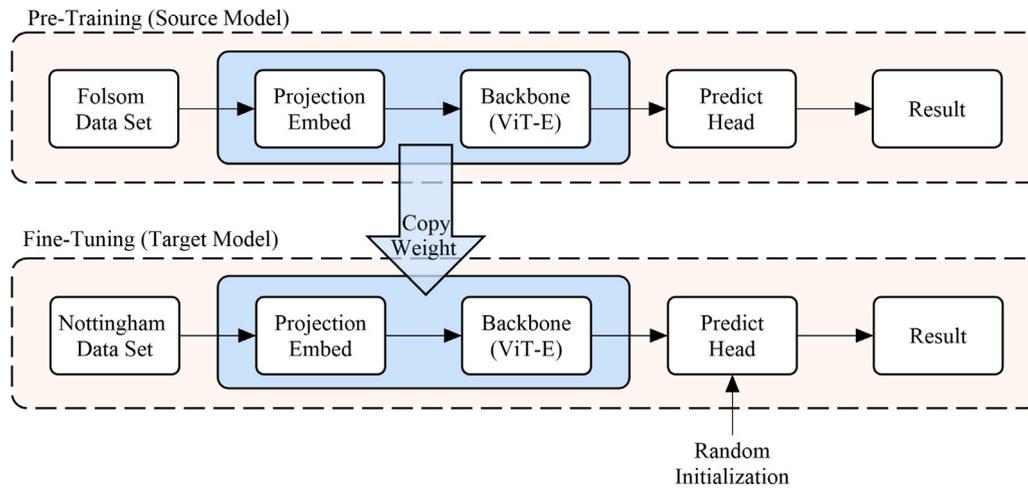


Fig. 1. The proposed transfer learning strategy. The beige framework is the model generic framework, which is presented in Section 2.1. The blue part is the transfer learning strategy, which is presented in Section 2.2.

It is worth noting that in this work, the forecasting target is the same as in [25], i.e., the Clear Sky Index (CSI), which is the current Global Horizontal Irradiance (GHI) relative to the clear sky GHI. The clear sky GHI is generated using the McClear [45] clear sky model.

More details of the model can be found in the work of Zhang et al. [25].

2.2. Transfer learning in solar irradiance forecasting

Transfer learning aims to enhance the performance of a target decision function, denoted as F_T , in a target domain by leveraging the knowledge encapsulated in a source domain. Formally, given a specific source domain and task (D_S, \mathcal{T}_S) , and a target domain and task (D_T, \mathcal{T}_T) , the objective is to improve F_T by utilising the learned features and representations from D_S and \mathcal{T}_S .

Inspired by the inductive transfer learning approach [46], this implies that labels are available in both the source and target domains. In our proposed deep learning-based very-short-term solar forecasting model, the transfer learning is categorised as homogeneous, signifying that the feature spaces and label spaces in both domains are consistent, i.e., $\mathcal{X}^S = \mathcal{X}^T$ and $\mathcal{Y}^S = \mathcal{Y}^T$. It is crucial to note that this assumption holds true only when the prediction target is the Clear Sky Index (CSI). The CSI normalisation process is assumed to effectively mitigate the impact of irradiance cycles and geographical factors on solar irradiance. Conversely, if the prediction target is solar irradiance itself, the outputs may differ due to meteorological and geographical variations, even under identical input conditions, i.e., $\mathcal{Y}^S \neq \mathcal{Y}^T$.

As illustrated in Fig. 1, the transfer learning process is executed through two main steps: weight transfer and fine-tuning. Initially, the model undergoes comprehensive pre-training on the source domain, specifically the Folsom dataset [33], utilising the ViT-E architecture. Upon completion of this phase, the optimised weights, denoted as W_S^* , are saved for subsequent use. These weights are then loaded into the target model, which is trained on the target domain, namely the Nottingham dataset, thereby facilitating knowledge transfer. Importantly, the transferred weights encompass only the projection embedding and the backbone layers, while the prediction head is initialised randomly for each new training session. The model then undergoes a full training cycle on the Nottingham dataset, using the pre-trained weights as an initialisation point. This subsequent training is commonly referred to as fine-tuning.

2.3. Evaluation metrics for solar irradiance forecasting

This section delineates the evaluation metrics employed in this study, encompassing both the assessment of model feature similarity and the quantitative and qualitative evaluation of model performance.

Assessment of model feature similarity. In this study, the similarity between different models' weights is gauged by comparing their respective feature vectors. Specifically, for models trained under identical architectures but with varying training datasets, the similarity is quantified based on the angle between the two feature vectors in a high-dimensional space. A smaller angle indicates higher similarity. This angle is computed using the cosine similarity metric, as formalised below:

$$\cos \Theta = \frac{\mathbf{a} * \mathbf{b}}{|\mathbf{a}| * |\mathbf{b}|}$$

A higher cosine similarity value, approaching 1, signifies that the feature vectors are more similar, and thus the angle between them is smaller. Conversely, a lower cosine similarity value, approaching -1 , indicates a larger angle and less similarity between the feature vectors.

Quantitative performance evaluation. The quantitative performance of the forecasting model is assessed using the Forecast Skill (FS) metric, a conventional method in the solar energy domain. This metric evaluates a model's performance relative to a baseline model, commonly the Smart Persistence Model (SPM). The SPM is a naive model that assumes no changes in meteorological indices during the forecast period. Specifically, it assumes that the Clear Sky Index (k_t), defined as the ratio of the current irradiance to the irradiance predicted by a clear sky model under current conditions, remains constant. The corresponding equation is:

$$\hat{y}_{\text{SPM}}(t + \Delta T) = k_t(t) \times y_{\text{clear}}(t + \Delta T)$$

$$k_t = \frac{I}{I_{\text{clear}}}$$

The McClear model serves as the clear sky model in this study. The FS metric is computed based on the above equation and offers a site- and dataset-agnostic parameter for model comparison. A positive FS value indicates superior performance compared to the SPM, with values closer to 1 suggesting that the model's predictions are nearing the ground truth.

$$\text{Forecast Skill} = \left(1 - \frac{RMSE_{\text{Model}}}{RMSE_{\text{Baseline}}}\right) \times 100\%$$

Qualitative performance evaluation. Recent studies have highlighted certain limitations of using FS as the sole predictive metric. Aiming for higher FS values can lead to conservative model behaviour, thereby neglecting transient solar phenomena such as solar ramp events (REs). REs are characterised by rapid changes in solar irradiance, often caused

by fast-moving clouds or other obstructions, leading to a sudden drop in solar irradiance by a significant percentage.

In this context, an RE is defined as a transient change in solar irradiance that reaches at least 10% of the current clear sky irradiance within a one-minute interval. REs can be further categorised based on their directionality into increasing or decreasing irradiance events. Consequently, each prediction has both a quantitative and a categorical component: it either corresponds to an upward RE, remains constant, or corresponds to a downward RE.

To capture the model's efficacy in predicting REs, this study employs the F1-score in a multi-class classification setting. The F1-score for each category is computed by first calculating the precision and recall. Due to the significantly lower frequency of RE events compared to constant events, an unweighted average of the F1-scores across all categories is used to provide a balanced performance measure. For each class i , the F_{1i} score is calculated as:

$$F_{1i} = 2 \times \frac{\text{Precision}_i \times \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}$$

where,

$$\text{Precision}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FP}_i}$$

$$\text{Recall}_i = \frac{\text{TP}_i}{\text{TP}_i + \text{FN}_i}$$

where, TP_i is the number of true positives for class i . FP_i is the number of false positives for class i . FN_i is the number of false negatives for class i .

The Balanced F_1 score is then calculated as the average of the F_1 scores for each class:

$$\text{Balanced } F_1 = \frac{1}{N} \sum_{i=1}^N F_{1i}$$

where N is the number of classes, which is 3 in this ternary classification example.

3. Experimental setup

This section outlines the experimental design for implementing transfer learning in very-short-term solar irradiance forecasting. We commence by conducting a comprehensive comparison between the source and target domain datasets, encompassing both meteorological numerical data and image data. Subsequently, we design a series of experiments to test and validate the efficacy of employing transfer learning under varying meteorological conditions, the impact of pre-training in the source domain on transfer learning, different transfer learning strategies, and the performance of transfer learning with limited target domain data. Two sites under different location and climatic conditions were selected for the proposed test: one is at Folsom dataset in the United States with more sunny and clear day over the whole year, the other one is a dataset in Nottingham, UK with more cloudy days annually.

The modelling was executed on a PC equipped with a 3.8 GHz AMD Ryzen 9 3900X CPU and a GeForce RTX 2080 SUPER GPU, utilising the Tensorflow [22] 2.5 platform with built-in Keras [47]. To mitigate errors inherent to the modelling process, such as the randomisation of observation order in mini-batch calculations and the use of a random number generator during training, the results presented are averaged over five repeated trials for each image model.

3.1. Dataset and data comparison

In this work, we utilised two datasets. One is a dataset collected explicitly for solar forecasting, released by the University of California, San Diego (UCSD) [33] This dataset covers data from 2014 to 2016 with a 1-minute interval, including sky images, solar irradiance data, and

Table 1
Summary of differences in meteorological data information between datasets.

	Folsom	Nottingham
Longitude and latitude	38.642°N 121.148°W	52.952°N 1.184°W
Köppen climate classification	Csa	Cfb
GHI measuring instruments	LI-200SZ Pyranometers	Calculated
DNI measuring instruments	Calculated	RaZON+ PH1 Pyrheliometer
DHI measuring instruments	LI-200SZ Pyranometers	RaZON+ PR1 Pyranometer
Classification to ISO 9060:1990	~ ± 5% Typical error compare to First class	Second class
Data set size	656k	96k
Duration of data set collection	3 Years	6 Months
Train/val/test set size	21K/25K/23K	58K/19K/19K

meteorological data. The other dataset is a local Nottingham dataset, collected based on the solar observation equipment at the University of Nottingham. The data collection started in November 2021 and includes sky images, solar irradiance data, meteorological data, solar spectral data, and PV output data.

This subsection delves into the disparities and similarities between the source domain, represented by the Folsom dataset, and the target domain, represented by the Nottingham dataset. Although both datasets encompass the same categories of data, they exhibit significant variances in specifics, attributable to differences in observation equipment, installation standards, and geographical locations. The comparative analysis focuses on the geographical impact on meteorological features, the congruencies and divergences in data collection instruments, and an evaluation of image data features in relation to the accuracy and distribution of meteorological data.

3.1.1. Comparison in meteorological data

This part of the subsection specifically examines the meteorological data contained in both datasets. The aim is to identify how geographical factors influence meteorological variables and to assess the quality and distribution of the data collected. This will provide a foundation for understanding the challenges and opportunities associated with applying transfer learning techniques for solar irradiance forecasting in different geographical settings.

The geographical locations of the observation stations inherently influence the sample distribution of the datasets. Table 1 presents the geographical conditions at the Folsom and Nottingham data collection sites. For instance, the Folsom dataset from 2015 and the Nottingham dataset from 2022 exhibit significant monthly data collection disparities, as depicted in Fig. 2(a). This discrepancy is primarily due to the latitude-induced variations in daylight duration; the UK dataset experiences elongated summers and truncated winters. The effective viewing angle limitation of the fisheye lens (Solar Zenith Angle, SZA, less than or equal to 75 degrees) further amplifies this difference. In Nottingham, for example, the solar zenith angle seldom surpasses 75 degrees in December, leaving only 87 samples post-quality control for the entire month.

Climatic classifications also contribute to these differences. Folsom falls under the Csa category (C = temperate climate, s = dry summer, a = hot summer) in the Köppen climate classification, whereas Nottingham is categorised as Cfb (C = temperate climate, f = no dry season, b = warm summer). These climatic distinctions manifest in the datasets, as illustrated in Fig. 2(b) and (c). In Fig. 2(b), Folsom generally exhibits higher monthly irradiance levels than Nottingham.

Table 2
Summary of differences in image data between the data sets.

	Folsom	Nottingham
Camera model	Vivotek FE8171V	MOBOTIX Q25
Sensor size	1/2" CMOS	1/1.8" CMOS
Original resolution	2480 × 1536	3072 × 2048
Aperture size	f/2.8	f/2.0
Light sensitivity	1.17 Lux	0.1 Lux
Output resolution	1536 × 1536	1028 × 1028
Orientation	15 degrees west of north	Due West
Sun marker	Yes	No

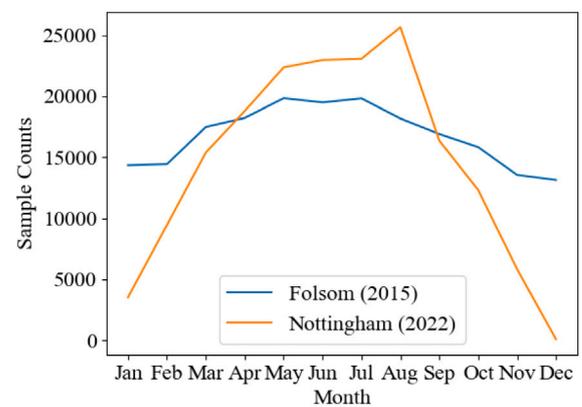
Even during summer, when daylight duration and solar angles are marginally lower in Folsom, the arid and hot climate conditions yield higher average irradiance values despite fewer extremes. Fig. 2(c) reveals that the GHI density distribution in Folsom is relatively uniform, whereas Nottingham displays a higher density of low irradiance values, attributable to its wetter and cloudier climate. The two-dimensional visual heat maps in Fig. 3 corroborate these trends. Nottingham's data lacks year-round continuity, particularly for Direct Normal Irradiance (DNI), indicating frequent cloud cover interruptions. Conversely, Folsom's dry summer conditions facilitate consistent and uninterrupted sunlight, as evidenced by minimal obstructions in the summer of 2016.

Beyond data distribution, the instrumentation for solar irradiance and meteorological data collection varies between the two sites. The Folsom site employs two LI-200SZ Pyranometers [48] for GHI and Diffuse Horizontal Irradiance (DHI) data collection, subsequently calculating DNI based on these measurements and solar angles [33]. In contrast, the Nottingham site utilises the Razon+ automatic solar tracker [49] to gather DHI and GHI data, with GHI values derived from the sum of solar angles. Notably, the pyranometers at these sites differ significantly in accuracy. The LI-200SZ, a photodiode pyranometer, has a GHI measurement error of 4.4% under clear sky conditions [50], aligning with the typical 5% error according to ISO 9060:2018 First Class standards [33]. In contrast, the thermopile pyranometers and pyrhemometers PR1 and PH1 have clear sky GHI errors of 0.3% and 0.03%, respectively [49]. According to ISO 9060:2018 [51], they qualify as Second Class instruments with typical errors under 1%. The authors note that although calibration can improve the LI-200SZ's accuracy, its error remains an order of magnitude higher than thermopile pyranometers due to its inherent limitations in the low-error field of view (60 degrees).

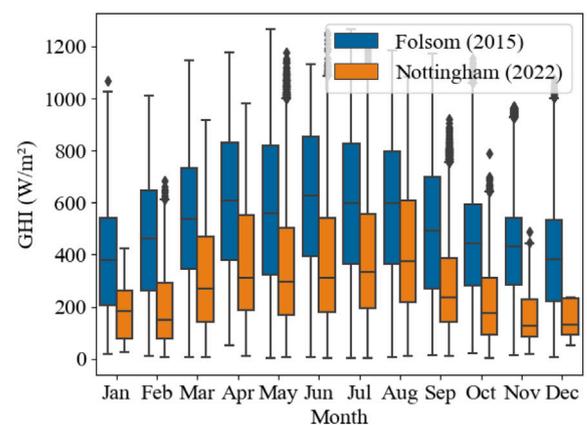
3.1.2. Comparison in image data

This section delves into the disparities and commonalities between image samples from the two datasets. As delineated in Table 2, the cameras employed in the Folsom and Nottingham datasets differ in specifications. The camera used for the Nottingham dataset boasts superior attributes, including higher original pixel resolution, a larger aperture, and enhanced light sensitivity. Significant differences in the actual quality of image output warrant discussion, as exemplified in Fig. 4. In the Nottingham dataset, intense sunlight often leads to overexposure in the solar region, thereby obfuscating details in this area. Conversely, the Folsom dataset mitigates this issue by blackening the fully overexposed (RGB values at 255) solar region. This treatment allows for the discernment of whether the sun is directly visible. For example, as depicted in the Cloudy Sky image on the right side of Fig. 4, the Folsom dataset retains information about the sun's position and its visibility through thin clouds. Such information is entirely absent in the Nottingham dataset. Attempts to ascertain the solar azimuth in the Nottingham dataset using similar methods were impractical due to camera quality limitations and extensive overexposure.

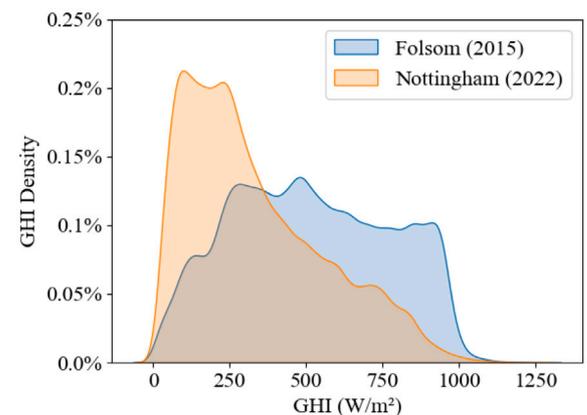
Another point of interest is the noticeable image noise in the Folsom dataset, particularly evident in the bottom-left image of Fig. 4. This



(a) Data volume of raw data.



(b) Distribution of monthly data GHI.

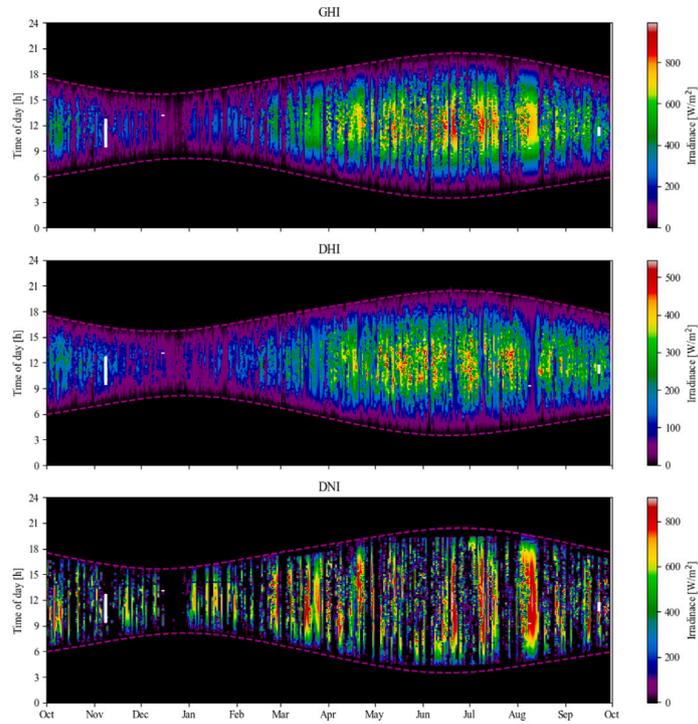


(c) Density distribution of annual GHI data

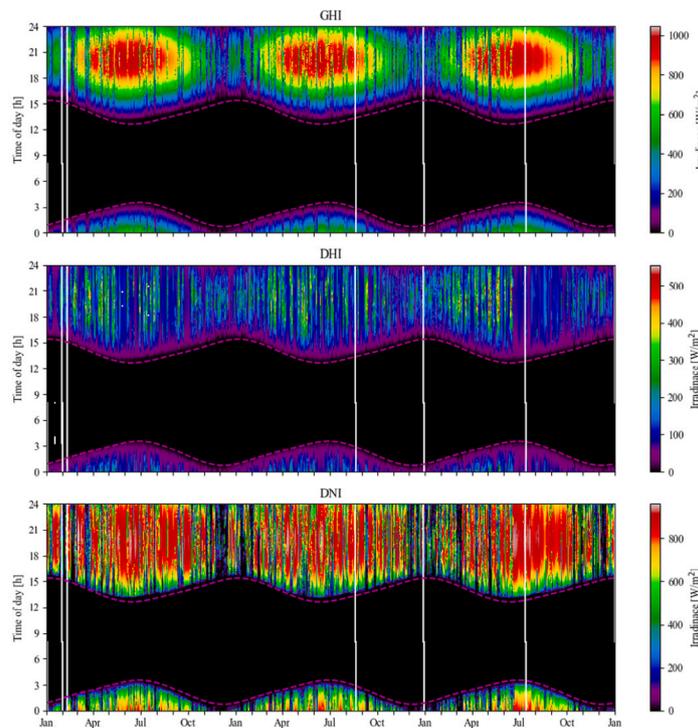
Fig. 2. Comparison of sample size and distribution between the Folsom dataset and the Nottingham dataset, for 2015 and 2022, respectively.

noise is speculated to arise from the degradation of the camera's transparent protective shell. The image distinctly reveals noise induced by sunlight refraction on the protective shell and lens flare reflections originating from the shell's imperfect transparency. This issue is especially pronounced on sunny days.

In summary, the divergences between the Folsom and Nottingham datasets can be primarily attributed to geographical locations



(a) Visual heat maps of Nottingham dataset, from 2021 Oct. to 2022 Oct.



(b) Visual heat maps of Folsom dataset, from 2014 to 2016.

Fig. 3. Two-dimensional visual heat maps of Folsom and Nottingham dataset (UTC Time Zone). Figure generated by Pvlib python library [52].

and the observational equipment employed. Geographical factors contribute to the variations in meteorological features, while differences in equipment lead to disparities in data accuracy and distribution. Comprehending these nuances is crucial for the effective application of transfer learning techniques and for making accurate performance comparisons between models trained on different datasets.

3.2. Experiment setup

3.2.1. Experiment 1: Assessing the efficacy of transfer learning

Experiment 1 primarily validates the efficacy of transfer learning. Here, efficacy refers to the preservation of model weight features from the source task during the transfer learning process, rather than being

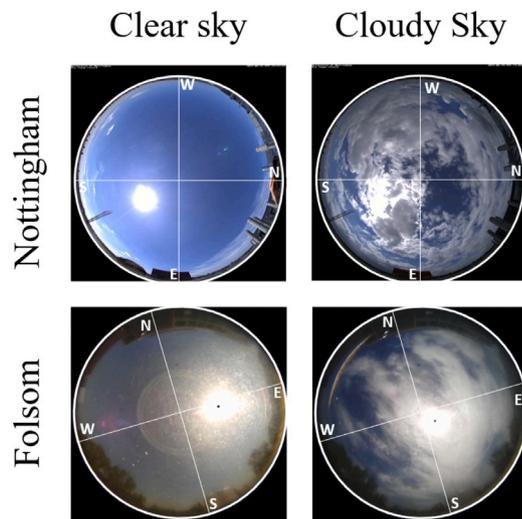


Fig. 4. Schematic of the sky images for the two data sets under different weather conditions.

entirely overwritten during training. We exploit the inherent randomness in model training to design this experiment. During the repeated training of the source-target model, the inevitable randomness in model training results in subtle variations in the trained model weights. We use the vector distribution of a specific structure, namely the output vector of the backbone network, as an explicit representation of the model feature space. We employ cosine similarity, as described in the evaluation metrics section, to compute the vector distribution across different models.

The specific experimental design is as follows. First, the model undergoes five independent repeated experiments in the source domain, yielding five sets of pre-trained weights for transfer from the source task, along with extracted feature vectors from the source task. Subsequently, five independent transfer experiments are conducted, resulting in five target domain task models and their corresponding feature vectors. By comparing the distribution of feature vectors between the source and target tasks, we assess whether the transfer training has inherited the model features from the source domain. A comprehensive analysis of the efficacy of transfer learning is then performed by comparing the performance of the five models from the source tasks with their corresponding models in the target tasks.

3.2.2. Experiment 2: Evaluating different fine-tuning strategies in transfer learning

Experiment 2 primarily investigates the impact of various fine-tuning strategies on transfer learning, as shown in Fig. 5. Multiple methods are available for fine-tuning the model in the target domain [37]. In this study, we employ a feature space adaptive method. This method initialises the target task by reusing the source task weights and subsequently adapts these weights in the target domain using a lower learning rate. Additionally, before the fine-tuning process begins, different modules can be frozen. Here, ‘freezing’ refers to fixing the source task weights to prevent further training, a strategy considered to preserve the inference mechanisms from the source task without additional adaptation to the target domain. Moreover, given that the ViT-E model’s inference module comprises two parts—the projection embedding and the backbone interface layer (i.e., Transformer Encoder)—we separately test the freezing of different components to facilitate a detailed comparison.

3.2.3. Experiment 3: Limited target domain dataset

One potential advantage of transfer learning is that when the size of the source domain dataset is larger than that of the target domain

Table 3

Details of the setup of each experiment in this work.

	Source model used	Target dataset size	Layers subject to freezing
Experiment 1	5 Trained models	55k	–
Experiment 2	Best model from Exp. 1	55k	No layers frozen Projection layer frozen Inference layer frozen All layers frozen
Experiment 3(a)	Best model from Exp. 1	25k to 2.5k	No layers frozen Projection layer frozen Inference layer frozen All layers frozen
Experiment 3(b)	Best model from Exp. 1	6.4k (Mar 4–18)	No layers frozen All layers frozen
	Best model from Exp. 1	9.6k (Jun 7–21)	No layers frozen All layers frozen
	Best model from Exp. 1	6.8k (Sep 9–23)	No layers frozen All layers frozen

dataset, the transferred knowledge can still achieve better generalisation in the target domain with a limited dataset. To verify this, we further reduced the size of the dataset. Specifically, we consider that in a complete one-year dataset, the trend of improving the model by enriching the dataset size begins to slow down starting from a dataset size of 25K samples [20]. In other words, the model’s generalisation reaches a bottleneck at 25K samples, making it improve continues more difficult. However, due to limitations in data collection and data quality, the Nottingham dataset actually only contains six months of data, so we still used the entire dataset for fitting experiments. Overall, we started with the entire Nottingham dataset (a total of 55K samples), randomly downsampled the dataset size to 25K, and further randomly downsampled to 12.5K, 7.5K, 5K, and 2.5K. We compared transfer learning with learning from scratch, testing the fitting performance of transfer learning in datasets with insufficient sample sizes.

In addition considering the limitations of random downsampling. This approach has a limitation in that random sampling preserves the diversity of the dataset to the maximum extent, thereby enabling the model to achieve better performance. Therefore, we used another downsampling method, namely, random continuous sampling. We randomly extracted three consecutive 14-day subsets from the original dataset as the dataset, simulating whether the model can achieve training with extremely limited dataset sizes through transfer learning in real situations. Specifically, we used data segments starting from March 4, June 7, and September 9, 2022, as training sets, without applying clear sky filters to the data. Ultimately, the three datasets contained 6.4K, 9.6K, and 6.8K data points, respectively, as shown in Fig. 6.

A summary of the experimental setups employed in this paper is presented in Table 3. Experiment 1 validates the effectiveness of transfer learning and identifies the best-performing model among five source models for subsequent transfer. Experiment 2 explores different transfer strategies based on the best-performing model identified in Experiment 1. Experiment 3(a) tests the efficacy of transfer learning under limited data conditions achieved through random sampling. Experiment 3(b) investigates the performance of transfer learning in real-world scenarios with limited datasets, specifically using three different continuous 14-day samples.

4. Results

4.1. Effectiveness of transfer learning

Results for the criteria used to evaluate the effectiveness of transfer learning of the ViT-E model under source domain, Folsom Dataset and target domain, Nottingham dataset are summarised in Figs. 7 and 8.

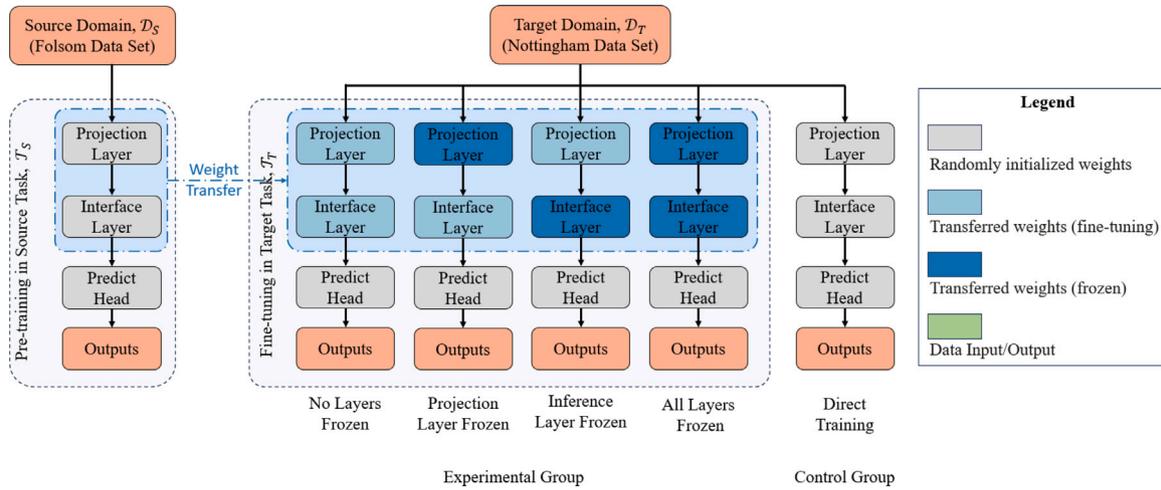


Fig. 5. Architectural overview of transfer learning from source to target domain, comparing direct transfer of frozen, fine-tuned, and randomly initialised weights against training only on the target dataset.

The results of the model effectiveness verification are shown in Fig. 7. The cosine similarity is calculated based on the angle between feature vectors, so the closer the result is to 1, the higher the similarity. Fig. 7(a) shows the cosine similarity between 5 source domain models, #1 to #5. It can be observed that despite the inevitable randomness in the training process, the impact of the model's randomness on the final trained model is minimal. All models exhibit consistency in the latent semantic space.

Fig. 7(b) shows the similarity between the target domain models and the source domain models after further training on the target domain dataset using the five source domain models. As can be seen from the figure, on the one hand, the target domain models after transfer learning are mostly more similar to their corresponding source domain models. On the other hand, the diversity trend between source domain models is preserved after transfer learning. For example, the high similarity between $y_{\#2}^S$ and $y_{\#4}^S$ is consistent with the similarity between the post-transfer learning $y_{\#2}^T$ and $y_{\#4}^T$.

Fig. 8 illustrates the influence of performance gaps, resulting from the randomness of 5 distinct source domain models, on the target domain models during transfer learning. Fig. 8(a) presents a comparison between the performance of source and target domain models. As evident in the figure, the impact of randomness on model performance is comparable across different domains, with an approximate error of 2%. Furthermore, the source and target domain models exhibit no consistency in performance. For instance, the top-performing model #3 in the source domain lags after transfer learning, whereas the underperforming model #4 in the source domain excels in the target domain. Fig. 8(b) highlights the influence of source domain model performance on transfer learning duration.

Notably, compared to target domain model performance, the transfer training duration exhibits a stronger correlation with source domain model performance. The top and bottom-performing models #3 and #2 in the source domain demonstrate a significant difference in retraining time. Despite model #3 in the source domain having a performance gap of 2.3% compared to model #2, its transfer learning time is reduced by one-third.

4.2. The impact of different fine-tuning strategies and dataset sizes in transfer learning

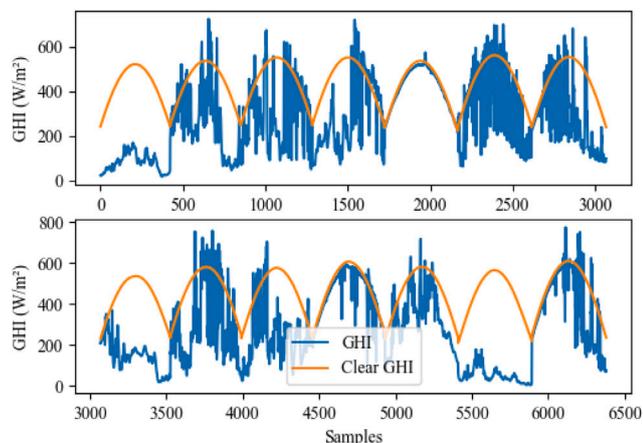
Results for the criteria used to evaluate the performance transfer learning of the ViT-E model under source domain, Folsom Dataset and target domain, Nottingham dataset are summarised in Fig. 8. The results are based on five repetitions of the experiment with the source task model #3 in the previous section as the pre-training weights.

Fig. 9(a) present the performance of models in terms of FS across various dataset sizes and transfer learning strategies. In Experiment 2, we evaluated multiple transfer learning strategies on a complete dataset consisting of 55,000 data points. The results reveal that the group that adapted the inference layer to the target domain achieved a performance comparable to training a new model from scratch. Conversely, groups that froze the inference layer experienced a decline in performance to varying extents. Remarkably, the group that completely froze all layers – essentially inheriting the source model's inference mechanism while only training the prediction head – saw a significant performance drop. Despite this setback, its FS still outperformed models based on CNNs, which based on single-location trained model [25].

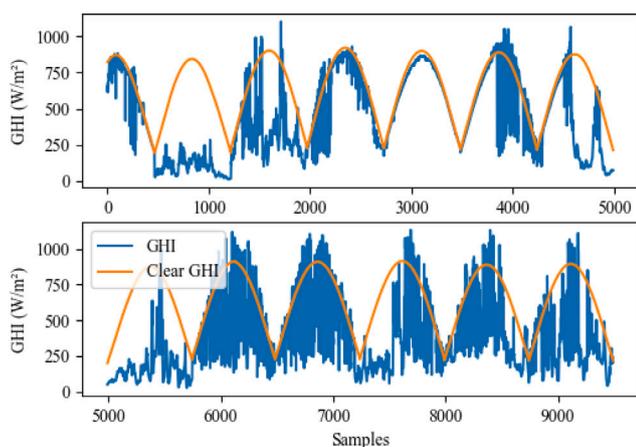
In Experiment 3(a), the data shows a clear trend: as the dataset size decreases, the FS of the models correspondingly declines. Most notably, the strategy of training new models from scratch without leveraging transfer learning becomes increasingly unstable and fails to maintain model accuracy when the dataset is limited. For instance, when the sample size is reduced to 7.5k, models trained from scratch display a significantly lower FS compared to those utilising transfer learning, with some even showing a drastic decline. As the sample size shrinks further to 5k, the FS of models trained from scratch begins to deteriorate noticeably. Although some models still manage to perform reasonably well due to the inherent randomness in the training process, the overall stability of the training becomes compromised.

On the other hand, even when the sample size is as small as 2.5k, all four transfer learning methods manage to maintain relatively high performance levels, despite the general decline in model performance. Intriguingly, the method that freezes all layers while training on the 2.5k sample set even surpasses the performance of a CNN-LG model trained on the full dataset. When comparing different transfer learning methods, Feature Space Adaptation techniques – such as not freezing any layers or only freezing the projection layer while keeping the core inference layer unfrozen – yield performance levels similar to training from scratch when ample data is available. In contrast, methods that lock the core inference layer perform relatively poorly when data is abundant. However, an exception is observed with the 2.5k dataset, where the two methods that fully inherit the source domain model's core inference layer achieve the best performance. We speculate that this may be because the smaller dataset is insufficient for the model to learn the underlying patterns in the target domain during adaptation. Consequently, models that completely freeze all layers, thereby avoiding any learning adaptation in the target domain, achieve superior performance.

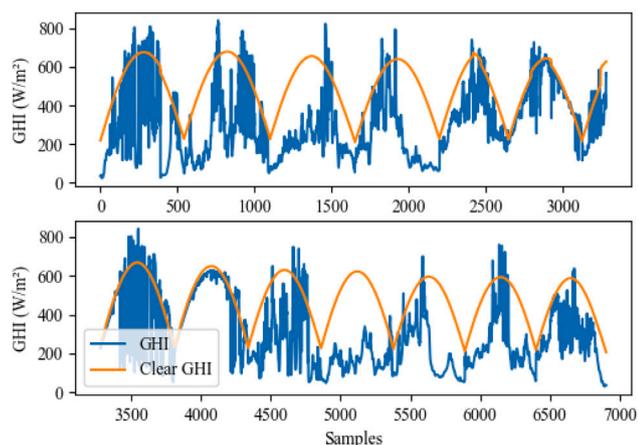
Fig. 9(b) present the F_1 scores of the model in the context of RE forecasting. Unlike the FS, which is directly optimised through loss



(a) Irradiance and Clear sky Irradiance during Mar 4th to 18th.



(b) Irradiance and Clear sky Irradiance during Jun 7th to 21st.



(c) Irradiance and Clear sky Irradiance during Sep 9th to 23rd.

Fig. 6. Three consecutive two-week datasets of the downsampled dataset in the mock-up experiment.

function constraints, the F_1 score is derived from further computations on the regression forecasts. Consequently, its behaviour is less regular than that of FS, particularly in relation to dataset size. However, the figure still reveals a discernible trend: akin to FS, the F_1 scores of the

model also experience varying degrees of decline as the dataset size decreases. The strategy of freezing all layers exhibits a performance pattern similar to that observed for FS; it underperforms when trained on the complete dataset but shows improved results when the data is limited.

Fig. 10 delineates the computational time required for training the models under various conditions. Generally, using smaller datasets can expedite the training process, albeit at the expense of model performance. Conversely, transfer learning strategies that freeze the core inference layer can substantially reduce training time, especially when larger datasets are employed. For instance, with a dataset comprising 55,000 samples, freezing just the inference layer can cut training time by approximately 10%. Freezing both the inference and projection layers can yield even greater time savings, reducing the training duration by around 40%. When the dataset size shrinks below 12,500 samples, all transfer learning approaches lead to shorter training times. Specifically, at a dataset size of 2500 samples, employing transfer learning methods reduces the training time to less than an hour, compared to the three hours needed for training a model from scratch—resulting in a time-saving of over 60

Another noteworthy observation is that the use of feature adaptation methods to continue training the inference module actually requires more computational time on the original dataset than training a new model from scratch. This suggests a trade-off between training time and model performance when employing different transfer learning strategies.

4.3. Performance under simulated real-world conditions

Fig. 11 illustrates the training outcomes under conditions designed to emulate real-world scenarios. For a meaningful comparison of model performance, we used the 7.5K-sized dataset from Experiment 3 as a benchmark, given its similar data volume to the three test datasets. The results reveal that models trained from scratch on these simulated datasets exhibit inconsistent performance. Specifically, the model trained on the two-week dataset from March yields an FS of approximately 4%, which is only comparable to the performance of the Standard Persistence Model (SPM).

In contrast, models trained via transfer learning on the two-week dataset from June demonstrate performance metrics closely aligned with those achieved using the full-year downsampled dataset. Notably, the June dataset contains approximately 30% more data points than the datasets from March and September, totalling 9.6K data points. This discrepancy is largely attributed to the longer duration of sunlight during June.

Furthermore, the solar zenith angle (SZA) is hypothesised to influence the diversity of the captured sky images. For instance, the minimum SZA in the March dataset is considerably high at 54 degrees, compared to 29 degrees for June and 47.7 degrees for September. Lower SZA values indicate that the sun is positioned closer to the centre of the image, thereby affecting the diversity of solar positions in the sky images. Consequently, the March dataset exhibits limited diversity in terms of sun positions, which could be a contributing factor to the model's suboptimal performance when trained on this dataset.

5. Discussion

This study delves into the intricacies of transfer learning within the VST-DL-GSI-SolarForecast framework, shedding light on its implications for model performance, training duration, and data prerequisites. Our findings underscore the efficacy of transfer learning in enhancing the ViT-E model's performance, especially when confronted with limited training data. The data visualisations in Figs. 7, 8, and 9 elucidate the advantages of transfer learning, particularly when dealing with constrained datasets.

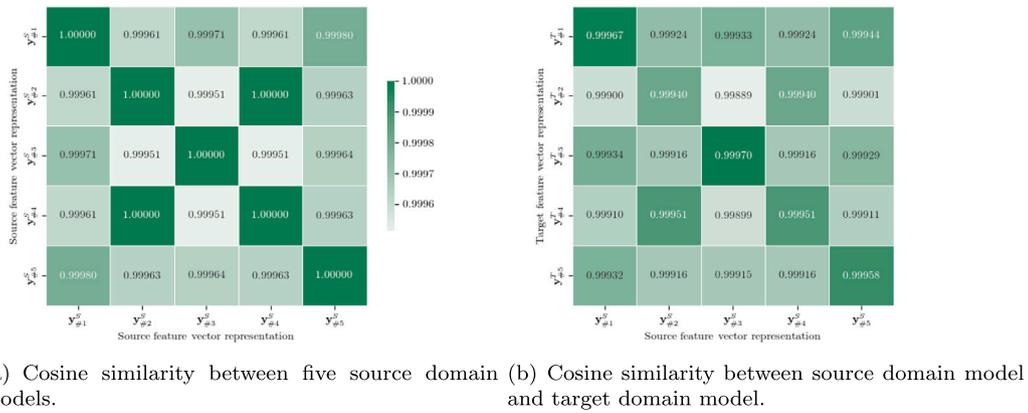


Fig. 7. Results on the validity of transfer learning based on cosine similarity. Note that the actual value of $y_{\#2}^S$ and $y_{\#4}^S$ similarity in figure (a) is 0.9999998987, limited by the display digits, which shows 1.00000.

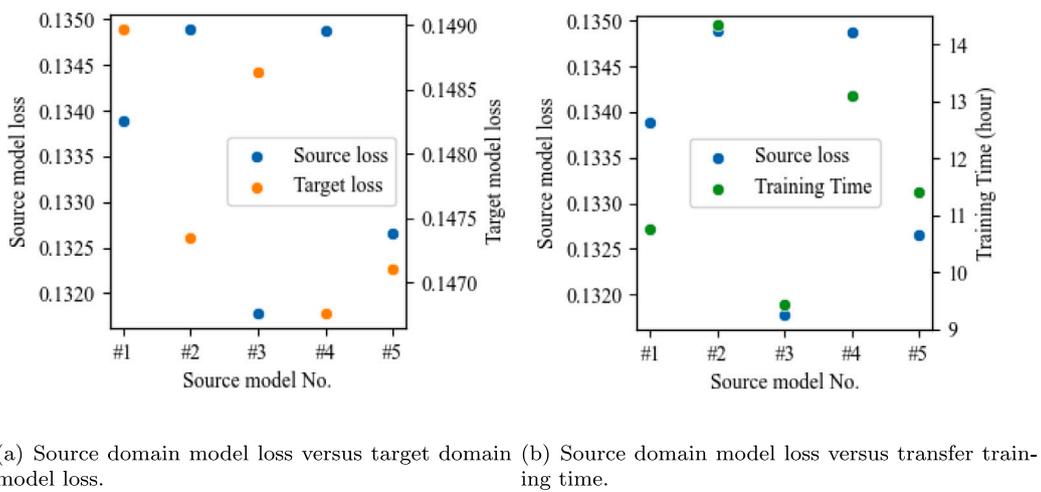


Fig. 8. Comparison of source domain model performance in terms of target domain model performance and training efficiency.

Fig. 12, which juxtaposes training duration against model accuracy, accentuates the time-saving potential of transfer learning. The graph reveals that transfer learning can consistently outperform traditional learning techniques, even with a mere 4.5% of the total dataset and just a tenth of the training time.

While transfer learning offers a promising avenue for optimising both model performance and training duration, the consistency of performance between source and target domain models remains an area ripe for exploration, as depicted in Fig. 8.

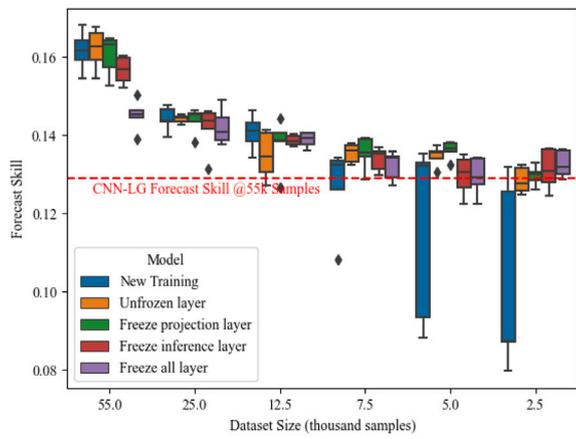
Emerging research should pivot towards intricate transfer learning mechanisms, such as shared representations or alignment techniques. These could potentially guide the target domain model during its training phase, thereby bridging the performance gap between source and target domain models. The potential of zero-shot or few-shot learning in the VST-DL-GSI-SolarForecast domain also beckons further exploration, promising innovative methodologies for broader application.

Transfer learning’s dual advantages of time and cost efficiency are evident in the significant reduction in training duration and data collection periods. However, our investigation also highlights certain transfer learning limitations, especially concerning generalisation when fine-tuning with real-world data. The observed performance disparities across datasets, particularly the June dataset, are likely attributed to the limited diversity of images. This limitation underscores the need for future research to focus on enhancing feature space adaptation during transfer learning, possibly by leveraging prior knowledge constraints on the target domain.

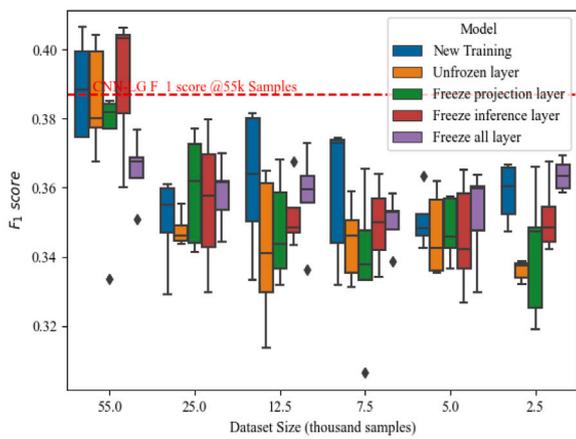
However, existing studies have several limitations. First, this study only discusses the transferability of model knowledge based on the currently available two datasets. Given that deep learning models heavily rely on data and considering the long time dimensions required for meteorological data collection, the feasibility and applicability of transfer learning in VST-DL-GSI-SolarForecast are still unclear. This is especially true for validation experiments under varying climatic conditions. Second, similar to other solar energy prediction modelling methods, this study was conducted under ideal conditions, using past data to predict historical future conditions. This approach overlooks potential issues that may arise during practical field deployment, such as sand erosion and equipment stability. For example, the sky images in the Folsom dataset used in this study already show signs of sand erosion on the acrylic casing protecting the lens. However, since there is currently no comparison between image quality and prediction performance, this phenomenon cannot be analysed and compared.

6. Conclusion

The essence of harnessing and generalising empirical knowledge from existing solar energy forecasting datasets is vital for the progression of Deep learning based Solar Forecast research. Prioritising model transferability can catalyse research advancements, streamline communication, and foster time and cost efficiencies in practical deployments. Our study highlights the ViT-E model’s superiority for transfer learning, particularly in transfer the model from a more steady solar availability



(a) Forecast skill



(b) F_1 Score

Fig. 9. Qualitative and quantitative results based on different subdataset sizes and transfer learning methods for ViT models. Represented by the red dashed line in the figure is the best model performance in the original paper [25] based on CNN architecture trained from scratch. For detailed information, please refer to the tables in Appendix.

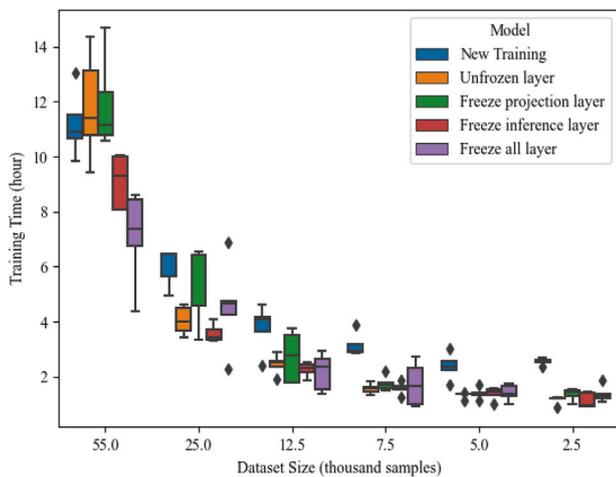


Fig. 10. Training Time for different dataset sizes and transfer learning approaches.

region to a region with more dynamic solar irradiance variations like Nottingham, UK.

Our transfer experiments employed a rudimentary weight transfer technique. However, our findings affirm the potential of weight transfer-based methods in assimilating prior knowledge from pre-trained models across diverse climatic conditions. This assimilation proves instrumental in VST-DL-GSI-SolarForecast tasks across varying climates. Moreover, by amalgamating different transfer strategies, models can be effectively trained with minimal datasets. This efficiency translates to significant savings in model training time and data collection efforts.

This research also amalgamates insights from diverse fields to offer a comprehensive analysis of transfer learning within the VST-DL-GSI-SolarForecast domain. While our simulation experiments showcased the potential of using continuous two-week data for transfer learning, the performance did not match the potential exhibited by datasets with robust generalisability. Future endeavours should pivot towards more advanced transfer learning strategies to ensure efficient transfers, even in datasets with limited temporal diversity.

CRedit authorship contribution statement

Liwenbo Zhang: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Robin Wilson:** Writing – review & editing, Supervision. **Mark Sumner:** Writing – review & editing, Supervision, Resources, Funding acquisition. **Yupeng Wu:** Writing – review & editing, Visualization, Validation, Supervision, Resources, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Yupeng Prof. Wu reports financial support was provided by Engineering and Physical Sciences Research Council. Yupeng Prof. Wu reports financial support was provided by Department for Energy Security and Net Zero.

Data availability

The authors do not have permission to share data.

Acknowledgements

This study has been funded by the Engineering and Physical Sciences Research Council, UK [grant number EP/W028581/1] and Department for Energy Security and Net Zero [grant number 208857380].

Appendix

The following tables provide detailed tabulated data corresponding to the results presented in Fig. 9 (see Tables A.1 and A.2).

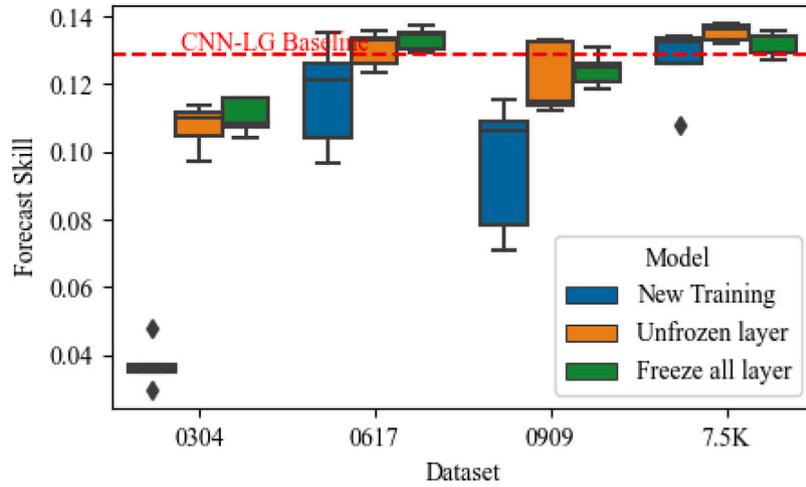


Fig. 11. The model’s results trained by collecting data for two consecutive weeks are compared with a full year of data downsampled to 7.5K data. Represented by the red dashed line in the figure is the best model performance in the original paper [25] based on CNN architecture trained from scratch.

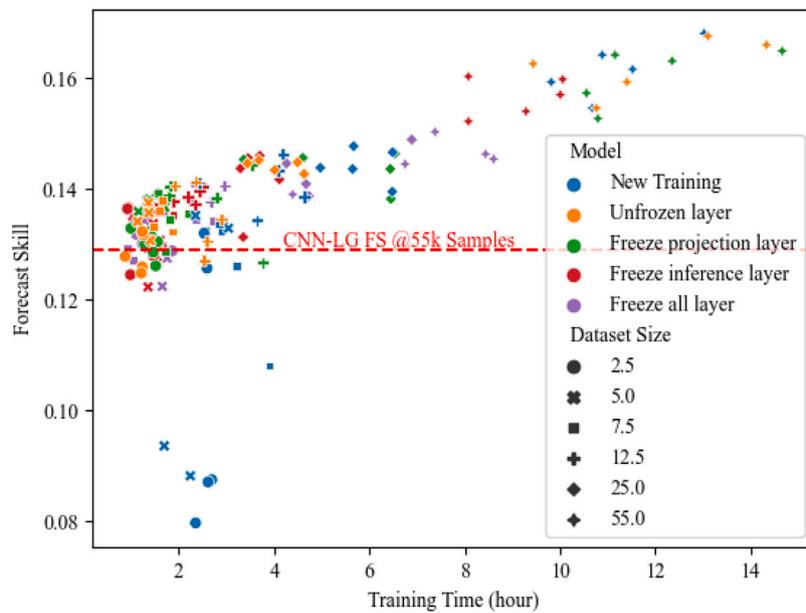


Fig. 12. Training time versus forecast skill for all models and all datasets. Represented by the red dashed line in the figure is the best model performance in the original paper [25] based on CNN architecture trained from scratch.

Table A.1

FS for different transfer learning strategies and dataset sizes, Based on SPM with RMSE 117.89 W/m².

Dataset size	Forecast skill (%)				
	New training	Unfrozen layer	Freeze inference layer	Freeze projection layer	Freeze all layer
2.5k	10.24 ± 2.44	12.85 ± 0.33	13.12 ± 0.52	12.96 ± 0.25	13.26 ± 0.34
5k	11.64 ± 2.34	13.47 ± 0.25	12.97 ± 0.52	13.62 ± 0.23	12.94 ± 0.49
7.5k	12.69 ± 1.09	13.54 ± 0.24	13.38 ± 0.30	13.53 ± 0.42	13.22 ± 0.37
12.5k	14.07 ± 0.45	13.48 ± 0.61	13.87 ± 0.13	13.77 ± 0.65	13.88 ± 0.19
25k	14.42 ± 0.31	14.41 ± 0.10	14.16 ± 0.60	14.38 ± 0.33	14.21 ± 0.46
55k	16.15 ± 0.50	16.20 ± 0.52	15.66 ± 0.35	16.04 ± 0.52	14.51 ± 0.40

Table A.2
 F_1 score for different transfer learning strategies and dataset sizes.

Dataset size	F_1 Score (%)				
	New training	Unfrozen layer	Freeze inference layer	Freeze projection layer	Freeze all layer
2.5k	35.85 ± 0.84	35.14 ± 1.01	33.60 ± 0.28	34.11 ± 1.90	36.36 ± 0.45
5k	35.05 ± 0.80	34.58 ± 1.57	34.65 ± 1.21	34.78 ± 0.91	35.22 ± 1.39
7.5k	35.94 ± 2.00	34.94 ± 1.17	34.45 ± 1.12	33.81 ± 2.16	35.00 ± 0.73
12.5k	36.18 ± 2.04	35.21 ± 0.94	34.22 ± 2.15	34.78 ± 1.52	35.75 ± 1.35
25k	35.04 ± 1.30	35.59 ± 2.00	34.78 ± 0.47	35.94 ± 1.61	35.82 ± 0.97
55k	38.87 ± 1.43	39.10 ± 2.00	38.57 ± 1.55	37.24 ± 2.19	36.54 ± 0.95

References

- [1] Yang Dazhi, Wang Wenting, Gueymard Christian A, Hong Tao, Kleissl Jan, Huang Jing, Perez Marc J, Perez Richard, Bright Jamie M, Xia Xiang'ao. A review of solar forecasting, its dependence on atmospheric sciences and implications for grid integration: Towards carbon neutrality. *Renew Sustain Energy Rev* 2022;161:112348.
- [2] Peng Zhenzhou, Yu Dantong, Huang Dong, Heiser John, Yoo Shinjae, Kalb Paul. 3D cloud detection and tracking system for solar forecast using multiple sky imagers. *Sol Energy* 2015;118:496–519.
- [3] Wang Kejun, Qi Xiaoxia, Liu Hongda. A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. *Appl Energy* 2019;251:113315.
- [4] Gandhi Oktoviano, Kumar Dhivya Sampath, Rodríguez-Gallegos Carlos D, Srinivasan Dipti. Review of power system impacts at high PV penetration part I: Factors limiting PV penetration. *Sol Energy* 2020;210:181–201.
- [5] Kumar Dhivya Sampath, Gandhi Oktoviano, Rodríguez-Gallegos Carlos D, Srinivasan Dipti. Review of power system impacts at high PV penetration part II: Potential solutions and the way forward. *Sol Energy* 2020;210:202–21.
- [6] Kiritmat Ayca, Tasgetiren M Fatih, Brida Peter, Krejcar Ondrej. Control of PV integrated shading devices in buildings: A review. *Build Environ* 2022;214:108961.
- [7] Du Plessis AA, Strauss JM, Rix AJ. Short-term solar power forecasting: Investigating the ability of deep learning models to capture low-level utility-scale photovoltaic system behaviour. *Appl Energy* 2021;285:116395.
- [8] Du Wenping, Li Ming, Wang Yunfeng, Ma Xun, Hu Chengzhi, Zhang Ying, Zhang Zhuoli. Dynamic energy efficiency characteristics analysis of a distributed solar photovoltaic direct-drive solar cold storage. *Build Environ* 2021;206:108324.
- [9] Paletta Quentin, Terrén-Serrano Guillermo, Nie Yuhao, Li Binghui, Bieker Jacob, Zhang Wenqi, Dubus Laurent, Dev Soumyabrata, Feng Cong. Advances in solar forecasting: Computer vision with deep learning. *Adv Appl Energy* 2023;100150.
- [10] Li Qingyong, Lu Weita, Yang Jun. A hybrid thresholding algorithm for cloud detection on ground-based color images. *J Atmos Ocean Technol* 2011;28(10):1286–96.
- [11] Mueller Richard, Trentmann Jörg, Träger-Chatterjee Christine, Posselt Rebekka, Stöckli Reto. The role of the effective cloud albedo for climate monitoring and analysis. *Remote Sens* 2011;3(11):2305–20.
- [12] Sawant Manisha, Shende Mayur Kishor, Feijóo-Lorenzo Andrés E, Bokde Neeraj Dhanraj. The state-of-the-art progress in cloud detection, identification, and tracking approaches: a systematic review. *Energies* 2021;14(23):8119.
- [13] Nouri Bijan, Wilbert Stefan, Segura Luis, Kuhn P, Hanrieder Natalie, Kazantzidis A, Schmidt Thomas, Zarzalejo L, Blanc Philipp, Pitz-Paal Robert. Determination of cloud transmittance for all sky imager based solar nowcasting. *Sol Energy* 2019;181:251–63.
- [14] Jumper John, Evans Richard, Pritzel Alexander, Green Tim, Figurnov Michael, Ronneberger Olaf, Tunyasuvunakool Kathryn, Bates Russ, Židek Augustin, Potapenko Anna, et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021;596(7873):583–9.
- [15] Liang Percy, Bommasani Rishi, Lee Tony, Tsipras Dimitris, Soylu Dilara, Yasunaga Michihiro, Zhang Yian, Narayanan Deepak, Wu Yuhuai, Kumar Ananya, et al. Holistic evaluation of language models. 2022, arXiv preprint arXiv:2211.09110.
- [16] Paletta Quentin, Hu Anthony, Arbod Guillaume, Blanc Philippe, Lasenby Joan. SPIN: Simplifying polar invariance for neural networks application to vision-based irradiance forecasting. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2022, p. 5182–91.
- [17] Sun Yuchi, Venugopal Vignesh, Brandt Adam R. Convolutional neural network for short-term solar panel output prediction. In: *2018 IEEE 7th world conference on photovoltaic energy conversion (WCPEC)(A joint conference of 45th IEEE PVSC, 28th PVSEC & 34th EU PVSEC)*. IEEE; 2018, p. 2357–61.
- [18] Feng Cong, Zhang Jie, Zhang Wenqi, Hodge Bri-Mathias. Convolutional neural networks for intra-hour solar forecasting based on sky image sequences. *Appl Energy* 2022;310:118438.
- [19] Zhao Xin, Wei Haikun, Wang Hai, Zhu Tingting, Zhang Kanjian. 3D-CNN-based feature extraction of ground-based cloud images for direct normal irradiance prediction. *Sol Energy* 2019;181:510–8.
- [20] Paletta Quentin, Arbod Guillaume, Lasenby Joan. Benchmarking of deep learning irradiance forecasting models from sky images—an in-depth analysis. *Sol Energy* 2021;224:855–67.
- [21] Gao Huiyu, Liu Miaomiao. Short-term solar irradiance prediction from sky images with a clear sky model. In: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*. 2022, p. 2475–83.
- [22] Abadi Martín, Barham Paul, Chen Jianmin, Chen Zhifeng, Davis Andy, Dean Jeffrey, Devin Matthieu, Ghemawat Sanjay, Irving Geoffrey, Isard Michael, et al. {TensorFlow}: a system for {large-scale} machine learning. In: *12th USENIX symposium on operating systems design and implementation*. OSDI 16, 2016, p. 265–83.
- [23] Wang Fei, Zhang Zhanyao, Chai Hua, Yu Yili, Lu Xiaoxing, Wang Tiegqiang, Lin Yuzhang. Deep learning based irradiance mapping model for solar PV power forecasting using sky image. In: *2019 IEEE industry applications society annual meeting*. IEEE; 2019, p. 1–9.
- [24] Kumari Pratima, Toshniwal Durga. Long short term memory-convolutional neural network based deep hybrid approach for solar irradiance forecasting. *Appl Energy* 2021;295:117061.
- [25] Zhang Liwenbo, Wilson Robin, Sumner Mark, Wu Yupeng. Advanced multimodal fusion method for very short-term solar irradiance forecasting using sky images and meteorological data: A gate and transformer mechanism approach. *Renew Energy* 2023;216:118952.
- [26] Paletta Quentin, Hu Anthony, Arbod Guillaume, Lasenby Joan. ECLIPSE: Environmental cloud induced perturbations in solar energy. Vol. 326, Elsevier; 2021, 119924, arXiv preprint arXiv:2104.12419.
- [27] Guen Vincent Le, Thome Nicolas. Disentangling physical dynamics from unknown factors for unsupervised video prediction. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020, p. 11474–84.
- [28] Le Guen Vincent, Thome Nicolas. A deep physical model for solar irradiance forecasting with fisheye images. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 2020, p. 630–1.
- [29] Nie Yuhao, Zelikman Eric, Scott Andea, Paletta Quentin, Brandt Adam. SkyGPT: Probabilistic short-term solar forecasting using synthetic sky videos from physics-constrained VideoGPT. 2023, arXiv preprint arXiv:2306.11682.
- [30] Yang Dazhi. Estimating 1-min beam and diffuse irradiance from the global irradiance: A review and an extensive worldwide comparison of latest separation models at 126 stations. *Renew Sustain Energy Rev* 2022;159:112195.
- [31] Sun Yuchi, Venugopal Vignesh, Brandt Adam R. Short-term solar power forecast with deep learning: Exploring optimal input and output configuration. *Sol Energy* 2019;188:730–41.
- [32] Feng Cong, Zhang Jie. SolarNet: A deep convolutional neural network for solar forecasting via sky images. In: *2020 IEEE power & energy society innovative smart grid technologies conference*. ISGT, IEEE; 2020, p. 1–5.
- [33] Pedro Hugo TC, Larson David P, Coimbra Carlos FM. A comprehensive dataset for the accelerated development and benchmarking of solar forecasting methods. *J Renew Sustain Energy* 2019;11(3):036102.
- [34] Nie Yuhao, Li Xiatong, Scott Andea, Sun Yuchi, Venugopal Vignesh, Brandt Adam. SKIPP'D: a SKY images and photovoltaic power generation dataset for short-term solar forecasting. 2022, arXiv preprint arXiv:2207.00913.
- [35] Nie Yuhao, Li Xiatong, Paletta Quentin, Aragon Max, Scott Andea, Brandt Adam. Open-source ground-based sky image datasets for very short-term solar forecasting, cloud analysis and modeling: A comprehensive survey. 2022, arXiv preprint arXiv:2211.14709.
- [36] Feng Cong, Zhang Jie. SolarNet: A sky image-based deep convolutional neural network for intra-hour solar forecasting. *Sol Energy* 2020;204:71–8.
- [37] Zhuang Fuzhen, Qi Zhiyuan, Duan Keyu, Xi Dongbo, Zhu Yongchun, Zhu Hengshu, Xiong Hui, He Qing. A comprehensive survey on transfer learning. *Proc IEEE* 2020;109(1):43–76.
- [38] Sarmas Elissaios, Dimitropoulos Nikos, Marinakis Vangelis, Mylona Zoi, Doukas Haris. Transfer learning strategies for solar power forecasting under data scarcity. *Sci Rep* 2022;12(1):14643.
- [39] Zeiler Matthew D, Fergus Rob. Visualizing and understanding convolutional networks. In: *Computer vision—ECCV 2014: 13th European conference, Zurich, Switzerland, September 6–12, 2014, proceedings, Part 13*. Springer; 2014, p. 818–33.

- [40] Shin Hoo-Chang, Roth Holger R, Gao Mingchen, Lu Le, Xu Ziyue, Nogues Isabella, Yao Jianhua, Mollura Daniel, Summers Ronald M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans Med Imaging* 2016;35(5):1285–98.
- [41] Abubakr Mohamed, Akoush Bassem, Khalil Adel, Hassan Muhammed 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(4):474.
- [42] Niu Tong, Li Jinkai, Wei Wei, Yue Hui. A hybrid deep learning framework integrating feature selection and transfer learning for multi-step global horizontal irradiation forecasting. *Appl Energy* 2022;326:119964.
- [43] Tang Yugui, Yang Kuo, Zhang Shujing, Zhang Zhen. Photovoltaic power forecasting: A hybrid deep learning model incorporating transfer learning strategy. *Renew Sustain Energy Rev* 2022;162:112473.
- [44] Nie Yuhao, Paletta Quentin, Scotta Andea, Pomares Luis Martin, Arbod Guillaume, Sgouridis Sgouris, Lasenby Joan, Brandt Adam. Sky-image-based solar forecasting using deep learning with multi-location data: training models locally, globally or via transfer learning? 2022, arXiv preprint arXiv:2211.02108.
- [45] Lefevre Mireille, Oumbe Armel, Blanc Philippe, Espinar Bella, Gschwind Benoît, Qu Zhipeng, Wald Lucien, Schroedter-Homscheidt Marion, Hoyer-Klick Carsten, Arola Antti. Mcclear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions. *Atmos Meas Tech* 2013;6(9):2403–18.
- [46] Tan Chuanqi, Sun Fuchun, Kong Tao, Zhang Wenchang, Yang Chao, Liu Chunfang. A survey on deep transfer learning. In: *Artificial neural networks and machine learning–ICANN 2018: 27th international conference on artificial neural networks*, Rhodes, Greece, October 4-7, 2018, proceedings, part III 27. Springer; 2018, p. 270–9.
- [47] Chollet Francois, et al. Keras. 2015, URL <https://github.com/fchollet/keras>.
- [48] Ltd LI-COR Biosciences UK. Light sensor brochure 15539, LI-200r. 2023, URL https://www.licor.com/env/products/light/pyranometer.html?gclid=CjwKCAjwiOCgBhAgEiwAjl5whHyBdT7XrpVTI-8QjRuk6iH8oeiNP7VOZvE23GeX6q6hIVABwwMljhoC7u0QAvD_BwE.
- [49] Kipp & Zonen BV. Instruction manual, RaZON+ solar monitoring system. 2023, URL <https://www.kippzonen.com/Download/908/Manual-RaZON-ALL-IN-ONE-Solar-Monitoring-System>.
- [50] Wilbert Stefan, Jessen Wilko, Forstinger Anne, Driesse Anton, Habte Aron M, Sengupta Manajit, Marzo Aitor, Vignola Frank, Zarzlejo Luis. Application of the clear-sky spectral error for radiometer classification in ISO 9060. Technical report, National Renewable Energy Lab.(NREL), Golden, CO (United States); 2019.
- [51] for Standardization International Organization. Solar energy — Specification and classification of instruments for measuring hemispherical solar and direct solar radiation. Standard, 2000, Geneva, CH: International Organization for Standardization; 2018.
- [52] Holmgren William F, Hansen Clifford W, Mikofski Mark A. Pvlib python: A python package for modeling solar energy systems. *J Open Source Softw* 2018;3(29):884.