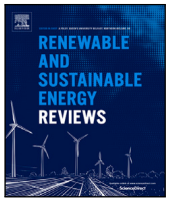
Renewable and Sustainable Energy Reviews 198 (2024) 114391

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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



Review article

A review of distributed solar forecasting with remote sensing and deep learning

Yinghao Chu [a], Yiling Wang [b,c], Dazhi Yang [d], Shanlin Chen [b], Mengying Li [b,*]

[a] Department of Systems Engineering, City University of Hong Kong, Hong Kong Special Administrative Region
[b] Department of Mechanical Engineering & Research Center on Data Science and Artificial Intelligence, The Hong Kong Polytechnic University, Hong Kong Special Administrative Region
[c] School of Energy and Environment, City University of Hong Kong, Hong Kong Special Administrative Region
[d] School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, Heilongjiang, China

ARTICLE INFO

Keywords:
Review
Solar integration
Spatial solar forecasting
Remote sensing
Deep learning
Hybrid methods

ABSTRACT

The rapidly growing capacity of globally distributed solar generation systems (DSGs) has imposed new challenges for solar forecasting research: the need for high-fidelity spatial solar forecasts across utility-scale areas with minimized capital, generalization, and maintenance costs. The majority of solar forecasting approaches were developed for centralized solar power plants, which only concern one or a few locations. Therefore, this work reviews the state-of-the-art methods for spatial solar forecasting that integrate deep learning and remote sensing, potentially capable of serving numerous DSGs simultaneously. This work has four missions: (1) provide a review of available remote-sensing- and deep-learning-based spatial solar forecasting methods; (2) provide suggestions of practical tools to accelerate the research and deployment of spatial solar forecasting methods; (3) identify challenges of spatial solar forecasting for sparsely distributed DSGs; and (4) discuss prospective approaches to further enhance both the performance and value of spatial solar forecasts, such as the attention mechanism, sequence analysis, or probabilistic forecasts. This work reveals that practical spatial solar forecasting for DSGs is still in its infancy, thus more research efforts should be involved to develop a new generation of forecasting engines, which could cost-effectively address the real-time needs of integrating massive regional DSGs.

1. Introduction

Motivated by the societal need for sustainable development around the world, the global installed capacity of solar photovoltaic (PV) has grown rapidly over the past decade and is expected to increase by 20-fold and reach 22% of the electricity market by 2050 [1,2]. A variety of studies have confirmed that the accurate forecasting of weather-dependent solar fluctuations is key to integrating the volatile and non-dispatchable solar energy into the power grid [3–5]. However, accurately forecasting either solar irradiance or solar power is a challenging task because of the complex meteorological processes. Specifically, the properties and dynamics of clouds, which exhibit high spatial and temporal variability, contribute to this complexity [6]. In addition, the rapidly growing installation of PV systems, particularly distributed solar generation systems (DSGs) over large areas, has imposed new challenges to solar forecasting and grid integration. Compared to centralized solar energy systems, DSGs offer several benefits including proximity to consumers, minimized transmission loss, lower investment entry barriers, local grid stabilization, and flexible installation options that optimize land use, such as rooftop or floating PV [7,8]. Therefore, recent installations of DSGs have markedly outpaced those of centralized solar systems, especially in densely populated regions with limited land availability [9]. With hundreds of grid-connected DSGs in operation [10], regional-scale spatial solar forecasting is essential for the planning, integrating, regulating, and managing of solar power generations over a large area [11].

Motivated by the increasing demands of solar integration, different solar forecasting approaches have been proposed and discussed in the literature, such as numerical weather prediction (NWP), physical models that are based on either local- or remote-sensing techniques, data-driven methods based on regressive, machine learning, and deep learning approaches, and hybrid approaches that integrate multiple methods to maximize the forecasting performance [2,3,12]. The commonly employed data sources include in-situ measurement of solar irradiance [13] and DSGs power generation [14], sky images [11,15],

* Corresponding author.
E-mail address: mengying.li@polyu.edu.hk (M. Li).

https://doi.org/10.1016/j.rser.2024.114391
Received 24 November 2022; Received in revised form 13 November 2023; Accepted 24 March 2024
Available online 25 April 2024

Nomenclature

Abbreviations

ABI	Advanced Baseline Imager
ADAM	Adaptive moment estimation
AGRI	Advanced Geostationary Radiation Imager
AHI	Advanced Himawari Imager
AMI	Advanced Meteorological Imager
ANN	Artificial neural network
ARIMA	Autoregressive integrated moving average
biLSTM	Bidirectional long short-term memory
CCI	Cloud clearness index
CMA	China Meteorological Administration
CNN	Convolutional neural network
COMS	Communication, Ocean and Meteorological Satellite
DES	Double exponential smoothing
DHI	Diffuse horizontal irradiance
DNI	Direct normal irradiance
DNN	Deep neural network
DSG	Distributed solar generation
ECMWF	European Centre for Medium-range Weather Forecasts
ELM	Extreme learning machine
EMOS	Ensemble model output statistics
EUMETSAT	European Organization for the Exploitation of Meteorological Satellite
FFNN	Feed-forward neural networks
GAN	Generative adversarial network
GBR	Gradient boosting regression
GFS	Global Forecast System
GHI	Global horizontal irradiance
GK2A/2B	Geostationary Korean Multi-purpose Satellite 2A/2B
GNN	Graph neural network
GOES	Geostationary Operational Environmental Satellites
GRAPES	Global and Regional Assimilation and Prediction System
GRU	Gated recurrent unit
HARMONIE	Hirlam–Aladin Research towards Mesoscale Operational NWP In Euromed
HRRR	High-Resolution Rapid Refresh
KMA	Korea Meteorological Administration
kNN	k Nearest neighbor
KSI	Kolmogorov–Smirnov integral
LSTM	Long short-term memory
LSTM-FC	Long short-term memory-fully connected
MAE	Mean absolute error
MBE	Mean bias error
MLP	Multilayer perceptron
MODIS	Moderate resolution imaging spectroradiometer
NAM	North American Mesoscale
NMSC	National Meteorological Satellite Center
NSRDB	National Solar Radiation Database
NWP	Numerical weather prediction
PV	Photovoltaic
RAP	Rapid Refresh model
ReLU	Rectified linear unit
RF	Random forest
rMAE	Relative mean absolute error
RMSE	Root mean square error
RNN	Recurrent neural network
ROI	Region of interest
rRMSE	Relative root mean square error
SARIMAX	Seasonal autoregressive integrated moving average with exogenous factors
SEVIRI	Spinning Enhanced Visible and infrared Imager
SVM	Support vector machine
SVR	Support vector regression
WRF	Weather Research and Forecasting
XGBoost	Extreme gradient boosting

Notations

α	Coefficient of relative positions between satellite and target location
γ_h	The autocorrelation at lag h
$\hat{\kappa}$	Clear-sky index forecast
$\hat{I}$	Solar irradiance/power forecast
κ	Clear-sky index
μ	Climatology reference term
ρ	Correlation coefficient
σ	Standard deviation
θ_z	Solar zenith angle
$\tilde{L}$	Normalized pixel values
c	Clear-sky expectation of irradiance/power
H_a	Rényi entropy
I	Solar irradiance/power
L	Pixel value of satellite images
L'	Normalized pixel values by solar zenith angle
N	Number of days
R^2	Coefficient of determination
s	Forecast skill

Subscripts

f	Forecasts of interest
h	Forecast horizon
i	Number of row in a matrix
j	Number of column in a matrix
p	Perfect forecasts
r	Reference forecasts
t	Time

Superscripts

b	Spectral band

meteorological variables (e.g., temperature, relative humidity, wind direction or wind speed) that are measured or predicted by NWP models [16–18], and remote-sensing satellite data products [19–21]. For solar forecasting research, the spatial and temporal scales of the aforementioned data sources are illustrated in Fig. 1. For spatial forecasts over hundreds of square kilometers, the choices of data are remote-sensing imagery and NWP output. Since comprehensive details of state-of-the-art NWP models are widely accessible elsewhere, such as the articles by Wang et al. [22], Yang et al. [23] and Jimenez et al. [24], this work only briefly summarizes a few NWP models commonly used in solar forecasting research, as presented in Appendix A.

Remote sensing has emerged as a reliable and informative data source for solar resource assessment and forecasting. Irradiance derived thereof has wide geographical coverages and thus partly eliminates the need for local instrumentation investment [26]. Since the presence

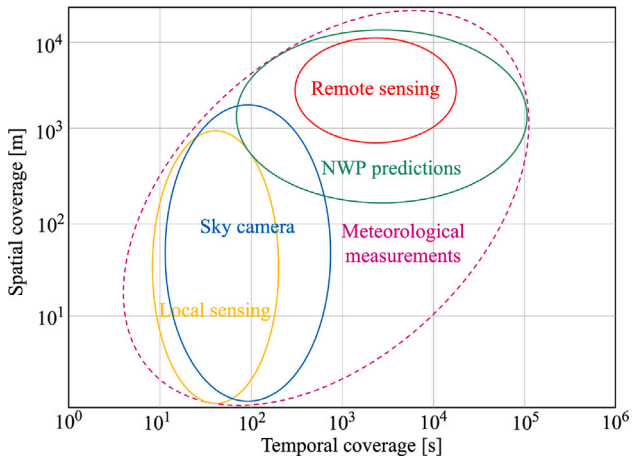


Fig. 1. Temporal and spatial coverage of commonly employed techniques for obtaining exogenous data.
Source: Modified based on [2,25].

and movement of clouds are highly relevant to surface irradiance and thus solar power [27], the inclusion of cloud information derived from satellite imagery as exogenous inputs has been used to enhance forecasting performance [3,28–30]. Images from the latest-generation geostationary satellites are currently being sampled every 5 to 15 min, which enables not only intra-day but also intra-hour solar forecasting, which is particularly useful for power system operations, such as load following and real-time electricity trading [30,31]. On the other hand, the spatial coverage of satellite-based solar forecasts is beneficial for the utility-scale market operators to balance grids for cities or even states [32,33]. Traditionally, remote-sensing-based solar forecasting is performed under a physical setting, in which the cloud advection and diffusion processes are captured and described [25, 34]. However, statistically determining the cloud dynamics and its impact on ground irradiance challenge these physics-based models, often resulting in large forecast errors [35]. Therefore, data-driven methods, including deep learning, have been introduced to enhance the remote-sensing-based solar forecasting models [12].

Deep learning methods (e.g., deep neural networks, or DNNs) have brought breakthroughs in many fields such as computer vision and speech recognition [36]. Evaluated with various datasets, deep learning methods have shown superior performance in terms of accuracy, generalization ability, and robustness, as compared to conventional rule-based models or machine learning models [37–40]. The convolutional neural networks (CNNs) and recurrent neural networks (RNNs) are particularly useful in computer vision and time series analysis, respectively [41,42]. With no exception, solar forecasting, as a fast-advancing field, has already seen deep learning applications, which have shown superior performance over physical or conventional machine learning methods on many occasions [12,43,44].

However, due to the end-to-end (also known as "black box") nature of data-driven techniques, deep-learning-based solar forecasting methods are mostly developed using locally collected data as training targets. Therefore, most of these models are developed for a limited number of point locations and can hardly be transferred or generalized to new locations beyond the training set [45]. This is also confined by the fact that deep learning models, unlike statistical regression models (e.g., linear regression), do not have ease-to-express function forms. So models fitted in one scenario are not conducive to adoption in another scenario, unless the code and data are passed on in their exact form, which has rarely been the case under the current "publish or perish" research regime. Consequently, only a few studies have utilized both satellite imagery and deep learning to perform spatial/regional forecasting for solar irradiance/power [46,47]. For example, by the year 2022, searching keywords such as *deep learning* with *solar forecast* or *PV forecast* returns more than 2000 publications from the Web of Science Core Collection. However, when searching *deep learning* with *spatial solar forecast*, the number of return publications decreases to just 45, among which a fraction use satellite data. Consequently, how to best perform satellite-based spatial solar forecasting using deep learning techniques has not been standardized and the related research gaps have not been identified.

Therefore, in this work, a thorough literature review of spatial solar forecasting methods that utilize both satellite imagery and deep learning is conducted, particularly for large-area applications such as grid integration. The main contributions of this work are summarized as follows:

- This work provides a comprehensive summary of spatial solar forecasting methods that leverage satellite data and deep learning, a topic not previously reviewed in detail.
- Practical recommendations on end-to-end data-driven forecasting and essential model development tools are provided, such as remote-sensing satellite data sources or deep learning implementation frameworks. This information is beneficial for researchers to conduct research and model development along this line more efficiently.
- Existing challenges of spatial solar forecasting are identified, providing valuable insights for the development and optimization of future methods for hundreds of DSGs across a large spatial area.
- To overcome the identified challenges, several potential approaches are provided that could potentially improve the performance and the utility/practical value of DSG power forecasts. These approaches mainly refer to attention mechanisms, sequence analyses, large-scale models, and spatial probabilistic forecasting techniques.

In what follows, basic considerations for spatial solar forecasts are summarized in Section 2. Spatial solar forecasting methods are reviewed in Sections Section 3. In Section 4, major challenges are identified and an outlook of potential solutions is provided. The conclusions are summarized in Section 5. It should be again highlighted that the term "solar forecasting", whenever used, denotes both solar irradiance forecasting and solar power forecasting, not just in this work, but in this field in general [43].

2. Basic considerations of spatial solar forecasting

In this section, the data-driven spatial solar forecasting procedure is firstly presented (Section 2.1). Then the most advanced geostationary satellite data around the globe are thoroughly summarized (Section 2.2), followed by common practices to process the satellite data before feeding into the forecasting models (Section 2.3). The forecasting performance assessment methods are briefly summarized in Section 2.4.

2.1. Overview of data-driven spatial forecasting procedure

The general data-driven spatial solar forecasting procedure is illustrated in Fig. 2. The procedure can be divided into three blocks, including (1) data sources, (2) data processing methods, and (3) forecasting model development, assessment, and optimization. The data employed predominantly originate from local sensors, satellite observations, and numerical weather prediction (NWP) models. Then, the data are processed by methods such as filtering, normalization, or feature extraction, during which the physical considerations such as clear-sky modeling are integrated, to isolate the known (i.e., calculable) deterministic trends and seasonality in data as far as possible. The time stamps of the processed data from different sources are then matched to serve as inputs and training targets for forecasting models. Different algorithms can be employed to perform either single-location forecasts

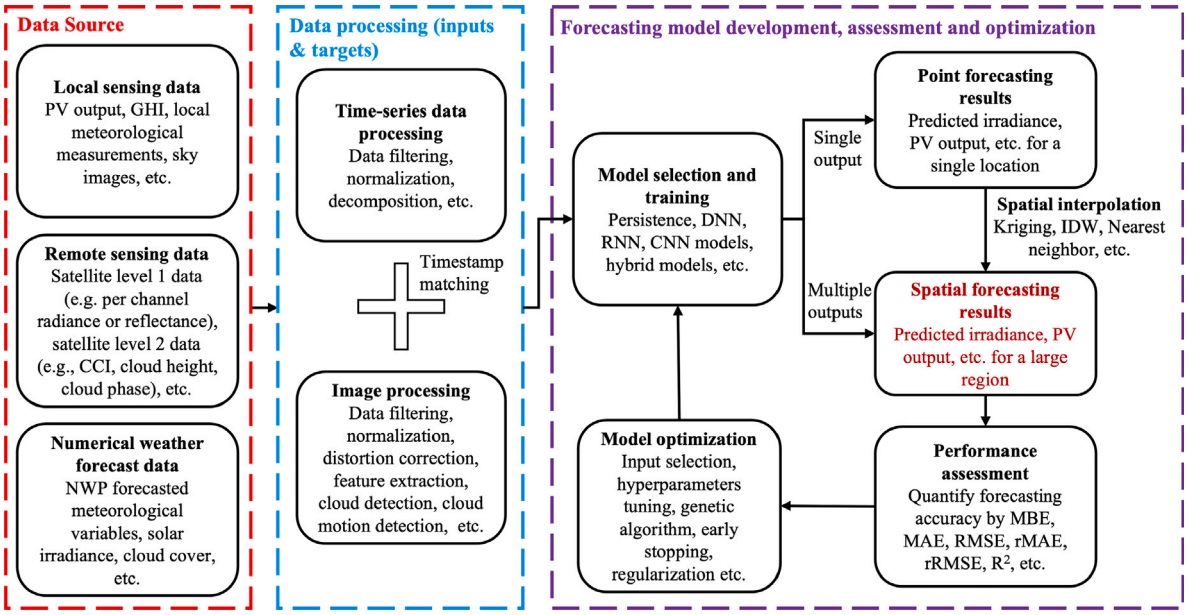


Fig. 2. Data-driven spatial forecasting procedure.

or spatial forecasts. The forecast accuracy is quantified by different metrics and the preliminary model is further optimized to achieve the best performance. Due to the scarcity of local-sensing data over a large area and the recent advancements of geostationary satellite imagers, remote-sensing data sources have become more applicable for spatial forecasting applications. Solar irradiance predictions from NWP can be used either as model inputs or as validation benchmarks for the developed forecasting models.

The spatio-temporal resolution of satellite-based forecasts is intrinsically tied to the resolution of the satellite data employed—the highest-resolution product has a native resolution of 500 m [48] with most being 2–4 km. The irradiance prediction for a satellite pixel is applicable to all DSGs within the coverage area of that pixel. However, such a resolution may be inadequate for forecasts that necessitate a higher spatio-temporal resolution, such as intra-hour forecasting [12], particularly under partly cloudy conditions where ground irradiance can substantially decrease within a minute [49,50]. To accommodate forecasting systems requiring higher spatio-temporal resolutions, ground-based sensors or cameras are advantageous for capturing minute-scale fluctuations in cloud cover and ground-based irradiance. Subsequent spatial analysis methods, including inverse distance weighting (IDW), kriging, or nearest neighbor methods, are implemented to derive values for locations without ground-based sensors. The forecasting performance of these ground-based systems is notably influenced by the density of the sensor network [11]. However, the installation and maintenance of a dense sensor network are both costly and resource-intensive. Consequently, selecting the appropriate data sources and forecasting methods based on application requirements proves to be of paramount importance.

2.2. Geostationary satellite data

Major geostationary satellite data sources that are commonly employed for solar forecasting are summarized here, including Geostationary Operational Environmental Satellites (GOES), Meteosat, Himawari, Fengyun, and Chollian. It should be noted that each of these mentioned names refers to a series of satellites rather than just a single one. The satellites are operated and managed by different authorities, with different onboard imagers, data resolutions, and coverages, as presented in Table 1 and Fig. 3. The advanced satellite imagers are capable of capturing upwelling radiance data from different spectral bands, which enables the monitoring of different atmospheric constituents, especially the optical and physical properties of clouds (as shown in Fig. 4). The comparison of visible, near-infrared, and infrared spectral bands of different satellites are illustrated in Fig. 4 and Table A.2.

The GOES series has been operated by the National Oceanic and Atmospheric Administration, and the National Aeronautics and Space Administration of the United States since 1975.

- Imager: Advanced Baseline Imager (ABI).
- Level 1 data: the upwelling radiance of each spectral band.
- Level 2 data: aerosol detection, aerosol optical depth, clear sky masks, cloud layers/heights, cloud and moisture imagery, cloud optical depth, cloud particle size distribution, cloud top height, cloud top phase, cloud top pressure, cloud top temperature, derived motion winds, derived stability indices, downward shortwave radiation at surface, fire/hot spot characterization, land surface albedo, land surface bidirectional reflectance factor, land surface temperature (skin), legacy vertical moisture profile, legacy vertical temperature profile, radiances, rainfall rate, reflected shortwave radiation from the top of atmosphere, sea and lake ice age, concentration and motion, sea surface temperature, snow cover, total precipitable water, volcanic ash detection and height [53].
- Data retrieval: open source for non-commercial use and can be downloaded directly from Amazon Web Services.[1]

European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) is an intergovernmental organization that is mostly funded by European Union countries. The Meteosat satellite series operated by EUMETSAT provides useful data products for meteorology and climate applications. Meteosat-1 is the first satellite of the series and was launched in November 1977. There are seven satellites of the first generation of Meteosat and four satellites of the second generation of Meteosat, including the currently operating Meteosat-9 and Meteosat-11.

- Imager: Spinning Enhanced Visible and InfraRed Imager (SEVIRI).
- Level 1.5 data: the image radiometry with linearization and equalization of each spectral band.

[1] https://noaa-goes16.s3.amazonaws.com/index.html, https://noaa-goes17.s3.amazonaws.com/index.html.

Table 1
Metadata of different operational geostationary satellites.

Satellite name	Position	Launch time	Number of bands	Spatial resolution	Observation area and temporal resolution
GOES-16 GOES-17[a]	75.2°W 137.3°W	Nov, 2016 Mar, 2018	16	b1–b2: 0.5–1 km b3–b6: 1–2 km b7–b16: 2 km	Full disk (15 min) USA area (5 min) Selected areas (0.5 min)
Meteosat-9 Meteosat-11	45.5°E 0°	Dec, 2005 Jul, 2015	12	b1–b3: 1 km b4–b12: 3 km	Full disk (15 min) European area (5 min)
Himawari-8	140.7°E	Oct, 2014	16	b1–b3: 0.5–2 km b4–b6: 1–2 km b7–b16: 2 km	Full disk (10 min) Japan area (2.5 min) Target area (0.5 min) Lanmark areas (0.5 min)
Fengyun-4A Fengyun-4B	104.7°E 123.5°E	Dec, 2016 Jun, 2021	14	b1–b2: 0.5–1 km b3–b6: 2–4 km b7–b14: 4 km	Full disk (15 min) China area (4–5 min)
GK2A	128.2°E	Dec, 2018	16	b1–b4: 0.5–1 km b5–b6: 2 km b7–b16: 2 km	Full disk (10 min) South Korea area (2.5 min)

[a] GOES-17 is replaced by GOES-18 in 2023, and its position is at 137.0°W.

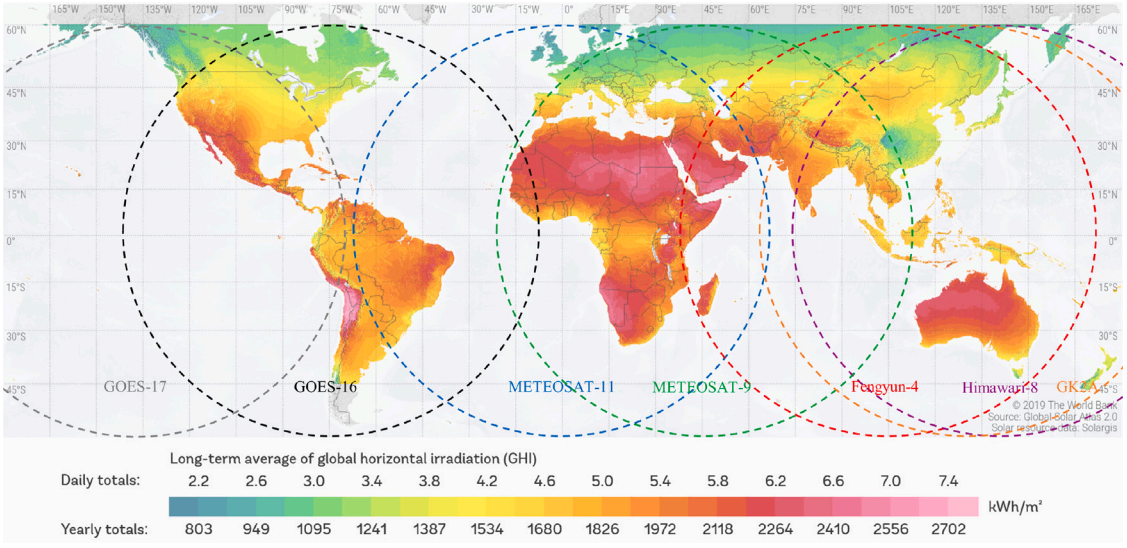


Fig. 3. The spatial observation area of the major operational geostationary satellites. The depicted global annual irradiance map is retrieved from the Solargis website.

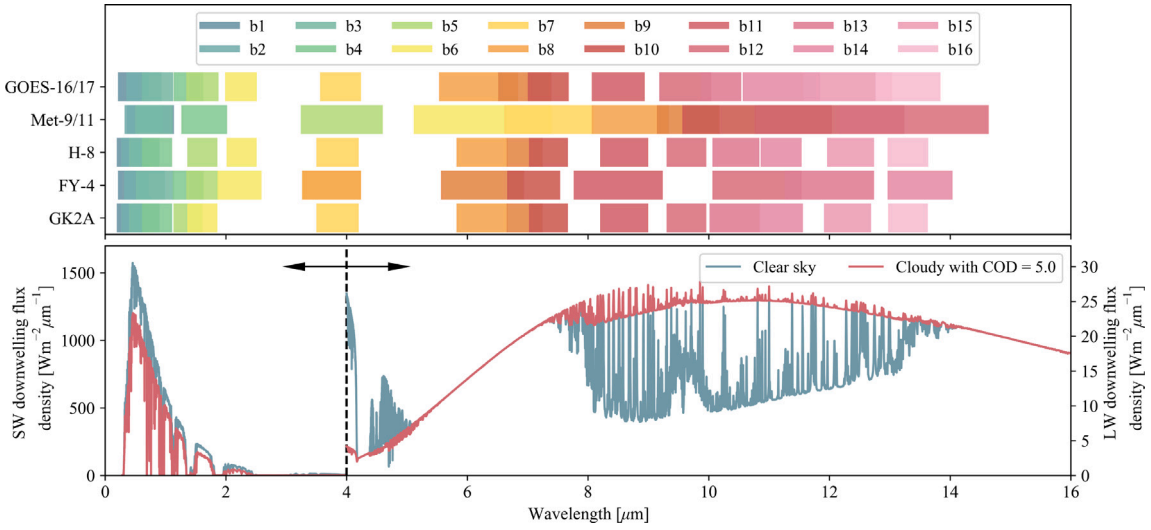


Fig. 4. (a) Illustration of different spectral bands of different satellites. (b) Downwelling ground-level solar shortwave ($\lambda < 4\,\mu m$) and atmospheric longwave ($\lambda > 4\,\mu m$) flux density for clear and cloudy conditions. The spectral longwave and shortwave flux densities are calculated using the radiative transfer models presented in [51,52], respectively.

- Level 2 data: aerosol properties over the sea, cloud analysis, cloud Analysis image, clear-sky reflectance map, divergence, high-resolution precipitation index, normalized difference vegetation index/normalized difference vegetation index decadal, tropospheric humidity [54].
- Data retrieval: open source and can be downloaded from the EUMETSAT website[2] for non-commercial use.

Himawari is a series of geostationary weather satellites launched and operated by the Japan Meteorological Agency. Himawari-8 was launched in October 2014 and has been operational since July 2015, aiming to provide data related to weather forecasting and climate change prediction.

- Imager: Advanced Himawari Imager (AHI).
- Level 1 data: albedo in solar spectrum, brightness temperature in the atmospheric longwave spectrum, satellite zenith angle, satellite azimuth angle, solar zenith angle, solar azimuth angle, and observation hour.
- Level 2 and level 3 data: aerosol property, sea surface temperature, nighttime sea surface temperature, shortwave radiation/photosynthetically available radiation, chlorophyll-a, cloud property (day-time only) and wildfire [55].
- Data retrieval: open source and can be downloaded from the Japan Aerospace Exploration Agency website[3] for non-commercial use.
- Notes: Himawari-8 has both full-disk and regional (i.e., over greater Japan) scans, with the latter being more frequent and more refined in resolution than the former. The scanning for the target area and landmark areas is determined based on changes in weather conditions.

The Fengyun series satellites are meteorological satellites developed by the Shanghai Institute of Aerospace Technology, China. The Fengyun series is mainly used for services such as weather analysis and forecasting, or environmental and disaster monitoring. There are several pre-determined positions for Fengyun satellites, such as 104.7°E, 86.5°E, or 123.5°E. For Fengyun-4A, the fixed position has been at 104.7°E since 2017.

- Imager: Advanced Geostationary Radiation Imager (AGRI).
- Level 1 data: geo-located and calibrated spectral radiance.
- Level 2 and level 3 data: cloud mask, cloud type, total cloud amount, rainfall rate/quantitative precipitation estimate, atmospheric motion vector, outgoing longwave radiation, black body brightness temperature, surface solar irradiance, legacy vertical moisture profile, layer precipitable water, aerosol detection (including smoke and dust), sea surface temperature (skin), snow cover, land surface (skin) temperature, cloud-top height, cloud-top pressure, cloud optical depth, cloud liquid water, cloud particle size distribution, cloud phase, downward longwave radiation at surface, upward longwave radiation at surface, reflected shortwave radiation at top of atmosphere, aerosol optical depth, convective initiation, fire/hot spot characterization, fog detection, land surface emissivity, land surface temperature, land surface albedo and tropopause folding turbulence prediction [56].
- Data retrieval: open source and can be downloaded from the China Meteorological Administration (CMA) website[4] with the requested permit for non-commercial use.

Chollian is the geostationary satellite series launched and operated by South Korea, which is used for observing oceanography and meteorology. The series contains 3 satellites: Communication, Ocean and Meteorological Satellite (COMS), Geostationary-Korean Multi-purpose Satellite-2A (Geo-KOMPSAT-2A, or GK2A) and Geo-KOMPSAT-2B (GK2B), all operated by the Korea Aerospace Research Institute.

- Imager: Advanced Meteorological Imager (AMI).
- Level 1A data: units of calibrated radiance without navigation.
- Level 1B data: calibrated, navigated, and in-pixel-shape spectral radiation.
- Primary meteorological products and meteorological products: land surface emissivity, surface albedo, snow depth, sea surface current, cloud type, cloud amount, cloud optical depth, cloud effective radius, cloud liquid water path, cloud layer/height, probability of rainfall, potential accumulated rainfall, angstrom exponent product, visibility, reflected shortwave radiation, downward shortwave radiation at surface, absorbed shortwave radiation at surface, downward longwave radiation at surface, upward longwave radiation at surface, outgoing longwave radiation, icing, overshooting top, SO_2 detection, total precipitable water, tropopause folding and turbulence detection [57].
- Data retrieval: open source and can be downloaded from National Meteorological Satellite Center[5] (NMSC) for non-commercial use.

2.3. Processing of satellite data

Since the observation areas of a satellite are usually much larger than the region of interest (ROI) to spatial solar forecasting, such as the region covering a cluster of DSG systems connected to the same feeder node, the initial step in the processing of satellite data involves the selection of an appropriate rectangular region that fully encompasses the ROI. To enhance the performance and robustness of solar forecasting applications, it is essential to pre-process the satellite data prior to developing the forecasting model [31].

A fundamental and time-tested approach for normalization is the *minmax* method [58], which is a similar but weaker version of the well-known Heliosat-2 method in this field. It takes the form:

$$\tilde{L}^{b}_{ij,t} = \frac{L^{b}_{ij,t} - \min\left(L^{b}\right)}{\max\left(L^{b}\right) - \min\left(L^{b}\right)}, \tag{1}$$

where $L^{b}_{ij,t}$ is the value of the pixel situated at the ith-row and jth-column in a satellite image of spectral band b at time t. This value may represent radiance, albedo, or brightness temperature, contingent on the data product sourced from the specific satellite. $\min\left(L^{b}\right)$ and $\max\left(L^{b}\right)$ denote the minimum and maximum pixel values across all images of band b in the entire dataset, respectively.

In addition to this standard normalization typically employed in computer vision, seasonality removal should be considered to enhance model performance and robustness. Although the utilization of clear-sky models for eliminating seasonality from solar irradiance time series has become a standard practice in solar resource assessment and forecasting [59], a significant portion of data-driven forecasting models in the extant literature neglects to incorporate seasonality removal when using satellite data as inputs. This omission leads to a failure to acknowledge the influence of inherent diurnal and seasonal variations on the upwelling radiance within the solar spectrum. Ensuring the independence of external cloud-related inputs, such as the cloud index (CI), from the time of year and season is crucial for enhancing the robustness of solar forecasting [60]. For instance, high-albedo pixels in satellite images, which could be snow or ice, may be incorrectly identified as clouds. Therefore, these upwelling radiance values recorded in satellite data are also subject to inherent diurnal and seasonal variations, potentially introducing noise into the input data and causing significant prediction errors.

[2] https://www.eumetsat.int/eumetsat-data-store.
[3] https://www.eorc.jaxa.jp/ptree/.
[4] https://satellite.nsmc.org.cn/PortalSite/Data/Satellite.aspx.
[5] https://datasvc.nmsc.kma.go.kr/datasvc/html/main/main.do.

To remove the seasonality in satellite data, Matsunobu et al. [61] proposed a method that weighs each satellite spectral band and removes diurnal effects (only for bands within the solar spectrum). This normalization method is mathematically expressed as:

$$\tilde{L}^b_{ij,t} = 1 - \frac{L^b_{ij,t} - \min\left(L^b\right)}{\cos\theta_{z,t}\left[\max\left(L^b\right) - \min\left(L^b\right)\right]} \tag{2}$$

where $\theta_{z,t}$ denotes the solar zenith angle at time t.

Furthermore, Si et al. [19] proposed an advanced pre-processing method designed to remove diurnal effects while preserving valuable information. The initial step in this method involves determining the coefficient (α) related to the relative positions between the satellite and target locations. This is optimally achieved by minimizing the Euclidean distance between the satellite albedo image and ground albedo images, under the assumption that the ground albedo within the target area remains constant over a short time period [19]. The value of a single pixel is then normalized by the solar zenith angle as:

$$L'^b_{ij,t} = \frac{L^b_{ij,t}}{(\cos\theta_{z,t})^\alpha}. \tag{3}$$

Subsequently, $L'^b_{ij,t}$ is further normalized to $\tilde{L}^b_{ij,t}$ employing the *minmax* method as presented in Eq. (1). The method presented by Si et al. [19] is explicitly designed for visible spectral bands, and its applicability to other bands cannot be assured without further research and experimentation.

Additionally, the semi-empirical Heliosat method can be utilized for deriving a cloud index map from satellite images as a data processing method [48,62–64]. For instance, this can be represented as:

$$\mathrm{CI}_{ij,t} = \frac{L^b_{ij,t} - \min\left(L^b_{ij,t-N:t}\right)}{\max\left(L^b_t\right) - \min\left(L^b_{ij,t-N:t}\right)} \tag{4}$$

where $\mathrm{CI}_{ij,t}$ is the cloud index of the pixel situated at the i-row and j-column in a satellite image at time t. $\min\left(L^b_{ij,t-N:t}\right)$ is the minimum pixel value over the preceding N days ($N = 10$ as per [62]) at the same time, which approximates the corresponding clear sky values, and $\max\left(L^b_t\right)$ is the maximum pixel value in the current image.

An alternative approach for removing the diurnal and seasonal variations inherent in satellite data involves the implementation of cloud detection. This technique involves the segmentation or identification of cloud cover in satellite imagery, which is subsequently utilized as input for forecasting models. Cloud detection can be achieved through traditional pixel color ratio methodologies [65], or through the application of more sophisticated end-to-end deep learning techniques [61,66–68]. These methods generate binary or three-state cloud maps by denoting pixels as clear skies, thin clouds, or thick clouds. A more comprehensive overview of cloud detection methodologies can be found in [69]. Notwithstanding, the incorporation of a cloud-detection process complicates the overarching forecasting system, potentially resulting in extended inference times. Moreover, the accuracy of the cloud detection process itself may be susceptible to the influence of the diurnal and seasonal variations of satellite data.

Despite these various efforts, a majority of the solar forecasting literature appears to have largely overlooked the importance of pre-processing satellite data, particularly in terms of seasonality removal, prior to the model estimation phase.

2.4. Performance assessment of spatial solar forecasts

At the current stage of development of the field, the performance assessment of spatial forecasts adapts and resembles that of forecasts for point locations. The assessment is usually performed by comparing the model predictions $\hat{I}$ against measured values I, which are also referred to as the ground truth or labels. Solar forecasts are continuous in values, and there are many scoring functions (i.e., error metrics) that can be used to quantify the forecast quality. Some commonly used metrics are presented in Table A.3, among which a highly recommended scoring function for solar forecasting is the root mean square error (RMSE) [59].

However, forecast accuracies such as RMSE attained from different forecasting contexts cannot be compared directly, because accuracy is highly influenced by factors such as diverse geographic, seasonal, climatic, or particularly meteorological conditions [12,70]. Therefore, reference models often serve as the benchmark to compare the performance of models proposed in different studies, which use different datasets for model development and evaluation. The irradiance forecasted by NWP is sometimes used as a benchmark for forecasts generated by data-driven and/or hybrid methods. The performance of NWP models has been often found to be outperformed by satellite-based or data-driven methods when the forecast horizons are less than four hours [71,72]. In addition, advanced deep learning methods usually employ conventional machine learning models as the reference, and hybrid approaches often employ their individual building blocks as the reference [3,12]. To offer a relative independent baseline, a once-popular standard of reference is *persistence*, which assumes that the current situation persists into the future:

$$\hat{I}_{t+h} = I_t, \tag{5}$$

where I could be either solar irradiance or power. I_t is I at current time t, $\hat{I}$ is the forecast of the persistence model, and h is the forecast horizon. To consider the inherent diurnal cycles in solar irradiance or power time series, the smart persistence (SP) [73,74] has been proposed:

$$\hat{\kappa}_{t+h} = \kappa_t, \tag{6}$$

where κ is the clear-sky index [3], which is defined as solar irradiance or power divided by its clear-sky counterpart:

$$\kappa_t = I_t / c_t, \tag{7}$$

where c_t represents the clear-sky expectation of the forecast variable in question.

Note that the persistence model has shown excellent performance during clear weather conditions. However, the error of the persistence model increases with the forecast horizon rapidly. Therefore, the optimal convex climatology–persistence combination has been advocated by Yang [75,76]:

$$\hat{\kappa}_{t+h} = \gamma_h \kappa_t + (1 - \gamma_h)\mu, \tag{8}$$

where γ_h is the autocorrelation at lag h with a value between −1 and 1, which statistically approaches zero with the increase of forecast horizon. More details about the practical calculation of γ_h can be found in [75,76], which is important because the solar time series contains nighttime gaps. μ is the climatology reference term [76], which is single-valued internal climatology (i.e., the mean value of κ_t in the verification set). The error of the convex model can be mathematically shown to be always less than both the persistence and the climatology references, thus serving as a better standard of reference [59].

Based on the aforementioned reference models, forecast skill (s) is proposed to quantify the performance of a newly developed model when compared with the persistence model [73]:

$$s = \frac{\mathrm{RMSE}_f - \mathrm{RMSE}_r}{\mathrm{RMSE}_p - \mathrm{RMSE}_r}. \tag{9}$$

where the RMSE_f, RMSE_p, RMSE_r are performances of the forecasts from the model of interest, perfect forecasts, and reference forecasts, respectively. Since perfect forecasts are usually unattainable—except when dynamical ensemble NWP forecasts are available [77]—one has no choice but to assume the perfect forecasts have an RMSE of zero. In that, the above skill score equation is reduced to:

$$s = 1 - \frac{\mathrm{RMSE}_f}{\mathrm{RMSE}_r}. \tag{10}$$

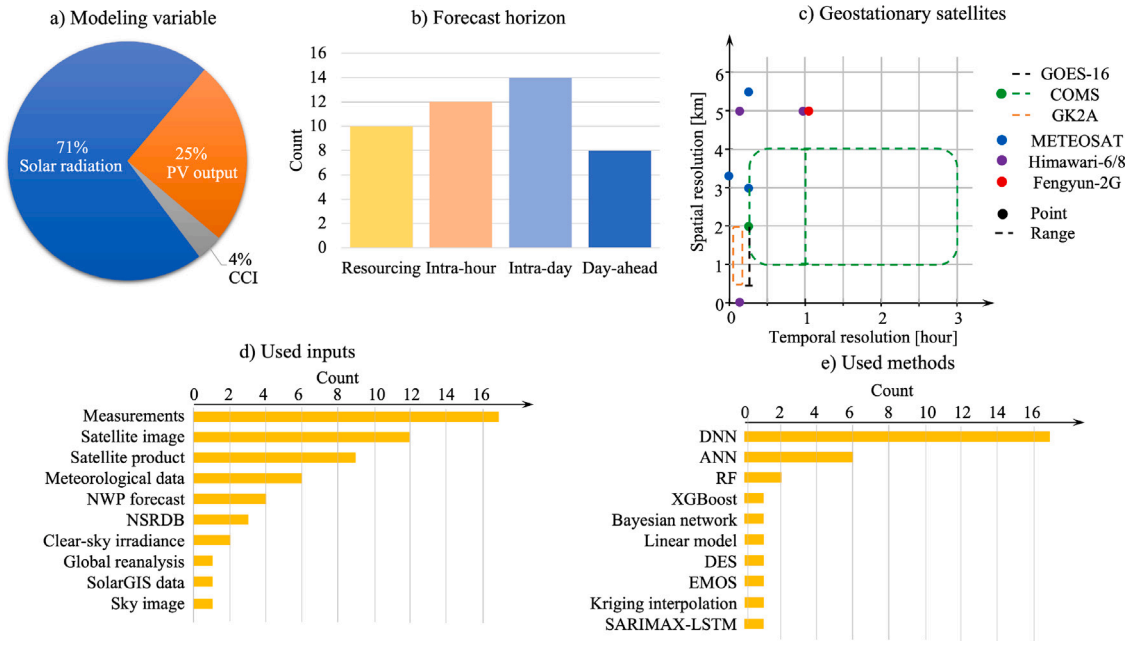


Fig. 5. Summary of the reviewed references. (a) Forecast variables, where solar radiation includes global horizontal irradiance (GHI), direct normal irradiance (DNI) and diffuse horizontal irradiance (DHI). (b) Forecast horizon: For papers with two or more forecast horizons, those are counted separately. (c) Geostationary satellites: All geostationary satellites used in the literature are listed, and their temporal and spatial resolutions are annotated with points (for a single value) and dash lines (for a range). Points with only temporal or spatial information are placed on the axes. (d) Input types: Measurements include historical irradiance data and PV output, and satellite products mean the data derived from satellite measurements. (e) Methods: CNN, long short-term memory (LSTM), graph neural network (GNN) and their combinations are counted as DNN, whereas multilayer perceptron (MLP) is placed under artificial neural networks (ANN).

In general, a positive s indicates the model has more accurate forecasts than the reference model, and vice versa. When compared with other statistical metrics, s is more independent of the forecasting objects (irradiance or power) and climatic or meteorologic factors. Therefore, s can be used to compare the performance of different models from different works that are developed and validated using different data.

3. Spatial solar forecasting with deep learning techniques

A literature search was conducted on the Web of Science using the keywords *deep learning* and *spatial solar forecast*. A total of 45 publications were found, among which only 20 considered satellite data. These references are selected, thoroughly reviewed, and summarized in Fig. 5 and Table 3.

3.1. Overview of deep learning techniques

Deep learning methods, which inherit the structures of ANN [36], are strong analytic tools for data-driven applications. Deep learning methods have been widely adopted in a range of applications, as one of the most promising state-of-the-art computer science technologies. One of the most established and popularly employed ANNs in both research and application is the MLP [12]. For time series analysis, RNNs are often found advantageous. For the analysis of spatial information, CNNs are often perceived as more suitable.

As an effective tool, RNN is capable of processing multiple inputs and thus can analyze temporally dynamic behaviors [78]. More advanced variants of RNN, such as LSTM [79] or gated recurrent unit (GRU) [80], are developed for long time series analysis. More specifically, compared to basic RNN units, LSTM/GRU units have gates to control information reserving and discarding, which is useful to solve the problem of vanishing gradient when the input time sequence is long. LSTM has hitherto been popular in the domain of solar forecasting, but GRU, which has fewer parameters than LSTM, is more friendly towards computational efficiency.

CNNs are effective in extracting spatial features from images. Additionally, CNNs are able to analyze spatially correlated data, such as multi-dimensional image/video data in computer vision tasks. Combining both facts, CNNs align well with the utilization of sky-camera and satellite images, which act as exogenous inputs facilitating spatial solar forecasting. Since processes of solar irradiance or cloud fields are often spatially correlated, significant enhancements in forecasting performance of both centralized or spatially distributed systems have been observed [12]. Commonly used CNN architectures include, but are not limited to, AlexNet [81], ResNet [41], DenseNet [82], and Inception net [83]. More details about CNN can be found in [41,84,85], and other comprehensive reviews of deep learning methods can be found in [36,86].

While deep-learning-based methods for solar forecasting have demonstrated promising results in the literature, particularly due to their ability to bypass error-prone intermediate physical processes and directly produce the final prediction, these methods do come with limitations. First, such models require large amounts of high-quality data and robust computational resources, which can significantly increase overall system costs. As a result, their use is often confined to locations with ample local measurements for model development and validation. Second, deep learning models can be prone to overfitting, leading to a decrease in performance when addressing unseen data. For instance, predictions pertaining to sudden weather changes that are absent from the training dataset are likely to exhibit substantial errors. Third, models refined for one location may not perform as well in a different location due to variations in local conditions and system configurations. The dynamic nature of city-scale DSGs, including new installations, decommissioning, and maintenance events, presents a significant challenge to the adaptive capabilities of data learning models. Last, the "black box" nature of these models undermines their transparency and trustworthiness, a factor that is crucial when understanding the decision-making process is important. Consequently, grid operators may have concerns regarding the explainability of model predictions and the diagnosability of potential model failures.

Table 2
Summary of deep learning frameworks.

Frameworks	Initial release	Platform	Written language	Open source	Official link
Theano	2007	Cross-platform	Python	Yes	www.deeplearning.net/software/theano
Deeplearning4j	2014	Linux, MacOS, Windows, Android (Cross-platform)	C++, Java	Yes	www.deeplearning4j.org
TensorFlow	2015	Linux, MacOS, Windows, Android, JavaScript, etc.	Python, C++	Yes	www.tensorflow.org
Keras	2015	Linux, MacOS, Windows, Cross-platform.	Python	Yes	keras.io
MXNet	2015	Linux, MacOS, Windows	C++, Python, R, Java, Julia, JavaScript, Scala, Go, Perl	Yes	mxnet.apache.org
Chainer	2015	Linux, MacOS	Python	Yes	chainer.org
PyTorch	2016	Linux, MacOS, Windows, etc.	Python, C++	Yes	pytorch.org
CNTK	2016	Linux, Windows	C++	Yes	www.microsoft.com/en-us/cognitive-toolkit
Caffe	2017	Linux, MacOS, Windows	C++	Yes	caffe.berkeleyvision.org
Flux	2017	Linux, MacOS, Windows (Cross-platform)	Julia	Yes	https://fluxml.ai
ONNX	2017	Windows, Linux	Python, C++	Yes	onnx.ai
Matlab Deep Learning Toolbox	–	Linux, MacOS, Windows	C, C++, Java, MATLAB	No	mathworks.com/products/deeplearning

3.2. Tools to implement deep learning methods

Building a deep learning model from scratch is neither efficient nor economical. Therefore, most researchers place their attention on the specific applications, rather than implementation. On the contrary, utilizing well-tested deep learning frameworks to develop the model of interest has been found sufficient in most cases. Although other options are available (see Table 2), the TensorFlow (Keras) and PyTorch frameworks are the most representative and therefore recommended, in consideration of their open-source nature, as well as rich community and documentation support. Indeed, most of the state-of-the-art deep learning methods and architectures are openly available in these frameworks. Detailed mathematical theories behind several representative deep learning algorithms are presented in Appendix B. In what follows, a brief account of Tensorflow and PyTorch is given.

TensorFlow supports a variety of platforms, such as Linux, Windows, MacOS, iOS, or Android, and it can work with a range of programming languages, such as Python, C++, or GO. TensorFlow is particularly popular in commercial and industrial applications due to its computational efficiency. TensorFlow is based on a symbolic math library with excellent documentation on its official website containing all modules. Therefore, researchers can customize deep learning models with TensorFlow, but its computational graph is not easy to use for beginners. For beginners, Keras, which is used on top of TensorFlow, is thought to be the better option. Keras is an open-source high-level deep learning API, written in Python. Keras allows fast experimentation, so researchers can quickly prototype and validate their ideas and concepts. However, Keras is less configurable to build large or new deep learning models.

PyTorch is developed based on Torch, which is mainly used for sequence models, reinforcement learning models, computer vision models, and relational models. It is written in Python, and its API is similar to other deep learning frameworks like TensorFlow. Similar to TensorFlow, PyTorch is compatible with all mainstream platforms and is strongly supported by great documentation and developer community. Whereas both PyTorch and TensorFlow are popular in research and development communities, the former has advantages in terms of flexibility and ease of use.

3.3. Solar forecasting for multiple point locations

For spatial solar forecasting methods using deep learning, early works were mostly interested in forecasting separately at point locations within an area. The general procedure first trains an MLP or an extended version of that (i.e., DNN) to provide deterministic forecasts for point locations [87,88], and then derives the spatial forecast maps based on spatial statistics, interpolation, or other space-filling strategies [11,88–90] (as shown in Fig. 2). Some proposed models are optimized to provide forecasts for multiple locations simultaneously, which can significantly lower the efforts required for model training and testing [87]. These works mostly employ historical irradiance/power time series and satellite images as default model inputs. To further enhance the model performance, some works employ NWP forecasts and meteorological data, such as temperature and relative humidity, as additional exogenous inputs [87,90]. At present, most methods available in the literature are developed at an hourly resolution.

The aforementioned deep learning models are either in the form of MLP or as hybrids with other methods that integrate conventional machine learning and/or data analysis, such as support vector machine (SVM) [91], principal component analysis, or wavelet transform analysis [92]. These methods are developed and validated independently for locations in different regions with different local climates, such as the Brazilian Northeastern region [89], Italy [90], Netherlands [87], Greece [93], Reunion island [92], and South Korea [88,94]. Consequently, the reported accuracies of these models are not directly comparable. When validated using local data, most works suggest that deep-learning-based forecasting methods have superior performance over conventional methods, such as random forest (RF), support vector regression (SVR), or unprocessed physical models [88,94].

3.4. Advanced approaches to analyze spatial–temporal information

To investigate the spatial–temporal correlations of solar irradiance/power time series measured at different locations, more complex deep learning architectures are employed for spatial solar forecasting. In this section, RNNs for analysis of time series and CNNs for analysis

of spatial information, particularly for satellite image feature extraction and analysis, are to be discussed.

Since the data involved in forecasting, whether as input or output, is often time series, it is commonplace to employ an RNN to describe the input–output relationship. RNNs are capable of predicting the sequences of solar irradiance or power based on input sequences of historical solar irradiance, satellite and local-sensing images, power measurement, or other meteorological variables. For example, Srivastava and Lessmann [95] used an LSTM model to forecast day-ahead solar irradiance using satellite data. The proposed model was applied and validated at 21 locations (16 in Europe and 5 in the United States). The forecasting results suggested that the LSTM model outperformed gradient boosting regression (GBR) and feed-forward neural networks (FFNN) in terms of day-ahead GHI forecasting. LSTM and other RNN models can also be integrated with conventional machine learning models to perform solar forecasting. For instance, Kim et al. [96] applied a stacking ensemble technique to develop a hybrid seasonal autoregressive integrated moving average with exogenous factors (SARIMAX)-LSTM model that used satellite images and local PV power data as inputs. This model analyzed and predicted spatial and temporal characteristics for PV power in various regions of South Korea. The SARIMAX-LSTM model had superior performance to the benchmark DNN, LSTM, and SARIMAX models.

Unlike LSTM, CNN is usually combined with a regression (e.g., MLP) to conduct solar forecasting [19,97]. The purpose of CNN is to analyze the satellite images and extract features (e.g., cloud cover information) from those. Subsequently, these extracted features are combined with locally measured irradiance or meteorological data to form the inputs to the regression to generate solar forecasts. The CNN-derived feature map consists of spatial information, which could be used to estimate a spatially continuous forecast map to cover multiple locations [97]. Most CNN+MLP structures are developed for hourly spatial irradiance forecasting. The CNN integrated models have been reported to achieve improved performance against stand-alone CNN, MLP, and unprocessed commercial NWP solutions, such as the European Centre for Medium-Range Weather Forecasts (ECMWF) [97–99].

The above works employ either CNNs or RNNs independently, whereas methods were also proposed to combine both CNN and RNN structures, such as CNN+LSTM/GRU, to simultaneously investigate spatial and temporal information in an end-to-end fashion. Although named differently, the basic structure of these hybrid models resembles those CNN+MLP structures discussed above. The CNN parts of these models are used to analyze satellite images for feature extraction, while the MLP part is replaced by the RNN unit, which is capable of analyzing time series and generating sequential forecasts during the inference stage. Therefore, satellite images, PV power, irradiance, and other meteorological data can be prepared in the form of time sequences to estimate the model parameters in an end-to-end learning approach. Among these hybrid models, most works employ LSTM while only a few works [100] employ GRU.

Hybrid models have been developed to provide spatial–temporal forecasts for both PV power [14,100] and surface irradiance [21,101–103] with forecast horizons ranging from one hour to one day. Some hybrid structures can serve forecasting for up to 400 DSGs with carefully prepared datasets [100]. Developed and validated in different locations, such as the United States [14], China [100], Brazil [103] and South Korea [101], hybrid model architectures have shown significant advantages in improving the overall forecasting performance over a set of reference models, such as stand-alone autoregressive integrated moving average (ARIMA), MLP, RNN, GRU and LSTM [104,105].

3.5. Other state-of-the-art approaches

In the realm of solar forecasting, researchers are progressively adopting state-of-the-art deep learning methodology, such as innovative learning strategies and application frameworks. For example, the deep generative framework, which usually has encoder–decoder architectures [106] where the backbones of the encoder or decoder are usually U-net or other CNN variants. These models are able to cooperatively process both spatial meteorological data and satellite image data in order to directly generate spatial solar forecasting. Compared with other machine learning models, such as ARMA, ARIMA, GRU, LSTM-fully connected (LSTM-FC), or CNN-bidirectional LSTM (CNN-biLSTM), these models are reported to have better performance for various forecast resolution and are capable of serving up to 50 real-world PV stations [13,106]. Other innovative approaches for satellite-based spatial forecasts include, but are not limited to graphical learning framework [72] and extreme learning machine (ELM) [107], which can investigate areas as large as a country (e.g. Australia). To address the challenge that spatial distributed PV systems have different capacities and configurations, a machine learning framework [108], which integrates different data sources to construct a database for regional models, has been proposed to capture the point-to-point relation between satellite data and in-situ measurements. This method is suggested to be site-adaptive to any region where both satellite and ground data are available.

In addition to statistical forecast accuracy, some works investigate the enhancements of other utilization values of forecasts. For example, most satellite-image-based works are developed for intra-day forecasting with hourly predicting resolution. To achieve even higher resolutions (e.g. intra-hour forecasting), Cheng et al. [72] proposed a spatial–temporal graphical learning framework (i.e., GNN) for intra-hour PV power prediction using historical image sequences as inputs. This framework used bi-directional extrapolation to simulate cloud motion and generated a directed graph to represent the shapes and motion directions of the regions of interest. It is suggested that GNN was more flexible for varying sizes of inputs, decreased the number of model inputs and parameters, reduced the data storage requirements, and was more practical for deployment of intra-hour and intra-day PV power forecast with horizons from 30 min to 3 h. For even higher resolution, Yao et al. [106] proposed a hybrid model that integrated both U-net and encoder–decoder architecture to achieve intra-hour forecasting of PV power. This model cooperatively processed both time series of historical data and satellite image data to learn both the spatial and temporal features. Compared with other machine learning models, such as ARMA, ARIMA, GRU, LSTM-FC, or CNN-biLSTM, the hybrid model is reported to have the best performance within a 1-h horizon at 15-min resolution. On the contrary, some works develop models for much longer time intervals. For example, Deo et al. [107] developed an ELM that was regionally adaptive to achieve monthly-scale irradiance forecasting. Using eight Moderate Resolution Imaging Spectroradiometer (MODIS) satellite data variables and geo-topographical site characteristics as inputs, this model was validated and deployed at 41 local sites distributed over Australia.

Another approach to enhance the utilization values for spatial solar forecasting is to estimate the uncertainty associated with the prediction map. There are always errors associated with forecasts in practice [109,110]. Therefore, probability density function forecasts convey more useful information for the decision-making of grid operators [111,112]. For example, Yang [113] proposed a probabilistic ensemble model output statistics (EMOS) method to provide regional scale solar irradiance forecasts for intra-hour horizons. The EMOS model allowed variance scaling and smoothing to optimize the prediction methods. However, to the best knowledge of the authors, very few research has investigated probabilistic spatial solar forecasting. More discussion about probabilistic solar forecasting for point locations can be found in [12,50].

In general, a hybrid architecture or learning framework, such as hybrid CNN+RNN models or attention-mechanism-based methods, is able to analyze time series data and generate sequential predictions. These hybrid models have demonstrated advantages in improving spatio-temporal forecasting performance and have been reported to outperform simple machine learning models, such as ARMA, ARIMA, GRU,

LSTM-FC, or CNN. It is important to note that the performance of these models may vary depending on the specific dataset, location, and climate conditions. Further research and validation are necessary to determine the best-performing model for a given circumstance.

In summary, enhanced with deep learning techniques, spatial forecasting has been proposed in the literature to issue forecast maps for cloud, solar irradiance, and distributed PV power, using satellite data with or without local measurements [93]. Related references are summarized and compared in Fig. 5 and Table 3. However, spatial forecasting for DSG is relatively challenging since individual DSG sites may have different configurations, such as capacity, efficiency, or tracking methods. Therefore, available methods usually develop forecast frameworks that could be quickly adapted to new DSG sites within the investigated regions when locally measured data (e.g. power) becomes available. Spatial forecasting of cloud and/or solar irradiance is sometimes employed as exogenous inputs to enhance the performance of these frameworks [114].

4. Outlooks for future research

The majority of solar forecasting research focused on predicting irradiance/power output for a single/point location. However, it is challenging to deploy these point-location-oriented forecasting methods for a large spatial area because: (1) Most available forecasting engines require extensive local-sensing data to develop and validate the models, particularly for deep learning methods. However, it is very costly to deploy a large number of distributed local-sensing systems to collect the required data. (2) The relationship between solar irradiance and solar power from individual PV systems is highly dependent on system configurations such as panel orientations, local shading, local meteorological conditions, and sun-tracking mechanism [116,117]. Therefore, for solar power forecasting, the models developed for one site may not be suitable to be directly applied to other sites. (3) Once a data-driven model is trained and deployed in real-time operation, it may not adapt to unexpected noise or changes in the system or environment, such as efficiency degeneration due to dirt or aging, partial system malfunction, and shading from growing trees or new buildings. Currently, most available solar forecasting methods are developed based on the assumption that the system will be operating stably without considering the conditions of partial malfunctions or unexpected service disruptions. Therefore, more advanced and practical forecasting methods that are capable of addressing the above issues should be developed. Here, a few potential approaches for spatial solar forecasting are presented in the following with the aim of overcoming the challenges.

One potential approach is to develop practical and economical solutions that provide full-spectrum forecasts with minimum requirements of local-sensing equipment. Although local-sensing systems have significant advantages in terms of data fidelity and temporal resolutions, especially for intra-hour and intra-day forecast horizons, installation, maintenance, data synchronizing, and preprocessing costs of local-sensing systems will scale up dramatically with the number of sites of interest. In addition, deployment of local-sensing systems usually requires complicated procedures to obtain hardware installation permits due to considerations including but not limited to security, safety, and weatherproof. Therefore, with the advancement of new generations of geostationary satellites, utilizing remote-sensing satellite data is potentially an alternative approach to deliver intra-hour and intra-day solar forecasts for numerous DSGs over large areas. Yagli et al. [118] investigated the performance difference between ground-based and satellite-based forecasting models. The results suggested that satellite-based forecasts have comparable performance to ground-based forecasts. Therefore, with the advancement of deep learning and spatial interpolation techniques, forecast engines using only exogenous inputs such as satellite data may be proposed. The advanced methods are expected to be quickly adapted to any location with simple system configuration information as local inputs (e.g. panel orientations and shading factor).

In recent years, the deep learning research community has been rapidly evolving and innovative methods have emerged. Therefore, the employment of more pioneering deep learning methods to analyze the spatially distributed solar resource or power is suggested. For example, the attention mechanism [119] based methods could address the spatial solar forecasting problems. The deep-learning-based attention model mimics the biological cognitive attention mechanism that focuses on specific aspects of a complex input [120]. As a result, the attention mechanism divides a complicated problem into smaller and simpler tasks [121]. The attention mechanism has been successfully employed in computer vision tasks, such as classification, object detection, semantic segmentation, video understanding, image generation, and 3D vision [122–125]. There are two major types of attention mechanisms for computer vision tasks: soft attention and hard attention [126]. Soft attention usually employs a group of filters to create a blurring effect that the background is faded or blurred while the ROI is in the focus. For hard attention, it detects the ROI and discards the entire background. More information about computer-vision-orientated attention mechanism can be found in [127]. The representative attention methods are promising for spatial solar forecasting include but are not limited to SENet [128], Image GPT [129], Vision Transformer [130], and Swin Transformer [131].

Another potential approach is to employ advanced deep learning methods that are capable of video or image sequence forecasts. Compared to other meteorologic factors such as temperature, pressure, and aerosol concentrations, volatile cloud fields are the major factors that cause intra-hour and intra-day solar ramps [65]. Therefore, the temporal–spatial features of cloud distribution and movement are useful exogenous inputs for intra-day solar forecasting. Satellite image sequence contains essential information to extract useful cloud properties and cloud movement features. For example, Lu et al. [132] proposed a hybrid method that integrated both Cascade Causal LSTM and Super-Resolution Network to predict the shape and speed of cloud motion based on sky images. For remote-sensing-based methods, Xu et al. [133] predicted the next few images of the satellite image sequence based on a hybrid of generative adversarial networks (GAN) and LSTM. This work suggested that the GAN-LSTM model is capable of combining the generating ability of the GAN with the forecasting ability of the LSTM. Similarly, Rüttgers et al. [134] predicted a typhoon track using a GAN that analyzes the sequence of satellite images. However, how to enhance the aforementioned approaches for spatial solar forecasts requires further investigation. Therefore, it is suggested to develop end-to-end spatial solar forecasting approaches using the analysis of both spatial and temporal features extracted from the sequence of satellite images as inputs.

In addition, probability forecasting is another potential approach to further assist solar integration [12]. Most available solar forecasting methods, either for the point location or the large areas, predict deterministic values. However, there are always inherent and irreducible errors associated with deterministic predictions regardless of the data processing, training methods, and model mechanisms [109,110]. Therefore, the quantified solar forecast uncertainty is useful information for grid operators to integrate solar power while minimizing the adversary impacts from forecast errors. The methods to estimate solar forecast uncertainty include but not limited to delta techniques [135], Bayesian methods [112], bootstrap method [136], bootstrap-ANN method [137], quantiles methods [138,139], lower upper bound estimation method [110], kNN ensemble model [50], naïve Bayes classifier and Bayesian models [140,141]. More details of probabilistic forecasting approaches can be found in [142–144]. Here, employing the aforementioned methods for spatial solar forecasting and generating spatial distributed forecast uncertainty map is suggested for future work.

Table 3
Summary of the publications of spatial solar irradiance or power forecasting based on remote sensing and deep learning.

Authors	Year	Forecast variables and horizons	Temporal and spatial resolution of satellite data		Input local-sensing data	Input remote-sensing data	Methods
Lima et al. [89]	2016	24-h-ahead GHI	Not used	Not used	2 years of hourly GHI measurement data of 110 weather stations in northeastern Brazil	Not used	Weather Research and Forecasting (WRF) and ANN
Pierro et al. [90]	2017	0 to 48-h-ahead PVoutput	Not provided	Not provided	2 years of PV output data from 11 sites in the south of Italy and irradiance data forecast by WRF-AWR	2 years of satellite-derived irradiance data from METEOSAT-10	ANN, referenced by smart persistent model
Lago et al. [87]	2018	0 to 6-h-ahead GHI	Not provided	Not provided	4 years of daily ECMWF forecasts, Ineichen–Perez clear-sky irradiance	4 years of satellite-derived irradiance data from Royal Netherlands Meteorological Institute	DNN, referenced by persistence model, linear model, extreme gradient boosting (XGBoost) and ECMWF forecasts
Srivastava and Lessmann [95]	2018	0 to 24-h-ahead GHI	Not provided	Not provided	Not used	10 years of satellite-based GHI data from 21 stations in Europe and US	LSTM, referenced by FFNN, GBR and persistence model
Zhang et al. [91]	2019	0 to 30-min-ahead PV output	Not used	Not used	1 year of PV output data and 6 types of meteorological data from two open data portals	Not used	Bayesian Network, referenced by ARIMA, k nearest neighbor (kNN), LSTM, persistent model and spatio-temporal baseline methods
Yeom et al. [94]	2019	GHI estimation	1 h	1 to 4 km	6 years of 35 ground pyranometers from Korea Meteorological Administration (KMA)	6 years of satellite images from COMS	Physical model, ANN, RF, SVR and DNN, RF and DNN show better results
Nikitidou et al. [93]	2019	0 to 6 h-ahead cloud clearness index (CCI)	15 min	5.5 km	Not used	3 years of satellite-derived CCI from METEOSAT Second Generation	ANN
Jeong and Kim [14]	2019	0 to 6-h-ahead PV output	Not used	Not used	1 year of PV output from National Renewable Energy Laboratory in the USA	Not used	CNN, referenced by AR, FFNN, and LSTM
Liu et al. [100]	2019	3-h-ahead GHI	Not used	Not used	2 years of hourly GHI and meteorological data from National Solar Radiation Database (NSRDB)	Not used	Variational Bayesian convolutional GRU, referenced by simple DNN, RNN and LSTM
Jiang et al. [98]	2019	GHI estimation	Not provided	Not provided	2 years of hourly GHI of 98 ground irradiance stations from CMA	2 years of satellite images from Himawari-6	DNN, referenced by DNN with less inputs and ANN
Khodayar et al. [13]	2019	0 to 2-h-ahead GHI	Not used	Not used	18 years of GHI data in 75 sites near Lake Michigan from NSRDB	Not used	Convolutional graph autoencoder, referenced by space–time Copula, spatio-temporal lasso quantile regression, Compressive Spatio-temporal Forecasting and spatio-temporal SVR
Doorga et al. [115]	2019	0 to 5-day-ahead GHI	Not provided	3.3 km	3 years of solar irradiance in the island of Mauritius	3 years of satellite data from METEOSAT First Generation	ARMA, nonlinear autoregressive neural network and double exponential smoothing (DES), DES has better results
Deo et al. [107]	2019	GHI	Not provided	Not provided	4 year of GHI from Scientific Information for Land Owners in Australia	4 years of satellite-derived data from MODIS Terra sensor	ELM (originated from ANN), referenced by RF, M5 Tree, multivariate adaptive regression spline
Yang [113]	2020	GHI estimation	Not used	Not used	6 satellite-derived databases and 2global reanalysis databases	Not used	EMOS technique
Li et al. [92]	2020	GHI estimation	Not used	Not used	18 years of global solar irradiance data from CM SAF	Not used	ANN
Kim et al. [88]	2020	1 to 2-h-ahead PV output	15 min	0.5 to 2 km	1 year data of PV power output from SK Telecom	1 year of satellite images from Himawari-8 and COMS	ANN, DNN and SVM, DNN has better results

(*continued on next page*)

Table 3 (*continued*).

Authors	Year	Forecast variables and horizons	Temporal and spatial resolution of satellite data		Input local-sensing data	Input remote-sensing data	Methods
Si et al. [19]	2020	0 to 4-h-ahead GHI	1 h	5 km	2 years of hourly GHI data and meteorological data in Shandong from CMA	2 years of satellite images from Fengyun-2G	CNN+ANN, referenced by very deep convolutional networks, LSTM+ANN
Pérez et al. [97]	2020	0 to 6-h-ahead GHI	15 min	3 km	2.5 years of GHI data in the southern part of France from a pyranometer	2.5 years of satellite images from METEOSAT Second Generation	DNN, referenced by smart persistent model
Yeom et al. [101]	2020	GHI estimation and 1-h-ahead forecast	15 min to 3 h	1 to 4 km	4.75 years of GHI measurement from 33 ground station from KMA in Korea	4.75 years of satellite images from COMS	ConvLSTM, referenced by RF and ANN
Hong et al. [102]	2020	0 to 24-h-ahead GHI	1 h	Not provided	1 year of GHI data from the SolarGIS database	1 year of satellite-derived GHI data	ConvLSTM, referenced by CNN-LSTM, ARIMA and LSTM
Kim et al. [96]	2021	0 to 1-h-ahead PV output	15 min	2 km	4 years of PV output data in South Korea from Open Data Porta, meteorological data	4 years of satellite images from COMS	SARIMAX, SVR, LSTM, DNN, RF and SARIMAX-LSTM, SARIMAX-LSTM shows better results
Narvaez et al. [108]	2021	GHI estimation and 24 to 168-h-ahead GHI forecasts	30 min	111 km	10 years of 30 min GHI, DNI, DHI and meteorological data in Columbia, South America	10 years of satellite data from NSRDB	RF for estimation, LTSM+GRU for forecast
Yao et al. [106]	2021	0 to 1-h-ahead PV output	10 min	Not provided	9 months of PV output data from 50 PV stations, NWP products, meteorological data	9 months of satellite images from Himawari-8	CNN+LSTM+AM, referenced by ARMA, ARIMA, GRU, LSTM-FC, CNN-biLSTM
Cheng et al. [72]	2021	0.5 to 3-h-ahead PV output	10 min	5 km	2 years of PV output from DKA, clear-sky irradiance, meteorological data	2 years of satellite images from Himawari-8	GNN, referenced by smart persistence, deep belief network, ConvLSTM, 3D-CNN, and 2D-MC CNN
Chu et al. [11]	2022	GHI, DNI and DHI estimation	Not used	Not used	2 months of images from sky cameras in California	Not used	Linear model for DHI, MLP for GHI and DNI, and kriging for spatial GHI
Jang et al. [99]	2022	GHI estimation	2 to 10 min	0.5 to 2 km	1 year of GHI measurement from 81 ASOS stations	1 year of satellite images from GK2A	CNNs
Qin et al. [21]	2022	0–6-h-ahead GHI	1 h	5 km	1 year of GHI measurement from 10 ground stations in China	1 year of satellite images from Himawari-6	CNN+LSTM, referenced by CLSTM, ConvLSTM and smart persistence
Rocha and Santos [103]	2022	GHI and DNI estimation	15 min	0.5 to 2 km	GHI, DNI DHI measurement data from SONDA station in Brazil	Satellite images from GOES-16	XGBoost and CNN-LSTM

Since 2021, large-scale models have emerged in leading journals for weather forecasting based on satellite imagery [145,146]. These pioneering techniques provide promising opportunities for developing innovative forecasting strategies that can effectively support a wide range of real-world DSG operations. For instance Pangu-Weather, a deep learning-driven weather forecasting system, utilizes reanalysis weather data to predict future weather conditions. It employs a 3D Earth-specific transformer architecture and hierarchical temporal aggregation, enhancing accuracy while reducing computational time. Tested on ERA5 data, this system has demonstrated notable improvements in accuracy compared to other models [147]. Similarly, the FourCastNet model, a data-centric, high-resolution weather forecasting solution, is engineered to predict variables such as surface winds and precipitation. It offers superior resolution compared to traditional models and delivers performance comparable with physics-based models [148]. Other large models for precipitation forecasting are also discussed in the literature [145,149]. Despite these advancements, there remain substantial research gaps in the application of these models to solar forecasting, which has higher variability when compared with other weather variables. These issues include low data quality and availability, high computational resource requirements and high spatio-temporal resolution requirements of solar forecasting. As such, the adaptability of these weather-oriented large-scale models to predict DSG generations with diverse parameters still need further research efforts.

In summary, there are still gaps between state-of-the-art spatial solar forecasting methods and grid demands, particularly in the era of massive deployment of distributed solar power techniques in different regions. In addition to solely focus on statistical accuracy, the application scenarios and associated grid operational policy should be carefully analyzed to identify the optimal forecasting approach. In the future, research of solar forecasting shall be more application orientated, cost-effective, and practical to be deployed in order to satisfy the immediate needs of power integration.

5. Conclusion

This review focuses on the current state-of-the-art spatial solar forecasting methods, particularly those that employ deep learning and remote-sensing techniques, with a special emphasis on grid-scale DSGs over large areas. Fundamental considerations for spatial solar forecasting, including data-driven forecasting procedures, satellite data sources and processing, tools for implementing deep learning methods, and performance assessments, are provided in a comprehensive summary to facilitate further research. Subsequently, the available spatial solar forecasting methods, their inputs, and associated deep learning models are reviewed, compared, and discussed.

At present, forecasting solutions for utility-scale DSGs are still in the early stages of research and implementation. There has been limited

investigation into this field due to various challenges, such as high scale-up costs, limited flexibility for adaptation to real-life operations, and the complexity involved in considering the diverse configurations of installed DSGs. However, advances in remote-sensing techniques and deep learning methods have led to the identification of several promising approaches that could enhance spatial solar forecasting solutions. There is a need for more research efforts to develop next-generation adaptive spatial solar forecasting methodologies, which should be capable of meeting both the technological and commercial demands posed by the substantial integration of DSG systems.

CRediT authorship contribution statement

Yinghao Chu: Conceptualization, Methodology, Formal analysis, Validation, Writing – original draft. **Yiling Wang:** Investigation, Writing – original draft. **Dazhi Yang:** Validation, Investigation, Writing – review & editing. **Shanlin Chen:** Visualization, Writing – review & editing. **Mengying Li:** Methodology, Software, Validation, Investigation, Resources, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

Declaration of competing interest

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

Data availability

Data will be made available on request.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the revision process of this work, the authors used GPT-4 in order to improve the language. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Acknowledgments

Yinghao Chu is substantially supported by a grant from City University of Hong Kong (Project No. 9610625). Mengying Li is substantially supported by a grant from the Research Grants Council of Hong Kong (Project No. 25213022).

Appendix A. Supplemental information

Commonly employed NWP models for solar forecasting include the ECMWF [16], the global forecast system (GFS) [150], the Hirlam–Aladin research towards mesoscale operational NWP in Euromed (HARMONIE) model, the rapid refresh (RAP) model, the North American mesoscale (NAM) model, the high resolution rapid refresh (HRRR) model, the WRF model and its solar parameterization [151,152], and the global and regional assimilation and prediction system - global forecast system (GRAPES_GFS) [153]. The comparison of different NWP models are presented in Table A.1, and the start, central and end wavelengths (in μm) of each band of different satellite imagers are compared in Table A.2.

The commonly used accuracy evaluation metrics are summarized in Table A.3. The European and International Energy Agency recommends the use of mean biased error (MBE), RMSE, and Kolmogorov–Smirnov integral (KSI) to assess the solar forecast performance [154]. Industrial/utility regulations frequently require relative errors such as relative mean absolute error (rMAE) and relative RMSE (rRMSE) as mandatory assessment metrics [3]. Zhang et al. [155] evaluated statistical metrics based on Western Wind and Solar Integration Study Phase 2 data [156] and suggested that most of these metrics are effective in assessing uniform forecast improvements. Therefore, the assessment of solar forecasts is application-oriented that based on local regulations and application/operation requirements.

Appendix B. Deep learning algorithms

Deep learning methods [36] are strong analytic tools for data-driven learning, which have been widely applied for a range of applications as one of the most promising state-of-the-art technologies. Here, the mathematical theories behind several representative deep learning methods are presented. More comprehensive review of deep learning methods can be found in [36,86].

B.1. Artificial neural networks (ANNs)

ANNs are inspired by the biological neural systems [159]. Computations of each ANN neuron, which is responsible for signal processing tasks, are mathematically defined as

$$y = \sigma_A\left(\sum_{i=1}^{N} w_i x_i + w_0\right), \tag{B.1}$$

where x is the neuron input and y is the neuron output, w_i and w_0 are the weights and bias, σ_A is the activation function, such as sigmoid, hyperbolic tangent, rectified linear unit (ReLU) and Leaky ReLU functions [160,161]. The widely used activation functions and corresponding derivatives are summarized in Table B.4, while more detailed discussion of deep learning activation functions can be found in [160,162]. One of the most popularly employed ANNs in both research and application is the MLP [12]. The MLP neurons are placed in layers, and neurons in one layer are fully forward connected with neurons in the following layer. The outputs of neurons in one layer are used as the inputs for neurons in the following layer.

B.2. ANN parameter estimation

The weights and bias of ANNs are mostly estimated using back propagation learning [163,164], which is an supervised learning method. The general procedure of back propagation training is:

1. Set the number of layers N_L and the number of neurons N_N for each layer to define the architecture of the ANN.
2. Set the training parameters, such as tolerance parameter $e_T > 0$, the learning rate α_L, and the maximum epochs of data learning.
3. Initialize the ANN weight vectors $w^l_{k_l k_{l-1}}$ for all layers $l = 1, 2, \ldots, N_L$ and all neurons k_l in layer l and k_{l-1} in layer $l-1$. Initialization can be done randomly or by using pre-trained weights from another dataset.
4. For each input vector x, compute the output of each neuron $y^l_{k_l} = \sigma_A\left(\sum_n w^l_{k_l k_{l-1}} y^{l-1}_{k_{l-1}}\right)$ for all layers l and all neurons k_l in layer l.
5. Compute the loss ε of the current iterative training epoch. The loss function could be the mean square error or cross-entropy loss between the network outputs and the training targets.
6. Compute the derivative of the cost function with respect to the output of each neuron in the output layer, then compute the error $\delta^{N_L}_{k_{N_L}}$ for each output neuron k_{N_L} as: $\delta^{N_L}_{k_{N_L}} = \partial\varepsilon/\partial y^{N_L}_{k_{N_L}}\, \sigma'_A\left(z^{N_L}_{k_{N_L}}\right)$, where z represents the input of the specific neuron.
7. For each layer $l = N_L - 1, N_L - 2, \ldots, 1$, compute the error $\delta^l_{k_l}$ for each neuron k_l as: $\delta^l_{k_l} = \left(\sum_{k_{l+1}} w^{l+1}_{k_l k_{l-1}} \delta^{l+1}_{k_{l+1}}\right)\sigma'\left(z^l_{k_l}\right)$.
8. Update the neuron weights by: $w^l_{k_l k_{l-1}} = w^l_{k_l k_{l-1}} - \alpha_L \delta^l_{k_l} y^{l-1}_{k_{l-1}}$.

Table A.1
Comparison of different NWP models.

Name	Spatial resolution [km]	Atmo-sphere layers	Update frequency [h]	Forecast horizon [h]	Forecast resolution [h]	Covered region	Main output variables
ECMWF(HRES)	9	137	6	90	1	Global	GHI, DNI, extraterrestrial irradiance, surface pressure, total column water vapor, total column ozone, forecast albedo, cloud cover, and air temperature
GFS	13	127	1	120/240 /384	1/3 /12	Global	Air temperature, wind, precipitation, soil moisture, and atmospheric ozone concentration
HARMONIE	2.5	65	6	48	1	Europe	GHI, DNI, air temperatures, relative humidity, cloud cover, pressure, and precipitable water
RAP	13	51	1	18	1	North American	Wind, precipitation, air temperature and cloud cover
NAM	12	60	1/3	84	1/3	North American	GHI, DNI, DHI and air temperature
HRRR	3	51	1	18/48	0.25/1	North American	Air temperature, geopotential height, dew point temperature, graupel, cloud mixing ratio and cloud ice
WRF-Solar	1–36	30–50	1–6	72–96	1–6	Regional	GHI, DNI, and DHI
GRAPES_GFS	28	60	6	240	6	Global	Double-moment cloud physics, cloud macrophysics and prognostic cloud

Table A.2
Start, central and end wavelengths (in μm) of each spectral band of different satellite imagers. Here, VIS denotes visible, NIR demotes near infrared and IR demotes infrared.

	GOES-16 ABI			Meteosat-11 SEVIRI			Himawari-8 AHI			Fengyun-4 AGRI			GK2A AMI		
	Start	Central	End	Start	Central	End	Start	Central	End	Start	Central	End	Start	Central	End
VIS	0.45	0.47	0.49	0.60	0.75	0.90	0.43	0.47	0.48	0.45	0.47	0.49	0.43	0.46	0.48
	0.59	0.64	0.69	0.56	0.63	0.71	0.50	0.51	0.52	0.55	0.65	0.75	0.50	0.51	0.52
				0.74	0.81	0.88	0.63	0.64	0.66				0.63	0.64	0.66
													0.85	0.86	0.87
NIR	0.85	0.86	0.89	1.50	1.64	1.78	0.85	0.86	0.87	0.75	0.83	0.90	1.37	1.38	1.38
	1.37	1.38	1.39				1.60	1.60	1.62	1.36	1.38	1.39	1.60	1.61	1.62
	1.58	1.61	1.64				2.25	2.30	2.27	1.58	1.61	1.64			
	2.23	2.26	2.28							2.10	2.25	2.35			
IR	3.80	3.90	4.00	3.48	3.92	4.36	3.74	3.90	3.96	3.50	3.75	4.00	3.74	3.85	3.96
	5.77	6.15	6.60	5.35	6.25	7.15	6.06	6.24	6.43	3.50	3.75	4.00	6.06	6.24	6.43
	6.75	7.00	7.15	6.85	7.38	7.85	6.89	6.94	7.01	5.80	6.25	6.70	6.89	6.95	7.01
	7.24	7.40	7.44	8.30	8.70	9.10	7.26	7.35	7.43	6.90	7.10	7.30	7.26	7.35	7.43
	8.30	8.50	8.70	9.38	9.66	9.94	8.44	8.59	8.76	8.00	8.50	9.00	8.44	8.60	8.76
	9.42	9.70	9.80	9.80	10.80	11.80	9.54	9.64	9.72	10.30	10.70	11.30	9.54	9.63	9.72
	10.10	10.30	10.60	11.00	12.00	13.00	10.30	10.41	10.60	11.50	12.00	12.50	10.25	10.43	10.61
	10.80	11.20	11.60	12.40	13.40	14.40	11.10	11.24	11.30	13.20	13.50	13.80	11.08	11.20	11.32
	11.80	12.30	12.80				12.20	12.38	12.50				12.15	12.30	12.45
	13.00	13.30	13.60				13.20	13.28	13.40				13.21	13.30	13.39

[a] Although the visible spectrum is from 0.4 μm to 0.7 μm, and the near-infrared spectrum is from 0.7 μm to 2.5 μm, the classification for visible bands is obtained from NMSC).
[b] The band 7 and band 8 of Fengyun-4 have the same spectral range, but the spatial resolution of band 7 is 2 km and that of band 8 is 4 km. Band 7 is for fire observation while band 8 is for land surface observation [56].

Table A.3
Statistical metrics for performance assessment of solar forecasts.

	Definition	Notes
Mean bias error (MBE)	$\mathrm{MBE} = n^{-1}\sum_{i=1}^{n}\left(\hat{I}_i - I_i\right)$	
Mean absolute error (MAE)	$\mathrm{MAE} = n^{-1}\sum_{i=1}^{n}\lvert\hat{I}_i - I_i\rvert$	
Root mean square error (RMSE)	$\mathrm{RMSE} = \sqrt{n^{-1}\sum_{i=1}^{n}\left(\hat{I}_i - I_i\right)^2}$	
Relative mean absolute error (rMAE)	$\mathrm{rMAE} = \mathrm{MAE}/\left(n^{-1}\sum_{i=1}^{n} I_i\right)$	The denominator can also be averaged irradiance, peak nominal irradiance, or clear sky index
Relative root mean square error (rRMSE)	$\mathrm{rRMSE} = \mathrm{RMSE}/\left(n^{-1}\sum_{i=1}^{n} I_i\right)$	The denominator can also be averaged irradiance, peak nominal irradiance, or clear sky index
Coefficient of determination (R^2)	$R^2 = 1 - \mathrm{Var}\left(\hat{\boldsymbol{I}} - \boldsymbol{I}\right)/\mathrm{Var}\left(\boldsymbol{I}\right)$	
Correlation coefficient (ρ)	$\rho = \mathrm{Cov}\left(\hat{\boldsymbol{I}}, \boldsymbol{I}\right)/\sqrt{\mathrm{Var}(\hat{\boldsymbol{I}})\mathrm{Var}(\boldsymbol{I})}$	
Kolmogorov–Smirnov integral (KSI)	$\mathrm{KSI} = \int D_I\, d\boldsymbol{I}$	D_I is the discrepancy in cumulative distributions between the predictions and the measurements
Error standard deviation (σ)	$\sigma = \sqrt{n^{-1}\sum_{i=1}^{n}\left(\varepsilon_i - \mu\right)^2}$	Where $\varepsilon_i = \hat{I}_i - I_i$, and μ is the mean value of ε_i
Rényi entropy (H_a) [157]	$H_a = (1-a)^{-1}\log_2\left(\sum_{j=1}^{m} p_j^a\right)$	p_j is the probability density for jth section of the error distribution, a is the order of H_a, and higher magnitude of a puts higher weight on more probable events [158]

Table B.4
Summary of activation functions of ANNs.

	Expression	Derivative
ReLU	$\sigma_A(z) = \begin{cases} z, & \text{if } z \geq 0 \\ 0, & \text{if } z < 0 \end{cases}$	$\delta(z) = \begin{cases} 1, & \text{if } z \geq 0 \\ 0, & \text{if } z < 0 \end{cases}$
Leaky ReLU	$\sigma_A(z) = \begin{cases} z, & \text{if } z \geq 0 \\ -\epsilon z, & \text{if } z < 0 \end{cases}$	$\delta(z) = \begin{cases} 1, & \text{if } z \geq 0 \\ \epsilon, & \text{if } z < 0 \end{cases}$
Sigmoid	$\sigma_A(z) = \frac{1}{1+e^{-z}}$	$\delta(z) = \sigma_A(z)(1-\sigma_A(z))$
Hyperbolic tangent	$\sigma_A(z) = \tanh(z)$	$\delta(z) = (1-\lvert\sigma_A(z)^2\rvert)$
Gaussian radial basis	$\sigma_A(z) = \exp(-\lvert\lvert z-m\rvert\rvert^2/\sigma^2)$	$\delta(z) = -2(z-m)\sigma_A(z)/\sigma^2$
Unipolar step	$\sigma_A(z) = H(z) = \begin{cases} 1, & \text{if } z > 0 \\ 0, & \text{if } z < 0 \end{cases}$	$\delta(z) = \begin{cases} 0, & \text{if } z \neq 0 \\ \infty, & \text{if } z = 0 \end{cases}$
Bipolar step	$\sigma_A(z) = \text{sign}(z) = 2H(z) - 1$	$\delta(z) = \begin{cases} 0, & \text{if } z \neq 0 \\ \infty, & \text{if } z = 0 \end{cases}$
Unipolar linear	$\sigma_A(z) = \begin{cases} 0, & \text{if } z < -1 \\ (z+1)/2, & \text{if } \lvert z\rvert < 1 \\ 1, & \text{if } z > 1 \end{cases}$	$\delta(z) = [H(z+1) - H(z-1)]/2$
Bipolar linear	$\sigma_A(z) = \begin{cases} -1, & \text{if } z < -1 \\ z, & \text{if } \lvert z\rvert < 1 \\ 1, & \text{if } z > 1 \end{cases}$	$\delta(z) = H(z+1) - H(z-1)$

9. Repeat the steps 4 to 8 until convergence is achieved, i.e. the preset maximum epochs has been reached or $\epsilon^j - \epsilon^{j-1} \leq e_T$ where j denotes the epoch number.
10. Output the trained ANN with the final weights.

The initialization of the weights and bias can employ random method, the Xavier method [165], or the HE method [166]. Another commonly used parameters initialization methods is fine-tuning with pretrained weights [167]. First, an ANN is pretrained on a large dataset to optimize its parameters. Second, the pretrained parameters will be used as the initialization parameters for new tasks [168,169]. Then, the ANNs will be trained with the training data from new tasks, which could have a relatively small learning rate.

The training or fine tuning process of ANN can be enhanced using training optimizers. Adaptive moment estimation algorithm (ADAM) [170] is one of the most commonly used optimizer, which is employed as the default choice for many tasks. More details of other training optimizers can be found in [3,159,171,172].

B.3. Spatial data analysis based on CNNs

CNN stands as one of the most effective deep learning tools for analyzing spatially correlated data, such as multi-dimensional image/video data in computer vision tasks. As such, CNNs are commonly employed in solar forecasting methods for sky/satellite image analysis, significantly enhancing the forecasting performance of both centralized and spatially distributed systems [12]. The concept of CNNs was first proposed in the 1980s [173]. Training these networks efficiently was a significant challenge of that era. However, this problem was later addressed with the introduction of the gradient backpropagation algorithm proposed by Al-Saffar et al. [174]. Since then, CNNs have demonstrated remarkable performance on datasets such as the handwritten digit dataset MNIST [175]. Several CNN architectures are frequently used, including but not limited to AlexNet [81], ResNet [41], DenseNet [82], and inception net [83]. For more comprehensive information about CNNs, readers are directed to the works of [41,84,85].

Most CNN architectures comprise convolutional, pooling, and classifier layers [36]. Convolutional layers examine input images using receptive fields, which function as feature extractors to derive image features or representations from different portions of an image. These feature extractors, with their distinct weights, allow the extraction of multiple features from the same location on an input image. This process can be mathematically defined as [176]:

$$y_k = \sigma_A(w_k * x), \tag{B.2}$$

where x represents the input image, y_k denotes the image features extracted using the convolutional layer, $\sigma_A(\cdot)$ is a nonlinear activation function, w_k is the trainable parameters of the kth feature extractor of the convolutional layer, and $*$ represents the convolutional operation. Convolutional layers are proposed based on the assumption that different parts of an image may share similar image features or representations, allowing the same feature extractor to analyze different positions on an image [42].

Pooling layers are often grouped with convolutional layers to form modules [176]. These modules are typically stacked to form a deep structure for image feature extraction [177]. A pooling layer is mathematically defined as a nonlinear down-sampling method, which calculates the maximum or average of features from the same local area [81,83]. In practice, a Max Pooling Layer, which forwards the maximum value of the receptive field to the next layer, is more commonly used. Pooling layers offer several benefits, including the reduction of computational efforts, mitigation of overfitting, and enhancement of spatial invariance [178].

The modules of convolutional and pooling layers generate a high-dimensional matrix of image features. These image features are further analyzed by the classifier layers to produce the final outputs. Common classifiers include SVM, MLP, or global pooling layers [179]. MLPs, which include fully connected layers, are commonly employed due to their capability for arbitrary nonlinear mappings and their flexibility in real-life application scenarios [3]. The number of neurons in the final output layer usually equals the number of classification labels in the training dataset. Further details on MLPs can be found in Appendix B.2.

B.4. Time series analysis with RNNs

RNNs can process multiple inputs, enabling the analysis of temporally dynamic behaviors [78]. RNNs have been effectively employed to model and predict solar irradiance or power output time series. The basic RNN unit is mathematically defined as:

$$a_t = \sigma_A(w_{aa}a_{t-1} + w_{ax}x_t + \beta_a), \tag{B.3}$$

$$y_t = \sigma_A(w_{ya}a_t + \beta_y), \tag{B.4}$$

where x_t represents the inputs, y_t is the output, and a_t denotes the hidden state activation at input sequence step t. σ_A is the activation function. The w_{aa}, w_{ax}, w_{ya}, β_a, and β_y are trainable weights or parameters.

LSTM [79] and GRU [80] are variants of RNN, developed to address the issue of vanishing gradient when the input time sequence is long. LSTM units, compared with basic RNN units, have an input gate (i_t), a forget gate (f_t), and an output gate (o_t). New information can be stored in the cell state using i_t, existing information can be discarded from the cell state using f_t, and the final output of the LSTM unit is determined using o_t. The mathematical definitions of these three gates are:

$$i_t = \sigma_A(w_i[h_{t-1}, x_t] + \beta_i), \tag{B.5}$$

$$f_t = \sigma_A(w_f[h_{t-1}, x_t] + \beta_f), \tag{B.6}$$

$$o_t = \sigma_A(w_o[h_{t-1}, x_t] + \beta_o), \tag{B.7}$$

where w and β represent the weights and biases of the three gates, respectively, and h_{t-1} is the hidden state output from the previous time step. For the next time step, the memory state m_t, candidate cell state $\tilde{m}_t$, and the output h_t are derived as:

$$\tilde{m}_t = \tanh(w_c[h_{t-1}, x_t] + \beta_c), \tag{B.8}$$

$$m_t = f_t m_{t-1} + i_t \tilde{m}_t, \tag{B.9}$$

$$h_t = o_t \tanh(m_t), \tag{B.10}$$

where tanh is the hyperbolic tangent function.

For computational efficiency, GRU has fewer parameters than LSTM. Instead of using three gates, GRU employs an update gate z_t, a reset gate r_t, and a memory content $\tilde{h}_t$:

$$z_t = \sigma_A(w_z[h_{t-1}, x_t]), \tag{B.11}$$

$$r_t = \sigma_A(w_r[h_{t-1}, x_t]), \tag{B.12}$$

$$\tilde{h}_t = \tanh(w_h[r_t h_{t-1}, x_t]), \tag{B.13}$$

With z_t, r_t, and $\tilde{h}_t$, the final GRU output is:

$$h_t = (1 - z_t)h_{t-1} + z_t \tilde{h}_t, \tag{B.14}$$

References

[1] Bouckaert S, Pales AF, McGlade C, Remme U, Wanner B, Varro L, D'Ambrosio D, Spencer T. Net zero by 2050: a roadmap for the global energy sector. Technical report, International Energy Agency; 2021.

[2] Yang D, Wang W, Gueymard CA, Hong T, Kleissl J, Huang J, Perez MJ, Perez R, Bright JM, Xia X, et al. 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.

[3] Inman RH, Pedro HTC, Coimbra CFM. Solar forecasting methods for renewable energy integration. Prog Energy Combust Sci 2013;39:535–76.

[4] Widén J, Carpman N, Castellucci V, Lingfors D, Olauson J, Remouit F, Bergkvist M, Grabbe M, Waters R. Variability assessment and forecasting of renewables: A review for solar, wind, wave and tidal resources. Renew Sustain Energy Rev 2015;44:356–75.

[5] Yang D, Wang W, Xia X. A concise overview on solar resource assessment and forecasting. Adv Atmos Sci 2022;39:1239–51. http://dx.doi.org/10.1007/s00376-021-1372-8.

[6] Lave M, Kleissl J. Solar variability of four sites across the state of Colorado. Renew Energy 2010;35:2867–73.

[7] Xin-gang Z, Zhen W. Technology, cost, economic performance of distributed photovoltaic industry in China. Renew Sustain Energy Rev 2019;110:53–64.

[8] Zakeri B, Gissey GC, Dodds PE, Subkhankulova D. Centralized vs. distributed energy storage–benefits for residential users. Energy 2021;236:121443.

[9] Jiang S, Wan C, Chen C, Cao E, Song Y. Distributed photovoltaic generation in the electricity market: Status, mode and strategy. CSEE J Power Energy Syst 2018;4:263–72.

[10] Chu Y, Pedro HTC, Kaur A, Kleissl J, Coimbra CFM. Net load forecasts for solar-integrated operational grid feeders. Sol Energy 2017;158:236–46.

[11] Chu Y, Li M, Pedro HTC, Coimbra CFM. A network of sky imagers for spatial solar irradiance assessment. Renew Energy 2022;187:1009–19.

[12] Chu Y, Li M, Coimbra CFM, Feng D, Wang H. Intra-hour irradiance forecasting techniques for solar power integration: A review. iScience 2021;24:103136.

[13] Khodayar M, Mohammadi S, Khodayar ME, Wang J, Liu G. Convolutional graph autoencoder: A generative deep neural network for probabilistic spatio-temporal solar irradiance forecasting. IEEE Trans Sustain Energy 2019;11:571–83.

[14] Jeong J, Kim H. Multi-site photovoltaic forecasting exploiting space–time convolutional neural network. Energies 2019;12:4490.

[15] Gao H, Liu M. 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.

[16] Yang D, Wang W, Hong T. A historical weather forecast dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF) for energy forecasting. Sol Energy 2022;232:263–74.

[17] Mayer MJ, Yang D. Probabilistic photovoltaic power forecasting using a calibrated ensemble of model chains. Renew Sustain Energy Rev 2022;168:112821.

[18] Agoua XG, Girard R, Kariniotakis G. Probabilistic models for spatio-temporal photovoltaic power forecasting. IEEE Trans Sustain Energy 2018;10:780–9.

[19] Si Z, Yu Y, Yang M, Li P. Hybrid solar forecasting method using satellite visible images and modified convolutional neural networks. IEEE Trans Ind Appl 2020;57:5–16.

[20] Benamrou B, Ouardouz M, Allaouzi I, Ben Ahmed M. A proposed model to forecast hourly global solar irradiation based on satellite derived data, deep learning and machine learning approaches. J Ecol Eng 2020;21.

[21] Qin J, Jiang H, Lu N, Yao L, Zhou C. Enhancing solar PV output forecast by integrating ground and satellite observations with deep learning. Renew Sustain Energy Rev 2022;167:112680.

[22] Wang W, Yang D, Hong T, Kleissl J. An archived dataset from the ECMWF ensemble prediction system for probabilistic solar power forecasting. Sol Energy 2022;248:64–75.

[23] Yang D, Wang W, Bright JM, Voyant C, Notton G, Zhang G, Lyu C. Verifying operational intra-day solar forecasts from ECMWF and NOAA. Sol Energy 2022;236:743–55.

[24] Jimenez PA, Hacker JP, Dudhia J, Haupt SE, Ruiz-Arias JA, Gueymard CA, Thompson G, Eidhammer T, Deng A. WRF-Solar: Description and clear-sky assessment of an augmented NWP model for solar power prediction. Bull Am Meteorol Soc 2016;97:1249–64.

[25] Yang Dazhi, Kleissl Jan. Solar Irradiance and Photovoltaic Power Forecasting. CRC Press; 2024.

[26] Yang D, Perez R. Can we gauge forecasts using satellite-derived solar irradiance? J Renew Sustain Energy 2019;11:023704.

[27] Yagli GM, Yang D, Srinivasan D. Automatic hourly solar forecasting using machine learning models. Renew Sustain Energy Rev 2019;105:487–98.

[28] Zagouras A, Pedro HTC, Coimbra CFM. Clustering the solar resource for grid management in island mode. Sol Energy 2014;110:507–18.

[29] Taravat A, Del Frate F, Cornaro C, Vergari S. Neural networks and support vector machine algorithms for automatic cloud classification of whole-sky ground-based images. Geosci Remote Sens Lett IEEE 2015;12:666–70.

[30] Pedro HTC, Larson DP, Coimbra CFM. A comprehensive dataset for the accelerated development and benchmarking of solar forecasting methods. J Renew Sustain Energy 2019;11:036102.

[31] Larson DP, Coimbra CFM. Direct power output forecasts from remote sensing image processing. J Sol Energy Eng 2018;140.

[32] Kaur A, Nonnenmacher L, Pedro HTC, Coimbra CFM. Benefits of solar forecasting for energy imbalance markets. Renew Energy 2016;86:819–30.

[33] Catalina A, Alaíz CM, Dorronsoro JR. Combining numerical weather predictions and satellite data for PV energy nowcasting. IEEE Trans Sustain Energy 2019;11:1930–7.

[34] Kleissl J. Solar Energy Forecasting and Resource Assessment. Academic Press; 2013.

[35] Chu Y, Urquhart B, Gohari SMI, Pedro HTC, Kleissl J, Coimbra CFM. Short-term reforecasting of power output from a 48 MWe solar PV plant. Sol Energy 2015;112:68–77.

[36] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.

[37] Weimer D, Scholz-Reiter B, Shpitalni M. Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. CIRP Ann 2016;65:417–20.

[38] Ren R, Hung T, Tan KC. A generic deep-learning-based approach for automated surface inspection. IEEE Trans Cybern 2017;48:929–40.

[39] Masci J, Meier U, Ciresan D, Schmidhuber J, Fricout G. Steel defect classification with max-pooling convolutional neural networks. In: Proc. int. joint conf. neural netw.. IJCNN, 2012, p. 1–6.

[40] Park J-K, Kwon B-K, Park J-H, Kang D-J. Machine learning-based imaging system for surface defect inspection. Int J Precis Eng Manuf-Green Technol 2016;3:303–10.

[41] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proc. IEEE comput. vis. pattern recognit.. CVPR, 2016, p. 770–8.

[42] Al-Saffar AAM, Tao H, Talab MA. Review of deep convolution neural network in image classification. In: 2017 international conference on radar, antenna, microwave, electronics, and telecommunications. ICRAMET, IEEE; 2017, p. 26–31.

[43] Yang D, Kleissl J, Gueymard CA, Pedro HTC, Coimbra CFM. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. Sol Energy 2018;168:60–101.

[44] Anagnostos D, Schmidt T, Cavadias S, Soudris D, Poortmans J, Catthoor F. A method for detailed, short-term energy yield forecasting of photovoltaic installations. Renew Energy 2019;130:122–9.

[45] Liu C, Li M, Yu Y, Wu Z, Gong H, Cheng F. A review of multi-temporal and multi-spatial scales photovoltaic forecasting methods. IEEE Access 2022.

[46] Zhu XX, Tuia D, Mou L, Xia G-S, Zhang L, Xu F, Fraundorfer F. Deep learning in remote sensing: A comprehensive review and list of resources. IEEE Geosci Remote Sens Mag 2017;5:8–36.

[47] Carneiro TC, de Carvalho PCM, Alves dos Santos H, Lima MAFB, d. S. Braga AP. Review on photovoltaic power and solar resource forecasting: current status and trends. J Sol Energy Eng 2022;144.

[48] Huang C, Shi H, Yang D, Gao L, Zhang P, Fu D, Xia X, Chen Q, Yuan Y, Liu M, Hu B, Lin K, Li X. Retrieval of sub-kilometer resolution solar irradiance from Fengyun-4A satellite using a region-adapted Heliosat-2 method. Sol Energy 2023;264:112038.

[49] Chu Y, Li M, Coimbra CFM. Sun-tracking imaging system for intra-hour DNI forecasts. Renew Energy 2016;96:792–9.

[50] Chu Y, Coimbra CFM. Short-term probabilistic forecasts for direct normal irradiance. Renew Energy 2017;101:526–36.

[51] Li M, Liao Z, Coimbra CFM. Spectral model for clear sky atmospheric longwave radiation. J Quant Spectrosc Radiat Transfer 2018;209:196–211.

[52] Li M, Liao Z, Coimbra CFM. Spectral solar irradiance on inclined surfaces: A fast Monte Carlo approach. J Renew Sustain Energy 2020;12:053705.

[53] Schmit TJ, Griffith P, Gunshor MM, Daniels JM, Goodman SJ, Lebair WJ. A closer look at the ABI on the GOES-R series. Bull Am Meteorol Soc 2017;98:681–98.

[54] Schmetz J, Pili P, Tjemkes S, Just D, Kerkmann J, Rota S, Ratier A. An introduction to Meteosat second generation (MSG). Bull Am Meteorol Soc 2002;83:977–92.

[55] Bessho K, Date K, Hayashi M, Ikeda A, Imai T, Inoue H, Kumagai Y, Miyakawa T, Murata H, Ohno T, et al. An introduction to Himawari-8/9—Japan's new-generation geostationary meteorological satellites. J Meteorol Soc Jpn Ser II 2016;94:151–83.

[56] Yang J, Zhang Z, Wei C, Lu F, Guo Q. Introducing the new generation of Chinese geostationary weather satellites, Fengyun-4. Bull Am Meteorol Soc 2017;98:1637–58.

[57] Kim D, Gu M, Oh T-H, Kim E-K, Yang H-J. Introduction of the advanced meteorological imager of Geo-Kompsat-2a: In-orbit tests and performance validation. Remote Sens 2021;13:1303.

[58] Jo J-M. Effectiveness of normalization pre-processing of big data to the machine learning performance. J Korea Inst Electron Commun Sci 2019;14:547–52.

[59] Yang D, Alessandrini S, Antonanzas J, Antonanzas-Torres F, Badescu V, Beyer HG, Blaga R, Boland J, Bright JM, Coimbra CFM, et al. Verification of deterministic solar forecasts. Sol Energy 2020;210:20–37.

[60] Ineichen P, Perez R. Derivation of cloud index from geostationary satellites and application to the production of solar irradiance and daylight illuminance data. Theor Appl Climatol 1999;64:119–30.

[61] Matsunobu LM, Pedro HTC, Coimbra CFM. Cloud detection using convolutional neural networks on remote sensing images. Sol Energy 2021;230:1020–32.

[62] Paletta Q, Arbod G, Lasenby J. Omnivision forecasting: Combining satellite and sky images for improved deterministic and probabilistic intra-hour solar energy predictions. Appl Energy 2023;336:120818.

[63] Rigollier C, Lefèvre M, Wald L. The method Heliosat-2 for deriving shortwave solar radiation from satellite images. Sol Energy 2004;77:159–69.

[64] Perez R, Ineichen P, Moore K, Kmiecik M, Chain C, George R, Vignola F. A new operational model for satellite-derived irradiances: Description and validation. Sol Energy 2002;73:307–17.

[65] Chu Y, Pedro HTC, Nonnenmacher L, Inman RH, Liao Z, Coimbra CFM. A smart image-based cloud detection system for intra-hour solar irradiance forecasts. J Atmos Ocean Technol 2014;31:1995–2007.

[66] Berthomier L, Pradel B, Perez L. Cloud cover nowcasting with deep learning. In: 2020 tenth international conference on image processing theory, tools and applications. IPTA, IEEE; 2020, p. 1–6.

[67] Wang Z, Zhao J, Zhang R, Li Z, Lin Q, Wang X. UATNet: U-shape attention-based transformer net for meteorological satellite cloud recognition. Remote Sens 2021;14:104.

[68] Roy R, Ahan M, Soni V, Chittora A. Towards automatic transformer-based cloud classification and segmentation. In: NeurIPS 2021 workshop on tackling climate change with machine learning. Vol. 2021, 2021, p. 60.

[69] Li L, Li X, Jiang L, Su X, Chen F. A review on deep learning techniques for cloud detection methodologies and challenges. Signal Image Video Process 2021;15:1527–35.

[70] Marquez R, Coimbra CFM. Proposed metric for evaluation of solar forecasting models. ASME J Sol Energy Eng 2013;135:0110161–9.

[71] Voyant C, Notton G, Kalogirou S, Nivet M-L, Paoli C, Motte F, Fouilloy A. Machine learning methods for solar radiation forecasting: A review. Renew Energy 2017;105:569–82.

[72] Cheng L, Zang H, Wei Z, Ding T, Sun G. Solar power prediction based on satellite measurements-a graphical learning method for tracking cloud motion. IEEE Trans Power Syst 2021.

[73] Marquez R, Coimbra CFM. Intra-hour DNI forecasting methodology based on cloud tracking image analysis. Sol Energy 2013;91:327–36.

[74] Chu Y, Pedro HTC, Coimbra CFM. Hybrid intra-hour DNI forecasts with sky image processing enhanced by stochastic learning. Sol Energy 2013;98:592–603.

[75] Yang D. Making reference solar forecasts with climatology, persistence, and their optimal convex combination. Sol Energy 2019;193:981–5.

[76] Yang D. Standard of reference in operational day-ahead deterministic solar forecasting. J Renew Sustain Energy 2019;11:053702.

[77] Liu B, Yang D, Mayer MJ, Coimbra CF, Kleissl J, Kay M, Wang W, Bright JM, Xia X, Lv X, Srinivasan D, Wu Y, Beyer HG, Yagli GM, Shen Y. Predictability and forecast skill of solar irradiance over the contiguous United States. Renew Sustain Energy Rev 2023;182:113359.

[78] Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature 1986;323:533–6.

[79] Hochreiter S, Schmidhuber J. Long short-term memory. Neural Comput 1997;9:1735–80.

[80] Medsker L, Jain LC. Recurrent neural networks: design and applications. CRC Press; 1999.

[81] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. In: Proc. adv. neural inform. process. syst.(neurIPS). 2012, p. 1097–105.

[82] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proc. IEEE comput. vis. pattern recognit.. CVPR, 2017, p. 4700–8.

[83] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: Proc. IEEE comput. vis. pattern recognit.. CVPR, 2015, p. 1–9.

[84] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. In: Proc. europ. conf. comp. vis.. ECCV, 2014, p. 818–33.

[85] Chen CW. Internet of video things: Next-generation IoT with visual sensors. IEEE Internet Things J 2020;7:6676–85. http://dx.doi.org/10.1109/JIOT.2020.3005727.

[86] Ahmed R, Sreeram V, Mishra Y, Arif M. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. Renew Sustain Energy Rev 2020;124:109792.

[87] Lago J, De Brabandere K, De Ridder F, De Schutter B. Short-term forecasting of solar irradiance without local telemetry: A generalized model using satellite data. Sol Energy 2018;173:566–77.

[88] Kim M, Song H, Kim Y. Direct short-term forecast of photovoltaic power through a comparative study between COMS and himawari-8 meteorological satellite images in a deep neural network. Remote Sens 2020;12:2357.

[89] Lima FJ, Martins FR, Pereira EB, Lorenz E, Heinemann D. Forecast for surface solar irradiance at the Brazilian northeastern region using NWP model and artificial neural networks. Renew Energy 2016;87:807–18.

[90] Pierro M, De Felice M, Maggioni E, Moser D, Perotto A, Spada F, Cornaro C. Data-driven upscaling methods for regional photovoltaic power estimation and forecast using satellite and numerical weather prediction data. Sol Energy 2017;158:1026–38.

[91] Zhang R, Ma H, Hua W, Saha TK, Zhou X. Data-driven photovoltaic generation forecasting based on a Bayesian network with spatial–temporal correlation analysis. IEEE Trans Ind Inf 2019;16:1635–44.

[92] Li P, Bessafi M, Morel B, Chabriat J-P, Delsaut M, Li Q. Daily surface solar radiation prediction mapping using artificial neural network: the case study of Reunion Island. J Solar Energy Eng 2020;142.

[93] Nikitidou E, Zagouras A, Salamalikis V, Kazantzidis A. Short-term cloudiness forecasting for solar energy purposes in Greece, based on satellite-derived information. Meteorol Atmos Phys 2019;131:175–82.

[94] Yeom J-M, Park S, Chae T, Kim J-Y, Lee CS. Spatial assessment of solar radiation by machine learning and deep neural network models using data provided by the COMS MI geostationary satellite: A case study in south Korea. Sensors 2019;19:2082.

[95] Srivastava S, Lessmann S. A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data. Sol Energy 2018;162:232–47.

[96] Kim B, Suh D, Otto M-O, Huh J-S. A novel hybrid spatio-temporal forecasting of multisite solar photovoltaic generation. Remote Sens 2021;13:2605.

[97] Pérez E, Pérez J, Segarra-Tamarit J, Beltran H. A deep learning model for intra-day forecasting of solar irradiance using satellite-based estimations in the vicinity of a PV power plant. Sol Energy 2021;218:652–60.

[98] Jiang H, Lu N, Qin J, Tang W, Yao L. A deep learning algorithm to estimate hourly global solar radiation from geostationary satellite data. Renew Sustain Energy Rev 2019;114:109327.

[99] Jang J-C, Sohn E-H, Park K-H. Estimating hourly surface solar irradiance from GK2A/AMI data using machine learning approach around Korea. Remote Sens 2022;14:1840.

[100] Liu Y, Qin H, Zhang Z, Pei S, Wang C, Yu X, Jiang Z, Zhou J. Ensemble spatiotemporal forecasting of solar irradiation using variational Bayesian convolutional gate recurrent unit network. Appl Energy 2019;253:113596.
[101] Yeom J-M, Deo RC, Adamowski JF, Park S, Lee C-S. Spatial mapping of short-term solar radiation prediction incorporating geostationary satellite images coupled with deep convolutional LSTM networks for South Korea. Environ Res Lett 2020;15:094025.
[102] Hong Y-Y, Martinez JJF, Fajardo AC. Day-ahead solar irradiation forecasting utilizing gramian angular field and convolutional long short-term memory. IEEE Access 2020;8:18741–53.
[103] Rocha PA, Santos VO. Global horizontal and direct normal solar irradiance modeling by the machine learning methods XGBoost and deep neural networks with CNN-LSTM layers: a case study using the GOES-16 satellite imagery. Int J Energy Environ Eng 2022;1–16.
[104] Oh M, Kim CK, Kim B, Yun C, Kang Y-H, Kim H-G. Spatiotemporal optimization for short-term solar forecasting based on satellite imagery. Energies 2021;14:2216.
[105] Kumari P, Toshniwal D. Deep learning models for solar irradiance forecasting: A comprehensive review. J Clean Prod 2021;318:128566.
[106] Yao T, Wang J, Wu H, Zhang P, Li S, Xu K, Liu X, Chi X. Intra-hour photovoltaic generation forecasting based on multi-source data and deep learning methods. IEEE Trans Sustain Energy 2021;13:607–18.
[107] Deo RC, Şhin M, Adamowski JF, Mi J. Universally deployable extreme learning machines integrated with remotely sensed MODIS satellite predictors over Australia to forecast global solar radiation: A new approach. Renew Sustain Energy Rev 2019;104:235–61.
[108] Narvaez G, Giraldo LF, Bressan M, Pantoja A. Machine learning for site-adaptation and solar radiation forecasting. Renew Energy 2021;167:333–42.
[109] Carney JG, Cunningham P, Bhagwan U. Confidence and prediction intervals for neural network ensembles. In: Neural networks, 1999. iJCNN'99. International joint conference on. Vol. 2, IEEE; 1999, p. 1215–8.
[110] Khosravi A, Nahavandi S, Creighton D. Prediction intervals for short-term wind farm power generation forecasts. IEEE Trans Sustain Energy 2013;4:602–10.
[111] Pinson P, Nielsen HA, Møller JK, Madsen H, Kariniotakis GN. Non-parametric probabilistic forecasts of wind power: required properties and evaluation. Wind Energy 2007;10:497–516.
[112] Bracale A, Caramia P, Carpinelli G, Di Fazio AR, Ferruzzi G. A Bayesian method for short-term probabilistic forecasting of photovoltaic generation in smart grid operation and control. Energies 2013;6:733–47.
[113] Yang D. Ensemble model output statistics as a probabilistic site-adaptation tool for satellite-derived and reanalysis solar irradiance. J Renew Sustain Energy 2020;12:016102.
[114] Wang P, van Westrhenen R, Meirink JF, van der Veen S, Knap W. Surface solar radiation forecasts by advecting cloud physical properties derived from Meteosat Second Generation observations. Sol Energy 2019;177:47–58.
[115] Doorga JRS, Dhurmea KR, Rughooputh S, Boojhawon R. Forecasting mesoscale distribution of surface solar irradiation using a proposed hybrid approach combining satellite remote sensing and time series models. Renew Sustain Energy Rev 2019;104:69–85.
[116] Khoo YS, Nobre A, Malhotra R, Yang D, Ruther R, Reindl T, Aberle AG. Optimal orientation and tilt angle for maximizing in-plane solar irradiation for pv applications in singapore. IEEE J Photovolt 2014;4:647–53.
[117] Yoshida S, Ueno S, Kataoka N, Takakura H, Minemoto T. Estimation of global tilted irradiance and output energy using meteorological data and performance of photovoltaic modules. Sol Energy 2013;93:90–9.
[118] Yagli GM, Yang D, Gandhi O, Srinivasan D. Can we justify producing univariate machine-learning forecasts with satellite-derived solar irradiance? Appl Energy 2020;259:114122.
[119] Niu Z, Zhong G, Yu H. A review on the attention mechanism of deep learning. Neurocomputing 2021;452:48–62.
[120] Zhao B, Feng J, Wu X, Yan S. A survey on deep learning-based fine-grained object classification and semantic segmentation. Int J Autom Comput 2017;14:119–35.
[121] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I. Attention is all you need. In: Advances in neural information processing systems. 2017, p. 5998–6008.
[122] Wang F, Jiang M, Qian C, Yang S, Li C, Zhang H, Wang X, Tang X. Residual attention network for image classification. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017, p. 3156–64.
[123] Mnih V, Heess N, Graves A, et al. Recurrent models of visual attention. In: Advances in neural information processing systems. 2014, p. 2204–12.
[124] Itti L, Koch C. Computational modelling of visual attention. Nat Rev Neurosci 2001;2:194–203.
[125] Chen L-Q, Xie X, Fan X, Ma W-Y, Zhang H-J, Zhou H-Q. A visual attention model for adapting images on small displays. Multimedia Syst 2003;9:353–64.
[126] Xu K, Ba J, Kiros R, Cho K, Courville A, Salakhudinov R, Zemel R, Bengio Y. Show, attend and tell: Neural image caption generation with visual attention. In: International conference on machine learning. 2015, p. 2048–57.
[127] Guo M-H, Xu T-X, Liu J-J, Liu Z-N, Jiang P-T, Mu T-J, Zhang S-H, Martin RR, Cheng M-M, Hu S-M. Attention mechanisms in computer vision: A survey. Comput Vis Media 2022;1–38.
[128] Hu J, Shen L, Sun G. Squeeze-and-excitation networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, p. 7132–41.
[129] Chen M, Radford A, Child R, Wu J, Jun H, Luan D, Sutskever I. Generative pretraining from pixels. In: International conference on machine learning. PMLR; 2020, p. 1691–703.
[130] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, et al. An image is worth 16x16 words: Transformers for image recognition at scale. 2020, arXiv preprint arXiv:2010.11929.
[131] Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. 2021, p. 10012–22.
[132] Lu Z, Wang Z, Li X, Zhang J. A method of ground-based cloud motion predict: CCLSTM+ SR-net. Remote Sens 2021;13:3876.
[133] Xu Z, Du J, Wang J, Jiang C, Ren Y. Satellite image prediction relying on GAN and LSTM neural networks. In: ICC 2019-2019 IEEE international conference on communications. ICC, IEEE; 2019, p. 1–6.
[134] Rüttgers M, Lee S, Jeon S, You D. Prediction of a typhoon track using a generative adversarial network and satellite images. Sci Rep 2019;9:1–15.
[135] Lu T, Viljanen M. Prediction of indoor temperature and relative humidity using neural network models: model comparison. Neural Comput Appl 2009;18:345–57.
[136] Dybowski R, Roberts S. Confidence intervals and prediction intervals for feed-forward neural networks. Clin Appl Artif Neural Netw 2001;298–326.
[137] Chu Y, Li M, Pedro HTC, Coimbra CFM. Real-time prediction intervals for intra-hour DNI forecasts. Renew Energy 2015;83:234–44.
[138] Bremnes JB. Probabilistic wind power forecasts using local quantile regression. Wind Energy 2004;7:47–54.
[139] Bremnes JB. A comparison of a few statistical models for making quantile wind power forecasts. Wind Energy 2006;9:3–11.
[140] Nam S, Hur J. Probabilistic forecasting model of solar power outputs based on the naive Bayes classifier and kriging models. Energies 2018;11:2982.
[141] Doubleday K, Jascourt S, Kleiber W, Hodge B-M. Probabilistic solar power forecasting using Bayesian model averaging. IEEE Trans Sustain Energy 2020;12:325–37.
[142] Pedro HTC, Coimbra CFM, David M, Lauret P. Assessment of machine learning techniques for deterministic and probabilistic intra-hour solar forecasts. Renew Energy 2018;123:191–203.
[143] Lauret P, David M, Pinson P. Verification of solar irradiance probabilistic forecasts. Sol Energy 2019;194:254–71.
[144] Doubleday K, Hernandez VVS, Hodge B-M. Benchmark probabilistic solar forecasts: Characteristics and recommendations. Sol Energy 2020;206:52–67.
[145] Ravuri S, Lenc K, Willson M, Kangin D, Lam R, Mirowski P, Fitzsimons M, Athanassiadou M, Kashem S, Madge S, et al. Skilful precipitation nowcasting using deep generative models of radar. Nature 2021;597:672–7.
[146] Burke M, Driscoll A, Lobell DB, Ermon S. Using satellite imagery to understand and promote sustainable development. Science 2021;371:eabe8628.
[147] Bi K, Xie L, Zhang H, Chen X, Gu X, Tian Q. Accurate medium-range global weather forecasting with 3D neural networks. Nature 2023;619:533–8.
[148] Kurth T, Subramanian S, Harrington P, Pathak J, Mardani M, Hall D, Miele A, Kashinath K, Anandkumar A. FourCastNet: Accelerating global high-resolution weather forecasting using adaptive fourier neural operators. In: Proceedings of the platform for advanced scientific computing conference. 2023, p. 1–11.
[149] Zhang Y, Long M, Chen K, Xing L, Jin R, Jordan MI, Wang J. Skilful nowcasting of extreme precipitation with NowcastNet. Nature 2023;619:526–32.
[150] NOAA. The global forecast system (GFS). 2023, Available at: https://www.emc.ncep.noaa.gov/emc/pages/numerical_forecast_systems/gfs/documentation.php, Accessed on 2023-11-08.
[151] Incecik S, Sakarya S, Tilev S, Kahraman A, Aksoy B, Caliskan E, Topcu S, Kahya C, Odman MT. Evaluation of WRF parameterizations for global horizontal irradiation forecasts: A study for Turkey. Atmósfera 2019;32:143–58.
[152] Gueymard C, Jimenez P. Validation of real-time solar irradiance simulations over Kuwait using WRF-solar. In: 12th international conference on solar energy for buildings and industry, euroSun2018. Rapperswil, Switzerland; 2018, p. 0130.
[153] Shen X, Wang J, Li Z, Chen D, Gong J. Research and operational development of numerical weather prediction in China. J Meteorol Res 2020;34:675–98.
[154] IEA. Solar heating and cooling programme, task 36, subtask a – standard qualification for solar resource products. Technical report, IEA; 2012.
[155] Zhang J, A F, Hodge B, Lu S, Hamann H, Banunarayanan V, Brockway A. A suite of metrics for assessing the performance of solar power forecasting. Sol Energy 2015;111:157–75.
[156] Lew D, Brinkman G, Ibanez E, Hodge B, King J. Western wind and solar integration study phase 2. Contract 2013;303:275–3000.
[157] Hodge B-M, Orwig K, Milligan MR. Examining information entropy approaches as wind power forecasting performance metrics. National Renewable Energy Laboratory; 2012.
[158] Bessa RJ, Miranda V, Botterud A, Wang J. 'Good' or 'bad' wind power forecasts: A relative concept. Wind Energy 2011;14:625–36.

[159] Mellit A, Kalogirou SA. Artificial intelligence techniques for photovoltaic applications: A review. Prog Energy Combust Sci 2008;34:574–632.
[160] Xu B, Wang N, Chen T, Li M. Empirical evaluation of rectified activations in convolutional network. 2015, arXiv preprint arXiv:1505.00853.
[161] Nair V, Hinton GE. Rectified linear units improve restricted Boltzmann machines. In: Proc. int. conf. mach. learn.. ICML, 2010, p. 807–14.
[162] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. In: Proc. 14th int. conf. artificial intell. stat. 2011, p. 315–23.
[163] Werbos PJ. Beyond regression: new tools for prediction and analysis in the behavioral sciences (Ph.D. thesis), Harvard University; 1974.
[164] Parker DB. Learning logic. Technical report TR-47, Center for Computational Research in Economics and Management Science, MIT; 1985.
[165] Glorot X, Bengio Y. Understanding the difficulty of training deep feedforward neural networks. In: Proc. int. conf. artificial intell. stat.. ICAIS, 2010, p. 249–56.
[166] He K, Zhang X, Ren S, Sun J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: Proc. IEEE int. conf. comput. vis. 2015, p. 1026–34.
[167] Pan SJ, Yang Q. A survey on transfer learning. IEEE Trans Knowl Data Eng 2009;22:1345–59.
[168] Yosinski J, Clune J, Bengio Y, Lipson H. How transferable are features in deep neural networks? In: Proc. adv. neural inform. process. syst. 2014, p. 3320–8.
[169] Tan C, Sun F, Kong T, Zhang W, Yang C, Liu C. 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 (2018) 4-7, proceedings, part III 27. Springer; 2018, p. 270–9.
[170] Kingma DP, Ba J. Adam: A method for stochastic optimization. 2014, arXiv preprint arXiv:1412.6980.
[171] Haykin S. Neural networks: a comprehensive foundation. 3rd ed.. Upper Saddle River, NJ: Pearson Prentice Hall; 2008.
[172] Lakhmi CJ, Martin NM. Fusion of neural networks, fuzzy systems and genetic algorithms: industrial applications. CRC Press; 1998.
[173] Fukushima K, Miyake S. Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. Pattern Recognit 1982;15:455–69.
[174] Al-Saffar AAM, Tao H, Talab MA. Review of deep convolution neural network in image classification. In: Proc. int. conf. radar, antenna, microw. electron. telecommun.. ICRAMET, 2017, p. 26–31.
[175] Deng L. The mnist database of handwritten digit images for machine learning research [best of the web]. IEEE Signal Process Mag 2012;29:141–2.
[176] LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proc IEEE 1998;86:2278–324.
[177] Rawat W, Wang Z. Deep convolutional neural networks for image classification: A comprehensive review. Neural Comput 2017;29:2352–449.
[178] Ranzato M, Huang FJ, Boureau Y-L, LeCun Y. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In: Proc. IEEE comput. vis. pattern recognit.. CVPR, 2007, p. 1–8.
[179] Lin M, Chen Q, Yan S. Network in network. 2013, arXiv preprint arXiv:1312.4400.