



Distributed solar generation forecasting using attention-based deep neural networks for cloud movement prediction

Maneesha Perera ^a,^{*}, Julian De Hoog ^a, Kasun Bandara ^c, Hansani Weeratunge ^b,
Saman Halgamuge ^a

^a Department of Mechanical Engineering, School of Electrical, Mechanical and Infrastructure Engineering, The University of Melbourne, Melbourne, Australia

^b Department of Mechanical Engineering, Sri Lanka Institute of Information Technology, Sri Lanka

^c School of Computing and Information Systems, Melbourne Centre for Data Science, The University of Melbourne, Melbourne, Australia

ARTICLE INFO

Keywords:

Solar photovoltaic power forecasting
Deep neural networks
Infrared satellite imagery
Satellite image forecasting
Cloud forecasting

ABSTRACT

Accurate forecasts of distributed solar generation are necessary to maintain grid stability amid the increased uptake of distributed solar photovoltaic (PV) systems. However, the high variability of solar generation over short time intervals (seconds to minutes) caused by cloud movement makes this forecasting task difficult. To address this, using cloud images, which capture the second-to-second changes in cloud cover affecting solar generation, has shown promise. Recently, deep neural networks with *attention* that focus on important regions of an image have been applied with success in many computer vision applications. However, whether such methods provide meaningful benefits for cloud movement forecasting, and how such improvements propagate through to downstream solar generation forecasting accuracy, remains under-explored. In this study, we conduct a large-scale empirical investigation of the impact of attention-based cloud forecasting on solar generation forecasting, addressing a gap that has been overlooked in the literature. To this end, we develop a pipeline that incorporates an attention-enhanced convolutional long short-term memory network and an existing self-attention-based video prediction method to forecast cloud movement using satellite imagery. The effectiveness of the resulting cloud forecasts is evaluated through their downstream impact on solar forecasting across 50 PV sites in Australia. We further provide insights into the cloud conditions under which attention-based cloud forecasting methods yield the most significant improvements in downstream solar forecasting accuracy. We find that for clouds at high altitudes, the cloud predictions obtained using attention-based methods result in solar forecast skill score improvements of 5.86% or more compared to non-attention-based methods.

1. Introduction

The number of small-scale distributed solar photovoltaic (PV) installations on the rooftops of residential and commercial buildings (i.e., behind-the-meter PV systems) has increased over the past decade and is expected to grow exponentially in the coming years [1,2]. These PV systems contribute to decarbonizing our electricity sector by generating clean and renewable energy. However, integrating large amounts of behind-the-meter PV systems into existing electricity networks poses challenges due to the variable and non-dispatchable nature of solar generation [2,3]. Fortunately, accurate predictions of future solar generation (i.e., solar PV power forecasts) can reduce this uncertainty and enable energy stakeholders to better plan and manage electricity grid

operation and energy market participation [4,5]. In particular, short-term solar power forecasts, ranging from seconds, minutes and up to 1–6 h ahead, are critical for demand response, scheduling spinning reserves, and maintaining grid stability and power quality [6–8].

Cloud changes in the sky are the primary driver of solar power fluctuations over short time intervals [9,10]. Short-term solar forecasting methods typically forecast changes in clouds in two forms: (i) by implicitly forecasting cloud changes using data such as cloud cover measurements provided by nearby weather observation stations, or (ii) by explicitly forecasting cloud movement using images of the sky. The latter enables a more accurate understanding of the movement and formation of clouds, since images of the sky can provide detailed information about clouds at high temporal resolutions (e.g., seconds, minutes)

^{*} Corresponding author.

E-mail addresses: maneesha.perera1@unimelb.edu.au (M. Perera), julian.dehoog@unimelb.edu.au (J. De Hoog), kasun.bandara@unimelb.edu.au (K. Bandara), hansani.w@sliit.lk (H. Weeratunge), saman@unimelb.edu.au (S. Halgamuge).

<https://doi.org/10.1016/j.energy.2026.141985>

Received 10 April 2026; Received in revised form 21 June 2026; Accepted 15 July 2026

Available online 21 July 2026

0360-5442/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

as compared to coarse-grained measures such as cloud cover which is usually available at lower temporal resolutions (e.g., hourly) [9,11].

Imagery of clouds, such as ground-based sky images or satellite images (visible and infrared), are useful data sources to learn changes in clouds that affect solar generation [10,12]. Many regions of the earth are now regularly imaged by geostationary satellites (for example, Himawari-8 for Japan and Australia), and this kind of data is becoming a promising source to monitor changes in clouds for behind-the-meter solar PV systems [12]. Forecasting solar generation using such satellite images commonly involves a two step learning process [9]. In the first step, future cloud movement is forecast using past imagery; in the second step, the predicted cloud information is combined with other relevant attributes (e.g., past power generation values, temperature) to forecast solar generation [11,12]. Much focus in this learning process is given to accurately forecasting cloud movement, since this is the primary driver of short term solar variability [13]. This has also been the main motivation for this work.

Various methods, such as optical flow [14] or particle image velocimetry [13], have been used to forecast cloud movement in the past. However, such techniques generally assume that clouds move in a linear manner, and that cloud shape and thickness remain unchanged over time. Deep learning methods, however, can overcome these limitations since they are well suited to learning highly non-linear spatiotemporal relationships in sequential image data. Therefore, more recently deep learning methods have been adapted in the cloud forecasting literature.

The commonly adapted deep learning methods for cloud forecasting include variants of convolutional neural networks (CNN) or combinations of CNN and long short-term memory networks due to their abilities to capture spatial and temporal relationships in the data. However, the field of deep learning has seen rapid advances, introducing architectures and components that have demonstrated strong performance across a broad range of applications [15–17]. One such mechanism that has proven successful in many applications is known as “attention model” or “attention mechanism” [18]. The main concept behind attention is giving more importance to useful parts (or areas) of an input to assist the task at hand and focusing less on other areas. In the context of solar forecasting, application of attention mechanisms has been investigated in several studies [19–21]. However, these studies primarily apply attention to identify important regions of power generation time-series data rather than to cloud images. Few studies have explored applying attention mechanisms in cloud forecasting to identify regions of interests in cloud images. One notable example is Cheng et al. [13], where a CNN-based attention module is used to refine cloud feature maps by weighting satellite image features alongside past power generation values and clear sky predictions, producing a single channel attention-refined feature map for solar power forecasting. While these studies provide valuable insights into how attention mechanisms provide improvements in solar forecasting, several important questions remain underexplored. First, the impact of improvements in attention-based cloud forecasting on the accuracy of downstream solar forecasting has not been systematically investigated. Furthermore, given the variety of solar forecasting methods that exist in the literature, it remains unknown whether these improvements are consistent across different forecasting models. Second, there has been limited research on the specific sky conditions in which attention mechanisms offer the most significant benefits.

In this paper, we aim to address the above questions. To do so, we take ConvLSTM — a network that laid the foundation for many subsequent architectures proposed for spatiotemporal applications [17, 22] as our baseline for cloud movement forecasting. Building on this, we develop an attention-based deep neural network, CBAMConvLSTM, which integrates a convolutional block attention module (CBAM) within ConvLSTM to capture the spatiotemporal evolution of cloud patterns in satellite imagery. We further evaluate SAConvLSTM [23], an existing attention-based model originally proposed for video prediction that replaces standard ConvLSTM operations with self-attention, for the

task of cloud movement forecasting. The primary motivation of this work is to systematically investigate how improvements in cloud movement forecasting from attention-based deep neural networks translate into improved downstream solar generation forecasting performance. To isolate and attribute forecast improvements specifically to cloud prediction quality rather than architectural differences in the solar forecasting stage, we use well-established solar forecasting networks as downstream models, a multilayer perceptron (MLP), a 1D convolutional neural network (CNN), and a long short-term memory network (LSTM). The cloud forecasts generated by each cloud forecasting model is used as input to each of these downstream networks. We conduct a comprehensive evaluation using a large scale dataset consist of 50 small-scale rooftop solar PV sites in Australia. To further understand in which situations attention provides the greatest benefit, we analyze how cloud conditions corresponding to low and high cloud altitudes (inferred from infrared satellite image pixel values) influence solar forecasting performance.

In summary, this study provides the following key contributions:

- We empirically investigate how attention-based improvements in cloud movement forecasting propagate through to downstream solar generation forecasting. To this end,
 - We propose CBAMConvLSTM an attention-based extension for ConvLSTM (an existing method widely adapted for spatio-temporal applications) to forecast cloud movement.
 - We adapt and evaluate an existing self-attention-based network (SAConvLSTM) originally proposed for video prediction to forecast cloud movement.
- We systematically evaluate how cloud movement forecasts obtained from attention- and non-attention-based networks influence the short-term solar power forecasting performance using a large scale dataset consisting of 50 small-scale rooftop solar PV sites.
- We provide novel insights under which cloud altitude conditions (high or low) attention-based cloud forecasting methods provide the greatest benefit to downstream solar generation forecasting accuracy.

The remainder of the paper is structured as follows. Section 2 discusses the related work; Section 3 presents solar power and satellite image data we use in this work; Section 4 outlines the overall methodology and discusses solar forecasting and cloud forecasting neural networks; Sections 5 and 6 present the evaluation and discuss the results and findings; Section 7 provides the conclusions of this work.

2. Background and related work

“Solar forecasting” in the literature may refer to forecasting either solar irradiance at a given location or solar PV power generated by a PV system [2,5,9,24–26]. A wider range of forecasting methods such as time series forecasting [27], traditional machine learning (e.g., support vector regression or tree-based methods) [28] and neural networks [29] have been applied to both irradiance and power forecasting. However, forecasting PV power typically requires either historical power generation data from the system, or system specific information (e.g., panel tilt and orientation) to convert irradiance forecasts to power forecasts. The input data to any solar forecasting approach can be broadly categorized as data-driven and/or image-based. Data-driven inputs include historical power generation data from the PV site or historical irradiance data, weather observation data, and numerical weather predictions [9,25,30, 31]. Typical image-based inputs include cloud images such as ground-based sky images (GSI) or satellite imagery, which are more commonly used for short-term solar forecasting [9].

GSI and satellite images differ based on the techniques used to capture the images and the available temporal and spatial coverage.

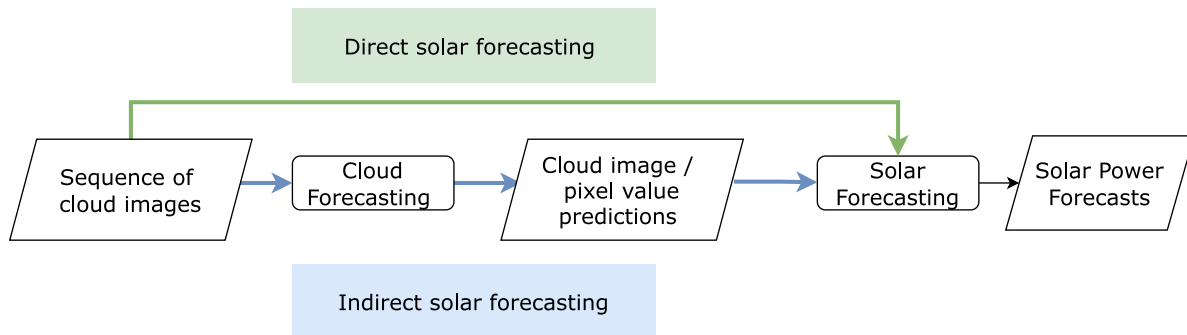


Fig. 1. Steps involved for two different approaches (direct vs. indirect) of forecasting solar generation using satellite or ground-based sky images. Indirect solar forecasting involves a two step process: forecasting cloud movement and using cloud forecasts as an input to the solar forecasting step. Direct solar forecasting approaches bypass the cloud forecasting step and builds a direct mapping between the cloud images and solar generation.

Satellite images are captured using satellite-sensors while GSI are acquired through digital cameras installed on the ground closer to the PV site [32,33]. GSI typically offer higher temporal resolutions, ranging from seconds to minutes compared to satellite images that are available at resolutions ranging from minutes to hours. Despite the higher temporal and spatial resolution of GSI, their utility may be limited due to expensive cameras that need to be installed close to a PV site [11]. Furthermore, installing cameras close to hundreds of behind-the-meter PV systems in a neighborhood or suburb is not practical. As discussed above, geostationary satellites (such as Himawari-8) now provide images for extensive geographical areas and therefore are becoming a promising source of data to monitor changes in clouds [12]. Therefore, in this work, we use satellite images to explore cloud movement forecasting.

The impact of clouds on solar generation is not straightforward as there are several attributes in clouds such as the thickness, altitude (i.e., height from the ground) that can affect the generation from a PV site [34]. For example, Barbieri et al. [7] discussed seven types of clouds that can affect solar generation. Infrared and visible satellite images are useful data sources for inferring cloud attributes such as cloud thickness and altitude levels [10,12]. Visible images are captured through sensors that measure the reflected solar radiation from clouds, while infrared images are captured by measuring the heat radiating from the cloud tops. As a result, infrared images can depict cloud cover during both day and night as they are not dependent on the solar radiation. Infrared images are particularly useful to distinguish high vs. low altitude clouds while thin vs. thick clouds are distinguishable using visible images. While many studies have focused on visible satellite images, limited work has focused on infrared images [10,13] that are useful in inferring some information on cloud altitudes. Despite using infrared imagery, studies in the literature have not investigated the impact of different cloud values inferred from infrared imagery that may correspond to high or low clouds. Thus, through this work, we also intend to shed some light on this direction.

To effectively use cloud attribute information for solar forecasting, it is necessary to accurately model and forecast the spatiotemporal evolution of clouds. Various methods have been used to forecast cloud movement that range from traditional image processing methods (e.g., optical flow method [14], particle image velocimetry [13]) to more recently deep learning methods [12]. In the following section, we specifically discuss solar forecasting studies using cloud imagery and deep learning methods that are relevant to the work presented in this paper.

2.1. Related work

As shown in Fig. 1, solar generation forecasting using cloud imagery can be categorized in two main forms (i) indirect forecasting and (ii) direct forecasting.

The indirect approach involves two steps where the first step involves forecasting the cloud movement. In some studies, the output of this step is a set of pixel values in the image corresponding to the clouds affecting the PV site, while in others the output is a sequence of cloud images where pixels corresponding to the clouds affecting the PV site are extracted afterwards [12]. In the second step, the cloud information predicted in the first step is used as an input to forecast solar generation. In contrast, direct forecasting bypasses the cloud movement forecasting and builds a direct mapping between past cloud images and solar predictions [35]. Although direct forecasting studies alleviate the two step process they make an assumption that the future solar generation can be explicitly learned using past cloud images. Therefore, in this work, we adapt an indirect solar generation forecasting approach similar to studies [11] and [12] to capture cloud dynamics and examine their impact on solar forecasting performance.

Regardless of the approach (indirect or direct), cloud modeling is essentially a spatiotemporal prediction problem. Therefore, some of the commonly adapted deep neural networks for cloud images include variants of convolutional neural networks (CNN) due to their high spatial feature extraction capabilities in images and combinations of CNN and long short-term memory networks (LSTM) due to their ability to model temporal relationships in data. These networks include 2D CNNs, 3D CNNs, CNN + LSTM where the output of the CNN is an input to the LSTM and architectures such as convolutional long short-term memory networks (ConvLSTM) where the long short-term memory network is further modified capture both spatial and temporal dynamics. Table 1 shows some of the solar forecasting studies using cloud images and deep learning.

In [11], the authors applied a CNN + LSTM to predict three channels of a multispectral satellite image corresponding to a PV system's location given a sequence of past satellite images. The output from the CNN + LSTM, together with historical power generation and temperature data, was used to train a SVR to predict the future solar generation. Si et al. [12] adapted a ConvLSTM (proposed for spatiotemporal applications such as video prediction, precipitation forecasting) to forecast cloud movement and combined the cloud forecasts from ConvLSTM and historical power generation data to forecast solar generation using XGBoost. A study by [36] investigated both CNN + LSTM and ConvLSTM. Both networks were used to extract features from GSI combined with a simple multi-layer perceptron network (MLP) to capture features from past PV power data. While both networks demonstrated comparable performance it was observed that ConvLSTM had superior performance

Table 1

Relevant solar forecasting studies using cloud images and deep neural networks. Deep neural networks applied to process cloud images are marked in boldface under the column Method(s) Under Study.

Literature	Input data	Method under study
[11]	Satellite images, Historical PV power data and temperature data	CNN + LSTM + SVR
[12]	Satellite images, Historical PV power data	ConvLSTM + XGBoost
[13]	Satellite images, Historical PV power data	Auto-encoder with Attention in decoder
[35]	GSI and Historical PV power data	CNN + MLP CNN + LSTM + MLP
[36]	GSI and Historical PV power data	CNN + MLP CNN + LSTM + MLP ConvLSTM + MLP PredNet + MLP
[10]	GSI, Irradiance data	CNN + LSTM CNN + LSTM + LSTM
[14]	GSI	Convolutional Auto-encoders
[38]	GSI	3D CNN + AR
	Irradiance data	3D CNN + MLP
[39]	Satellite images, Historical PV power data, NWP data	ConvGRU + GRU Auto-encoder

over CNN + LSTM. ConvLSTM is one of the earliest deep learning architectures that combined the advantages of convolutional and recurrent networks into a single neural network. ConvLSTM laid the foundation for many subsequent neural network architectures that were proposed for spatiotemporal applications [22,37]. Therefore, in our work, we use ConvLSTM as the baseline network to forecast cloud movement.

Attention in deep learning is one of the most prominent mechanisms that has boosted the performance of deep neural networks for many applications [18,40]. Attention can be applied to different data modalities (e.g., sequential data, images) and multiple forms of attention have been proposed in literature. The intuition behind including any attention mechanism into learning process of a neural network is to focus on informative regions/parts of the data (i.e., focus on more important features) that can assist with the task at hand [41]. This essentially means assigning a higher weight to important features relative to less relevant ones.

Including attention mechanisms has shown promising results in several solar forecasting studies. Qu et al. [19] applied a simple attention mechanism using two fully connected layers (i.e., dense layers) and softmax activation to forecast solar generation. The attention module was used to weight (i.e., give importance to) short-term and long-term periodicity of previous solar power and temperature time series that can assist future predictions. It was observed that the attention-based method showed superior performance compared to non-attention based methods. Another attention-based method was proposed by Lai et al. [20], where features extracted from an irradiance time series were weighted using a similar attention module consisting of fully connected layers. The weighted features were passed through a gated recurrent unit network to forecast solar irradiance. When attention mechanisms are applied on image data, they are often referred to as visual attention.

Visual attention mechanisms mimic the human visual system by focusing on relevant regions within an image to make decisions [40]. Cheng et al. [13] proposed a direct solar forecasting method using satellite images and visual attention. In their work, the attention module was a CNN that weights satellite image feature maps (i.e., to give high importance to relevant cloud regions) produced by another CNN. The attention module returns a single channel feature map as the output by weighting the satellite image features maps, power at previous time steps and clear sky predictions. These features maps with attention was then used to predict the PV power using a LSTM. It was reported that 1-step ahead forecasts were comparable among attention-based and non-attention based methods, while attention-based method

showed improvements for multi-step ahead forecasts. Although, including visual attention has shown a promising direction to improve solar forecasts using cloud images for short-term forecasting, limited work has investigated how these improvements translate into downstream solar generation forecast performance. As shown in Table 1, existing studies employ a variety of methods to generate the downstream solar forecast, making it difficult to isolate and attribute performance differences to the quality of the cloud forecast alone.

3. Solar PV sites and satellite data

The solar power generation data we use in this work consists of 50 solar power generation time series obtained from residential and commercial buildings located in Perth, Western Australia. Each solar generation time series consists of one full year of data from February 2020 to February 2021 at 15 min time intervals. The location of the sites on a map of Perth are shown in Fig. 2(a). We extract the infrared satellite images from the Meteorological Satellite Center of Japan Meteorological Agency (JMA) website¹ for the same time period. The satellite images are captured by the Himawari-8 geostationary satellites which takes images of the Earth's surface in 16 spectral bands ranging from 0.47 μm to 13.3 μm . The images we study in this work correspond to the infrared channel (band 13) ranging from 10.4 μm to 11.2 μm . The satellite images have a spatial resolution of 2 km (i.e., each pixel in the image covers an area of 2 km $\times$ 2 km) and are available at 10 min time intervals. Therefore, we re-sampled the solar generation time series to 10 min intervals to ensure consistency with the satellite images. Fig. 2(b) shows an infrared satellite image for entire Australia from the JMA website. In our work, we extract a 60 $\times$ 60 pixel area from the satellite images such that the center of the image corresponds to the latitude and longitude of the locations of PV sites in Perth as shown in Fig. 2(c). As the spatial resolution of the satellite image is within few kilometers, the location of a PV site may be represented using a single pixel in the image [11,12]. Therefore, the pixel value in the satellite image reflects the clouds that are vertically above the PV site of interest.

As can be seen in Fig. 2(b), the infrared satellite image is a grayscale image where each pixel value shows an integer between 0 and 255 (0-black, 255-white). As we discussed before, clouds are identified in infrared images using temperature sensors that measure the heat radiating from the cloud tops. Since the temperature decreases with increasing height of the troposphere, clouds that reside at higher altitudes (such as *Cirrus-cirrostratus* clouds) have cold tops, while lower altitude clouds (such as *Cumulus*, *Stratocumulus* clouds) have warmer tops [7]. Clouds that have very cold tops are shown in bright white in the infrared images, while clouds with warmer tops will not appear as white in the image. Therefore, infrared images can be used to infer information about cloud altitudes, where high altitude clouds may appear closer to white color, while low altitude clouds may not appear very white. Moreover, black shows that there are no clouds present (i.e., clear sky).

Fig. 3 shows the solar power generation time series of one PV site and the corresponding satellite image (where the location of the PV site in the satellite image is marked with a red rectangle) for the timestamp marked in red in the time series. To better represent the different cloud altitude levels that may be present in the satellite image, we use a color scale to show the pixel values of the original grayscale image. Fig. 3(a) shows a day where the solar power generation from the PV site has a complete power generation profile without any fluctuations during the day. As can be seen in the corresponding satellite image, there are no clouds covering the PV site and hence the power generation is not interrupted. However, as shown in Fig. 3(b), where the power generation is interrupted during the middle of the day at multiple time intervals, the corresponding satellite image shows clouds covering the location of the PV site.

¹ <https://www.data.jma.go.jp/mscweb/data/himawari/index.html>.

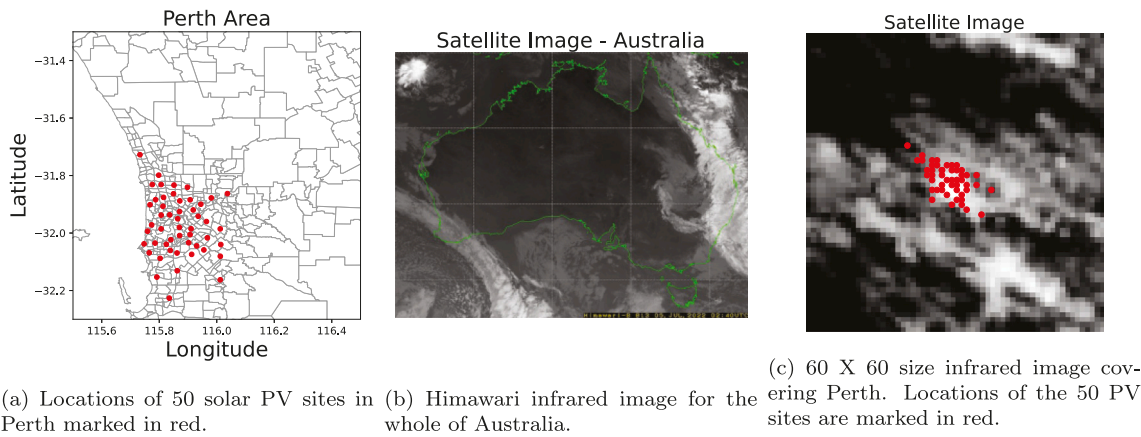


Fig. 2. Locations of the 50 PV sites in Perth and infrared satellite image from Himawari. Locations shown in Figs. 2(a) and 2(c) are in fact the same, but look different due to different projections.

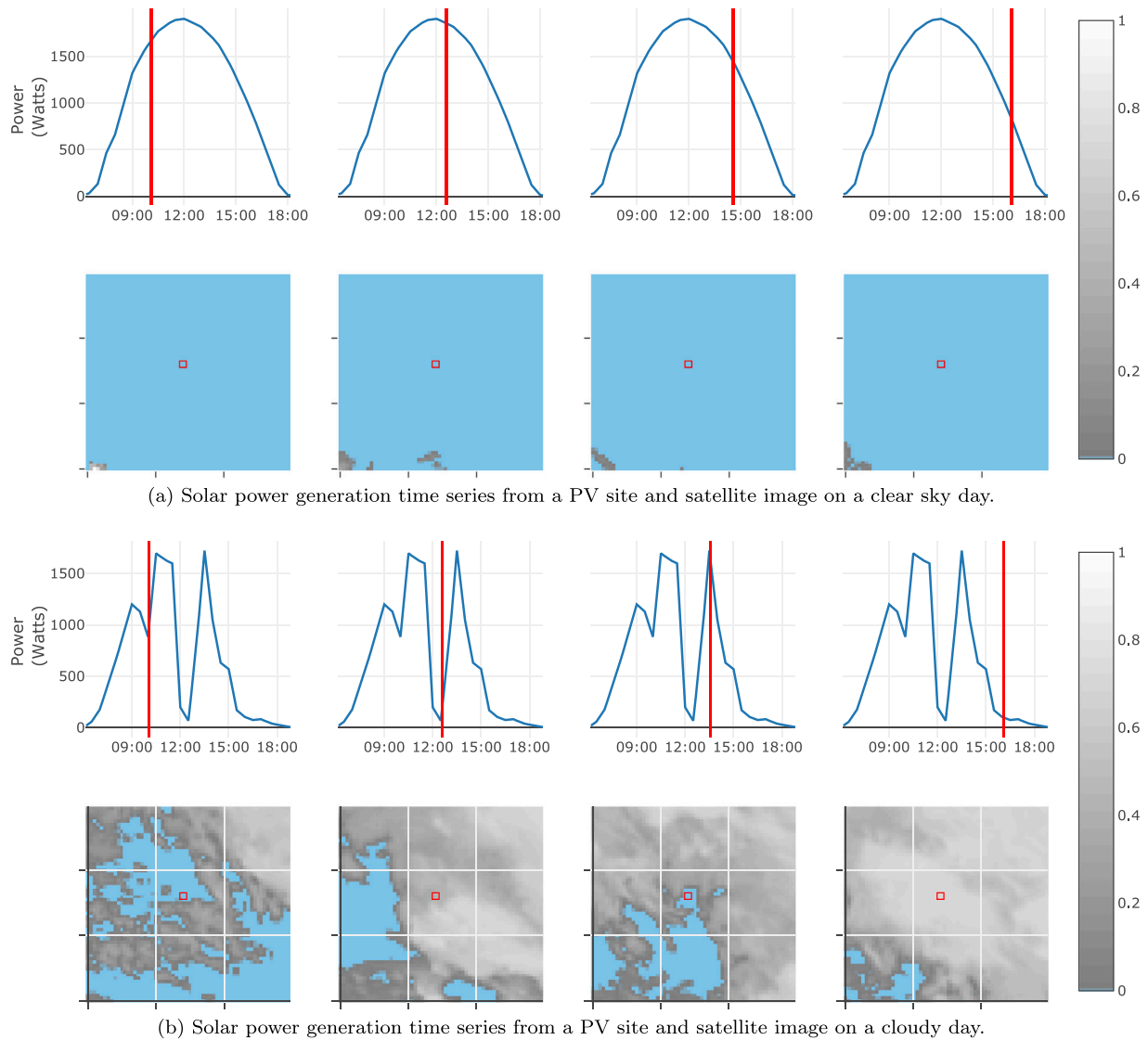


Fig. 3. Solar power generation time series from a PV site and infrared satellite images corresponding to the time marked in red in the PV power generation time series. The location of the PV site in the satellite image is marked in a red rectangle which indicates the cloud condition vertically above the PV site. Areas without clouds are shown in blue color. Color scale of the satellite images refer to the pixel value that correlates to altitude levels of clouds (1- high, 0-low).

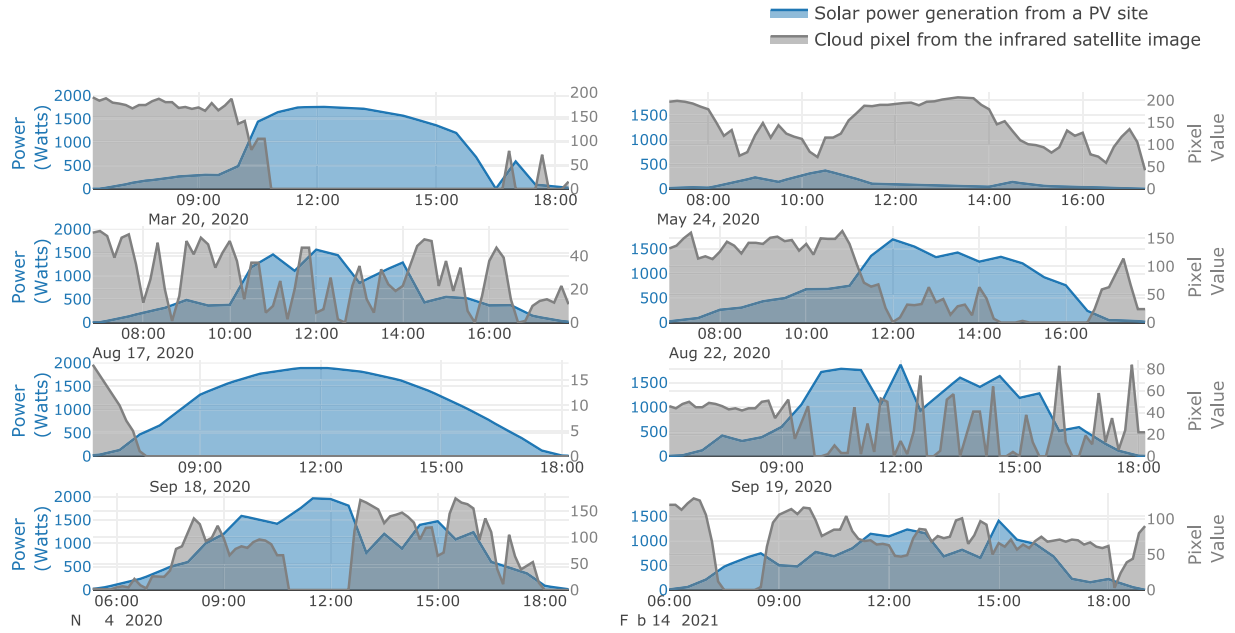


Fig. 4. Solar power generation of a PV site in Perth on eight different days and the pixel value of the infrared satellite image corresponding to the cloud condition above the site.

Similar observations can be made by examining Fig. 4 which shows the solar power generation for one PV site and the pixel value from the satellite image showing the altitude of the cloud vertically above the PV site on eight different days. We can see that in most scenarios (e.g., Mar 20 2020 between 12:00–16:00, Sept 18 2020, Nov 4 2020 around 12:00) when the pixel value is 0 (no clouds or clear sky) the solar power generation from the site is not interrupted. In contrast, during periods where the pixel value is not 0, indicating the presence of clouds, fluctuations of the power generation are visible (e.g., Mar 20 2020 between 8:00–11:00, May 24 2020, Sept 19 2020). However, a clear correlation between the solar power generation and cloud pixel (i.e., the cloud altitude) cannot be derived solely through visual observations of the data where high altitude may have a lower impact on the solar generation (due to the scattering effect of the sunlight) and low altitude clouds may have a high impact.

4. Methodology

4.1. Overall framework

Fig. 5 shows the overall framework of our work which consists of two stages (i) Training Stage, (ii) Forecasting Stage. We consider a short-term forecast horizon of 1 h at 10 min time intervals (i.e., 6 time steps into the future) where the primary driver of solar power fluctuations during these time intervals are cloud movements [9].

Both the *Training Stage* and the *Forecasting Stage* (i.e., testing phase) involve training and forecasting two categories of networks (for cloud forecasting and solar forecasting), as shown in Fig. 5. During the Training Stage the cloud forecasting and solar forecasting networks are trained separately. However, during the Forecasting Stage the output of the cloud forecasting networks are used as an input to the solar forecasting networks.

1. Cloud movement forecasting:

- (a) **Training Stage:** To forecast cloud movement, we train the deep neural networks discussed in Section 4.2. The inputs

to the cloud forecasting networks include 6 satellite images of size 60×60 corresponding to the clouds in the past hour as shown in Fig. 5. Since a 60×60 satellite image covers the locations of all PV sites under study, we train a single cloud forecasting network (of the same type) to forecast the movement of clouds that affect all 50 PV sites.

- (b) **Forecasting Stage:** We obtain the infrared satellite image predictions from the trained cloud forecasting network for the forecasting hour (shown as step 1 in Fig. 5).

2. Solar generation forecasting:

- (a) **Training Stage:** We train multiple neural networks discussed in Section 4.3 to forecast the power generation at each of the PV sites. For each PV site, we train a separate network of the same type where the inputs to the network include power generation values of the past hour and ground-truth cloud pixel values (corresponding to the PV site extracted from the infrared satellite image) for the prediction time period $(t+1, \dots, t+6)$ — as shown in Fig. 5.
- (b) **Forecasting Stage:** As shown in step 2 in Fig. 5, using the outputs from the cloud movement forecasting (from above), together with the power generation values of the past hour as inputs, we generate solar forecasts using the trained solar forecasting network corresponding to the PV site.

4.2. Deep neural networks for cloud forecasting

Forecasting cloud movement is a spatiotemporal prediction problem where the time or temporal factor relates to the changes observed across continuous satellite images and the space or spatial factor relates to the formations of clouds within a satellite image [12]. Several deep neural networks have been proposed for spatiotemporal learning in the deep learning literature [22]. One of the widely adapted structures is

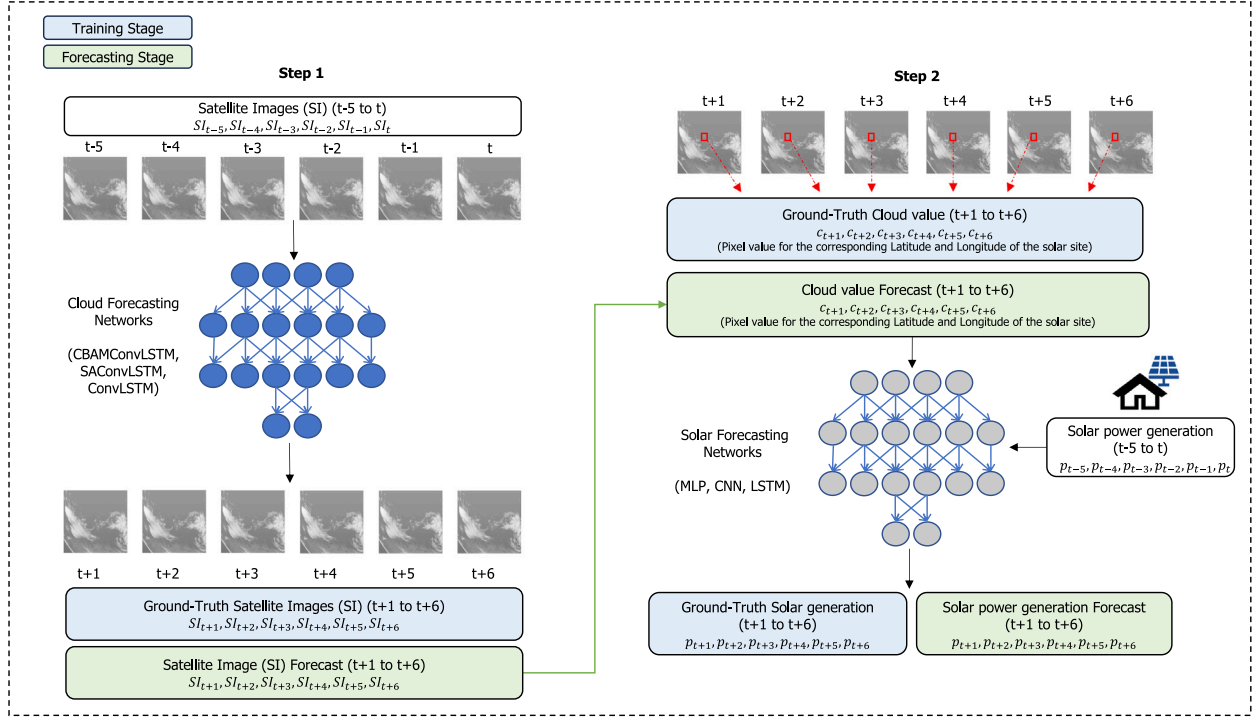


Fig. 5. Solar power forecasting framework to study the impact of cloud movement predictions from attention and non-attention-based networks on downstream solar forecasting performance.

the combination of CNN and LSTMs, which combines the advantages of convolutional and recurrent neural networks. *Convolutional LSTM* (i.e., ConvLSTM) proposed by Shi et al. [37] is one of the very early deep neural networks that is widely adapted in spatiotemporal applications and has also been adapted to forecast cloud movement in multiple solar forecasting studies [12,36,42].

In ConvLSTM, the convolutional and recurrent neural networks are combined together by replacing the matrix multiplication within a standard LSTM cell with the convolution operation to capture spatial dependencies in addition to the long and short-term modeling of the LSTM. Eqs. (1a)–(1e) shows the operations of a ConvLSTM and the matrix multiplications replaced with convolution operations are marked in boldface. i_t , o_t , f_t , C_t and h_t represent the input gate, output gate, forget gate, cell state and hidden state at time t . $W_{gate/cell}$, $b_{gate/cell}$ are the weight matrices for the respective gate or cell state associated with the hidden state or input state and $b_{gate/cell}$ indicates the corresponding bias vectors.

$$i_t = \sigma(W_{ih} * h_{t-1} + W_{ix} * x_t + b_i) \quad (1a)$$

$$o_t = \sigma(W_{oh} * h_{t-1} + W_{ox} * x_t + b_o) \quad (1b)$$

$$f_t = \sigma(W_{fh} * h_{t-1} + W_{fx} * x_t + b_f) \quad (1c)$$

$$C_t = i_t \odot \tanh(W_{ch} * h_{t-1} + W_{cx} * x_t + b_c) + f_t \odot C_{t-1} \quad (1d)$$

$$h_t = o_t \odot \tanh(C_t) \quad (1e)$$

Compared to the traditional LSTM that ignores spatial information, ConvLSTM can extract spatial dependencies and is therefore suitable for spatiotemporal applications where both spatial and temporal dynamics of the data need to be extracted. We train a ConvLSTM to forecast the satellite image sequence of the next hour in 10 min time intervals (i.e., to predict 6 images). We use this network as a benchmark method for cloud forecasting to evaluate the attention-based networks we investigate in this work.

4.2.1. Attention-based deep neural networks

Attention model or attention mechanism in neural networks in a simple form refers to giving attention (i.e., weight) to important parts of an input that can assist the task at hand and ignore (or give less focus) to the other parts and thereby enhancing the ability of a neural network to process and learn from data [18].

CBAM ConvLSTM

In this paper, we integrate an attention mechanism into the standard ConvLSTM structure to improve the ability of ConvLSTM to process spatiotemporal information by focusing on more important parts of an input image. We specifically consider the task of predicting a sequence of infrared satellite images where our objective by adding an attention mechanism to the ConvLSTM is to focus and learn the cloud regions of the infrared image, in contrast to the non-cloud regions. We adapt an attention module proposed in [43] known as convolutional block attention module (i.e., CBAM). CBAM is an attention module that can be seamlessly integrated into any convolutional neural network architecture to refine intermediate feature maps generated through a convolution layer by adding attention to these feature maps.

The basic structure of the CBAM is shown in Fig. 6. The attention module focuses on adding attention to two dimensions — the channel and spatial dimension. The channel attention module focuses on what to focus on in the image (i.e., what is meaningful). In contrast, the spatial attention module is responsible for identifying where to focus (i.e., where is the informative part in the image).

As can be seen in Fig. 6, first, the input feature map from a convolution layer is passed to the channel attention module. In this module, to focus on the channel dimension, first, the spatial information of an image is aggregated through global average and max pooling. In global average/max pooling the average/maximum across the spatial dimension is derived to convert a $H \times W \times C$ dimensional input to $1 \times 1 \times C$ dimensional output where H , W , C indicates the size of each dimension. Afterwards, these down-sampled feature maps

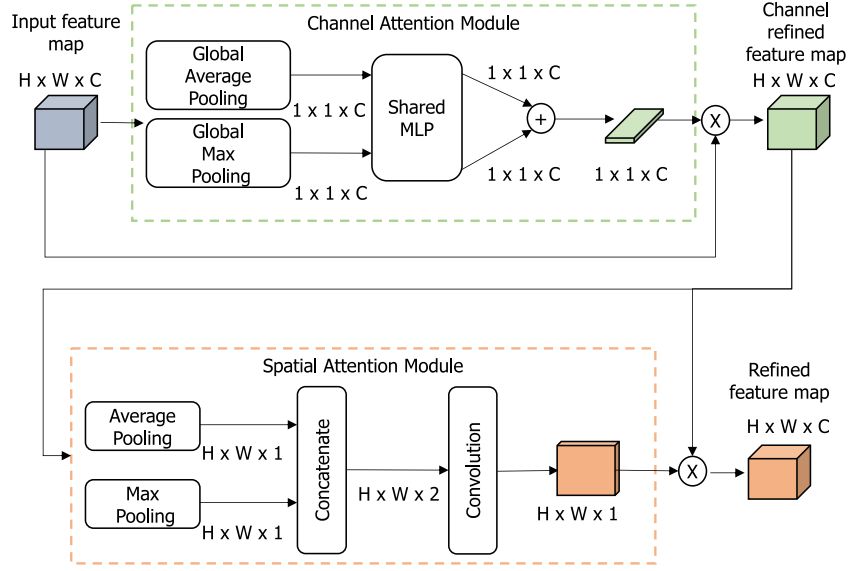


Fig. 6. Basic structure of the Convolutional Block Attention Module (CBAM). H, W, C indicates the size of three different dimensions. $\otimes$ shows element wise multiplication and $+$ indicates addition.

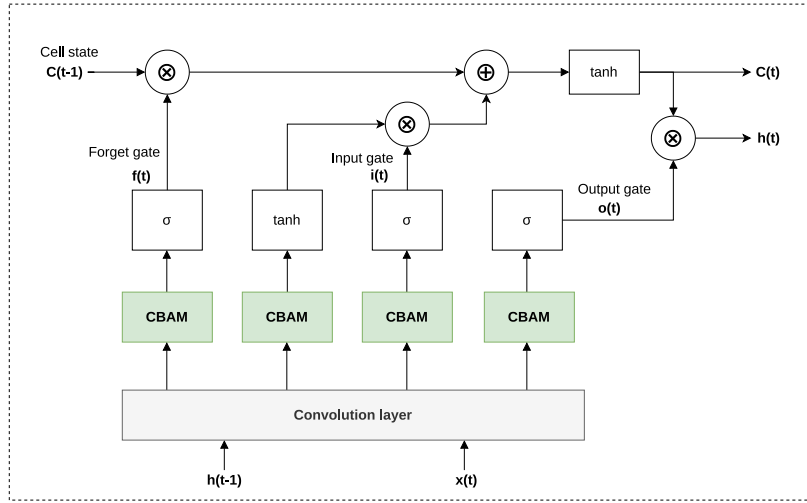


Fig. 7. CBAM ConvLSTM architecture. The addition of the CBAM modules to the ConvLSTM architecture is shown in green color. $\otimes$ shows element wise multiplication and $+$ indicates addition.

are passed through a shared MLP to produce a refined feature map across the channel dimension. The final channel refined feature map is obtained by multiplying the original feature map with the attention map obtained from the channel attention module.

The channel refined feature map is then passed through the spatial attention module, where at first, the feature map is down-samples across the channel dimension through average and max pooling. Afterwards, these two feature maps are combined together, and a convolution operation is applied to obtain an attention map across the spatial dimension. Finally, the spatial attention map is multiplied with the channel refined attention map to derive the final refined feature map with attention across both channel and spatial dimensions.

In this work, we integrate a CBAM-based attention mechanism to the ConvLSTM structure after each convolution operation within the

ConvLSTM to refine (through attention) the intermediate feature maps that are generated after each convolution operation. Fig. 7 shows the integration of the CBAM module to the current ConvLSTM architecture. Eqs. (2a)–(2e) show the operations of CBAMConvLSTM where the changes to the standard ConvLSTM operations by adding the CBAM are shown in boldface.

$$i_t = \sigma(\text{CBAM}(W_{ih} * h_{t-1}) + \text{CBAM}(W_{ix} * x_t) + b_i) \quad (2a)$$

$$o_t = \sigma(\text{CBAM}(W_{oh} * h_{t-1}) + \text{CBAM}(W_{ox} * x_t) + b_o) \quad (2b)$$

$$f_t = \sigma(\text{CBAM}(W_{fh} * h_{t-1}) + \text{CBAM}(W_{fx} * x_t) + b_f) \quad (2c)$$

$$C_t = i_t \odot \tanh(\text{CBAM}(W_{ch} * h_{t-1}) + \text{CBAM}(W_{cx} * x_t) + b_c) + f_t \odot C_{t-1} \quad (2d)$$

$$h_t = o_t \odot \tanh(C_t) \quad (2e)$$

SAConvLSTM

In this work, we further adapt a self-attention based ConvLSTM architecture (SAConvLSTM) that was proposed by Lin et al. [23] for video prediction tasks to forecast cloud movement using infrared satellite images. SAConvLSTM relies on an attention mechanism known as self-attention (also refereed to as intra-attention) where the main idea is to capture global dependencies in an input sequence [15]. In the context of images, self-attention mechanism calculates a correlation among pixels (i.e., pairwise relations) in an image to capture global spatial dependencies [23].

SAConvLSTM embeds the self-attention mechanism to the standard ConvLSTM to capture global spatial dependencies and further introduces an additional memory module to remember global spatial dependencies captured in past time steps when predicting an image sequence. To relate this concept to the context of cloud forecasting in a simplified form, capturing global dependencies in an image refers to identifying how each pixel (showing clouds in a particular location) relates to every other pixel (clouds in another location) in the image, and memorizing global spatial dependencies refers to remembering relevant spatial features captured from the past cloud images in addition to the spatial features of the current cloud image. To enable capturing such long-range spatial dependencies, SAConvLSTM has a memory module named self-attention memory module that is embedded to the standard ConvLSTM.

The following equations show the operations within the addition memory module which takes the input hidden state h_t of the current time step calculated from standard ConvLSTM operations and M_{t-1} (newly introduced memory) from the last time step. Furthermore to memorize global dependencies an aggregated feature map Z is generated by applying self-attention on the hidden state h_t and M_{t-1} features. The following equations shows the calculations of newly introduced memory M in [23].

$$i'_t = \sigma(W_{i'h} * h_t + W_{i'z} * Z + b_{i'}) \quad (3a)$$

$$o'_t = \sigma(W_{o'h} * h_t + W_{o'z} * Z + b_{o'}) \quad (3b)$$

$$M_t = i'_t \odot \tanh(W_{mh} * h_t + W_{mz} * Z + b_m) + (1 - i'_t) \odot M_{t-1} \quad (3c)$$

$$\hat{h}_t = o'_t \odot \tanh(M_t) \quad (3d)$$

Similar to the ConvLSTM, we train the attention-based ConvLSTMs (both *CBAMConvLSTM*, *SAConvLSTM*) to forecast the satellite images for the next hour (6 time steps ahead). During the forecasting stage, for all cloud forecasting networks, we conduct an autoregressive prediction to forecast the next 6 satellite images from $t+1$ to $t+6$. In this approach the predicted cloud images from $t+1$ to $t+(k-1)$ are given as inputs to predict $t+k$ image for all $k > 1$.

We train the networks using the *Adam* optimizer to minimize the structural similarity index measure (SSIM) loss which takes into account luminance (i.e., brightness), contrast and structure. SSIM is more suitable to measure the similarity of the images compared to the mean squared error loss that only measures the difference between corresponding pixels of the ground-truth and predicted image [44]. Using the SSIM loss enables the network to learn to generate images structurally similar to the ground-truth images.

Due to the computational complexity of training cloud forecasting networks using multiple hyperparameters, we limit the hyperparameter search to a fixed set of values. In particular, we tune the hyperparameters shown in Table 2 using a grid search approach where a fixed number of values to explore for each hyperparameter is given. Moreover, to ensure the network does not over-fit to the training images, we train all networks for 200 epochs with early stopping to stop training if there is no improvement in the validation loss in 10 consecutive epochs. At the beginning of training, the learning rate of all networks is set to 0.001, and a learning rate scheduler is used to reduce the learning rate by a factor of 0.1 if there is no improvement in 5 consecutive epochs.

Table 2

Hyperparameters and parameter ranges provided for cloud forecasting networks.

Hyperparameter	Range
Number of layers	1, 2, 3, 4, 5, 6
Number of hidden states per layer	32, 64
Batch size	16, 32

Table 3

Hyperparameters tuned in this work and the hyperparameters ranges provided to find the optimal hyperparameters for each neural network.

Network	Hyperparameter	Range
MLP	Learning rate	0.0001–0.1
	Number of hidden layers	1–5
	Number of neurons per layer	1–256
	Number of epochs	100–2000
	Learning rate	0.0001–0.1
1D CNN	Batch size	1–size of training samples
	Drop out rate	0–0.5
	Learning rate	0.0001–0.1
	Kernel/filter size	2, 3
	Number of filters	32, 64
LSTM	Number of convolution layers	1–5
	Drop out rate	0–0.5
	Number of epochs	500–2000
	Batch size	1–size of training samples
	Learning rate	0.0001–0.1
	Number of lstm layers	1–5
	Number of neurons per layer	32, 64
	Number of epochs	500–2000
	Learning rate	0.0001–0.1
	Batch size	1–size of training samples

Table 4

Number of training, validation and test samples.

Stage	Number of samples
Training	16,915
Validation	4229
Testing/Forecasting	2289

4.3. Solar forecasting neural networks

To forecast solar generation and to evaluate the impact of cloud predictions on their solar forecasting performance, we use three widely used state-of-the-art neural networks in the solar forecasting literature [45,46]: Multi-Layer Perceptron (MLP), 1D Convolutional Neural Network (CNN), Long Short-Term Memory Network (LSTM).

The first network is a *MLP* which is a simple feed forward neural network architecture. In feed forward neural networks the information is only passed in one direction from input neurons of the network to the output neurons. The information is passed through several hidden layers of the neural network in a MLP structure to learn a non-linear mapping between the input and output [47]. As we discussed above MLP is adapted in several forecasting studies in the literature to forecast solar generation and is one the common neural networks used to compare against more advanced networks [27,46,48].

The second network is a *LSTM* and is a type of recurrent neural network architecture that consists of feedback loops to carry information from one input step to the next in contrast to traditional feed forward neural networks. Feedback loops in RNN cells enables to capture temporal patterns in data allowing the information flow from one time step to the other and therefore, is adapted in many sequence prediction problems including forecasting studies [49]. LSTM cells have an additional gating mechanism that allows to control the flow of information by preserving relevant information and forgetting irrelevant information which enables LSTMs to capture both long and short term temporal dynamics in the data compared to traditional RNNs.

The third network we use for solar power forecasting is a 1D CNN. CNNs are a class of deep neural networks that are translation-invariant, powerful feature extractors widely adapted in computer vision applications [50]. The filter (i.e., kernel) that is essentially a weight vector is the main component in a CNN and it slides through the input and learns to recognize specific patterns in the data. Depending on the dimensions of the input and the features need to be extracted different filter dimension are used. For example 2-dimensional filters are more commonly used to capture feature across the height and width dimensions of an image. For 1-dimensional data such as time series, relationships among data points exist in the temporal dimension and therefore 1-dimensional filters are more commonly applied [51].

In this work, as discussed above, we train a separate neural network: MLP, LSTM, CNN for each PV site and train these networks to minimize the mean squared error loss with the *Adam* optimizer. Moreover, we tune the hyperparameters shown in Table 3 for each of the networks. For hyperparameter tuning we use an optimization method *Bayesian optimization* from Hyperopt python package [52]. Compared to other hyperparameter tuning methods such as grid search that conduct an exhaustive search to find the optimal hyperparameters, Bayesian optimization method limits the number of exhaustive evaluations by using information from prior evaluations and is therefore much more efficient to tune a multiple hyperparameter combinations with large ranges of possible values.

5. Evaluation

During the *Forecasting Stage*, we evaluate the impact of cloud forecasts produced by the cloud forecasting networks under the following three benchmark scenarios.

1. *Ground truth clouds* - The first scenario uses cloud pixels from the ground-truth satellite images as an input to the solar forecasting networks. This condition shows the best or “ideal” solar power forecasts that can be achieved as the cloud conditions in the future time instants are known.
2. *Persistence clouds* - The second scenario uses the cloud image observed at last timestamp before the start of the forecasting hour as inputs to the solar forecasting network without predicting the satellite images. This condition acts as a benchmark to determine if the image predictions from the cloud forecasting networks have improved over simply using the last observed image for the prediction time period.
3. *No clouds included* - The third scenario includes training the same solar forecasting networks discussed in Section 4.3 with power generation data from the past hour as the sole input to the networks. This condition is used to determine if including cloud values derived from the infrared satellite images have an impact on forecasting solar power generation.

Table 4 shows the number of training, validation and testing samples of the solar power forecasting networks per PV site. For the three cloud forecasting networks, we use similar numbers of training, validation and test image samples.

To evaluate the solar power forecasting performance, we use two error measures — Root Mean Squared Error Skill Score (RMSE Skill Score) and Mean Absolute Error Skill Score (MAE Skill Score) commonly used in the literature [53,54]. RMSE Skill Score and MAE Skill Score measure the quality of forecasts compared to a reference model, and therefore provide better error interpretability and comparison across data with different scales. The most commonly used reference model is the persistence method where the forecasts are similar to the last observed power generation value [11]. The following shows the calculation of the above error metrics where h indicates the forecast

horizon (6 steps in our case) and $\hat{P}_i$, P_i refers to the predicted and actual power generation values at a given time i .

$$RMSE = \sqrt{\frac{1}{h} \sum_{i=t+1}^{t+h} (\hat{P}_i - P_i)^2} \quad (4a)$$

$$RMSE \text{ Skill Score} = \left(1 - \frac{RMSE_{forecasting \text{ method}}}{RMSE_{persistence}}\right) * 100\% \quad (4b)$$

$$MAE = \frac{1}{h} \sum_{i=t+1}^{t+h} |\hat{P}_i - P_i| \quad (5a)$$

$$MAE \text{ Skill Score} = \left(1 - \frac{MAE_{forecasting \text{ method}}}{MAE_{persistence}}\right) * 100\% \quad (5b)$$

A higher value for the RMSE/MAE skill score indicates that the forecasting method performs better than a persistence method. A skill score of 0% indicates the forecasts quality is similar to the quality of a persistence method. We calculate the RMSE and MAE skill scores across all test samples of a PV site and then take the average error across all test samples of a PV site. Finally, we calculate the average RMSE/MAE skill scores across all 50 PV sites obtain the final errors.

6. Results and discussion

6.1. Cloud forecasts — Visualizations

Fig. 8 shows the predictions from cloud forecasting deep neural networks for four randomly selected test samples with varying cloud conditions. We can observe that all cloud forecasting networks produce output images similar to the ground truth image in the immediate future time instant ($t + 1$). However, we can see that as the number of time instants to the future increase the similarity with the ground truth image is reduced. We quantitatively evaluate the quality of these cloud forecasts and their impact on quality of the solar power forecasts considering all PV sites under study in Section 6.2. As the solar forecasting network parameters are fixed after training and the only varying input is the source of the cloud forecast, any observed differences in solar forecast accuracy can be directly attributed to the quality of the cloud predictions, and therefore provides a clear and interpretable indicator of cloud forecast quality.

6.2. Impact of cloud forecasts on solar power forecasts

Table 5 shows the average RMSE Skill Score across the 50 PV sites considering all test samples for different cloud inputs during the forecasting stage. As discussed above, “Ground-truth clouds” shows the best solar forecasts (i.e., “ideal” scenario) that can be achieved as the ground-truth cloud values are known. Therefore, if the cloud forecasts are similar to the ground-truth, the skill score will be closer to the ground-truth clouds scenario. The RMSE Skill Score difference in comparison to using ground-truth clouds is also shown in the Table. A lower value in the Skill Score Difference column therefore, indicates the solar forecasts are closer to the best skill score that can be achieved. MAE skill scores are presented in Appendix.

When comparing the ground-truth clouds and no clouds scenarios, we can see that including clouds as an input can improve the solar forecasts. However, the LSTM trained without clouds can achieve a RMSE skill score of 36.60%, which is greater than MLP and CNN networks using ground-truth clouds as inputs. This may relate to the ability of LSTMs to capture long-term temporal patterns of time series data compared to other networks and, therefore, have been able to forecast the next-hour by effectively learning temporal patterns in the past hour. Nonetheless, we can see that including clouds to the LSTM network further improve the solar forecasts.

Considering the scenarios where cloud forecasts are given as an input, we can see that irrespective of the type of solar forecasting

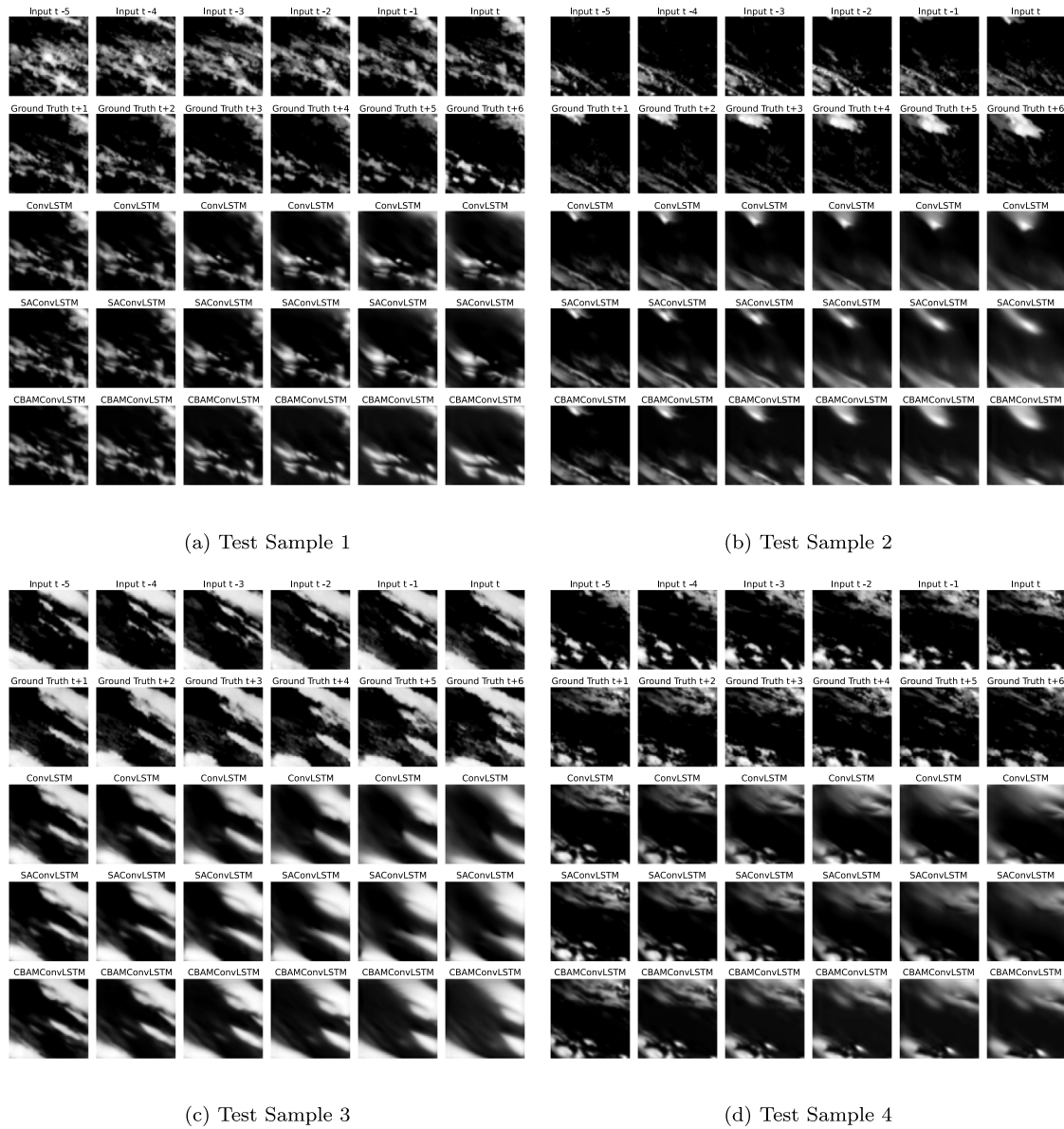


Fig. 8. Visualization examples of predicted cloud images from ConvLSTM, SAConvLSTM and CBAMConvLSTM for 4 randomly selected test samples with varying cloud conditions. The first row shows the input images to the network, the second row shows the ground truth satellite images for the forecasting period and the next three rows shows the predictions from ConvLSTM, SAConvLSTM and CBAMConvLSTM networks respectively.

Table 5

Average RMSE skill score across 50 rooftop solar PV sites considering all test samples of a site and the respective skill score difference with regard to using *Ground-truth clouds* showing the best skill score that can be achieved. The lowest skill score difference is marked in boldface.

Cloud input during Forecasting stage	Solar power forecasting network					
	MLP		CNN		LSTM	
	RMSE skill score (%)	Skill score difference	RMSE skill score (%)	Skill score difference	RMSE skill score (%)	Skill score difference
Ground-truth clouds	34.26	–	32.79	–	40.28	–
Forecasted clouds from ConvLSTM	30.00	4.26	28.85	3.94	35.46	4.82
Forecasted clouds from SAConvLSTM	31.33	2.93	30.02	2.77	36.82	3.46
Forecasted clouds from CBAMConvLSTM	31.57	2.69	30.39	2.40	37.15	3.13
Persistence clouds	28.63	5.63	27.02	5.77	34.02	6.26
No clouds included	18.27	15.99	30.87	1.92	36.60	3.68

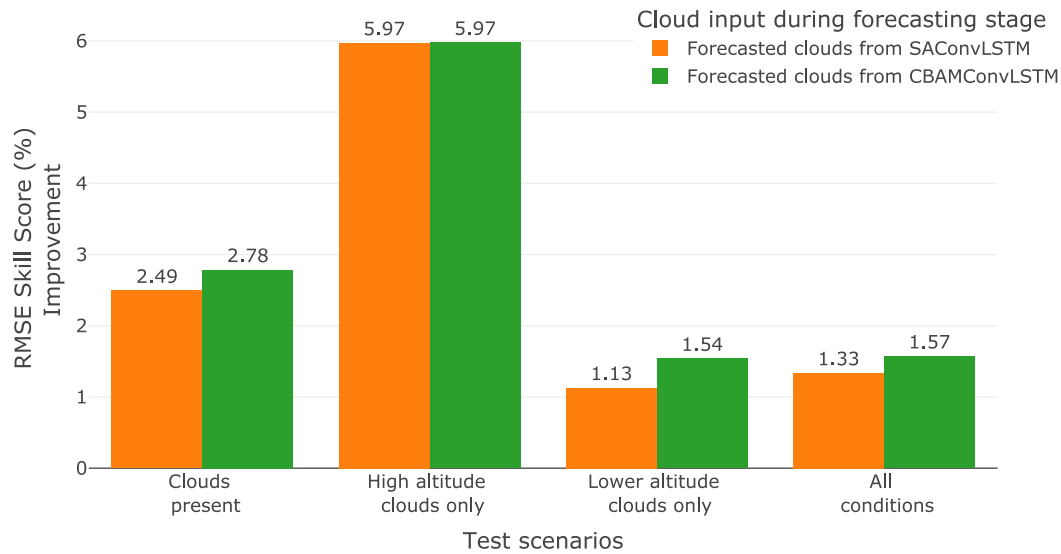


Fig. 9. Improvement in the RMSE skill score of the MLP network under different test scenarios when using cloud forecasts from attention-based methods — SAConvLSTM and CBAMConvLSTM compared to using cloud forecasts from ConvLSTM.

Table 6

Average percentage of clear sky, high altitude clouds, low altitude clouds, and all cloudy conditions present across all test samples of the 50 PV sites under study.

Sky condition	Percentage of test samples
Clear sky	46.57%
High altitude clouds	16.24%
Low altitude clouds	37.19%
Cloudy (all clouds)	53.43%

network, including cloud forecasts from ConvLSTM, SAConvLSTM and CBAMConvLSTM can improve the forecasts compared to using persistence cloud forecasts as an input. Furthermore, cloud forecasts produced by the attention-based deep neural networks (SAConvLSTM and CBAMConvLSTM) lead to improved solar forecasts compared to the standard ConvLSTM derived cloud forecasts, and using CBAMConvLSTM cloud forecasts improving over SAConvLSTM.

An interesting observation can be made regarding the importance of the quality of the cloud forecasts based on the type of solar forecasting network. We can see that for a simple MLP network including cloud predictions always improve the solar forecasts compared to not including clouds. However, for CNN and LSTM networks including cloud forecasts does not always improve the solar forecasts which indicates the importance of the cloud prediction accuracy (i.e., forecasted cloud values being closer to the ground-truth cloud values).

We further investigate this observation by evaluating the impact of cloud forecasts under different sky conditions inferred through the satellite images:

1. Clear sky (i.e., pixel value of 0) — only considering test samples with clear sky conditions, and
2. Cloudy (i.e., pixel value >0) — only considering test samples where clouds are present.

We further divide the cloudy condition into two conditions as high altitude clouds and low altitude clouds. For this categorization we consider a pixel value of 50, if the cloud pixel covering the PV site is greater than 50, we consider it as a high altitude cloud condition (as we discussed above a brighter color in the infrared image reflects higher altitude clouds), while a pixel value less than or equal 50 and greater than 0 is a low altitude cloud condition. Table 6 shows the

percentage of test samples available for the four sky conditions from all test samples under study.

Tables 7, 8, 9, 10 shows the RMSE skill scores for clear sky, high altitude clouds, low altitude clouds and cloudy conditions respectively. Under clear sky conditions, forecasting clouds using a deep neural network or persistence clouds does not make a difference as expected. Moreover, a higher skill score can be achieved as there are no clouds that impact solar generation. LSTM and CNN trained without including clouds achieve a skill score closer to the scenario of including ground-truth clouds. This indicates the ability of CNN and LSTM networks to capture temporal patterns in the data and forecast the solar generation under clear sky conditions learning only through the past hour's power generation patterns. On the contrary, the MLP network can achieve a higher skill score if trained using cloud inputs. This could be due to the inability of the MLP network to capture temporal patterns in the solar generation time series and therefore, challenging to learn only using past hour's solar generation information.

Under high altitude cloud conditions (Table 8), solar forecast from all three networks are improved by including ground-truth clouds which indicate that if the cloud predictions are closer to the ground-truth cloud values the solar forecasts can be improved. For all three networks, using CBAMConvLSTM cloud forecasts as an input result in a skill score closer to the ground-truth cloud condition followed by the SAConvLSTM. However, including cloud forecasts from ConvLSTM or persistence clouds deteriorates the forecasts compared to training without any clouds. These observations emphasize to the importance of having accurate cloud forecasts when forecasting solar generation.

Under low altitude cloud conditions (Table 9), similar to high altitude conditions, solar forecasts from all three networks are improved when ground-truth cloud are included. However, including cloud forecasts have increased the skill score closer to ground-truth cloud's skill score only for the MLP network. For LSTM and CNN networks using cloud predictions from any deep neural network or persistence clouds deteriorate the solar forecasts. This could also be because when clouds affecting the solar generation are at lower altitudes, the impact on the solar generation is high (i.e., the generation is lower) and therefore, when the cloud input is deviated from the ground-truth value, CNN and LSTM networks are unable predict the low solar generation accurately. Nonetheless, the LSTM network using cloud predictions from attention-based methods, CBAMConvLSTM and SAConvLSTM achieves higher skill scores in all cloud input scenarios compared to the MLP network.

Table 7

Average RMSE skill score across 50 rooftop solar PV sites considering only clear sky test samples of a site and the respective skill score difference with regard to using *Ground-truth clouds* showing the best skill score that can be achieved. The lowest skill score difference is marked in boldface.

Cloud input during Forecasting stage	Solar power forecasting network					
	MLP		CNN		LSTM	
	RMSE skill score (%)	Skill score difference	RMSE skill score (%)	Skill score difference	RMSE skill score (%)	Skill score difference
Ground-truth clouds	51.11	–	51.41	–	62.47	–
Forecasted clouds from ConvLSTM	50.56	0.55	50.78	0.63	61.80	0.67
Forecasted clouds from SAConvLSTM	50.41	0.70	50.91	0.50	61.60	0.87
Forecasted clouds from CBAMConvLSTM	50.56	0.55	50.99	0.42	61.84	0.63
Persistence clouds	50.36	0.75	50.59	0.82	61.67	0.80
No clouds included	29.27	21.84	50.23	1.18	59.87	2.60

Table 8

Average RMSE skill score across 50 rooftop solar PV sites considering only high altitude cloud condition test samples of a site and the respective skill score difference with regard to using *Ground-truth clouds* showing the best skill score that can be achieved. The lowest skill score difference is marked in boldface.

Cloud input during Forecasting stage	Solar power forecasting network					
	MLP		CNN		LSTM	
	RMSE skill score (%)	Skill score difference	RMSE skill score (%)	Skill score difference	RMSE skill score (%)	Skill score difference
Ground-truth clouds	17.56	–	13.23	–	20.7	–
Forecasted clouds from ConvLSTM	9.76	7.80	6.20	7.03	11.79	8.91
Forecasted clouds from SAConvLSTM	15.73	1.83	11.64	1.59	18.09	2.61
Forecasted clouds from CBAMConvLSTM	15.74	1.82	12.06	1.17	18.46	2.24
Persistence clouds	10.38	7.18	5.07	8.16	11.04	9.66
No clouds included	3.36	14.20	11.28	1.95	16.17	4.53

Table 9

Average RMSE skill score across 50 rooftop solar PV sites considering only low altitude cloud condition test samples of a site and the respective skill score difference with regard to using *Ground-truth clouds* showing the best skill score that can be achieved. The lowest skill score difference is marked in boldface.

Cloud input during Forecasting stage	Solar power forecasting network					
	MLP		CNN		LSTM	
	RMSE skill score (%)	Skill score difference	RMSE skill score (%)	Skill score difference	RMSE skill score (%)	Skill score difference
Ground-truth clouds	22.99	–	20.7	–	24.42	–
Forecasted clouds from ConvLSTM	16.16	6.83	14.34	6.36	16.79	7.63
Forecasted clouds from SAConvLSTM	17.28	5.71	15.22	5.48	17.90	6.52
Forecasted clouds from CBAMConvLSTM	17.69	5.30	15.75	4.95	18.29	6.13
Persistence clouds	12.81	10.18	10.64	10.06	13.72	10.70
No clouds included	12.12	10.87	18.05	2.65	20.00	4.42

Table 10

Average RMSE skill score across 50 rooftop solar PV sites considering only sky conditions where clouds are present in test samples of a site and the respective skill score difference with regard to using *Ground-truth clouds* showing the best skill score that can be achieved. The lowest skill score difference is marked in boldface.

Cloud input during Forecasting stage	Solar power forecasting network					
	MLP		CNN		LSTM	
	RMSE skill score (%)	Skill score difference	RMSE skill score (%)	Skill score difference	RMSE skill score (%)	Skill score difference
Ground-truth clouds	21.5	–	18.64	–	23.39	–
Forecasted clouds from ConvLSTM	14.39	7.11	12.08	6.56	15.4	7.99
Forecasted clouds from SAConvLSTM	16.88	4.62	14.25	4.39	17.96	5.43
Forecasted clouds from CBAMConvLSTM	17.17	4.33	14.74	3.90	18.35	5.04
Persistence clouds	12.15	9.35	9.10	9.54	12.97	10.42
No clouds included	9.67	11.83	16.17	2.47	18.93	4.46

Table 10 summarizes the results considering all cloudy test samples. The results show a similar observations as the low altitude cloud test samples where including ground-truth clouds can improve the solar forecasts but using cloud forecasts can deteriorate the solar forecasts for CNN and LSTM networks. These results are consistent with low altitude clouds as majority of the cloudy test samples consist of low altitude cloud conditions as seen in Table 6. It is important to note that ground-truth cloud observations are unavailable at the time of prediction in real-world operational settings. In practice, backtesting forecast models across historical periods allows to systematically understand the

performance of each method under different conditions and inputs. In such scenarios, where systematic variations are observed across models under different conditions, an ensemble framework where the outputs of multiple models utilizing different inputs are collectively used in generating the final forecast can be considered as a practical solution.

6.3. Comparison of attention-based and non-attention cloud forecasting

Figs. 9, 10, 11 show the RMSE skill score improvements when using cloud forecasts from CBAMConvLSTM, SAConvLSTM compared

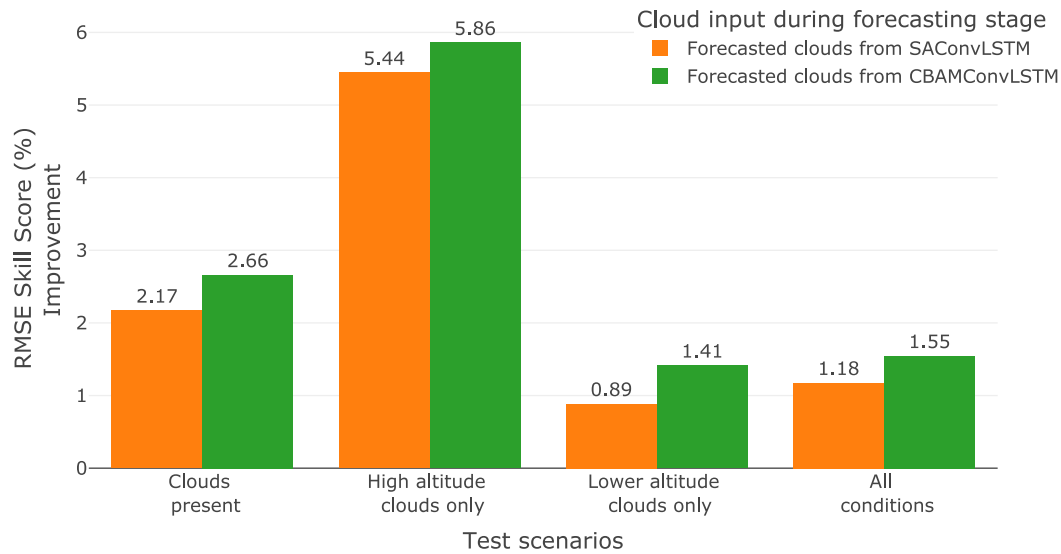


Fig. 10. Improvement in the RMSE skill score of the CNN network under different test scenarios when using cloud forecasts from attention-based methods — SAConvLSTM and CBAMConvLSTM compared to using cloud forecasts from ConvLSTM.

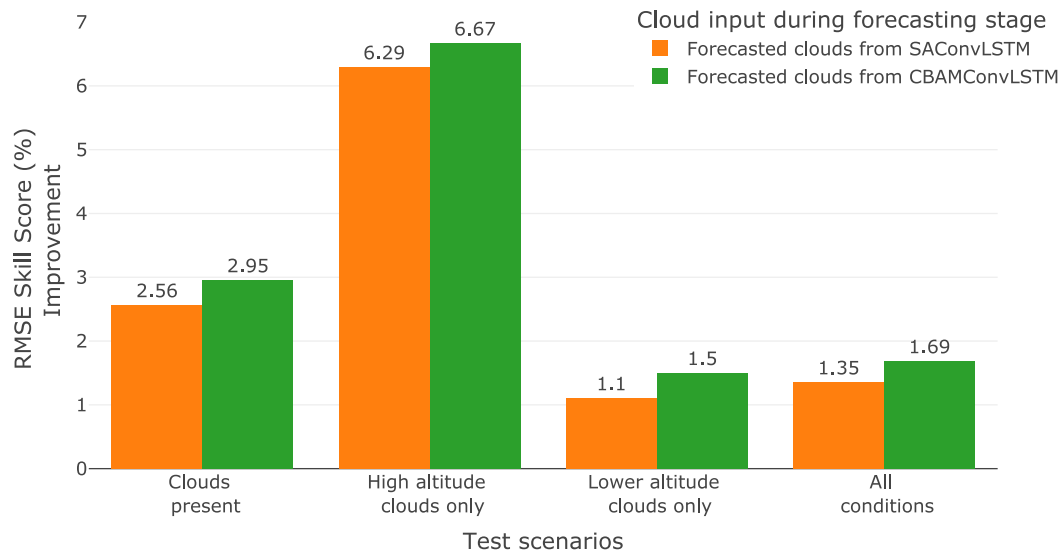
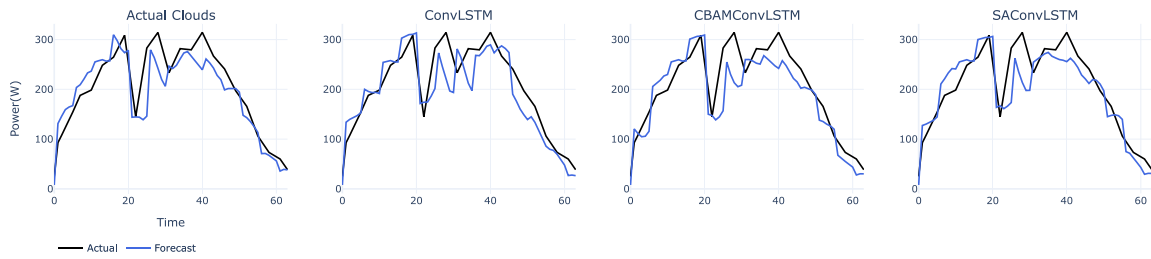


Fig. 11. Improvement in the RMSE skill score of the LSTM network under different test scenarios when using cloud forecasts from attention-based methods — SAConvLSTM and CBAMConvLSTM compared to using cloud forecasts from ConvLSTM.

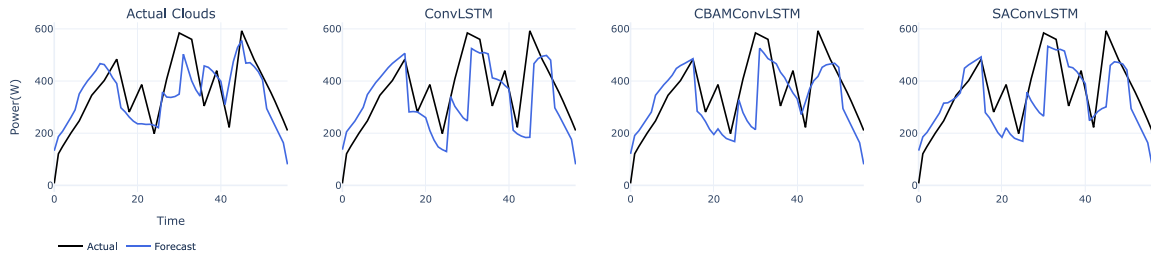
to ConvLSTM. We can see that using cloud forecasts from CBAMConvLSTM and SAConvLSTM have improved the skill scores compared to using cloud forecasts from ConvLSTM under all sky conditions. In test samples where clouds are present, CBAMConvLSTM increases the RMSE skill score by 2.78, 2.66, 2.95 for MLP, CNN, LSTM networks respectively in comparison to ConvLSTM. We can observe the best improvement with attention-based networks in high altitude cloud test samples where CBAMConvLSTM cloud inputs result in a skill score increase of 5.97 (MLP), 5.86 (CNN), 6.67 (LSTM) and SAConvLSTM result in a skill score increase of 5.97 (MLP), 5.44 (CNN), 6.29 (LSTM). Under low altitude cloud test samples the skill score increase is comparatively lower where CBAMConvLSTM cloud inputs result in a skill score increase of 1.54 (MLP), 1.41 (CNN), 1.50 (LSTM) and SAConvLSTM

result in a skill score increase of 1.13 (MLP), 0.89 (CNN), 1.10 (LSTM). It is noteworthy, that we consider high altitude cloud scenarios in this work when pixel values greater than 50. These values appear in colors closer to white in the infrared satellite image. Therefore, it is possible that the attention-based networks have been successful predicting pixel values closer to shades of white in the satellite image compared to other areas contributing to the skill score improvements in high altitude cloud scenarios.

Fig. 12 presents a comparison of actual solar generation at two sites against solar forecasts produced by the LSTM network, using both actual cloud observations and cloud forecasts derived from ConvLSTM, CBAMConvLSTM, and SAConvLSTM. As shown in Fig. 12(a), CBAMConvLSTM and SAConvLSTM more accurately capture the rapid



(a) Solar PV site 1.



(b) Solar PV site 2.

Fig. 12. A comparison of actual solar generation from two solar sites with forecasts generated by the LSTM solar forecasting network using actual cloud observations and cloud forecasts from ConvLSTM, CBAMConvLSTM and SAConvLSTM.

fluctuations in solar generation (e.g., time indices 20–50) compared to ConvLSTM-based cloud forecasts. Similarly, Fig. 12(b) demonstrates that the fluctuations around time index 40 are more accurately captured when using cloud forecasts from CBAMConvLSTM and SAConvLSTM, whereas ConvLSTM consistently under-forecasts during this period.

6.4. Limitations and future directions

In this work, as discussed above in Section 4.1, we focused on a two-step process to forecast solar generation using cloud predictions. Therefore, we trained attention-based networks to predict a sequence of infrared satellite images given a past sequence of the images. Subsequently, a pixel in the predicted images corresponding to the latitude and longitude of a PV site was extracted as the clouds covering the PV site to train the solar forecasting methods. These extracted pixels represent the clouds vertically above the PV site as we discussed before. However, clouds covering the sunlight may not be vertically above the site depending on the angle of sunlight reaching the ground [12]. As the satellite images covered a 2 km spatial resolution, a pixel in the image covered a relatively large area and therefore, taking the pixel matching the latitude and longitude was reasonable. However, this approach may be extended by finding the exact cloud pixel covering the PV site using information such as solar azimuth, zenith, and cloud heights.

We further divided the cloud conditions as high and low altitude clouds based on the pixel values of the satellite image. However, this categorization may not precisely reflect high and low altitude clouds as presented in some works such as [7] as the cloud altitude is not explicitly calculated in our work.

The work presented in this paper focuses on infrared satellite images (band 13) that can be used to infer information on cloud altitudes. As a natural extension, future studies could explore other satellite data sources, such as visible imagery, which captures cloud thickness information, or *Himawari-based Shortwave Radiation* products that more directly represent the solar irradiance reaching earth surface. Additionally, given that brightness temperature observations in infrared images may carry residual diurnal patterns beyond cloud effects, investigating

pre-processing strategies to better isolate cloud-related information represents another worthwhile direction.

Furthermore, this work focuses on short-term solar generation forecasting via an indirect two-stage approach, in which cloud forecasting serves as an interpretable intermediate step, enabling the impact of attention-based mechanisms on cloud prediction to be systematically assessed. As discussed in Section 2, direct forecasting methods also exist that consolidate the cloud and solar forecasting stages into a single end-to-end pipeline. A rigorous comparison with such direct attention-based methods for solar generation forecasting, together with a systematic analysis of computational costs across approaches, represents a valuable and natural extension of this work.

Recently, novel deep neural network architectures, such as PreDRNN [22] have been proposed for spatio-temporal applications to overcome the existing limitations of standard ConvLSTM. As future work, it would be interesting to investigate the adaption of such networks and their ability to forecast cloud movement using satellite images.

7. Conclusions

The movement of clouds is the primary driver of short-term fluctuations of solar generation. In this work, we studied the impact of using cloud forecasts generated from attention and non-attention based deep neural networks on the down-stream task of short-term solar generation forecasting. To empirically investigate how attention-based improvements in cloud forecasting impact solar forecasting, we proposed an attention-based deep neural network (CBAMConvLSTM) which extends the ConvLSTM network widely used for spatio-temporal prediction. In addition, we adapted another attention-based deep neural network (SAConvLSTM) originally developed for video prediction. Both attention-based networks were compared against the non-attention based baseline ConvLSTM. The impact of cloud forecasts when forecasting solar generation was evaluated for 50 small-scale rooftop solar PV sites in Australia, using three state-of-the-art neural networks (MLP, CNN, LSTM) trained with and without cloud values as an input to the networks. The quality of cloud predictions from CBAMConvLSTM,

Table A.11

Average MAE skill score across 50 rooftop solar PV sites considering all test samples of a site and the respective skill score difference with regard to using *Ground-truth clouds* showing the best skill score that can be achieved. The lowest skill score difference is marked in boldface.

Cloud input during Forecasting stage	Solar power forecasting network					
	MLP		CNN		LSTM	
	MAE skill score (%)	Skill score difference	MAE skill score (%)	Skill score difference	MAE skill score (%)	Skill score difference
Ground-truth clouds	34.58	–	32.84	–	41.64	–
Forecasted clouds from ConvLSTM	30.65	3.93	29.17	3.67	37.20	4.44
Forecasted clouds from SConvLSTM	31.90	2.68	30.28	2.56	38.45	3.19
Forecasted clouds from CBAMConvLSTM	32.10	2.48	30.62	2.22	38.76	2.88
Persistence clouds	29.23	5.35	27.29	5.55	35.66	5.59
No clouds included	18.25	16.33	31.76	1.08	38.57	3.07

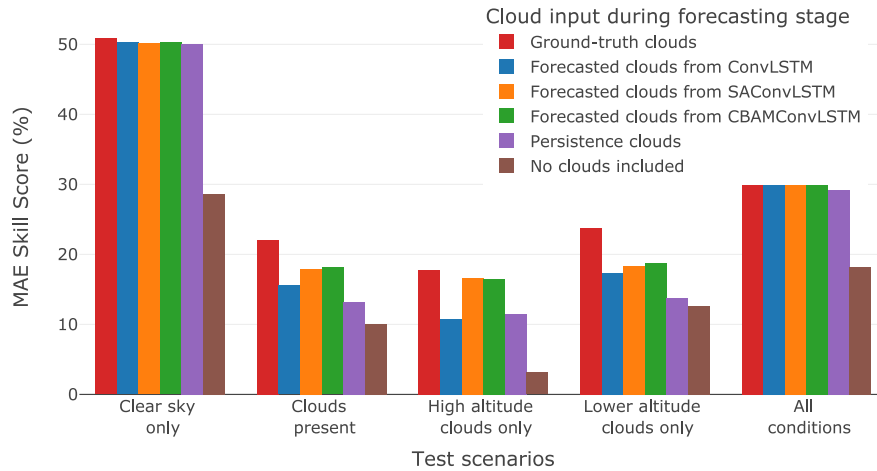


Fig. A.13. Average MAE Skill Score across 50 PV sites from the MLP network under different test scenarios and cloud inputs.

SConvLSTM and ConvLSTM were evaluated against using ground-truth and persistence cloud values as inputs to the solar forecasting networks.

Our results demonstrated that using cloud forecasts from attention-based networks achieved a higher RMSE solar forecast skill score compared to using cloud forecasts from non-attention-based ConvLSTM and using LSTM with cloud predictions from CBAMConvLSTM as input to the LSTM network achieves the best solar power forecasts. When analyzed across different sky conditions, it was observed that in clear sky conditions, the inclusion of cloud forecasts (ground-truth or predicted) does not make much of a difference in the solar forecast accuracy. However, for intervals having cloudy skies, including cloud forecasts as an input to solar forecasting methods resulted in higher solar forecast skill score in most scenarios. In particular, using cloud predictions from the proposed CBAMConvLSTM improved the RMSE skill score by 5.86%, 5.97%, 6.67% for MLP, CNN and LSTM networks respectively, while SConvLSTM improved the RMSE skill score 5.44%, 5.97%, 6.29% under high altitude cloud scenarios (i.e., pixel values greater than 50 in a gray-scale infrared image). However, the improvements under lower altitude conditions (i.e., pixel values less than 50) were minimal.

CRediT authorship contribution statement

Maneesha Perera: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Julian De Hoog:** Writing – review & editing, Supervision, Methodology, Conceptualization. **Kasun Bandara:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Hansani Weeratunge: Writing – review & editing, Conceptualization. **Saman Halgamuge:** Writing – review & editing, Supervision, Methodology, Conceptualization.

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.

Acknowledgments

The authors are grateful to Solar Analytics for providing anonymized solar power generation data to conduct this research. This work was supported by the Melbourne Research Scholarship awarded to the first author. SH and JDH acknowledge Australian Research Council grant DP220101035. The research was undertaken using the LIEF HPC-GPGPU Facility hosted at the University of Melbourne (this facility was established with the assistance of LIEF Grant LE170100200).

Appendix. Average MAE skill scores

Table A.11 shows the average RMSE skill score considering all test samples under all sky conditions. Figs. A.13, A.15, A.14 shows the average MAE skill scores for MLP, CNN and LSTM networks under test scenarios of different sky conditions studied in this work.

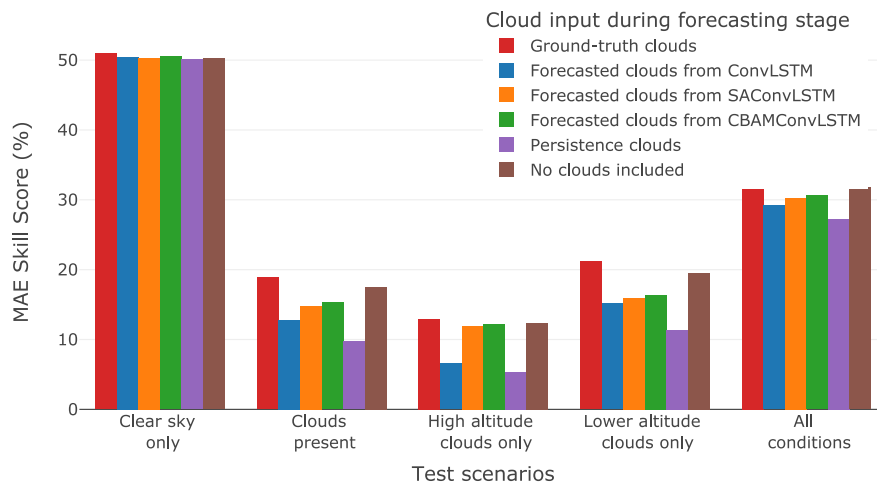


Fig. A.14. Average MAE Skill Score across 50 PV sites from the CNN network under different test scenarios and cloud inputs.

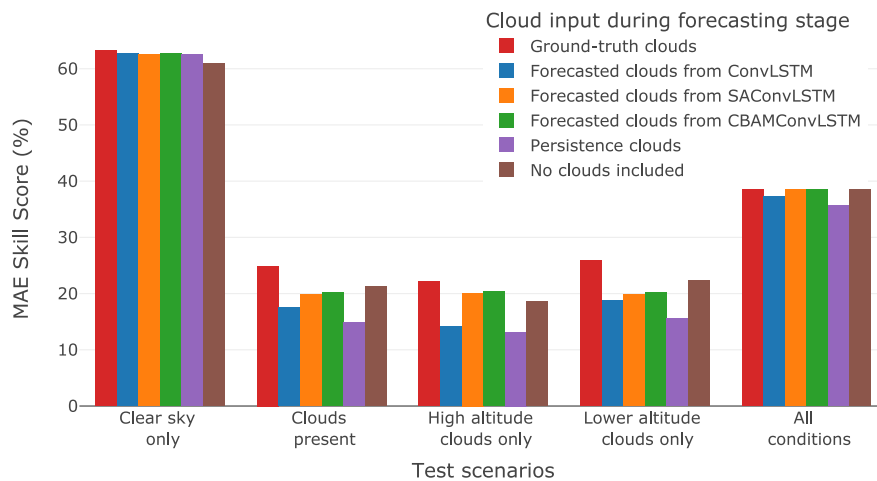


Fig. A.15. Average MAE Skill Score across 50 PV sites from the LSTM network under different test scenarios and cloud inputs.

Data availability

Data will be made available on request.

References

- [1] IEA. World energy outlook 2021. Technical Report, Paris; 2021, p. 1–386, URL: <https://www.iea.org/reports/world-energy-outlook-2021>.
- [2] Erdener BC, Feng C, Doubleday K, Florita A, Hodge BM. A review of behind-the-meter solar forecasting. *Renew Sustain Energy Rev* 2022;160(November 2021):112224. <http://dx.doi.org/10.1016/j.rser.2022.112224>.
- [3] Hou Q, Zhang N, Du E, Miao M, Peng F, Kang C. Probabilistic duck curve in high PV penetration power system: Concept, modeling, and empirical analysis in China. *Appl Energy* 2019;242:205–15. <http://dx.doi.org/10.1016/J.APENERGY.2019.03.067>.
- [4] Yang D, Wu E, Kleissl J. Operational solar forecasting for the real-time market. *Int J Forecast* 2019;35(4):1499–519. <http://dx.doi.org/10.1016/j.ijforecast.2019.03.009>.
- [5] Antonanzas J, Osorio N, Escobar R, Urraca R, Martinez-de Pison FJ, Antonanzas-Torres F. Review of photovoltaic power forecasting. *Sol Energy* 2016;136:78–111. <http://dx.doi.org/10.1016/j.solener.2016.06.069>.
- [6] Nespoli A, Niccolai A, Ogliaresi E, Perego G, Collino E, Ronzio D. Machine Learning techniques for solar irradiation nowcasting: Cloud type classification forecast through satellite data and imagery. *Appl Energy* 2022;305(April 2021):117834. <http://dx.doi.org/10.1016/j.apenergy.2021.117834>.
- [7] Barbieri F, Rajakaruna S, Ghosh A. Very short-term photovoltaic power forecasting with cloud modeling: A review. *Renew Sustain Energy Rev* 2017;75:242–63. <http://dx.doi.org/10.1016/j.rser.2016.10.068>.
- [8] Perera M, De Hoog J, Bandara K, Halgamuge S. Multi-resolution, multi-horizon distributed solar PV power forecasting with forecast combinations. *Expert Syst Appl* 2022;205:117690.
- [9] Lin F, Zhang Y, Wang J. Recent advances in intra-hour solar forecasting: A review of ground-based sky image methods. *Int J Forecast* 2022. <http://dx.doi.org/10.1016/j.ijforecast.2021.11.002>.
- [10] Ajith M, Martínez-Ramón M. Deep learning based solar radiation micro forecast by fusion of infrared cloud images and radiation data. *Appl Energy* 2021;294. <http://dx.doi.org/10.1016/J.APENERGY.2021.117014>.
- [11] Bansal AS, Bansal T, Irwin D. A moment in the sun: Solar nowcasting from multispectral satellite data using self-supervised learning. In: Proceedings of the thirteenth ACM international conference on future energy systems. 2022, p. 251–62, URL: <http://arxiv.org/abs/2112.13974>.
- [12] Si Z, Yang M, Yu Y, Ding T. Photovoltaic power forecast based on satellite images considering effects of solar position. *Appl Energy* 2021;302(May):117514. <http://dx.doi.org/10.1016/j.apenergy.2021.117514>.
- [13] Cheng L, Zang H, Wei Z, Ding T, Xu R, Sun G. Short-term solar power prediction learning directly from satellite images with regions of interest. *IEEE Trans Sustain Energy* 2022;13(1):629–39. <http://dx.doi.org/10.1109/TSTE.2021.3123476>.
- [14] Fu Y, Chai H, Zhen Z, Wang F, Xu X, Li K, Shafie-Khah M, Dehghanian P, Catalao JP. Sky image prediction model based on convolutional auto-encoder for minutely solar PV power forecasting. *IEEE Trans Ind Appl* 2021;57(4):3272–81. <http://dx.doi.org/10.1109/TIA.2021.3072025>.
- [15] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser {L}, Polosukhin I. Attention is all you need. In: Advances in neural information processing systems. Vol. 30, 2017, <http://dx.doi.org/10.1109/2943.974352>.
- [16] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: Proceedings - 30th IEEE conference on computer vision and pattern recognition. CVPR 2017, Vol. 2017-Janua, 2017, p. 2261–9. <http://dx.doi.org/10.1109/CVPR.2017.243>.

- [17] Wang Y, Long M, Wang J, Gao Z, Yu PS. PredRNN: Recurrent neural networks for predictive learning using spatiotemporal LSTMs. *Adv Neural Inf Process Syst* 2017;2017-Decem(Nips):880–9.
- [18] Chaudhari S, Mithal V, Polatkan G, Ramanath R. An attentive survey of attention models. *ACM Trans Intell Syst Technol* 2021;12(5):53. <http://dx.doi.org/10.1145/3465055>.
- [19] Qu J, Qian Z, Pei Y. Day-ahead hourly photovoltaic power forecasting using attention-based CNN-LSTM neural network embedded with multiple relevant and target variables prediction pattern. *Energy* 2021;232:120996. <http://dx.doi.org/10.1016/j.energy.2021.120996>.
- [20] Lai CS, Zhong C, Pan K, Ng WW, Lai LL. A deep learning based hybrid method for hourly solar radiation forecasting. *Expert Syst Appl* 2021;177:114941. <http://dx.doi.org/10.1016/j.eswa.2021.114941>.
- [21] Yu D, Choi W, Kim M, Liu L. Forecasting day-ahead hourly photovoltaic power generation using convolutional self-attention based long short-term memory. *Energies* 2020;13(15):4017. <http://dx.doi.org/10.3390/EN13154017>, URL: <https://www.mdpi.com/1996-1073/13/15/4017/htm> <https://www.mdpi.com/1996-1073/13/15/4017>.
- [22] Wang Y, Wu H, Zhang J, Gao Z, Wang J, Yu PS, Long M. PredRNN: A recurrent neural network for spatiotemporal predictive learning. *IEEE Trans Pattern Anal Mach Intell* 2022;1–16. <http://dx.doi.org/10.1109/TPAMI.2022.3165153>, URL: <http://arxiv.org/abs/2103.09504>.
- [23] Lin Z, Li M, Zheng Z, Cheng Y, Yuan C. Self-attention ConvLSTM for spatiotemporal prediction. In: *AAAI 2020 - 34th AAAI conference on artificial intelligence*. 2020, p. 11531–8. <http://dx.doi.org/10.1609/aaai.v34i07.6819>.
- [24] Yang D, Kleissl J, Gueymard CA, Pedro HT, Coimbra CF. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. *Sol Energy* 2018;168(October 2017):60–101. <http://dx.doi.org/10.1016/j.solener.2017.11.023>.
- [25] Mayer MJ, Gróf G. Extensive comparison of physical models for photovoltaic power forecasting. *Appl Energy* 2021;283(November 2020). <http://dx.doi.org/10.1016/j.apenergy.2020.116239>.
- [26] Barancsuk L, Groma V, Kocziha B. Hybrid ultra-short term solar irradiation forecasting using resource-efficient multi-step long-short term memory. *Renew Energy* 2025;247:122962. <http://dx.doi.org/10.1016/j.renene.2025.122962>, URL: <https://www.sciencedirect.com/science/article/pii/S096014812500624X>.
- [27] Fouilloy A, Voyant C, Notton G, Motte F, Paoli C, Nivet ML, Guillot E, Duchaud JL. Solar irradiation prediction with machine learning: Forecasting models selection method depending on weather variability. *Energy* 2018;165:620–9. <http://dx.doi.org/10.1016/j.energy.2018.09.116>.
- [28] Visser L, Alskaf T, van Sark W. Operational day-ahead solar power forecasting for aggregated PV systems with a varying spatial distribution. *Renew Energy* 2022;183:267–82. <http://dx.doi.org/10.1016/j.renene.2021.10.102>.
- [29] Kumari P, Toshniwal D. Long short term memory–convolutional neural network based deep hybrid approach for solar irradiance forecasting. *Appl Energy* 2021;295(January):117061. <http://dx.doi.org/10.1016/j.apenergy.2021.117061>.
- [30] Dolara A, Leva S, Manzolini G. Comparison of different physical models for PV power output prediction. *Sol Energy* 2015. <http://dx.doi.org/10.1016/j.solener.2015.06.017>.
- [31] Partio M, Hieta L, Kokkonen A. CloudCast—Total cloud cover nowcasting with machine learning. *Artif Intell Earth Syst* 2025;4(3):e240104. <http://dx.doi.org/10.1175/AIES-D-24-0104.1>, URL: <https://journals.ametsoc.org/view/journals/aies/4/3/AIES-D-24-0104.1.xml>.
- [32] Sobri S, Koohi-Kamali S, Rahim NA. Solar photovoltaic generation forecasting methods: A review. *Energy Convers Manage* 2018;156(December 2017):459–97. <http://dx.doi.org/10.1016/j.enconman.2017.11.019>.
- [33] Zhoolideh Haghighi MH, Ghasriamaneh A, Khosroshahi H. Deep learning-based identification of precipitation clouds from all-sky camera data for observatory safety. *Mach Learn Appl* 2025;20:100640. <http://dx.doi.org/10.1016/j.mlwa.2025.100640>, URL: <https://www.sciencedirect.com/science/article/pii/S2666827025000234>.
- [34] Hou X, Papachristopoulou K, Saint-Drenan Y-M, Kosmopoulos PG, Psiloglou BE, Kazadzis S. Effects of clouds, aerosols and shadows on solar energy potential on urban rooftops using earth observation data and digital surface models. *Appl Energy* 2025;397:126243. <http://dx.doi.org/10.1016/j.apenergy.2025.126243>, URL: <https://www.sciencedirect.com/science/article/pii/S0306261925009730>.
- [35] Zhang J, Verschae R, Nobuhara S, Lalonde JF. Deep photovoltaic nowcasting. *Sol Energy* 2018;176(October):267–76. <http://dx.doi.org/10.1016/j.solener.2018.10.024>.
- [36] Kong W, Jia Y, Dong ZY, Meng K, Chai S. Hybrid approaches based on deep whole-sky-image learning to photovoltaic generation forecasting. *Appl Energy* 2020;280(August):115875. <http://dx.doi.org/10.1016/j.apenergy.2020.115875>.
- [37] Shi X, Chen Z, Wang H, Yeung DY, Wong WK, Woo WC. Convolutional LSTM network: A machine learning approach for precipitation nowcasting. In: *Advances in neural information processing systems*. Vol. 2015-Janua, 2015, p. 802–10.
- [38] Zhao X, Wei H, Wang H, Zhu T, Zhang K. 3D-CNN-based feature extraction of ground-based cloud images for direct normal irradiance prediction. *Sol Energy* 2019;181(February):510–8. <http://dx.doi.org/10.1016/j.solener.2019.01.096>.
- [39] Wang K, Wang T, Shan S, Dou W, Zhang J, Wei H, Zhang K. Multimodal ensemble photovoltaic power forecasting incorporating diffusion-based satellite image prediction. *IEEE Trans Sustain Energy* 2025.
- [40] Guo MH, Xu TX, Liu JJ, Liu ZN, Jiang PT, Mu TJ, Zhang SH, Martin RR, Cheng MM, Hu SM. Attention mechanisms in computer vision: A survey. *Comput Vis Media* 2022;8(3):331–68. <http://dx.doi.org/10.1007/s41095-022-0271-y>.
- [41] Obeso AM, Benois-Pineau J, García Vázquez MS, Acosta AÁR. Visual vs internal attention mechanisms in deep neural networks for image classification and object detection. *Pattern Recognit* 2022;123:108411. <http://dx.doi.org/10.1016/j.patcog.2021.108411>.
- [42] Paletta Q, Arbod G, Lasenby J. Benchmarking of deep learning irradiance forecasting models from sky images – An in-depth analysis. *Sol Energy* 2021;224(June):855–67. <http://dx.doi.org/10.1016/j.solener.2021.05.056>.
- [43] Woo S, Park J, Lee JY, Kweon IS. CBAM: Convolutional block attention module. In: *Proceedings of the European conference on computer vision. ECCV, 2018*, p. 3–19. http://dx.doi.org/10.1007/978-3-030-01234-2_1.
- [44] Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: From error visibility to structural similarity. *IEEE Trans Image Process* 2004;13(4):600–12. <http://dx.doi.org/10.1109/TIP.2003.819861>.
- [45] Perera M, De Hoog J, Bandara K, Senanayake D, Halmuge S. Day-ahead regional solar power forecasting with hierarchical temporal convolutional neural networks using historical power generation and weather data. *Appl Energy* 2024;361:122971.
- [46] Castangia M, Aliberti A, Bottaccioli L, Macii E, Patti E. A compound of feature selection techniques to improve solar radiation forecasting. *Expert Syst Appl* 2021;178:114979. <http://dx.doi.org/10.1016/j.eswa.2021.114979>.
- [47] Svozil D, Kvasnička V, Pospíchal J. Introduction to multi-layer feed-forward neural networks. *Chemometr Intell Lab Syst* 1997;39(1):43–62. [http://dx.doi.org/10.1016/S0169-7439\(97\)00061-0](http://dx.doi.org/10.1016/S0169-7439(97)00061-0).
- [48] du Plessis AA, Strauss JM, Rix AJ. Short-term solar power forecasting: Investigating the ability of deep learning models to capture low-level utility-scale Photovoltaic system behaviour. *Appl Energy* 2021;285(December 2020):116395. <http://dx.doi.org/10.1016/j.apenergy.2020.116395>.
- [49] Hewamalage H, Bergmeir C, Bandara K. Recurrent neural networks for time series forecasting: Current status and future directions. *Int J Forecast* 2021;37(1):388–427. <http://dx.doi.org/10.1016/j.ijforecast.2020.06.008>.
- [50] Liu W, Wang Z, Liu X, Zeng N, Liu Y, Alsaadi FE. A survey of deep neural network architectures and their applications. *Neurocomputing* 2017;234:11–26. <http://dx.doi.org/10.1016/j.neucom.2016.12.038>.
- [51] Mellit A, Pavan AM, Lughi V. Deep learning neural networks for short-term photovoltaic power forecasting. *Renew Energy* 2021;172:276–88. <http://dx.doi.org/10.1016/j.renene.2021.02.166>.
- [52] Bergstra J, Yamins D, Cox D. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. 2013, URL: <https://proceedings.mlr.press/v28/bergstra13.html>.
- [53] Voyant C, Notton G, Kalogirou S, Nivet ML, Paoli C, Motte F, Fouilloy A. Machine learning methods for solar radiation forecasting: A review. *Renew Energy* 2017;105:569–82. <http://dx.doi.org/10.1016/j.renene.2016.12.095>.
- [54] Yang D, Alessandrini S, Antonanzas J, Antonanzas-Torres F, Badescu V, Beyer HG, Blaga R, Boland J, Bright JM, Coimbra CF, David M, Frimane A, Gueymard CA, Hong T, Kay MJ, Killinger S, Kleissl J, Lauret P, Lorenz E, van der Meer D, Paulescu M, Perez R, Perpiñán-Lamigueiro O, Peters IM, Reikard G, Renné D, Saint-Drenan YM, Shuai Y, Urraca R, Verbois H, Vignola F, Voyant C, Zhang J. Verification of deterministic solar forecasts. *Sol Energy* 2020;210(February):20–37. <http://dx.doi.org/10.1016/j.solener.2020.04.019>.