

# 3D-CNN-based feature extraction of ground-based cloud images for direct normal irradiance prediction

Xin Zhao<sup>a</sup>, Haikun Wei<sup>a,\*</sup>, Hai Wang<sup>b</sup>, Tingting Zhu<sup>a</sup>, Kanjian Zhang<sup>a</sup>

<sup>a</sup> Key Laboratory of Measurement and Control of CSE, Ministry of Education, School of Automation, Southeast University, Nanjing 210096, PR China

<sup>b</sup> Sobey School of Business, Saint Mary's University, Halifax, Canada

## ARTICLE INFO

### Keywords:

Direct normal irradiance  
3D convolutional neural network  
Feature extraction  
Ground-based cloud image

## ABSTRACT

Cloud cover and cloud motion have a large impact on solar irradiance. One of the effective ways for direct normal irradiance (DNI) prediction is to use cloud features, which has been extensively studied. A Convolutional Neural Network (CNN) has the advantage of automatic features extraction by using strong computing capabilities. In this paper, a novel 3D-CNN method is proposed by processing multiple consecutive ground-based cloud (GBC) images in order to extract cloud features including texture and temporal information. The resulting features and the DNI data are then used to establish a DNI forecasting model. The experiments are carried out to evaluate the performance of the proposed forecasting method by using the data from January 1, 2013 to December 31, 2014. The experimental results show that the proposed method coupled with the multilayer perceptron (MLP) model achieves forecast skill of 17.06% for 10-minute ahead DNI prediction.

## 1. Introduction

With large-scale solar power generation grid integration, the randomness and fluctuation of the output power have affected the security, stability and the economic operation of the power grid (Martín et al., 2010; Singh, 2013). The accurate prediction of the output of solar power generation can provide important decision support to power dispatching, and can effectively reduce the operation cost of the power system (Bacher et al., 2009; Su et al., 2012). The surface solar irradiance has a direct influence on the output of solar power generation. Direct normal irradiance (DNI) is the dominant factor for predicting the output of concentrating solar power (CSP). In order to obtain high CSP forecasting accuracy, an estimation of the local DNI is usually necessary (Behar et al., 2015; Chu and Coimbra, 2017).

In recent years, many studies have been published on cloud images for DNI prediction, including cloud identification, cloud classification and cloud motion. The partly prior literature for DNI prediction based on GBC images used cloud transmittance combined with cloud motion vector (CMV) method (Marquez and Coimbra, 2013; Quesada-Ruiz et al., 2014). Cloud identification was the first step, followed by calculation of cloud transmittance coefficients or clear-sky indexes (Chu et al., 2013). Then, multiple consecutive images were used for cloud velocity derivation. The common velocity derivation methods include: the Scale-invariant feature transform (SIFT), X-correlation (X-corr),

Optical flow (OF) and Particle image velocimetry (PIV) (Li et al., 2016). Due to the rapid changes and diversity of clouds, the cloud tracking is highly dependent on the specific cloud image. It is difficult to know what kind of features and tracking methods are suitable for a GBC image. Especially for the fast and complex motion, the motion detection techniques have difficulty in tracking the clouds.

Other prior literature focused on the image-based feature extraction for DNI prediction, which showed that using image-based features could improve forecasting performance (Fu and Cheng, 2013; Cheng et al., 2014; Cheng and Yu, 2015). In traditional pattern recognition, features are firstly obtained by various feature extraction techniques and then identified by a classification technique. The difference between deep learning and traditional pattern recognition methods is that the features are learnt from large data automatically rather than manually. Deep learning-based classifiers can learn features and achieve better accuracy (Shaheen and Verma, 2017).

The convolutional neural network (CNN) is one of the most representative deep learning methods for image classification (Schmidhuber, 2015; Krizhevsky et al., 2012; Li et al., 2017). As a state-of-the-art method, it has been successfully applied to GBC images classification (Shi et al., 2017; Ye et al., 2017). Shi et al. (2017) used the deep CNN features for GBC image classification. The results showed that the cloud classification accuracy of CNN improved significantly, further demonstrating the superiority of CNN over hand-engineered

\* Corresponding author.

E-mail address: [hkwei@seu.edu.cn](mailto:hkwei@seu.edu.cn) (H. Wei).

<https://doi.org/10.1016/j.solener.2019.01.096>

Received 11 September 2018; Received in revised form 27 January 2019; Accepted 29 January 2019  
0038-092X/ © 2019 International Solar Energy Society. Published by Elsevier Ltd. All rights reserved.

features for cloud classification. Recently, Ye et al. (2017) adopted CNN for cloud feature extraction. They used the fisher vector encoding for high-dimensional feature mapping and spatial feature aggregation. They demonstrated that CNN consistently outperformed traditional methods, especially for the fine-grained categorization cases.

The traditional CNN is mainly used to process static image. In this study, the CNN model is used to automatically extract features of GBC images for DNI prediction. Because cloud motion is one of the main influence factors for DNI prediction, it is necessary to consider the cloud motion when establishing a CNN model. This requires selecting multiple consecutive GBC images for analysis. The CNN frame with a 3D kernel (3D-CNN) has been proved that it can extract features from both spatial and temporal dimensions (Ji et al., 2010; Tran et al., 2014). Therefore 3D-CNN frame is suitable for automatic feature extraction of multiple consecutive GBC images with texture and temporal information at the same time. In order to make the automatic extracted features useful for DNI prediction, the classification category is by discretizing the attenuation rate of DNI into  $n$  levels, which is different from the common cloud shape classification approaches (Zhu et al., 2016). The final forecasting model with the inputs of GBC image features and DNI sequence is used for intra-hour DNI prediction. In this paper, the application of CNN architecture to DNI prediction is the first time. In order to verify the validity of the proposed method, experiments are conducted using two-year DNI data from the National Renewable Energy Laboratory (NREL). The main contributions of the paper include: (1) a 3D-CNN-based approach for automatically extracting features of consecutive GBC images, which are suitable for DNI prediction; (2) DNI forecasting models by combining GBC image features with sequence data.

The remainder of this paper is organized as follows. Section 2 presents the 3D-CNN-based approach for extracting features of GBC images suitable for DNI prediction. Section 3 describes the experiments for evaluating the proposed method. Section 4 discusses the results and performance analysis. Finally, Section 5 summarizes the conclusions.

## 2. 3D-CNN approach to GBC image

### 2.1. Clear-sky indexes

In this study, the clear-sky indexes are used as the classification criterion for GBC image feature extraction. The clear-sky index is defined as the ratio between the measured DNI and the estimated theoretical clear-sky DNI as follow (Engerer and Mills, 2014)

$$k = \frac{I_m}{I_{cs}}, \quad (1)$$

where  $I_m$  is the measured DNI value,  $I_{cs}$  is the theoretical clear-sky DNI.

The Linke formulation with turbidity coefficient is used to calculate theoretical clear-sky DNI  $I_{cs}$  (Ineichen and Perez, 2002)

$$I_{cs} = (0.664 + 0.163 \cdot \exp(h/8000)) \cdot I_0 \cdot \exp(-0.09 \cdot m \cdot (T_L - 1)) \quad (2)$$

where  $T_L$  is the Linke turbidity coefficient,  $h$  is the altitude of the power plant,  $I_0$  is the extraterrestrial solar irradiance,  $m$  is the air mass. The atmospheric turbidity coefficient ( $T_L$ ) is calculated by inverting Eq. (2) as follows

$$T_L = 1 + \left( \ln \left( \frac{(0.664 + 0.163 \cdot \exp(h/8000)) \cdot I_0}{I_{cs}} \cdot \frac{1}{0.09 \cdot m} \right) \right). \quad (3)$$

Fig. 1 shows the profiles of DNI and  $T_L$  during a clear-sky. In the 9:00–15:00 of the local time per day, the  $T_L$  is almost unchanged. Therefore, the mean of  $T_L$  in the six center hours, named  $T_{LD}$ , is used to calculate the  $I_{cs}$ . Due to the long time span of experimental data, the theoretical clear-sky DNI fitted with a fixed  $T_{LD}$  is not ideal. The 48 days of clear-sky DNI data selected manually from the experimental dataset are used for  $T_{LD}$  dynamic adjustment. The theoretical clear-sky DNI  $I_{cs}$

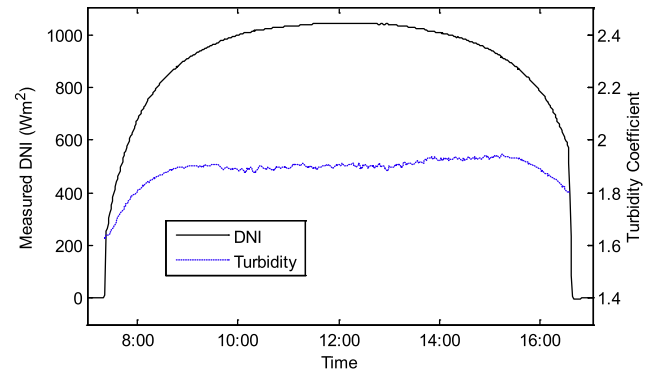


Fig. 1. The DNI and atmospheric turbidity coefficient under a whole clear-sky weather type. The black line is the measured DNI value. The blue dotted line is the  $T_L$  calculated by formula. The  $T_L$  is almost unchanged in the 9:00–15:00.

per day is calculated by using the  $T_{LD}$  of the latest day.

### 2.2. 2D convolution and 3D convolution

Compared with traditional image classification algorithms, CNN requires little pre-processing (Krizhevsky et al., 2012; Lawrence et al., 1997). One of the major advantages of CNN is that prior knowledge and manual intervention are not required (Karpathy et al., 2014). The hidden layers of a CNN consist of convolutional layers, pooling layers, normalization layers and fully-connected layers. The process of convolution computation is also called filter and the essence is to extract features of different frequency bands (Simonyan and Zisserman, 2014). The data of 2D convolution at the point  $(x, y)$  on the  $j$ th feature map in the  $i$ th layer is denoted as

$$v_{ij}^{xy} = f \left( b_{ij} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} w_{ij}^{pq} v_{(i-1)m}^{(x+p)(y+q)} \right), \quad (4)$$

where  $f(\cdot)$  is the activation function,  $b_{ij}$  is the bias,  $m$  indexes over the set of feature maps in the  $(i-1)$ th layer connected to the current feature map,  $P_i$  and  $Q_i$  are the height and width of the kernel, respectively, and  $w_{ij}^{pq}$  is the value at the position  $(p, q)$  of the kernel connected to the  $j$ th feature map.

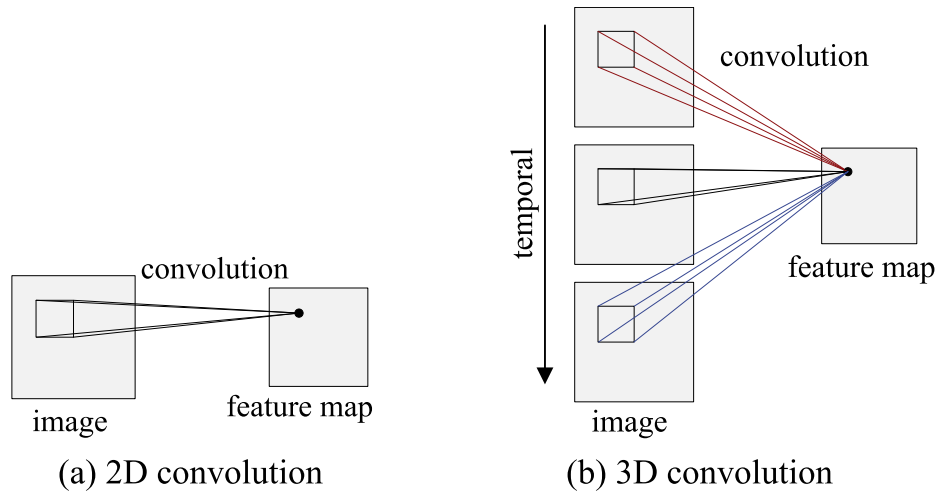
The 3D convolution approach is applied across the temporal dimension by using a 3D kernel to convolve multiple consecutive images together. By this construction, the first convolutional layer is connected to multiple consecutive images, so that the local motion information can be obtained. The data at a point  $(x, y, z)$  on the  $j$ th feature map in the  $i$ th layer is given by

$$v_{ij}^{xyz} = f \left( b_{ij} + \sum_m \sum_{p=0}^{P_i-1} \sum_{q=0}^{Q_i-1} \sum_{r=0}^{R_i-1} w_{ij}^{pqr} v_{(i-1)m}^{(x+p)(y+q)(z+r)} \right), \quad (5)$$

where  $R_i$  is the size of the 3D kernel along the temporal dimension, and  $w_{ij}^{pqr}$  is the value at the position  $(p, q, r)$  of the kernel connected to the  $j$ th feature map. A comparison of 2D and 3D convolution is illustrated in Fig. 2.

### 2.3. 3D-CNN architecture

Based on 3D convolution, a 3D-CNN is used to extract features of GBC images. The 3D-CNN is implemented by extending the filters on the first convolutional layer to be  $M \times M \times N \times T$ , where  $M \times M \times N$  is the spatial dimension ( $M$  is filter window size and  $N$  is image channel) and  $T$  is the temporal dimension extent (Tran et al., 2014). The GBC images are RGB images and therefore the image channel  $N$  is 3. For action recognition in videos, the 3D-CNN method takes all video frames as inputs and does not rely on any preprocessing. However, when using 3D-CNN to extract GBC image features, it needs to



**Fig. 2.** Comparison of (a) 2D and (b) 3D convolution. The feature map is obtained by convoluting the local region at the same position in the previous layer. The 2D convolution handles static images. The 3D convolution handles multiple consecutive images simultaneously to extract motion features.

determine the temporal dimension  $T$  which is similar to frames in a video segment. The multiple consecutive cloud images in a fixed time dimension are considered as a video segment.

The 10-fold cross-validation is performed to determine the dimension  $T$ . Each model with different dimension  $T$  has an average cross-validation accuracy. The model with the highest cross-validation accuracy is the best and the corresponding  $T$  is desired. The average cross-validation accuracy is shown in Table 1. NREL public database provides high-quality DNI data and GBC images with a sampling interval of 10 min. In this paper, the data from the NREL database are used for experiments. Our 3D-CNN approach is a general method for features extraction of GBC images, and may be used for multiple consecutive images at any short-term interval.

The fusion of consecutive images enables the network to accurately detect the motion information. To obtain different types of features from the GBC images, multiple 3D convolutions with different kernels are applied to the same location in the input layer. All the weights of convolutional connections are not shared, the feature extraction by multiple 3D convolutions are shown in Fig. 3.

The number of valid samples is approximately 13,000, which is insufficient to support the adequate training of parameters in a deep network structure. We refer to the CNN model (Ji et al., 2010) with similar data magnitude and classification categories, and finally the network structure is selected after many trial runs. This architecture consists of 3 convolution layers, 2 pooling layers, and 1 fully-connected layer. The configurations of the 3D-CNN model are outlined in Table 2. All layers are initialized with the Xavier approach (Glorot and Bengio, 2010). The activation function is the ReLU function and the loss function uses cross entropy. The model is trained by using stochastic gradient descent (SGD) with an initial learning rate of  $1e-4$ , momentum of 0.9 and mini-batch size of 128 examples. The experiment is performed on the PC with Intel Core E5-2620 v4 CPU @ 2.10 GHz.

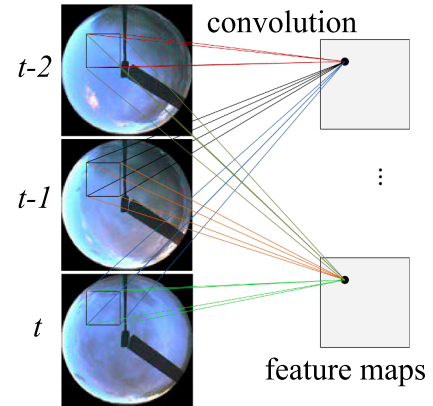
#### 2.4. DNI forecasting model

In this study, the classification criterion of 3D-CNN model is based on the clear-sky index to be predicted (DNI attenuation). The clear-sky

**Table 1**

Average cross-validation accuracy with different dimension  $T$ . The parameter  $T$  corresponding to the highest accuracy is the desired.

$T$	2	3	4	5	6
Accuracy (%)	62.2	68.7	66.9	63.2	59.3



**Fig. 3.** Feature extraction by multiple 3D convolutions in this paper. The feature maps are obtained by implementing multiple 3D convolutions in the input layer. The 3D convolution is applied to handle multiple images simultaneously to extract motion features.

index  $k$  (calculated by Eq. (1)) is divided into 5 categories, including completely overcast and rainy ( $k \in [0, 0.1]$ ), strong attenuation ( $k \in [0.1, 0.4]$ ), medium attenuation ( $k \in [0.4, 0.7]$ ), weak attenuation ( $k \in [0.7, 0.95]$ ) and clear-sky ( $k$  is greater than or equal to 0.95). Class 1 through Class 5 correspond to completely overcast and rainy, strong attenuation, medium attenuation, weak attenuation and clear-sky (Martínez-Chico et al., 2011). The cloud images at time  $t-2$ ,  $t-1$ ,  $t$  are used as inputs, and the class of  $k(t+i)$  is as the classification output. The 3D-CNN classification model has established the mapping between the images at previous moments and the predicted DNI. Therefore, the automatically extracted GBC image features are related to the DNI to be predicted.

The final forecasting model with the input of fully-connected features and DNI data is used for DNI predication. The Bayesian information criterion (BIC) (Burnham and Anderson, 2004) for the DNI training set is used to determine the input dimension of forecasting model. The input dimension is chosen to be 3, as shown in Fig. 4.

Typical linear autoregressive model (AR) (Dambreville et al., 2014) and nonlinear multilayer perceptron model (MLP) (Azimi et al., 2016) are selected as forecasting models. Date normalization is carried out before training and testing. For the AR model, the least squares method is used to obtain the model parameters. For the MLP model, the MLP structure is with two hidden layers consisting of 30 neurons respectively. The proposed DNI forecasting model is shown in Fig. 5.

**Table 2**

The architecture of the 3D-CNN model. The Conv1 is 3D convolution. The pooling operator is max pooling.

Layer name	Conv1	Pool2	Conv3	Pool4	Conv5	Fc6
Kernel	$7 \times 7 \times 3 \times 3$	$3 \times 3$	$5 \times 5 \times 16$	$3 \times 3$	$5 \times 5 \times 16$	$1 \times 1 \times 10$
Num. filters	16	–	16	–	10	5
stride	3	2	3	2	1	1

### 3. Experimental data and preparation

#### 3.1. Data sets

In this study, all the data are from the NREL database. The GBC images are taken by a Total Sky Imager (TSI-880), and the DNI values are measured by a Kipp & Zonen pyrliometer mounted on a Sci-Tec tracer (Zhu et al., 2017). The image resolution is  $352 \times 288$  pixels and the GBC images in different weather conditions are shown in Fig. 6. The shadow band in the image is used to protect the CCD sensor from the direct sun. When the shadow band malfunctions, the pointer is inaccurate, and the area around the sun is overexposed.

The GBC images and DNI data from January 1, 2013 to December 31, 2014 are used for the experiment, exclusive of the fault moment of the shadow band. The model is trained with the data in 2013. The 10% data randomly selected from 2014 are used as validation and the remaining used to test the proposed method. Fig. 7 shows the distribution of the DNI in a week.

#### 3.2. Image preprocessing

The image processing procedures are used to extract the regions of interest (ROI) in the GBC images, so that the shadow band can be discarded. The masks are matrices with binary elements corresponding to locations of the image. The element values are used to indicate true or false pixel values. Regions indicated as true are the ROI and only these pixels are used as the input of the 3D-CNN model. The binary masks are created by a circle with a radius of 125 pixels and centered at the center pixel of the picture. The shadow band is not included in the ROI. The original image resolution is  $352 \times 288$  pixels, the image size after extraction ROI is  $251 \times 251$  pixels. The images are shown in Fig. 8.

#### 3.3. Image augmentation

In training set, the GBC image samples in Class 2–4 are relatively few. Image augmentation is used to overcome the sample unbalance. Each GBC image is specificity with different sun's position and cloud density, which corresponds to different irradiance. Therefore, only the contrast transformation is used to augment the training set. The adjustment maps the values in initial GBC image to new values by Eq. (7) and the same mapping is applied for each color channel, which is as follow

$$x_{new} = \begin{cases} 0 & x < t_l \times 255 \\ \frac{1}{t_h - t_l}(x - t_l \times 255) & t_l \times 255 < x < t_h \times 255 \\ 255 & x > t_h \times 255 \end{cases} \quad (7)$$

where  $x$  is the values in initial GBC image,  $x_{new}$  is the values in augmented images,  $t_l$  and  $t_h$  correspond to the low threshold and high threshold with values of 0.15 and 0.85.

This adjustment increases the contrast of the new image. The partly samples in Class 2–4 are performed this operation. Fig. 9 shows one original image is processed by preprocessing and contrast enhancement.

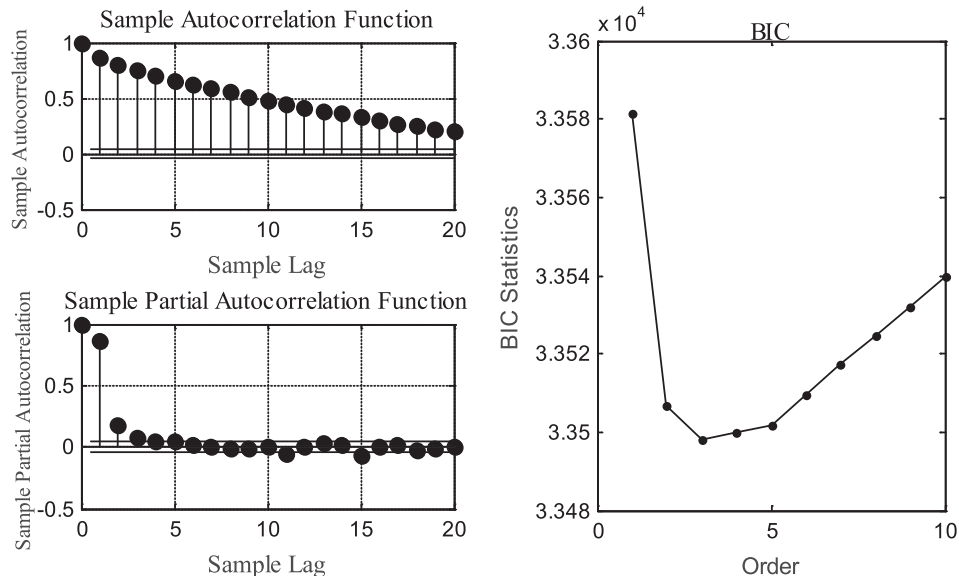
#### 3.4. Forecasting performance evaluation

The models are evaluated synthetically using the following evaluation criteria:

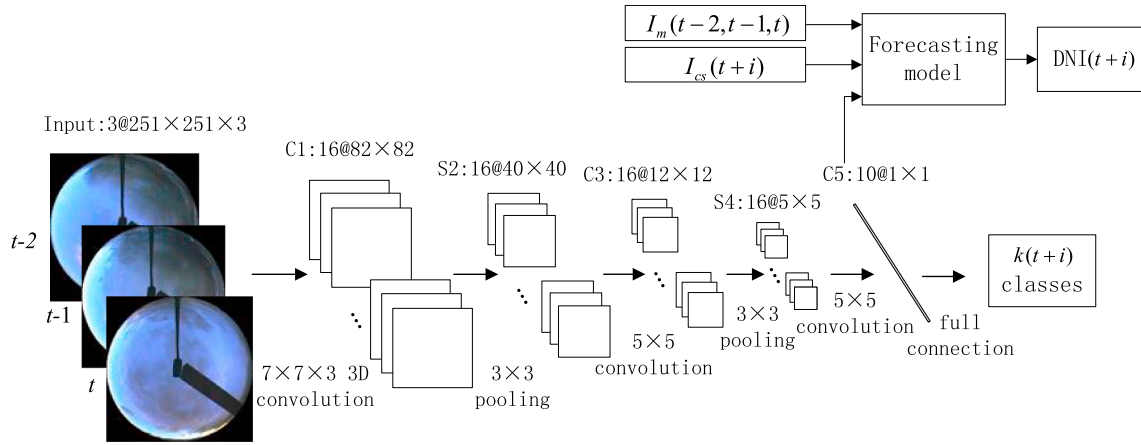
- (1) Normalized root mean square error (nRMSE)

$$\text{nRMSE} = \frac{1}{\sqrt{N}} \times \sqrt{\sum_{t=1}^N (I_t - \hat{I}_t)^2 / I} \times 100\%, \quad (8)$$

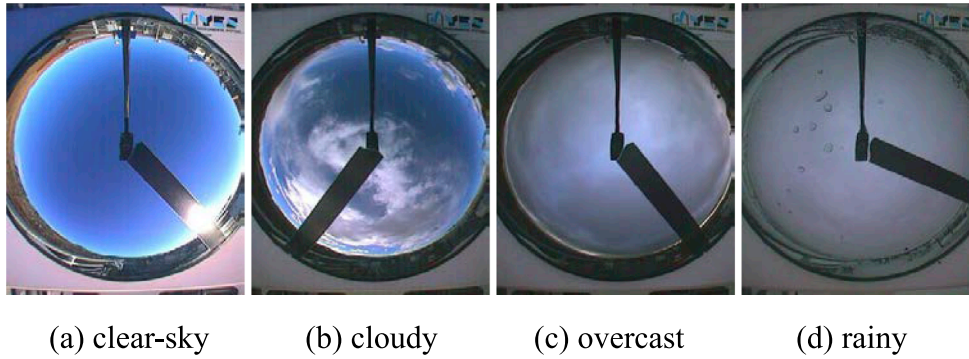
- (2) Normalized mean absolute error (nMAE)



**Fig. 4.** The ACF, PACF, and BIC of DNI sequence. The ACF decays slow and the PACF decays fast to essentially zero. The order corresponding to the minimum BIC statistic value is selected as the optimal model input dimension.



**Fig. 5.** The DNI forecasting model. The 3D-CNN architecture consists of 3 convolution layers, 2 downsampling layers, and 1 fully-connected layer. The features in fully-connected layer and DNI value are used as forecasting model inputs.



**Fig. 6.** The GBC images in NREL database under different weather condition (a) clear-sky (b) cloudy (c) overcast (d) rainy.

$$\text{nMAE} = \frac{1}{N} \times \sum_{t=1}^N |I_t - \hat{I}_t| / \bar{I} \times 100\%, \quad (9)$$

(3) Normalized mean bias error (nMBE)

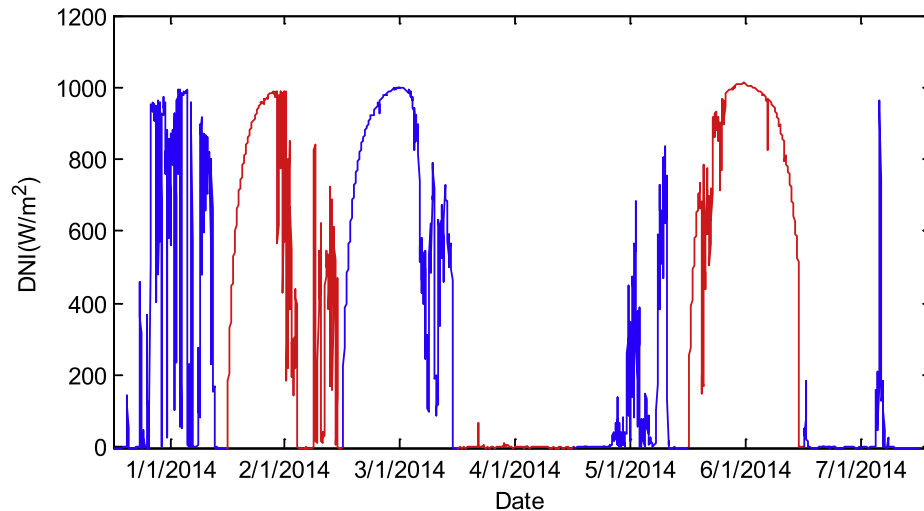
$$\text{nMBE} = \frac{1}{N} \times \sum_{t=1}^N (I_t - \hat{I}_t) / \bar{I} \times 100\%, \quad (10)$$

(4) Mean absolute percentage error (MAPE)

$$\text{MAPE} = \frac{1}{N} \times \sum_{t=1}^N \left| \frac{I_t - \hat{I}_t}{I_t} \right|, \quad (11)$$

where  $I_t$  and  $\hat{I}_t$  are the measured and predicted DNI at time  $t$ ,  $\bar{I}$  is the mean of all the testing measured DNI ( $607.88 \text{ W/m}^2$ ),  $N$  is the number of test data.

The lower the value of nRMSE, nMAE and MAPE are, the better the prediction is. In addition, the forecast skill  $s$ , which indicates the forecasting improvement against the persistence model (constant clear-sky index) is computed as



**Fig. 7.** The distribution of the DNI in a week. The DNI values of consecutive two days use different colors.



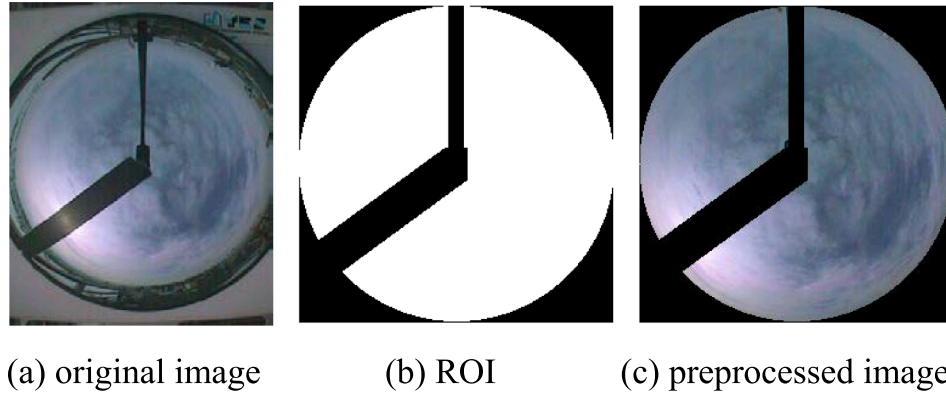


Fig. 8. The GBC image preprocessing uses a binary mask to extract the regions of interest. Figure (c) is obtained by multiplying the region of interest figure (b) and original figure (a).

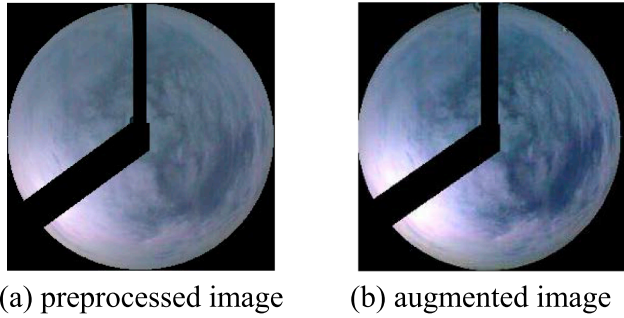


Fig. 9. The GBC image augmentation uses contrast transformation. Figure (a) is the preprocessed image, the contrast transformation is used to increase the preprocessed image contrast as the augmented image figure (b).

$$s = \left( 1 - \frac{\text{nRMSE}_f}{\text{nRMSE}_p} \right) \times 100\%, \quad (12)$$

where  $\text{nRMSE}_f$  and  $\text{nRMSE}_p$  are the nRMSE for any forecasting model and persistence model.

#### 4. Results and discussions

The proposed method uses automatically extracted features from the multiple consecutive GBC images by 3D-CNN for DNI prediction. The clear-sky index (DNI attenuation) to be predicted is used as the category of 3D-CNN classification model. Through the supervised training model, the automatically extracted features are related to the DNI to be predicted. The better the classification results are, the stronger the correlation between the extracted features and the DNI predictions is. In order to prevent over-fitting, the 10% data in 2014 is used as the validation set. Fig. 10 shows the loss and accuracy learning curves of training and validation set. It can be seen that the validation loss curve has risen after 32 epoch and the validation accuracy has declined at the same time. Therefore, the training model at 32 epoch is selected as the final model.

The handcrafted features for cloud classification and DNI prediction proposed by Cheng and Yu (2015) are used to compare with the 3D-CNN automatically extracted features. The classification results of different methods for testing set are summarized in Table 3. The Top1 and Top2 accuracy of each class are given by

$$\text{Accuracy}_{\text{Top}i}^j = \frac{\text{Cor}_{ij}}{\text{All}_j}, \quad i = 1, 2; \quad j = 1, 2, \dots, 5 \quad (13)$$

where  $\text{All}_j$  indicates the number of testing samples of Class  $j$ ,  $\text{Cor}_{ij}$  is the number of that the target label  $j$  is one of the top  $i$  predictions (the  $i$  ones with the highest probabilities) in testing samples of Class  $j$ .

For the Class 1 (completely overcast and rainy condition), the clear-sky index is  $k \in [0, 0.1)$ , which means the clouds are thick, shading, and the brightness is relatively low. For the Class 5 (clear-sky condition), the clear-sky index is greater than or equal to 0.95, which means there is translucent or no cloud. These two classes can be well recognized and the probability of mutation is small. Hence, the classification results in Class 1 and Class 5 are better. The Top1 average accuracy of 3D-CNN model is 67.0%, and the Top2 average accuracy is 82.6%. Compared to the handcrafted features proposed by Cheng and Yu (2015), the performance of 3D-CNN-based features is better, especially in Class 2–4. Overall, the 3D-CNN model has a good performance on feature extraction for DNI attenuation classification.

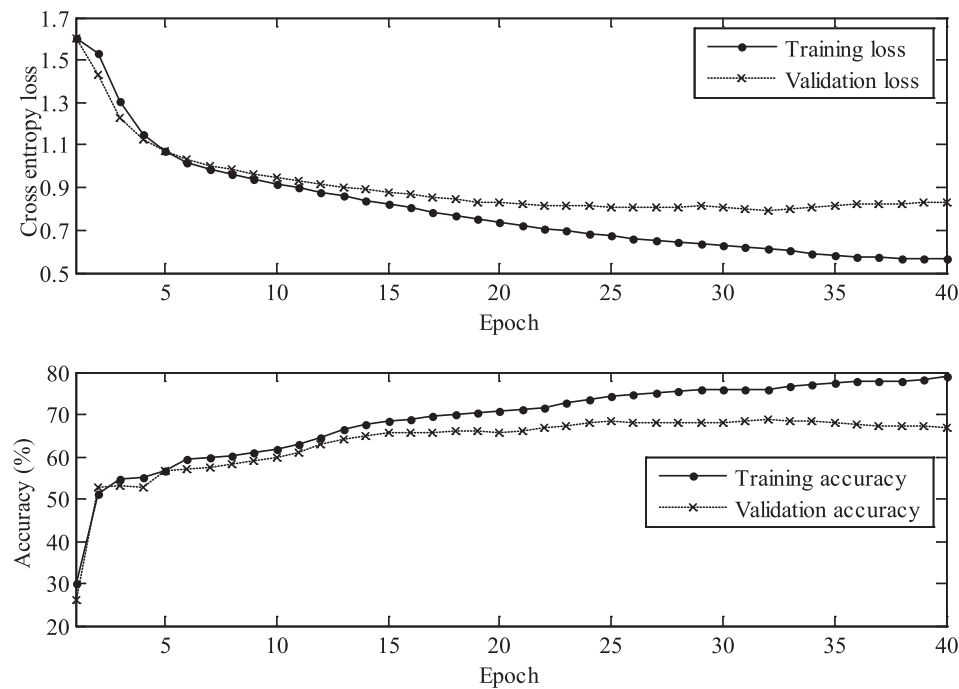
To evaluate the performance of the proposed 3D-CNN-based features extraction method for predicting DNI, the two groups of experiments were carried out by using: (1) only historical measured DNI time series and (2) historical measured DNI, theoretical clear-sky DNI and extracted GBC images features as inputs. Table 4 shows the results for the testing set with different methods and inputs.

From Table 4, it can be seen that the forecasting model with the input of GBC image features and DNI outperforms the forecasting model with the input of only historical DNI. For different forecasting models, the non-linear MLP is better than the linear AR. The improvement of MLP outperforms that of AR by adding GBC image features.

The manually extracted features by Cheng and Yu (2015) (represented as 'M') are used to quantitatively compare with the automatically extracted features by the proposed method (represented as 'A'). After adding the GBC images features to the input, the model performance is significantly improved. The forecasting results of AR-A are close to the AR-M results. However, MLP-A performance is better than MLP-M. The MLP model coupled the automatically extracted GBC images features with DNI data has the best forecasting results with nRMSE of 30.82% and nMAE of 17.91%, the forecast skill is 17.06%. Fig. 11 shows the nRMSE of proposed MLP-A method and persistence method with different forecast horizon. The MLP-A method always achieves lower nRMSE than the Persistence forecast.

For the clear-sky and rainy days, there are almost no changes in the cloud. Therefore, 127 days under mostly broken-sky condition are selected to further verify the proposed algorithm. The selection criterion is that broken-sky condition (including strong attenuation, medium attenuation and weak attenuation) occupies more than half of the effective time of a day. Table 5 describes the DNI forecasting results for 127 broken-sky days.

Compared with Table 4, the forecasting results achieve a little bit worse performance. For the clear-sky and rainy days, all the forecasting models have small errors. Since the clear-sky and rainy data are removed from the annual data, the DNI forecasting results will become worse than the results of the whole data. Similar to the results in Table 4, the forecasting models are better after adding GBC image



**Fig. 10.** The loss and accuracy learning curves of training and validation set. The validation set is used to avoid over-fitting. At the 32 epoch, the validation results are the best and the training model at present are selected as the final model.

**Table 3**

The classification accuracy (%) of different methods. *Class 1* through *Class 5* correspond to completely overcast and rainy, strong attenuation, medium attenuation, weak attenuation and clear-sky. Top1 accuracy is the correct rate of classification. Top2 accuracy is the correct rate of the correct label among the two labels considered most probable by the model.

	Accuracy	Class 1	Class 2	Class 3	Class 4	Class 5	Mean
3D-CNN	Top1	78.2	63.8	56.8	56.1	80.2	67.0
	Top2	86.3	78.2	76.4	75.6	90.0	82.6
Cheng and Yu (2015)	Top1	73.6	54.3	51.4	50.3	81.2	63.1
	Top2	79.3	62.6	61.3	61.7	88.4	73.5

features. The MLP-A model still has the best forecasting results with nRMSE of 40.15% and nMAE of 26.10%. It has more improvement of forecasting skill for broken-sky days. The forecasting skill of AR-A and MLP-A are 18.28% and 28.89%, respectively, and the effect is improved significantly. Therefore, comparing the results of Tables 4 and 5, the 3D-CNN-based feature extraction is suitable to follow the DNI changes.

In order to visually compare the DNI forecasting results, time-series plots of the measured values, the persistence results, and the MLP-A results on January 1, 2014 are shown in Fig. 12.

Fig. 13 intuitively shows the forecast skills of all forecasting results. It is clear that the best forecasting model is MLP-A for whole days or

broken-sky days.

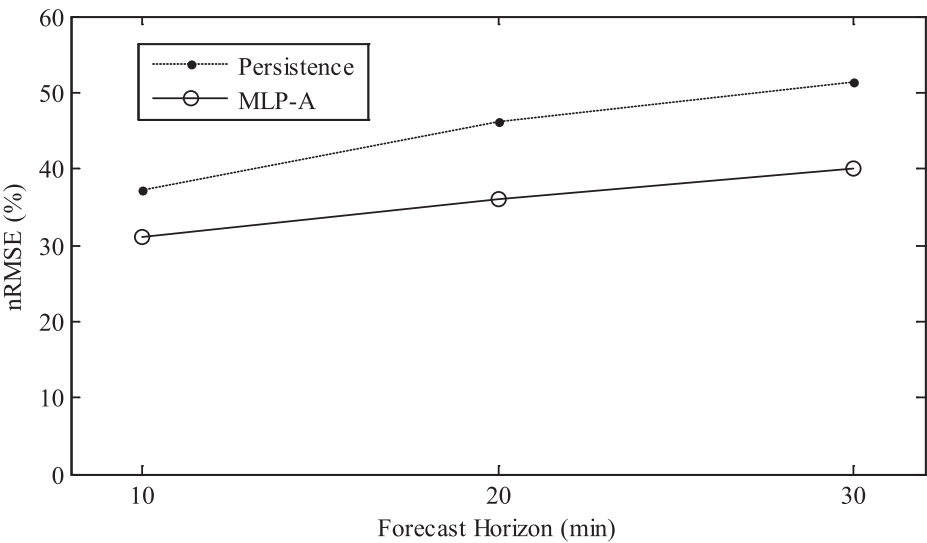
The forecast skill of the MLP-A for all sky conditions and broken-sky conditions is 17.06% and 28.89%. Some prior literature focused on the ground-based CMV also achieved significant improvement over the persistence model. Marquez and Coimbra (2013) obtained the improvement between 20 and 40% for 3 to 15 min forecast horizons by using PIV to calculate the cloud velocity. Chu et al. (2013) combined cloud motion velocity with ANN to predict DNI. The forecast skill for both 5 and 10 min forecast horizons on RMSE is about 20%. Bone et al. (2018) used adaptive clear-sky model and cloud tracking for DNI prediction and had 13% improvement when taking the average nRMSE values across 1–15 min forecasting horizons for days with an average cloud fraction of 0.2 or higher. The research showed that the performance of ground-based CMVs is better than persistence under broken-sky conditions and worse than persistence when sky conditions had low variability (clear-sky or rainy conditions) (Law et al., 2014). Compared with the prediction results of the prior works, it can be seen that the proposed work is competitive.

By quantitative comparison, the automatically extracted features by 3D-CNN are more effective than the handcrafted ones for DNI prediction. Moreover, the proposed method enables generalizations to multiple sites easily by adjusting model structure.

**Table 4**

Results of persistence method, basic method, present method (Cheng and Yu, 2015) and proposed novel method for 10-minute ahead (one-step ahead) DNI prediction. The input types include the measured DNI, theoretical clear-sky DNI and image features. The images features of present method are compared to that of the proposed method.

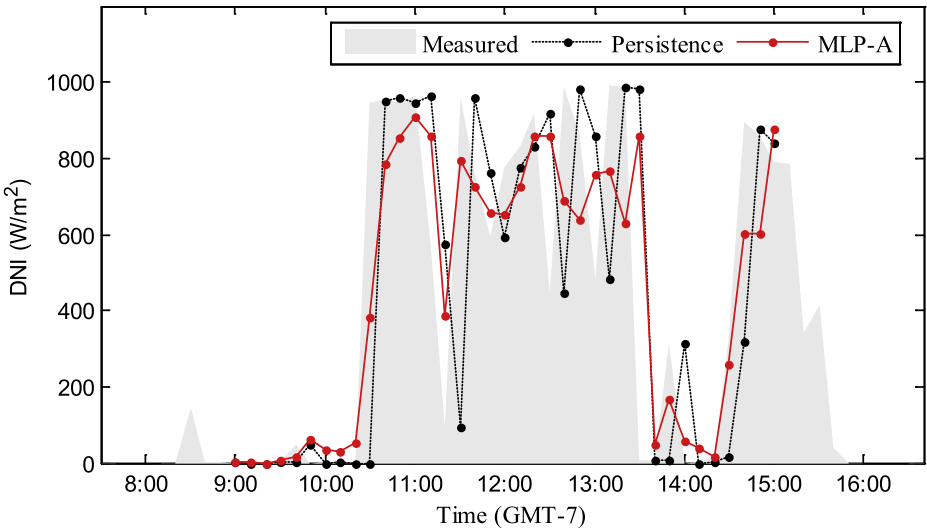
Inputs	Models	nRMSE (%)	nMAE (%)	nMBE (%)	MAPE (%)	s (%)
DNI	Persistence	37.16	18.73	−0.14	38.34	–
	AR	36.20	22.56	−0.86	47.17	2.58
	MLP	34.81	21.07	0.62	40.80	6.32
Cheng and Yu (2015)	AR-M	34.96	22.03	−0.75	46.39	5.92
	MLP-M	33.23	19.96	−0.09	44.28	10.58
Proposed	AR-A	34.97	21.98	0.65	44.21	5.89
	MLP-A	30.82	17.91	0.21	37.80	17.06



**Fig. 11.** DNI forecast nRMSE of proposed MLP-A method and persistence method with different forecast horizon. The temporal resolution is 10 min. The figure shows one-step ahead, two-step ahead and three-step ahead forecast (10-minute, 20-minute and 30-minute forecast horizon).

**Table 5**  
The 10-minute ahead (one-step ahead) DNI forecasting results of persistence method, basic method and proposed method for 127 selected broken-sky days. The DNI changes are obviously in the selected broken-sky days.

Inputs	Models	nRMSE (%)	nMAE (%)	nMBE (%)	MAPE (%)	s (%)
DNI	Persistence	56.46	33.37	−0.18	41.57	–
	AR	52.67	31.65	−0.61	41.24	6.71
	MLP	51.13	32.09	0.39	44.88	9.44
Cheng and Yu (2015)	AR-M	48.59	30.49	−0.57	36.40	13.94
	MLP-M	43.78	29.26	−0.13	36.87	22.46
Proposed	AR-A	46.14	31.84	−0.11	34.83	18.28
	MLP-A	40.15	26.10	0.26	35.22	28.89



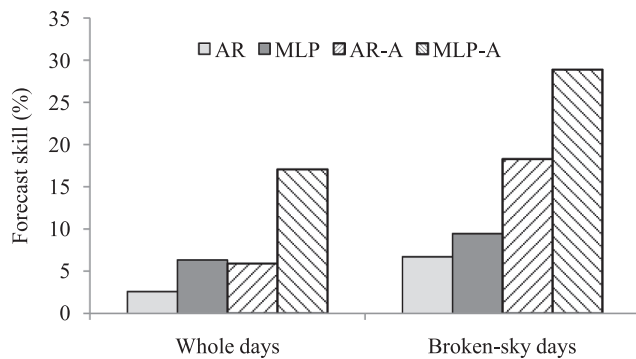
**Fig. 12.** DNI forecasting results for 1/1/2014. The nRMSE for this period are 58.31% (Persistence) and 44.01% (MLP-A).

5. Conclusion

In this study, a CNN-based feature extraction of GBC images for 10–30 min ahead DNI prediction was proposed. The 3D-CNN model was first used to establish the relationship between GBC images and the clear-sky indexes for forecasting DNI. The features in the last fully-connected layer were as one part inputs of the DNI forecasting models. Multiple input variables, including historical measured DNI, theoretical clear-sky DNI and extracted image features, were employed as the

forecasting model inputs. Data from the NREL dataset in the whole year of 2013 were used to train the forecasting model, while the 2014 data were used for testing. Experimental results showed that the 3D-CNN could automatically extract features of GBC images and the 3D-CNN-based features were useful for forecasting DNI. The MLP model coupled with the 3D-CNN-based features had the best forecasting accuracy for different weather types, achieving 17.06% and 28.89% improvement on nRMSE over the persistent model, respectively.





**Fig. 13.** Forecast skills over persistence of different methods in whole days and broken-sky days. The order of the bars: AR = far left, MLP = middle left, AR-A = middle right, MLP-A = far right.

## Acknowledgment

This work was supported by the National Key Research and Development Program of China (Grant No. 2018YFB1500800), and by the National Natural Science Foundation of China (Grant No. 61773118, Grant No. 61703100 and Grant No. 61374006).

## References

Azimi, R., Ghayekhloo, M., Ghofrani, M., 2016. A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting. *Energy Convers. Manage.* 118, 331–344.

Bacher, P., Madsen, H., Nielsen, H.A., 2009. Online short-term solar power forecasting. *Sol. Energy* 83, 1772–1783.

Behar, O., Khellaf, A., Mohammedi, K., 2015. Comparison of solar radiation models and their validation under Algerian climate – The case of direct irradiance. *Energy Convers. Manage.* 98, 236–251.

Bone, V., Pidgeon, J., Kearney, M., Veeraragavan, A., 2018. Intra-hour direct normal irradiance forecasting through adaptive clear-sky modelling and cloud tracking. *Sol. Energy* 159, 852–867.

Burnham, K.P., Anderson, D.R., 2004. Multimodel inference understanding AIC and BIC in model selection. *Sociological Methods Res.* 33, 261–304.

Cheng, H., Yu, C., Lin, S., 2014. Bi-model short-term solar irradiance prediction using support vector regressors. *Energy* 70, 121–127.

Cheng, H., Yu, C., 2015. Multi-model solar irradiance prediction based on automatic cloud classification. *Energy* 91, 579–587.

Chu, Y., Coimbra, C.F.M., 2017. Short-term probabilistic forecasts for direct normal irradiance. *Renew. Energy* 101, 526–536.

Chu, Y., Pedro, H.T.C., Coimbra, C.F.M., 2013. Hybrid intra-hour DNI forecasts with sky image processing enhanced by stochastic learning. *Sol. Energy* 98 (Part C) 592–603.

Dambreville, R., Blanc, P., Chanussot, J., Boldo, D., 2014. Very short term forecasting of the global horizontal irradiance using a spatio-temporal autoregressive model.

Renew. Energy 72, 291–300.

Engerer, N.A., Mills, F.P., 2014.  $K_{PV}$ : A clear-sky index for photovoltaics. *Sol. Energy* 105, 679–693.

Fu, C., Cheng, H., 2013. Predicting solar irradiance with all-sky image features via regression. *Sol. Energy* 97, 537–550.

Glorot, X., Bengio, Y., 2010. Understanding the difficulty of training deep feedforward neural networks. *AISTATS* 249–256.

Ineichen, P., Perez, R., 2002. A new airmass independent formulation for the Linke turbidity coefficient. *Sol. Energy* 73, 151–157.

Law, E.W., Prasad, A.A., Kay, M., Taylor, R.A., 2014. Direct normal irradiance forecasting and its application to concentrated solar thermal output forecasting – A review. *Sol. Energy* 108, 287–307.

Ji, S., Xu, W., Yang, M., Yu, K., 2010. 3D convolutional neural networks for human action recognition. *ICML* 495–502.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Li, F., 2014. Large-scale video classification with convolutional neural networks. *IEEE CVPR* 1725–1732.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. *NIPS* 1097–1105.

Lawrence, S., Giles, C.L., Tsoi, A.C., Back, A.D., 1997. Face recognition: a convolutional neural-network approach. *IEEE T. Neural Network* 8, 98–113.

Li, M., Chu, Y., Pedro, H.T.C., Coimbra, C.F.M., 2016. Quantitative evaluation of the impact of cloud transmittance and cloud velocity on the accuracy of short-term DNI forecasts. *Renew. Energy* 86, 1362–1371.

Li, P., Chen, Z., Yang, L.T., Zhang, Q., Deen, M.J., 2017. Deep convolutional computation model for feature learning on big data in internet of things. *IEEE T. Ind. Inform.* 14, 790–798.

Marquez, R., Coimbra, C.F.M., 2013. Intra-hour DNI forecasting based on cloud tracking image analysis. *Sol. Energy* 91, 327–336.

Martínez-Chico, M., Batlles, F.J., Bosch, J.L., 2011. Cloud classification in a mediterranean location using radiation data and sky images. *Energy* 36, 4055–4062.

Martin, L., Zarzalejo, L.F., Polo, J., Navarro, A., Marchante, R., Cony, M., 2010. Prediction of global solar irradiance based on time series analysis: Application to solar thermal power plants energy production planning. *Sol. Energy* 84, 1772–1781.

Quesada-Ruiz, S., Chu, Y., Tovar-Pescador, J., Pedro, H.T.C., Coimbra, C.F.M., 2014. Cloud-tracking methodology for intra-hour DNI forecasting. *Sol. Energy* 102, 267–275.

Schmidhuber, J., 2015. Deep learning in neural networks: An overview. *Neural Networks* 61, 85–117.

Shaheen, F., Verma, B., 2017. An Ensemble of Deep Learning Architectures for Automatic Feature Extraction. *IEEE SSCI*.

Shi, C., Wang, C., Wang, Y., Xiao, B., 2017. Deep convolutional activations-based features for ground-based cloud classification. *IEEE Geosci. Remote S.* 14, 816–820.

Simonyan, K., Zisserman, A., 2014. Two-Stream convolutional networks for action recognition in videos. *IEEE CVPR* 568–576.

Singh, G.K., 2013. Solar power generation by PV (photovoltaic) technology: A review. *Energy* 53, 1–13.

Su, Y., Chan, L., Shu, L., Tsui, K.L., 2012. Real-time prediction models for output power and efficiency of grid-connected solar photovoltaic systems. *Appl. Energy* 93, 319–326.

Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M., 2014. C3D: generic features for video analysis. *ArXiv* 412.0767v2.

Ye, L., Cao, Z., Xiao, Y., 2017. DeepCloud: Ground-based cloud image categorization using deep convolutional features. *IEEE T. Geosci. Remote* 55, 5729–5740.

Zhu, T., Wei, H., Zhao, X., Zhang, K., Fang, S., 2016. A method of cloud classification based on DNI. *IEEE CCDC* 4155–4160.

Zhu, T., Wei, H., Zhao, X., Zhang, C., Zhang, K., 2017. Clear-sky model for wavelet forecast of direct normal irradiance. *Renew. Energy* 104, 1–8.