



ARTICLE

Research on Photovoltaic Power Forecasting Based on Ground-Based Cloud Images

Yong Xie^{1,#}, Shidong Feng^{1,#}, Wenzheng Yu^{2,*}, Junyu Zhu², Yifeng Ma^{2,3}, Jianzhou Lin², Zhiyuan You⁴, Zhanyuan Ma¹, Fa Wen¹, Zhidan Hui¹ and Guoyu Niu⁵

¹China Resources New Energy (Qinghai) Co., Ltd., Xining, China

²School of Geographical Sciences, Nanjing University of Information Science and Technology, Nanjing, China

³Faculty of Engineering, University of Nottingham, University Park, Nottingham, UK

⁴Yangzhou Meteorological Bureau, Yangzhou, China

⁵China Resources Wind Power (Da Qaidam) Co., Ltd., Haixi, China

*Corresponding Author: Wenzheng Yu. Email: ywzheng519@126.com

#These authors contributed equally to this work

Received: 11 August 2026; Accepted: 17 September 2026; Published: 18 October 2026

ABSTRACT: To address the challenges of photovoltaic (PV) power forecasting, this study proposes an improved forecasting model that aims to enhance prediction accuracy and cope with the uncertainty of the climate system. The core of the model consists of three parts: (1) A Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) network is employed to predict ground-based cloud images, demonstrating superior detail prediction capability over Generative Adversarial Networks (GAN). (2) An algorithm is designed to track the sun's position, partition the cloud image region, and construct quantitative indicators, which are verified through statistical tests to be highly correlated with direct solar radiation. (3) Top-of-atmosphere irradiance is predicted by incorporating meteorological parameters, and a Classification and Regression Tree (CART) regression tree model is subsequently used for power forecasting, achieving a Mean Absolute Error (MAE) of 4.43 and a Root Mean Square Error (RMSE) of 6.08.

KEYWORDS: CNN; LSTM; ground-based cloud images; CART

1 Introduction

Renewable energy sources such as solar energy have become pivotal to green and low-carbon development owing to their clean and sustainable nature [1]. However, the inherent variability of solar irradiance—largely driven by cloud dynamics—significantly affects photovoltaic (PV) power output, thereby influencing overall system efficiency and grid stability [2,3]. Accurate forecasting of solar irradiance and PV power generation is therefore essential for the efficient operation and optimal dispatch of modern power systems [4,5].

Regarding solar power forecasting, scholars worldwide have conducted extensive research and established three main approaches: physical model methods, traditional statistical methods, and machine learning methods. Physical models construct simulation frameworks based on solar cell module characteristics and environmental parameters. For instance, Zhai [6] employed a current model combined with parameters such as temperature and radiation for prediction; Emanuele et al. achieved relatively high prediction accuracy using circuit models [7]; and Li et al. developed a prediction model using satellite spectral image

data [8]. Traditional statistical methods rely on historical data. For example, Babatunde et al. utilized a multiple linear regression model to handle correlation indicators [9]. Additionally, various statistical techniques such as grey system models and moving average sequence models have been employed [10–12]. Machine learning methods, especially deep learning algorithms, have demonstrated significant advantages in the field of solar power forecasting. Malvoni et al. [13] combined LSSVM with PSO to enhance prediction performance and trained an SVM after dimensionality reduction using PCA; Xiang [14] utilized VAE and Long Short-Term Memory (LSTM) combined with meteorological and cloud image data to forecast power; Zhan et al. attempted to use Generative Adversarial Networks (GAN) to improve satellite cloud image prediction [15]; and Shree constructed a similar-day discrimination model based on weather classification and combined it with LSTM to forecast under different weather conditions [16]. Furthermore, studies on irradiance-driven weather classification, sky/meteorological feature modeling, and PV generation characteristics are also indispensable in improving power forecasting. Yeung et al. found that solar irradiance distribution varies with weather conditions, thereby affecting PV power output [17]. They further integrated weather classification into PV power forecasting, significantly reducing forecast errors and improving grid management [18].

A comparison of existing sky image-based PV forecasting studies reveals a lack of predictive modeling for the evolution of ground-based cloud images. In contrast, the proposed method utilizes an encoder-decoder Convolutional Neural Network-Long Short-Term Memory (CNN-LSTM) network to perform time-series forecasting of future ground-based cloud images, explicitly accounting for cloud motion and morphological evolution, the steps include: designing a LSTM model with an encoder-decoder architecture to predict future cloud images and optimize information extraction; using the Otsu algorithm and an adaptive algorithm to extract features of cloud and sky regions; and integrating real-time meteorological data to build a model that analyzes the influence of various factors on solar power generation, which is then applied to power forecasting. The study aims to provide more accurate data support and facilitate efficient dispatch and operation of the power system.

2 Feature Extraction Based on Ground-Based Cloud Images

2.1 Ground-Based Cloud Image Inpainting

This study employs a Huayun Shengda ASI-16 all-sky imager (ASI). To mitigate the impact of obstacles such as buildings and trees on the sky features captured by the ASI, this study proposes an image obstacle identification and interpolation algorithm. Utilizing clear-sky images captured by a fisheye lens, the obstacle contours are extracted by setting an appropriate threshold. Subsequently, a dilation operation is applied to select the edge transition regions. Finally, a recursive algorithm is employed to perform bilinear interpolation within the obstacle regions, thereby completing the missing information in the ground-based cloud images. The final result is shown in Fig. 1.

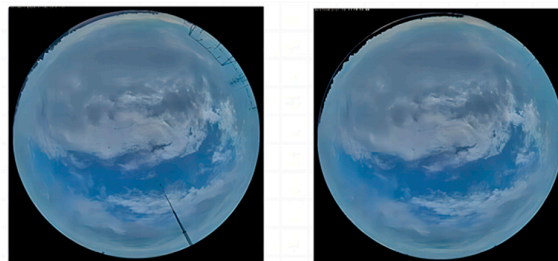


Figure 1: Comparison of ground cloud images before and after interpolation.

2.2 Calculation of Solar Azimuth and Zenith Angles

To better determine the solar position, a spherical coordinate system is established: taking the observation point as the origin O, the z-axis points vertically upward; the x-axis points due north, serving as the 0° azimuth reference; and the y-axis points due east. The solar azimuth angle γ_s is measured clockwise from due north, with a range of $[0^\circ, 360^\circ]$; the zenith angle θ_z is measured from the positive z-axis, with a range of $[0^\circ, 90^\circ]$; the solar elevation angle α_s and the zenith angle are complementary, i.e., $\alpha_s = 90^\circ - \theta_z$. All angular units in this study are in degrees, which are converted to radians when trigonometric functions are calculated. The latitude of the observation site φ is positive north of the equator and negative south of it; the declination angle δ and the hour angle ω are calculated according to astronomical conventions, where the hour angle is negative in the morning and positive in the afternoon. The specific calculation formulas are as follows:

$$\delta = 23.45 \sin\left(\frac{2\pi(284 + N)}{365}\right) \quad (1)$$

$$\theta_z = \arccos(\sin \varphi \sin \delta + \cos \varphi \cos \delta \cos \omega) \quad (2)$$

$$\gamma_s = \arccos\left(\frac{\cos \theta_z \sin \varphi - \sin \delta}{\sin \theta_z \cos \varphi}\right) \quad (3)$$

In the formulas, θ_z is the zenith angle; γ_s is the solar azimuth angle; φ is the latitude of the observation site; δ is the declination angle; and ω the hour angle.

Using the solar azimuth angle, the orientation of the sun's centroid relative to the center of the ground-based cloud image can be determined. Subsequently, by determining the distance from the sun's centroid to the center of the ground-based cloud image, the position of the sun's centroid in the image can be established for any given time.

2.3 Fisheye Lens Imaging Fitting

The interior of a fisheye lens typically consists of dozens of different lenses. Incoming light undergoes varying degrees of refraction as it passes through lenses of different thicknesses and sizes, and is ultimately projected onto the imaging plane. As a result, the field of view of a fisheye lens can reach up to 180° , far exceeding that of ordinary lenses. The perspective projection of fisheye lenses can be broadly classified into five types: perspective projection, equisolid angle projection, equidistant projection, stereographic projection, and orthographic projection. However, real lenses do not fully conform to these projection models, and from the perspective of correction, a general form of projection model can be adopted to accommodate different types of fisheye lenses [19].

Overall, this spatiotemporal module strikes a balance between extracting multi-scale spatial structural features and enabling efficient interaction of inter-frame dynamic information through concatenation and attention-based fusion. As a result, it accurately captures the motion evolution patterns across consecutive frames while preserving spatial details.

$$r(\theta) = k_0 + k_1\theta + k_2\theta^3 + k_3\theta^5 + k_4\theta^7 + \dots \quad (4)$$

where θ is the zenith angle (in degrees), and $r(\theta)$ represents the distance from the solar centroid to the image center (in pixels).

Using the calculated historical solar zenith angle data and the computed distances from the sun's centroid to the image center, a least-squares parameter fitting was performed on the formula. After least-squares fitting, the root mean square error (RMSE) of the affine model is 1.6 pixels and the maximum residual is 3.8 pixels; the coefficients of terms higher than the first order are close to zero, indicating that the simplified model does not introduce significant fitting error within the effective field-of-view radius and meets the accuracy requirement for solar region segmentation. To simplify the model while maintaining accuracy, this study adopts the affine transformation formula as the imaging principle of the fisheye lens, which is given as follows:

$$r(\theta) = -10.78 * \theta + 972.9 \quad (5)$$

where θ is the zenith angle (in degrees), and $r(\theta)$ represents the distance from the solar centroid to the image center (in pixels).

This simplified model is applicable when the distance from the solar centroid to the image center does not exceed the effective field-of-view radius of the image. For the all-sky imager used in this study, the effective imaging range of the solar centroid in the image covers the main observation period from after sunrise to before sunset.

After calculating the solar zenith angle, this equation determines the distance from the sun's centroid to the center of the ground-based cloud image. Combined with the solar azimuth angle, the exact imaging position of the sun's centroid within the ground-based cloud image can be established. Taking the sun's centroid as the center and the maximum radius of the solar region derived from historical data as the boundary, the ground-based cloud image is divided into a solar region quantitative indicator (SRQI) and a non-solar region quantitative indicator (NSRQI).

2.4 Normalized Red-Blue Ratio

In clear weather, large particulate matter such as aerosols is relatively scarce. According to Rayleigh scattering theory, the scattering intensity is inversely proportional to the fourth power of the wavelength. Consequently, blue light, with its shorter wavelength, is scattered more strongly, while red light is scattered to a lesser extent. In contrast, cloud masses contain abundant large water droplets and ice crystals. For particles comparable in size to the wavelength of radiation, the scattering intensity is proportional to the square of the frequency, causing clouds to appear white or gray. Therefore, quantitative analysis can be conducted using the red, green, and blue (RGB) color channels from ground-based cloud images. Many researchers adopt the ratio of the red channel to the blue channel as the feature space for threshold determination, because the scattering intensity of particles is wavelength-dependent. The specific definition of the red-blue ratio λ is as follows:

$$\lambda = \frac{B}{R} \quad (6)$$

where R is the pixel value of the red channel, and B is the pixel value of the blue channel.

To avoid a zero denominator, when R is 0 it is assigned a default value of 1, so that the red-blue ratio λ ranges within $[0, 255]$. The traditional red-blue ratio can effectively detect thick cloud information in ground-based cloud images and performs particularly well in clear or overcast/rainy weather. However, this method suffers from poor visual contrast and noise robustness, especially in low-brightness areas where it is susceptible to instrument noise. To improve the visual contrast and noise robustness of the feature,

this study adopts the normalized red-blue ratio [20] as the feature space for threshold determination. The specific definition of the normalized red-blue ratio is as follows:

$$\bar{\lambda} = \frac{B - R}{B + R} \quad (7)$$

The normalized red-blue ratio $\bar{\lambda}$ ranges from -1 to 1 . By applying inverse normalization, the values can be remapped to the range $[0, 255]$.

2.5 Otsu's Method

Maximum between-class variance serves as a crucial criterion for automatic, non-parametric, and unsupervised clustering in statistical pattern recognition. The image thresholding method based on this principle was introduced by the Japanese scholar Nobuyuki Otsu [21] and is widely known as Otsu's method. The core principle involves partitioning the original image into two distinct classes—the target (foreground) and the background—based on their grayscale values, such that the between-class variance is maximized. From a statistical pattern recognition perspective, the threshold derived from this method maximizes the between-class variance, thereby yielding an optimal classification.

Furthermore, Otsu's method operates exclusively on the one-dimensional (1D) grayscale histogram of the image. Since calculating a 1D histogram is computationally straightforward and intuitive, the algorithm becomes completely independent of the original image dimensions once the histogram is acquired, ensuring exceptionally high computational efficiency.

The original Otsu algorithm performs remarkably well on images where the histogram exhibits distinct bimodal characteristics and the variance difference between the foreground and background is small. This effectiveness stems from the fact that its core criterion—maximizing the between-class variance—naturally aligns with the structural characteristics of a bimodal distribution, where the grayscale values of the foreground and background are concentrated respectively, leaving a distinct valley between the two peaks. When the threshold is located exactly at this valley, the two classes of pixels are correctly separated; the difference between their respective grayscale means is maximized, and the intra-class grayscale distribution is the most compact, thereby maximizing the between-class variance. Conversely, when these bimodal conditions are not met, the segmentation efficacy significantly deteriorates.

2.6 Sky Region Feature Analysis

Cloud image feature extraction is generally categorized into machine learning approaches and non-machine learning approaches. Machine learning methods rely on extensive, manually interpreted historical data and public cloud image datasets to train models. Conversely, non-machine learning methods typically employ fixed-threshold segmentation techniques to detect cloud regions. The threshold for cloud detection depends heavily on the regional climate and the specific camera equipment used for image capture. Furthermore, this process is influenced by various complex factors, resulting in the frequent absence of distinct threshold boundaries between cloud and sky elements [22]. Based on their statistical characteristics, cloud images can be broadly classified into single-peak images and double-peak images. Single-peak images consist of only one type of object (such as cloud or sky); cloud images under stratiform cloud cover or clear-sky conditions can usually be regarded as single-peak images, whereas cumuliform clouds and most cirriform clouds can be regarded as double-peak images [20]. By converting the original RGB images into specific feature images (e.g., the red-to-blue channel ratio), distinct differences can be observed between the histograms of unimodal and bimodal images. The histogram of a single-peak image typically has only one

peak and a relatively small standard deviation, while that of a double-peak image has two or more peaks and a larger variance. Although cloud and sky elements exhibit significant differences to the naked eye, they largely overlap in most feature spaces. This substantial overlap renders fixed-threshold methods generally ineffective when processing bimodal images. Fig. 2 shows cloud images of stratiform and cirriform clouds along with their statistical histograms.

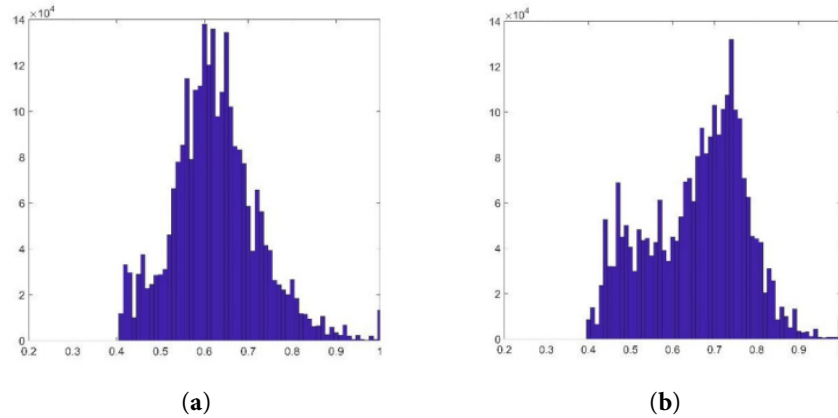


Figure 2: (a) Histograms of stratocumulus cloud; (b) Histograms of cumulus cloud.

Due to the significant physical and statistical discrepancies between unimodal and bimodal cloud images, distinct algorithms are employed for their respective processing. From the perspective of image processing, adaptive threshold algorithms tend to yield more accurate results for double-peak images [23]. For single-peak images, because the sky conditions are relatively uniform, a fixed threshold can perform well; however, fixed thresholds are less effective for thin clouds and most cirriform cloud images. In summary, this study establishes a hybrid strategy for determining thresholds in ground-based cloud images, with the specific workflow illustrated in Fig. 3.

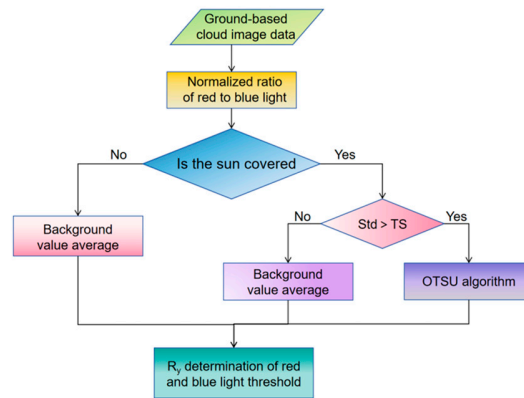


Figure 3: Flowchart of threshold determination algorithm for ground-based cloud images.

3 Forecasting Model

3.1 CNN and LSTM Algorithm Models

CNN is one of the primary algorithms in deep learning. By replacing traditional linear models with convolution operations, CNN possesses a deeper structure and is capable of extracting data features and geometric regularities, making it suitable for processing high-dimensional image data. The convolution

operation proceeds across the image in a left-to-right, top-to-bottom order according to set parameters. The convolution calculation involves multiplying the pixel values of the covered image region by the corresponding kernel weights, summing these products, adding a bias term, and finally obtaining the pixel value of the target. Let X_i be the input, w_{ij} be the weight connecting the i -th input to the j -th neuron, and b_j be the bias of the j -th neuron; the expression for its output value is then:

$$y_j = F\left(b_j + \sum_{i=1}^n (X_i \times w_{ij})\right) \quad (8)$$

where F is the activation function of the neuron. Different activation functions are often adopted according to different computational requirements; the main ones include the Sigmoid function, Tanh function, ReLU function, etc.

LSTM is a specialized recurrent neural network (RNN) architecture that can handle long-term dependencies much more effectively than traditional RNNs. The specific mathematical expressions are as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (9)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \bar{C}_t \quad (10)$$

where x_t is the input of the unit at time t ; h_t is the output of the neuron at time t ; W is the weight parameter matrix within the neuron; b is the bias vector within the neuron; σ is the activation function; f_t is the output value of the forget gate; and $\odot$ denotes the element-wise product.

The input gate is used to determine how much of the current input should be retained in the cell state. That is, it selectively records new information into the cell and decides which new information to keep. The input gate generally consists of a Sigmoid layer and a Tanh layer; the Sigmoid layer determines which values to update, while the Tanh layer generates new candidate memories, supplementing and discarding attribute values. The specific mathematical expressions are as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (11)$$

$$\bar{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (12)$$

The output gate is used to control the cell state. That is, it decides whether new information should be added to the state at the current time step, and the final output is obtained by multiplying the results of the Sigmoid layer and the Tanh layer. The specific mathematical expressions are as follows:

$$O_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (13)$$

$$h_t = O_t \odot \tanh(C_t) \quad (14)$$

where O_t is the output of the output gate at time t , and h_t is the output of the neuron at time t ; $\bar{C}_t$ denotes the candidate cell state and C_t the updated cell state. The parameters within the unit are iteratively updated during the training process, and the activation function serves to enhance the model's nonlinear expressive capability.

The CNN-LSTM forecasting model constructed in this study adopts an encoder–decoder structure. The specific network configuration is as follows. The input consists of 16 consecutive historical ground-based cloud images, each resized to 256×256 pixels and normalized to the range $[0, 1]$. In the encoder, a shared-weight convolutional feature extraction module first extracts spatial features from each frame. This module contains three convolutional blocks: the first block comprises a convolutional layer with 32 kernels of size 3×3 , stride 1, and padding 1, followed by batch normalization, ReLU activation, and 2×2 max pooling; the second block increases the number of kernels to 64 while keeping the remaining configuration unchanged; the third block increases the number of kernels to 128 while maintaining the same configuration. After the three convolutional blocks, each $256 \times 256 \times 3$ image is mapped to a feature map of size $32 \times 32 \times 128$. Subsequently, the feature maps at the 16 time steps are flattened into a sequence of vectors, and each vector is reduced to 512 dimensions through a fully connected layer, forming a temporal sequence of shape $(16, 512)$. This sequence is fed into a two-layer LSTM with 256 hidden units per layer, and the hidden state at the last time step is taken as the encoded representation. In the decoder, a fully connected layer first maps the 256-dimensional hidden state to 512 dimensions, which is then reshaped and upsampled through a series of transposed convolutional layers to reconstruct a $256 \times 256 \times 3$ image. Specifically, the 512-dimensional vector is mapped to $128 \times 16 \times 16$ via a fully connected layer and reshaped to $(16, 16, 128)$, followed by upsampling (2×2) and transposed convolution (64 kernels of size 3×3 , stride 2, padding 1), batch normalization, and ReLU activation; then upsampling (2×2) and transposed convolution (32 kernels of size 3×3 , stride 2, padding 1), batch normalization, and ReLU activation; and finally a 3×3 convolutional layer with 3 kernels, stride 1, and padding 1 with Tanh activation to output the predicted image. The model is trained using the mean squared error loss function and the Adam optimizer with a learning rate of 0.001, $\beta_1 = 0.8$, and $\beta_2 = 0.99$. The batch size is set to 64, the model is trained for 5000 epochs, and the model with the lowest validation loss is saved.

3.2 Encoder-Decoder Architecture

The encoder maps the original data into a low-dimensional vector space for feature extraction from the original data. As the name implies, the encoder encodes the input original data X , transforming the input sentence into an intermediate semantic representation C through nonlinear transformations:

$$C = F(x_1, x_2, \dots, x_m) \quad (15)$$

The decoder then restores this vector to the original data or produces prediction results for a specific task. For the decoder, its task is to generate the output y_i at time i based on the intermediate semantic representation C of the input X and the previously generated historical information $y_1, y_2, \dots, y_{i-1}$:

$$y_i = g(C, y_1, y_2, \dots, y_{i-1}) \quad (16)$$

Each y_i is generated sequentially in this manner, so that the entire system appears to produce the target output Y from the input X .

3.3 Construction of the Optimized CART Regression Tree Model

CART (Classification and Regression Tree) is a method that recursively partitions a dataset into different subsets using a binary tree structure to build a predictive model. It continuously performs binary splits using the Gini coefficient until all features are exhausted. The Gini coefficient can be expressed as:

$$G(x, y) = \frac{x_1}{x} G(x_1) + \frac{x_2}{x} G(x_2) \quad (17)$$

where G is the Gini coefficient; x is the output dataset; x_1, x_2 are the two sub-datasets formed by splitting the output dataset; and y is the feature vector.

After completing the feature classification for the solar power generation correlation model using the Gini coefficient, the optimal feature is selected as the root node of the CART tree. Parallelization and a loss function are then introduced to construct the basic CART model. On this basis, the leaf nodes of the regression tree are determined. During the generation of the regression tree, the original sample space is recursively partitioned according to the differences among sample attributes. The degree of difference between sample attributes can be quantified by the following formula:

$$D = \zeta(z) - \frac{|z_1|}{|z|}\zeta(z_1) - \frac{|z_2|}{|z|}\zeta(z_2) \quad (18)$$

where D is the degree of difference of the input sample attributes; $\zeta(z_2)$ is the standard deviation of the target attribute of the total sample; z is the sample space; z_1 and z_2 are the sample subspaces. After the sample space partitioning is completed, the leaf nodes need to be determined. We adopt the leaf node regression equation defined by the SVM regression algorithm as follows:

$$f(x_i) = \tau K(x_i) + b \quad (19)$$

where f is the optimal separating hyperplane; K is the kernel function of the indicator; b is the bias term of the regression equation; and τ is the normal vector.

To address the nonlinear relationship between the input feature vectors and various indicators, the original data are mapped into a high-dimensional space to transform the problem into a linear one. The specific optimization objective of this model is as follows:

$$\min\left(\frac{1}{2}\|\tau\|^2 + \Phi \sum_{i=1}^n (\delta_i + \delta_i^*)\right) \quad (20)$$

subject to:

$$\begin{cases} f(x_i) - [\tau K(x_i)] - b \leq \varepsilon + \delta_i \\ [\tau K(x_i)] + b - f(x_i) \leq \varepsilon + \delta_i^* \\ \delta_i \geq 0 \ (i = 1, 2, K, n) \\ \delta_i^* \geq 0 \ (i = 1, 2, K, n) \end{cases} \quad (21)$$

where f is the optimal separating hyperplane; K is the kernel function of the indicator; b is the bias term of the regression equation; τ is the normal vector; Φ is the penalty parameter; ε is the regularization term used to balance the error; and δ_i, δ_i^* are the slack variables.

By solving the problem through the Lagrangian function objective, and under the premise of nonlinear separability, the following optimal hyperplane decision function can be obtained:

$$f(x) = uK(x) + b^* \quad (22)$$

where u is the Lagrange multiplier; b^* is the optimized bias term.

After the above steps, we successfully constructed a CART regression tree model suitable for characterizing the correlation conditions of solar power generation. To improve the prediction accuracy, the model needs to be pruned. By using a bottom-up recursive method, redundant subtrees are replaced with leaf

nodes, thereby achieving the construction of an optimal decision tree. After pruning, smoothing operations are also required to ensure continuity between adjacent leaf nodes. By applying the optimized CART regression tree model, we can effectively analyze the influence relationships of various indicators on solar power generation. In addition to analyzing these influence relationships, the trained CART regression tree model can also be used for simple data prediction to support the feasibility of the proposed research approach.

3.4 Evaluation Metrics

The goodness of fit between model predictions and training values is often assessed using the mean squared error (MSE) as the loss function; a smaller loss value generally indicates better fitting performance. Other indicators that also reflect the model's fitting performance include the mean absolute error (MAE) and the root mean square error (RMSE). The specific calculation formulas are as follows:

$$E_{MSE} = \frac{1}{N} \sum_{i=1}^N (p_i - P_i)^2 \quad (23)$$

$$E_{MAE} = \frac{1}{N} \sum_{i=1}^N |p_i - P_i| \quad (24)$$

$$E_{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (p_i - P_i)^2} \quad (25)$$

4 Case Study

4.1 Data Processing

This study presents a case study based on data from the Ecological Fishery Photovoltaic Power Generation "Leader" Demonstration Base in Baoying County, Yangzhou City, Jiangsu Province. Since January 2020, this demonstration base has continuously collected ground-based cloud images, meteorological data, and solar power generation data at the Liubao Photovoltaic Power Station in Baoying County, Yangzhou City, Jiangsu Province. The Liubao Solar Power Station is equipped with meteorological monitoring equipment and an ASI-16 all-sky imager. The meteorological equipment collects multi-dimensional meteorological data every 15 min, and the ASI-16 captures ground-based cloud image data with a resolution of 1960×1960 every 10 min. Data collection was conducted daily between 06:00 and 19:50 from 30 April to 29 July 2020. To prepare the data for model training, all variables were scaled to a $[0, 1]$ range using min-max normalization. Both the training and validation datasets were partitioned from this continuous observation period.

All data were sorted chronologically and partitioned into strictly non-overlapping training, validation, and test sets, with the test set positioned at the end of the temporal sequence. Specifically, the training set spanned from 30 April to 15 July 2020; the validation set from 16 July to 27 July 2020; and the test set comprised data from 28 July 2020. For the cloud image forecasting task, each sample sequence consisted of 17 consecutive frames, where the first 16 frames served as inputs to predict the 17th frame. Sequence generation strictly adhered to chronological order, ensuring that the time windows of the test sequences fell entirely after those of the training and validation sets. To prevent temporal data leakage, all normalization statistics were calculated exclusively from the training set before being applied to the validation and test sets. Similarly, model hyperparameters were tuned using only the training and validation data, reserving

the test set strictly for unbiased final evaluation. A CNN-LSTM network was employed for forecasting. The model was optimized using Adam with a learning rate of 0.001, a batch size of 64, and decay rates of $\beta_1 = 0.8$ and $\beta_2 = 0.99$. Optimal model weights were obtained and saved after 5000 iterations. To ensure a fair and rigorous evaluation, both the GAN and CNN-LSTM models were fed identical input sequences and evaluated under the exact same prediction horizon. Finally, a comparison with the GAN prediction results is illustrated below.

As shown in Fig. 4, both the CNN-LSTM and the GAN perform well in predicting the global contour features of ground-based cloud images. However, the CNN-LSTM yields superior predictions for localized details—such as sun position, cloud distribution, cloud morphology, and color contrast—while also exhibiting more realistic texture features. Conversely, the GAN primarily captures the approximate trajectory of the sun and struggles to accurately predict critical spatial information like cloud position and morphology. Consequently, the CNN-LSTM architecture was selected for forecasting future ground-based cloud images in this study.

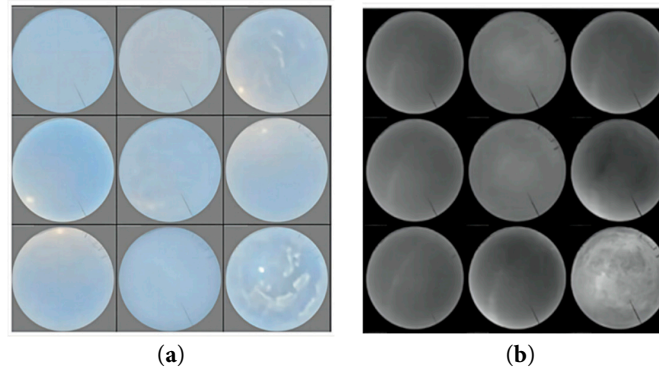


Figure 4: (a) Prediction results of CNN-LSTM; (b) Prediction results of GAN.

4.2 Quantitative Indices for Ground-Based Cloud Images

Analysis of ground-based cloud imagery reveals that the red channel exhibits the maximum separability among the solar, cloud, and sky regions. Typically, red channel intensity peaks in the solar region, while the cloud and sky regions register at approximately one-half and one-third of this maximum value, respectively. However, across different images, the absolute red channel values for these regions fluctuate significantly due to varying capture times and meteorological conditions, resulting in inconsistent background baselines. To robustly quantify these images for integration into downstream correlation models, the derived index must accurately capture the distinct features of the cloud, sun, and sky regions. Furthermore, its numerical values should ideally exhibit an ordinal relationship (e.g., sun > sky > cloud). These essential quantification criteria, combined with the empirically observed numerical characteristics of the color channels across the three regions, establish the foundational rationale for the proposed methodology detailed below.

To effectively highlight anomalous regions in the sky (i.e., clouds and the solar disk) and eliminate background baseline interference, the absolute mean deviation of the cloud image channel data was calculated. The specific mathematical formulation is as follows:

$$\tilde{x} = |x - \mu| \quad (26)$$

where $\tilde{x}$ is the processed absolute deviation data, x is the original sample data, and μ is the mean of all pixels in the sample image.

Based on the observed channel characteristics of the sun, cloud, and sky regions, and since the red channel is the most sensitive to various region types, this study adopts red band pixel values as the primary data source for quantifying ground-based cloud images. By incorporating the absolute mean deviation principle, a quantitative index Z is established. The index Z quantifies the degree to which a pixel or region in the red channel of a ground-based cloud image deviates from the “cloud–sky critical state.” Its physical meaning can be summarized as the red-channel weighted anomaly intensity, and the following formula is proposed:

$$Z = R \left| R - R_y + \frac{R_b}{2} \right| \quad (27)$$

where R is the red-band pixel gray value of the ground-based cloud image recorded by the imaging sensor, R_y is the red-band threshold for separating cloud from sky (i.e., the cloud detection threshold), and R_b is the red-band pixel value of the clear, cloud-free sky (i.e., the sky background reference value). R , R_y , and R_b are all 8-bit grayscale values ranging from 0 to 255. The absolute value term in the equation represents the distance by which the current red-channel value R deviates from the reference value $R_y - \frac{R_b}{2}$. After multiplication by R , regions that are both bright and strongly deviating obtain larger Z values, thereby highlighting the solar region and thick cloud regions while suppressing the sky background. Therefore, a larger Z value indicates that the region is more likely to be the solar region or thick cloud, whereas a smaller Z value indicates that the region is closer to the clear-sky background.

4.3 Solar Power Generation Correlation Model Based on the Regression Tree Algorithm

The quantitative indices extracted from the cloud images were resampled to a 10-min resolution using linear interpolation and aligned with the meteorological data. Using R programming software, the solar region quantitative indicator (SRQI), non-solar region quantitative indicator (NSRQI), core temperature (CT), temperature (T), wind speed (WS), and solar power generation (SPG) were input into the CART model. The relationship between the cross-validation relative error (X-val relative error) and the complexity parameter (CP) of the model is shown in Fig. 5. As can be seen from the figure, the minimum cross-validation error is 0.022, at which point the CART model achieves optimal performance; that is, when $CP = 0.01$, the hierarchical result of the CART regression tree attains the highest accuracy with the least complexity. Therefore, the CP value should be set to 0.01, meaning that the first two levels of the CART regression tree are retained while the remaining subtrees are pruned.

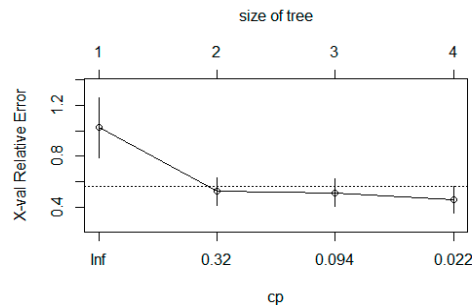


Figure 5: cp value and its corresponding error.

Based on the above results, the CART tree was pruned, yielding the tree shown in Fig. 6. The tree has a depth of 3, a total of 7 nodes, of which 4 are terminal nodes. Each path from the root node to a leaf node

represents a set of fluctuation conditions affecting solar power generation. Within each node box are the mean value (Z) of the response variable for the classified samples, the total number of samples (n), and the percentage of samples. The variables between nodes represent the main influencing predictors extracted by the model, and the corresponding splitting values ($<$ or $\geq$) are the classification thresholds for these predictors.

The root node uses the NSRQI as the splitting criterion, with less than 40% of the samples assigned to one intermediate node. This intermediate node then uses T as the splitting criterion: samples with $T < 32$ (27%) are assigned to node 6, and the remaining samples (12%) are assigned to node 7. The remaining 61% of the samples at the root node are assigned to the other intermediate node. This other intermediate node again uses NSRQI as the splitting criterion: 32% of the samples with $\text{NSRQI} \geq 411$ are assigned to node 5, and the remaining 29% with $\text{NSRQI} < 411$ are assigned to node 4.

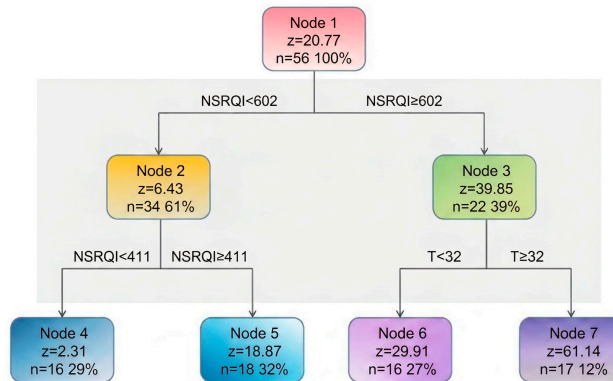


Figure 6: Solar power generation decision tree structure.

These results indicate that among the influencing factors of solar power generation input to the CART model, two are dominant: NSRQI (non-solar region quantitative index) and T (temperature). NSRQI has the greatest impact on solar power generation in the study area. When $\text{NSRQI} < 602$, converting the threshold of the classification condition (standardized data) back to the original data yields solar power generation $\text{SPG} < 23.7$. The Z values of solar power generation within this node are generally small, with an average Z value of only 8.43, indicating relatively low solar power output. When $\text{NSRQI} \geq 602$, the main factor affecting the variation in solar power generation is T , which is the second most important fluctuation factor. When temperature $T \geq 32$, the overall solar power generation in the study area is the highest, with an average Z value reaching 61.14. When temperature $T < 32$, solar power generation is at an intermediate level, with a mean value of 29.91.

From the perspective of PV power fluctuation, cloudy weather is characterized by rapid and severe power fluctuations. As widely recognized in the solar forecasting community, complex overcast conditions with rapidly evolving cloud patterns pose the greatest challenge for prediction and yield the lowest forecast accuracy. Accordingly, we select July 28, a typical cloudy day, as the test day. The trained regression tree was applied using R language to predict the solar power generation at the Liubao Photovoltaic Power Station in Baoying County, Yangzhou City, on 28 July 2020. The final prediction results are shown in Fig. 7.

The prediction results of the CART model were evaluated using standard formulas, and the forecasting errors of the proposed method were compared with those of alternative approaches. The calculated results are presented in Table 1.

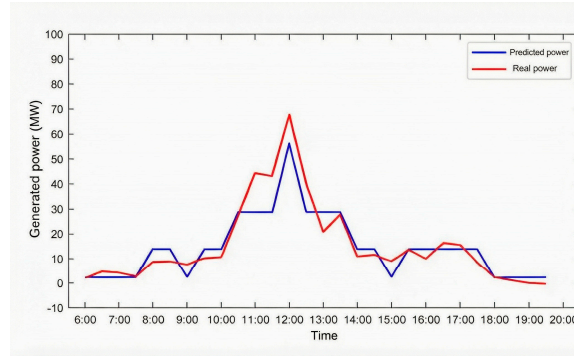


Figure 7: Solar power generation forecasting results chart.

Table 1: Forecasting error.

Model	Forecasted Day	E_{MAE}	E_{RMSE}
CART	7.28	4.43	6.08
LSTM	7.28	4.72	6.91
BP	7.28	5.61	7.88

The results indicate that the proposed method achieves a slight reduction in forecasting error compared to the models reported in the literature. However, given the limited size of the training dataset currently utilized for the CART model, a more substantial improvement in predictive performance is anticipated once additional data is incorporated in future work.

The regression tree model results demonstrate that the quantitative indices evaluated in this study—particularly the non-solar region index—significantly impact solar power generation. Higher values in the non-solar region index correspond to increased average power output, suggesting this metric effectively captures scattering and reflection features within ground-based cloud imagery. Additionally, temperature exhibits a strong positive correlation with solar power generation. Historical data indicates this is largely driven by the diurnal cycle: higher solar elevation angles at midday simultaneously increase ambient temperatures and radiation flux on the solar panels compared to morning and evening hours. Consequently, while temperature and power generation are statistically correlated, there is no direct causal relationship. Conversely, the solar region index did not demonstrate a significant impact on the model, likely due to a predominance of overcast conditions in the training dataset. Accurately characterizing the influence of the solar region index will therefore require incorporating more data from clear-sky conditions.

5 Conclusion

To address the challenges that the stochastic volatility of solar power poses to grid stability, this study proposes a comprehensive framework to enhance forecasting accuracy. The research methodology is structured around three key components. First, an encoder-decoder Convolutional LSTM model is developed to forecast future ground-based cloud imagery. By optimizing the network architecture and data preprocessing pipeline, the model effectively captures spatiotemporal features, significantly improving the predictive accuracy of cloud, sky, and sun positions. Second, we introduce a method to locate the solar centroid in ground-based images, facilitating their segmentation into distinct solar and non-solar regions. Quantitative indices are derived using a hybrid thresholding strategy and validated as reliable indicators of direct solar radiation. Finally, a CART model fuses this quantitative image data with meteorological variables

to forecast solar power generation. Experimental results demonstrate that the proposed multimodal model achieves high predictive accuracy, closely mirroring actual power generation profiles.

The findings of this work preliminarily demonstrate that the proposed approach offers a viable framework for short-term PV power forecasting. While we have validated the model's efficacy under highly fluctuating cloudy conditions, the evaluation of the proposed framework is currently based on only a single test day, this limited temporal coverage may restrict the statistical significance of the reported results and cannot fully capture the model's performance under different weather regimes or seasonal variations. Broadening the assessment to encompass multi-seasonal weather scenarios and supplementing rigorous statistical correlation analyses between the cloud index and solar irradiance will be key focuses of our future investigations.

Acknowledgement: Not applicable.

Funding Statement: This work was financially supported by the Qinghai Provincial Key Research and Development Program (First funding, Grant No. 2025-QY-244): Research on Novel Ultra-Short-Term Photovoltaic (Solar Thermal) Power Generation Output and Medium- and Long-Term Power Generation Forecasting and Its Application. This work was also supported by the Second Tibetan Plateau Scientific Expedition and Research Program (STEP) (Second funding, Grant No. 2019QZKK0804): Evaluation of Current Status and Long-Term Prospect of Clean Energy.

Author Contributions: The authors confirm contribution to the paper as follows: conceptualization, Yifeng Ma and Wenzheng Yu; methodology, Junyu Zhu and Yifeng Ma; software, Junyu Zhu; validation, Yifeng Ma and Zhiyuan You; formal analysis, Junyu Zhu and Guoyu Niu; investigation, Junyu Zhu; resources, Yong Xie, Shidong Feng, Zhiyuan You, Zhidan Hui, Zhanyuan Ma and Fa Wen; data curation, Shidong Feng and Yong Xie; writing—original draft preparation, Jianzhou Lin; writing—review and editing, Wenzheng Yu and Yifeng Ma; visualization, Jianzhou Lin; supervision, Wenzheng Yu; project administration, Wenzheng Yu; funding acquisition, Yong Xie, Shidong Feng, Zhidan Hui, Guoyu Niu, Zhanyuan Ma and Fa Wen. All authors reviewed and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this study are available from the Corresponding Author, Wenzheng Yu, upon reasonable request.

Ethics Approval: Not applicable.

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

Abbreviations

The following abbreviations are used in this manuscript:

Abbreviation	Interpretation
ARMA	Autoregressive Moving Average
ASI	All Sky Imager
BP	Back Propagation
CART	Classification and Regression Tree
CNN	Convolutional Neural Network
CP	Complexity Parameter
EMAE	Mean Absolute Error (used as a forecast error metric, equivalent to MAE)
ERMSE	Root Mean Square Error (used as a forecast error metric, equivalent to RMSE)
WS	Wind speed
GAN	Generative Adversarial Network
SPG	Solar power generation
CT	Core temperature
LSSVM	Least Squares Support Vector Machine
LSTM	Long Short-Term Memory

MAE	Mean Absolute Error
MSE	Mean Squared Error
PCA	Principal Component Analysis
PSO	Particle Swarm Optimization
PV	Photovoltaic
RGB	Red Green Blue
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
STEP	Second Tibetan Plateau Scientific Expedition and Research Program
SVM	Support Vector Machine
SRQI	Solar region quantitative indicator
NSRQI	Non-solar region quantitative indicator
VAE	Variational Autoencoder
T	Temperature

References

1. Sundaram GA, Muniyandi GR. Advancements in renewable energy: Innovations and sustainable solutions for a greener future. *Next Res.* 2025;2(3):100482. [[CrossRef](#)].
2. Siddaraj U, Yaragatti UR, Paragonda LRS, Tangi S. Adaptive MPPT control for reliable transitions between grid connected and islanded operations in PV battery microgrids. *Sci Rep.* 2026;16:7613. [[CrossRef](#)].
3. Ladjal B, Nadour M, Bechouat M, Hadroug N, Sedraoui M, Rabehi A, et al. Hybrid deep learning CNN-LSTM model for forecasting direct normal irradiance: A study on solar potential in Ghardaia, Algeria. *Sci Rep.* 2025;15:15404. [[CrossRef](#)].
4. Rathnayake N, Wijewardane S. Machine learning-based Direct Normal Irradiance (DNI) forecasting using satellite data for Concentrated Solar Power (CSP) plants with Thermal Energy Storage (TES). *Sci Rep.* 2026;16:11257. [[CrossRef](#)].
5. Mohamad Radzi PNL, Mekhilef S, Mohamed Shah N, Akhter MN, Seyedmahmoudian M, Stojcevski A. Optimizing solar power forecasting with metaheuristic algorithms and deep learning models for photovoltaic grid connected systems. *Sci Rep.* 2025;15:40045. [[CrossRef](#)].
6. Zhai Z. Prediction of the output performance of photovoltaic arrays under arbitrary conditions [dissertation]. Hefei, China: University of Science and Technology of China; 2008. (In Chinese).
7. Ogliari E, Dolara A, Manzolini G, Leva S. Physical and hybrid methods comparison for the day ahead PV output power forecast. *Renew Energy.* 2017;113:11–21. [[CrossRef](#)].
8. Li Y, Su Y, Shu L. An ARMAX model for forecasting the power output of a grid connected photovoltaic system. *Renew Energy.* 2014;66:78–89. [[CrossRef](#)].
9. Babatunde AA, Abbasoglu S. Predictive analysis of photovoltaic plants specific yield with the implementation of multiple linear regression tool. *Environ Prog Sustain Energy.* 2019;38(4):13098. [[CrossRef](#)].
10. Chang GW, Lu HJ. Integrating gray data preprocessor and deep belief network for day-ahead PV power output forecast. *IEEE Trans Sustain Energy.* 2020;11(1):185–94. [[CrossRef](#)].
11. Alvarado-Barrios L, Rodríguez del Nozal Á, Boza Valerino J, García Vera I, Martínez-Ramos JL. Stochastic unit commitment in microgrids: Influence of the load forecasting error and the availability of energy storage. *Renew Energy.* 2020;146:2060–9. [[CrossRef](#)].
12. Kushwaha V, Pindoriya NM. A SARIMA-RVFL hybrid model assisted by wavelet decomposition for very short-term solar PV power generation forecast. *Renew Energy.* 2019;140:124–39. [[CrossRef](#)].
13. Malvoni M, de Giorgi MG, Congedo PM. Data on Support Vector Machines (SVM) model to forecast photovoltaic power. *Data Brief.* 2016;9:13–6. [[CrossRef](#)].
14. Xiang M. Probabilistic power forecasting of photovoltaic generation based on ground-based cloud images [dissertation]. Hangzhou, China: Zhejiang University; 2022. (In Chinese). [[CrossRef](#)].
15. Xu Z, Du J, Wang J, Jiang C, Ren Y. Satellite image prediction relying on GAN and LSTM neural networks. In: *Proceedings of the ICC 2019—2019 IEEE International Conference on Communications (ICC)*; 2019 May 20–24; Shanghai, China. p. 1–6. [[CrossRef](#)].

16. Acharya SK, Wi YM, Lee J. Day-ahead forecasting for small-scale photovoltaic power based on similar day detection with selective weather variables. *Electronics*. 2020;9(7):1117. [[CrossRef](#)].
17. Yeung C, Wang J, Du Y, Cao J, Zhou Q, Ding Y, et al. Comprehensive analysis of solar irradiance and photovoltaic power generation: Case studies in Wuhan and Zhangbei, China. *IEEE Trans Ind Applicat*. 2025;61(6):8307–17. [[CrossRef](#)].
18. Yeung C, Wang J, Du Y, Cao J, Zhou Q, Du Z, et al. Clearness index cluster analysis for photovoltaic weather classification based on solar irradiation measurement data: Theory and application. *Energy* 360. 2024;2:100010. [[CrossRef](#)].
19. Kannala J, Brandt S. A generic camera calibration method for fish-eye lenses. In: *Proceedings of the 17th International Conference on Pattern Recognition 2004 (ICPR 2004)*; 2004 Aug 26; Cambridge, UK. p. 10–3. [[CrossRef](#)].
20. Li Q, Lu W, Yang J. A hybrid thresholding algorithm for cloud detection on ground-based color images. *J Atmos Ocean Technol*. 2011;28(10):1286–96. [[CrossRef](#)].
21. Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern*. 1979;9(1):62–6. [[CrossRef](#)].
22. Long CN, Sabburg JM, Calbó J, Pagès D. Retrieving cloud characteristics from ground-based daytime color all-sky images. *J Atmos Ocean Technol*. 2006;23(5):633–52. [[CrossRef](#)].
23. Sankur B. Survey over image thresholding techniques and quantitative performance evaluation. *J Electron Imaging*. 2004;13(1):146. [[CrossRef](#)].