Smart Energy System Research

SCIEPublish

Article

Improving the Short-Term Photovoltaic Power Forecasting Accuracy Based on Deep Learning Hybrid Model

Zahra Jalali Lychaei and Seyed-Masoud Moghaddas-Tafreshi *

Department of Electrical Engineering, Faculty of Engineering, University of Guilan, Rasht 41996-13776, Iran; zahra.jalali144@gmail.com (Z.J.L.)

* Corresponding author. E-mail: tafreshi@guilan.ac.ir (S.-M.M.-T.)

Received: 2 July 2026; Revised: 4 August 2026; Accepted: 2 September 2026; Available online: 15 September 2026

ABSTRACT: The increasing penetration of renewable resources in smart energy systems has increased the need for accurate forecasting of photovoltaic (PV) power to reduce uncertainty and improve operational planning. Rapid changes in atmospheric conditions, especially cloud movement, cause severe fluctuations and pose a significant challenge to PV power forecasting in the short term. In this paper, an integrated Deep Learning-based framework for short-term PV power forecasting is proposed that models the relationships among sky conditions, solar irradiance, and PV power. First, the cloud movement is forecasted based on the Optical Flow algorithm using sky images. Then, the forecasted images are fed into a ResNet50 network to estimate the solar irradiance. Finally, the forecasted solar irradiance, meteorological parameters, and historical PV power data are fed into the CNN-BiGRU-Attention model, whose hyperparameters are optimized using Bayesian Optimization to forecast the PV power for the next hour. The performance of the proposed model was evaluated under different atmospheric conditions and compared with other benchmark models. The results showed that the proposed model achieves high forecasting accuracy, obtaining a R^2 of 0.987 and RMSE, MAE, and MAPE error values of 0.231 Kw, 0.147 Kw, and 8.69%, respectively, and outperforming benchmark models.

Keywords: Photovoltaic power forecasting; Deep learning; Gated recurrent units; Convolutional neural network; Solar irradiance forecasting; Sky images

1. Introduction

The increasing global energy demand and environmental consequences of fossil fuel consumption have increased the attention to renewable energy sources, especially solar energy. However, the intermittent and unstable nature of PV power due to solar irradiance variations makes accurate forecasting of its generation essential to reduce uncertainty, increase grid reliability, and optimize energy system management. Therefore, in order to investigate the performance and feasibility of these systems, a grid-connected rooftop PV system was designed and simulated on the rooftop of the University of Guilan, Iran [1,2].

1.1. Forecasting Cloud Motion Using Sky Images

Analyzing sequential sky images enables the forecasting of cloud movement and shape, enhancing the precision of short-term solar irradiance forecasting [3]. Traditional Cloud Movement Vector (CMV) methods for 1 to 5-h solar irradiance forecasting estimate the movement and dynamics of clouds by analyzing the sequence of satellite images and making forecasts accordingly. But due to the low spatial resolution of these images, the accuracy of the model decreases in rapid atmospheric changes, and CMV faces a challenge in adapting to the fast and nonlinear dynamics of clouds [4,5].

Compared to satellite images, ground-based images provide high-resolution, real-time data on cloud structures. This makes them more suitable for hourly PV power forecasts than satellite images, as they are more sensitive to local radiation changes [6–9]. Recent research has increasingly integrated Deep Learning (DL) and machine vision models to overcome the limitations of CMV. For instance, in [9], the ConvLSTM and TrajGRU architectures were found to enhance short-term forecasting by capturing spatiotemporal dependencies of cloud motion. Also, in [10], unsupervised learning combined with the Lucas–Kanade algorithm has been utilized to extract motion vectors with minimal human labeling. In [11], cloud optical thickness (COT) from multispectral images was used in conjunction with Optical Flow to track cloud masses. In [12], the Lucas–Kanade method was utilized to track cloud motion, even in various atmospheric conditions. Expanding on this, Ref. [13] demonstrated a significant advancement in detecting thin and translucent clouds by utilizing a clear sky library and a two-stage segmentation approach. Although most previous studies have demonstrated the effectiveness of cloud motion tracking methods, limited attention has been paid to utilizing forecasted cloud motion for short-term solar irradiance and PV power forecasting.

1.2. Solar Irradiance Forecasting

New (Machine Learning) ML and DL methods using satellite and sky images have offered innovative solutions for short-term solar irradiance forecasting by capturing cloud motion and spatiotemporal atmospheric patterns [14,15]. In [16,17], short-term solar irradiance was estimated using features extracted from sky images through CNN-based neural networks.

Recent models have focused on learning spatial-temporal dependencies of cloud and irradiance data. In this regard, combining CNN with recurrent architectures such as CNN-LSTM [18], has shown promising performance in short-term solar irradiance forecasting by capturing both spatial and temporal features. In [19], BiLSTM-BiGRU and CNN-LSTM models were used to forecast irradiance in Quetta, Pakistan. Also, the use of attention and Seq2Seq structures, such as the attention-sequence-to-sequence CNN model [20], has further improved the accuracy of the models.

In [21], the use of PredRNN++ and indicators like SBI demonstrated more accurate performance in forecasting GHI, DHI, and DNI at a 10-min horizon. Additionally, Ref. [22] introduced an ATT-CNN model that incorporated an attention mechanism into convolutional channels to emphasize important image regions and improve forecasting accuracy.

A group of studies has used cloud motion vector extraction, estimating cloud movement through methods such as Optical Flow, as an intermediate step for solar irradiance forecasting using ML or LSTM models, particularly for very short-term forecasting [23]. Existing studies mainly focus on direct irradiance forecasting from sky images or meteorological data, whereas the causal relationships among cloud dynamics, solar irradiance, and PV power have not been comprehensively modeled within an integrated DL framework.

1.3. PV Power Forecasting

Statistical models like AR, ARIMA, and SARIMA were first used to analyze PV power time series, proving effective for multi-hour and daily forecasts [24]. Also, other methods have used meteorological parameters such as temperature and solar irradiance, employing the copula technique and Monte Carlo methods to forecast PV power [25].

Statistical models were acceptable for short-term forecasts but were unable to extract complex features of long time series. These limitations prompted the introduction of ML algorithms that improved PV power forecasting by modeling nonlinear relationships and handling large datasets. In [26], a model using a Kohonen self-organizing map (SOM), a feedforward neural network (FFNN), and K-nearest neighbor (KNN) algorithms for PV power forecasting is proposed.

A hybrid model combining variational mode decomposition (VMD) and extreme learning machine (ELM) was introduced for short-term PV power forecasting [27]. In [28], a hybrid MLP-FGWO-PSO approach is presented for PV power forecasting. This model modifies the Gray Wolf Optimization (GWO) technique and integrates it with Particle Swarm Optimization (PSO) and a Multilayer Perceptron (MLP) model.

The emergence of DL with its ability to automatically extract features and model complex nonlinear relationships was a turning point in PV power forecasting [29]. CNNs have been used to identify spatial patterns of radiation and the influence of meteorological parameters [30], and to combine them with wavelet transform (WT) [31] or variable mode decomposition (VMD) [32] for hourly PV power forecasting.

Recurrent neural networks (RNNs) such as LSTM are well suited for PV power forecasting due to their ability to understand temporal dependencies [33,34]. In various studies, hybrid models using methods such as wavelet packet decomposition (WPD) and LSTM [35], or the TFC-RIME-LSTM model, which exploits total factor (TFC) features and generates similar uncertainty scenarios, have been proposed to improve the accuracy and reduce the complexity of PV power forecasting [36]. The LSTM disregards details from both the past and the future in time series forecasting. In order to address this issue, BiLSTM have been developed. The DAFA-BiLSTM [37], which incorporates deep autoregression for more accurate PV power forecasting. Also, the SSA-BO-BiLSTM model [38] improves PV power forecasting by decomposing the time series using Singular Spectrum Analysis (SSA) and using Bayesian optimization (BO) to optimize the hyperparameters of the BiLSTM.

Hybrid models based on CNN-LSTM and ConvLSTM have improved the forecast accuracy, especially in variable weather conditions, by simultaneously extracting spatial and temporal features [35,39]. Subsequently, more advanced models, such as the combination of GNN and BiLSTM, have been developed to model spatial dependencies between power plants and satellite images of clouds [40]. In [41], a short-term PV power forecasting method based on improved CNN-LSTM and cascade learning was presented, in which six effective meteorological features were selected.

GRU is a simplified version of LSTM with faster performance. In [42], a two-stage GRU-Informer and SVR model is proposed for PV power forecasting, considering weather feature screening by XGBoost and weather triple clustering by K-means. The PV power forecasting model combines ET-EEMD and GRU by extracting visual sky features, applying LAB color quantification, using K-means clustering, and selecting effective variables via GRA and PCA, thereby increasing accuracy [43]. In [44], an ultra-short-term PV power forecasting method for single-axis trackers is presented using a radiation transform and a CNN-BiGRU hybrid model, in which atmospheric conditions are classified with FCM, and a model is trained to suit each situation.

The use of sky imagery in short-term forecasting of PV power has been expanded. In [45,46], sky images were used to forecast PV Power using neural networks. In [47], a new DL model for forecasting PV power using sky images was presented. This model connects DL to the physical processes of PVs by introducing the Radiant Transmission Index (RTI) to measure the effects of clouds on solar irradiance. In a study [48], Gaussian mixture probability hypothesis density (GM-PHD)-based a random hypersurface model (RHM) model was used to forecast the movement and time of cloud-sun collision, and the ARIMA and backpropagation (BP) models were used to forecast the PV power.

In [49], a model called SkyNet was proposed, which is an integrated DL architecture for multimodal intra-hourly solar forecasting based on sky images. It can simultaneously perform time, probabilistic, and spatial series forecasts.

In [50], a framework for deterministic and probabilistic PV power forecasting is presented based on the weather regime. First, the sky images are divided into three regimes: sunny, cloudy, and completely cloudy, and cloud cover prediction using the condensed light flux is enabled only in cloudy conditions. Then, the WRF model meteorological data and historical power are fed into a four-stage DL based on quantile regression (QR) to produce point forecasts and calibrated confidence intervals. However, the main focus of this study is on classifying the climate regime and predicting the cloud cover ratio as an auxiliary intermediate variable, rather than explicitly predicting the future sky image or directly extracting radiance from it. In [51], Sky-BiMamba was introduced as an efficient framework for PV power forecasting. This model integrates sky images and PV data in two stages using the Mamba architecture. Sky-BiMamba outperforms the baseline models by extracting cloud texture and motion features and incorporating a time-aware intermodal attention mechanism.

1.4. Limitations of Existing Models and Contributions of This Study

Despite significant advances in PV power forecasting using ML and DL methods, several limitations remain in existing studies. Many studies rely on measured solar irradiance values, meteorological data. As a result, these methods may not accurately capture the effects of rapid changes in cloud movement and the spatiotemporal variability of solar irradiance, and may therefore fail to fully reflect their impact on PV system performance.

Furthermore, although several studies have employed sky images to track cloud movement or directly forecast solar irradiance, integrated frameworks in which future sky conditions are first forecast, solar irradiance is then extracted from the forecast images, and the resulting forecast irradiance is finally used as an input to a PV power forecasting model have rarely been investigated [14–16,18–23,48].

Other studies have used cloud-cover classification as an auxiliary regime indicator [50]. Table 1 summarizes the methodological approaches, key contributions, and limitations of existing studies.

Table 1. Comparison of representative studies with the proposed framework.

Literature	Cloud Motion Forecasting	Solar Irradiance Forecasting	PV Power Forecasting	Integrated Causal Framework	Main Methodological Contribution
Ref. [18]	-	✓	-	-	Hybrid DL for solar irradiance forecasting using sky images.
Ref. [20]	-	✓	-	-	Attention-fused Seq2Seq CNN for solar irradiance forecasting using sky images.
Ref. [23]	✓	✓	-	-	Optical-flow-based cloud-motion extraction for irradiance forecasting.
Ref. [41]	-	-	✓	-	CNN–LSTM combined with cascading learning and meteorological feature selection for PV power forecasting
Ref. [44]	-	✓	✓	-	Ultra-short-term PV forecasting using irradiance transformation, weather classification, and CNN–BiGRU.
Ref. [48]	✓	-	✓	-	GM-PHD-based RHM for cloud dynamics and collision prediction, followed by ARIMA–BP PV forecasting
Ref. [49]	-	✓	-	-	SkyNet for multimodal intra-hour solar forecasting using sky images

Ref. [50]	-	-	✓	-	Regime-aware deterministic and probabilistic ultra-short-term PV forecasting using weather classification and dynamic cloud tracking
Ref. [51]	-	-	✓	-	Sky-BiMamba, a multimodal PV forecasting framework integrating sky images and PV data through time-aware intermodal attention
Proposed Model	✓	✓	✓	✓	Three-stage causal framework integrating cloud-motion prediction using Dense Optical Flow, ResNet50-based irradiance forecasting, and Bayesian-optimized CNN–BiGRU–Attention PV forecasting.

The main innovation of the proposed framework lies in establishing an explicit causal chain between future sky conditions, forecasted solar irradiance, and PV power. Unlike approaches that use measured irradiance directly as input to the power forecasting model, in this framework, the future sky state and cloud movement over the short-term horizon are first forecast; then, the solar irradiance corresponding to the forecasted sky images is estimated; and finally, this forecasted irradiance is transferred to the PV power forecasting stage along with other meteorological variables and historical power information. In this way, the effects of cloud dynamics and movement are taken into account in the forecasting process before they are reflected in changes in measured irradiance and, subsequently, PV power. In other words, by directly linking sky image information to irradiance forecasting and then to power forecasting, this framework creates an integrated and causal forecasting pathway that simultaneously considers spatial and temporal information related to cloud changes.

In the first step, short-term cloud movement and displacement are forecast using the Dense Optical Flow algorithm to generate future sky images. In the second step, the forecasted sky images are fed into a ResNet50 to forecast the solar irradiance. In the third step, the forecasted solar irradiance, along with other meteorological variables and historical PV power values, is fed into a CNN–BiGRU-Attention model to forecast PV power one hour ahead. In addition, to increase forecasting accuracy and improve model tuning, an Attention Mechanism is used to focus on temporal information and relevant features, while Bayesian Optimization is employed to optimize the model hyperparameters. Accordingly, the most important achievements and innovations of this research in addressing the existing gaps in the field of short-term PV power forecasting are summarized as follows:

(1) A novel three-step framework for PV power forecasting is designed by combining sky image data and meteorological data. First, the movement and direction of clouds are modeled; then solar irradiance is forecast based on the resulting images; and finally, PV power for the next hour is estimated using a CNN–BiGRU hybrid model.
(2) By integrating the attention mechanism and Bayesian optimization to adjust hyperparameters, forecasting accuracy is improved, overfitting is prevented, and model stability is improved in variable climate conditions.
(3) The CNN-BiGRU architecture simultaneously extracts spatial and temporal features and models the long-term dependencies of time series data in two directions. It represents the nonlinear and oscillatory behavior of PV power more accurately.

The rest of this paper is organized as follows: In Section 2, a brief explanation of the proposed model framework and the method used to tune its hyperparameters is provided. This section also reviews the overall architecture of the system and explains the theoretical foundations of its key components, including ResNet50, the dense optical flow algorithm, BiGRU, the attention mechanism, and the BO method. In Section 3, the overall simulation process is described, including data collection, preprocessing of sky images, preparation of numerical datasets, outlier identification, and analysis of correlations between meteorological variables and PV power. In Section 4, the structure of the proposed model is explained in detail. In Section 5, the performance of the model under different atmospheric conditions is examined, and its results are compared with benchmark models based on the statistical criteria R^2, RMSE, MAE, and MAPE. In the final section, a summary of the research results is presented, which includes a summary of the most important achievements and references to suggested directions for future research.

2. Proposed Methodology

This section presents a detailed description of the PV power forecasting method based on the proposed model. First, we provide an overview of the system architecture. Then we describe the theoretical foundations and give a brief overview of its components, including ResNet50, the Dense Optical Flow algorithm, BiGRU, attention mechanisms, and Bayesian optimization.

2.1. System Architecture

In this study, a three-stage framework for short-term PV power forecasting is presented, as shown in Figure 1. The first stage is dedicated to cloud motion modeling. In this stage, raw sky images are prepared through preprocessing, including noise removal and contrast enhancement in bright areas, to improve cloud resolution.

Then, the Farneback Dense Optical Flow algorithm is used to estimate the motion vector field between consecutive frames. This motion field represents cloud displacement and is used to forecast spatial changes in the next time step. Based on the motion pattern obtained from the consecutive frames, the future sky image is estimated based on motion information. This output is used as input to the next stage to forecast solar irradiance.

In the second stage, the ResNet50 is used to forecast solar irradiance using sky images. This network forecasts solar irradiance by extracting cloud features from forecast images. It should be noted that the cloud motion modeling and solar irradiance forecasting section in this study is based on the approach presented in Ref. [23]. However, in the present model, improved preprocessing methods for sky images are used, including enhancement of cloud contrast and color normalization based on spectral features. These improvements enhance feature extraction quality and forecasting performance.

In the third step, the output of the previous step, which includes the solar irradiance forecast, is fed to the CNN–BiGRU hybrid model, along with other meteorological variables, to forecast PV power. In this model, CNN extracts and compresses the spatial features of the input data, while BiGRU uses two-way temporal dependencies to consider past and future data. The attention mechanism also dynamically weights the inputs, highlighting important information and preventing the influence of noise and irrelevant data, which ultimately improves the accuracy, stability, and convergence speed of the model. BO is used to determine the optimal configuration of the hyperparameters.

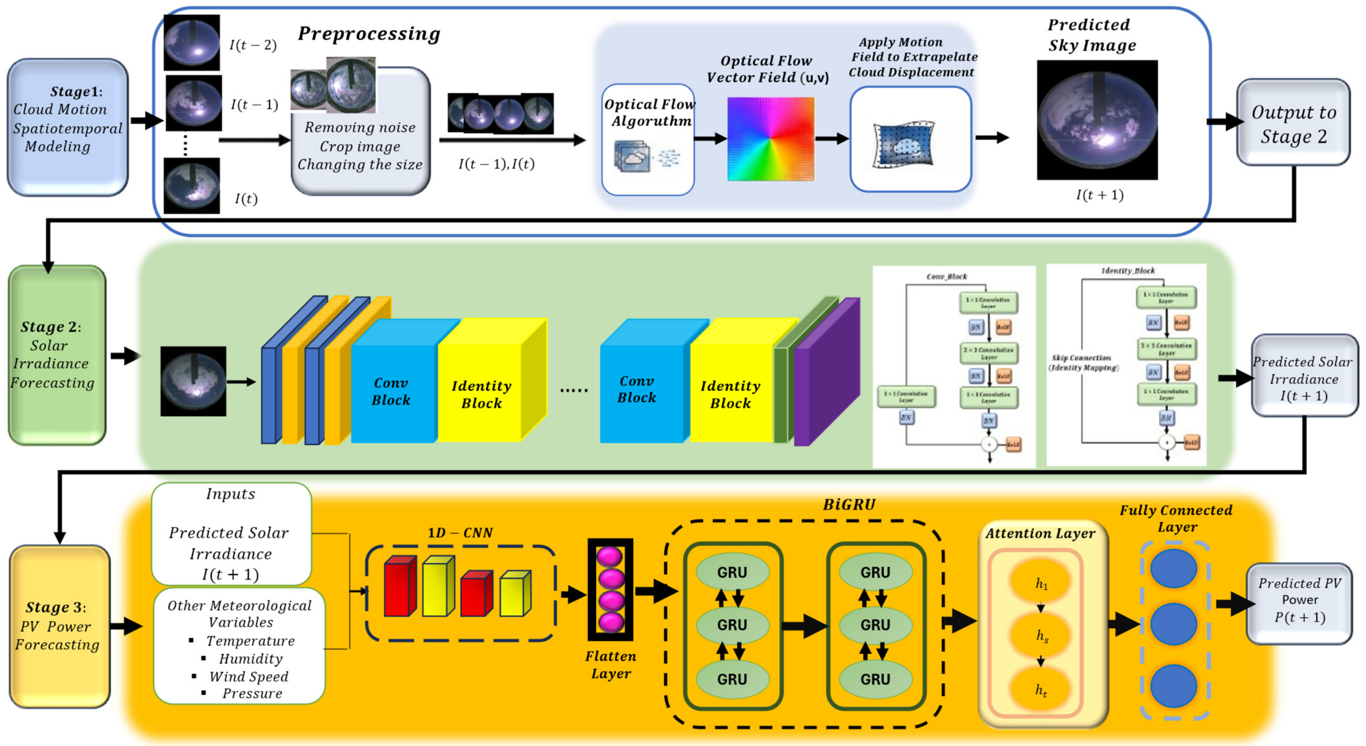


Figure 1. Graphical representation of the overall structure of the fundamental aspects and main content of the study.

2.2. Components of the Proposed Model

2.2.1. Dense Optical Flow Algorithm

The optical flow algorithm estimates pixel motion and the apparent displacement of objects between successive images. The sequence of images enables the forecasting of object movements at future time steps. Optical flow methods attempt to calculate the movement (velocity) between two frames of images taken at times t and $t + \Delta t$. These methods are often based on the assumption of brightness constancy, meaning the intensity of a point in the image does not change as it moves. Under this assumption, if a point at position (x, y) in the image at time *t* with intensity $I(x, y, t)$ moves to a new position $(x + \Delta x, y + \Delta y)$ at time $t + \Delta t$, then the new position of the image is according to Equation (1):

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \tag{1}$$

Assuming small displacements $(\Delta x, \Delta y, \Delta t)$, the intensity at the new position can be expanded using a first-order Taylor series approximation around the original position (x, y, t) as obtain Equation (2):

$$I(x + \Delta x, y + \Delta y, t + \Delta t) \approx I(x, y, t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t \tag{2}$$

Combining this with the brightness constancy assumption (Equation (2)), Equation (3) can be calculated:

$$I(x, y, t) = I(x, y, t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t \tag{3}$$

This can be simplified to the optical flow constraint Equation (4):

$$\frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t = 0 \tag{4}$$

Dividing by Δt (assuming $\Delta t \neq 0$) and considering that as $\Delta t \rightarrow 0$, $\frac{\Delta x}{\Delta t} \rightarrow \frac{dx}{dt} = v_x$ and $\frac{\Delta y}{\Delta t} \rightarrow \frac{dy}{dt} = v_y$, where v_x and v_y are the components of the velocity vector in the x and y directions, respectively, can be obtained by Equation (5):

$$\frac{\partial I}{\partial x} v_x + \frac{\partial I}{\partial y} v_y + \frac{\partial I}{\partial t} = 0 \tag{5}$$

This is a basic formula for optic flow. In this equation, v_x and v_y represent the velocity components of the luminous flux at the point (x, y, t), and $(\frac{\partial I}{\partial X}, \frac{\partial I}{\partial y}, \frac{\partial I}{\partial t})$ are the partial derivatives of the image intensity with respect to the spatial coordinates (x, y) and time (t). These derivatives represent the image gradients in the corresponding directions [52].

To solve the optical flow constraint Equation (6) in practice, a dense optical flow approach proposed by Gunnar Farnebäck is employed. Unlike sparse methods that estimate motion only at selected feature points, Farnebäck's method computes motion vectors for all pixels in the image by modeling local neighborhoods with quadratic polynomial expansions. In this approach, the intensity function in a local window around each pixel is approximated as:

$$I(x, y) \approx X^T A X + b^T X + c \tag{6}$$

where $X = [x, y]^T$, A is a symmetric matrix describing the local curvature of the intensity surface, b is a vector representing the linear component, and c is a constant term. Assuming that the local intensity structure undergoes a displacement (d_x, d_y) between two consecutive frames, the relationship between the polynomial representations at time t and $t + \Delta t$ can be expressed as:

$$I_{t+\Delta t}(x, y) \approx I_t(x - dx, y - dy) \tag{7}$$

The displacement vector (d_x, d_y) is then estimated by minimizing the difference between the polynomial approximations of the two frames, typically in a least-squares sense over a local neighborhood. This results in a dense motion field that provides a robust estimation of optical flow, even in regions with smooth intensity variations.

Thus, the Farnebäck method provides an efficient numerical solution to the optical flow estimation problem defined in Equation (7), enabling the computation of dense velocity fields (v_x, v_y) across the entire image.

In this study, the Farneback Dense Optical Flow algorithm was implemented using OpenCV's calcOpticalFlowFarneback function with the following parameter settings: pyramid scale factor = 0.5, number of pyramid levels = 5, averaging window size = 21, number of iterations at each pyramid level = 10, size of the pixel neighborhood used for polynomial expansion = 7, and standard deviation of the Gaussian used to smooth derivatives = 1.5.

2.2.2. ResNet50

To effectively capture spatial environmental factors affecting PV power, CNNs were introduced by LeCun et al. [53]. However, training deeper CNN architectures often encounters challenges such as vanishing or exploding gradients. To address these issues, ResNet was proposed in 2015, incorporating skip connections or residual blocks that enable direct signal flow to deeper layers, thereby facilitating the training of substantially deeper networks for complex PV forecasting tasks [54].

In this paper, the ResNet50 network is used to forecast solar radiation for 1-h ahead. Although the ResNet50 architecture is typically known for its success in classification tasks, its use in this regression problem of solar irradiance forecasting requires specific structural adjustments, as illustrated in Figure 2. The specifics of the design and operation of each layer in this network are outlined below:

- Convolutional Layer: This initial layer uses 64 filters of size 7 × 7 with a stride 2 to extract local nonlinear features while reducing spatial dimensions and computational cost.
- Batch Normalization Layer: This layer is employed to enhance stability and accelerate training by normalizing layer activations, ensuring consistent learning rates and quicker convergence.
- Activation Function Layer: The Rectified Linear Unit (ReLU) activation function is chosen for its computational simplicity and prevention of vanishing gradients. Since the aim of this study is to forecast solar irradiance, and the output should consist of real, non-negative values, employing a function that sets negative values to zero aligns with the nature of the problem. Potential 'Dying ReLU' issues are mitigated through careful learning rate adjustment, and batch normalization enhances the efficiency and stability of the learning. Specifically, the Dying ReLU problem occurs when the pre-activation input of a neuron consistently becomes negative, causing its output and corresponding gradient to remain zero, thereby preventing weight updates during backpropagation. In this study, Batch Normalization layers play a central role in mitigating this risk by stabilizing the distribution of inputs to each activation function and reducing the likelihood that pre-activation values remain systematically negative. Furthermore, the warm-up and cosine-decay learning rate scheduling strategies generally contribute to the stability of the training process, which can indirectly diminish the risk of neuron deactivation as well.
- Max Pooling Layer: This layer is used to reduce the spatial dimensions (width and height) of the features extracted by the previous layers. A Max Pooling layer with dimensions 3 × 3 and step 2 is implemented at this stage, which reduces the spatial dimensions, reduces computational complexity, and improves the robustness to spatial variations, thereby helping to reduce overfitting. The feature maps obtained from this layer are passed to the following convolution blocks in the network.

ResNet50 leverages residual blocks, incorporating Skip Connections (also known as Identity Mapping), to address performance degradation in deep networks. These connections are crucial for efficient gradient propagation and successful training of deeper architectures. ResNet50 implements residual blocks as "Bottleneck Blocks", each consisting of three convolutional layers:

- 1 × 1 Convolution Layer for channel reduction
- 3 × 3 Convolution Layer for feature extraction
- 1 × 1 Convolution Layer (Dimension Restoration): Another 1 × 1 convolutional layer to restore the number of channels to their original or desired state.

In ResNet50, operations that reduce spatial dimensions (height and width) and increase the number of channels occur sequentially across different stages of the network. These changes typically occur at the beginning of each new stage via a convolution with a stride of 2, thereby halving the spatial dimensions and often doubling the number of channels, to maintain the network's capacity to learn more complex features. The ResNet50 structure is composed of four main stages of residual blocks in addition to its initial layers, culminating in a total of 50 layers with trainable weights:

- Initial Layer: 7 × 7 convolution with 64 filters followed by Batch Normalization and ReLU.
- Stage 1: 3 bottleneck blocks (64, 64, 256), totaling 9 convolutional layers.
- Stage 2: 4 bottleneck blocks (128, 128, 512), totaling 12 convolutional layers.
- Stage 3: 6 bottleneck blocks (256, 256, 1024), totaling 18 convolutional layers.
- Stage 4: 3 bottleneck blocks (512, 512, 2048), totaling 9 convolutional layers.

For regression, the final classification head is replaced with an Average Pooling layer followed by a Fully Connected layer with one neuron and linear activation to produce continuous solar irradiance forecating [54].

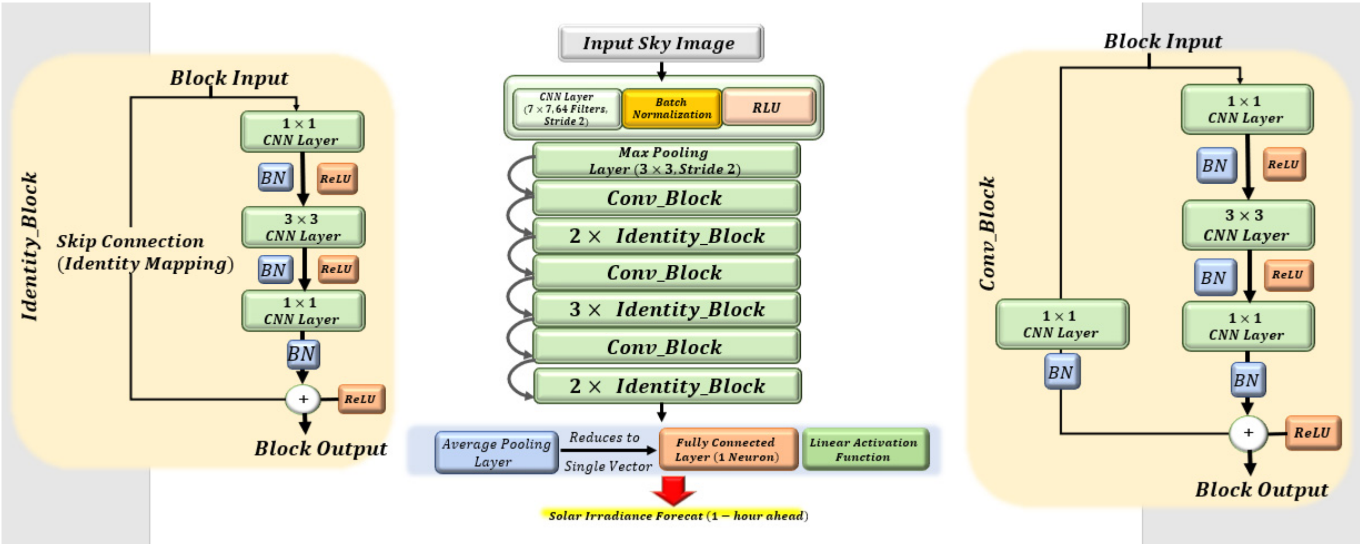


Figure 2. Overview of the ResNet50 model architecture.

ResNet50 Implementation and Training Configuration

The ResNet50 architecture was optimized using AdamW with a weight decay of 1×10^{-4}. A warm-up followed by a cosine-decay learning-rate schedule was employed, with an initial learning rate of 1×10^{-3}, a minimum learning rate of 1×10^{-6}, and a 5-epoch warm-up period. Huber loss ($\delta = 1$) was used as the loss function, selected over standard MSE due to its robustness to outliers commonly present in solar irradiance data caused by abrupt cloud movement, while MAE and RMSE were monitored during training.

The model was trained for a maximum of 100 epochs with a batch size of 16. Early stopping was applied based on validation loss with a patience of 15 epochs, and the weights corresponding to the best validation performance were restored.

2.2.3. Bidirectional Gated Recurrent Unit (BiGRU)

RNNs process time series datasets, they face challenges with vanishing or exploding gradients. LSTM [55] and GRU [56] are types of RNNs that solve this problem. LSTMs use three gates to control the flow of information. A simpler alternative to LSTMs, GRUs have fewer parameters, enabling more efficient computation and faster training. The GRU has two gates; the update gate, which determines how much new information to keep, and the reset gate, which decides how much old information to discard.

The GRU only extracts information in one direction, while the bidirectional GRU processes data in both forward and backward directions and extracts more complete information from the time series.

2.2.4. Attention Mechanism

Attention mechanisms enhance the accuracy of neural networks by enabling them to selectively weight and focus on crucial parts of the input data [57]. The CNN-BiGRU model, utilizing the self-attention mechanism, dynamically determines the importance of different parts of the sequence while extracting spatial features with the CNN and temporal dependencies with the BiGRU. By focusing on relevant information and modeling long-term dependencies, this mechanism reduces the limitation of fixed-length context vectors in traditional encoder-decoder architectures and can lead to improved information extraction and forecasting accuracy.

The combined feature vector generated by CNN and BiGRU layers is fed into the self-attention mechanism. The output of the CNN-BiGRU hidden layer can be denoted as H = $\{h_1, h_2, .., h_t\}$, where each h_i is a vector of dimension. This vector is nonlinearly transformed to obtain a set of raw attention scores

$E = [e_1, e_2, e_3, \ldots, e_n]$, where each e_i represents the relevance of h_i to other hidden states within the same sequence.

The CNN layers capture local patterns from both the PV power time series and environmental inputs, while the BiGRU layers model long-range temporal dependencies. As a result, each hidden state h_i encodes rich contextual information that reflects both temporal and environmental influences. The self-attention mechanism then assigns adaptive weights to these hidden states, allowing the model to focus on the most informative time steps for accurate forecasting. To achieve this, the self-attention mechanism is used to generate an attention weight vector $[\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_n]$, which helps determine the significance of each intermediate state in the forecasting process [57]. The detailed formula is given as Equations (8) and (10):

$$e_i = \tanh(w_h h_i + b_h), e_i \in [-1, 1] \quad (8)$$

$$\alpha_i = \frac{\exp(e_i)}{\sum_{i=1}^{t} \exp(e_i)}, \sum_{i=1}^{t} \alpha_i = 1 \quad (9)$$

where w_h is the weight matrix of h_i, and b_h is the bias. The attention vector $H' = \{h'_1, h'_2, \ldots, h'_t\}$ can be can be calculated by multiplying attention weight α_i and h_i.

$$h'_i = \alpha_i \cdot h_i \quad (10)$$

The resulting attention-weighted vectors, $H' = \{h'_1, h'_2, \ldots, h'_t\}$, are aggregated typically through summation or averaging to form a fixed-size context vector emphasizing the most relevant information in the input. This context vector is then passed to a fully connected forecasting layer (Dense layer) to generate the final PV power forecast.

2.2.5. Bayesian Optimization

BO was used to determine the optimal hyperparameter values. BO is particularly suitable for black-box objective functions and high-dimensional hyperparameter spaces, as it balances exploration and exploitation by modeling the relationship between hyperparameters and model performance. BO is a sequential optimization method that starts with a surrogate model of the objective function and iteratively updates it using new evaluation points to identify the best hyperparameter settings [58].

For optimization, a Gaussian Process (GP) was used as a surrogate model to estimate the relationship between the CNN-BiGRU model's hyperparameters and the validation error. The Expected Improvement (EI) acquisition function was used to select the next configuration; this function balances the selection of configurations with low error (exploitation) against the exploration of less-tested points in the search space (exploration). At each iteration, the surrogate model was updated with the new result, and the next configuration was selected accordingly.

The evaluation criterion was the validation RMSE, and to avoid information leakage, the test data were not used at this step. The process started with 10 initial random configurations and then continued for another 30 Bayesian-guided evaluations (up to 40 configurations in total). Each configuration was trained for a maximum of 100 epochs, with early stopping applied after 10 epochs without improvement to prevent overfitting.

Furthermore, if the best RMSE did not improve by at least 0.0001 over 10 consecutive evaluations, the process was stopped early. The entire search took 2595 s, and in the end, the configuration with the lowest validation RMSE was selected to train the final model. The BO algorithm settings used for the proposed CNN-BiGRU model are summarized in Table 2.

Table 2. Settings used for the BO of the model.

Component	Setting
Acquisition function	Expected Improvement (EI), based on a Gaussian Process (GP) surrogate model
Optimization iterations	10 initial (random) evaluations + up to 30 additional Bayesian-guided evaluations, for a maximum total of 40 configurations
Stopping criteria	Terminated upon reaching (i) the maximum of 40 evaluations, or (ii) no improvement $\geq 10^{-4}$ in validation RMSE over 10 consecutive evaluations, whichever occurred first
Computational budget	NVIDIA Tesla T4 GPU (Google Colab); total wall-clock time 2595 s

3. The Architecture of the Proposed Model

CNN-BiGRU

The proposed model consists of CNN and BiGRU based on a self-attention mechanism. It is used to forecast PV power for 1-h ahead. This model processes the input data in three stages: First, a two-layer 1D-CNN extracts spatial features from the sequential PV power data. Next, a two-layer BiGRU processes the output to capture temporal dependencies. After the BiGRU layer, a self-attention mechanism is integrated. BiGRU neural networks are well-suited for accurate, high-resolution PV power forecasting because they can process sequential data while maintaining long-term dependencies. The mechanism assigns weights to different time stamps, emphasizing the most relevant parts of the input sequence and providing more accurate forecasting. The final multi-step output is generated by a fully connected (FC). The overall model design integrates spatial and temporal information processing to deliver a high-resolution, day-ahead PV power forecast. BO was used to select the optimal values, including the number of CNN filters, CNN kernel size, BiGRU units, dropout rate, and batch size from the search space shown in Table 3 to construct the proposed model. The description of each layer of the model is provided in the following explanation:

(1) Input layer: The dataset contains five additional features. Four of these input variables are weather parameters such as temperature, humidity, wind speed, and solar irradiance forecasted. The fifth variable is the corresponding PV power. After pre-processing, they are used as input for the forecast model and then fed into the CNN layer. To prepare the input data for the proposed model, a sliding window approach with a fixed length of 24 h was employed. In this scheme, at each time step *t*, the input vector fed to the model consists of the values of five variables solar irradiance (G), wind speed (W), humidity (H), temperature (T), and PV power (P) corresponding to the past 24 h from $t-24$ to $t-1$. These values are arranged as a two-dimensional matrix (number of features × window length) and passed to the CNN layer. The corresponding output of this window is the forecasted PV power for the next hour, $\hat{P}_t$. This procedure is repeated sequentially across the entire test period, generating a one-step-ahead forecast at each iteration $(\hat{P}_{t+1}, \hat{P}_{t+2}, \ldots, \hat{P}_{t+k})$. Figure 3 illustrates this sliding window mechanism.

(2) CNN layer: This model consists of two convolutional layers, two MaxPooling layers, and one Flatten layer. MaxPooling layers with a window size of 3 and the same padding are used to reduce the dimensions and number of parameters. The ReLU activation function is used in the convolutional layers. Finally, the extracted features after passing through the Flatten layer are sent as input to the BiGRU model.

(3) BiGRU layer: Two BiGRU layers are utilized to extract temporal features. ReLU is applied as the activation function for these layers. The output of the last BiGRU is then passed on to the attention layer.

(4) Attention layer: The network consists of two DL models that generate a combined feature vector. However, this combined vector contains redundant information, making the network computationally intensive and leading to poor convergence and limited performance. To address this issue, an attention layer is introduced. This layer improves convergence and boosts the performance of the model by filtering out unnecessary information and focusing on the key data.

(5) Dense layer: The optimal output of the attention layer is then fed into a fully connected layer, which consists of a single neuron in a dense layer.

In this study, the Adam optimizer was used to train the model parameters, together with MSE as the loss function. Instead of a fixed learning rate, a two-stage step-based learning rate schedule was employed via a custom learning rate scheduler: the learning rate was set to 1×10^{-5} for the first 50 epochs and increased to 1×10^{-4} for the remaining epochs. To avoid overfitting or underfitting, an early stopping strategy was implemented, monitoring the validation loss with a patience of 10 epochs; if no improvement was observed within this window, training was halted, and the weights corresponding to the best validation performance were restored.

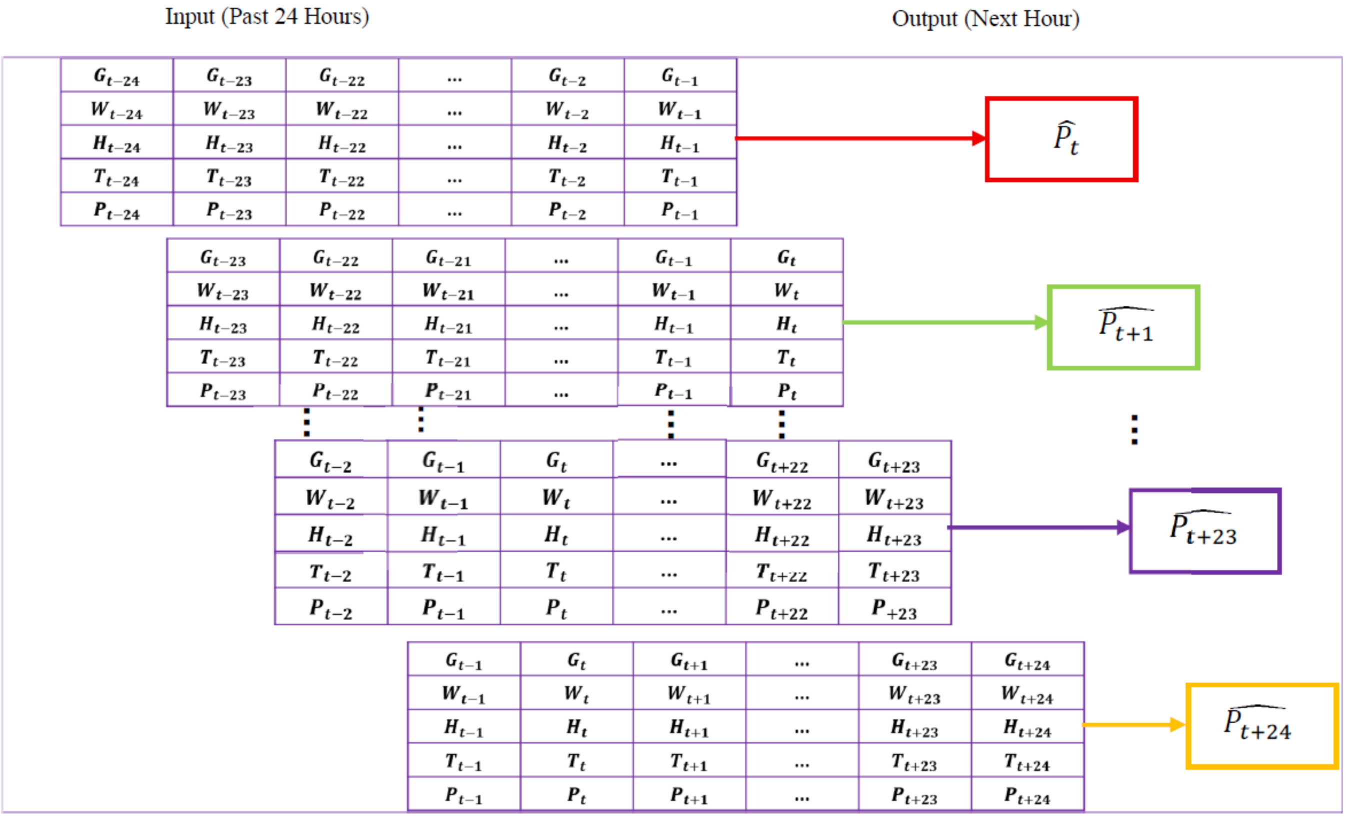


Figure 3. Schematic of sliding time window.

Table 3. Using Bayesian Optimization to tune the hyperparameters of the proposed model.

Type	Filter Size	Kernel Size	Number of Units	Best Hyperparameters
1D-CNN	[16, 256]	[2, 10]	-	Filter Size = 210, Kernel Size = 3
BiGRU	-	-	[16, 100]	Number of Units = 33
Dropout rate	-	-	[0.1, 0.5]	Dropout rate = 0.28
Batch size	-	-	[16, 64]	Batch Size = 32
Epochs	-	-	-	Epochs = 100

4. Overall Procedure of the Simulation

The proposed hybrid model consists of five main steps: data collection and processing, cloud motion Forecasting with the Optical Flow algorithm, solar irradiance forecasting with ResNet50, and PV power forecasting with CNN-BiGRU. The proposed model was implemented in Python (version 3.9.11) using Keras (version 2.9.0) with TensorFlow (version 2.9.1) as the DL framework. Initial development and debugging were performed on a local system running Windows 10 with a GeForce MX150 GPU. However, all final model training, hyperparameter optimization, and reported results were obtained using Google Colab with an NVIDIA Tesla T4 GPU.

4.1. Data Collection

The data for this study were obtained from the National Renewable Energy Laboratory (NREL) website and included sky images recorded by ground-based cameras, meteorological parameters (humidity, solar irradiance, wind speed, and temperature), and the PV power of a 10 MW system. Data from (1 January 2017 to 31 December 2017) were used for training and validation, where the first two months of each season were considered as training data and the last month of each season as validation data. The dataset for 2018 was reserved for testing. The 10-min data were resampled to hourly intervals to reduce computational load, and sky images and numerical data were preprocessed separately. A fixed random seed of 42 was used consistently across all stages of the framework.

4.1.1. Preprocessing of Sky Images

Image preprocessing was performed using the Pillow and OpenCV libraries in Python programming. Figure 4 shows the overall steps of this process. The first step involves resizing the original raw images, which are of 288 × 352 pixels, to 500 × 500 pixels. The next step is to eliminate irrelevant areas. In some images, areas outside the circular camera frame may show cloud-like colors. During training, these areas could be mistakenly identified as clouds. The next step is noise removal.

Sky images contain noise, which reduces the accuracy of cloud motion forecasting using the dense optical flow algorithm. In this study, a combination of morphological operations and noise removal methods (such as fast averaging) is used to reduce noise. Contrast enhancement and brightening were used to highlight cloud structures and improve their tracking across successive frames. This preprocessing allowed the Optical Flow algorithm to extract cloud motion vectors more accurately. Finally, suitable preprocessed images are extracted as input to the dense optical flow algorithm for cloud motion tracking.

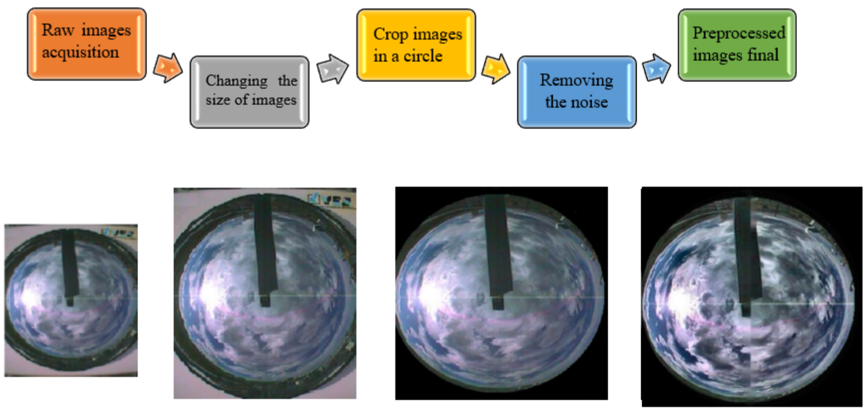


Figure 4. Image preprocessing steps.

4.1.2. Numerical Dataset and Preprocessing

The dataset consists of historical meteorological parameters and PV power. The data were preprocessed using normalization, missing-value imputation, and outlier removal techniques. Table 4 shows the number of training, validation, and test samples after the final data preprocessing.

Data Normalization

To enhance the effectiveness of the DL model, min-max normalization was implemented. The objective of this method is to convert the value of a characteristic into a range from 0 to 1 [59].

Missing Values

Due to device errors, the PV power datasets contain missing values. To address this issue, the imputation technique of K-Nearest Neighbors (KNN) was applied. KNN identifies similar examples in the training dataset and averages their values to replace the missing data point. In this study, this technique is implemented in Python using the KNN imputation from the scikit-learn library, with the number of neighbors set to 3.

Detection & Removal of Outliers

Outliers are values that can indicate measurement errors or unusual patterns. Identifying, removing, or adjusting these data points is a crucial step in data preprocessing to avoid unbalanced and biased results in statistical analyses and modeling. In this paper, tools such as Pandas and Scikit-learn libraries, which provide powerful techniques in this field, have been used to identify and manage outliers. Additionally, the Z-score method has been used as an effective approach to detect such data. The Z-score method is based on the standard deviation and assumes that the data have an approximately normal distribution. In this study, the Z-score was used to detect outliers. The Z-score is a statistical tool that indicates how far a data point (x_i) is from the mean (μ) of the data set in terms of the standard deviation (σ). For each data point (x_i), the Z-score is calculated using Equation (11):

$$Z_i = \frac{(x_i - \mu)}{\sigma} \tag{11}$$

where x_i is the individual data point, μ is the mean, and σ is the standard deviation of the dataset. To identify outliers, a threshold was defined based on the Z-score values. Data points with a Z-score above this threshold were classified as outliers. Specifically, in this study, a threshold of ±1.96 standard deviations from the mean was used. This criterion indicates that data points outside this range (in either direction) were flagged as potential outliers. In a standard normal distribution, about 5% of the data fall outside this range, which corresponds to a significance level of 0.05 in a two-tailed Z-test. The upper and lower threshold intervals for outlier identification are presented in Equations (12) and (13), respectively:

$$\mu + (1.96 \times \sigma) \tag{12}$$

$$\mu - (1.96 \times \sigma) \tag{13}$$

Table 4. Total number of valid samples in each of the train, validation, and test.

Split	The Number of Datasets	Time Window
Training data	2932	1 January 2017 07:00:00 → 31 December 2017 19:00:00
Validation data	1429	-
Test data	4290	1 January 2018 08:00:00 → 31 December 2018 20:00:00

4.1.3. Analaysing Correlations Between Meteorological Parameters and PV Power Output

The three-dimensional plots in Figure 5 show the statistical analysis of the relationships between humidity, irradiance, temperature, wind speed, and the PV power. In all sub-figures, the x, y, and z axes represent environmental variables and the PV power. In Figure 5a, the highest PV power is generally observed under conditions of low humidity and moderate wind speed, indicating the nonlinear behavior of the effect of wind speed on the system performance. Figure 5b shows that solar irradiance is the most important factor affecting the power, while humidity plays a secondary and condition-dependent role at high irradiance levels. In Figure 5c, an increase in temperature under high irradiance conditions leads to a decrease in PV power, while at low irradiance levels, the effect of temperature is less noticeable. Figure 5d shows that wind speed under high irradiance conditions can improve the system performance to some extent, which is probably due to the cooling effect of convective flow. In Figure 5e, the interaction between wind speed and temperature is observed, such that increasing wind speed can reduce the negative effect of high temperature to some extent. Figure 5f shows that the PV power is usually higher under low temperature and low humidity conditions, while the simultaneous increase of these two variables causes a decrease in the power.

Table 5 shows the Pearson correlation coefficient and variance inflation factor (VIF) of each meteorological input feature and historical PV power. Solar irradiance has the highest linear correlation with PV power (r = 0.964). However, the VIF value calculated for solar irradiance (16.368), indicates the existence of significant multiple collinearity between this feature and other inputs, especially PV power. In contrast, temperature (VIF = 1.744), humidity (VIF = 1.511), and wind speed (VIF = 1.168) have low VIF values, indicating that these variables add relatively independent and non-additive information to the model and, despite a weaker linear correlation with PV power (r = 0.181, −0.195, and 0.325, respectively), play a complementary role in the forecasting process. To quantitatively assess the impact of this multicollinearity on forecasting performance, an ablation study was conducted in Section 5.4.2, in which the model's accuracy was compared across configurations with and without the high-VIF feature.

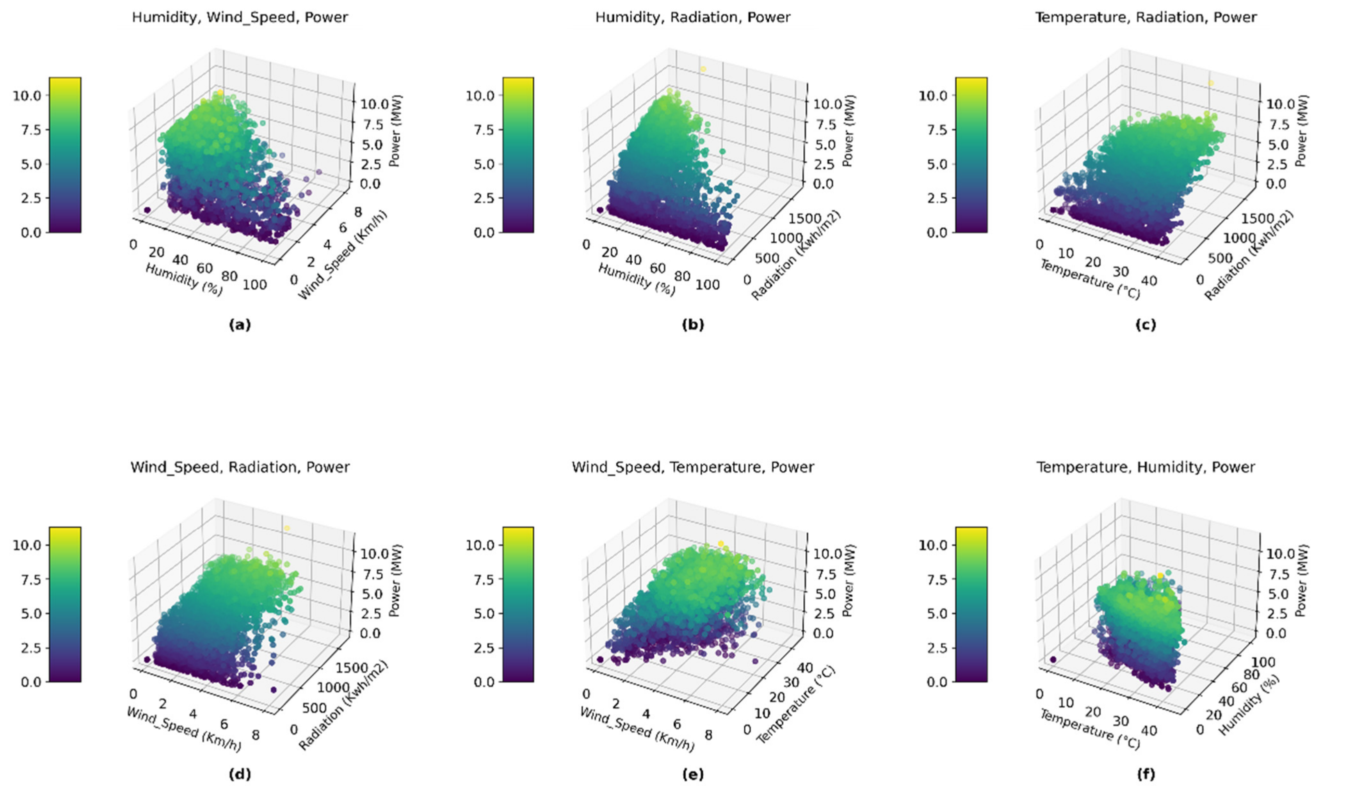


Figure 5. Statistical analysis of the impact of meteorological variables on PV power. (**a**) Power variations versus humidity and wind speed; (**b**) power variations versus humidity and irradiance; (**c**) power variations versus temperature and irradiance; (**d**) power variations versus wind speed and irradiance; (**e**) power variations versus wind speed and temperature; (**f**) power variations versus temperature and humidity.

Table 5. Pearson correlation with PV power and Variance Inflation Factor (VIF) of meteorological input features.

Feature	Correlation with PV Power (r)	VIF
Solar Irradiance	0.964	16.368
Temperature	0.181	1.744
Humidity	−0.195	1.511
Wind Speed	0.325	1.168

5. Forecasting Results of the Proposed Model

This section presents a comprehensive analysis of the simulation results obtained from the proposed model. The model consists of three stages. The first stage focuses on forecasting cloud tracking and displacement over a 1-h-ahead horizon using the Dense Optical Flow algorithm. The second stage employs the ResNet50 neural network to forecast solar irradiance. Finally, the third stage leverages a CNN-BiGRU-Attention mechanism architecture to forecast the PV power.

5.1. Performance Metrics

The effectiveness of the proposed models is assessed using four standard forecasting metrics: Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Coefficient of Determination (R^2). These evaluation metrics are mathematically expressed in Equations (14)–(17).

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}_i| \tag{14}$$

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{15}$$

$$\text{MAPE} = \frac{100\%}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i - y_i|}{y_i} \tag{16}$$

$$R^2 = 1 - \frac{\sum_i(\hat{y}_i - \bar{y}_i)^2}{\sum_i(y_i - \bar{y}_i)^2} \tag{17}$$

Computational Complexity Analysis

A comprehensive analysis of the computational cost of the steps is presented in Table 6. The proposed framework consists of three consecutive steps. For each step, the training time, inference time, hardware requirements, computational cost, and model complexity are examined.

The first step, Dense Optical Flow, is a classical image processing-based method and does not require a training process. Therefore, there is no cost for training this part, and only the computational time related to estimating the motion field and generating the forecasted sky image is calculated in the inference step.

In the second step, the ResNet50 model has a deep structure and a significant number of parameters, but the training process is performed only once, and in the exploitation step, its inference time is 0.009 s for each image. The computational cost of this model is 39,800 MFLOPs, and its number of parameters is 25,589,761 million.

The results in the table show that the CNN-BiGRU-Attention model has a much lower computational cost than ResNet50, with only 0.81 MFLOPs and 1920.52 trainable parameters, and an inference time of 0.005 s per instance.

Table 6. Computational complexity and efficiency of each stage of the proposed framework.

Stage	Type	Training Time (Total/per Epoch)	Inference Time (Each Sample)	GPU	Computational Cost	Model Complexity
1. Dense Optical Flow	Classical (non-trainable)	-	0.02	GPU T4	-	-
2. ResNet50	Deep Learning	12,600 s	0.009	GPU T4	39,800 $MFLOPs$	Total parameters = 25,589,761
3. CNN_BiGRU-Attention	Deep Learning	1920.52 s	0.005	GPU T4	0.81 $MFLOPs$	Total parameters = 204,511

5.2. Results from the Proposed Model's First Stage: Cloud Movement Tracking with Dense Optical Flow

Figure 6 shows the qualitative evaluation of the proposed cloud motion forecasting framework based on the Dense Optical Flow algorithm in four different sky cover conditions, including clear, partly cloudy, completely cloudy, and mostly cloudy. In each case, two consecutive sky images at $t - \Delta t$ and t are used as input to estimate the cloud motion. The extracted motion vectors are used to estimate cloud displacement and generate the forecasted cloud image for the next time step. To enhance visual interpretation, Dense Optical Flow is represented as a color map and the flow-vector magnitudes as a spatial intensity map, enabling the identification of local variations in cloud motion speed and direction. The third column of the figure shows the cloud motion field estimated by the Dense Optical Flow algorithm, where color changes indicate differences in the direction and magnitude of cloud movement. The fourth column shows the Optical Flow Magnitude Map, which identifies areas of more intense cloud motion. Higher values correspond to larger estimated cloud displacements between two consecutive images, while lower values indicate areas of limited or near-stationary motion.

Under clear-sky conditions, the forecast image closely matches the actual image due to the absence of significant cloud structures and limited motion patterns. Under fully cloudy conditions, the relatively homogeneous cloud cover and smoother spatial variations facilitate the estimation process and reduce the differences between forecasted and actual images. More complex cloud behavior is observed in partly cloudy and mostly cloudy conditions.

In these conditions, the presence of heterogeneous structures provides suitable spatial features for optical flow estimation. The forecasted images in these conditions preserve the dominant cloud structures and their spatial distribution with reasonable accuracy. The visual comparison of the forecasted and actual images in the last column shows acceptable agreement. The results indicate that the forecasted cloud field largely preserves the overall morphology, spatial continuity, and movement of cloud structures. Although slight differences are observed in cloud boundary regions and areas affected by rapid local changes, the overall cloud evolution is captured satisfactorily. These results indicate the appropriate capacity of the proposed framework for application in short-term PV power forecasting systems based on sky images.

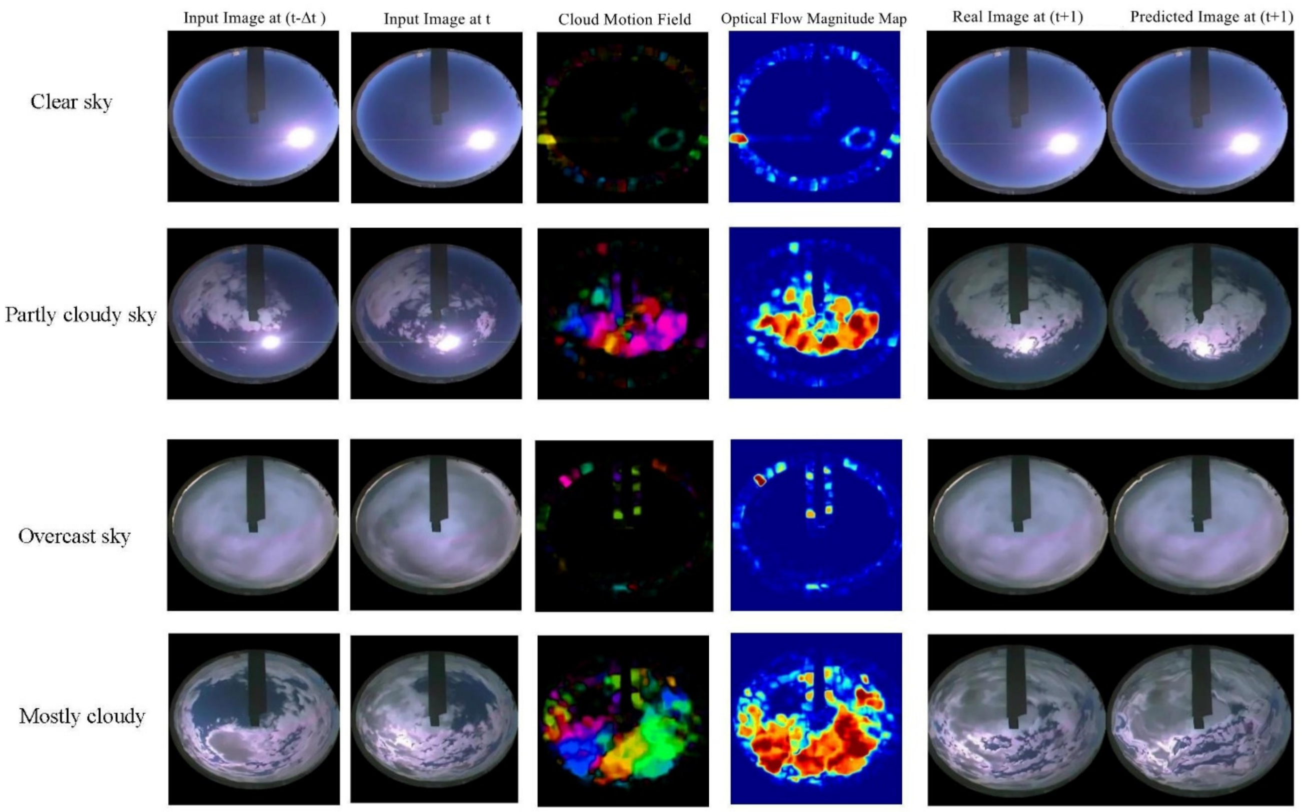


Figure 6. The results of Dense Optical Flow under different conditions (clear, partly cloudy, cloudy, and mostly cloudy) are shown. For each condition, the input image, heat map, final output, and a comparison between the forecasted and actual images are presented.

5.3. Results from the Proposed Model's Second Stage: Solar Irradiance Forecasting Using the ResNet50

The validation results of the ResNet50 model in four weather conditions, including clear sky, cloudy, partly cloudy, and rainy, are presented in the Figure 7a–d; so that the actual and forecasted values of solar irradiance are displayed for each scenario and Figure 7e also presents the comparison of the MAE, RMSE, and R^2 indices. In clear sky conditions, the model shows the highest accuracy, and a very good mapping between the actual and forecasted values is observed, such that the lowest error values (MAE = 4.20 and RMSE = 5.50 W/m^2) and the highest R^2 = 0.92 are obtained. In cloudy conditions, the model performance is still satisfactory, although due to temporal changes in cloud cover and increased fluctuations in solar irradiance, the error rate increased slightly and the MAE and RMSE values reached 6.84 and 7.56 W/m^2, respectively.

As can be seen in Figure 7c, in partly cloudy conditions, due to intermittent changes in cloud cover and rapid fluctuations in solar irradiance, the difference between the actual and forecasted values increases, especially in the middle hours of the day. This is confirmed by the increase in MAE and RMSE to 10.81 and 14.74 W/m^2, respectively, and the decrease in the R^2 to 0.82 in Figure 7e. The highest forecasting error is observed in rainy conditions, especially in the time interval from 09:00 to 11:00, where MAE and RMSE reach 18.68 and 20.73 W/m^2, and the R^2 value decreases to 0.80. This performance loss can be attributed to the unstable nature of precipitation, the presence of dense clouds, and the complex effects of atmospheric humidity on solar irradiance. Despite this, the model can still track the overall trend in irradiance changes, especially between 12:00 and 15:00, with acceptable accuracy.

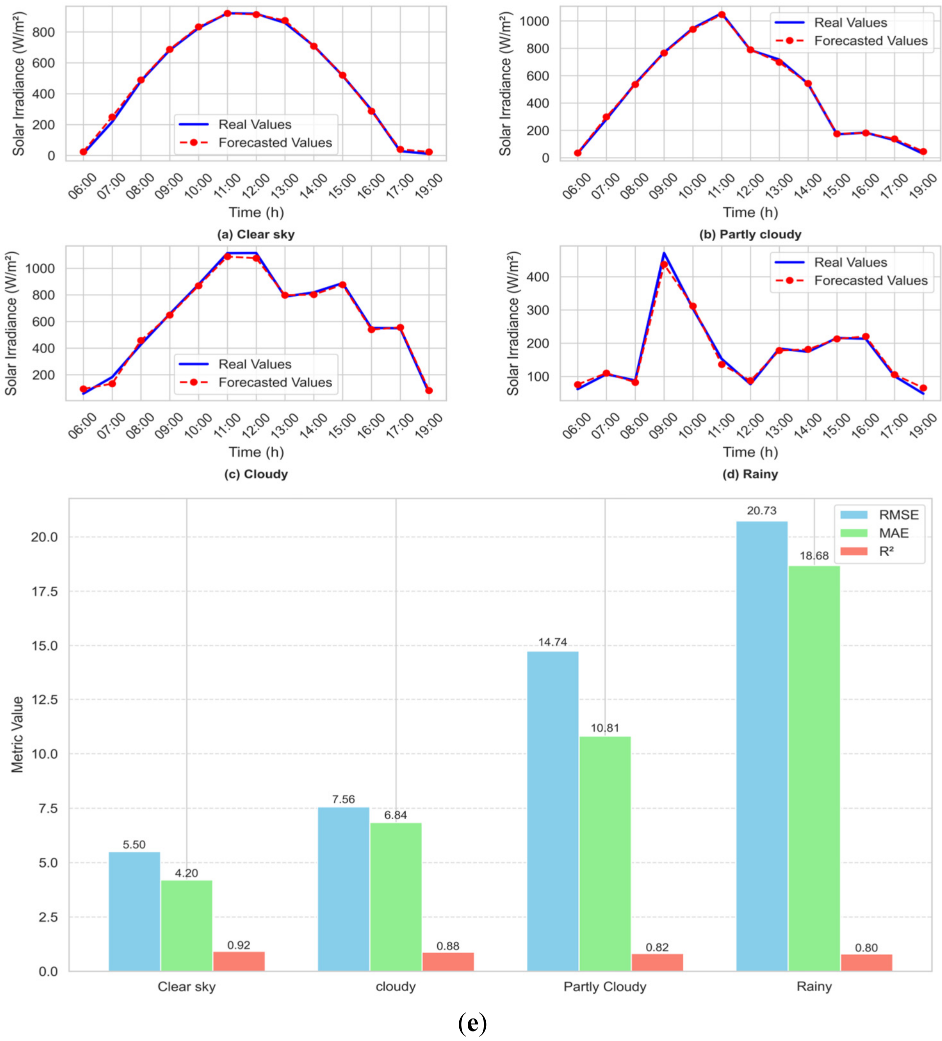


Figure 7. Comparison of real and forecasted solar irradiance values under different weather conditions: (**a**) clear sky, (**b**) cloudy, (**c**) partly cloudy, and (**d**) rainy conditions; (**e**) comparison of statistical performance metrics (MAE, RMSE, and R^2).

Model Performance Comparison

The results of comparing the proposed model (Model 2) with the reference model [23] (Model 1) in Figure 8 and Table 7 show that both models have a high correlation between the actual and forecasted values of solar irradiance, but the proposed model provides higher accuracy with more data density around the ideal line ($y = x$) and less dispersion. The analysis of the KDE plot also indicates that the error distribution of the proposed model has a narrower and more concentrated peak around zero, lower variance and a lower probability of large errors. The quantitative results in Table 7 confirm these observations; so that the RMSE, MAE, and MAPE values have decreased from 20.326 to 7.702 W/m^2, from 9.322 to 5.045 W/m^2 and from 3.455% to 1.921%, respectively. Also, R^2 has increased from 0.94 to 0.99. These results show that the proposed model, despite the complexity and variability of atmospheric conditions, has greater accuracy and stability in forecasting solar irradiance than the (Model 1).

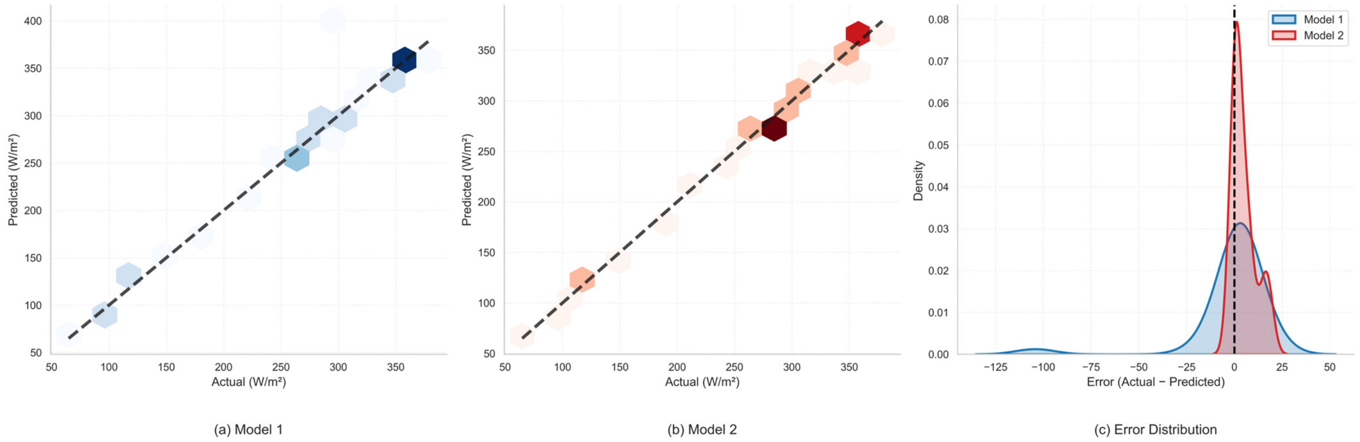


Figure 8. Comparison of the performance of the proposed model and the model presented in the study [26] in forecasting daily solar irradiance in June. (**a**) Forecasted versus actual values for Model 1; (**b**) Forecasted versus actual values for Model 2; (**c**) Distribution of forecast errors for Model 1 and Model 2.

Table 7. Comparison of forecasting performance metrics between the proposed model (Model 2) and the reference model [26] (Model 1).

	RMSE	MAE	MAPE (%)	R^2
Model 1	20.326	9.322	3.455	0.94
Model 2	7.702	5.045	1.921	0.99

5.4. Results from the Proposed Model's Third Stage: PV Power Forecasting Using CNN-BiGRU

In this section, the performance of the proposed model is evaluated. Figure 9 compares the actual and forecasted PV power over five days (one hour ahead), demonstrating that the model accurately captures overall trends and daily variations. Table 8 lists statistical indices (RMSE, MAPE, MAE, and R^2) for each day.

The results demonstrate that the model accurately forecasted daily trends and overall PV power generation patterns over five days. Particularly during peak hours of solar irradiance, the model forecasted production peaks consistent with actual values, with only minor instantaneous deviations. The model also responded well to sudden atmospheric changes, such as cloud cover. The forecasted PV power values were more accurate on sunny days (3 April, MAPE = 1.434%) than on cloudy days (4 April, MAPE = 4.199%).

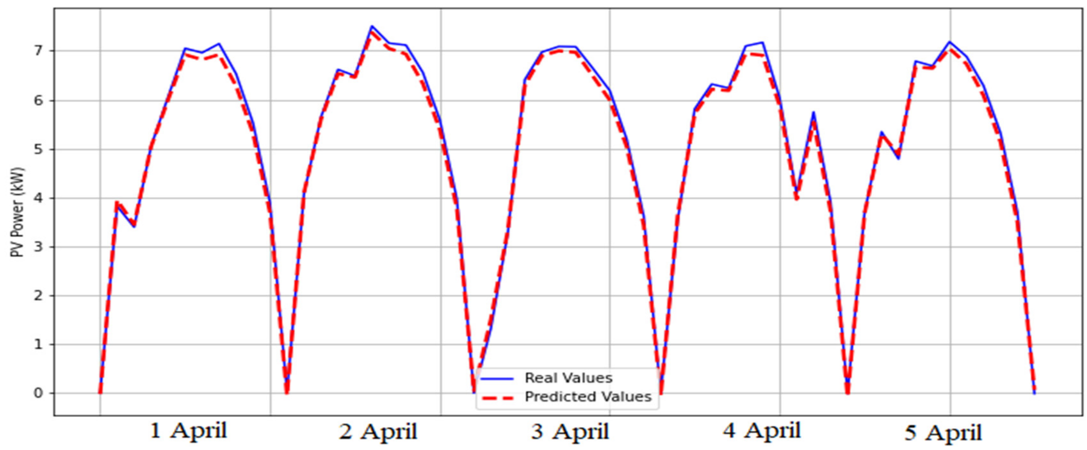


Figure 9. PV Power Forecasting Results and Evaluation Metrics over Five Days.

Table 8. Performance evaluations of the proposed model for seven days consecutive.

Days	RMSE	MAE	MAPE (%)	R^2
1 April	0.048	0.045	1.221	0.9996
2 April	0.0507	0.047	1.804	0.9995
3 April	0.0398	0.0326	1.434	0.9995
4 April	0.0784	0.072	4.199	0.988
5 April	0.1078	0.0966	4.178	0.97

5.4.1. Accuracy Comparison with the Other Developed Deep Learning Models

To evaluate the performance of the proposed model in forecasting PV power for one hour ahead, its results were compared with eight benchmark models (GRU, BiGRU, CNN-GRU, CNN-BiGRU, LSTM, BiLSTM, Transformer, and Informer). Among them, CNN-BiGRU is the closest basic architecture to the proposed model but lacks the Attention mechanism; therefore, comparing these two models allows for a direct assessment of the contribution of the attention mechanism in improving the forecast accuracy. Also, adding Transformer and Informer as independent self-attention architectures allows for the assessment of whether the superiority of the proposed model is solely due to the presence of the attention mechanism or is derived from the simultaneous combination of spatial feature extraction CNN, BiGRU, and attention.

The details of the architectures and optimal hyperparameters of these models, all of which were tuned using BO, are presented in Tables A1–A6 of the Appendix. All models were trained using the same training dataset and a common preprocessing pipeline to ensure a fair and reliable comparison with the proposed model.

The monthly results are presented in Table 9, and their visual representation is shown in Figure 10. The proposed model achieved the lowest MAE and MAPE values in most months, particularly from February to September and in November. In some months, CNN-BiGRU performs slightly better; however, the annual average MAPE of the proposed model (3.14%) is lower than CNN-BiGRU (3.38%), Transformer (6.43%), and Informer (18.42%). Also, the proposed model obtained the lowest RMSE in most months, although CNN-BiGRU performed slightly better in a few months, and Transformer and Informer performed slightly better in January and February. The findings show that adding Attention to CNN-BiGRU generally reduces MAPE and MAE, but the improvements in RMSE and R^2 are not uniform across all months; in some cases, CNN-BiGRU performs similarly or slightly better. This suggests that Attention is more effective in reducing relative and average errors than large errors. However, the proposed model outperforms the other models compared overall and in almost all months.

Performance Analysis and Model Failure Analysis

In Figure 11, the performance of the proposed model is examined under four weather conditions (sunny, partly cloudy, cloudy, and rainy). In each of these four cases, a sequence of hourly images is presented first, resulting from cloud movement forecasting using the Optical Flow algorithm and showing the position and movement of clouds throughout the day. A comparison curve of the actual and forecasted solar irradiance is presented, followed by a comparison curve of the actual and forecasted PV power.

In sunny conditions, the sequence of optical flow images shows that the Optical Flow algorithm produces a stable and predictable pattern of cloud movement, allowing the CNN-BiGRU-Attention model to forecast irradiance and power with very high accuracy and almost perfect agreement. This case shows the high ability of the model to learn smooth and low-volatility patterns. According to Table 10, the sunny case has the lowest PV power forecasting error (RMSE = 0.0651 kW, MAE = 0.0503 kW, MAPE = 3.71%) and also shows the lowest error in solar irradiance forecasting (RMSE = 5.4155 W/m^2, MAE = 4.7203 W/m^2). Therefore, this scenario shows the best performance of the model.

In partly cloudy conditions, the sequence of optical flow images shows the rapid and irregular movement of clouds. In this situation, the model's weakness in responding to sudden and short-term changes becomes apparent. The largest difference between the actual and forecasted values is observed in the zoomed-in range. According to Table 10, the RMSE of PV power increases to 0.11204 kW and the RMSE of solar irradiance increases to 6.6408 W/m^2, (MAPE of solar irradiance: 30.131%). These results show the sensitivity of the model to fluctuations caused by the rapid passage of clouds.

In cloudy conditions, the sequence of optical flow images shows relatively uniform, stable cloud cover. Despite the reduction in the absolute level of irradiance, the temporal pattern of the signal is much smoother and less volatile than that observed in the partly cloudy case. This uniformity in cloud movement allows the model to achieve good relative accuracy. According to Table 10, although the RMSE of irradiance is slightly higher in this case (6.723 W/m^2), the MAPE of PV power is the lowest among the four scenarios with a value of 3.3687%. This result indicates that the uniformity of cloud movement, even with higher cloud cover, helps to improve the model accuracy.

In rainy conditions, which are considered the most critical scenario for the model, the sequence of optical flow images shows dense and stable cloud cover with short, sudden, and unpredictable gaps. These transient gaps are not well captured by the cloud motion vectors, and consequently, the model is unable to fully reproduce the sudden changes in irradiance and power. As is clearly evident in the zoomed-in view, the largest gap between the actual and forecasted values among the four scenarios occurs in this case.

This finding is consistent with Table 10, which shows the cloudy scenario shows the highest errors for both solar irradiance (RMSE = 6.686 W/m^2, MAPE = 36.14%) and PV power (RMSE = 0.139 kW, MAPE = 10.54%). The PV power MAPE is approximately 2.8 times higher than in the sunny case, indicating that the main weakness of the proposed model lies in tracking rapid, short-term fluctuations caused by transient cloud movements.

Table 9. Comparison of the forecast accuracy of the proposed model with benchmark models in 2018.

Model	Metrics	January	February	March	April	May	June	July	August	September	October	November	December
GRU	MAE (kW)	0.5315	0.6046	0.4859	0.4438	0.4557	0.6361	0.6284	0.6905	0.7922	0.8803	0.8490	0.832
	MAPE (%)	27.6948	34.1671	64.1708	24.0541	14.1005	34.1286	28.9061	32.2516	31.584	32.59	76.9896	53.26
	RMSE (kW)	0.6944	0.8968	0.5952	0.5311	0.5466	0.7403	0.7059	0.7895	0.9119	0.845	0.7505	0.85
	R^2	0.9260	0.8968	0.9455	0.9570	0.9339	0.7080	0.7111	0.4782	0.892	0.845	0.782	0.81
BiGRU	MAE (kW)	0.3708	0.4130	0.3142	0.2846	0.2888	0.3982	0.3912	0.4435	0.4846	0.5122	0.4816	0.5177
	MAPE (%)	15.3690	21.0765	37.4171	12.5804	8.8017	20.6089	17.4490	20.1743	48.4304	41.6854	36.3305	53.7136
	RMSE (kW)	0.5148	0.6354	0.4184	0.3564	0.3508	0.4604	0.4363	0.4985	0.5562	0.5955	0.5922	0.6364
	R^2	0.9593	0.9371	0.9731	0.9806	0.9728	0.8871	0.8896	0.7920	0.8331	0.8223	0.8589	0.8493
LSTM	MAE (kW)	0.7137	0.7488	0.6793	0.5978	0.6194	0.7711	0.6813	0.7499	0.7794	0.8430	0.85344	0.8160
	MAPE (%)	46.8838	47.2500	85.7083	37.8060	17.6178	50.7248	33.8481	52.8486	88.4977	82.4038	96.9273	80.25
	RMSE (kW)	0.9116	0.9688	0.8629	0.7496	0.7607	0.9772	0.8344	0.8606	0.8870	0.9980	0.9982	0.986
	R^2	0.8725	0.8539	0.8855	0.9144	0.8719	0.4911	0.5964	0.3800	0.5755	0.5009	0.5992	0.6379
BiLSTM	MAE (kW)	0.3902	0.3956	0.3056	0.2713	0.3125	0.4919	0.4618	0.5285	0.5391	0.6056	0.5526	0.5869
	MAPE (%)	18.2279	24.8096	35.7739	13.9360	9.4053	25.3105	19.8029	25.3239	42.3281	44.6695	39.9474	57.9765
	RMSE (kW)	0.5443	0.6332	0.4201	0.3570	0.3988	0.5927	0.5330	0.6123	0.6440	0.7205	0.7005	0.7352
	R^2	0.9545	0.9376	0.9729	0.9806	0.9648	0.8128	0.8353	0.6861	0.7763	0.7399	0.8025	0.7990
Informer	MAE (kW)	0.2331	0.2514	0.1532	0.1318	0.1098	0.1474	0.1227	0.1657	0.2462	0.2649	0.2580	0.3002
	MAPE (%)	11.4059	12.7230	23.1460	6.2533	3.6981	10.0143	8.7833	7.2005	39.1369	28.5615	27.0837	43.0387
	RMSE (kW)	0.3762	0.5380	0.2364	0.1826	0.1512	0.1961	0.1694	0.2228	0.3376	0.3731	0.3484	0.4174
	R^2	0.9783	0.9549	0.9914	0.9949	0.9949	0.9795	0.9834	0.9584	0.9385	0.9303	0.9512	0.9352
Transformer	MAE (kW)	0.1476	0.1945	0.0841	0.0809	0.0706	0.0598	0.0507	0.0434	0.0492	0.0524	0.0651	0.0681
	MAPE (%)	7.5230	5.9166	8.8770	5.8876	1.7914	2.8457	1.7243	4.1936	7.9920	6.4055	9.1058	14.9264
	RMSE (kW)	0.3363	0.5538	0.1941	0.0990	0.0891	0.0803	0.0734	0.0552	0.0624	0.0671	0.0857	0.089
	R^2	0.9827	0.9522	0.9942	0.9985	0.9982	0.9966	0.9969	0.9975	0.9979	0.9971	0.9970	0.9971
CNN-GRU	MAE (kW)	0.3723	0.3901	0.3073	0.2953	0.3343	0.5667	0.5503	0.6274	0.6451	0.6601	0.5901	0.5978
	MAPE (%)	13.9395	18.0708	32.7344	12.4169	10.4380	28.4411	23.4774	23.3245	51.127	45.9531	43.7058	55.6963
	RMSE (kW)	0.5179	0.6274	0.4010	0.3662	0.4194	0.6595	0.6189	0.7169	0.7437	0.7768	0.7172	0.7390
	R^2	0.9589	0.9387	0.9753	0.9796	0.9611	0.7683	0.7779	0.5698	0.017	0.6977	0.7969	0.7969
CNN-BiGRU	MAE (kW)	0.1178	0.1694	0.0552	0.0476	0.0303	0.0313	0.0259	0.0268	0.0274	0.0305	0.0409	0.0471
	MAPE (%)	4.0854	4.0796	2.7895	2.1224	0.8555	2.4410	1.4380	1.5882	2.6626	3.3600	5.9419	9.2296
	RMSE (kW)	0.3501	0.5719	0.1776	0.0650	0.0398	0.0404	0.0322	0.0328	0.0348	0.0430	0.0626	0.0930
	R^2	0.9812	0.9491	0.9951	0.9994	0.9997	0.9991	0.9994	0.9991	0.9993	0.9991	0.9984	0.9968

Proposed Model	MAE (kW)	0.1178	0.1518	0.0532	0.0413	0.0330	0.0450	0.0355	0.0204	0.0208	0.0311	0.0359	0.055
	MAPE (%)	4.3159	4.0111	2.5924	1.5644	0.9293	2.3767	1.9583	1.1134	2.0498	3.7913	4.4641	8.5233
	RMSE (kW)	0.3482	0.5536	0.1860	0.0546	0.0393	0.0572	0.0443	0.0317	0.0318	0.0578	0.0502	0.0884
	R^2	0.9814	0.9470	0.9947	0.9995	0.999	0.9983	0.9989	0.9980	0.9979	0.9970	0.9981	0.9971

Table 10. Performance comparison across weather conditions.

Condition	Solar Irradiance Forecasting			PV Power Forecasting		
	RMSE (W/m^2)	MAE (W/m^2)	MAPE (%)	RMSE (W/m^2)	MAE (W/m^2)	MAPE (%)
Sunny	5.4155	4.7203	28.596	0.0651	0.0503	3.7066
Partly Cloudy	6.6408	4.762	30.131	0.11204	0.1052	3.577
Cloudy	6.723	6.323	30.063	0.0673	0.0599	3.3687
Rainy	6.6856	5.859	36.1431	0.1392	0.0767	10.5424

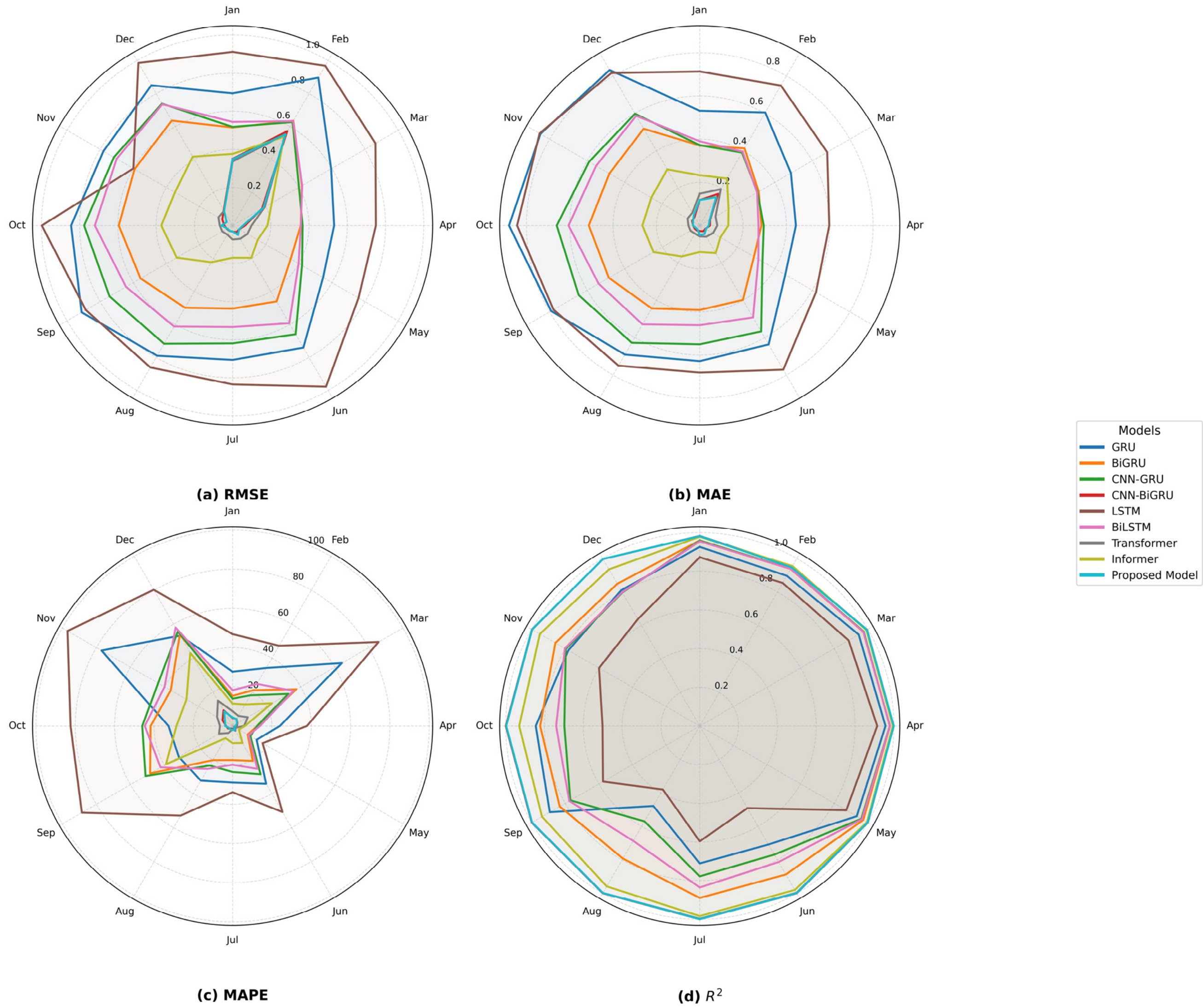
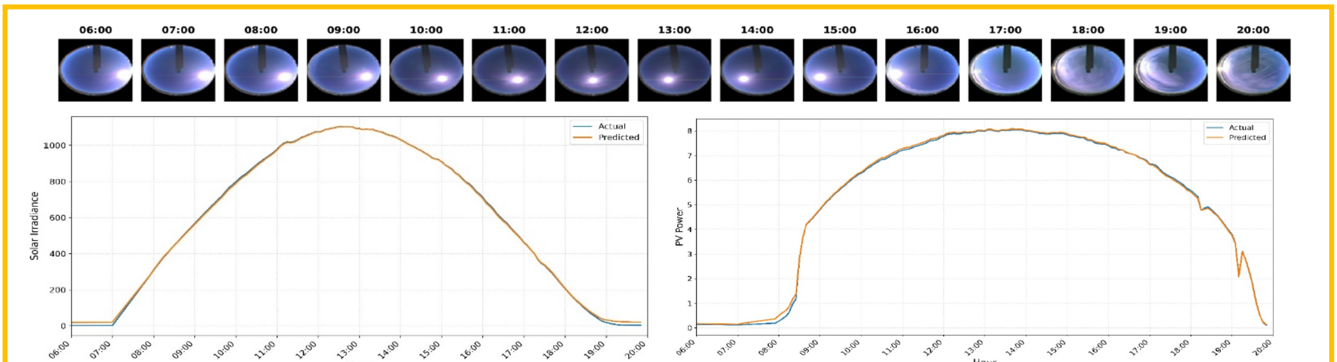


Figure 10. Radar graphs of RMSE, MAPE, MAE, and R^2 performance measures for different models across months of the year.



(a) Sunny

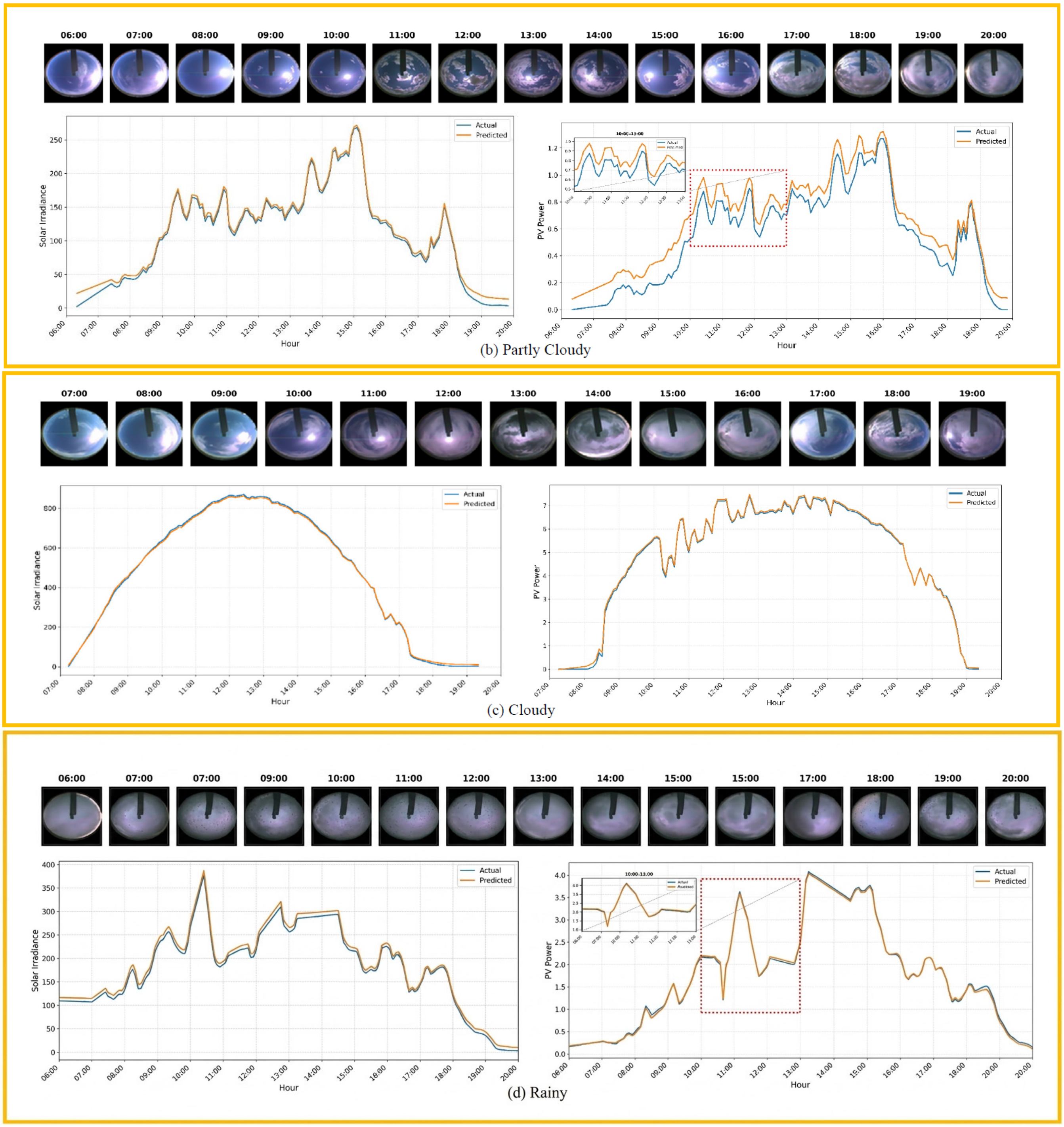


Figure 11. Model performance under sunny, partly cloudy, cloudy and rainy conditions, including Cloud motion, solar irradiance, and PV power forecasting.

Statistical Significance Analysis

To assess the statistical significance of the proposed model's performance improvements over the baseline models, the nonparametric paired Wilcoxon signed-rank test was applied at a significance level of $\alpha = 0.05$. The test was conducted on the paired forecasting errors from all samples in the test dataset for the proposed and baseline models. By calculating Cohen's d in addition to p-value, the effect size and practical significance of the differences were evaluated. The results in Table 11 show that the proposed model has a significant improvement ($p < 0.05$, indicated by *) in all three metrics, RMSE, MAE, and MAPE, compared to the seven baseline models. The largest effect size was observed in comparison with LSTM ($d = 4.71$ for

RMSE and d = 7.35 for MAE), indicating a very large performance difference. Similarly, the GRU model exhibits large effect sizes across all three metrics (d = 3.30 for RMSE, d = 3.47 for MAE, and d = 2.05 for MAPE), indicating a substantial and consistent improvement of the proposed model over the GRU baseline across the evaluation period.

In contrast, comparing the proposed model with CNN-BiGRU showed that the differences in none of the three metrics are statistically significant (RMSE: p = 0.8501, MAE: p = 0.8984, and MAPE: p = 0.2036), and the effect size is also negligible to small (d = 0.01–0.43). This finding is also consistent with Figure 10, which shows that CNN-BiGRU has the most overlap in error distribution with the proposed model. Architecturally, this result is also logical, since CNN-BiGRU forms the core of the proposed model, and the attention mechanism primarily improves performance by adaptively weighting more important temporal dependencies.

Table 11. Statistical significance analysis of the proposed model compared with benchmark models.

Benchmark Model	RMSE	MAE	MAPE
GRU	p = 0.0005 * | d = 3.30	p = 0.0005 * | d = 3.47	p = 0.0005 * | d = 2.05
BiGRU	p = 0.0005 * | d = 2.38	p = 0.0005 * | d = 3.68	p = 0.0005 * | d = 1.79
LSTM	p = 0.0005 * | d = 4.71	p = 0.0005 * | d = 7.35	p = 0.0005 * | d = 2.31
BiLSTM	p = 0.0005 * | d = 2.19	p = 0.0005 * | d = 3.00	p = 0.0005 * | d = 2.05
Informer	p = 0.0010 * | d = 1.39	p = 0.0005 * | d = 2.20	p = 0.0005 * | d = 1.26
Transformer	p = 0.0049 * | d = 1.06	p = 0.0005 * | d = 2.74	p = 0.0010 * | d = 1.44
CNN-GRU	p = 0.0005 * | d = 2.04	p = 0.0005 * | d = 2.69	p = 0.0005 * | d = 1.82
CNN-BiGRU	p = 0.8501 | d = 0.01	p = 0.8984 | d = 0.09	p = 0.2036 | d = 0.43

5.4.2. Comparative Analysis of Measured and Forecasted Irradiance for PV Power Forecasting

In order to evaluate the effectiveness of the proposed framework, the performance of the proposed CNN-BiGRU-Attention model was examined in two scenarios. In the first scenario, according to the framework presented in this study, the forecasted solar irradiance values were used to forecast the PV power. In the second scenario, the measured solar irradiance was considered as the input to the model, which can be considered as the ideal state and reference for the system performance.

According to the results in Table 12, using the forecasted irradiance instead of the measured irradiance increased the MAE by about 12.21%, the MAPE by 11.12%, and the RMSE by 11.59%, while the R^2 value decreased slightly from 0.991 to 0.987. These results indicate that the effect of the error caused by the irradiance forecasting stage is limited, and the proposed framework was able to retain a major part of the information affecting the PV power forecasting.

Table 12. Performance comparison of the proposed CNN-BiGRU model using measured and forecasted solar irradiance inputs.

Irradiance Source	MAE (kW)	MAPE (%)	RMSE (kW)	R^2
Measured Solar Irradiance	0.131	7.82	0.207	0.991
Forecasted Solar Irradiance	0.147	8.69	0.231	0.987
Relative Change (%)	+12.21	+11.12	+11.59	−0.4

Comparative Analysis Using Reduced Feature Sets

After the redundancy analysis presented in Section 4.1.3, a cumulative feature addition experiment was conducted to investigate the effect of gradually adding meteorological variables on model performance. The features were added to the model input in descending order of their correlation with PV power: solar irradiance, wind speed, temperature, and humidity. Historical PV power was kept constant across all scenarios. The RMSE, MAE, MAPE, and R^2 values for each configuration are presented in Table 13.

The results show that the model performance improved with the gradual addition of meteorological features, with RMSE decreasing from 0.241 to 0.231 kW, R^2 increasing from 0.957 to 0.980, and MAPE decreasing from 9.00% to 8.69%.

However, the addition of each meteorological variable resulted in a further and continuous reduction in the forecast error, and none of the scenarios showed a performance degradation after adding the new feature. In particular, humidity, despite having the weakest correlation and only a negative correlation with PV power, still improved the model performance the most when added to the last configuration. The model is capable of extracting nonlinear relationships, conditional interactions, and complex dependencies between variables.

Table 13. Cumulative feature-addition analysis: forecasting performance as meteorological features are added in descending order of correlation with PV power.

Scenario	Input Feature	RMSE (kW)	MAE (kW)	MAPE (%)	R^2
Scenario 1	Power + solar Irradiance	0.241	0.1512	9.002	0.957
Scenario 2	Power + solar Irradiance + Wind speed	0.239	0.1497	8.806	0.9750
Scenario 3	Power + solar Irradiance + Wind Speed + Temperature	0.236	0.1490	8.728	0.9756
Scenario 4	Power + solar Irradiance + Wind Speed + Temperature + Humidity	0.231	0.147	8.69	0.98

5.4.3. Discussion on Practical Applicability and Future Directions

By establishing a forecasting chain linking cloud dynamics, solar irradiance, and PV power, the framework provides information beyond historical numerical data and instantaneous irradiance measurements. Unlike approaches that directly use measured irradiance as an input for PV power forecasting, the proposed method first forecasts cloud movement and future sky conditions and then forecasts PV power for the next hour based on the forecasted solar irradiance. Therefore, the proposed framework can provide information on short-term changes in PV power and enhance the ability to make operational decisions before significant changes in power output occur.

By reducing the uncertainty of PV power forecasting, this capability can play an effective role in the optimal management and operation of solar power systems and grids with high renewable energy penetration. The forecasted power can be used in energy management systems (EMS) for short-term generation planning and in battery energy storage systems (BESS) to optimize the charging and discharging process. Also, increasing forecast accuracy helps plan storage capacity, manage generation and consumption imbalances, determine more appropriate rotating storage, and ultimately reduce grid balancing costs and increase the utilization of renewable resources.

The proposed framework shows good performance stability under different seasonal conditions, particularly from summer to early autumn. In addition, the use of dense optical flow enables the forecasting of cloud movement and the early detection of changes in solar irradiance and PV power. This capability can be effective in detecting and managing ramp events and taking corrective actions before severe changes in power generation occur, particularly in grid-connected power plants.

From a computational perspective, the proposed three-stage architecture has good scalability. As shown in Table 6, the CNN-BiGRU-Attention model has a lightweight structure with 204,511 trainable parameters, a computational cost of 0.81 MFLOPs, and an inference time of 0.005 s per sample, allowing it to run on readily available hardware. This feature enables the model to be used in online applications and even in PV power plants with limited computational resources. However, in practical implementation, the computational costs of the preceding stages must also be considered. Given the inference times of 0.02 s for Optical Flow and 0.009 s for ResNet50, executing the different stages in a parallel or asynchronous pipeline can reduce the cumulative latency, enabling the timely provision of forecasts to energy management systems.

Despite the advantages of the proposed framework, its deployment in real environments is challenging because, unlike methods based solely on numerical data, it requires sky camera images and, consequently, proper installation, calibration, and maintenance at the power plant site. Factors such as dust, precipitation, direct sunlight reflections, lighting changes, and severe atmospheric fluctuations can reduce image quality and the accuracy of cloud movement forecasting. In this study, to reduce these effects, removing irrelevant areas, reducing noise, increasing the contrast of cloud areas, and color normalization were used as preprocessing steps. However, in practical applications, continuous monitoring of image quality and detection of invalid data can further increase the reliability of the framework.

On the other hand, the chain nature of the proposed framework can cause error transmission between stages, such that cloud motion forecasting errors affect solar irradiance forecasting and ultimately PV power forecasting. Therefore, quantifying uncertainty and monitoring the outputs of different stages can increase the reliability of the framework. Also, the temporal synchronization of images and meteorological data.

Based on the findings and limitations of this study, three main directions for future research are suggested. First, the proposed framework should be evaluated in several PV systems and different climatic regions, including desert, coastal, and mountainous regions, in order to investigate the generalizability and robustness of the model. In this regard, the use of transfer learning and domain adaptation methods can enable the rapid adaptation of the model to new sites with minimal local data. Second, hybrid methods should be developed to more comprehensively model cloud motion by accounting for changes in cloud displacement, shape, density, and structure. This can improve forecast accuracy under variable weather conditions, especially during periods of severe fluctuations in cloud cover, enhance the quality of input data, and reduce the performance gap relative to advanced architectures. Third, the framework should be extended to multi-step forecasting at 30 min to 6 h horizons to increase its applicability to operational planning and energy management. In this context, it is particularly important to investigate and quantify the cumulative effect of early-stage errors, especially cloud motion forecasting errors, on the accuracy of the final power forecasts over longer time horizons.

6. Conclusions

This Paper presents a novel three-stage DL framework for PV power forecasting for one-hour ahead, in which cloud motion forecasting, solar irradiance forecasting, and PV power forecasting are combined into an integrated structure. In contrast to conventional methods that usually rely on measured irradiance data or meteorological variables directly, this approach first forecast the future sky state using dense optical flow-based cloud motion tracking, then estimates future solar irradiance with a ResNet50, and finally uses the irradiance output along with meteorological data and historical PV power as input to the CNN-BiGRU-Attention model for power forecasting.

Combining sky images and forecasted solar irradiance allows modeling the causal relationship between cloud movement, solar irradiance changes, and PV power. Bayesian optimization is also used to automatically adjust hyperparameters, and a self-attention mechanism is used to extract important temporal features while reducing the impact of irrelevant information.

The results of the first stage showed that the Dense Optical Flow algorithm forecasted cloud movement patterns well in clear, cloudy, partly cloudy, and almost cloudy conditions.

In the second step, the ResNet50 model achieved accurate solar irradiance forecasting under different atmospheric conditions. The best performance was achieved under clear sky conditions with MAE = 4.20 W/m^2, RMSE = 5.50 W/m^2, and R^2 = 0.92, while acceptable performance was also maintained under cloudy, partly cloudy, and rainy conditions despite increasing atmospheric variations.

The third stage of the framework demonstrated the effectiveness of the proposed CNN-BiGRU-Attention architecture for PV power forecasting. The hybrid model successfully captured spatial correlations between meteorological variables and long-term temporal dependencies in PV power.

Comparative studies with the benchmark GRU, BiGRU, LSTM, BiLSTM, Informer, Transformer, CNN-GRU, and CNN-BiGRU models showed that the proposed model consistently achieved the best forecasting performance in all months of the test data.

The model achieved monthly R^2 values ranging from 0.947 to 0.999, MAE values ranging from 0.0204 to 0.1518 kW, RMSE values ranging from 0.0393 to 0.5536 kW, and MAPE values ranging from 0.9293% to 8.5233%. For one-hour ahead forecasting using the forecasted irradiance inputs, the proposed framework achieved an overall R^2 of 0.987 with RMSE, MAE, and MAPE values of 0.231 kW, 0.147 kW, and 8.69%, respectively. Furthermore, the comparison with the ideal scenario based on measured irradiance showed only a slight performance degradation, confirming that the irradiance forecasting step effectively retained most of the information needed to accurately forecast PV power.

Overall, the proposed framework shows that the integration of cloud motion forecasting, image-based irradiance forecasting, and attention-based DL significantly improves the accuracy and stability of short-term PV power forecasting under diverse weather conditions.

Despite the favorable performance of the proposed framework, this study has some limitations, the most important of which is the validation of the model based on data from a single PV power plant in a single geographical location; although the 12-month data cover a diverse range of seasonal and atmospheric conditions, this limitation may reduce the generalizability of the results to regions with different climatic conditions and irradiance patterns. Therefore, in future research, it is necessary to validate the framework using data from multiple PV power plants across different geographical and climatic regions, particularly under variable atmospheric conditions and rapid cloud changes. Also, the development of more advanced sky image processing methods to more accurately extract cloud features, improve cloud motion and shape modeling, and more accurately consider the effect of cloud shadow on solar panels, along with the integration of this information with meteorological data, can help increase the accuracy of solar irradiance forecasting and ultimately improve the performance of the PV power forecasting model.

Appendix A

The configuration and optimal hyperparameters selected for single and hybrid deep learning models used for comparison in this study are shown in Tables A1–A8.

Table A1. Specifications of the GRU model.

Component	Parameters	Value
GRU layer	Hidden unit Activation	64 Relu
GRU layer	Hidden unit Activation	19
Dense layer	Neuron	1
Dropout		0.19
Batch size		16
Epoch		100

Table A2. Specifications of the BiGRU model.

Component	Parameters	Value
BiGRU layer	Hidden unit Activation	68 Relu
BiGRU layer	Hidden unit Activation	33 Relu
Dense layer	Neuron	1
Dropout		0.13

Batch size	32
Epoch	100

Table A3. Specifications of the LSTM model.

Component	Parameters	Value
GRU layer	Hidden unit	50
	Activation	Relu
GRU layer	Hidden unit	50
	Activation	Relu
Dense layer	Neuron	1
Dropout		0.24
Batch size		32
Epoch		100

Table A4. Specifications of the BiLSTM model.

Component	Parameters	Value
GRU layer	Hidden unit	50
	Activation	Relu
GRU layer	Hidden unit	50
	Activation	Relu
Dense layer	Neuron	1
Dropout		0.24
Batch size		32
Epoch		100

Table A5. Specifications of the Informer model.

Component	Parameters	Value
Embedding Layer	Dimension	64
Encoder	Number of layers	3
	Distilling operations	2
ProbSparse Attention	Number of heads	4
	Sampling factor	5
Feed-forward layer	Hidden dimension	128
	Activation	Relu
Decoder	Number of layers	1
Dropout		0.1
Dense layer	Neuron	1
	Activation	linear

Table A6. Specifications of the Transformer model.

Component	Parameters	Value
Input projection (Dense)	Dimension	64
Positional Embedding	Type	Learnable (Embedding layer)
Encoder	Number of layers	2
Multi-Head Attention	Number of heads	4
Feed-forward layer	Hidden dimension	64
	Activation	Relu
Dropout	-	0.15
Dense layer	Neuron	20
	Activation	Relu

Dense layer	Neuron	1
	Activation	linear

Table A7. Specifications of the CNN-GRU model.

Component	Parameters	Value
Conv 1D	Filters	64
	Kernel size	3 × 3
	Activation	Relu
MaxPooling	Pool size	3 × 3
	Padding	Same
Conv 1D	Filters	64
	Kernel size	3 × 3
	Activation	Relu
MaxPooling	Pool size	3 × 3
	Padding	same
GRU layer	Hidden unit	18
	Activation	Relu
GRU layer	Hidden unit	18
	Activation	Relu
Dense layer	Neuron	1

Table A8. Specifications of the CNN-BiGRU model.

Component	Parameters	Value
Conv 1D	Filters	64
	Kernel size	3 × 3
	Activation	Relu
MaxPooling	Pool size	3 × 3
	Padding	same
Conv 1D	Filters	64
	Kernel size	3 × 3
	Activation	Relu
MaxPooling	Pool size	3 × 3
	Padding	same
BiGRU layer	Hidden unit	20
	Activation	Relu
BiGRU layer	Hidden unit	20
	Activation	Relu
Dense layer	Neuron	1

Statement of the Use of Generative AI and AI-Assisted Technologies in the Writing Process

During the preparation of this manuscript, the authors used Grammarly and ChatGPT in order to improve the English language and readability of the text. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Author Contributions

Conceptualization, Z.J.L. and S.-M.M.-T.; Methodology, Z.J.L. and S.-M.M.-T.; Software, Z.J.L.; Validation, Z.J.L. and S.-M.M.-T.; Formal Analysis, Z.J.L.; Investigation, Z.J.L.; Data Curation, Z.J.L.; Writing—Original Draft Preparation, Z.J.L.; Writing—Review & Editing, Z.J.L. and S.-M.M.-T.; Visualization, Z.J.L.; Supervision, S.-M.M.-T.; Project Administration, S.-M.M.-T.

Ethics Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Funding

This research received no external funding.

Declaration of Competing Interest

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

Nomenclature and Abbreviations

DL	Deep Learning	WRF	Weather Research & Forecasting
ML	Machine Learning	SAT	Single-Axis Tracker
CNN	Convolutional Neural Network	SAA	Singular Spectrum Analysis
GRU	Gated Recurrent Unit	VMD	Variational Mode Decomposition
BiGRU	Bidirectional GRU	WT	Wavelet Transform
LSTM	Long Short-Term Memory	WPD	Wavelet Packet Decomposition
BiLSTM	Bidirectional LSTM	CMV	Cloud Motion Vector
MLP	Multilayer Perceptron	COT	Cloud Optical Thickness
FFNN	Feedforward Neural Network	NWP	Numerical Weather Prediction
RNN	Recurrent Neural Network	KDE	Kernel Density Estimation
FC	Fully Connected (layer)	QR	Quantile Regression
SVM	Support Vector Machine	RTI	Radiant Transmission Index
SVR	Support Vector Regression	RHM	Regression Hypersurface Model
ELM	Extreme Learning Machine	FCM	Fuzzy C-Means
GNN	Graph Neural Network	SOM	Self-Organizing Map
SAT	Single-Axis Tracker	ARIMA	Autoregressive Integrated Moving Average
BO	Bayesian Optimization	SARIMA	Seasonal Autoregressive Integrated Moving Average
GP	Gaussian Process	G	Solar Irradiance
EI	Expected Improvement	T	Air temperature
KNN	K-Nearest Neighbours	H	Relative humidity
PCA	Principal Component Analysis	W	Wind Speed
GRA	Grey Relational Analysis	P	Historical PV power
VIF	Variance Inflation Factor	**Variables**	
PSO	Particle Swarm Optimization	$I(x, y, t)$	Image intensity
GWO	Grey Wolf Optimizer	vx, vy	Optical flow velocity components
NREL	National Renewable Energy Laboratory	dx, dy	Displacement vector
FLOP	Floating-Point Operations	e_i	Raw attention score
α_i	Attention weight	Z_i	Z-score
h_i'	Attention-weighted hidden state	y_i	Actual value
wh	Attention weight matrix	$\hat{y}_i$	Forecasted value
bh	Attention bias vector	$\bar{y}_i$	Mean of actual values
$\hat{P}_t$	Forecasted PV power	N	Number of samples

x_i	Individual data point	r	Pearson correlation coefficient
μ	Mean	σ	Standard deviation

References

1. Krishnan N, Kumar KR, Inda CS. How solar radiation forecasting impacts the utilization of solar energy: A critical review. *J. Clean. Prod.* **2023**, *388*, 135860. DOI:10.1016/j.jclepro.2023.135860
2. Jalali Z, Moghaddas-Tafreshi SM. Optimal Design and Simulation of a Rooftop Photovolataic System in Faculty of Engineering Guilan University. In Proceedings of the 2023 10th Iranian Conference on Renewable Energy & Distributed Generation (ICREDG), Shahrood, Iran, 15–16 March 2023; pp. 1–5. DOI:10.1109/ICREDG58341.2023.10092071
3. Tzoumanikas P, Nikitidou E, Bais AF, Kazantzidis A. The effect of clouds on surface solar irradiance, based on data from an all-sky imaging system. *Renew. Energy* **2016**, *95*, 314–322. DOI:10.1016/j.renene.2016.04.026
4. Aicardi D, Musé P, Alonso-Suárez R. A comparison of satellite cloud motion vectors techniques to forecast intra-day hourly solar global horizontal irradiation. *Sol. Energy* **2022**, *233*, 46–60. DOI:10.1016/j.solener.2021.12.066
5. Aguiar LM, Pereira B, Lauret P, Díaz F, David M. Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting. *Renew. Energy* **2016**, *97*, 599–610. DOI:10.1016/j.renene.2016.06.018
6. Caldas M, Alonso-Suárez R. Very short-term solar irradiance forecast using all-sky imaging and real-time irradiance measurements. *Renew. Energy* **2019**, *143*, 1643–1658. DOI:10.1016/j.renene.2019.05.069
7. Trigo-González M, Cortés-Carmona M, Marzo A, Alonso-Montesinos J, Martínez-Durbán M, López G, et al. Photovoltaic power electricity generation nowcasting combining sky camera images and learning supervised algorithms in the Southern Spain. *Renew. Energy* **2023**, *206*, 251–262. DOI:10.1016/j.renene.2023.01.111
8. Logothetis SA, Salamalikis V, Wilbert S, Remund J, Zarzalejo LF, Xie Y, et al. Benchmarking of solar irradiance nowcast performance derived from all-sky imagers. *Renew. Energy* **2022**, *199*, 246–261. DOI:10.1016/j.renene.2022.08.127
9. Yu Z, Tan Z, Ma S, Yan W. Nowcast for cloud top height from Himawari-8 data based on deep learning algorithms. *Meteorol. Appl.* **2023**, *30*, e2130. DOI:10.1002/met.2130
10. Terrén-Serrano G, Martínez-Ramón M. Detection of clouds in multiple wind velocity fields using ground-based infrared sky images. *Knowl.-Based Syst.* **2023**, *274*, 110628. DOI:10.1016/j.knosys.2023.110628
11. Kosmopoulos P, Kouroutsidis D, Papachristopoulou K, Raptis PI, Masoom A, Saint-Drenan YM, et al. Short-Term Forecasting of Large-Scale Clouds Impact on Downwelling Surface Solar Irradiation. *Energies* **2020**, *13*, 6555. DOI:10.3390/en13246555
12. Mondragón R, Alonso-Montesinos J, Riveros-Rosas D, Bonifaz R. Determination of Cloud Motion Applying the Lucas-Kanade Method to Sky Cam Imagery. *Remote Sens.* **2020**, *12*, 2643. DOI:10.3390/rs12162643
13. Niu Y, Song J, Zou L, Yan Z, Lin X. Cloud detection method using ground-based sky images based on clear sky library and superpixel local threshold. *Renew. Energy* **2024**, *226*, 120452. DOI:10.1016/j.renene.2024.120452
14. Al-lahham A, Theeb O, Elalem K, Alshawi TA, Alshebeili SA. Sky Imager-Based Forecast of Solar Irradiance Using Machine Learning. *Electronics* **2020**, *9*, 1700. DOI:10.3390/electronics9101700
15. Kumari P, Toshniwal D. Deep learning models for solar irradiance forecasting: A comprehensive review. *J. Clean. Prod.* **2021**, *318*, 128566. DOI:10.1016/j.jclepro.2021.128566
16. Feng C, Zhang J. SolarNet: A sky image-based deep convolutional neural network for intra-hour solar forecasting. *Sol. Energy* **2020**, *204*, 71–78. DOI:10.1016/j.solener.2020.03.083
17. Nie Y, Paletta Q, Scott A, Pomares LM, Arbod G, Sgouridis S, et al. Sky image-based solar forecasting using deep learning with heterogeneous multi-location data: Dataset fusion versus transfer learning. *Appl. Energy* **2024**, *369*, 123467. DOI:10.1016/j.apenergy.2024.123467
18. Ansong M, Huang G, Nyang'onda TN, Musembi RJ, Richards BS. Very short-term solar irradiance forecasting based on open-source low-cost sky imager and hybrid deep-learning techniques. *Sol. Energy* **2025**, *294*, 113516. DOI:10.1016/j.solener.2025.113516
19. Khan AW, Duan J, Nawaz F, Lu W. Novel hybrid BiLSTM-BiGRU and CNN-LSTM architectures for enhanced solar irradiance forecasting in semi-arid climates. *Int. J. Electr. Power Energy Syst.* **2025**, *172*, 111380. DOI:10.1016/j.ijepes.2025.111380
20. Ukwuoma CC, Cai D, Bamisile O, Yin H, Nneji GU, Monday HN, et al. An attention fused sequence-to-sequence convolutional neural network for accurate solar irradiance forecasting and prediction using sky images. *Renew. Energy* **2024**, *237*, 121692. DOI:10.1016/j.renene.2024.121692
21. Rajagukguk RA, Lee H. Sky-image-based sun-blocking index and PredRNN++ for accurate short-term solar irradiance forecasting. *Build. Environ.* **2025**, *274*, 112429. DOI:10.1016/j.buildenv.2024.112429

22. Jonathan AL, Cai D, Ukwuoma CC, Nkou NJJ, Huang Q, Bamisile O. A radiant shift: Attention-embedded CNNs for accurate solar irradiance forecasting and prediction from sky images. *Renew. Energy* **2024**, *234*, 121133. DOI:10.1016/j.renene.2024.121133
23. Tiwari S, Sabzehgar R, Rasouli M. Short Term Solar Irradiance Forecast based on Image Processing and Cloud Motion Detection. In Proceedings of the 2019 IEEE Texas Power and Energy Conference (TPEC), College Station, TX, USA, 7–8 February 2019; pp. 1–6. DOI:10.1109/TPEC.2019.8662134
24. Bouzerdoum M, Mellit A, Massi Pavan A. A hybrid model (SARIMA–SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant. *Sol. Energy* **2013**, *98*, 226–235. DOI:10.1016/j.solener.2013.10.002
25. Rajabalizadeh S, Tafreshi SMM. A Practicable Copula-Based Approach for Power Forecasting of Small-Scale Photovoltaic Systems. *IEEE Syst. J.* **2020**, *14*, 4911–4918. DOI:10.1109/JSYST.2019.2961661
26. Isapour E, Moghaddas Tafreshi SM. A Hybrid Model Based on Neural networks to Predict the Power Output of Photovoltaic systems. In Proceedings of the 2019 Iranian Conference on Renewable Energy & Distributed Generation (ICREDG), Tehran, Iran, 11–12 June 2019; pp. 1–6. DOI:10.1109/ICREDG47187.2019.194227
27. Abed JN, Abdoos AA, Ebadi A. A new hybrid intelligent method for short-term photovoltaic power forecasting based on combination of ELM, VMD and PCA. *Electr. Power Syst. Res.* **2026**, *250*, 112104. DOI:10.1016/j.epsr.2025.112104
28. Cui C, Wu H, Jiang X, Jing L. Short- and medium-term forecasting of distributed PV output in plateau regions based on a hybrid MLP-FGWO-PSO approach. *Energy Rep.* **2024**, *11*, 2685–2691. DOI:10.1016/j.egyr.2024.01.080
29. Cecaj A, Lippi M, Mamei M, Zambonelli F. Comparing Deep Learning and Statistical Methods in Forecasting Crowd Distribution from Aggregated Mobile Phone Data. *Appl. Sci.* **2020**, *10*, 6580. DOI:10.3390/app10186580
30. Wang H, Yi H, Peng J, Wang G, Liu Y, Jiang H, et al. Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network. *Energy Convers. Manag.* **2017**, *153*, 409–422. DOI:10.1016/j.enconman.2017.10.008
31. Huang CJ, Kuo PH. Multiple-Input Deep Convolutional Neural Network Model for Short-Term Photovoltaic Power Forecasting. *IEEE Access* **2019**, *7*, 74822–74834. DOI:10.1109/ACCESS.2019.2921238
32. Zang H, Cheng L, Ding T, Cheung KW, Liang Z, Wei Z, et al. Hybrid method for short-term photovoltaic power forecasting based on deep convolutional neural network. *IET Gener. Transm. Distrib.* **2018**, *12*, 4557–4567. DOI:10.1049/iet-gtd.2018.5847
33. Qing X, Niu Y. Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM. *Energy* **2018**, *148*, 461–468. DOI:10.1016/j.energy.2018.01.177
34. Luo X, Zhang D, Zhu X. Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge. *Energy* **2021**, *225*, 120240. DOI:10.1016/j.energy.2021.120240
35. Wang K, Qi X, Liu H. A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network. *Appl. Energy* **2019**, *251*, 113315. DOI:10.1016/j.apenergy.2019.113315
36. Ma Y, Huang Y, Yuan Y. The short-term forecasting of distributed photovoltaic power considering the sensitivity of meteorological data. *J. Clean. Prod.* **2025**, *486*, 144599. DOI:10.1016/j.jclepro.2024.144599
37. Wang H, Zhang Y, Liang J, Liu L. DAFA-BiLSTM: Deep Autoregression Feature Augmented Bidirectional LSTM network for time series prediction. *Neural Netw.* **2023**, *157*, 240–256. DOI:10.1016/j.neunet.2022.10.009
38. Guo X, Mo Y, Yan K. Short-Term Photovoltaic Power Forecasting Based on Historical Information and Deep Learning Methods. *Sensors* **2022**, *22*, 9630. DOI:10.3390/s22249630
39. Lee W, Kim K, Park J, Kim J, Kim Y. Forecasting Solar Power Using Long-Short Term Memory and Convolutional Neural Networks. *IEEE Access* **2018**, *6*, 73068–73080. DOI:10.1109/ACCESS.2018.2883330
40. Elizabeth Michael N, Mishra M, Hasan S, Al-Durra A. Short-Term Solar Power Predicting Model Based on Multi-Step CNN Stacked LSTM Technique. *Energies* **2022**, *15*, 2150. DOI:10.3390/en15062150
41. Guo F, Yang C, Xia D, Xu J. Short-Term Prediction of Photovoltaic Power Based on Improved CNN-LSTM and Cascading Learning. *Energy* **2025**, *122*, 1975–1999. DOI:10.32604/ee.2025.062035
42. Xu X, Guan L, Wang Z, Yao R, Guan X. A double-layer forecasting model for PV power forecasting based on GRU-Informer-SVR and Blending ensemble learning framework. *Appl. Soft Comput.* **2025**, *172*, 112768. DOI:10.1016/j.asoc.2025.112768
43. Guo S, Fan H, Huang J. Ultra-short-term PV power prediction based on an improved hybrid model with sky image features and data two-dimensional purification. *Energy* **2025**, *331*, 137018. DOI:10.1016/j.energy.2025.137018
44. Xia Y, Su Y, Yu X, Wang Z, Su J, Zhu S, et al. Ultra-Short-Term Power Forecasting of Single-Axis Tracking PV System Based on Irradiance Transformation and CNN-BiGRU Model. *IET Renew. Power Gener.* **2026**, *20*, e70214. DOI:10.1049/rpg2.70214

45. Zhang X, Gao R, Zhu C, Liu C, Mei S. Ultra-short-term prediction of regional photovoltaic power based on dynamic graph convolutional neural network. *Electr. Power Syst. Res.* **2024**, *226*, 109965. DOI:10.1016/j.epsr.2023.109965
46. Hendrikx NY, Barhmi K, Visser LR, de Bruin TA, Pó M, Salah AA, et al. All sky imaging-based short-term solar irradiance forecasting with Long Short-Term Memory networks. *Sol. Energy* **2024**, *272*, 112463. DOI:10.1016/j.solener.2024.112463
47. Li H, Wu W, Chen W, Zhang M. RTI-Net: Physics-informed deep learning for photovoltaic power forecasting. *Renew. Energy* **2026**, *256*, 124152. DOI:10.1016/j.renene.2025.124152
48. Zhang S, Dong Z, Qi D, Wang M, Xu Z, Han Y, et al. Intra-Hour PV Power Forecasting Technique Based on Total-Sky Images. *CSEE J. Power Energy Syst.* **2026**, *12*, 210–219. DOI:10.17775/CSEEJPES.2022.07770
49. Ruan G, Chen X, Li Y, Lim EG, Fang L, Jiang L, et al. SkyNet: A Deep Learning Architecture for Intra-hour Multimodal Solar Forecasting with Ground-based Sky Images. *Renew. Energy* **2026**, *256*, 124354. DOI:10.1016/j.renene.2025.124354
50. Jittratorn N, Chi W, Huang CM, Yang HT. Ultrashort-term PV power forecasting based on weather classification and dynamic cloud tracking: Deterministic and probabilistic modeling. *Energy Convers. Manag. X* **2026**, *30*, 101713. DOI:10.1016/j.ecmx.2026.101713
51. Zhang Q, Zeng Z, Zhou Q, Ding X, Dong Z, Tian M, et al. Multimodal bidirectional MAMBA framework for ultra-short-term photovoltaic power forecasting using sky images. *Renew. Energy* **2026**, *256*, 124361. DOI:10.1016/j.renene.2025.124361
52. Walker J, Gupta A, Hebert M. Dense Optical Flow Prediction from a Static Image. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015; pp. 2443–2451. DOI:10.1109/ICCV.2015.281
53. LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* **2015**, *521*, 436–444. DOI:10.1038/nature14539
54. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. DOI:10.1109/CVPR.2016.90
55. Hochreiter S, Schmidhuber J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. DOI:10.1162/neco.1997.9.8.1735
56. Cho K, van Merrienboer B, Gulcehre C, Bahdanau D, Bougares F, Schwenk H, et al. Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), Doha, Qatar, 25–29 October 2014; pp. 1724–1734. DOI:10.3115/v1/D14-1179
57. Qin Y, Song D, Chen H, Cheng W, Jiang G, Cottrell GW. A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction. In Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, Melbourne, Australia, 19–25 August 2017; pp. 2627–2633. DOI:10.24963/ijcai.2017/366
58. Victoria AH, Maragatham G. Automatic tuning of hyperparameters using Bayesian optimization. *Evol. Syst.* **2021**, *12*, 217–223. DOI:10.1007/s12530-020-09345-2
59. Yuzer EO, Bozkurt A. Deep learning model for regional solar radiation estimation using satellite images. *Ain Shams Eng. J.* **2023**, *14*, 102057. DOI:10.1016/j.asej.2022.102057