

## Article

# Solar Power Forecasting Using CNN-LSTM Hybrid Model

Su-Chang Lim <sup>1</sup>, Jun-Ho Huh <sup>2</sup> , Seok-Hoon Hong <sup>1</sup> , Chul-Young Park <sup>1,\*</sup>  and Jong-Chan Kim <sup>3,\*</sup> 

<sup>1</sup> R&D Center, TEF Co., Ltd., 60-12 Suncheon-ro, Seo-Myeon, Suncheon 57906, Korea

<sup>2</sup> Department of Data Science, Korea Maritime and Ocean University, Busan 49112, Korea

<sup>3</sup> Department of Computer Engineering, Sunchon National University, Suncheon 57992, Korea

\* Correspondence: naksu21@sunchon.ac.kr (C.-Y.P.); seaghost@sunchon.ac.kr (J.-C.K.)

**Abstract:** Photovoltaic (PV) technology converts solar energy into electrical energy, and the PV industry is an essential renewable energy industry. However, the amount of power generated through PV systems is closely related to unpredictable and uncontrollable environmental factors such as solar radiation, temperature, humidity, cloud cover, and wind speed. Particularly, changes in temperature and solar radiation can substantially affect power generation, causing a sudden surplus or reduction in the power output. Nevertheless, accurately predicting the energy produced by PV power generation systems is crucial. This paper proposes a hybrid model comprising a convolutional neural network (CNN) and long short-term memory (LSTM) for stable power generation forecasting. The CNN classifies weather conditions, while the LSTM learns power generation patterns based on the weather conditions. The proposed model was trained and tested using the PV power output data from a power plant in Busan, Korea. Quantitative and qualitative evaluations were performed to verify the performance of the model. The proposed model achieved a mean absolute percentage error of 4.58 on a sunny day and 7.06 on a cloudy day in the quantitative evaluation. The experimental results suggest that precise power generation forecasting is possible using the proposed model according to instantaneous changes in power generation patterns. Moreover, the proposed model can help optimize PV power plant operations.



**Citation:** Lim, S.-C.; Huh, J.-H.; Hong, S.-H.; Park, C.-Y.; Kim, J.-C. Solar Power Forecasting Using CNN-LSTM Hybrid Model. *Energies* **2022**, *15*, 8233. <https://doi.org/10.3390/en15218233>

Academic Editor: Alon Kuperman

Received: 25 September 2022

Accepted: 2 November 2022

Published: 4 November 2022

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** PV system; PV power generation forecasting; AI; deep learning; CNN; LSTM network

## 1. Introduction

The growing energy demand and development of various energy resources worldwide because of the increasing global population and industrialization have raised the amount of electrical power generated [1]. Renewable energy is considered an alternative to fossil fuels. Importantly, renewable energy, such as solar, wind, hydroelectric, biomass, and geothermal energy, is replenishable and is directly procured from nature without environmental destruction.

Photovoltaic (PV) energy generation is a crucial component of renewable energy generation. PV energy is abundant, clean, and environment-friendly; further, it has experienced a gradual increase in use in recent years [2–4]. Moreover, there is a significant acceleration in the installation of PV power generation systems worldwide. PV power generation involves directly converting solar energy into electrical energy. When solar rays are irradiated onto a semiconductor (*n*- and *p*-type silicon), electrons move between electrodes, generating electricity. Unlike fossil fuel power plants, the construction of a PV power plant is less complicated, emits zero pollution, and is environment-friendly [5].

However, the power generated from a PV power plant can vary since it is affected by the construction location, time, panel capability, and size [6,7]. Moreover, intermittent variability of a PV power plant occurs depending on weather conditions, such as solar radiation, temperature, humidity, cloud cover, and wind speed, that are unpredictable and uncontrollable. Notably, changes in temperature and solar radiation can substantially affect power generation [8,9].

The nature of such variables can lead to unstable PV power generation, causing a sudden surplus or reduction in power output. Furthermore, it may cause an imbalance between power generation and load demand, inducing control and operation problems in the power grid [10,11]. If the amount of power generation can be accurately forecasted, operation optimization strategies, such as peak shaving and reducing the uncertainty of a power generation system, can be effectively applied [12]. Therefore, a method for accurately predicting the amount of energy produced is crucial for PV-based power generation systems [13–15]. Accurate forecasting can improve the quality of power provided to a power grid and help reduce the costs related to general variability [16]. In addition, it can be used in various operation and control activities, including power scheduling in power distribution and transmission grids [17].

Research on PV power generation forecasting has recently gained considerable interest. Primarily, studies that employ deep learning models are receiving increasing attention [18–20]. Machine learning (ML)-based deep learning models were developed to solve complicated problems by extracting meaningful data from big data. Deep learning models are different from other theoretical ML models. Various hierarchical structures of deep learning models enable the automatic learning of the methods required to extract semantic features from raw data and find useful patterns.

Solar power generation has intermittent characteristics and is highly correlated with dependence on meteorological parameters. The use of various meteorological parameters can improve the forecasting accuracy of the model. Most conventional methods use multivariate regression, which requires collecting multiple relevant data such as solar radiation, temperature, and power generation. However, most PV plants in Korea have no environmental sensors installed owing to cost considerations; thus, the factors affecting PV power generation, such as temperature, humidity, cloud cover, and solar radiation, are challenging to identify. Therefore, only power generation data can be obtained. The Korea Meteorological Administration provides temperature and solar radiation data but it is not suitable for use due to location errors of PV plants. Motivated by this, we aim to present a deep model that forecasts power generation by analyzing time-series patterns in the data from a PV power plant at which environmental sensors are not installed.

We propose a hybrid model of a long short-term memory (LSTM) and convolutional neural network (CNN) model specialized for time-series data analysis and forecasting to improve the accuracy of power generation forecasting. The proposed network adopts a parallel structure of branched CNN-LSTM. First, a CNN offers the advantage of higher accuracy since different patterns can be identified depending on weather conditions based on the pattern extraction schemes. A proposed CNN model classifies the daily weather as sunny or cloudy by analyzing historical patterns in raw data. Then, the LSTM part is split into two models trained separately on sunny and cloudy day data and proceeds to extract long-term dependent features from the raw data. The LSTM model can more accurately predict the power generation by individually learning the power generation pattern according to each weather condition type.

The main purpose of this study is to enhance short-term PV forecasting accuracy by using the proposed hybrid model based on a univariate data approach. To train and evaluate the proposed method, using the power generation data collected from a PV power plant in Busan, Korea. The data were collected from 10 September 2019 to 22 July 2021. The data defined as corresponding to sunny and cloudy weather conditions by the Korea Meteorological Administration were used. The contributions obtained through the proposed method are as follows. We develop an accurate and simple CNN-LSTM model for PV forecasting that relies on historical PV data and does not rely on specific sensors or environment data.

The remaining paper is organized as follows. Section 2 discusses previous studies on power generation forecasting algorithms. In Section 3, the proposed hybrid CNN-LSTM model is introduced. The overview of the proposed analytical method, data cleaning, re-definition, and design verification of the LSTM model is presented. The performance of

the proposed model is evaluated using objective indicators in Section 4, and the conclusions are presented in Section 5.

## 2. Related Work

A decrease in the PV module price, timely forecasting of power generation, and policies related to renewable energy have demonstrated a synergistic effect. Renewable-energy power plants are actively being constructed, thereby increasing energy production. Power generation forecasting plays a vital role from an economic perspective. Moreover, PV power generation forecasting helps rationally plan power generation to effectively solve problems such as system stability and power generation balance. The power generation forecasting of renewable energy using physical, statistical, ML, and artificial neural network (ANN)-based methods have been consistently researched.

Physical models forecast PV power generation based on numerical weather prediction (NWP) and physical principles of PV cells [21]. An input of a physical model consists of dynamic information, such as NWP and environmental monitoring data, and static information, such as the installation angle of PV panels and the conversion efficiency of PV cells [22]. Though the method does not require past information, it depends on geographic information of PV panels and detailed weather data.

Statistical methods set mapping relationships between past time-series data and PV energy outputs [23]. The model performance depends on the relationship between past power generation observations and weather/climate parameters. Generally used statistical methods include autoregressive moving average, Kalman filter, and Markov chain [24–26].

An ML method forecasts power generation by analyzing the correlations in nonlinear data. Such a method involves learning the characteristics of sample data and making predictions using a neural network or training model such as a support vector machine (SVM). The cross-validated model in [27] used the characteristic statistical parameters of illumination intensity and surrounding temperature as the input vector. In [28], a dataset corresponding to four different environments—sunny, cloudy, foggy, and rainy—was used. The model used an SVM and weather data to classify a specific environment and select an appropriate model to predict power generation. A hybrid ML model combining an SVM and the random forest algorithm was developed in [29]. SVM predicts power generation, while the random forest was applied with an ensemble learning method wherein predicted values are combined and analyzed. Past and present generated power and weather data were combined as the input.

Artificial neural networks (ANNs), a subcategory of ML models, are widely used for forecasting [30]. An ANN is a nonlinear model consisting of completely connected nodes. The node connections consist of weights used to analyze specific patterns within the data. An ANN predicts PV power generation through nonlinear approximation. An ANN-based prediction algorithm obtained favorable results using a shallow neural network in the initial steps. For the real-time prediction of PV solar energy, a hybrid method combining ANN and random forest has been proposed [31]. An improved ANN model was successfully developed to enhance the PV energy forecasting accuracy [32] despite uncertainty in solar radiation. Another study predicted power generation using solar radiation and panel temperature as independent variables [33]. Furthermore, [34] used the nonlinear characteristics of ANNs to predict solar radiation using multiple variables, such as average temperature, relative humidity, and daytime hours.

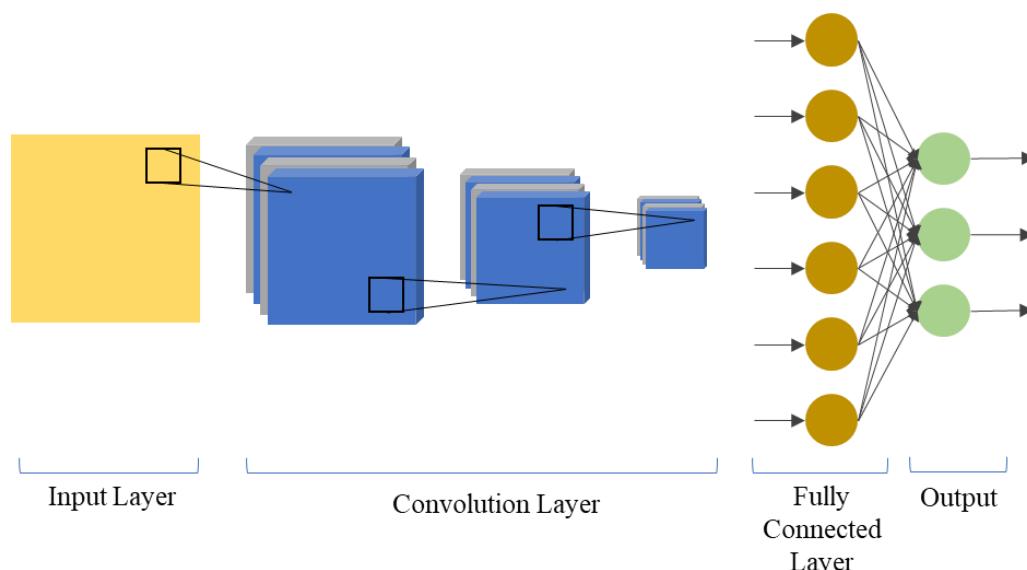
In recent years, numerous studies have been conducted on applying deep learning, which has shown excellent performance in diverse applications to power generation forecasting. Unlike ML algorithms, deep learning learns data and patterns using a complicated neural network architecture. The deep-learning-based recurrent neural network (RNN) is an ANN-type network that uses sequential or time-series data [35]. This model uses information from the previous input to determine the current input and output. A feedback loop for the recurrent layer saves information about a point in time in the past time-series data of a memory cell. An RNN contains a shared parameter in each layer. A feed-forward

network has different weights for each node, but an RNN shares the same weight parameter in each network layer. This weight is adjusted through backpropagation and gradient descent during training. Such characteristics of RNNs have been applied to PV power generation forecasting [36,37].

### 3. Proposed Method

#### 3.1. CNN for Feature Extraction

CNN, a deep learning algorithm frequently used for image, text, and signal inputs, consists of stacked layers that extract object features. Figure 1 shows the basic architecture of a CNN.

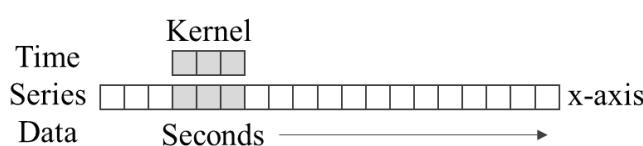


**Figure 1.** Standard convolutional neural network (CNN) architecture.

The model performance is based on the number of stacked layers and the type and size of the kernel. The data in the input layer goes through convolution and pooling layers to extract deep features. These features are then input in the fully connected layer, where the result values are classified.

A CNN is used for extracting hierarchical features from an image. Accordingly, a CNN can extract important information from the one-dimensional sequential and two-dimensional input data.

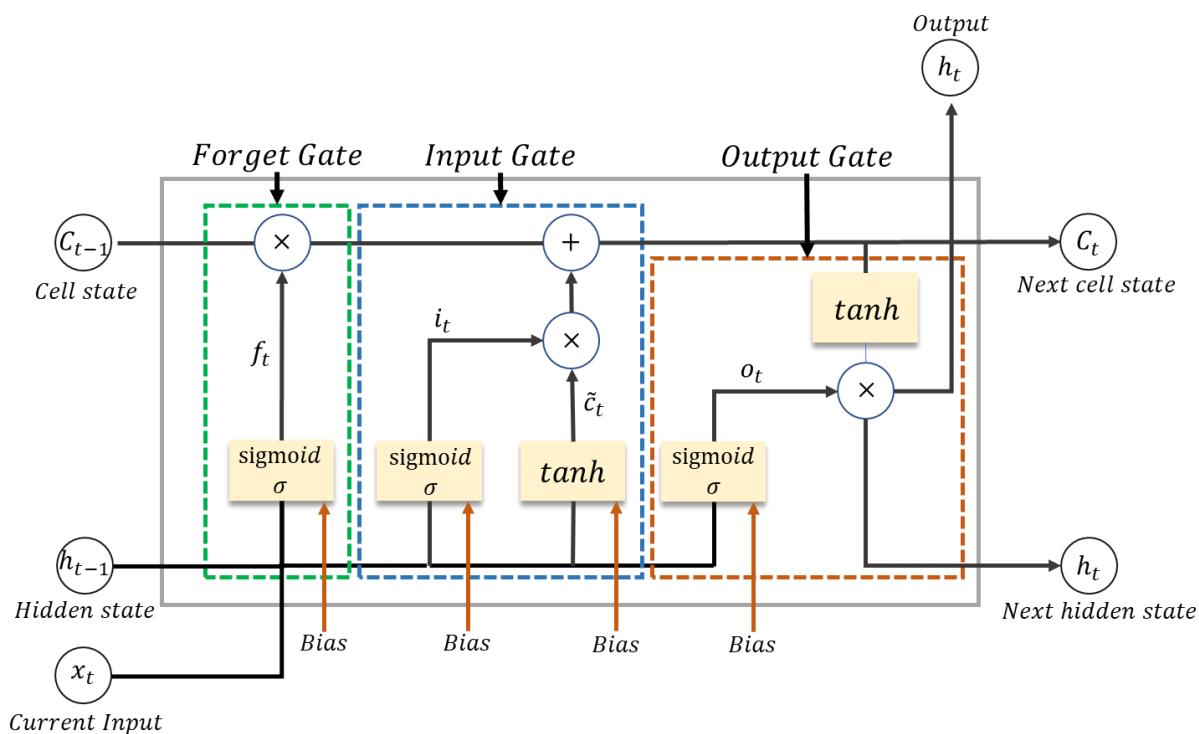
1D-CNNs are frequently used in natural language or time series processing since they can handle sequential data. Unlike in 2D-CNNs, the kernel executing convolution and the sequence of data being applied have a one-dimensional form in 1D-CNNs. As shown in Figure 2, a 1D-CNN has a kernel moving along a single dimension.



**Figure 2.** Computation process of 1D-CNN.

#### 3.2. Long Short-Term Memory

LSTM is one of the RNN architectures used for processing time-series data while solving the problem of gradient exploding or gradient vanishing. This network forgets unnecessary information while storing information for a prolonged time. Figure 3 shows the cell structure of LSTM.



**Figure 3.** Cell diagram of LSTM.

LSTM receives current input data and long- and short-term memory of the previous cell in each time step. Short-term memory represents the hidden state and represents  $h_{t-1}$ , while long-term memory represents the cell state  $C_{t-1}$ . A cell adjusts the information to be maintained or discarded in each time step before delivering short-term and long-term information to the next cell using a gate. This gate is called the input, forget, or output gate and accurately performs filtering through training.

The first step of LSTM is identifying and removing unnecessary information in a memory cell. This process is performed in the forget gate that determines an output value between 0 and 1 based on a sigmoid function. The closer the value is to 1, the more information about a previous state is maintained. The information of a previous state is forgotten as the value approaches 0, and omitted parts are decided. After passing the forget gate, the information to be stored is selected. If the previous time is forgotten, new information to be remembered is added, and the value of each element is decided as the newly added information. In this case, new information is not stored in the memory cell unconditionally; instead, an appropriate value is selected using an input gate.

The sigmoid function is added with the last LSTM cell and the current state and time activation feature. The value passed through the sigmoid layer is expressed as a number between 0 and 1 and indicates the degree of new information being updated. The value that has passed through the tanh function has a value between  $-1$  and  $1$ . Then, the output is multiplied, and the final value is stored in the long-term memory. The following process is used to select the output information. The output gate generates a new short-term memory (hidden state) to be delivered to the cell in the next step using the current input, previous short-term memory, and newly generated long-term memory. The output of the current time step can also be imported from the hidden state. The short- and long-term values generated by this gate are transferred to the next cell as the process is repeated. The output of each time step can be obtained from short-term memory or a hidden state.

$$f_t = \sigma(x_t w_f + h_{t-1} w_f + bias_f) \quad (1)$$

$$g_t = \tanh(x_t w_g + h_{t-1} w_g + bias_g) \quad (2)$$

$$i_t = \sigma(x_t w_i + h_{t-1} w_i + bias_i) \quad (3)$$

$$c_t = f_t \circledast c_{t-1} + g_t \circledast i_t \quad (4)$$

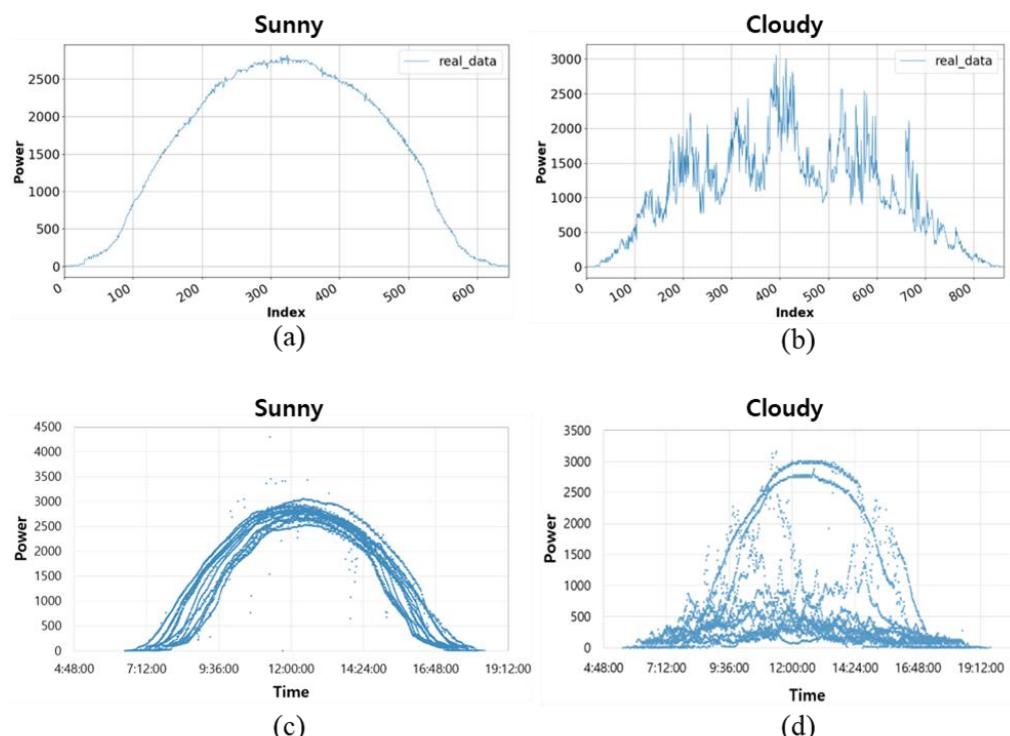
$$o_t = \sigma(x_t w_o + h_{t-1} w_o + bias_o) \quad (5)$$

$$h_t = o \circledast \tanh(c_t) \quad (6)$$

Equations (1)–(6) are applied to each gate. Equation (1) shows the process of the forget gate. Input data are updated through Equations (2)–(4). The data undergo an operation, as in Equation (5), in the last output gate and are updated as per the final Equation (6). Here,  $\sigma$  is the sigmoid function,  $w$  is the weight matrix,  $h$  and  $c_t$  represent the hidden and cell states, respectively.

### 3.3. Adaptive Selector

This paper proposed a hybrid CNN-LSTM model to forecast power generation. The CNN classifies weather type, while the LSTM predicts power generation. The LSTM model learns the power generation patterns according to weather type, which reduces the complexity and variability of data fitting to improve prediction accuracy. Figure 4 is the graph of power generation on sunny and cloudy days.

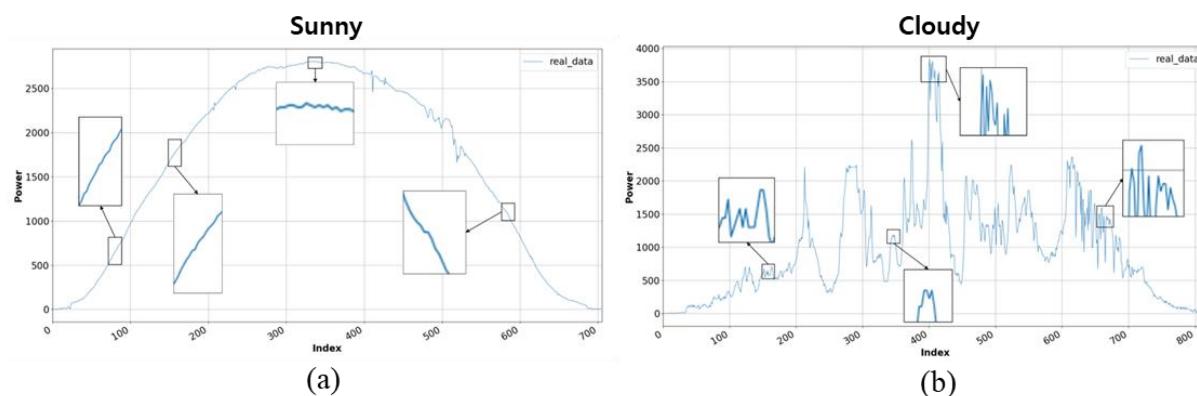


**Figure 4.** PV power generation graphs; (a) sunny day, (b) cloudy day, (c) scatter plot of sunny days, (d) scatter plot of cloudy days.

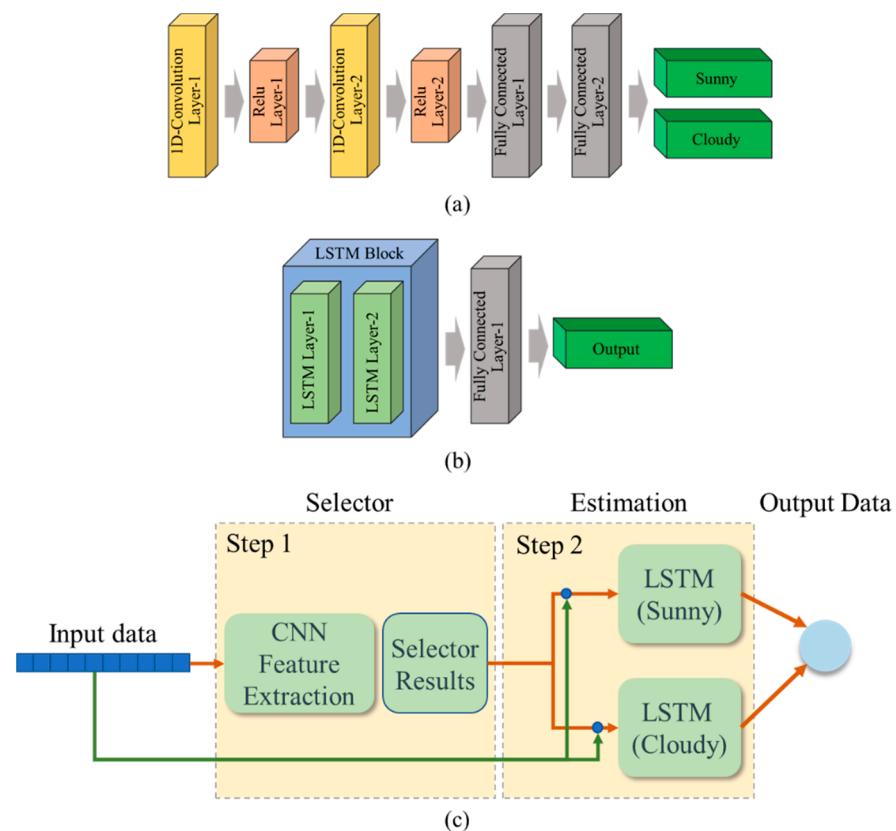
The power generation differs on a sunny and cloudy day, as shown in Figure 4a,b, respectively. The power generation on a cloudy day is low and highly inconsistent. On a sunny day, the graph has a semi-circular data distribution affected by sunrise and sunset; the power generation is the highest around noon, as indicated by the high elevation. On a cloudy day, the variability of power generation is severe due to the changes in solar radiation. Figure 4c,d shows that the power generation data on clear and cloudy days are superimposed and expressed as a scatter plot. In Figure 4c, it can be seen that the power generation production pattern on a clear day is regularly generated within a certain area. In this figure, there are cases in which the power generation data is out of range by one point. This is an outlier due to equipment error or electrical noise. In Figure 4d, it can be seen that

the power generation pattern at the time of cloudy days has high variability according to the non-linear characteristics.

Figure 5a,b show the sectional power generation patterns during sunny and cloudy days, respectively. On a sunny day, the power generation gradually increases or decreases, but the variability abruptly changes on a cloudy day. Since such data patterns also influence continuous time-series patterns, using the LSTM model for training can complicate data convergence. Therefore, the current weather conditions are first classified according to the patterns in Figure 5, and the power generation is forecasted using the individual LSTM model according to the classification results. Boxes in the Figure 5 show the detailed pattern form of the collected power generation data. The data of sunny day is generated with a regular period, whereas data of cloudy day has a large fluctuation range and does not show periodicity. The proposed hybrid CNN-LSTM model is shown in Figure 6.



**Figure 5.** Sectional patterns of PV power generation; (a) sunny day, (b) cloudy day.



**Figure 6.** Proposed CNN-LSTM hybrid model; (a) CNN architecture for weather classification, (b) LSTM architecture for PV power generation forecasting, (c) Overall CNN-LSTM hybrid model.

Figure 6a shows the CNN architecture used to classify weather conditions. This architecture consists of two layers of convolution with relu layer and two layers of fully connected layers. The CNN model is trained to classify sunny and cloudy days based on the data pattern of weather conditions. The last layer outputs the results of the weather conditions. In the inner process of the neural network, the difference between the predicted values and the labels is calculated using a loss function. It then uses a backpropagation algorithm to minimize the loss function so that the predicted values are as close to the labels as possible. In the course of training a neural network, parameters, including weights and biases, are fine-tuned and updated to compare predicted values to labels to produce better predictions at every epoch. The optimizer used Adam and set learning rate = 0.01, the coefficient for primary momentum  $\beta_1 = 0.9$ , the coefficient for secondary momentum  $\beta_2 = 0.999$ , and epsilon =  $10^{-8}$ . The weight initialization used kaiming uniform initializer and adopted the cross-entropy function as the loss function. Figure 6b shows the LSTM architecture used to forecast PV power generation. This architecture consists of two layers of LSTM and one layer of the fully connected layer. The last layer outputs the predicted power generation. The two LSTM models with the same structure are trained on sunny and cloudy days data, respectively. In the case of the LSTM model, the input, forget and output gates and cell-state help the model to determine which values to preserve for keeping long-term memory during the computational process. The hyper-parameters used for network training are as follows. The optimizer used Adam and set learning rate = 0.001, the coefficient for primary momentum  $\beta_1 = 0.9$ , the coefficient for secondary momentum  $\beta_2 = 0.999$ , and epsilon =  $10^{-8}$ . The weight initialization used kaiming uniform initializer and adopted MSE (Mean Square Error) function as the loss function. Figure 6c shows the overall structure of the CNN-LSTM hybrid model. Select an LSTM model according to the output of CNN and predict PV power using time series data.

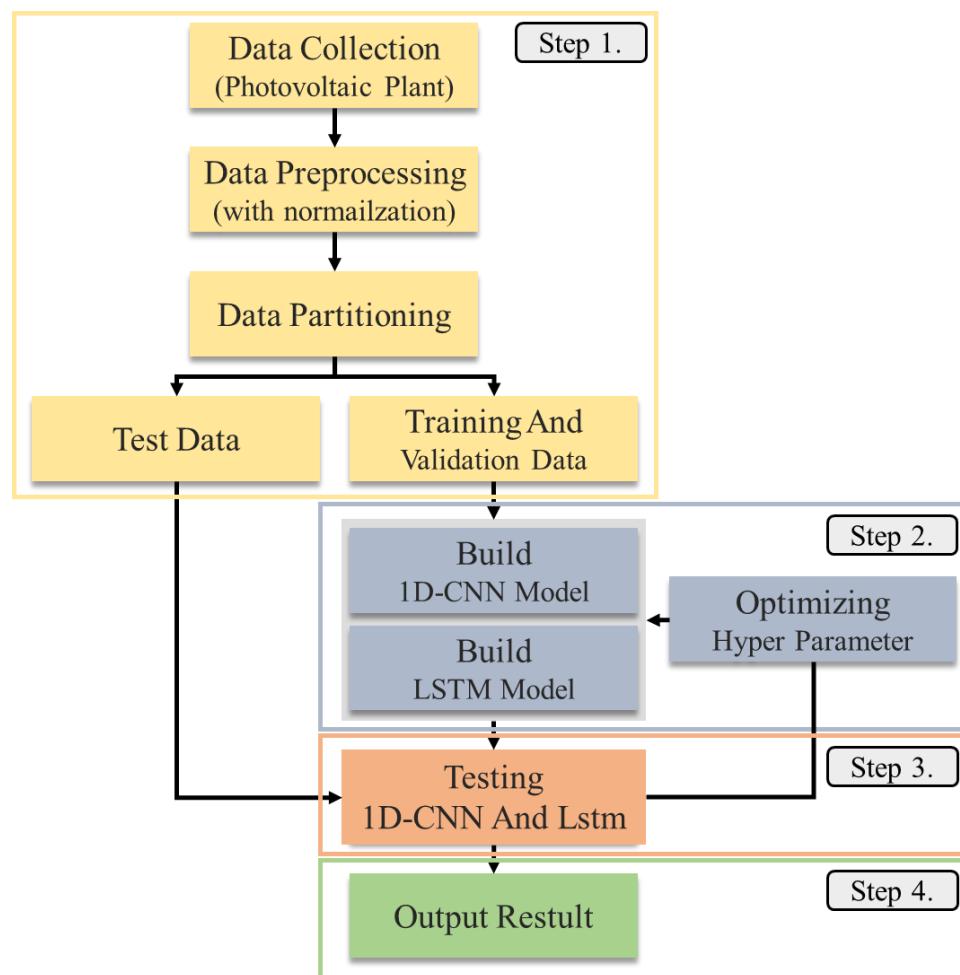
#### 4. Experiment

In order to evaluate the effectiveness of the proposed deep model, we designed the following experiments using collected PV power datasets. The proposed hybrid model was trained and tested using the Python 3.8-based Pytorch deep learning framework as a software platform along with an AMD Ryzen 5 5600X CPU operating at 3.70 GHz, NVIDIA RTX 2070 GPU, and 16 GB RAM hardware environment.

The proposed hybrid CNN- LSTM model performs PV power generation forecasting, as shown in Figure 7. In step 1, the power generation data generated from a PV power plant are classified into sunny and cloudy day data. In step 2, the collected data are pre-processed to remove noise elements, such as missing values or outliers affecting the experiment. Normalization is then applied to use the data as a network input. In step 3, the CNN and LSTM are trained using the training data. Each model only uses power generation data. In step 4, the results of the forecasting model are tested and verified using various matrices.

##### 4.1. Data Collection and Pre-Processing

This study used the power generation data collected from a PV power plant in Busan, Korea. The average annual PV power generation time of the region is 3.5 h, the average wind speed is 4–5 m/s, and the meridian altitude is 53° in spring (March–May), 76.5° in summer (June–August), 53° in fall (September–November), and 29.5° in winter (December–January). The data were collected from 10 September 2019 to 22 July 2021. The data defined as corresponding to sunny and cloudy weather conditions by the Korea Meteorological Administration were used. The Korea Meteorological Administration defines weather conditions in 11 steps. In this study, steps 1–5 were classified as sunny, while steps 6–11 were classified as cloudy [38]. Table 1 presents the schema of the collected data.



**Figure 7.** Flowchart of the proposed PV power generation forecasting method.

**Table 1.** Schema of data collected in Busan.

Column	Data Type	Default
Date	Date	Null
Time	Time	Null
Cumulative Power	Double (22, 0)	"0"

Data and time in Table 1 represented when the data were collected; cumulative power indicates cumulative power generation. The power generation data of 682 days were used for the experiment. The dataset consists of 446 and 236 sunny and cloudy day data points, respectively. Data imbalance may occur since the number of sunny day data is 1.88 times greater than the cloudy day data. Therefore, 100 data from each class were used for network training, while the remaining data were used for testing.

Abnormal elements in the input data of a forecasting model can cause high forecasting errors. Pre-processing the input data can improve model accuracy by reducing computation costs and solving the inappropriate training problem. PV power data sets can contain outliers due to solar equipment problems, collection system errors, and software system problems. Therefore, to eliminate outliers, the datasets are cleansing by calculating the 75% quantile.

$$\text{Quartile3} + 1.5 \times \text{IQR} \quad (7)$$

The Cumulative power data in Table 1 is preprocessed using the interquartile range (IQR). If the value of data was higher than that provided by Equation (7), it was determined as an outlier and removed.

Missing values in the collected data can occur due to problems in the inverter or data collection device. Since the proposed model predicts the data of a future point in time using continuous time-series data, missing values at  $x_t$  can affect the model's performance. Thus, time interpolation using the values at times  $x_{t-1}$  and  $x_{t+1}$  were applied.

The numerical scale of the power generation data used in this study is significant. For example, power generation is less than 10 W near 06:00 when the amount of solar radiation is low and more than 2500 W between 12:00–14:00 when the amount of solar radiation is the highest. A significant difference between the minimum and maximum values can affect the training speed and convergence accuracy of a network. Therefore, using Equation (8), normalization was performed to improve computation speed and accelerate the network training convergence speed.

$$z = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (8)$$

Here,  $x$  is the actual data, and  $z$  indicates a normalized value.

The data used in this study are the power generation data per minute, obtained by dividing the cumulative power generation data by power generation time. This study assumed a power plant without environmental sensors and predicted power generation using time-series analysis characteristics. Clouds and rain that occur locally reduce the amount of solar radiation, thus affecting the PV power generation efficiency. However, they are difficult to predict when environmental sensors are unavailable. Therefore, this study aimed to predict power generation based only on power generation patterns. The network receives 20 time-series data and outputs one data. The hyper-parameters used for training the network are as follows. The mean squared error (MSE) was used as the loss function, and the Adam optimizer was used for optimization. A learning rate of 0.01 was applied to network optimization.

#### 4.2. Model Evaluation

In this study, a hybrid model combining CNN and LSTM is established. The general indicators used for evaluating the performance of a time-series data forecasting model include mean absolute error (MAE), root mean square error (RMSE), and mean absolute percentage error (MAPE). Let  $N$  be the number of test data,  $x_{pred}$  be the value predicted by the proposed algorithm, and  $x_{act}$  be the actual value in the quantitative indicators.

MAE measures the error between predicted and actual values and can vary depending on the measure of continuous variables, as shown in Equation (9). A smaller value indicates a higher accuracy.

$$MAE = \frac{1}{N} \sum_{i=1}^n |x_{pred} - x_{act}| \quad (9)$$

RMSE, defined as in Equation (10), measures the difference between predicted and actual values; it is a measure of the deviation between predicted and actual values. An RMSE closer to 0 indicates better performance. The difference between RMSE and MAE is that RMSE is sensitive to outliers, making it more susceptible to significant deviation. Since the error is squared, a larger error leads to weight being reflected a greater extent.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_{predicted} - x_{real})^2} \quad (10)$$

MAPE, given by Equation (11), indicates the extent of error in the predicted value. A value closer to 0 can be interpreted as a more outstanding forecasting model performance.

MAPE is robust to outliers, but in this case, it is more challenging to check for errors intuitively than in the case of MAE.

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{x_{predicted} - x_{real}}{x_{real}} \right| \times 100\% \quad (11)$$

The power generation forecasted using the LSTM entails errors. The magnitude of errors varies depending on the density of the predicted power generation data. The error between predicted and actual values is smaller when the density of predicted values is dense. An error becomes more significant when the density is sparse. The accuracy of a forecasting model is also determined by density. As shown in Equation (12),  $R^2$  measures the strength of a correlation between the predicted and actual values.  $R^2$  has the range of  $0 \leq R^2 \leq 1$ . A value closer to 0 indicates very low accuracy, and a value closer to 1 indicates higher accuracy of the forecasting model.

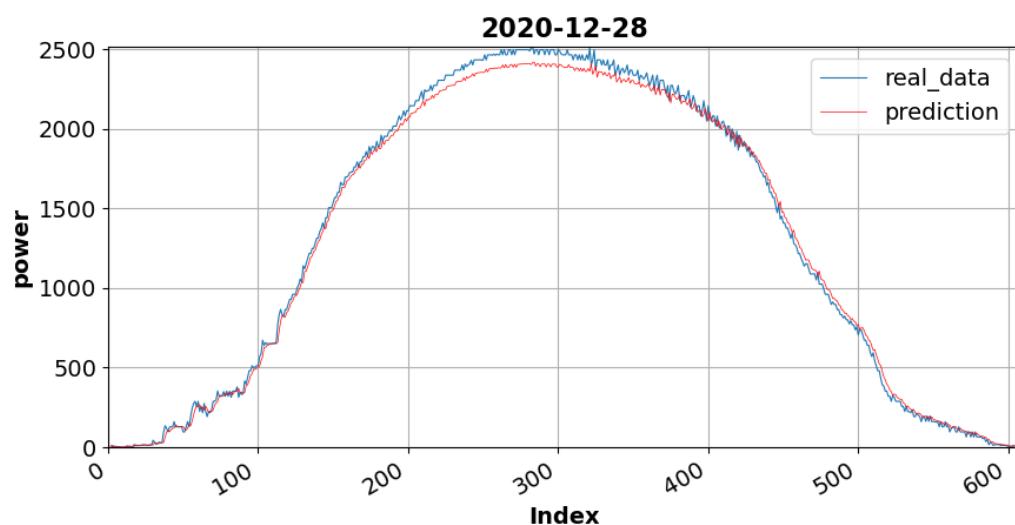
$$R^2 = 1 - \frac{\sum_{i=1}^n (x_{predicted} - x_{real})^2}{\sum_{i=1}^n (\bar{x}_{real} - x_{real})^2} \quad (12)$$

where  $\bar{x}_{real}$  is the average of the actual power generation data.

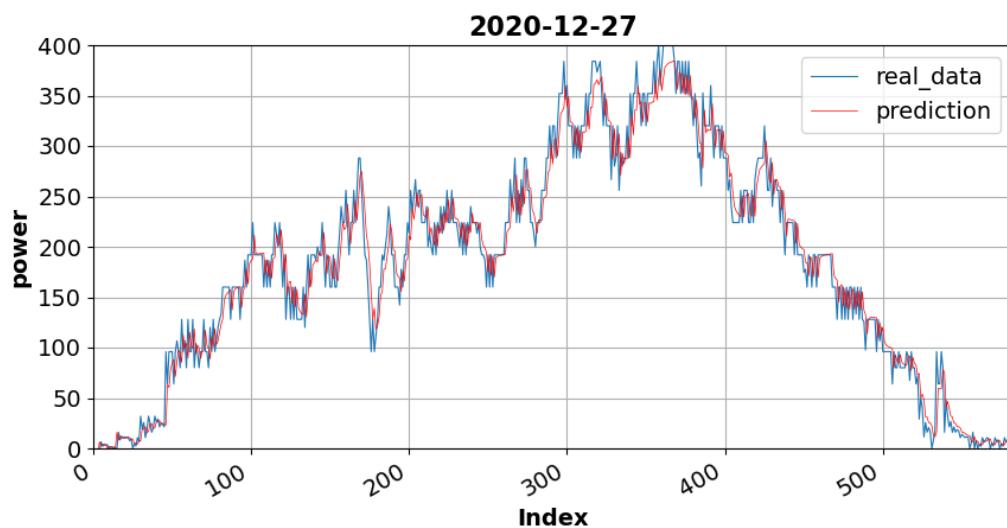
#### 4.3. Experimental Results

Quantitative and qualitative evaluations were performed using sunny and cloudy days' data for model verification. The data of the same month (sunny day: 28 December 2020, cloudy day: 27 December 2020) were used for uniformity of the validation process.

Figure 8 is the graph of the sunny day dataset; certain sections demonstrate power generation fluctuations due to clouds. Figure 9 is the graph of the cloudy day dataset in which the maximum power output is 400 W. In contrast, on a sunny day, the maximum power output is 2500 W. This result implies that the amount of solar radiation is minimal, resulting in severe fluctuations in overall power generation. The blue line shows the measured data from the device, while the red line shows the prediction data. A quantitative evaluation was performed using MAPE, RMSE, and MSE to validate the LSTM forecasting model. The LSTM model adequately followed the trend of the observation data for accurate forecasting.



**Figure 8.** Forecasting result of sunny day power generation data.

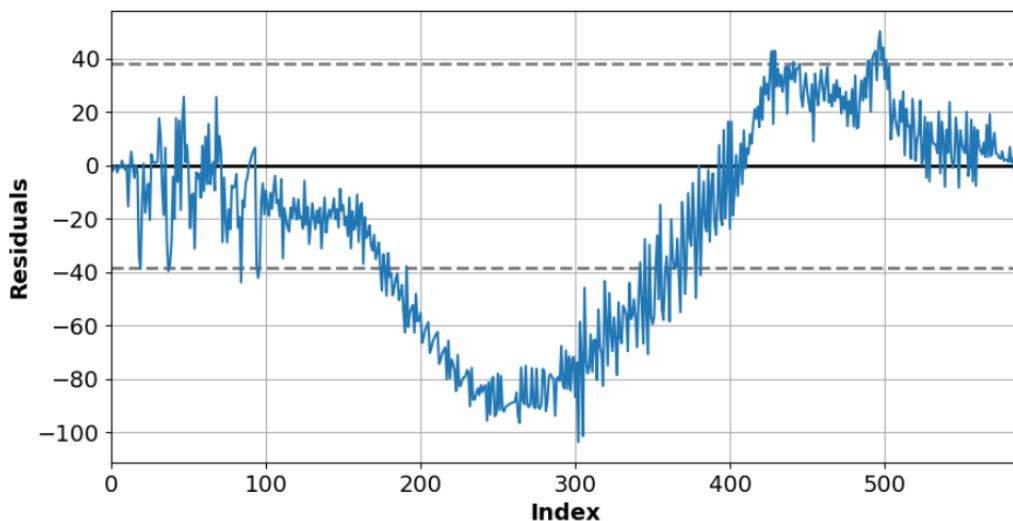


**Figure 9.** Forecasting result of cloudy day power generation data.

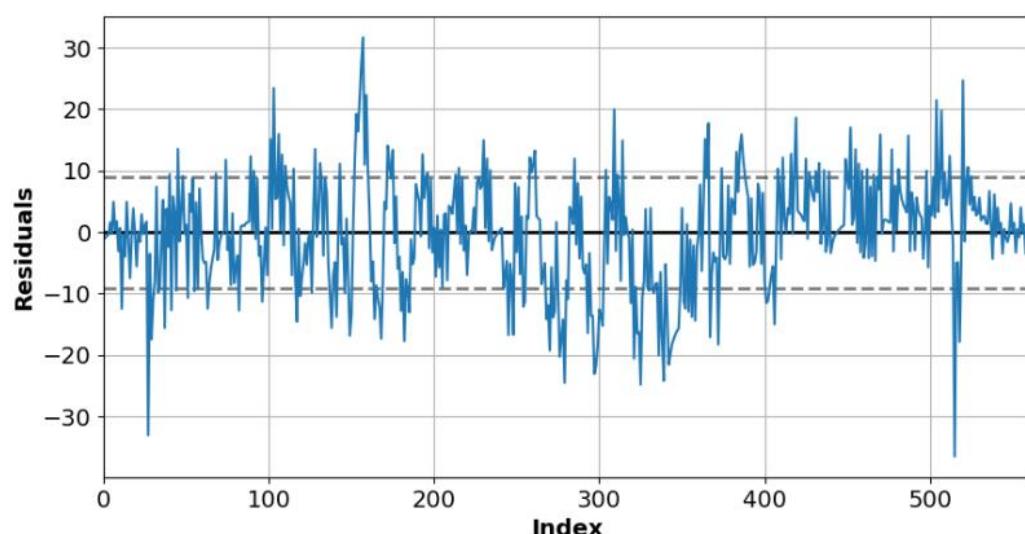
Table 2 presents the results of a quantitative evaluation of the test data for sunny and cloudy days. The sunny day data showed a MAPE of 4.58, which is lower than that of the cloudy day data, 7.06. In the cloudy day data, peak point errors occurred in the fluctuation amplitude according to the power production period, which increased the MAPE value compared to that of the sunny day data. The RMSE and MAE were lower in the cloudy day data, thus exhibiting more outstanding results. Figures 10 and 11 plot the residuals between observed and forecasted values for the sunny and cloudy day test data. The dotted line indicates the standard deviation (SD) corresponding to the residuals.

**Table 2.** Quantitative validation results of the power generation forecasting model.

Column	MAPE	RMSE	MAE	R <sup>2</sup>
Sunny	4.58	43.87	34.00	0.99
Cloudy	7.21	9.09	6.97	0.99



**Figure 10.** Residual graph of forecasting results using power generation data of sunny days.



**Figure 11.** Residual graph of forecasting results using power generation data of cloudy days.

Table 3 presents the residual comparison results for a sunny day. The SD of residual is 38.25, and the mean is 34.00 on a sunny day. Table 4 presents the residual comparison results of a cloudy day. The SD of residual is 27.42, and the mean is 17.13 on a cloudy day. A relatively large number of the residual values of sunny days deviate from the standard deviation section for the number of data, which can be observed between index 200 and 400 in Table 3. This section is the peak point region where the power generation is the largest in Figure 8.

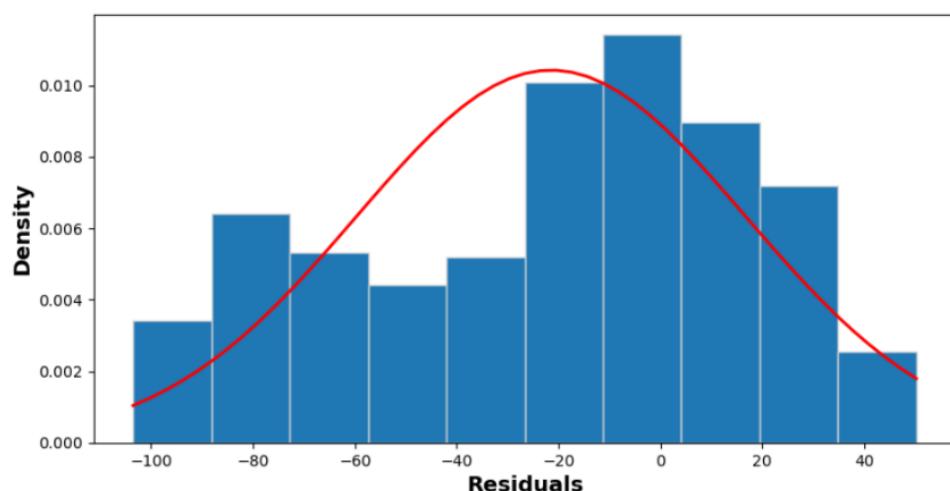
**Table 3.** Comparison between observed and hybrid-model estimated values of DC power: Sunny day.

Index	DC Power (Observed)	Model Estimated	Difference
1	9	9	0
100	864	854	10
200	2272	2216	56
300	2464	2377	87
400	1984	1960	24

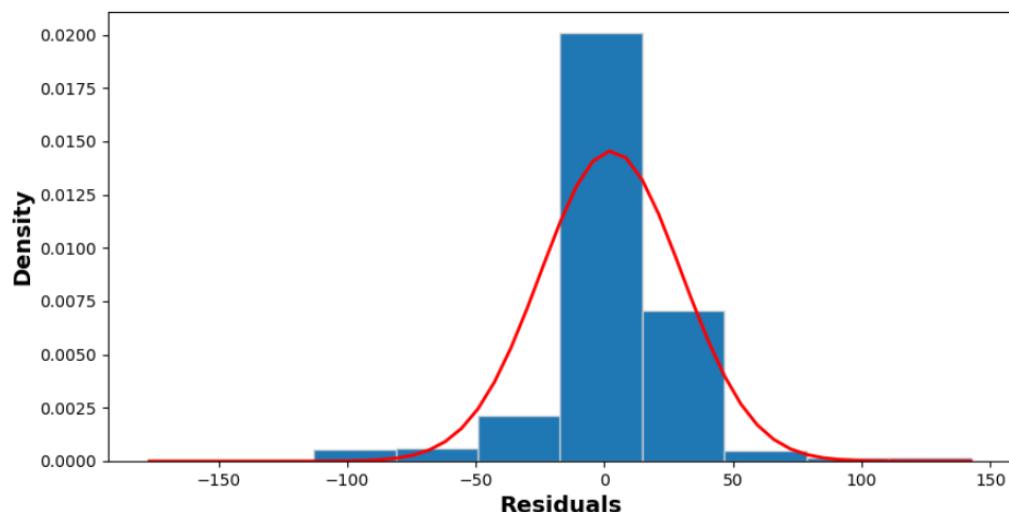
**Table 4.** Comparison between observed and hybrid-model estimated values of DC power: Cloudy day.

Index	DC Power (Observed)	Model Estimated	Difference
1	10	9	1
100	192	197	-5
200	213	214	-1
300	373	360	13
400	266	258	8

The forecasting model can generate errors in predicting the maximum power generation value. The error in the cloudy day data is 10.83 less than that for the sunny day data. The number of data deviating from the standard deviation section is less than the number of data used in the experiment. Figures 12 and 13 show the frequency distribution of residuals with an SD of 38.25 and a mean of 34 for sunny days and an SD of 27.42 and a mean of 17.13 for cloudy days, respectively.



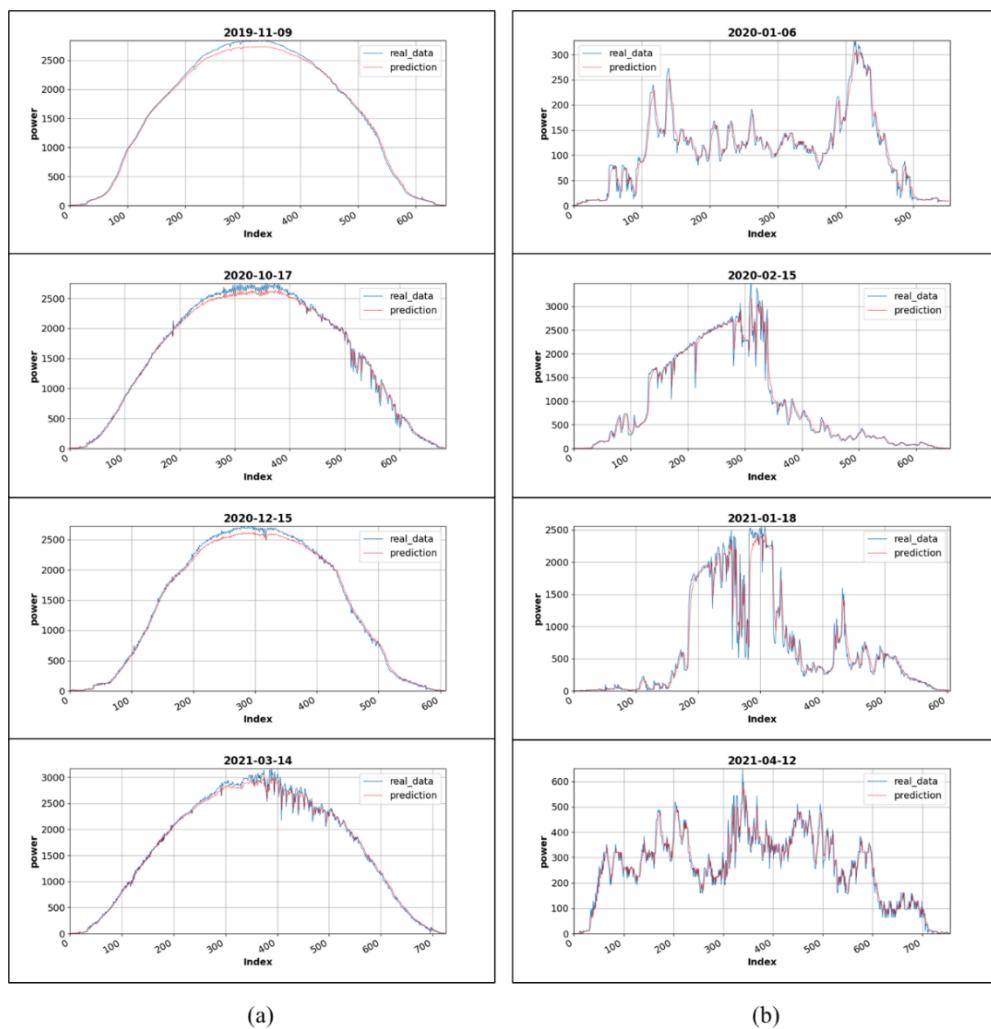
**Figure 12.** Histogram of residuals between predicted and actual values on a sunny day.



**Figure 13.** Histogram of residuals between predicted and actual values on a cloudy day.

The advantage of the LSTM model is further highlighted when using unfavorable weather patterns. Figure 14a,b show the qualitative evaluation results of various test data for the sunny and cloudy day data, respectively. The proposed algorithm adequately follows the trends in various power generation patterns for accurate forecasting. However, errors occur at peak points for the cloudy data with irregular patterns and large fluctuations. However, the algorithm is robust to abrupt changes in the patterns and fluctuation periods, thus reasonably following the trend for reliable forecasting. A qualitative evaluation confirms the high efficiency and improved performance of the proposed model in solving the PV power generation forecasting problem.

The power generation forecasting performance of the proposed model differs depending on the weather patterns and validation datasets. The characteristics of power generation fluctuation patterns are sufficiently captured by the CNN, while the LSTM finds long-term dependencies in the time-series input. In other words, the CNN-LSTM hybrid model may not produce the same result in an environment having different weather conditions. For example, the proposed model produced errors in the power generation peak points for the sunny day data but made accurate predictions by following the fluctuations pattern for the cloudy day data having abrupt changes.



**Figure 14.** Qualitative evaluation of trends in validation data; (a) sunny day, (b) cloudy day.

## 5. Conclusions

A PV power generation forecasting model can improve prediction accuracy according to weather conditions and enhance the planning, operation, and stability of PV power systems. However, PV power generation forecasting can be challenging owing to intermittency in weather conditions. A statistical method for inferring dependency between past and short-term observed data is required to build an effective forecasting model that depends only on past data, excluding the solar radiation data highly correlated with power production.

This paper proposes a CNN-LSTM hybrid model for PV power generation forecasting. The proposed model overcomes the drawbacks of the individual models while preserving their advantages. Because training one LSTM model using different time-series data may affect network convergence, separate models were built according to weather conditions. Weather patterns were classified using CNN, and the LSTM model was applied to the classification results. The power generation patterns can clearly distinguish between sunny and cloudy days. The proposed forecasting model can sufficiently reflect the fluctuations in power generation for accurate forecasting. A qualitative evaluation confirmed that the forecasted power output signals react to fluctuations and adequately follow the actual power output signal trend.

Moreover, a quantitative evaluation demonstrated that the proposed model has RMSEs of 4.58 and 27.55, MAEs of 34.00 and 17.13, and MAPEs of 4.58 and 7.06 for sunny and cloudy day data, respectively. The residual distributions had an SD of 38.25 and a mean of 34.00 for sunny days and an SD of 27.42 and a mean of 17.13 for cloudy days, respectively,

thus establishing the validity of the proposed model. The characteristics of collected data and power generation values may vary depending on the inverter manufacturer. Since the power generation capacity of a PV power plant system varies depending on installation scale and weather conditions, the model requires reconfiguration according to the power generation capacity. Future research should focus on methods for forecasting power generation by adaptively reflecting the power generation capacity of a PV power plant. This includes the application of optimization techniques that automatically perform model fitting according to the data accumulated in the system to improve the prediction accuracy; In addition, it is necessary to analyze factors affecting power generation efficiencies such as PV panel characteristics and solar radiation. Additionally, the forecasting results are expected to be used to understand the decrease in inverter efficiency over time. In cases where inverter power generation is less than the predicted generation, it will help determine whether the phenomenon is an efficiency decline due to the aging of the inverter.

**Author Contributions:** Conceptualization, S.-C.L., C.-Y.P. and J.-C.K.; Data curation, S.-C.L., J.-H.H., S.-H.H. and J.-C.K.; Formal analysis, S.-C.L. and J.-C.K.; Funding acquisition, C.-Y.P. and J.-C.K.; Investigation, S.-C.L., J.-H.H., S.-H.H., C.-Y.P. and J.-C.K.; Methodology, S.-C.L., J.-H.H., S.-H.H., C.-Y.P. and J.-C.K.; Project administration, S.-H.H., C.-Y.P. and J.-C.K.; Resources, S.-C.L., J.-H.H., S.-H.H., C.-Y.P. and J.-C.K.; Software, S.-C.L., J.-H.H., S.-H.H., C.-Y.P. and J.-C.K.; Supervision, S.-H.H., C.-Y.P. and J.-C.K.; Validation, S.-C.L., S.-H.H., C.-Y.P. and J.-C.K.; Visualization, S.-C.L., J.-H.H. and J.-C.K.; Writing—original draft, S.-C.L., J.-H.H., S.-H.H., C.-Y.P. and J.-C.K.; Writing—review & editing, S.-C.L., J.-H.H., C.-Y.P. and J.-C.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korean government (MOTIE) (20203040010130, Development and demonstration of remote intelligent operation and maintenance (O&M) technology of MW-class solar power plant using 5G technology).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Antonanzas, J.; Osorio, N.; Escobar, R.; Urraca, R.; Martinez-De-Pison, F.J.; Antonanzas-Torres, F. Review of Photovoltaic Power Forecasting. *Sol. Energy* **2016**, *136*, 78–111. [[CrossRef](#)]
2. Kim, J.C.; Huh, J.H.; Ko, J.S. Improvement of MPPT Control Performance Using Fuzzy Control and VGPI in the PV System for Micro Grid. *Sustainability* **2019**, *11*, 5891. [[CrossRef](#)]
3. Eseye, A.T.; Zhang, J.; Zheng, D. Short-Term Photovoltaic Solar Power Forecasting Using a Hybrid Wavelet-PSO-SVM Model Based on SCADA and Meteorological Information. *Renew. Energy* **2018**, *118*, 357–367. [[CrossRef](#)]
4. Wang, H.; Yi, H.; Peng, J.; Wang, G.; Liu, Y.; Jiang, H.; Liu, W. Deterministic and Probabilistic Forecasting of Photovoltaic Power Based on Deep Convolutional Neural Network. *Energy Convers. Manag.* **2017**, *153*, 409–422. [[CrossRef](#)]
5. Park, C.Y.; Hong, S.H.; Lim, S.C.; Song, B.S.; Park, S.W.; Huh, J.H.; Kim, J.C. Inverter Efficiency Analysis Model Based on Solar Power Estimation Using Solar Radiation. *Processes* **2020**, *8*, 1225. [[CrossRef](#)]
6. Pelland, S.; Remund, J.; Kleissl, J.; Oozeki, T.; Brabandere, K.D. Photovoltaic and Solar Forecasting: State of the Art. *IEA PVPS Task* **2013**, *14*, 1–36.
7. Pedro, H.T.C.; Coimbra, C.F.M. Assessment of Forecasting Techniques for Solar Power Production with no Exogenous Inputs. *Sol. Energy* **2012**, *86*, 2017–2028. [[CrossRef](#)]
8. Rettger, P.; Keshner, M.; Pligavko, K.A.; Moore, J.; Littmann, W.B. Dynamic Management of Power Production in a Power System Subject to Weather-Related Factors. U.S. Patent 2010/0198420 A, 5 August 2010.
9. Qing, X.; Niu, Y. Hourly Day-Ahead Solar Irradiance Prediction Using Weather Forecasts by LSTM. *Energy* **2018**, *148*, 461–468. [[CrossRef](#)]
10. Ko, J.S.; Huh, J.H.; Kim, J.C. Overview of Maximum Power Point Tracking Methods for PV System in Micro Grid. *Electronics* **2020**, *9*, 816. [[CrossRef](#)]
11. Sobri, S.; Koohi-Kamali, S.; Rahim, N.A. Solar Photovoltaic Generation Forecasting Methods: A review. *Energy Convers. Manag.* **2018**, *156*, 459–497. [[CrossRef](#)]

12. Husein, M.; Chung, I.-Y. Day-Ahead Solar Irradiance Forecasting for Microgrids Using a Long Short-Term Memory Recurrent Neural Network: A Deep Learning Approach. *Energies* **2019**, *12*, 1856. [[CrossRef](#)]
13. European Photovoltaic Industry Association. *Connecting the Sun: Solar Photovoltaics on the Road to Large-Scale Grid Integration*; Technical Report; EPIA: Brussels, Belgium, 2012.
14. Porter, K.; Rogers, J. *Survey of Variable Generation Forecasting in the West: August 2011–June 2012*; National Renewable Energy Laboratory (NREL): Golden, CO, USA, 2012.
15. Tapakis, R.; Charalambides, A. Equipment and Methodologies for Cloud Detection and Classification: A Review. *Sol. Energy* **2013**, *95*, 392–430. [[CrossRef](#)]
16. Yang, H.T.; Huang, C.M.; Huang, Y.C.; Pai, Y.S. A Weather-Based Hybrid Method for 1-Day Ahead Hourly Forecasting of PV Power Output. *IEEE Trans. Sustain. Energy* **2014**, *5*, 917–926. [[CrossRef](#)]
17. Eftekharnejad, S.; Heydt, G.T.; Vittal, V. Optimal Generation Dispatch with High Penetration of Photovoltaic Generation. *IEEE Trans. Sustain. Energy* **2015**, *6*, 1013–1020. [[CrossRef](#)]
18. Zhang, Q.; Tian, X.; Zhang, P.; Hou, L.; Peng, Z.; Wang, G. Solar Radiation Prediction Model for the Yellow River Basin with Deep Learning. *Agronomy* **2022**, *12*, 1081. [[CrossRef](#)]
19. Brahma, B.; Wadhvani, R. Solar Irradiance Forecasting Based on Deep Learning Methodologies and Multi-Site Data. *Symmetry* **2020**, *12*, 1830. [[CrossRef](#)]
20. Aslam, M.; Lee, J.M.; Kim, H.S.; Lee, S.J.; Hong, S. Deep Learning Models for Long-Term Solar Radiation Forecasting Considering Microgrid Installation: A Comparative Study. *Energies* **2020**, *13*, 147. [[CrossRef](#)]
21. Bakker, K.; Whan, K.; Knap, W.; Schmeits, M. Comparison of Statistical Post-Processing Methods for Probabilistic NWP Forecasts of Solar Radiation. *Sol. Energy* **2019**, *191*, 138–150. [[CrossRef](#)]
22. Yeom, J.M.; Deo, R.C.; Adamwoski, J.F.; Chae, T.; Kim, D.S.; Han, K.S.; Kim, D.Y. Exploring Solar and Wind Energy Resources in North Korea with COMS MI Geostationary Satellite Data Coupled with Numerical Weather Prediction Reanalysis Variables. *Renew. Sustain. Energy Rev.* **2020**, *119*, 109570. [[CrossRef](#)]
23. Das, U.K.; Tey, K.S.; Seyedmahmoudian, M.; Mekhilef, S.; Idris, M.Y.I.; Van Deventer, W.; Stojevski, A. Forecasting of Photovoltaic Power Generation and Model Optimization: A review. *Renew. Sustain. Energy Rev.* **2018**, *81*, 912–928. [[CrossRef](#)]
24. Kushwaha, V.; Pindoriya, N.M. A SARIMA-RVFL Hybrid Model Assisted by Wavelet Decomposition for Very Short-Term Solar PV Power Generation Forecast. *Renew. Energy* **2019**, *140*, 124–139. [[CrossRef](#)]
25. Lamsal, D.; Sreeram, V.; Mishra, Y.; Kumar, D. Kalman Filter Approach for Dispatching and Attenuating the Power Fluctuation of Wind and Photovoltaic Power Generating Systems. *IET Gener. Transm. Distrib.* **2018**, *12*, 1501–1508. [[CrossRef](#)]
26. Miao, S.; Ning, G.; Gu, Y.; Yan, J.; Ma, B. Markov Chain Model for Solar Farm Generation and its Application to Generation Performance Evaluation. *J. Clean. Prod.* **2018**, *186*, 905–917. [[CrossRef](#)]
27. Wang, F.; Mi, Z.; Su, S.; Zhao, H. Short-Term Solar Irradiance Forecasting Model Based on Artificial Neural Network Using Statistical Feature Parameters. *Energies* **2012**, *5*, 1355–1370. [[CrossRef](#)]
28. Liu, J.; Fang, W.; Zhang, X.; Yang, C. An Improved Photovoltaic Power Forecasting Model with the Assistance of Aerosol Index Data. *IEEE Trans. Sustain. Energy* **2015**, *6*, 434–442. [[CrossRef](#)]
29. Abuella, M.; Chowdhury, B. Random Forest Ensemble of Support Vector Regression Models for Solar Power Forecasting. In Proceedings of the IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), Washington, DC, USA, 23–26 April 2017; pp. 1–5.
30. Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.L.; Paoli, C.; Motte, F.; Fouilloy, A. Machine Learning Methods for Solar Radiation Forecasting: A Review. *Renew. Energy* **2017**, *105*, 569–582. [[CrossRef](#)]
31. Benali, L.; Notton, G.; Fouilloy, A.; Voyant, C.; Dizene, R. Solar Radiation Forecasting Using Artificial Neural Network and Random Forest Methods: Application to Normal Beam, Horizontal Diffuse and Global Components. *Renew. Energy* **2019**, *132*, 871–884. [[CrossRef](#)]
32. Ghimire, S.; Deo, R.C.; Downs, N.J.; Raj, N. Global Solar Radiation Prediction by ANN Integrated with European Centre for Medium Range Weather Forecast Fields in Solar Rich Cities of Queensland Australia. *J. Clean. Prod.* **2019**, *216*, 288–310. [[CrossRef](#)]
33. Roumpakias, E.; Stamatelos, T. Prediction of a Grid-Connected Photovoltaic Park’s Output with Artificial Neural Networks Trained by Actual Performance Data. *Appl. Sci.* **2022**, *12*, 6458. [[CrossRef](#)]
34. Ahmed, R.; Sreeram, V.; Mishra, Y.; Arif, M.D. A Review and Evaluation of the State-of-the-Art in PV Solar Power Forecasting: Techniques and Optimization. *Renew. Sustain. Energy Rev.* **2020**, *124*, 109792. [[CrossRef](#)]
35. Noh, S.H. Analysis of Gradient Vanishing of RNNs and Performance Comparison. *Information* **2021**, *12*, 442. [[CrossRef](#)]
36. Beigi, M.; Beigi Harchegani, H.; Torki, M.; Kaveh, M.; Szymanek, M.; Khalife, E.; Dziwulski, J. Forecasting of Power Output of a PVPS Based on Meteorological Data Using RNN Approaches. *Sustainability* **2022**, *14*, 3104. [[CrossRef](#)]
37. Son, N. Comparison of the Deep Learning Performance for Short-Term Power Load Forecasting. *Sustainability* **2021**, *13*, 12493. [[CrossRef](#)]
38. Available online: <https://www.kma.go.kr/kma/biz/forecast05.jsp> (accessed on 26 August 2022).