Assessment of Different Deep Learning Methods of Power Generation Forecasting for Solar PV System

Wen-Chi Kuo 1, Chiun-Hsun Chen 2, Shih-Hong Hua 1 and Chi-Chuan Wang 1,*

1 Department of Mechanical Engineering, National Yang Ming Chiao Tung University, Hsinchu 300, Taiwan; wenchi.en08@nycu.edu.tw (W.-C.K.); hs860731abc.met08g@nctu.edu.tw (S.-H.H.)
2 Department of Aerospace and Systems Engineering, Feng Chia University, Taichung 407, Taiwan; chiunhchen@fcu.edu.tw
* Correspondence: ccwang@nycu.edu.tw

Abstract: An increase in renewable energy injected into the power system will directly cause a fluctuation in the overall voltage and frequency of the power system. Thus, renewable energy prediction accuracy becomes vital to maintaining good power dispatch efficiency and power grid operation security. This article compares the one-day-ahead PV power forecasting results of three models paired with three groups of weather data. Since the number, loss, and matching problem of weather data will all influence the training results of the model, a pre-processing data framework is proposed to solve the problem in this study. The models used are a deep learning algorithm-based artificial neural network (ANN), long short-term memory (LSTM), and gated recurrent unit (GRU). The weather data groups are Central Weather Bureau (CWB), local weather station (LWS), and hybrid data (the combination of CWB and LWS data). Compared to the other two groups, hybrid data showed a 5–8% improvement in measurements. In addition, when it comes to different weather conditions, the advantages of the LSTM model were highlighted. After further analysis, the LSTM model combined with hybrid data showed the most accurate measurements, which was proved through forecasting results for one month. Finally, the results indicate that when the amount of data is limited, using hybrid data and the five weather features is helpful for training the model. Accordingly, the proposed model shows better one-day-ahead PV forecasting.

Keywords: deep learning (DL); forecasting; neural network; renewable energy; solar power generation

1. Introduction

In recent decades, renewable energy development and carbon reduction have been energy policy targets in various countries in response to global warming. Due to the intrinsic intermittent and uncertain nature of renewable energy, previous experience shows that a high penetration percentage of renewable energy will affect the system’s voltage, frequency, and load demand fluctuations. In the state of California, the solar penetration rate of residential roofs has increased rapidly, causing the load to show a duck curve, which conflicts with conventional generation efficiency and the frequency of daily start-up/shut-down operations [1,2]. In this context, developing an accurate forecast of long-term and short-term renewable generation ability is vital for efficient power dispatch and operations. This paper compares different models and weather data to predict one-day-ahead hourly PV power generation.

Accurate prediction of PV power generation is challenging due to complex interactions between many environmental conditions and uncontrollable factors. Therefore, studying weather data is crucial to predicting the impact of solar power generation [3–5]. Sangrody et al. [6] selected the most influential variables, including sky coverage, relative humidity, and temperature, for forecasting model training. Sensitivity tests showed that relative humidity plays a dominant role in forecasting. Sangrody et al. [7] concluded that
the most common weather variables, including sky coverage, dew point, relative humidity, and temperature, can be used to predict solar PV generation. Rehman and Mohandes [8] estimated global solar radiation using relative humidity and daily mean temperature and showed that their model outperformed other cases ((1) daily maximum air temperature and (2) daily mean air temperature), with an absolute mean percentage error of 4.49%. Alluhaidah et al. [9] compared the impact of seven meteorological variables on the accuracy of predictions, including air temperature, relative humidity, pressure, cloud coverage, wind speed, and wind direction. The best results strongly depended on cloud coverage and relative humidity. The least influential variable was wind direction, followed by wind speed. As described by Ghanbarzadeh et al. [10], sunshine hours, daily mean air temperature, and relative humidity are key variables, and the proper processing of these variables on a specific timescale outperformed the other timescale, with an absolute mean percentage error of 8.84%. Giorgi et al. [11] determined the influence of ambient temperature on the forecast of PV power and solar irradiance and yielded the best predictive ability. Gigoni et al. [12] noted that temperature was not as sensitive as other weather parameters among several methods for solar energy prediction. Several studies indicated that weather parameters had a potential impact on PV power prediction, including the relative humidity, temperature, diffuse horizontal radiation, and global horizontal radiation, which significantly affected PV power output. Moreover, daily precipitation seems to be a less critical factor in PV power prediction [13–15].

Additionally, previous studies pointed out the importance of weather data classification, which significantly helped in improving the PV prediction effect. Liu and Zhang [16] proposed a novel weather classification process and a novel way to embed a physical model of PV units into the K-Nearest Neighbor (KNN) algorithm model. The results showed that the method could reduce the uncertainty of PV output forecasting. Hossain and Mahmood [17] used the K-means algorithm to classify different levels of irradiance, where the interval was based on the hour of the day and the season. Cheng et al. [18] proposed a new method, which was a factor-based graph modeling method using numerical weather prediction (NWP) data for PV power prediction. This method was used to explore and analyze a multi-graph model that can adapt to different weather conditions for PV forecasting. Yu et al. [19] pointed out that inaccurate predictions usually occur on cloudy days. Therefore, to improve prediction accuracy in cloudy weather, for the classification of weather types, cloudy days were divided into cloudy and mixed. Silvia et al. [20] proposed a new ensemble method based on the probabilistic distribution of trials using the probabilistic ensemble method (PEM) for PV power forecasting on cloudy days. The result showed that the nRMSE metric reached 4.79% in 2017 in the totally cloudy days class. This paper proposes a data classification method and finds the critical weather factor that affects PV power generation forecasting.

Furthermore, with advances in the calculation speed of computers, machine learning has been widely used in the past decade, especially in prediction-related literature using large amounts of complex historical information [21,22]. Yona et al. [23] proposed power output forecasting of a PV system based on 24 h ahead insolation forecasting by using reported weather data, fuzzy theory, and a neural network. Yang et al. [24] proposed a hybrid method for hourly forecasting of 1-day-ahead PV power. The historical data of PV output power were classified by the self-organizing map (SOM) and learning vector quantization (LVQ) networks. The results showed that the proposed approach achieved better predictive accuracy than the traditional ANN and simple Support Vector Regression (SVR) methods. Sorkun et al. [25] assessed deep learning with the time series method and claimed that it is suitable and competitive for solar forecasting. The results showed that LSTM and GRU models are suitable and competitive for solar radiation time series forecasting. Hanafi et al. [26] used a backpropagation neural network (BPNN) and extreme learning machine (ELM) for one-hour-ahead solar power forecasting. The results showed that the accuracy and computational time of extreme learning machine were better than those of backpropagation neural networks. Liu et al. [27] used weather conditions that
were categorized as sunny, cloudy, rainy, low light, sunny-cloudy, and rainy-sunny as the training dataset to carry out one-day-ahead solar power forecasting. The results showed that their proposed simplified LSTM model outperformed the multilayer perceptron (MLP) model. U. K. Das et al. [28] provided a comprehensive and systematic review of the direct forecasting of photovoltaic power generation. They also provided a critical analysis of recent work, including statistics and machine learning models based on historical data. Meftah et al. [29] compared the performance of LSTM and MLP for PV power forecasting in winter and summer. The results indicated that the effect using LSTM was better for short-term forecasting. A summary of different forecasting algorithms, data inputs, and horizons in the literature is shown in Table 1.

Table 1. Summary of different forecasting algorithms, data inputs, and horizons.

<table>
<thead>
<tr>
<th>Ref.</th>
<th>Year</th>
<th>Method</th>
<th>Inputs</th>
<th>Horizon</th>
<th>Best Results of Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>[10]</td>
<td>2009</td>
<td>ANN</td>
<td>Air temperature, relative humidity, and sunshine</td>
<td>Hour</td>
<td>MAPE (%): 8.84</td>
</tr>
<tr>
<td>[16]</td>
<td>2016</td>
<td>KNN</td>
<td>DNI, DHI, ambient temperature</td>
<td>Hour</td>
<td>MAPE (%): 18.25 MRAE 4 (%): 2.01</td>
</tr>
<tr>
<td>[17]</td>
<td>2020</td>
<td>LSTM</td>
<td>Solar irradiance, temperature, relative humidity, and wind speed</td>
<td>24 h</td>
<td>MAPE (%): 22.31 RMSE: 0.71 MAE: 0.36</td>
</tr>
<tr>
<td>[23]</td>
<td>2014</td>
<td>SVR</td>
<td>Temperature, probability of precipitation, and solar irradiance</td>
<td>One-day-ahead hourly</td>
<td>MRE 5 (%): 3.295</td>
</tr>
<tr>
<td>[26]</td>
<td>2019</td>
<td>BPNN ELM 1</td>
<td>Sunshine hours, global radiation, and UV index</td>
<td>One-hour-ahead</td>
<td>nRMSE 6 (%): 7.75</td>
</tr>
<tr>
<td>[30]</td>
<td>2018</td>
<td>DEPSO 2</td>
<td>Tipping Bucket Rain Gauge, wind speed, wind direction, air temperature, relative humidity, and solar radiation</td>
<td>Hour</td>
<td>RMSE (%): 4.4% MAE: 0.03 MBE: −1.63 VAR 7: 0.01 WME 8: 0.16 MRE (%): 3.1</td>
</tr>
<tr>
<td>[31]</td>
<td>2019</td>
<td>Hybrid GA-SVM 3</td>
<td>Tipping Bucket Rain Gauge, wind speed, wind direction, air temperature, relative humidity, and solar radiation</td>
<td>Hour</td>
<td>MAPE (%): 1.7052 RMSE: 11.226</td>
</tr>
<tr>
<td>[32]</td>
<td>2021</td>
<td>Two-stage attention-based encoder–decoder over LSTM</td>
<td>Forty-one features, including solar radiation, temperature, humidity, snowfall, albedo, etc.</td>
<td>One-day-ahead</td>
<td>RMSE: 0.0638 MAE: 0.0324</td>
</tr>
</tbody>
</table>

1 ELM: Elaboration Likelihood Model; 2 DEPSO: Discrete Evolutionary Particle Swarm Optimization; 3 GA-SVM: Genetic Algorithm and Support Vector Machine; 4 MRAE: Mean Absolute Relative Error; 5 MRE: Mean Relative Error; 6 nRMSE: normalized Root Mean Square Error; 7 VAR: variance of the prediction errors; 8 WME: Weighted Mean Errors.

The actual PV power generation site used in this study is in Taoyuan City, Taiwan. The number of equivalent sunshine hours is about 3 h. The weather in this city in the experimental year was relatively unstable. The objective of this study was to compare three kinds of deep learning models, namely, the artificial neural network (ANN), long short-term memory (LSTM), and gated recurrent unit (GRU), for one-day-ahead PV power generation forecasting. Furthermore, a data pre-processing framework is proposed to improve predictive accuracy and prevent the failure of one-day-ahead PV power forecasts. The structure of this article is as follows: Section 1 presents a brief review of weather factors and modeling research for PV power forecasting. The details of the various weather data resources used in this study are introduced in Section 2. Then, Section 3 proposes the framework for data pre-processing. Section 4 introduces the application of deep learning
modeling and evaluation methods. Section 5 presents the numerical PV power forecasting results, followed by the final section for the conclusion.

2. Data Description

This section introduces the sources of the weather and solar photovoltaic parameters and clarifies the means of obtaining various data. The architecture of data collection is shown in Figure 1. The data used in this study include weather data and PV power generation data. The measured weather forecast data and historical data can be easily accessed from the Central Weather Bureau (CWB) website. These data were used for model training and prediction for one-day-ahead PV power forecasting. In addition, the PV power data used in this study were obtained from the SQL server of the energy management system (EMS), which is used as the target of deep learning.

Figure 1. The architecture of data collection (including Central Weather Bureau and local micro weather station) and model forecasting structure.

2.1. Photovoltaic Generation Data

The actual PV module field is depicted in Figure 2. The PV module was installed on the rooftop of a fast-food restaurant in Taoyuan City, Taiwan, and the capacity is 10 kW. The number of effective sunshine hours is about 3 h in this area. The collection period of the photovoltaic generation data used in this study was from 1 August 2020 to 3 July 2021. The PV power generation data are sampled every minute in this field, and hourly PV power data were obtained as the average data for 60 min.

Figure 2. Rooftop solar PV with 30 solar panels and 10 kW capacity in Yangmei District, Taiwan.
2.2. Meteorological Data from the CWB

The experimental database of historical meteorological data used in this study was provided by the Xinwu Weather Station of CWB (about 26 km away from the site) from August 2020 to July 2021. The observed weather database has 55 different weather parameters and can be easily accessed from the CWB website. On the other hand, the weather forecast dataset has 16 different weather features, providing the next three days of weather forecast data, and the data interval is hourly. Details can be found on the CWB website [33,34].

2.3. AccuWeather Data

From the historical PV data, it was found that on rainy days, there was a drastic change or almost no PV power generation, as shown in Figure 3. The figure illustrates hourly power generation from 2 to 4 March 2021. The photo shows the weather condition captured by the sky camera at that time for reference. As a result, rainfall data are one of the important parameters for photovoltaic power forecasting. Since there were no rainfall data in the forecast data of the CWB, the rainfall data of Yangmei District in this article were obtained from the AccuWeather Data website, and the data interval is once per hour [35].

![Figure 3. The historical photovoltaic forecast data graph and the weather condition photo captured by the sky camera from 2–4 March 2021.](image)

2.4. Local Weather Station (LWS) and Pyrheliometer

The LWS data are from a micro weather station installed in the field of Yangmei District, Taiwan, which is a 6-in-1 weather station. It was installed on the platform less than 100 m from the PV panel. That equipment provides data that include temperature, relative humidity, average wind speed, wind direction, rainfall, and atmospheric pressure. The datasheet of the weather station is presented in Table 2. A pyrheliometer was installed in the field to measure and record the current solar radiation, and the datasheet is depicted in Table 3. The data are updated every minute. The hourly LWS data were obtained as the average data for 60 min.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature [°C]</td>
<td>0–60</td>
</tr>
<tr>
<td>Relatively Humidity [%]</td>
<td>0–100</td>
</tr>
<tr>
<td>Average Wind Speed [m/s]</td>
<td>0–60</td>
</tr>
<tr>
<td>Wind Direction [Degree]</td>
<td>0–360</td>
</tr>
<tr>
<td>Rainfall [mm/h]</td>
<td>0–200</td>
</tr>
<tr>
<td>Pressure [hPa]</td>
<td>600–1100</td>
</tr>
</tbody>
</table>
Table 3. Pyrheliometer specification sheet.

<table>
<thead>
<tr>
<th>Specification</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measurement Range [W/m²]</td>
<td>0–2000</td>
</tr>
<tr>
<td>Spectral Range [nm]</td>
<td>305–2800</td>
</tr>
</tbody>
</table>

3. Data Pre-Processing

The literature review showed that the acquisition of meteorological data has a critical influence on solar photovoltaic forecasting. The accuracy of the forecasting results will vary considerably and is subject to the quality of the measured parameters. These historical and forecast weather data may be lost during collection or unavailable because the website is inaccessible, and this will affect the forecasting results of photovoltaic power generation for the next day. Therefore, data pre-processing is the most important and critical factor for a one-day-ahead PV power forecasting study. The data pre-processing methods used in this study can be divided into three steps, which are data classification, data filtration, and missing data processing. The structure of data pre-processing used in this study is presented in Figure 4.

Data pre-processing

![Data pre-processing structure](image)

3.1. Data Classification

The main purpose of data classification is to classify weather data from different sources before performing forecast training. To study the influence of data from different sources on one-day-ahead solar photovoltaic forecasting, the data used in this study were divided into three sources: The first group is the historical and forecast data provided by the open platform of the CWB. The second group is the historical weather data from the on-site LWS equipment to carry out one-day-ahead forecasting. The third group is the combination of CWB and LWS data, which means using the same weather features and training data from the CWB and LWS, named hybrid weather data in this article.

3.2. Data Filtering

The data filtration process consists of the pre-processing of solar photovoltaic power data, historical weather data, and historical weather forecast data. The details are described as follows:

1. The filtration method of solar photovoltaic power data: When collecting solar photovoltaic power data, very little electricity consumption is needed to maintain the standby status of the inverter when no power generation occurs. In addition, solar photovoltaic power generation is too low in the early morning. These data not only affect the forecast calculation but are useless in the actual power generation forecast. Therefore, the pre-processing method of solar photovoltaic data entails replacing...
values less than 0.5 kW and the standby value of $-0.002$ W by 0. In the zoomed-in area of the collected data illustrated in Figure 5, the values are much lower than $-0.002$ W, so these values are replaced by 0.

2. The filtration method of weather feature data: There are 55 different weather feature parameters in CWB’s historical data but only 16 features from CWB’s weekly forecast data. Thus, 16-parameter weather forecast data were compared to 55-historical-parameter data to find exact matches. The eigenvalues used in the model training process must be the same as those used in the predictive model so that the calculated weights of the deep-learning model are consistent. After cross-checking with the complete data used in this research, there were only 5 parameters that could be used, as shown in Table 4. Among these 5 parameters, the value subject to rainfall has a certain influence on the forecast for the next day. It was obtained from the AccuWeather website and used as one of the parameters.

Table 4. Final features selected from the CWB.

<table>
<thead>
<tr>
<th>Feature</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>°C</td>
</tr>
<tr>
<td>Relative Humidity</td>
<td>%</td>
</tr>
<tr>
<td>Rainfall</td>
<td>mm</td>
</tr>
<tr>
<td>Average Wind Speed</td>
<td>m/s</td>
</tr>
<tr>
<td>UV Index</td>
<td>-</td>
</tr>
</tbody>
</table>

3.3. Missing Data Processing

In the process of data transmission and storage, data may be lost due to network disconnection, database maintenance, or some unexpected malfunction. Different methods were used to pre-process lost historical and forecast data, which are described as follows:

1. The pre-processing method of the historical data, including the data from the CWB and local weather station, involved directly deleting abnormal and missing data. The data for such days were erased when the historical data were $-9999$, null, or intermittently lost. The number of processed data points is shown in Figure 6. The hollow bars show the amount of data that can be used for one-day-ahead solar photovoltaic forecasting training after processing by the above method. There are 170 days of hybrid data, 308 days of local weather station data, and 181 days of CWB data. The original data are from 1 August 2020 to 20 June 2021, a total of 324 days.
2. However, the pre-processing method used for lost weather forecast data differs from that used for historical data. There were two methods used in this study. The first was the interpolation of data that were missing for one or two hours of the day. If more than one-fourth of the forecast data for a day were lost, the following method was used. The historical data of the next day’s forecast were searched on the CWB’s database, and similar weather data were directly used as the forecast weather factor for the next day. For example, if the forecast for the next day was mostly clear from 6:00 to 18:00, the forecast data for the same weather were searched in the historical database and used as the next day’s forecast data. The reason for adopting this method is that the unavailability of weather information for the next day will cause the predictive system to crash. As a result, it cannot be applied to actual cases, and it will be meaningless to introduce the predictive module into the energy management system.

![Figure 6. Missing data processing results (hollow bars: the amount of data after processing; the orange, green and blue bars are the amount of original data).](image)

4. Methods and Evaluation

We note that PV power generation and weather data are related to time series. Thus, the deep learning algorithms chosen for one-day-ahead PV power forecasting were LSTM and GRU neural network models, and the traditional ANN model was added to compare the accuracy of the three models in this study. Hyperparameter tuning of the deep learning network will directly affect the model’s performance and then influence the prediction results. This section introduces the method of hyperparameter tuning for each model. The accuracy evaluation methods used in this study are the most widely used metrics, namely, MAPE, RMSE, and MAE. They are described below.

4.1. Artificial Neural Network (ANN)

An artificial neural network is one of the main tools used in deep learning. It consists of an input layer, hidden layer, and output layer. The input layer comprises artificial input neurons and brings the initial data into the system for further processing by subsequent layers of artificial neurons. The hidden layer is the layer between input layers and output layers, where artificial neurons receive a set of weighted inputs and produce an output through an activation function. The output layer is the last layer of neurons and produces the given outputs for the program. For the neurons in each layer, the activation function measures the importance of the neuron output to improve the model’s effectiveness. The advantage of ANN is it can deal with complicated nonlinear problems and has excellent adaptive ability. Therefore, ANN is one of the most popular methods to realize predictions [36]. However, the performance will be relatively poor when there are less data.
4.2. Long Short-Term Memory (LSTM)

RNN (Recurrent Neural Network) is a neural network that can cause the vanishing and exploding gradient problem, so it cannot handle information for a long period of time. Therefore, long short-term memory was introduced by Hochreiter and Schmidhuber [37], who proposed that LSTM can solve the vanishing and exploding gradient problem during the training process. In 1999, Gers et al. [38] introduced the forget gate into the LSTM structure, which decides the ratio of information that needs to be kept in the memory cell. LSTM performs better when there are more data.

4.3. Gated Recurrent Unit (GRU)

The gated recurrent unit was introduced by Cho et al. [39]. GRU’s design is similar to that of LSTM, but without an output gate. It tries to solve the vanishing gradient problem by using an update gate and a reset gate. When dealing with small datasets, GRU is a better choice than LSTM because it has only two gates with a fast calculating speed.

4.4. Hyperparameter Adjustment Process

Hyperparameter tuning is a way to optimize the deep learning model, including the hidden unit, batch size, and epochs. Note that the batch size indicates the number of training samples. By increasing the number of samples, the training speed can be accelerated. If the batch size is too large, the memory will be insufficient. By contrast, too small a batch size may lead to underfitting. The hidden unit indicates the number of parameters. The higher the number of hidden layer units, the easier overfitting becomes. Epochs represent the number of training iterations for all of the batches. The training curve will descend when the epoch increases to optimize the loss value. Therefore, the batch size was fixed at the beginning, and three to five layers were tested in every step with the same epochs in the hyperparameter adjustment process.

4.5. Evaluation Indices

After obtaining every hourly forecast value of one-day-ahead PV power with the proposed model, it was necessary to validate the applicability of the model through certain indices. In machine learning, the most common indices used to evaluate the performance of models are MAE, RMSE, and MAPE [40].

1. Mean Absolute Error (MAE). MAE can be used to measure the error between predicted values and actual values. It depends on the scale of continuous variables. The lower the value, the higher the accuracy of the predicted model. The equation is given below:

\[
MAE = \frac{1}{N} \sum_{i=1}^{N} |\hat{X}_i - X_i|
\]  

where \(\hat{X}_i\) and \(X_i\) represent the \(i^{th}\) forecasted and actual values, respectively, and \(N\) is the size of the test dataset (\(N = 168\) for one week in this study).

2. Root Mean Squared Error (RMSE). RMSE can be used to measure the deviation between predicted values and actual values. The difference between RMSE and MAE is that RMSE is sensitive to outliers. That means that RMSE is easily influenced by large deviations. Hence, a smaller error indicates better performance. The equation is given below:

\[
RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\hat{X}_i - X_i)^2}
\]  

where \(\hat{X}_i\) and \(X_i\) represent the \(i^{th}\) forecasted and actual values, respectively, and \(N\) is the size of the test dataset (\(N = 168\) for one week in this study).
3. Mean Absolute Percentage Error (MAPE). MAPE measures the accuracy as a percentage. It can be used to judge the quality of the predicted result. The definition is:

$$\text{MAPE} = \left[ \frac{1}{N} \sum_{i=1}^{N} \frac{|X_i - \hat{X}_i|}{X_i} \right] \times 100\% \quad (3)$$

where $\hat{X}_i$ and $X_i$ represent the $i$th forecasted and actual values, respectively, and $N$ is the size of the test dataset ($N = 168$ for one week in this study).

5. Numerical Results

5.1. Results of Hyperparameter Adjustment

For the actual forecast from 4 February to 21 May 2021, five features (temperature, relative humidity, rainfall, average wind speed, and UV index) were employed with Adam optimizer to tune the hyperparameters. The test results of hyperparameters are shown in Tables 5–7. The tables show that the accuracy of hybrid weather data for the one-day-ahead forecast is higher than that of CWB and LWS data when applied to different models.

<table>
<thead>
<tr>
<th>Model</th>
<th>Input Time</th>
<th>Layers</th>
<th>Epochs</th>
<th>Learning Rate</th>
<th>Batch Size</th>
<th>MAE</th>
<th>MAPE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSTM</td>
<td>2 days</td>
<td>3</td>
<td>2000</td>
<td>$1 \times 10^{-3}$</td>
<td>8</td>
<td>0.5074</td>
<td>17%</td>
<td>0.9979</td>
</tr>
<tr>
<td>GRU</td>
<td>2 days</td>
<td>3</td>
<td>2000</td>
<td>$1 \times 10^{-3}$</td>
<td>8</td>
<td>0.4898</td>
<td>18%</td>
<td>0.8397</td>
</tr>
<tr>
<td>ANN</td>
<td>3 days</td>
<td>5</td>
<td>2000</td>
<td>$1 \times 10^{-3}$</td>
<td>8</td>
<td>0.8999</td>
<td>34%</td>
<td>1.6067</td>
</tr>
</tbody>
</table>

Table 6. The hyperparameter tuning results of LWS historical data.

<table>
<thead>
<tr>
<th>Model</th>
<th>Input Time</th>
<th>Layers</th>
<th>Epochs</th>
<th>Learning Rate</th>
<th>Batch Size</th>
<th>MAE</th>
<th>MAPE</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSTM</td>
<td>2 days</td>
<td>3</td>
<td>2000</td>
<td>$1 \times 10^{-3}$</td>
<td>8</td>
<td>0.3992</td>
<td>12%</td>
<td>0.7105</td>
</tr>
<tr>
<td>GRU</td>
<td>2 days</td>
<td>3</td>
<td>2000</td>
<td>$1 \times 10^{-3}$</td>
<td>8</td>
<td>0.4083</td>
<td>12%</td>
<td>0.7059</td>
</tr>
<tr>
<td>ANN</td>
<td>1 day</td>
<td>3</td>
<td>2000</td>
<td>$1 \times 10^{-3}$</td>
<td>8</td>
<td>0.4377</td>
<td>12%</td>
<td>1.3326</td>
</tr>
</tbody>
</table>

Table 7. The hyperparameter tuning results of hybrid data.

5.2. Forecast Performance with Different Weather Data Groups

After deep learning model training, these models were applied to execute solar photovoltaic forecasting from 27 June to 3 July 2021. This section describes the use of different deep learning methods (ANN, LSTM, and GRU) with different weather groups for solar photovoltaic forecasting. The forecasting results of the ANN model are presented in Figure 7. The solid blue line is the actual PV power generation, the green dashed line is the predicted result using LWS weather data, the red dashed line is the predicted result using hybrid weather data, and the yellow dashed line is the predicted result using CWB weather data. The lines used in the other graphs have the same meanings. It is apparent that when only CWB data are used, the prediction results of the ANN model severely deviate from the actual power generation during the peak period on 27 June and 29 June 2021. The results indicate that the ANN model does not provide good predictions, even under sunny conditions. On the other hand, when only historical LWS data are used, the results during
peak power generation periods deviate quite substantially from the actual values due to poor weather (e.g., 28 June to 29 June). In contrast, the overall prediction curve using hybrid data is closer to the actual power generation trend, although the results are still not good enough under poor weather conditions.

![Figure 7. The PV power forecasting results of the ANN model.](image)

The forecasting results of the LSTM model are shown in Figure 8. When CWB data are used for prediction, the accuracy of LSTM is better than that of ANN. This means that the learning mode with time series helps to improve the predictive accuracy. However, from the prediction results of LWS data, it is found that although the results are better than those of the ANN model (the trends on 28 and 29 June 2021 show significant improvement), there is still a big difference in the prediction results on 1 July 2021. This means that the predictive ability is still limited when only historical data are employed. Conversely, the forecast trend of the hybrid model is similar to the actual PV power generation results.

![Figure 8. The PV power forecasting results of the LSTM model.](image)

The forecasting results of the GRU model are shown in Figure 9. It is also evident from the results that when using CWB data, the GRU model can effectively improve the accuracy in comparison with ANN, especially on 27 June 2021. However, the results show that if only the LWS historical weather data are used for forecasting, the prediction results are still inaccurate with severe deviations in tendency. This consequence is the same as those of ANN and LSTM, e.g., on 28 June and 1 July 2021. In contrast, when the hybrid data are used, the one-day-ahead solar photovoltaic forecasting trend is better than when only using CWB or LWS data. The prediction results on 1 July 2021 show that hybrid data significantly help improve the effect of the one-day-ahead solar photovoltaic forecast.
Finally, the comprehensive comparison results from the trend of the PV power forecast curve or numerical statistical analysis show that using hybrid weather data can effectively improve the accuracy of solar photovoltaic forecasts. On the other hand, from the average prediction results of the three different groups of weather data, the effect of the LSTM model is the best, indicating that this model is more suitable as a solar photovoltaic prediction model in this study, with a MAPE of 20.0%, MSE of 1.158, and RMSE of 1.004. Based on the suggestion by Lewis [41], the accuracy of the predictive ability based on MAPE is classified as highly accurate (less than 10%), good (11% to 20%), reasonable (21% to 50%), and inaccurate (more than 50%). It is shown that the data pre-processing method proposed in this study combined with the LSTM deep learning forecasting model can achieve good prediction results.

From the analysis results mentioned above, the use of hybrid data can significantly improve the one-day-ahead solar photovoltaic forecast effect. Therefore, the results of one-day-ahead solar photovoltaic forecasting with the hybrid data group for one week were further analyzed in this study, as shown in Figure 10. The results of the one-week forecast statistical numerical analysis are presented in Table 9. In Figure 10, there is an obvious prediction difference among the three models on 2 and 3 July 2021. The forecast results of ANN are worse than those of LSTM and GRU. This means that when there is a

<table>
<thead>
<tr>
<th>Model</th>
<th>CWB Weather Data</th>
<th>Local Weather Data</th>
<th>Hybrid Weather Data</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAPE</td>
<td>MAE</td>
<td>RMSE</td>
<td>MAPE</td>
</tr>
<tr>
<td>ANN</td>
<td>29%</td>
<td>2.020</td>
<td>1.400</td>
<td>23%</td>
</tr>
<tr>
<td>LSTM</td>
<td>21%</td>
<td>0.931</td>
<td>0.940</td>
<td>23%</td>
</tr>
<tr>
<td>GRU</td>
<td>19%</td>
<td>1.083</td>
<td>1.003</td>
<td>23%</td>
</tr>
</tbody>
</table>

Table 8. Average one-day-ahead PV power forecasting results of each weather data group for one week.
big change in weather conditions, the difference between the forecast and actual power generation trends will become more pronounced.

![Real PV Curve](image.png)

**Figure 10.** Comprehensive comparison results of hybrid metadata.

**Table 9.** Daily one-day-ahead PV power forecasting results using hybrid weather parameters.

<table>
<thead>
<tr>
<th>Day</th>
<th>ANN Model</th>
<th>LSTM Model</th>
<th>GRU Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MAPE</td>
<td>MAE</td>
<td>RMSE</td>
</tr>
<tr>
<td>27 June 2021</td>
<td>15%</td>
<td>1.090</td>
<td>1.044</td>
</tr>
<tr>
<td>28 June 2021</td>
<td>35%</td>
<td>1.863</td>
<td>1.365</td>
</tr>
<tr>
<td>29 June 2021</td>
<td>31%</td>
<td>1.506</td>
<td>1.227</td>
</tr>
<tr>
<td>30 June 2021</td>
<td>20%</td>
<td>0.968</td>
<td>0.984</td>
</tr>
<tr>
<td>1 July 2021</td>
<td>21%</td>
<td>1.240</td>
<td>1.113</td>
</tr>
<tr>
<td>2 July 2021</td>
<td>18%</td>
<td>2.364</td>
<td>1.538</td>
</tr>
<tr>
<td>3 July 2021</td>
<td>36%</td>
<td>3.720</td>
<td>1.929</td>
</tr>
<tr>
<td>Average</td>
<td>24%</td>
<td>1.33</td>
<td>1.15</td>
</tr>
</tbody>
</table>

When comparing the daily forecast numerical analysis results of the LSTM and GRU models for one week, the results of the LSTM model are better than those of the GRU model, and the average forecast results are relatively stable. The average results of LSTM for a week have a MAPE of 16%, MSE of 0.71, and RMSE of 0.83. On the other hand, the GRU model is better than LSTM on sunny days. The MAPE, MAE, and RMSE of GRU are 8%, 0.263, and 0.513, and those of LSTM are 11%, 0.531, and 0.792, respectively, on 27 June 2021. In addition, the forecast results of GRU on 2 July are also better than those of LSTM. However, the advantages of the LSTM model are highlighted when the weather pattern is poor. From the forecasting results on 1 July, the MAPE of LSTM is about 8% higher than that of GRU, and MAE and RMSE are about 0.3–0.4 higher, respectively. Therefore, when selecting the prediction results for the next day, the weather forecast can be used as a reference. If the weather forecast for the next day is sunny, the results of the GRU model are more accurate. If it is cloudy, the results of LSTM are a better choice.

According to the aforementioned results, it is found that forecasting via the LSTM model combined with the hybrid weather data yields the best prediction results. Therefore, the LSTM model was used to predict the solar photovoltaic power for June 2021, and the results are shown in Figure 11. The results show that MAPE, MAE, and RMSE are 16.984%, 1.764, and 1.283, respectively. It is shown that the combination of CWB data and LWS data together with the deep learning model is helpful for improving the accuracy of the one-day-ahead solar photovoltaic forecast. Furthermore, a comparison with other one-day-ahead solar photovoltaic forecast methods from the literature is shown in Table 10. It can be found that the application of complex deep learning algorithms combined with
the weather data and data pre-processing method proposed in this study has a favorable effect on one-day-ahead solar photovoltaic forecasting.

![Figure 11. Forecasting results of the LSTM model for the month of June.](image)

<table>
<thead>
<tr>
<th>Paper</th>
<th>Method Used</th>
<th>Inputs</th>
<th>Horizon</th>
<th>Best Results for Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>[17]</td>
<td>LSTM NN</td>
<td>Solar irradiance, temperature, relative humidity, and wind speed</td>
<td>24 h</td>
<td>MAPE (%): 22.31&lt;br&gt;RMSE: 0.71&lt;br&gt;MAE: 0.36</td>
</tr>
<tr>
<td>[23]</td>
<td>Weather-based hybrid method: SOM, LVQ, and SVR</td>
<td>Temperature, probability of precipitation, and solar irradiance</td>
<td>One-day-ahead hourly</td>
<td>RMSE: 1.6811</td>
</tr>
<tr>
<td>This manuscript</td>
<td>LSTM</td>
<td>Temperature, relative humidity, rainfall, average wind speed, and UV index</td>
<td>One-day-ahead 24 h</td>
<td>MAPE (%): 16.984&lt;br&gt;RMSE: 1.764&lt;br&gt;MAE: 1.283</td>
</tr>
</tbody>
</table>

### 6. Conclusions

The acquisition of weather feature data greatly influences the one-day-ahead solar photovoltaic prediction results. The previous literature focused on a single data source, comparing different weather features, or used historical time-series data directly for one-day-ahead PV power forecasting. The objective of this study was to find the appropriate weather features and weather data group paired with a deep learning algorithm to implement one-day-ahead PV power forecasting. In this work, the application of ANN, LSTM, and GRU combined with different weather data (LWS, CWB, and hybrid data) was used for one-day-ahead photovoltaic power generation forecasting, and the forecasting effects were compared. The area of study covers a rooftop solar photovoltaic system in Yangmei District, Taiwan. The data pre-processing method proposed in this article can prevent the failure of forecasting results and solve the data unavailability problem. Hybrid weather data show up to 5–8% improvements in accuracy compared to the CWB and LWS. Considering all of the weather patterns, the LSTM model is more accurate for one-day-ahead PV power forecasting. MAPE, MAE, and RMSE are 16.984%, 1.764, and 1.283, respectively. The results indicate that when the amount of data is limited, using hybrid data and the five weather features is helpful for training the model. Accordingly, the proposed model shows better one-day-ahead PV forecasting.

The pre-processing data framework may remove a large number of historical data points, which may lead to insufficient data and data sample discontinuity. This causes a limitation in model training in this study. Nevertheless, the deep learning method has good generalization ability (the adaptability of machine learning algorithms to fresh samples). Therefore, keeping the valuable data will have the best effect on long-term predictions. This article shows that the validation data over the one-month study is stable. The MAE, MAPE,
and RMSE for one month are 16.984%, 1.764, and 1.283, respectively. These results support that the LSTM model combined with hybrid weather data is reliable for one-day-ahead PV power forecasting.

**Author Contributions:** This paper is the collaborative work of all authors. Conceptualization, C.-H.C. and C.-C.W.; methodology, W.-C.K. and S.-H.H.; software, W.-C.K. and S.-H.H.; validation, W.-C.K., C.-H.C. and S.-H.H.; formal analysis, W.-C.K.; investigation, W.-C.K. and C.-H.C.; resources, C.-H.C.; data curation, W.-C.K.; writing—original draft preparation, W.-C.K.; writing—review and editing, C.-C.W.; supervision, C.-H.C.; project administration, C.-H.C., C.-C.W. and W.-C.K. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the Ministry of Science and Technology, Taiwan, R.O.C., grant number MOST 110-2221-E-035-091 and MOST 110-2221-E-A49-079.

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

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The authors have not provided data associated with this article.

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

**References**


