Photovoltaic Power Output Prediction Based on TabNet for Regional Distributed Photovoltaic Stations Group

Dengchang Ma, Rongyi Xie, Guobing Pan, Zongxu Zuo, Lidong Chu and Jing Ouyang

Abstract: With the increasing proportion of distributed photovoltaic (DPV) installations in county-level power grids, to improve the centralized operation and maintenance of the stations and to meet the needs of power grid dispatching, the output of the county-level regional DPV stations group needs to be predicted. In this paper, the weather prediction information is used to predict the output based on the model input average strategy. To eliminate the effect of the selected non-optimal training sample collection period on the prediction accuracy, an ensemble prediction method based on the minimum redundancy maximum relevance criterion and TabNet model is carried out. To reduce the influence of weather prediction errors on the power output prediction, a modified model based on error prediction is proposed. The ensemble prediction model is used to predict the day-ahead output, and a combination prediction model based on the proposed ensemble prediction model and the proposed modified model is established to predict the hour-ahead output. The experimental results verify the effectiveness of the proposed models. Compared with the corresponding reference models, the proposed ensemble prediction method reduces the normalized mean absolute errors (nMAEs) and the normalized root mean square errors (nRMSEs) of the day-ahead output prediction results by 2.86% and 5.51%, respectively. The combination prediction model reduces the nMAE and nRMSE of the hour-ahead output prediction results by 3.05% and 3.05%, respectively. Therefore, the prediction accuracy can be improved by the proposed models.

Keywords: distributed photovoltaic; regional power output prediction; minimum redundancy maximum relevance criterion; TabNet; ensemble prediction

1. Introduction

As one of the most potential new energy utilization technologies, photovoltaic (PV) power generation has attracted the attention of most countries in the world [1]. PV systems are available in two forms: Centralized and distributed. Compared with the centralized, distributed PV (DPV) systems are constructed close to loads, and the system output can be absorbed locally, which helps to overcome the common defects of mismatch between actual distribution and application demand of PV resources. Therefore, in recent years, the Chinese DPV industry has been vigorously developed along with the whole county advance.

However, with the increasing proportion of DPV installations in county-level regions, the power grid operation is greatly affected. Considering the system construction and operation costs, the traditional DPV system operation and maintenance are extensive, resulting in low system benefits. The centralized operation and maintenance management of DPV systems in a county can improve the overall efficiency of the systems. To meet the needs of power grid dispatching and the centralized operation and maintenance of DPV
stations, the output of the regional DPV stations group needs to be predicted. Therefore, we take DPV systems in a county as a whole to study the regional DPV output prediction.

Regional PV prediction models can be divided into different types. Liu et al. [2] reviewed the research on regional PV prediction methods based on multiple time scales. Kim and Kim [3] divided the models into two categories: Type 1 is about the public utility-scale systems [4–13], Type 2 is about the system behind the meter [14–17]. Pierro et al. [9] proposed an interesting classification method based on prediction strategies:

1. Bottom-up strategy. In this strategy, firstly, the output of each PV system in the region needs to be predicted, and then the regional output can be achieved by accumulating these predicted values.

2. Upscaling strategy. The strategy can be further divided into models output average strategy and model input average strategy. Model output average strategy is based on the selection and output prediction of a subset of regional PV systems, which is taken as the representation of the whole systems. Then the predicted subset power output is rescaled based on the subset capacity and total capacity to predict the regional power output. In the model output average strategy, the PV output in the predicted area is taken as the output of a virtual PV system. Then, the regional PV power output is directly predicted.

For the model output average strategy, Shaker et al. [18] proposed a data-driven method to estimate the power output of invisible PV systems based on the measured values of a little of representative systems. The representative sites were selected using the proposed data dimension reduction model based on K-mean clustering and principal component analysis, and then regional PV power generation was obtained according to the mapping function. In addition to power generation estimation, Shaker et al. [17] also proposed a probabilistic prediction model based on a Fuzzy Arithmetic Wavelet Neural Network (FAWNN) to predict the power generation of a large number of small PV systems. Bright et al. [19] evaluated the satellite-only and upscaling-only PV output estimate methods, and the authors concluded that the method through combining the two methods is more beneficial. Saint-Drenan et al. [8] analyzed the performance of the upscaling strategy by using measured power data of a set of 366 PV systems. They found that the error decreases with an increasing number of reference systems and a decreasing number of un-metered systems, and the average distance between a reference and the unknown system has a great influence on the performance of a set of reference systems.

For the model input average strategy, Fonseca et al. [5] proposed a method based on principal component analysis, support vector regression, and weather prediction data. One-day ahead regional PV power outputs of the four main regions of Japan in 2009 were predicted with hourly power output data of 453 PV systems. Aillaud et al. [20] proposed a model through a combination of a convolutional neural network (CNN) with a long short-term memory architecture. The day-ahead regional PV power outs of Germany were predicted, and the main result of this study shows that the proposed model is more accurate than the Random Forest model. Moschella et al. [21] attempted to directly predict wind and solar power generation in each Italian region based on the model input average strategy. Based on the same strategy, Pierro [22] et al. conducted a more detailed study on the solar power generation prediction of six regions in Italy by comparing six different prediction models. Yu et al. [23] presented a probabilistic prediction method based on CNN and non-linear quantile regression (QR). The model was used to predict the regional PV power output of PV systems in the Weifang region of China, and the prediction result shows that the improved CNN can effectively process high-dimensional and complex input data and the non-linear QR model can provide quantile prediction information of regional PV power output.

The upscaling prediction strategy is improved in some interesting studies, such as the research of Pierro et al. [9], Wolff et al. [24], Saint-Drenan et al. [12], and Fu [10]. They first clustered the PV systems in the region and then used the upscaling prediction strategy on the clustered subsets, respectively, to predict the PV power output of the whole region.
There are also some studies in which different strategies were compared. Fonseca et al. [6] conducted a comparative study on the four models, and each prediction method assumed a different scenario regarding the data available to make the prediction. In view of the complete availability of regional PV power data, the strategy of direct prediction and then accumulation is adopted. A prediction model based on stratified sampling is proposed for the partial availability of regional PV power data. In light of the availability of regional aggregate PV power, the model input average strategy is adopted. In the case that the power data cannot be obtained, the strategy of indirect prediction and then accumulation is adopted. By comparison, in the region with a variety of weather conditions, the prediction methods based on single systems’ predictions and the one based on stratified sampling provided the best results. Zamo et al. [7] predicted the regional PV power generation in two counties based on the bottom-up strategy and the model input average strategy. By using a reference system to directly predict the regional aggregate PV power, the results can get an RMSE of about 6%, whatever the county and the RMSE can be reduced to about 5.8% by using the bottom-up strategy. Pierro et al. [9] firstly clustered PV systems in a region and then compared two strategies: (1) Calculate the average prediction results of each cluster to obtain the regional PV power (the models output average strategy), (2) the input variables of each cluster center are used to directly predict the regional PV power (the model input average strategy). The results show that the accuracy of the latter is a little better.

Saint-Drenan et al. [25] proposed a new strategy that can be used as an alternative to the upscaling strategy for the scenario where no or few power measurements are available. The strategy uses an average PV model to calculate the power output of the most frequent module orientation angles. The calculated power values are finally weighted according to their probability of occurrence to estimate the real power output. The basic condition of this strategy is that the physical model information of regional PV systems is available. However, in practice, it is usually difficult to meet this condition.

For regional PV output prediction, the bottom-up strategy needs to predict the output of all systems in the predicted area. It is necessary to establish a prediction model for each system and perform a lot of data processing and calculation. When some PV systems in the region lack historical data and cannot apply the data-driven model, it is necessary to adopt the prediction method based on the physical model. However, in reality, it is difficult to obtain the physical models of all PV systems in a region. Therefore, the bottom-up strategy is actually difficult to apply in practice [10]. To reduce the amount of calculation through simplified methods, the research focus of regional PV output prediction mainly focuses on the upscaling strategy [9].

Through the study of the previous research, we found that there are two problems with the county-level regional DPV output prediction. The first is about the available data resources for prediction. Most of the DPV systems in county-level regions are small rooftop PV systems. Considering the construction cost, there is generally no single output prediction device in these systems, which leads to the lack of predicted output data from single systems. For the same reason, there is also no meteorological data acquisition device in these systems, which leads to the lack of locally measured meteorological data for output prediction. Although the easily available weather prediction data can be used to predict the regional power output, the inherent weather prediction errors will affect the output prediction accuracy. The lack of available data resources and the weather prediction errors make it difficult to directly use the previously proposed models to predict the county-level regional DPV output. The second is about the prediction method. Most of the previously proposed deep neural network (DNN) architectures are successfully applied to images, text, and audio but are not well suited for tabular data [26]. Therefore, there are few studies on the regional PV output prediction method based on DNNs. As is the case with other data-driven models, when a new DNN for tabular data, such as TabNet, is applied to predict the county-level regional DPV output, the optimal training sample collection period (TSCP) is dynamically changing, and it is difficult to select this hyper-parameter, so
generally a fixed value is used, or all historical samples are taken as the training samples, which will reduce the accuracy of the predicted results.

In this paper, the weather prediction information is used to predict the county-level regional DPV output based on the model input average strategy. To eliminate the effect of the selected non-optimal TSCP on the prediction accuracy, an ensemble prediction method based on the minimum redundancy maximum relevance (mRMR) criterion and the TabNet model is carried out. To reduce the influence of weather prediction errors on the power output prediction, a modified model based on error prediction is proposed. The proposed ensemble prediction method is used to predict the day-ahead output, and a combination prediction model based on the proposed ensemble prediction method and the modified model is established to predict the hour-ahead output. Finally, the performance of the proposed models is verified by error analysis.

2. Theoretical Background

2.1. TabNet Model

TabNet model [26], a novel DNN for tabular data, is introduced in this paper to predict the county-level regional DPV output. As shown in Figure 1, TabNet’s encoding is based on sequential multi-step processing with \( N_{\text{steps}} \) decision steps and can be summarized as follows:

- Step 1:
  1. **Feature transformer**: Process the normalized features using a feature transformer and then split into two parts, \([d[i], a[i]]\), where \(d[i] \in \mathbb{R}^{B \times N_d}\) and \(a[i] \in \mathbb{R}^{B \times N_a}\).
  2. **Attentive transformer**: Take \(a[0]\) as the input of the attentive transformer to obtain the mask \(M[1] \in \mathbb{R}^{B \times D}\). Employ \(M[1]\) to select the salient features and then process the selected features by using the feature transformer. Split the processed features for the decision step output and information for the subsequent decision step, \([d[1], a[1]]\), where \(d[1] \in \mathbb{R}^{B \times N_d}\) and \(a[1] \in \mathbb{R}^{B \times N_a}\).

- Step 2:
  1. **Agg.**
  2. **ReLU**

(1) \(D\)-dimensional features \(f \in \mathbb{R}^{B \times D}\) are normalized by applying batch normalization (BN) and then passed to each decision step, where \(B\) is the batch size.

(2) Process the normalized features using a feature transformer and then split into two parts, \([d[0], a[0]]\), where \(d[0] \in \mathbb{R}^{B \times N_d}\) and \(a[0] \in \mathbb{R}^{B \times N_a}\).

(3) Perform decision Step 1. Take \(a[0]\) as the input of the attentive transformer to obtain the mask \(M[1] \in \mathbb{R}^{B \times D}\). Employ \(M[1]\) to select the salient features and then process the selected features by using the feature transformer. Split the processed features for the decision step output and information for the subsequent decision step, \([d[1], a[1]]\), where \(d[1] \in \mathbb{R}^{B \times N_d}\) and \(a[1] \in \mathbb{R}^{B \times N_a}\).

(4) Similarly, perform the other decision steps to obtain the split features \([d[i], a[i]]\), where \(d[i] \in \mathbb{R}^{B \times N_d}\) and \(a[i] \in \mathbb{R}^{B \times N_a}\).

(5) Construct the overall decision embedding as \(d_{\text{out}} = \sum_{i=1}^{N_{\text{steps}}} \text{ReLU}(d[i])\) and apply a linear mapping \(W_{\text{final}} d_{\text{out}}\) to get the output mapping.

Figure 1. A TabNet model example.
(6) Obtain the aggregate feature importance mask \( M_{agg-b,j} \) by the following formula:

\[
M_{agg-b,j} = \sum_{i=1}^{N_{steps}} \eta_b[i] M_{b,j}[i] / \sum_{i=1}^{D} \sum_{j=1}^{N_{steps}} \eta_b[i] M_{b,j}[i],
\]

where \( M_{b,j}[i] \) is the element in the \( b^{th} \) row and the \( j^{th} \) column of the mask in the \( i^{th} \) decision step. \( \eta_b[i] \), which denotes the aggregate decision contribution at the \( i^{th} \) decision step for the \( b^{th} \) sample, is calculated as follows:

\[
\eta_b[i] = \sum_{c=1}^{N_d} \text{ReLU}(d_{b,c}[i]),
\]

where \( d_{b,c}[i] \) is the element in the \( b^{th} \) row and the \( c^{th} \) column of the split features for the decision step \( i \).

According to the structure of the attentive transformer shown in Figure 2, the feature selection mask \( M[i] = \text{sparsemax}(P[i - 1] \cdot h_i(a[i - 1])) \), where sparsemax is the sparsemax normalization, \( a[i - 1] \) is the processed features from the preceding decision step, and \( h_i \) is a trainable function with a FC (fully-connected) layer followed by BN. \( P[i - 1] \), the prior scale term denoting how much a particular feature has been used previously, is defined as follows:

\[
P[i - 1] = \prod_{j=1}^{i-1} (\gamma - M[j]),
\]

where \( \gamma \) is a relaxation parameter. \( P[0] \) is initialized as all ones, \( 1^{B \times D} \).

![Figure 2. An attentive transformer block example.](image)

For parameter-efficient and robust learning with high capacity, a feature transformer should comprise layers that are shared across all decision steps, as well as decision step-dependent layers. Figure 3 shows the implementation as a concatenation of two shared layers and two decision step-dependent layers. Each FC layer is followed by BN and gated linear unit (GLU) nonlinearity [27], eventually connected to a normalized residual connection with normalization. Normalization with \( \sqrt{0.5} \) helps to stabilize learning by ensuring that the variance throughout the network does not change dramatically [28]. For faster training, the ghost BN [29] form is used in the feature transformer.

![Figure 3. A feature transformer block example.](image)
2.2. Mutual Information

In information theory, mutual information (MI) is used to represent the degree of dependence between two systems or the correlation between two variables [10]. MI can be represented based on system entropy, which expresses the complexity or uncertainty of a system. Assuming that the probability of system output \( Y = y \) is \( P(y) \), the system entropy \( H(Y) \) is defined as follows:

\[
H(Y) = -\sum_y P_Y(y) \log P_Y(y),
\]

(4)

When the system input \( X = x \) is known, the conditional entropy \( H(Y|X) \) is defined in Equation (5):

\[
H(Y|X) = -\sum_x P_X(x) \left( \sum_y P_{Y|X}(y|x) \log P_{Y|X}(y|x) \right),
\]

(5)

where \( P_{Y|X}(y|x) \) is the conditional probability of \( Y \) when the system input \( X = x \).

The system joint entropy is defined as follows:

\[
H(Y, X) = -\sum_{y,x} P_{YX}(y,x) \log P_{YX}(y,x),
\]

(6)

Since the known \( X = x \) can reduce the uncertainty of the system, the conditional entropy \( H(Y|X) \) is smaller than the system entropy \( H(Y) \). MI is used to quantify the degree of system uncertainty decrease, represented by \( I(X, Y) \) in Equation (7):

\[
I(X, Y) = H(X) + H(Y) - H(X|Y)
= H(X) - H(X|Y)
= H(Y) - H(Y|X)
= I(Y|X),
\]

(7)

Based on Equations (4), (5), and (7), \( I(X, Y) \) is calculated as follows:

\[
I(X, Y) = \sum_x \sum_y P_{XY}(x,y) \log \frac{P_{XY}(x,y)}{P_X(x)P_Y(y)},
\]

(8)

where \( P_{XY}(x,y) \) represents the joint probability distribution when \( X = x \) and \( Y = y \). When \( X \) and \( Y \) are continuous variables, \( I(X, Y) \) is calculated by Equation (9):

\[
I(X, Y) = \iint p_{XY}(x,y) \log \frac{p_{XY}(x,y)}{p_X(x)p_Y(y)} \, dx \, dy
\]

(9)

2.3. mRMR Criterion

The mRMR criterion [30] is an eigenvalue selection method, and its core idea is to establish the optimal feature subset based on the maximum relevance condition and the minimum redundancy condition [31]. Let the eigenvalue set consisting of \( m \) eigenvalues be \( F_m \), from which \( n(n \leq m) \) eigenvalues are selected to form a subset \( S_n \), where \( F_m = \{v_i, i = 1, 2, \ldots, m\} \) and \( S_n = \{v_j, j = 1, 2, \ldots, n\} \), and \( S_n \subseteq F_m \). According to the maximum relevance condition, the average value of the MI between the characteristic variables in the subset and the target variable is the largest, and the constraints are as follows:

\[
\max D(S_n, c), \quad D = \sum_{i=1}^{n} I(v_i, c),
\]

(10)

where \( I(v_i, c) \) is the MI between the \( i \)-th eigenvalue and the target variable \( c \).

Eigenvalues of the subset \( S_n \) obtained by Equation (10) may be highly correlated, which may introduce unnecessary redundant. Therefore, in addition to meeting the maximum
relevance condition, the mean value of MI between eigenvalues in the subset $S_n$ should be minimized. The minimum redundancy condition is as follows:

$$\min R(S_n), \quad R = \frac{1}{C_n^2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} I(v_i, v_j),$$

(11)

where $I(v_i, v_j)$ is the MI between the eigenvalues of the subset $S_n$. Based on Equations (10) and (11), the mRMR criterion is expressed by Equation (12):

$$\max \Phi(D, R), \quad \Phi(D, R) = D - R$$

(12)

### 3. Proposed Method

#### 3.1. Data Experiment Scheme

As shown in Figure 4, the steps of the data experiment are:

1. Collect the measured output data (sampling period: 15 min). Collect the weather prediction data (sampling period: 1 h), and then obtain the weather prediction data of the same period with the measured output data by linear interpolation. Preprocess and combine the output data and weather data to establish the experimental sample set.

2. Take a test sample.

3. For the test sample extracted in the previous step, $n$ fixed TSCPs are randomly generated. Based on different TSCPs, $n$ training sample sets are established by extracting qualified training samples from the experimental sample set.

4. Based on the $n$ training sample sets established in the previous step, TabNet model is trained to generate $n$ prediction models.

5. Taking the $n$ prediction models generated in the previous step as base predictors, an ensemble prediction model based on mRMR is established.

6. Error analysis of day-ahead output prediction results.

7. Error analysis of hour-ahead output prediction results.

8. Proposed modified model.


**Figure 4.** The data experiment scheme.

Step 1: Collect the measured output data (sampling period: 15 min). Collect the weather prediction data (sampling period: 1 h), and then obtain the weather prediction data of the same period with the measured output data by linear interpolation. Preprocess and combine the output data and weather data to establish the experimental sample set.

Step 2: Take a test sample.

Step 3: For the test sample extracted in the previous step, $n$ fixed TSCPs are randomly generated. Based on different TSCPs, $n$ training sample sets are established by extracting qualified training samples from the experimental sample set.

Step 4: Based on the $n$ training sample sets established in the previous step, TabNet model is trained to generate $n$ prediction models.

Step 5: Taking the $n$ prediction models generated in the previous step as base predictors, an ensemble prediction model based on mRMR is established.
Step 6: Based on the test sample taken in Step 2 and the ensemble prediction model established in the previous step, the day-ahead and hour-ahead outputs are predicted, respectively.

Step 7: Based on the hour-ahead output predicted in the previous step and the proposed modified model, the final hour-ahead output is obtained.

Repeat Step 2 to Step 7 \( m \) (test sample size) times to obtain the day-ahead and hour-ahead output prediction series, respectively. Step 8 and Step 9 are prediction error analyses.

The normalized mean absolute error (nMAE) calculated by Equation (13) and the normalized root mean square error (nRMSE) calculated by Equation (14) are imported to present the prediction errors in this paper.

\[
\text{nMAE} = \frac{1}{n} \sum_{i=1}^{n} |p_i - p'_i|, \tag{13}
\]

\[
\text{nRMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (p_i - p'_i)^2}, \tag{14}
\]

where \( n \) is the test sample size, \( p_i \) is the normalized predicted regional DPV output, and \( p'_i \) is the normalized measured output.

3.2. Proposed Ensemble Prediction Model

In predicting the regional DPV output based on the model input average strategy, the training sample set affects the prediction performance of the trained TabNet model. The training sample set depends on the TSCP. For the output prediction of a specific period, there is a relatively optimal training sample set, which corresponds to a specific TSCP and the historical samples in the TSCP. With the systems running, the newly generated samples will update the historical sample set in the previous optimal TSCP, which will cause the training sample set in this TSCP is no longer optimal. Therefore, the TSCP corresponding to the optimal training sample set is dynamic. However, it is difficult to select the optimal value of this hyper-parameter. The TSCP selected by traditional methods is often not optimal, which affects the prediction accuracy. To solve this problem, we proposed an ensemble prediction model based on mRMR criterion and TabNet model, and the specific steps of the model are as follows:

Step 1: Randomly generate \( n \) fixed TSCPs and establish \( n \) training sample sets for the day before the predicted day. Then, based on the established \( n \) training sample sets and TabNet model, the regional DPV output series \( \{p_1, p_2, \ldots, p_n\} \) is predicted.

Step 2: Calculate each \( I(p_i, p) \), the MI between the \( n \) regional DPV output prediction series in the previous step and calculate each \( I(p_i, c) \), the MI between the \( n \) predicted output series and the measured output series \( c \).

Step 3: Let \( F = \{p_1, p_2, \ldots, p_n\} \), select the \( p_i \) with the largest \( I(p_i, c) \), let \( S = \{p_i\} \), and then update \( F \) as follows: \( F = F - \{p_i\} \);

Step 4: Select \( p_k \) which meet the conditions expressed in Equation (12) from \( F \), then update \( S \) and \( F \) as follows: \( S = S + \{p_k\} \), \( F = F - \{p_k\} \);

Step 5: Repeat Step 4 for a total of \( m - 1 \) times to select the \( m \) output prediction series according to the mRMR criterion from the \( n \) output prediction series in Step 1 and constitute a set \( S_m = \{p_1, p_2, \ldots, p_m\} \). The \( m \) fixed TSCPs corresponding to the \( m \) output prediction series in the set \( S_m \) are extracted to constitute a set \( T_m = \{t_1, t_2, \ldots, t_m\} \). Then calculate the MI between the predicted output series and the measured output series to constitute a set \( I(P_m, c) = \{I(p_{1}, c), I(p_{2}, c), \ldots, I(p_{m}, c)\} \).

Step 6: Calculate the weights by Equation (15) and construct a weight vector \( \omega = [\omega_1, \omega_2, \ldots, \omega_m]^T \).

\[
\omega_i = \frac{I(p_i, c)}{\sum_{j=1}^{m} I(p_j, c)}, \tag{15}
\]
Step 7: Predict the regional DPV output series of the predicted day based on the fixed TSCPs in $T_m$ and TabNet model. Then, construct an output prediction matrix $P = [p'_1, p'_2, \ldots, p'_m]$.

Step 8: Calculate the county-level regional DPV output prediction series $p_w$ of the predicted day by Equation (16):

$$p_w = P \times w$$  \hspace{1cm} (16)

3.3. Proposed Modified Model

The weather prediction information is taken as the input to predict the power output of the regional DPV stations group. Therefore, weather prediction accuracy has a great influence on the prediction accuracy of power output. However, weather prediction errors are similar in adjacent time periods. In this paper, a modified model based on error prediction is established by mining this similarity. Based on the known prediction errors, the unknown prediction errors are predicted, thus the influence of weather prediction errors on the output prediction accuracy is reduced.

To describe the proposed modified model, the concepts of potential test sample (PTS), non-potential test sample (NPTS), and the closest similar sample (CSS) are defined. A test sample is defined as PTS if there are some historical samples with the same weather type on the same day, otherwise it is defined as NPTS. There is some potential that the power output prediction error in a PTS period can be reduced by the proposed modified model. The closest historical sample on the same day and with the same weather type of PTS is defined as the CSS of the PTS. As shown in Figure 5, the steps of the proposed modified model are:

![Figure 5. The proposed modified model.](image)

Step 1: Take a test sample and determine whether it is a PTS. If so, proceed to the next step, if not, return to take the next test sample.

Step 2: Extract the historical PTSs with the same weather type of the PTS taken in the previous step and the CSSs corresponding to the historical PTSs.
Step 3: Calculate the prediction errors in the periods of PTSs extracted in Step 2 by Equation (17), and transform the errors by Equation (18):

$$e = p_p - p_a,$$

$$e_T = \frac{e + 1}{2},$$

where $e$ is the prediction error, $p_p$ is the normalized predicted output, $p_a$ is the normalized measured output, and $e_T$ is the transformed error.

Step 4: Calculate the differences of extraterrestrial solar radiation in the periods of PTSs extracted in Step 2 by Equation (19), and transform the differences by Equation (20):

$$G_d = G_P - G_C,$$

$$G_T = \frac{G_d + 1}{2},$$

where $G_d$ is the difference of extraterrestrial solar radiation in one of the periods of PTSs extracted in Step 2, $G_P$ is the normalized extraterrestrial solar radiation in the period of the PTS, $G_C$ is the normalized extraterrestrial solar radiation in the period of the CSS corresponding to the PTS, and $G_T$ is the transformed difference of extraterrestrial solar radiation.

Step 5: Establish a training sample set in which a training sample $S_{\text{train}} = [G_T, e_C, e_T]$, where $e_C$ is the transformed error in the period of the CSS corresponding to the training sample.

Step 6: Train a TabNet model based on the training sample set established in the previous step, and then predict the error $e_p$.

Step 7: Transform the predicted error according to Equation (21) and modify the predicted power output according to Equation (22).

$$e_m = 2e_p - 1,$$

$$p_m = p_p - e_m,$$

where $e_m$ is the transformed predicted error, $p_m$ is the modified regional DPV output, and $p_p$ is the output of the proposed ensemble prediction model.

3.4. Experimental Data and DATA Preprocessing

The raw data used in the experiment include measured regional DPV output data and weather prediction data from 7 January 2021 to 30 September 2021. The measured output data (sampling period: 15 min) is collected from 27 DPV systems in Xiaoshan District, Hangzhou City, Zhejiang Province, China. Weather prediction data (sampling period: 1 h) was obtained from Xinzhi weather prediction platform. The weather prediction data (sampling period: 15 min) is obtained based on the original weather prediction data and linear interpolation. The experimental sample set is established by combining multisource data sets according to the time attribute of samples. The attributes of the established experimental sample set include: Extraterrestrial solar radiation, weather type, air temperature, air index, wind speed, and measured regional DPV output.

After establishing the experimental sample set, data preprocessing is carried out. In order to improve the convergence speed and accuracy of DNN models, Max-Min normalization is usually carried out on the input and output features as shown in Equation (23):

$$x_{\text{nor}} = \frac{x - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}},$$

where $x_{\text{nor}}$ is a normalized feature value, $x$ is the original feature value, $x_{\text{max}}$ is the maximum feature value, and $x_{\text{min}}$ is the minimum feature value.
4. Results and Discussion

4.1. Output Prediction

According to the proposed ensemble prediction model, 16 fixed TSCPs (1, 3, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 180 days) were randomly generated in this paper to establish training sample sets. In particular, the experimental data on 27 September 2021 was selected to verify the proposed models. For output prediction, based on the mRMR criterion, the first 14 optimal output prediction series were selected from the 16 series on the day before the predicted day to calculate weights. The weight calculation results are listed in Table 1.

<table>
<thead>
<tr>
<th>Extraction Order of Output Prediction Series</th>
<th>Fixed TSCP Corresponding to Output Prediction Series</th>
<th>MI</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>4.28</td>
<td>7.69%</td>
</tr>
<tr>
<td>2</td>
<td>7</td>
<td>3.87</td>
<td>6.96%</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>4.25</td>
<td>7.64%</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>3.61</td>
<td>6.50%</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>4.13</td>
<td>7.42%</td>
</tr>
<tr>
<td>6</td>
<td>14</td>
<td>3.63</td>
<td>6.53%</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>4.10</td>
<td>7.37%</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>4.19</td>
<td>7.53%</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>4.03</td>
<td>7.24%</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>3.51</td>
<td>6.31%</td>
</tr>
<tr>
<td>11</td>
<td>15</td>
<td>4.12</td>
<td>7.40%</td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>4.14</td>
<td>7.45%</td>
</tr>
<tr>
<td>13</td>
<td>6</td>
<td>3.74</td>
<td>6.72%</td>
</tr>
<tr>
<td>14</td>
<td>9</td>
<td>4.01</td>
<td>7.21%</td>
</tr>
</tbody>
</table>

The regional DPV output series on the predicted day were predicted based on the fixed TSCPs and TabNet model. Then based on these predicted output series and calculated weights, the day-ahead output prediction series was calculated according to Equation (16). As shown in Figure 6, the predicted values of the day-ahead outputs on this day are generally lower than the corresponding measured values, which may be caused by the similarity of the weather prediction errors of the day.

![Figure 6. Prediction results based on the proposed ensemble prediction model.](image-url)
On the basis of the day-ahead prediction results, the hour-ahead output prediction results were obtained through the proposed modified model. As shown in Figure 7, compared with the day-ahead output prediction results, the predicted hour-ahead outputs are closer to the corresponding measured outputs.

![Figure 7](image_url)

**Figure 7.** Prediction results based on the proposed combination prediction model.

### 4.2. Performance Analysis

In order to further verify the validity of the proposed models, test samples from 1 May 2021 to 30 September 2021 were used. Figure 8 presents the measured regional DPV outputs versus the predicted outputs based on the proposed and reference models. Power values are normalized from 0 to 1 to show a better comparison. The blue points represent the predicted power values and the corresponding measured power values at the same time. The red solid line $y = x$ provides a reference when the predicted value is equal to the measured value, so the dispersion degree of the blue points around the red line reflects the error between the measured power and the predicted power. The more intensively the blue points aggregate to the red line, the less prediction error the corresponding model shows.

![Figure 8](image_url)

**Figure 8.** Cont.
Figure 8. The predicted outputs versus the measured outputs: (a) using the TabNet-based combination prediction model; (b) using the TabNet-based ensemble prediction model; (c) using the SVM-based combination prediction model; (d) using the SVM-based ensemble prediction model; (e) using the hour-ahead output persistence prediction model; (f) using the day-ahead output persistence prediction model.

The proposed ensemble prediction strategy is used to predict the day-ahead output, and the proposed combination prediction strategy is used to predict the hour-ahead output. As shown in Figure 8a,c,e, the prediction errors of the models with the combination prediction strategy are less than that of the hour-ahead output persistence prediction model, and the prediction error of the TabNet-based combination prediction model is smaller than that of the SVM based combination prediction model. Figure 8b,d,f show that the prediction errors of the models with the ensemble prediction strategy are less than that of the day-ahead output persistence prediction model, and the prediction error of the TabNet-based ensemble prediction model is smaller than that of the SVM based ensemble prediction model.

As shown in Figure 8, the distribution of the measured power versus prediction power points reveals an intuitive comparison between different prediction models but is not reliable in quantitative evaluation. Therefore, the prediction errors based on different models are listed in Tables 2 and 3 and described in Figures 9 and 10. In Figure 9, the numbers 1, 3, . . ., 180 represent the fixed TSCPs, E represents the proposed ensemble prediction strategy, and C represents the proposed combination prediction strategy. In the subsequent figures of this section, this abbreviation way is also applied.
Table 2. Prediction errors for each TabNet-based model.

<table>
<thead>
<tr>
<th>Model</th>
<th>nMAE</th>
<th>nRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSCP = 1</td>
<td>16.09%</td>
<td>23.74%</td>
</tr>
<tr>
<td>TSCP = 3</td>
<td>14.01%</td>
<td>20.86%</td>
</tr>
<tr>
<td>TSCP = 5</td>
<td>12.09%</td>
<td>17.62%</td>
</tr>
<tr>
<td>TSCP = 10</td>
<td>11.74%</td>
<td>17.60%</td>
</tr>
<tr>
<td>TSCP = 20</td>
<td>10.16%</td>
<td>14.20%</td>
</tr>
<tr>
<td>TSCP = 30</td>
<td>9.53%</td>
<td>13.36%</td>
</tr>
<tr>
<td>TSCP = 40</td>
<td>9.49%</td>
<td>13.12%</td>
</tr>
<tr>
<td>TSCP = 50</td>
<td>9.07%</td>
<td>12.33%</td>
</tr>
<tr>
<td>TSCP = 60</td>
<td>9.11%</td>
<td>12.43%</td>
</tr>
<tr>
<td>TSCP = 70</td>
<td>8.85%</td>
<td>12.06%</td>
</tr>
<tr>
<td>TSCP = 80</td>
<td>9.78%</td>
<td>13.75%</td>
</tr>
<tr>
<td>TSCP = 90</td>
<td>9.60%</td>
<td>12.97%</td>
</tr>
<tr>
<td>TSCP = 100</td>
<td>8.98%</td>
<td>12.11%</td>
</tr>
<tr>
<td>TSCP = 110</td>
<td>9.40%</td>
<td>13.35%</td>
</tr>
<tr>
<td>TSCP = 120</td>
<td>9.79%</td>
<td>13.86%</td>
</tr>
<tr>
<td>TSCP = 180</td>
<td>9.52%</td>
<td>12.84%</td>
</tr>
<tr>
<td>Ensemble prediction model</td>
<td>8.40%</td>
<td>11.11%</td>
</tr>
<tr>
<td>Combination prediction model</td>
<td>6.90%</td>
<td>9.49%</td>
</tr>
</tbody>
</table>

Figure 9. Prediction errors based on different models. (a) nMAEs predicted by the TabNet models with the fixed TSCPs, TabNet-based ensemble prediction model, and TabNet-based combination prediction model; (b) nRMSEs predicted by the TabNet models with the fixed TSCPs, TabNet-based ensemble prediction model, and TabNet-based combination prediction model; (c) nMAEs predicted by the SVM models with the fixed TSCPs, SVM-based ensemble prediction model, and SVM-based combination prediction model; (d) nRMSEs predicted by the SVM models with the fixed TSCPs, SVM-based ensemble prediction model, and SVM-based combination prediction model.
Table 3. Prediction errors for each SVM-based model.

<table>
<thead>
<tr>
<th>Model</th>
<th>nMAE</th>
<th>nRMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>TSCP = 1</td>
<td>27.79%</td>
<td>34.57%</td>
</tr>
<tr>
<td>TSCP = 3</td>
<td>12.16%</td>
<td>16.31%</td>
</tr>
<tr>
<td>TSCP = 5</td>
<td>11.49%</td>
<td>15.48%</td>
</tr>
<tr>
<td>TSCP = 10</td>
<td>9.60%</td>
<td>12.46%</td>
</tr>
<tr>
<td>TSCP = 20</td>
<td>9.30%</td>
<td>11.91%</td>
</tr>
<tr>
<td>TSCP = 30</td>
<td>9.24%</td>
<td>11.94%</td>
</tr>
<tr>
<td>TSCP = 40</td>
<td>9.12%</td>
<td>11.75%</td>
</tr>
<tr>
<td>TSCP = 50</td>
<td>8.99%</td>
<td>11.59%</td>
</tr>
<tr>
<td>TSCP = 60</td>
<td>9.02%</td>
<td>11.64%</td>
</tr>
<tr>
<td>TSCP = 70</td>
<td>8.97%</td>
<td>11.59%</td>
</tr>
<tr>
<td>TSCP = 80</td>
<td>9.00%</td>
<td>11.62%</td>
</tr>
<tr>
<td>TSCP = 90</td>
<td>8.97%</td>
<td>11.57%</td>
</tr>
<tr>
<td>TSCP = 100</td>
<td>9.01%</td>
<td>11.66%</td>
</tr>
<tr>
<td>TSCP = 110</td>
<td>8.98%</td>
<td>11.62%</td>
</tr>
<tr>
<td>TSCP = 120</td>
<td>9.01%</td>
<td>11.80%</td>
</tr>
<tr>
<td>TSCP = 180</td>
<td>9.01%</td>
<td>11.80%</td>
</tr>
<tr>
<td>Ensemble prediction model</td>
<td>8.85%</td>
<td>11.35%</td>
</tr>
<tr>
<td>Combination prediction model</td>
<td>7.71%</td>
<td>9.97%</td>
</tr>
</tbody>
</table>

Figure 9. Prediction errors based on different models. (a) nMAEs predicted by the TabNet models with the fixed TSCPs, TabNet-based ensemble prediction model, and TabNet-based combination prediction model; (b) nRMSEs predicted by the TabNet models with the fixed TSCPs, TabNet-based ensemble prediction model, and TabNet-based combination prediction model; (c) nMAEs predicted by the SVM models with the fixed TSCPs, SVM based ensemble prediction model, and SVM based combination prediction model; (d) nRMSEs predicted by the SVM models with the fixed TSCPs, SVM based ensemble prediction model, and SVM based combination prediction model.

Figure 10. Histogram comparison of the prediction errors: (a) Of day-ahead output prediction results; (b) of hour-ahead output prediction results.

It can be seen from Figure 9 that, based on different data-driven models, the proposed ensemble prediction strategy can eliminate the influence of the non-optimal sampling period of training samples on the prediction accuracy, and the proposed combination prediction strategy can further improve the hour-ahead prediction accuracy.

Figure 10 presents a comparative analysis of the prediction errors between the proposed models and the reference models. For day-head output prediction, the nMAEs predicted by the TabNet-based ensemble prediction model, SVM-based ensemble prediction model, and day-ahead persistence prediction model are 8.40%, 8.85%, and 11.26% respectively. Meanwhile, the nRMSEs are 11.11%, 11.35, and 16.62%, respectively. For hour-head output prediction, the nMAEs predicted by the TabNet-based combination prediction model, SVM based combination prediction model, and hour-ahead persistence prediction model are 6.90%, 7.71%, and 9.95%, respectively. Meanwhile, the nRMSEs are 9.49%, 9.97%, and 12.54%, respectively. The experiment results verified that the proposed methods are more accurate than the corresponding reference models.

In order to analyze the stabilities of the proposed models, the daily errors are analyzed. Figures 11 and 12 present daily prediction error distributions of the TabNet-based models and the SVM-based models, respectively. It can be seen that the distributions of daily
nMAE and nRMSE using the ensemble prediction strategy are lower than that using the fixed TSCPs, and the distributions using the combination prediction strategy are lower than that using the ensemble prediction strategy.

**Figure 11.** Boxplots of daily prediction errors of TabNet-based models. (a) nMAE boxplot; (b) nRMSE boxplot.

**Figure 12.** Boxplots of daily prediction errors of SVM-based models. (a) nMAE boxplot; (b) nRMSE boxplot.

Figure 13 presents a comparative analysis about the distributions of daily day-ahead prediction errors between the proposed model and the reference models. The mean values of daily nMAE predicted by the TabNet-based ensemble prediction model, SVM-based ensemble prediction model, and day-ahead persistence prediction model are 8.22%, 8.77%, and 11.12%, respectively, and the median values are 7.95%, 8.39%, and 9.90%, respectively. The mean values of daily nRMSE predicted by the three models are 10.19%, 10.65%, and 14.39%, respectively, and the median values are 9.58%, 10.03%, and 9.90%, respectively. Meanwhile, the boxplots and the statistical indicators verified that the proposed TabNet-based ensemble prediction model is more stable than the reference day-ahead prediction models.
Figure 13. Boxplots of daily day-ahead prediction errors. (a) nMAE boxplot; (b) nRMSE boxplot.

Figure 14 presents a comparative analysis about distributions of daily hour-ahead prediction errors between the proposed model and the reference models. The mean values of daily nMAE predicted by the TabNet based combination prediction model, SVM based combination prediction model, and hour-ahead persistence prediction model are 6.77%, 7.65%, and 9.88% respectively, and the median values are 6.46%, 7.39%, and 10.77% respectively. The mean values of daily nRMSE predicted by the three models are 8.83%, 9.49%, and 11.91%, respectively, and the median values are 8.38%, 9.12%, and 12.78% respectively. Both the boxplots and the statistical indicators verified that the proposed TabNet-based combination prediction model is more stable than the reference hour-ahead prediction models.

Figure 14. Boxplots of daily hour-ahead prediction errors. (a) nMAE boxplot; (b) nRMSE boxplot.

In order to analyze the variations in the performance of the proposed models with time, the monthly errors are analyzed. Figures 15 and 16, respectively, present the monthly nMAE and nRMSE of TabNet-based models. The closer the color of the histogram in the figure is to yellow, the greater the error value is, and the closer it is to blue, the smaller the error value is. It can be seen that, compared with the models with fixed TSCPs, the monthly errors by using the ensemble prediction strategy are lower. By using the proposed combination prediction strategy, the monthly errors are further reduced. Figures 17 and 18, respectively, present the monthly nMAE and nRMSE of SVM-based models. The experimental results are similar to those of the TabNet-based models.
Figures 17 and 18, respectively, present the monthly nMAE and nRMSE of SVM-based models. The experimental results are similar to those of the TabNet-based models.

Figure 15. The monthly nMAE histogram of the TabNet-based models.

Figure 16. The monthly nRMSE histogram of the TabNet-based models.

Figure 17. The monthly nMAE histogram of the SVM-based models.
Figure 18. The monthly nRMSE histogram of the SVM-based models.

Figure 19 presents a comparative analysis of the day-ahead prediction models. In September, the nMAEs of the proposed day-ahead prediction model (TabNet-based ensemble prediction model) and the two reference models (SVM-based ensemble prediction model and day-ahead persistence prediction model) are close to each other, and the nRMSE of the proposed day-ahead prediction model is close to that of the SVM-based ensemble prediction model but significantly higher than that of the day-ahead persistence prediction model. In other months, considering the two error indicators comprehensively, the performance of the proposed day-ahead prediction model is better than that of the reference models.

Figure 19. Histogram of monthly errors of the day-ahead prediction models. (a) nMAE histogram; (b) nRMSE histogram.

Figure 20 presents a comparative analysis of the hour-ahead prediction models. Considering the two error indicators comprehensively, the proposed hour-ahead prediction model (TabNet-based combination prediction model) is better than the reference models.

The experimental results verify that the proposed day-ahead and hour-ahead prediction models are more accurate and stable than the corresponding reference models and show robust performance with monthly variations.
Thus, the influence of weather prediction errors on the power output prediction is reduced by the proposed modified model. The nMAE and nRMSE of the hour-ahead output prediction results obtained by this combination prediction model are 6.9% and 11.11%, respectively. Therefore, the effect of the selected non-optimal TSCP on the prediction accuracy can be eliminated by the proposed ensemble prediction model.

Taking into account the influence of weather prediction errors on the power output prediction, a modified model based on error prediction is proposed. Firstly, the functional relationship between the prediction errors of the same weather type on the same day is learned. Then, based on the functional relationship, the prediction error of the predicted period is predicted. Finally, the output prediction result of the proposed ensemble prediction model is modified by the predicted error. The nMAE and nRMSE of the hour-ahead output prediction results obtained by this combination prediction model are 8.4% and 11.11%, respectively. Therefore, the effect of the selected non-optimal TSCP on the prediction accuracy can be eliminated by the proposed ensemble prediction model.

According to the overall error analysis, compared with the reference day-ahead prediction model, the proposed ensemble prediction model reduces the nMAE by 2.86% and the nRMSE by 5.51%, respectively, and compared with the reference hour-ahead prediction model, the proposed combination prediction model reduces the nMAE by 3.05% and the nRMSE by 3.05%, respectively. Based on the daily error analysis, compared with the reference day-ahead prediction model, the proposed ensemble prediction model reduces the mean value of daily nMAE by 2.9% and daily nRMSE by 4.2%, respectively, and compared with the reference hour-ahead prediction model, the proposed combination prediction model reduces the mean value of daily nMAE by 3.0% and daily nRMSE by 3.0%.

Figure 20. Histogram of monthly errors of the hour-ahead prediction models. (a) nMAE histogram; (b) nRMSE histogram.

5. Conclusions

This paper presents a new prediction method for the output of the county-level regional DPV stations group, which aims to improve the centralized operation and maintenance of the stations and to meet the needs of power grid dispatching. The weather prediction information is used to predict the output based on the model input average strategy. Considering the effect of the selected non-optimal TSCP on the prediction accuracy, an ensemble prediction method based on the mRMR criterion and TabNet model is carried out to predict the day-ahead output. Firstly, multiple fixed TSCPs are randomly generated, and the output prediction value of the predicted period is obtained by the weighted average of the output prediction series of the previous day calculated according to the mRMR algorithm. Then, based on the fixed TSCPs and the TabNet model, the output vector of the predicted period is predicted. Finally, the output prediction value of the predicted period is obtained by the weighted average method. The nMAEs and nRMSEs of the prediction results based on the fixed TSCPs are in the interval (8.85%, 16.09%) and the interval (12.06%, 23.74%), respectively. The nMAE and nRMSE of the prediction results based on the proposed ensemble prediction model are 8.4% and 11.11%, respectively. Therefore, the effect of the selected non-optimal TSCP on the prediction accuracy can be eliminated by the proposed ensemble prediction model.
prediction model reduces the mean value of daily nMAE by 3.11% and daily nRMSE by 3.08%, respectively. In accordance with the monthly error analysis, compared with the reference day-ahead prediction model, the proposed ensemble prediction model reduces the mean value of monthly nMAE by 2.91% and monthly nRMSE by 5.3%, respectively, and compared with the reference hour-ahead prediction model, the proposed combination prediction model reduces the mean value of monthly nMAE by 3.02% and monthly nRMSE by 3.01%, respectively. Therefore, the proposed day-ahead and hour-ahead prediction models are more accurate and stable than the corresponding reference models and show robust performance with monthly variations.

Author Contributions: Each author contributed extensively to this research. G.P. and J.O. provided the research ideas and guided the revision of the draft; D.M. and R.X. designed the experiment; D.M., Z.Z. and L.C. performed the experiments; D.M., Z.Z. and L.C. analyzed the data; and D.M. wrote the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research is financially supported by Key R&D Program of Zhejiang Province, China (2022C01244); and The National Key Research and Development Program of China (2017YFA0700301).

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

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

References

1. Fichter, T.; Soria, R.; Szkel, A.; Schaeffer, R.; Lucena, A.F. Assessing the potential role of concentrated solar power (CSP) for the northeast power system of Brazil using a detailed power system model. *Energy* 2017, 121, 695–715. [CrossRef]


Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.