

Article

Wind and Photovoltaic Power Generation Forecasting for Virtual Power Plants Based on the Fusion of Improved K-Means Cluster Analysis and Deep Learning

Zhichao Qiu ¹, Ye Tian ^{2,*}, Yanhong Luo ¹, Taiyu Gu ² and Hengyu Liu ²

¹ College of Information Science and Engineering, Northeastern University, Shenyang 110819, China; qiu zhichao0901@163.com (Z.Q.); luoyanhong@ise.neu.edu.cn (Y.L.)

² Institute of Electric Power Science, State Grid Liaoning Electric Power Co., Ltd., Shenyang 110000, China; guty_ldk@ln.sgcc.com.cn (T.G.); lhy_ldk@ln.sgcc.com.cn (H.L.)

* Correspondence: tianyemailbox@126.com

Abstract: Virtual power plants (VPPs) have emerged as an innovative solution for modern power systems, particularly for integrating renewable energy sources. This study proposes a novel prediction approach combining improved K-means clustering with Time Convolutional Networks (TCNs), a Bi-directional Gated Recurrent Unit (BiGRU), and an attention mechanism to enhance the forecasting accuracy of wind and photovoltaic power generation in VPPs. The proposed TCN-BiGRU-Attention model demonstrates superior predictive performance compared to traditional models, achieving high accuracy and robustness. These results provide a reliable basis for optimizing VPP operations and integrating renewable energy sources effectively.

Keywords: virtual power plant; cluster analysis; wind and photovoltaic power generation; TCN-BiGRU-Attention prediction model



Citation: Qiu, Z.; Tian, Y.; Luo, Y.; Gu, T.; Liu, H. Wind and Photovoltaic Power Generation Forecasting for Virtual Power Plants Based on the Fusion of Improved K-Means Cluster Analysis and Deep Learning. *Sustainability* **2024**, *16*, 10740. <https://doi.org/10.3390/su162310740>

Academic Editor: Jack Barkenbus

Received: 4 November 2024

Revised: 28 November 2024

Accepted: 3 December 2024

Published: 7 December 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

China's traditional power system relies heavily on coal-fired generation. According to data provided by China's National Energy Administration, by the end of 2023, the installed size of wind and photovoltaic power generation had increased tenfold compared to a decade ago. The installed capacity of clean energy power generation accounted for 58.2% of the total installed capacity, and new clean energy power generation accounted for more than half of the incremental electricity consumption of society as a whole. The share of clean energy consumption in total energy consumption rose from 15.5% to 26.4%, and the share of coal consumption fell by 12.1 percentage points. New results have been achieved in the clean and efficient use of energy. Over the past decade, China has eliminated more than 100 million kilowatts of outdated coal-powered electricity production capacity, and emissions of pollutants from the power sector have been reduced by more than 90%. The rate of electrification of end-use energy in society as a whole has reached 28%. With the large-scale grid integration of new energy power, its inherent random volatility brings great challenges to the stable operation of the power system, resulting in insufficient flexibility, inefficient resource utilization, and high investment costs. The VPP mainly relies on advanced information and communication technologies and software systems to aggregate flexible resources such as traditional power plants, distributed power sources, controllable loads, and energy storage systems within a certain range. It realizes centralized control and operation through intelligent dispatching and optimized management and participates in grid operation. The VPP has the capacity of adjustment on a scale and is able to cope with the problems of short-term power supply and demand tension and the difficulty of new energy consumption. It can be used as a “positive power plant” to supply power to the system for peak shifting and also as a “negative power plant” to increase

the load consumption and cooperate with the system for valley shifting. It realized the change from the traditional “source follows the load” scheduling mode to the new “source and load interaction” mode. Compared with traditional power plants, it is significantly economical and has contributed to balancing the supply-and-demand relationship of the power grid and maintaining the stable operation of the power grid.

VPPs deliver optimal scheduling and management of energy resources through integrating distributed energy sources such as wind and photovoltaic power generation. However, due to the stochastic and fluctuating nature of wind and photovoltaic power generation, accurately predicting their power output becomes a challenge. A massive number of data brings great value to the prediction, aggregation, and optimal scheduling of VPPs but, at the same time, increases the computational burden. Navid R et al. developed a mixed-integer nonlinear programming model for a stochastic self-scheduling problem for a new type of VPP structure, the hydroelectric VPP, using a K-means data clustering method for scenario reduction [1]. Cao W et al., considering the uncertainty factors of wind and photovoltaic (PV) power generation, used Latin Hypercubic Sampling to simulate wind and PV output scenarios and combined it with the improved CLARA clustering algorithm to reduce the scenarios to form a classical scenario set, which reduces the impact of wind and PV output volatility [2]. Aksan F et al. applied the clustering analysis technique to assess the power quality (PQ) of a VPP and compared it with coalescing algorithms with different scale characteristics. The results indicated that the K-means algorithm outperforms the agglomerative algorithm in clustering the data points without additional features of the PQ data according to the Calinski–Harabasz index and the profile coefficients [3]. Zhang R et al. proposed a real-time distributed clustering algorithm for aggregating distributed energy storage systems into heterogeneous VPPs [4]. Yin L F et al. proposed a short-term wind power prediction model based on Densely Connected Convolutional Networks (DenseNet). By simplifying the DenseNet201 network, the model obtained the DenseNet160 network with an excellent dense connection structure and appropriate depth and width [5]. Wang W G et al. proposed a combination model based on variational mode decomposition, tent chaotic mapping, a random walk sparrow search algorithm, and least squares support vector machines to solve the problem of inaccuracy in wind power prediction due to fluctuation and stochasticity in wind power prediction [6]. Wang Y et al. proposed an adversarial graph neural network approach by considering data scarcity scenarios for ultra-short-term distributed photovoltaic power prediction in VPPs [7]. Yu D et al. proposed a long-term and short-term memory network based on convolutional self-attention for a photovoltaic power generation prediction model [8]. Malik H et al. proposed a hybrid model combining Empirical Modal Decomposition and Neural Networks, which is used for multi-step-ahead load forecasting [9]. And Y et al. proposed a probabilistic ensemble prediction model that integrates the modules of data preprocessing, non-stationary time period discrimination, feature extraction, deterministic prediction, uncertainty prediction, and optimization integration into a common framework in order to improve the prediction accuracy of non-stationary time period intervals in photovoltaic power generation [10]. A novel deep learning method, SL-Transformer, for wind and solar energy prediction was proposed by Zhu J et al. Techniques such as the Savitzky–Golay filter and the Local Outlier Factor algorithm were employed for data preprocessing and combined with a self-concern mechanism to improve prediction accuracy. The model shows excellent performance compared to other methods, with a SMAPE of 4.2156% for solar and 5.8507% for wind, which is a significant improvement over existing models [11]. Soret A et al. explore a sub-seasonal forecasting methodology for predicting wind and solar power generation in Europe. It combines meteorological modeling and energy system data to address the challenges posed by the variability and intermittency of renewable energy sources. The findings provide tools for optimizing grid operations and policy planning [12]. Bhutta M S et al. used a hybrid machine learning model to improve the efficiency of solar power generation in smart grids. The HCLN model outperforms other models with high accuracy in predicting power generation, irradiance, and performance

ratios, demonstrating its potential for optimizing renewable energy systems [13]. Hos-sain M A et al. presented a hybrid deep learning model combining convolutional and GRU layers to improve short-term wind power forecasting. The model was tested on two Australian wind farms and outperformed existing models with an 8.13% improvement in MAPE, proving its accuracy and robustness [14]. Teixeira R et al. provide a comprehensive review of forecasting techniques for wind and solar energy, focusing on the integration of climate variables, optimization algorithms, and pre-processing methods. The importance of selecting appropriate forecast ranges and incorporating machine learning techniques to improve accuracy and stability is emphasized [15].

As can be seen, a variety of methods can be used to cluster and analyze the data, and a variety of artificial intelligence methods can also be applied to forecast wind and photovoltaic power generation. In this paper, the two methods are combined for distributed wind and photovoltaic power generation prediction in a VPP. Common virtual power plant optimization methods include Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Reinforcement Learning (RL). GA is a global optimization method based on a biological evolution mechanism, which is commonly used to solve multi-objective optimization problems of virtual power plants, such as cost minimization and reliability maximization. Its advantage is its strong search ability, but it is easy to fall into a local optimum. PSO is a population behavior-based optimization method for virtual power plant scheduling problems. Its convergence is fast, but its performance may be limited for high-dimensional complex problems. RL combines dynamic optimization and self-learning mechanisms to dynamically adjust virtual plant scheduling strategies in changing environments. Its advantage is that it is highly adaptable, but the training time is longer, and it is more dependent on high-quality environmental data. A VPP integrates energy resources (e.g., wind, solar, storage systems, fuel cells, flexible loads, etc.) that are distributed in different locations to be managed and optimized as a whole [16,17]. For this characteristic, the large size of data within the VPP makes data processing relatively difficult. On the basis of this, the main contributions of this study are as follows:

- (1) The wind and photovoltaic power generation data participating in the VPP aggregation are downsampled using principal component analysis to extract key features. Subsequently, cohesive hierarchical clustering is utilized to ascertain the reasonable number of clusters for the data. Finally, the K-means clustering algorithm is combined to cluster the data participating in the aggregation.
- (2) According to the clustering results, we train the model. First, the non-stationarity of time series data is handled through capturing long-time dependencies using causal convolution at the TCN layer of the temporal convolutional network. Secondary, based on the improvement of the gated recurrent unit GRU model, the BiGRU is used to convey information in both forward and backward directions to improve the ability of the model to understand the time series data. In the end, the attention mechanism is combined to improve prediction accuracy by calculating attention weights and highlighting the time steps that have an important impact on prediction results.

The innovation of this paper is mainly in combining the improved cluster analysis method with the proposed TCN-BiGRU-Attention model. The articles that have appeared in the literature have focused on one particular method or model in their discussion, and with a single method or model that is not optimized, the scope of application is relatively limited. The aim of this study is to apply the innovative method of combining the two in a virtual power plant, enabling the prediction of wind and photovoltaic power generation. That is, once the data are acquired, predictions can be implemented directly, greatly improving efficiency and applicability.

This paper combines clustering algorithms and artificial intelligence algorithms, firstly, to downsize the data to reduce the computational burden and secondly, to use the intelligent method to predict the uncertain variables in the VPP in order to achieve the optimal scheduling of the resources. Figure 1 shows the flowchart of the operating mechanism.

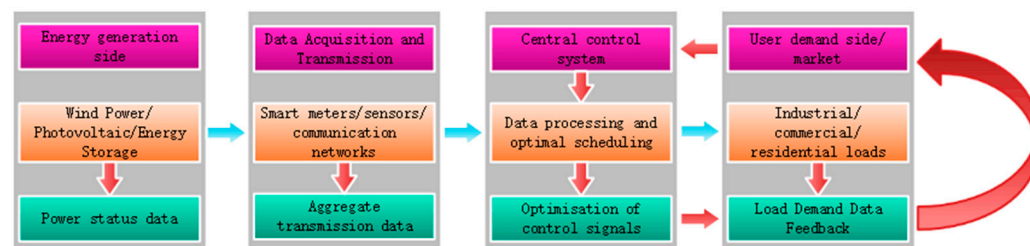


Figure 1. Flowchart of the operation mechanism.

2. Methods and Materials

The purpose of cluster analysis is to classify historical data into several classes of distinct patterns so that subsequent predictive models can be trained and predict corresponding to each class of patterns. Due to the huge number of distributed power data in the VPP, typical wind power and photovoltaic data are preselected to be analyzed in this paper. Combining the data obtained from the information–physical network, the wind and photovoltaic power generation data participating in the VPP aggregation are downscaled using principal component analysis to extract important features. Then, using a combination of cohesive hierarchical clustering and the K-means clustering algorithm, the data participating in the aggregation are clustered.

After clustering is completed, the TCN-BiGRU-Attention model is trained. First, the non-stationarity of time series data is handled through capturing long-time dependencies using causal convolution at the TCN layer of the temporal convolutional network. Next, based on the improvement of the gated recurrent unit GRU model, a bidirectional gated recurrent unit is used to convey information in both forward and backward directions so as to improve the ability of the model to understand the time series data. In the end, by calculating attention weights combined with the attention mechanism, the time steps that have an essential impact on prediction results are highlighted to enhance prediction accuracy.

2.1. Principal Component Analysis

Principal component analysis (PCA) is a statistical analysis method that classifies original multiple variables into a small number of composite variables. Principal component analysis is used for dimensionality reduction, which reduces the dimensionality of data by extracting the most representative features of the data [18]. This work analyzed the wind and photovoltaic power generation of a power plant in the Liaoning region of China for the whole year of 2019, which was recorded every 15 min, with a total of 35,040 data points for the whole year.

Wind power data are collected through specialized wind towers and hub-height sensors, and these data are collected in strict compliance with international wind power standards to ensure accuracy and consistency. Wind speed and direction were measured at 10 m, 30 m, 50 m, and 70 m in height, as well as at hub height. The data at different heights are used to assess the vertical variability of the wind resource and to ensure a comprehensive understanding of wind speed distribution. A precise mechanical anemometer, Thies Clima 4.3303.22.000, the manufacturer of this equipment is Thies Clima, Göttingen, Germany, is used to measure wind speed at different altitudes in m/s. The device is highly accurate and capable of measuring accurately at low wind speeds. Wind direction is measured in degrees (°) by means of a wind direction sensor, Thies Clima 4.3150.00.000; the manufacturer of the equipment is Thies Clima, located in Göttingen, Germany, mounted on an anemometer tower. The wind direction sensor responds quickly to changes in wind direction, ensuring accurate data collection in real time. Temperature (°C), barometric pressure (hPa), and humidity (%) are measured by weather station equipment mounted on the wind tower. All equipment is calibrated to meet meteorological monitoring specifications. The key parameters for wind power data collection are time, wind speed at 10 m, 30 m, 50 m, and 70 m in the height of the wind tower (m/s); wind speed at hub height (m/s); wind direction

at 10 m, 30 m, 50 m, and 70 m in the height of the wind tower ($^{\circ}$); wind direction at hub height ($^{\circ}$); the temperature ($^{\circ}\text{C}$); barometric pressure (hPa); and humidity (%).

The output power of a wind turbine can be calculated using the following formula:

$$P = \frac{1}{2} \cdot \rho \cdot A \cdot v^3 \cdot C_P \quad (1)$$

where P is the output power of the wind turbine (W); ρ is the air density (kg/m^3); A is the area swept by the wind turbine (m^2); v is the wind speed (m/s); and C_P is the power coefficient, which reflects the proportion of wind energy that can be captured by the turbine.

The air density ρ can be calculated from the temperature, air pressure, and humidity by the following formula:

$$\rho = \frac{p}{R \cdot T \cdot (1 + 0.378 \cdot e/p)} \quad (2)$$

Wherein ρ is the density of air (kg/m^3); p is the air pressure (Pa), 1 hPa = 100 Pa; T is the absolute temperature (K), i.e., the ambient temperature $T_{\text{ambient}} + 273.15$; R is the gas constant of air, which is usually $287.05 \text{ J}/(\text{kg} \cdot \text{K})$; and e is the water vapor pressure (Pa), which can be calculated from the humidity and temperature.

The water vapor pressure e is calculated as

$$e = h \cdot e_s \quad (3)$$

Here, e is the water vapor pressure (Pa); h is the relative humidity (%) expressed as a decimal (e.g., 50% = 0.5); and e_s is the saturated water vapor pressure (Pa), which can be calculated from the ambient temperature T_{ambient} using the following formula:

$$e_s = 6.112 \cdot \exp\left(\frac{17.67 \cdot T_{\text{ambient}}}{T_{\text{ambient}} + 243.5}\right) \quad (4)$$

The area A swept by the wind turbine is determined by the radius r of the wind turbine, which is calculated by the formula

$$A = \pi \cdot r^2 \quad (5)$$

The power generated by wind power is inversely proportional to the cube of the wind speed; thus, the wind speed v has a significant effect on the power. If the wind speed is low, the power output of the wind turbine will be drastically reduced. Typically, a wind turbine will have a minimum starting wind speed $v_{\text{cut-in}}$ and a maximum cut-out wind speed $v_{\text{cut-out}}$. If the wind speed is lower than the starting wind speed or higher than the cut-out wind speed, the wind turbine will not generate electricity, as per the following formula:

$$P = \begin{cases} 0, v < v_{\text{cut-in}} \text{ or } v > v_{\text{cut-out}} \\ \frac{1}{2} \cdot \rho \cdot A \cdot v^3 \cdot C_P, v_{\text{cut-in}} \leq v \leq v_{\text{rated}} \\ P_{\text{rated}}, v_{\text{rated}} < v \leq v_{\text{cut-out}} \end{cases} \quad (6)$$

In which $v_{\text{cut-in}}$ is the minimum starting wind speed, usually 3–4 m/s, and $v_{\text{cut-out}}$ is the maximum cut-out wind speed, usually 20–25 m/s; v_{rated} is the rated wind speed, and at this wind speed, turbine output is rated power P_{rated} .

Data acquisition for PV power generation is carried out at meteorological stations and with radiometric instruments installed in the vicinity of the PV plant. The total direct and scattered radiation received by the PV module was measured by a solar radiation meter, Kipp and Zonen CMP11, in W/m^2 , respectively. These devices are calibrated to ensure that the accuracy of measurements meets international standards. The PV module temperature and the ambient temperature are measured in $^{\circ}\text{C}$ using thermistors installed in the PV plant, ensuring the accurate monitoring of the module operating status and external environmental conditions. The barometric pressure (hPa) and humidity (%) are

measured by the weather station equipment to guarantee that the PV generation data match the environmental conditions. The key parameters for PV data collection are time, the module temperature ($^{\circ}\text{C}$), ambient temperature ($^{\circ}\text{C}$), air pressure (hPa), humidity (%), total radiation (W/m^2), direct radiation (W/m^2), and scattered radiation (W/m^2).

The data are automatically collected at high frequency and stored in the data acquisition system for subsequent analysis and use. Through data cleaning and filtering processing, outliers are removed to ensure data quality.

The actual power generated by the photovoltaic module can be calculated by the following formula:

$$P_{\text{output}} = G_{\text{inc}} \cdot A_V \cdot \eta r \quad (7)$$

P_{output} is the actual output power of the photovoltaic module (W); G_{inc} is the total radiation incident on the surface of the photovoltaic module (W/m^2); A_V is the area of the photovoltaic module (m^2); and ηr is the module efficiency corrected for the temperature.

The amount of radiation incident G_{inc} on the surface of the photovoltaic module is then determined:

$$G_{\text{inc}} = G_{\text{beam}} \cdot \cos(\theta) + G_{\text{diffuse}} + G_{\text{reflected}} \quad (8)$$

Wherein G_{total} denotes the total radiation from the sun, which includes direct exposure and scattering; G_{beam} is the radiation directly from the sun; and G_{diffuse} is the scattered radiation due to particulate matter in the atmosphere, among other things.

The efficiency of a photovoltaic module decreases as the temperature of the module increases. Therefore, temperature correction is required based on the provided module temperature T_{module} and the ambient temperature T_{ambient} . The module temperature can be calculated using the following equation:

$$T_{\text{module}} = T_{\text{ambient}} + \frac{G_{\text{total}}}{800} \times \Delta T \quad (9)$$

Here, ΔT is the temperature rise coefficient of the photovoltaic module, which is usually $20\text{--}30\text{ }^{\circ}\text{C}$. The efficiency of the photovoltaic module is then corrected using the temperature coefficient of the module, β , which is typically $0.4\text{--}0.5\%/^{\circ}\text{C}$, with the following equation:

$$\eta T = \eta_{\text{STC}} \times [1 - \beta \times (T_{\text{module}} - 25)] \quad (10)$$

In this case, η_{STC} indicates the conversion efficiency of the photovoltaic module under standard test conditions.

Environmental factors, such as barometric pressure and humidity, affect atmospheric conditions, which in turn affect the transmission rate of radiation. Typically, high humidity leads to an increase in scattered radiation and has a lesser effect on air pressure. The effects of these factors can be indirectly represented by radiation data.

The data were collected in 2019 from Liaoning, China, and have been put through a rigorous data processing and analysis process to ensure its accuracy and reliability. Due to the large number of data points, and in order to present the periodicity of the data and other characteristics, this study uses data from a particular day for analysis. The actual power generated from the wind power plant is shown in Figure 2a, and the actual power generated from the photovoltaic module is shown in Figure 2b.

Due to the huge amount of data and many variables, principal component analysis was chosen in this study to reduce the dimensionality of the data. Assuming that there are a total of n output curves of wind and photovoltaic participating in the scheduling of the VPP, each curve is a p -dimensional variable, and the sample matrix X consisting of these curves is

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \quad (11)$$

where x_{ij} is the value of the j th dimension of the i th sample. By calling the original variable indicators $x_1, x_2, \dots, x_p$, and setting new variables, i.e., composite indicator variables after the dimensionality reduction process, to be $z_1, z_2, \dots, z_m$ ($m \leq p$), we have

$$\begin{bmatrix} z_1^T \\ z_2^T \\ \vdots \\ z_m^T \end{bmatrix} = \begin{bmatrix} l_{11} & l_{12} & \cdots & l_{1p} \\ l_{21} & l_{22} & \cdots & l_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ l_{m1} & l_{m2} & \cdots & l_{mp} \end{bmatrix} \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_p^T \end{bmatrix} \quad (12)$$

where the new variables $z_1, z_2, \dots, z_m$ are the 1st, 2nd, $\dots$, m th principal components of the original variables $x_1, x_2, \dots, x_p$, and the coefficients l_{ij} are called principal component loadings.

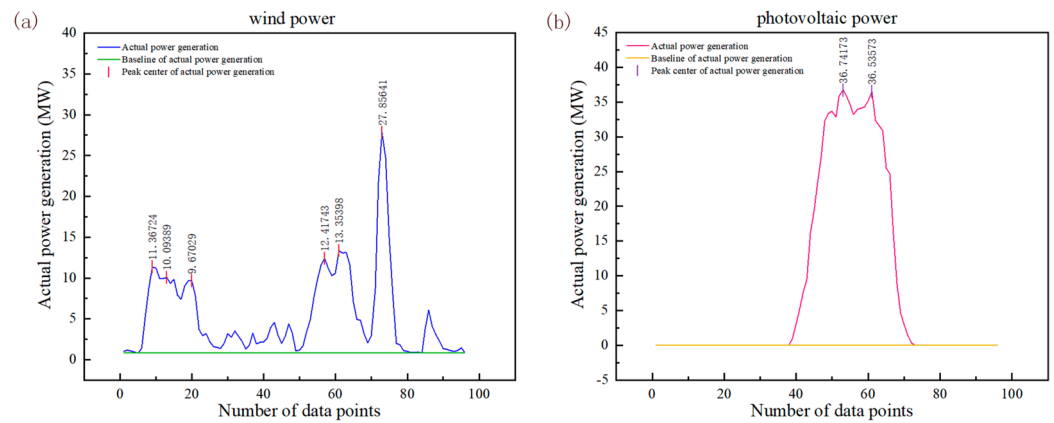


Figure 2. Actual power generated from wind and a photovoltaic module. (a) The actual power generated from the wind power plant. (b) The actual power generated from the photovoltaic module.

The matrix X is normalized as follows:

$$x_{ij}^* = \frac{(x_{ij} - \bar{x}_j)}{s_j}, i = 1, 2, \dots, n, j = 1, 2, \dots, p \quad (13)$$

$$\bar{x}_j = \frac{1}{n} \sum_{k=1}^n x_{kj} \quad (14)$$

$$s_j = \sqrt{\frac{1}{n-1} \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2} \quad (15)$$

where s_j is the standard deviation of the j th column of matrix X .

The calculation of the correlation coefficient matrix R_X of the normalized matrix is carried out as follows:

$$R_X = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix} \quad (16)$$

where r_{ij} is defined by the following equation:

$$r_{ij} = \frac{\sum_{k=1}^n (x_{ki} - \bar{x}_i)(x_{kj} - \bar{x}_j)}{\sqrt{\sum_{k=1}^n (x_{ki} - \bar{x}_i)^2 \sum_{k=1}^n (x_{kj} - \bar{x}_j)^2}}, i, j = 1, 2, \dots, p \quad (17)$$

By solving the eigenequation $|\lambda I - R| = 0$ by Jacobi's method, we find the eigenvalues, then arrange them in the order $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$, and find the unit eigenvector $e_i = [e_{i1}, e_{i2}, \dots, e_{ip}]^T$ ($i = 1, 2, \dots, p$) corresponding to λ_i .

The contribution rate α_i of the i th principal component is calculated as follows:

$$\alpha_i = \frac{\lambda_i}{\sum_{j=1}^p \lambda_j} \quad (18)$$

The cumulative contribution rate β_i is defined as

$$\beta_i = \frac{\sum_{k=1}^i \lambda_k}{\sum_{k=1}^p \lambda_k} \quad (19)$$

Generally, the eigenvalues with a cumulative contribution of 85–95%, corresponding to m principal components, respectively, are taken to describe the characteristics of the original sample and the loadings $l_{ij} = \sqrt{\lambda_i} e_{ij}$ ($i, j = 1, 2, \dots, p$).

The individual principal component scores can be acquired by the following matrix:

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1m} \\ z_{21} & z_{22} & \cdots & z_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nm} \end{bmatrix} \quad (20)$$

So far, the dimensionality reduction in each type of wind and photovoltaic output curve of the VPP has been realized, and in the subsequent calculations, these m principal components are used instead of the original complex data for the operation.

In this study, the number of principal components was selected so as to achieve a cumulative variance ratio of 95%. For wind power, the number of principal components chosen is 53. The explained variance ratio of principal components is shown in Figure 3a, which is used to show the contribution of each principal component to the variance of the data. The cumulative explained variance of principal components is shown in Figure 3b, which is used to determine how many principal components need to be selected to explain 95% of the variance in the data. For photovoltaic power, the number of principal components was chosen to be 11; the explained variance ratio of principal components is shown in Figure 4a, and the cumulative explained variance ratio of principal components is shown in Figure 4b.

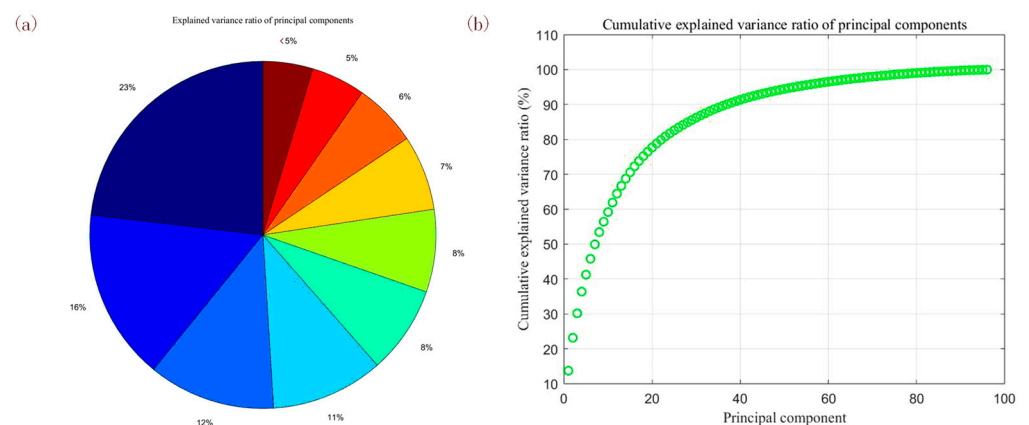


Figure 3. Wind power. (a) The explained variance ratio of principal components. (b) The cumulative explained variance of principal components.

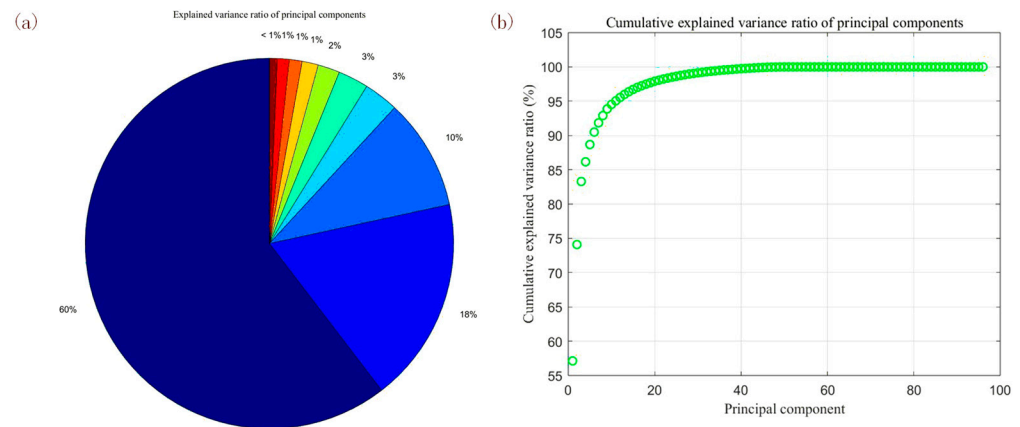


Figure 4. Photovoltaic power. (a) The explained variance ratio of principal components. (b) The cumulative explained variance ratio of principal components.

The difference in the number of principal components in the above results is mainly due to the complexity, variability, and structure of the data. Wind power generation data have higher dimensionality and complexity, such as higher variation in wind speed, which requires more principal components to capture the variability in the data. Whereas photovoltaic power generation data have less variability and simpler data patterns, thus only fewer principal components are required to explain the variance in the data.

2.2. Agglomerative Hierarchical Clustering

Agglomerative hierarchical clustering is a relatively common clustering method. The n objects are viewed as n clusters, and the Euclidean distance d_{ij} between the class clusters is calculated as follows:

$$d_{ij} = \sqrt{(z_{i1} - z_{j1})^2 + (z_{i2} - z_{j2})^2 + \cdots + (z_{im} - z_{jm})^2} \quad (21)$$

In which d_{ij} denotes the distance between the i th curve and the j th curve. $z_{i1}, z_{j1}, z_{i2}, z_{j2}, \dots, z_{im}, z_{jm}$ are the corresponding principal component scores in matrix Z . The n class clusters have a total distance of $n(n-1)/2$. Then, we arrange the distances from the smallest to the largest, with the arrangement containing the marking of the n objects. Eventually, in order for each distance to be calculated, the two class clusters a and b marked by it are compared. If it is in a different class cluster, it is coalesced into a single class cluster until the total number of class clusters is 1 or until the specific situation is reached [19].

Agglomerative hierarchical clustering is helpful in determining a reasonable number of clusters for the data. Through constructing a dendrogram, the results of clustering at different levels can be visualized, thereby providing a reasonable number of clusters for K-means called K [20]. In this way, bad clustering in K-means due to an improper choice of the K value can be avoided. For wind power, its hierarchical clustering dendrogram is shown in Figure 5a. For photovoltaic power, its hierarchical clustering dendrogram is shown in Figure 5b.

The performance of K-means depends heavily on the choice of initial cluster centers. The initial clusters obtained through agglomerative hierarchical clustering can be used as the initial cluster centers of K-means, which helps to improve the clustering effect of K-means, reduces its sensitivity to the initial value, and decreases the risk of falling into the local optimal solution.

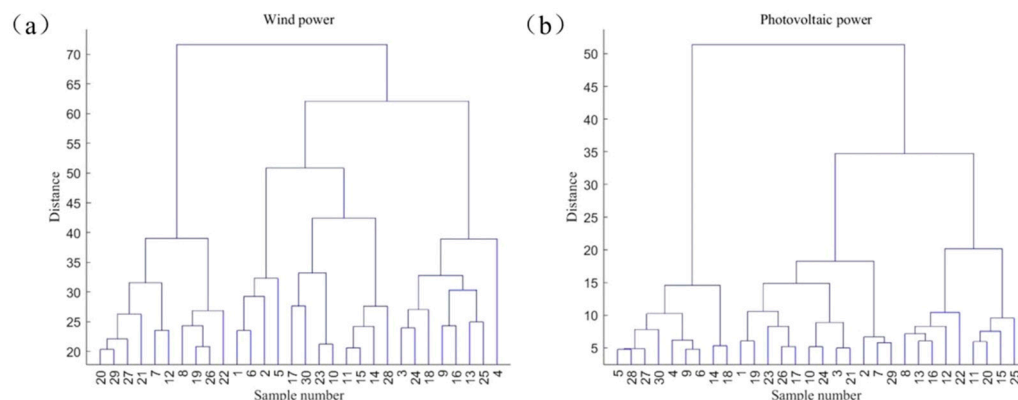


Figure 5. Hierarchical clustering dendrograms. (a) Hierarchical clustering dendrogram for wind power. (b) Hierarchical clustering dendrogram for photovoltaic power.

2.3. Improved K-Means Clustering

After the initial clustering center is determined from the above agglomerative hierarchical clustering, K-means clustering is performed. At first, the number of clusters K is determined, and the number of clusters chosen for this study is 10. Then, the matrix is input, and the dataset is subjected to agglomerative hierarchical clustering analysis, where the data of the obtained K class clusters are averaged to obtain the initial clustering centers. In the next step, K-means clustering is used to calculate the distance of each datum to all the cluster centers, and those data are classified into the class cluster where the cluster center nearest to it is located [21]. At last, after the delimitation is complete, the cluster centers of each cluster are recalculated until the cluster centers no longer change or no data are reassigned to another class cluster [22,23].

Figure 6 displays the distribution of the K-means clustering results in the space of the top two principal components with the highest explained variance, which capture the most important information about the variance in the data. The clustering results for the wind power data are shown in Figure 6a, and the clustering results for the photovoltaic power data are shown in Figure 6b.

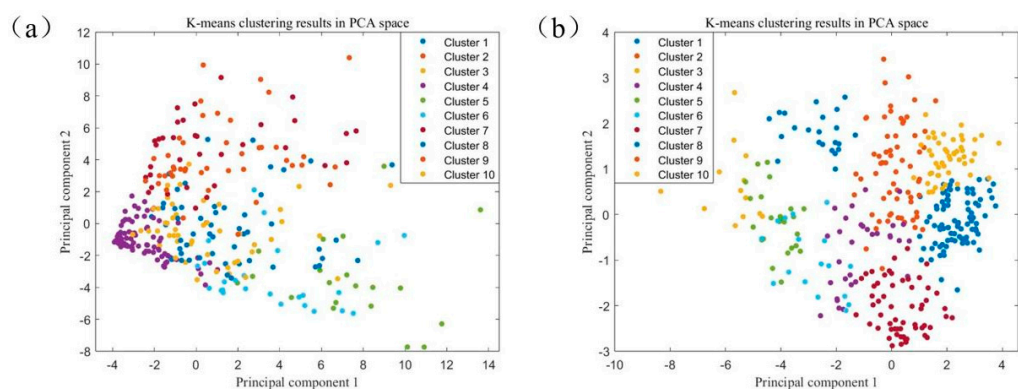


Figure 6. K-means clustering results in PCA space. (a) The clustering results for the wind power data. (b) The clustering results for the photovoltaic power data.

By adopting principal component analysis to reduce the dimensionality of the data and then combining agglomerative hierarchical clustering and K-means clustering methods to deal with the distributed and large-scale wind and photovoltaic power generation dataset in the VPP, the shortcomings of a single method are compensated [24]. This hybrid strategy exploits the advantages of agglomerative hierarchical clustering in the determination of the number of clusters and the selection of the initial cluster centers and, at the same time, plays the role of K-means in refining and optimizing cluster partitioning. The quality of the final clustering is improved, and better clustering results are obtained on complex

datasets. The final typical scenario after clustering of wind power is shown in Figure 7a. The final typical scenario after clustering of photovoltaic power is shown in Figure 7b. The differently colored curves in the figure indicate 1 to 10 different classes of clusters.

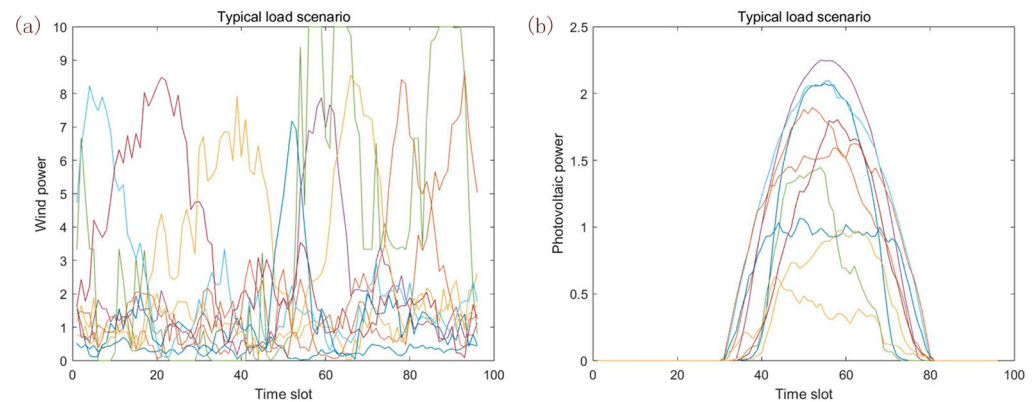


Figure 7. Typical scenario. (a) The final typical scenario after clustering of wind power. (b) The final typical scenario after clustering of photovoltaic power.

For the wind power dataset, the probabilities of occurrence of each scenario are 0.0767, 0.0274, 0.0849, 0.3836, 0.0548, 0.0685, 0.0932, 0.0685, 0.0904, and 0.0521, respectively. For the photovoltaic power dataset, the probabilities of occurrence of each scenario are 0.2384, 0.1123, 0.0411, 0.0849, 0.0658, 0.0466, 0.1562, 0.0603, 0.0603, and 0.1342, respectively. The clustered data are applied to the later wind and photovoltaic power generation prediction to simplify the complex data and improve computational efficiency.

2.4. Time Convolutional Network (TCN)

The TCN is a type of convolutional neural network for processing time series data. The TCN is based on one-dimensional convolutional operations, which have causal and inflationary convolutional features to ensure that the model captures long-term temporal dependencies [25]. The causal convolutional network ensures that convolution operations do not use future information and that the convolution kernel is only related to current and past time steps, avoiding information leakage. The expansion convolution network increases the receptive field by introducing expansion factors between convolution kernels, thus capturing dependencies over longer time scales without adding to the size of the convolution kernel. The TCN enables better training of deep networks by introducing a residual connection in order to avoid the problems of gradient vanishing and gradient explosion [26–29]. The structural framework of the TCN is shown in Figure 8.

2.5. Bidirectional Gated Recurrent Unit (BiGRU)

The Gated Recurrent Unit (GRU) is a recurrent neural network that processes time series data, which can capture short-term and long-term dependencies in time series [30]. It comprises a simple structure, few parameters, and fast training speed. However, the recursive structure of the GRU limits its efficiency and effectiveness in capturing long-distance dependencies. The model can handle sequential data only in the forward direction and utilizes information only up to the current time step, unable to utilize information after the current time step. Therefore, the BiGRU is introduced, which is able to capture information in the series more comprehensively through processing the time series data in both forward and backward directions [31,32]. The network structure of the BiGRU is shown in Figure 9 below.

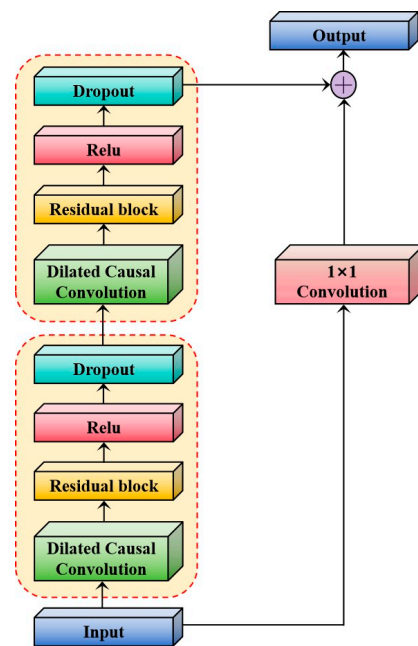


Figure 8. Structural framework of the TCN.

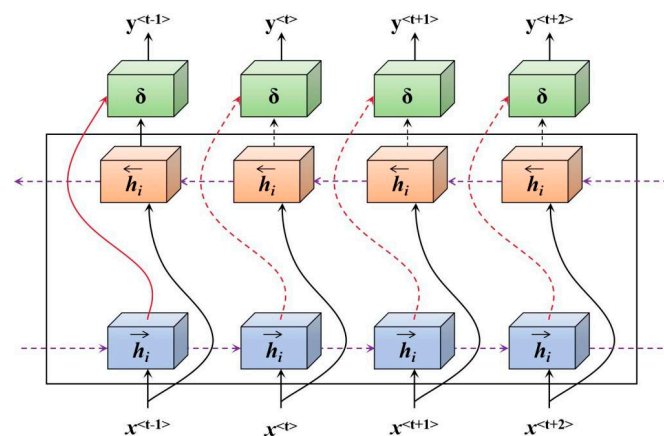


Figure 9. BiGRU network structure.

The formulae for the BiGRU are expressed as follows:

$$\vec{h}_t = \text{GRU}\left(x_t, \vec{h}_{t-1}\right) \quad (22)$$

$$\overleftarrow{h}_t = \text{GRU}\left(x_t, \overleftarrow{h}_{t-1}\right) \quad (23)$$

$$y_t = \vec{h}_t \oplus \overleftarrow{h}_t \quad (24)$$

In which $\vec{h}_t$ is the forward output of the BiGRU, $\overleftarrow{h}_t$ is the reverse output of the BiGRU, and y_t is the modeled output of the entire BiGRU.

2.6. Attention Mechanism

The attention mechanism is able to dynamically adjust the contribution of each part of the input sequence by calculating the importance of each time step [33–36]. In this way, the model can focus on more significant parts of the time series, thus raising prediction accuracy. The network block diagram of the attention mechanism is shown in Figure 10 below.

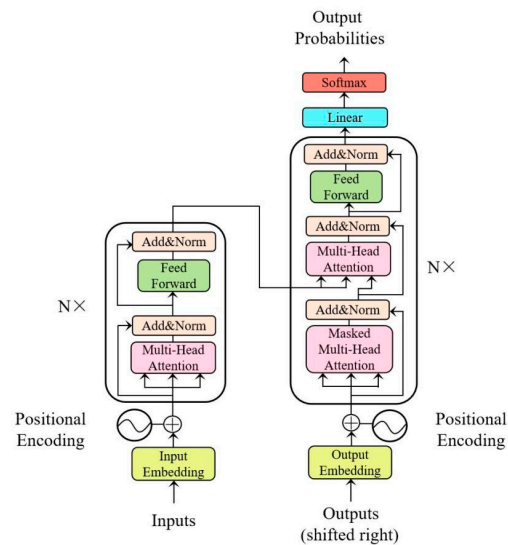


Figure 10. The transformer-model architecture.

We calculate the correlation between each group of hidden layer states h_j and h_i , as shown in Equation (25) below. We convert each correlation value into probability weights a_{ij} using SoftMax function, as shown in Equation (26) below. Finally, the effect of different h_j on h_i is calculated to obtain the output sequence of the augmentation function, as shown in Equation (27) below.

$$g_{ij} = \tanh(W_1 * h_i + W_2 * h_j + b) \quad (25)$$

$$a_{ij} = \text{soft max}(g_{ij}) = \frac{\exp(g_{ij})}{\sum_j \exp(g_{ij})} \quad (26)$$

$$H_i = \sum_j a_{ij} * h_j \quad (27)$$

where g_{ij} denotes the relationship between the state i and the state j , W_1 and W_2 denote the weights, and b denotes the offset vector. a_{ij} is the attentional weight from i to j . H_i describes the final output state of the attentional layer.

3. Results

3.1. Data Processing and Sample Set Partitioning

In this paper, the historical power generation data of wind farms and photovoltaic plants for the whole year from 1 January to 31 December 2019 of the VPP in Liaoning, China, are selected as the experimental dataset, with a sampling interval of 15 min and a total of 35,040 data points. Due to the VPP's requirement of the aggregation of multiple resources and a huge number of data, the data have been analyzed by clustering first in the previous Section 2. Therefore, in this section, the clustered data will be used as the dataset for the TCN-BiGRU-Attention model.

The dataset is divided into training and test sets in the ratio of 7:3, which is used to train and validate the proposed predictive model. At the same time, several comparison experiments with different models are set up to evaluate the prediction performance of the constructed models. Among them, the segmentation results for the wind power dataset are shown in Table 1, and those for the photovoltaic power dataset are shown in Table 2.

In order to make the model stable and promote the training speed, the wind and photovoltaic power generation data were normalized and converted into scalar values. The maximum–minimum normalization is used to normalize the data into the interval [0,1], and the calculation formula is shown below:

$$x' = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \quad (28)$$

Here, x' is the normalized value, x is the actual value of the sample, and $x_{\max}$ and $x_{\min}$ are the maximum and minimum values of the column in which these data are located.

Table 1. Statistics of the wind power dataset.

Dataset	NS ^a	XV ^b (%)	NV ^c (%)	AV ^d (%)	SD ^e (%)
Training set	672	10	0	2.0948	2.1816
Prediction set	288	8.8755	0.0092	1.7056	1.7966

NS ^a: Number of samples; XV ^b: Maximum value; NV ^c: Minimum value; AV ^d: Average value; SD ^e: Standard deviation.

Table 2. Statistics of the photovoltaic power dataset.

Dataset	NS ^a	XV ^b (%)	NV ^c (%)	AV ^d (%)	SD ^e (%)
Training set	672	2.2547	0	0.4155	0.6434
Prediction set	288	2.1030	0	0.5763	0.6720

NS ^a: Number of samples; XV ^b: Maximum value; NV ^c: Minimum value; AV ^d: Average value; SD ^e: Standard deviation.

3.2. Model Evaluation Indicator

So as to assess the performance of the model, the mean square error (MSE), mean absolute error (MAE), and the coefficient of determination (R^2) are used as evaluation metrics in this paper [37]. The corresponding equations are shown as Equations (27)–(30). Therein, the smaller values of MSE and MAE indicate a better prediction of the model. The MAE is less sensitive to outliers than the MSE because it calculates purely absolute errors without the amplifying effect of the squared term. The closer the R^2 value is to 1, the better the explanatory power of the model, i.e., the better the model fits the data.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (29)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (30)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (31)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (32)$$

in which n denotes the total number of data points involved, y_i denotes the actual value of the wind power sample, $\bar{y}_i$ denotes the average of the actual values, and $\hat{y}_i$ denotes the predicted value of wind and photovoltaic power.

3.3. Model Optimization and Training

With the aim of verifying the validity of the TCN-BiGRU-Attention model, several models of the GRU, BiGRU, TCN-GRU, TCN-BiGRU, and TCN-GRU-Attention were also selected for comparative experiments in this study. During the training process, the control variable method was adapted to adjust the network parameters. The prediction accuracy is judged by comparing the Root Mean Square Error ($RMSE$) until the model is optimal. A judicious increase in TCN blocks can enhance the prediction accuracy, and the $RMSE$

is relatively low when the number of TCN blocks is two, as shown in Figure 11a. After determining the optimal number of TCN blocks, the number of hidden neurons in the BiGRU is then adjusted. When 20 hidden neurons are applied, the *RMSE* value of the model is low, as shown in Figure 11b. After several parameter-tuning iterations, it was found that the model achieves the best prediction performance with a maximum number of iterations of 150, a gradient threshold of 1, an initial learning rate of 0.01, a learning rate tuning factor of 0.001, a regularization parameter of 0.001, and 20 hidden layer nodes in the BiGRU layer. To ensure comparability between experiments, the same dataset, optimal network structure, and basic hyperparameters were used for this study. This consistent set up contributes to a more accurate assessment of model performance differences and enables effective model comparison and analysis. Through a series of adjustments and comparisons, the optimal model parameters are determined, which in turn identifies a better model to be applied.

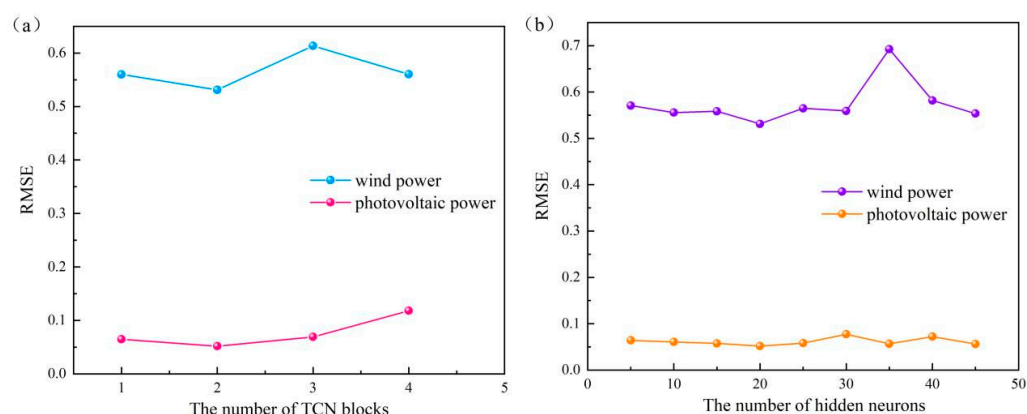


Figure 11. Determination of the number of TCN blocks and the number of hidden neurons. (a) Selection of the number of TCN blocks. (b) Selection of the number of hidden neurons.

From the above figure, it can be seen that when the same model is used to predict the power of wind power and photovoltaic power, the prediction error of the power of photovoltaic power is relatively small, which is mainly due to the fact that the data characteristics of photovoltaic power are more stable and regular. Photovoltaic power generation is mainly affected by solar irradiance, temperatures, and weather conditions, which are relatively stable and have obvious daily cyclical and seasonal characteristics, enabling the model to make more accurate predictions. In contrast, wind power generation is affected by a variety of factors, such as wind speed, wind direction, air temperatures, etc., which are more volatile and random, leading to highly nonlinear and uncertain changes in wind power generation power and increasing the difficulty of prediction. In addition, the physical processes of photovoltaic power generation are relatively simple; higher data quality and fewer missing values and noise make it easier for the model to fit in with the data characteristics of photovoltaic power generation. In contrast, data from wind power generation may have more missing values and noise, and the complex physical processes and high volatility further increase the difficulty of model prediction.

3.4. Results and Analysis

To select the optimal prediction model, the GRU, BiGRU, TCN-GRU, TCN-BiGRU, TCN-GRU-Attention, and TCN-BiGRU-Attention are compared and analyzed according to the evaluation index. For wind power generation, the prediction results are shown in Table 3. For photovoltaic power generation, the prediction results are shown in Table 4.

As can be seen in Table 3, for the wind power generation dataset, the TCN-BiGRU-Attention model is more effective than the GRU, BiGRU, TCN-GRU, TCN-BiGRU, and TCN-GRU-Attention models. The *MSE* values of the prediction set were reduced by 0.2201, 0.1739, 0.1750, 0.0687, and 0.0294, respectively; the *MAE* values were reduced by 0.1240,

0.1043, 0.1075, 0.0597, and 0.0258, respectively, and the R^2 values were improved by 0.0934, 0.0742, 0.0673, 0.0404, and 0.0140, respectively. The GRU, BiGRU, TCN-GRU, TCN-BiGRU, TCN-GRU-Attention, and TCN-BiGRU-Attention models were trained in 6.2422 s, 9.5600 s, 10.7462 s, 11.9641 s, 14.7634 s, and 16.5763 s, respectively.

Table 3. Modeling results of wind power data based on different algorithms.

Models	Training Set			Prediction Set		
	MSE _C	MAE _C	R ² _C	MSE _P	MAE _P	R ² _P
GRU	0.6431	0.5673	0.8314	0.5023	0.5112	0.8169
BiGRU	0.6060	0.5537	0.8422	0.4561	0.4915	0.8361
TCN-GRU	0.6052	0.5468	0.8424	0.4572	0.4947	0.8430
TCN-BiGRU	0.4203	0.4867	0.8954	0.3509	0.4469	0.8699
TCN-GRU-Attention	0.3954	0.4429	0.9049	0.3116	0.4130	0.8963
TCN-BiGRU-Attention	0.3660	0.4275	0.9145	0.2822	0.3872	0.9103

Table 4. Modeling results of photovoltaic data based on different algorithms.

Models	Training Set			Prediction Set		
	MSEC	MAEC	R2C	MSEP	MAEP	R ² _P
GRU	0.0104	0.0705	0.9733	0.0112	0.0823	0.9722
BiGRU	0.0047	0.0443	0.9883	0.0049	0.0514	0.9888
TCN-GRU	0.0034	0.0386	0.9915	0.0044	0.0492	0.9895
TCN-BiGRU	0.0033	0.0362	0.9918	0.0037	0.0443	0.9914
TCN-GRU-Attention	0.0035	0.0364	0.9913	0.0032	0.0395	0.9926
TCN-BiGRU-Attention	0.0030	0.0358	0.9924	0.0027	0.0368	0.9939

As can be seen in Table 4, for the photovoltaic power generation dataset, similarly, the TCN-BiGRU-Attention model works better than the GRU, BiGRU, TCN-GRU, TCN-BiGRU, and TCN-GRU-Attention models. The MSE values of the prediction set decreased by 0.0085, 0.0022, 0.0017, 0.0010, and 0.0005, respectively; the MAE values decreased by 0.0455, 0.0146, 0.0124, 0.0075, and 0.0027, respectively; and the R^2 values increased by 0.0217, 0.0051, 0.0044, 0.0025, and 0.0013, respectively. The GRU, BiGRU, TCN-GRU, TCN-BiGRU, TCN-GRU-Attention, and TCN-BiGRU-Attention models were trained in 6.3513 s, 9.3459 s, 11.6965 s, 14.9234 s, 16.2831 s, and 16.4604 s, respectively.

From the above, it can be seen that the training time of the TCN-BiGRU-Attention model is longer compared to the other models. Nevertheless, it achieves the best prediction performance; the model training can be performed offline, and the prediction time of around 16 s is acceptable.

The TCN-BiGRU-Attention model has an obvious advantage over the other models for wind power data, while its advantage is not obvious for photovoltaic power data. This is due to the difference in data characteristics and patterns. Wind power generation data typically have a strong time series dependency and complex seasonal and cyclical patterns. The TCN efficiently captures long-term dependencies through deep convolutional networks, while the BiGRU can handle bidirectional dependencies in time series. The attention mechanism can further focus on important points in time. Overall, TCN-BiGRU-Attention is better able to capture these complex patterns. In contrast, photovoltaic power generation data are affected by weather, temperatures, radiation, and other factors, and its data characteristics may be smoother and have relatively few seasonal fluctuations. Other models may have been able to capture the patterns of these data well enough that the advantages of TCN-BiGRU-Attention are less obvious.

For the wind power dataset, the resultant plot of true versus predicted values using multiple models is shown in Figure 12. It can be seen that the overall trend of the curves of the real and predicted values is consistent, and the points in time where the wind speed increases or decreases are congruent, which indicates that the model is able to predict the

main trend of wind power very well. And it can be intuitively seen that the predicted values of the proposed TCN-BiGRU-Attention model are the closest to the real values and have the best prediction effect.

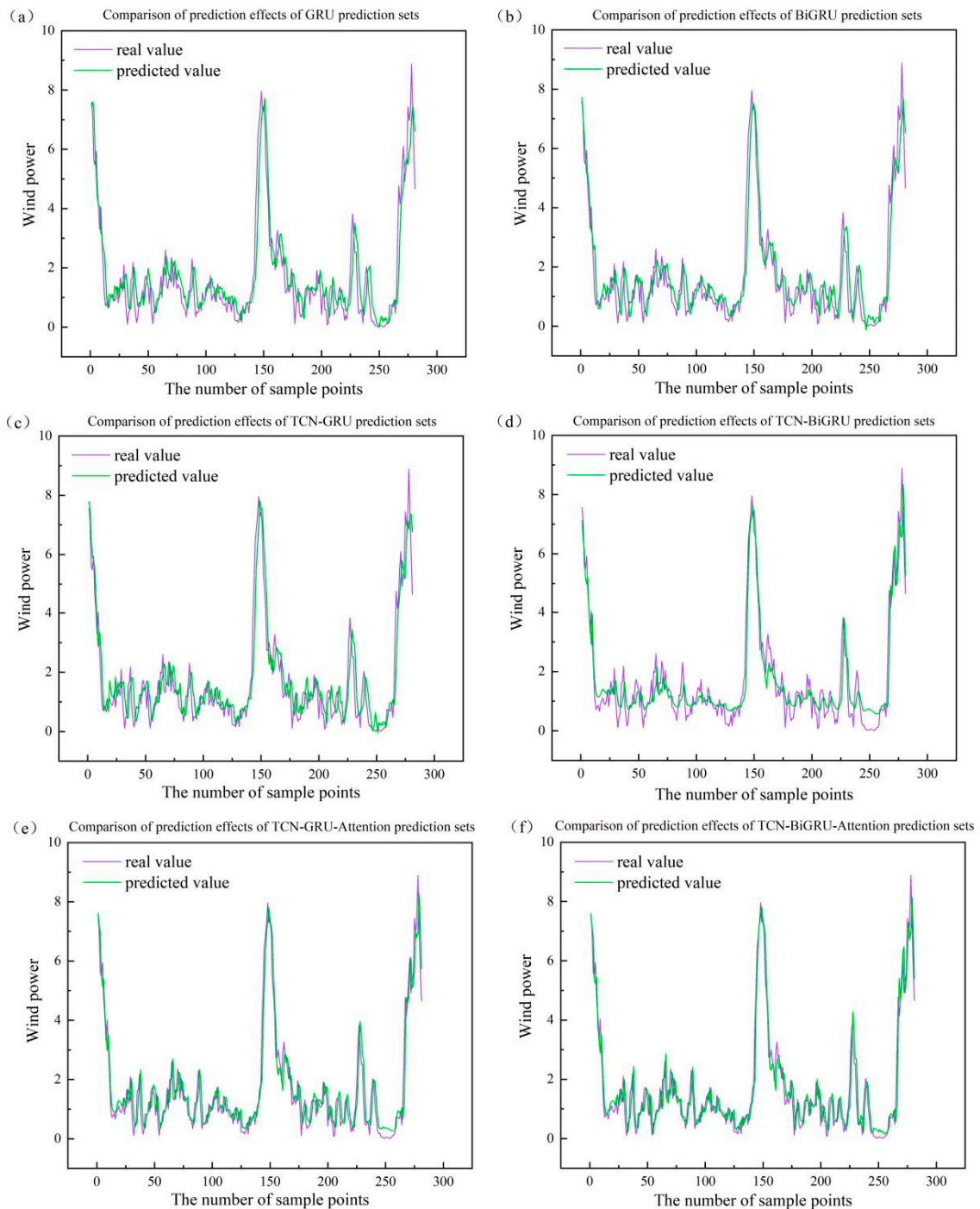


Figure 12. Comparison between real and predicted results of wind power generation. (a) The plot of true values versus predicted values output using the GRU algorithm. (b) The plot of true values versus predicted values output using the BiGRU algorithm. (c) The plot of true values versus predicted values output using the TCN-GRU algorithm. (d) The plot of true values versus predicted values output using the TCN-BiGRU algorithm. (e) The plot of true values versus predicted values output using the TCN-GRU-Attention algorithm. (f) The plot of true values versus predicted values output using the TCN-BiGRU-Attention algorithm.

The error results for wind power predicted using multiple models are plotted in Figure 13. The difference in the error curves shows that the GRU focuses only on past data, whereas the BiGRU is able to take into account future information as well, and the latter is better able to capture the bidirectional dependence of variables such as wind speed and direction at different points in time. The TCN captures long-time temporal dependencies and complex nonlinear relationships through convolution, which is more flexible than GRU processing sequences alone and therefore has fewer errors in capturing long-time wind power data patterns. The attention mechanism allows the model to focus on the input variables at key moments and therefore improves the model's performance at critical points in time, reduces errors, and performs well especially in the complex and variable environment of wind power data.

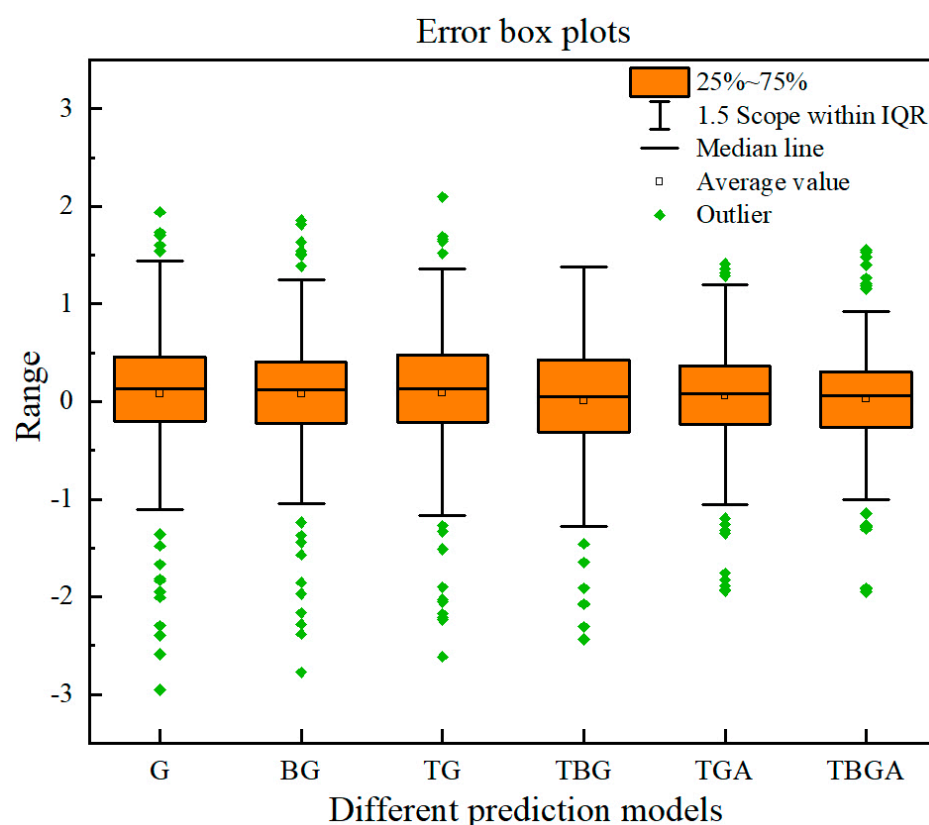


Figure 13. Graph of wind power error results, wherein G denotes the GRU model, BG denotes the BiGRU model, TG denotes the TCN-GRU model, TBG denotes the TCN-BiGRU model, TGA denotes the TCN-GRU-Attention model, and TBGA denotes the TCN-BiGRU-Attention model.

For the photovoltaic power dataset, the resultant plots of true versus predicted values using multiple models are shown in Figure 14. As can be seen, the photovoltaic power generation shows a clear daily cycle, i.e., the power generation increases during the day when the light is strong and decreases to zero at night. The proposed model is able to accurately capture such daily cycle variations and can effectively assess the model performance. The model predicts the true values relatively accurately in most cases and maintains a consistent predictive ability across conditions, which indicates the excellent robustness of the model.

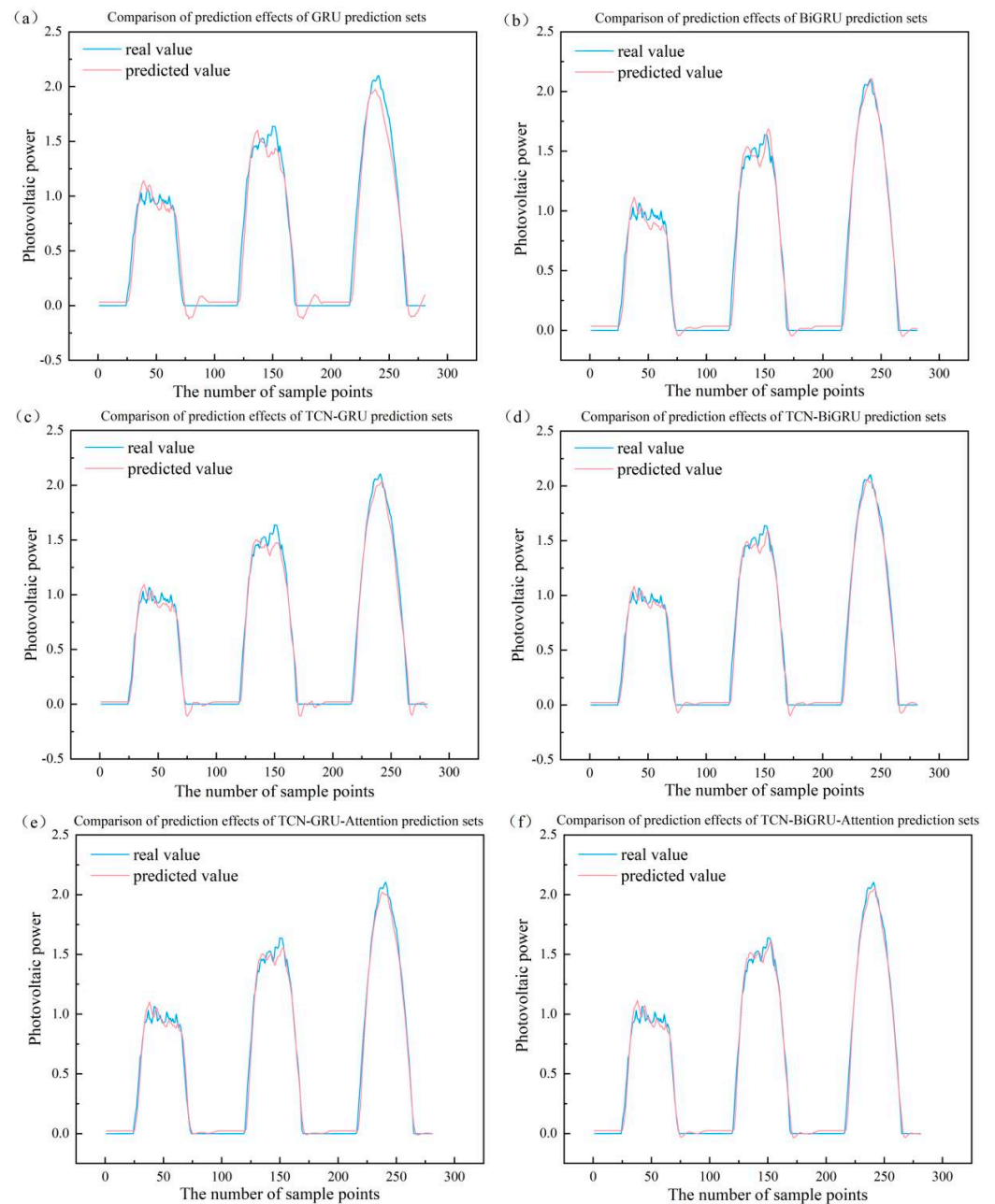


Figure 14. Comparison between real and predicted results of photovoltaic power generation. (a) The plot of true values versus predicted values output using the GRU algorithm. (b) The plot of true values versus predicted values output using the BiGRU algorithm. (c) The plot of true values versus predicted values output using the TCN-GRU algorithm. (d) The plot of true values versus predicted values output using the TCN-BiGRU algorithm. (e) The plot of true values versus predicted values output using the TCN-GRU-Attention algorithm. (f) The plot of true values versus predicted values output using the TCN-BiGRU-Attention algorithm.

The error results for photovoltaic power generation predicted using multiple models are shown in Figure 15. From the different error curves, it can be inferred that the TCN-BiGRU-Attention model tends to be able to combine the advantages of bidirectional temporal dependence, complex feature extraction, and important moments of attention. It performs best in dealing with time-varying and nonlinear features in the data and has the lowest prediction error.

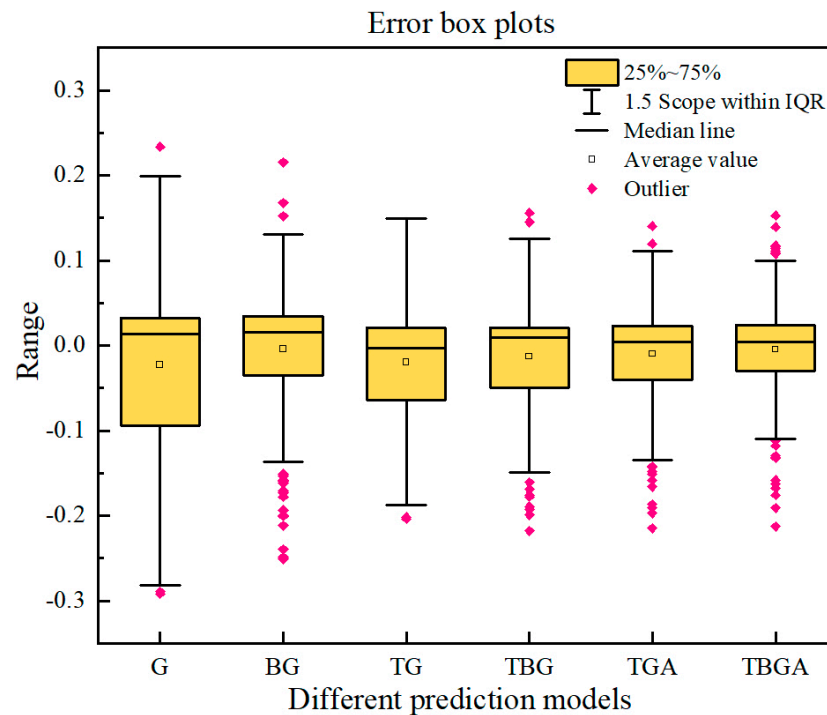


Figure 15. Graph of photovoltaic power error results, wherein G denotes the GRU model, BG denotes the BiGRU model, TG denotes the TCN-GRU model, TBG denotes the TCN-BiGRU model, TGA denotes the TCN-GRU-Attention model, and TBGA denotes the TCN-BiGRU-Attention model.

The above data processing and predictive modeling was carried out using MATLAB 2024a software (MathWorks, Natick, MA, USA) and plotted using Origin 2018 software. In addition, the computer CPU used in this study is Intel Core R7 5800U, 16 GB of RAM, the graphics card is MX450 and integrated graphics, the hard drive is 512G SSD, and the operating system is 64-bit, windows 11.

4. Discussion

From the analytical study in this paper, it can be concluded that the predictive performance of the TCN-BiGRU-Attention model proposed in this study is excellent. The integration of clustering analysis with deep learning not only improves the accuracy and interpretability of the predictions but also extends the application scope of the methodology. By pre-processing wind and photovoltaic power generation data through clustering analysis, the data heterogeneity and complexity were effectively reduced, resulting in more uniform input datasets. Principal Component Analysis, Agglomerative Hierarchical Clustering, and K-means clustering were adopted to determine an optimal data partitioning strategy, which enhanced model interpretability and training efficiency.

Based on clustering, the proposed TCN-BiGRU-Attention model demonstrated strong capabilities in capturing long-term time series dependencies, bidirectional temporal relationships, and key features through the attention mechanism. Experimental results confirmed the effectiveness of this model in predicting wind and photovoltaic power data, as evidenced by the following metrics:

For wind power dataset, $MSE = 0.2822$, $MAE = 0.3872$, $R^2 = 0.9103$.

For photovoltaic power dataset, $MSE = 0.0027$, $MAE = 0.0368$, $R^2 = 0.9939$.

These metrics highlight significant advantages over traditional GRU and other benchmark models, particularly in terms of accuracy and data fitting. However, the performance of the model is highly dependent on high-quality and sufficient historical data, which may lead to unsatisfactory predictions if the data are insufficient or noisy. Meanwhile, the hyperparameter tuning process is complex, time-consuming, and requires extensive experience.

Such improvements are of great importance for advancing smart grid technologies and facilitating the efficient integration of renewable energy sources.

Despite the promising results, several areas deserve further exploration. Future research could focus on optimizing model hyperparameters, integrating multiple prediction models, and incorporating external features to improve the stability and accuracy of predictions. Research on data preprocessing methods and real-time prediction systems can also enhance practical applications. Moreover, adopting advanced neural network architectures and reinforcement learning methods to create adaptive systems that continuously learn and optimize in dynamic environments could further elevate the model's predictive performance and robustness [38–40]. These directions will contribute to advancing the development of renewable energy forecasting in Virtual Power Plants (VPPs) and ensuring efficient grid management.

5. Conclusions

This study proposed and validated a method for predicting wind and photovoltaic power generation in VPPs based on clustering analysis and the TCN-BiGRU-Attention model. The main objectives were achieved, including improving forecasting accuracy, reducing data complexity, and demonstrating the applicability of the method in renewable energy scenarios.

Key findings and results are summarized as follows:

1. **Clustering Analysis:** By combining PCA, agglomerative hierarchical clustering, and K-means clustering, the data were effectively divided into homogeneous patterns, reducing data heterogeneity and enhancing model interpretability and training efficiency.
2. **Predictive Model:** The TCN-BiGRU-Attention model extracted long-term dependencies with TCN, captured bidirectional temporal relationships using BiGRU, and improved feature focus with the attention mechanism, leading to significant gains in prediction accuracy.
3. **Performance Metrics:** For wind power data, the model achieved $MSE = 0.2822$, $MAE = 0.3872$, and $R^2 = 0.9103$; for photovoltaic power data, $MSE = 0.0027$, $MAE = 0.0368$, and $R^2 = 0.9939$, surpassing traditional GRU and other benchmark models.
4. **Practical Applications:** The proposed method enhances the management efficiency and prediction accuracy of VPPs, improving their competitiveness in electricity markets while supporting the sustainable development of renewable energy.

Future research will focus on integrating adaptive learning mechanisms, expanding real-time and cross-domain prediction capabilities, and exploring hybrid approaches that combine clustering analysis with state-of-the-art neural network models. These efforts aim to further improve the prediction reliability and expand the practical applications of the method in renewable energy management.

Author Contributions: Conceptualization, Z.Q. and Y.T.; methodology, Z.Q. and Y.L.; software, Z.Q.; validation, Y.T., Y.L., T.G. and H.L.; data curation, Y.T.; writing—original draft preparation, Z.Q.; writing—review and editing, Z.Q.; visualization, Z.Q.; supervision, Y.L.; project administration, T.G. and H.L.; funding acquisition, Y.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially supported by the National Natural Science Foundation of China (62173082, U22B20115), and the Fundamental Research Funds for the Central Universities (N2304005).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

Acknowledgments: The authors would like to thank their research institutions and supervisors as well as the reviewers who provided valuable suggestions to this paper.

Conflicts of Interest: Author Zhichao Qiu and Yanhong Luo are from the College of Information Science and Engineering at Northeastern University., Ye Tian, Taiyu Gu, and Hengyu Liu were employed by the company Institute of Electric Power Science, State Grid Liaoning Electric Power Co. LTD. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Abbreviations

The Following Abbreviations are Used in this Manuscript

VPPs	Virtual power plants
TCNs	Time Convolutional Networks
GRU	Gated Recurrent Unit
BiGRU	Bi-directional Gated Recurrent Unit
PV	photovoltaic
PQ	power quality
P	the output power of the wind turbine
ρ	the air density
A	the area swept by the wind turbine
v	the wind speed
C_p	the power coefficient
p	the air pressure
T	the absolute temperature
R	the gas constant of air
e	the water vapor pressure
h	the relative humidity
e_s	the saturated water vapor pressure
r	the radius of the wind turbine
v_{cut-in}	the minimum starting wind speed
$v_{cut-out}$	the maximum cut-out wind speed
v_{rated}	the rated wind speed
P_{rated}	the rated power output of the turbine at the rated wind speed
P_{output}	the actual output power of the photovoltaic module
G_{inc}	the total radiation incident on the surface of the photovoltaic module
A_V	the area of the photovoltaic module
η_r	the module efficiency corrected for temperature
G_{total}	the total radiation from the sun
G_{beam}	the radiation directly from the sun
$G_{diffuse}$	the scattered radiation
T_{module}	the module temperature
$T_{ambient}$	the ambient temperature
ΔT	the temperature rise coefficient of the photovoltaic module
β	the temperature coefficient of the module
η_T	the efficiency of the photovoltaic module
η_{STC}	the conversion efficiency of the photovoltaic module under standard test conditions
X	the sample matrix
x_{ij}	the value of the j th dimension of the i th sample
$x_1, x_2, \dots, x_p$	the original variable indicators
$z_1, z_2, \dots, z_m$	the composite indicator variables after the dimensionality reduction process
l_{ij}	principal component loadings
s_j	the standard deviation of the j th column of the matrix X
R_X	the correlation coefficient matrix of the normalized matrix
α_i	the contribution rate of the i th principal component
β_i	the cumulative contribution rate
d_{ij}	the distance between the i th curve and the j th curve
$z_{i1}, z_{j1}, z_{i2}, z_{j2}, \dots, z_{im}, z_{jm}$	the corresponding principal component scores in the matrix Z

$\vec{h_t}$	the forward output of BiGRU
$\overleftarrow{h_t}$	the reverse output of BiGRU
y_t	the modeled output of the entire BiGRU
g_{ij}	the relationship between the state i and the state j
W_1, W_2	the weights
b	the offset vector
a_{ij}	the attentional weights from i to j
H_i	the final output state of the attentional layer
X'	the normalized value
x	the actual value of the sample
$x_{\max}, x_{\min}$	the maximum and minimum values of the column in which these data are located
MSE	the mean square error
$RMSE$	the root mean square error
MAE	the mean absolute error
R^2	the coefficient of determination
y_i	the actual value of the wind power sample
$\bar{y}_i$	the average of the actual values
$\hat{y}_i$	the predicted value of wind and photovoltaic power

References

- Navid, R.; Yasin, P.; Pir, R.-M. Optimal stochastic self-scheduling of a water-energy virtual power plant considering data clustering and multiple storage systems. *J. Energy Storage* **2023**, *65*, 107366. [\[CrossRef\]](#)
- Cao, W.; Wang, S.; Xu, M. Optimal Scheduling of Virtual Power Plant Based on Latin Hypercube Sampling and Improved CLARA Clustering Algorithm. *Processes* **2022**, *10*, 2414. [\[CrossRef\]](#)
- Aksan, F.; Jasiński, M.; Sikorski, T. Clustering methods for power quality measurements in virtual power plant. *Energies* **2021**, *14*, 5902. [\[CrossRef\]](#)
- Zhang, R.; Hredzak, B. Distributed dynamic clustering algorithm for formation of heterogeneous virtual power plants based on power requirements. *IEEE Trans. Smart Grid* **2020**, *12*, 192–204. [\[CrossRef\]](#)
- Yin, L.-F.; Meng, Y.-J. Short-term wind power forecasting based on DenseNet convolutional neural networks. *Integr. Intell. Energy* **2024**, *46*, 12–20. [\[CrossRef\]](#)
- Wang, W.-G.; Wei, Y.-B.; Teng, X.-D. Short-term wind power forecasting based on VMD-SSA-LSSVM. *Acta Energetica Solaris Sin.* **2023**, *44*, 204–211. [\[CrossRef\]](#)
- Wang, Y.; Fu, W.; Wang, J. Ultra-short-term distributed PV power forecasting for virtual power plant considering data-scarce scenarios. *Appl. Energy* **2024**, *373*, 123890. [\[CrossRef\]](#)
- Yu, D.; Choi, W.; Kim, M. Forecasting day-ahead hourly photovoltaic power generation using convolutional self-attention based long short-term memory. *Energies* **2020**, *13*, 4017. [\[CrossRef\]](#)
- Malik, H.; Alotaibi, M.-A.; Almutairi, A. A new hybrid model combining EMD and neural network for multi-step ahead load forecasting. *J. Intell. Fuzzy Syst.* **2022**, *42*, 1099–1114. [\[CrossRef\]](#)
- An, Y.; Dang, K.; Shi, X. A probabilistic ensemble prediction method for PV power in the nonstationary period. *Energies* **2021**, *14*, 859. [\[CrossRef\]](#)
- Zhu, J.; Zhao, Z.; Zheng, X.; An, Z.; Guo, Q.; Li, Z.; Sun, J.; Guo, Y. Time-series power forecasting for wind and solar energy based on the SL-transformer. *Energies* **2023**, *16*, 7610. [\[CrossRef\]](#)
- Soret, A.; Torralba, V.; Cortesi, N.; Christel, I.; Palma, L.; Manrique-Suñén, A.; Lledó, L.; González-Reviriego, N.; Doblas-Reyes, F.J. Sub-seasonal to seasonal climate predictions for wind energy forecasting. *J. Phys. Conf. Ser.* **2019**, *1222*, 012009. [\[CrossRef\]](#)
- Bhutta, M.S.; Li, Y.; Abubakar, M.; Almasoudi, F.M.; Alatawi, K.S.S.; Altimania, M.R.; Al-Barashi, M. Optimizing solar power efficiency in smart grids using hybrid machine learning models for accurate energy generation prediction. *Sci. Rep.* **2024**, *14*, 17101. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hossain, A.; Chakraborty, R.K.; Elsayah, S.; Ryan, M.J. Very short-term forecasting of wind power generation using hybrid deep learning model. *J. Clean. Prod.* **2021**, *296*, 126564. [\[CrossRef\]](#)
- Teixeira, R.; Cerveira, A.; Pires, E.J.S.; Baptista, J. Advancing renewable energy forecasting: A comprehensive review of renewable energy forecasting methods. *Energies* **2024**, *17*, 3480. [\[CrossRef\]](#)
- Venegas-Zarama, J.-F.; Muñoz-Hernandez, J.-I.; Baringo, L. A review of the evolution and main roles of virtual power plants as key stakeholders in power systems. *IEEE Access* **2022**, *10*, 47937–47964. [\[CrossRef\]](#)
- Yang, Y.; Wang, Y.; Wu, W. Allocating ex-post deviation cost of virtual power plants in distribution networks. *J. Mod. Power Syst. Clean Energy* **2022**, *11*, 1014–1019. [\[CrossRef\]](#)
- He, J.-H.; Lu, Y.-Z.; Lu, J.-Y. User Classification Method Based on ‘Evolution’ PCA and Its Application. *Electr. Power Constr.* **2017**, *38*, 101–107. [\[CrossRef\]](#)

19. Yang, L.; Li, C. Identification of vulnerable lines in smart grid systems based on improved agglomerative hierarchical clustering. *IEEE Access* **2023**, *11*, 13554–13563. [\[CrossRef\]](#)
20. Ai, X.; Yang, Z.-H.; Hu, Y.-Y. A Load Curve Clustering Method Based on Improved K-means Algorithm for Virtual Power Plant and Its Application. *Electr. Power Constr.* **2020**, *41*, 28–36. [\[CrossRef\]](#)
21. Li, H.; Zhao, L.; Peng, D. A Virtual Power Plant Load Curve Clustering Method Based on Improved K-means Algorithm and Its Application. *IOP Conf. Ser. Earth Environ. Sci.* **2020**, *619*, 012055. [\[CrossRef\]](#)
22. Sinaga, K.-P.; Yang, M.-S. Unsupervised K-means clustering algorithm. *IEEE Access* **2020**, *8*, 80716–80727. [\[CrossRef\]](#)
23. Yang, M.-S.; Hussain, I. Unsupervised multi-view K-means clustering algorithm. *IEEE Access* **2023**, *11*, 13574–13593. [\[CrossRef\]](#)
24. Yao, G.; Wu, Y.; Huang, X. Clustering of typical wind power scenarios based on K-means clustering algorithm and improved artificial bee colony algorithm. *IEEE Access* **2022**, *10*, 98752–98760. [\[CrossRef\]](#)
25. Bai, S.; Kolter, J.-Z.; Koltun, V. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling. *arXiv* **2018**, arXiv:1803.01271. [\[CrossRef\]](#)
26. Feng, Y.; Zhu, J.; Qiu, P. Short-term Power Load Forecasting Based on TCN-BiLSTM-Attention and Multi-feature Fusion. *Arab. J. Sci. Eng.* **2024**, 1–12. [\[CrossRef\]](#)
27. Li, L.; Li, Y.; Mao, R. Remaining useful life prediction for lithium-ion batteries with a hybrid model based on TCN-GRU-DNN and dual attention mechanism. *IEEE Trans. Transp. Electr.* **2023**, *9*, 4726–4740. [\[CrossRef\]](#)
28. Liu, M.; Qin, H.; Cao, R. Short-term load forecasting based on improved TCN and DenseNet. *IEEE Access* **2022**, *10*, 115945–115957. [\[CrossRef\]](#)
29. Li, Y.; Song, L.; Zhang, S. A TCN-based hybrid forecasting framework for hours-ahead utility-scale PV forecasting. *IEEE Trans. Smart Grid* **2023**, *14*, 4073–4085. [\[CrossRef\]](#)
30. Bi, S.; Wang, H. Network Traffic Prediction Based on the Multi-time Granularity GRU-BP Neural Network. *IEEE Access* **2024**, *12*, 96997–97003. [\[CrossRef\]](#)
31. Chen, J.; Zhang, J.-L. An attention-based TCN-BiGRU model for short-term wind power forecasting. *J. Tianjin Univ. Technol.* **2024**, 1–6. Available online: <https://link.cnki.net/urlid/12.1374.N.20240328.1409.057> (accessed on 3 November 2024).
32. Guo, Y.; Zhou, J.; Dong, Z. Rolling bearing RUL prediction based on Fusion of Multi-Head Attention and improved TCN-BiLSTM. *IEEE Access* **2024**, *12*, 95641–95658. [\[CrossRef\]](#)
33. Zhang, Y.; Ma, W.; Huang, F. A Novel Air Target Intention Recognition Method Based on Sample Reweighting and Attention-Bi-GRU. *IEEE Syst. J.* **2023**, *18*, 501–504. [\[CrossRef\]](#)
34. Vaswani, A. Attention is all you need. *arXiv* **2017**, arXiv:1706.03762. [\[CrossRef\]](#)
35. Aslam, M.; Lee, S.-J.; Khang, S.-H. Two-stage attention over LSTM with Bayesian optimization for day-ahead solar power forecasting. *IEEE Access* **2021**, *9*, 107387–107398. [\[CrossRef\]](#)
36. Hao, C.-L.; Zhang, J. Traffic flow prediction based on self-attention mechanism TCN-BiGRU. *Electr. Meas. Technol.* **2024**, *47*, 61–68. [\[CrossRef\]](#)
37. Guo, Y.; Li, Y.; Qiao, X. BiLSTM multitask learning-based combined load forecasting considering the loads coupling relationship for multienergy system. *IEEE Trans. Smart Grid* **2022**, *13*, 3481–3492. [\[CrossRef\]](#)
38. Al-Dahidi, S.; Madhiarasan, M.; Al-Ghussain, L. Forecasting Solar Photovoltaic Power Production: A Comprehensive Review and Innovative Data-Driven Modeling Framework. *Energies* **2024**, *17*, 4145. [\[CrossRef\]](#)
39. Yilmaz, M. Real measure of a transmission line data with load fore-cast model for the future. *Balk. J. Electr. Comput. Eng.* **2018**, *6*, 141–145. [\[CrossRef\]](#)
40. Wang, Y.; Wu, Z.; Ni, D. Large-Scale Optimization among Photovoltaic and Concentrated Solar Power Systems: A State-of-the-Art Review and Algorithm Analysis. *Energies* **2024**, *17*, 4323. [\[CrossRef\]](#)

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.