

Artificial Intelligence-Driven Ultra-Short-Term Photovoltaic Power Forecasting: A Comprehensive Review

Wenbo Sun^{1,a,*}

¹School of Electrical and Control Engineering, Beijing, China

a. 1627236022@qq.com

**corresponding author*

Abstract: Photovoltaic (PV) power prediction is an important tool for optimizing PV power generation efficiency and grid operation. PV power is affected by weather conditions, time series, and other complex factors, which requires accurate data processing and feature extraction. This paper firstly introduces data processing methods in PV power prediction, such as missing value processing, outlier detection and correction, and data normalisation, to ensure data integrity and improve model training efficiency. Secondly, the applications of several machine learning methods are discussed, including linear regression (LR), support vector machines, and random forests for handling nonlinear relationships and assessing feature importance. Finally, the advantages of deep learning methods in PV power prediction are analyzed. Convolutional neural networks are used for extracting spatial features. Long short-term memory networks (LSTM) and Bi-Directional LSTM Networks effectively handle time-series data and their global dependencies. This significantly improves prediction accuracy and stability. Finally, it concludes and looks at future research trends in Ultra-Short-Term Forecasting of Photovoltaic Power.

Keywords: photovoltaic power prediction, data preprocessing, feature engineering, machine learning, deep learning.

1. Introduction

With the growing global demand for clean energy, PV power generation, as an important renewable energy source, has gained widespread attention and application. PV power, which makes use of solar energy resources and converts solar radiation into electricity through photovoltaic panels[1], is one of the most important ways to achieve the goal of carbon neutrality[2]. The promotion of PV power generation not only helps to reduce greenhouse gas emissions, but also alleviates the environmental pollution problems caused by fossil fuels[3]. Therefore, promoting the development and optimization of PV technology is one of the important tasks in the current energy sector.

However, PV power generation is characterized by significant intermittency and fluctuation, which makes its application in the grid face many challenges. Since PV power generation depends on solar radiation, which is an uncontrollable natural factor, its output power may fluctuate greatly in different time periods. This volatility poses a serious threat to the stable operation of the power system, especially in the scheduling decision-making process, accurate power prediction is particularly important[4]. Ultra-short-term PV power prediction, as a key technology to solve this problem, can

accurately predict the power output of PV power generation in a short period of time, thus providing reliable support for the flexible scheduling and stable operation of the power system[5-7].

In order to improve the accuracy of photovoltaic power prediction, artificial intelligence technology has been gradually introduced into this field in recent years[8]. Traditional PV power prediction methods mainly rely on physical models and statistical methods. Although these methods can reflect the changing laws of PV power generation to a certain extent, they often struggle with complex and variable meteorological conditions. This is due to the complexity of the models and their dependence on the quality of input data. In contrast, artificial intelligence technology, especially deep learning algorithms, has powerful data processing capabilities and adaptivity. These technologies can automatically learn the complex relationship between PV power generation and its influencing factors based on large amounts of historical data. This significantly improves the accuracy and reliability of predictions. Therefore, this paper focuses on the research of ultra-short-term PV power prediction technology based on artificial intelligence and explores its application and development prospects in the field of photovoltaic power generation[9].

2. Data processing and feature engineering in PV power prediction

2.1. Data pre-processing methods

2.1.1. Missing data

Missing data are often encountered in PV power prediction datasets. Missing data may be caused by sensor failures, data transmission errors, or other unforeseen reasons. Commonly used missing data processing methods include linear interpolation, forward padding, backward padding, and so on. Linear interpolation is suitable for situations where the data changes smoothly, while forward padding and backward padding can maintain data consistency when the data changes drastically.

2.1.2. Outlier Detection and Correction

Outliers are usually caused by equipment faults, external disturbances and other factors, which may introduce errors in the prediction model and reduce the prediction accuracy of the model. Commonly used outlier detection methods include the 3σ principle. This method identifies possible outliers by calculating the standard deviation of the data. The box-and-whisker plot (box plot) method also detects outliers. It visualizes the location of an outlier by plotting the quartiles of the data.

2.1.3. Data Normalisation

As there may be significant differences in the magnitude of different data features, such as temperature (degrees Celsius) and solar radiation intensity (Watts/m²), these differences may affect the convergence and prediction accuracy of the model. Data normalisation can improve the efficiency of model training by scaling data with different features to a similar range. Commonly used normalisation methods include Min-Max normalisation, which scales the data to the [0,1] interval, and Z-score normalisation, which standardises the data based on mean and standard deviation.

2.2. Feature extraction methods

Historical power data is one of the important features in PV power prediction, and by analysing the trend of historical data, the model can effectively predict the future power output. Utilizing historical data not only helps capture daily cyclical changes but also identifies long-term seasonal patterns and anomalies. Additionally, when combined with weather forecasts and other environmental factors, historical data can enhance the model's generalization ability, allowing for more accurate predictions

under varying weather conditions. In-depth analysis of historical data can further optimize the operational efficiency of PV systems and improve the reliability and flexibility of grid management.

3. Machine Learning Approaches in PV Power Prediction

3.1. Linear Regression (LR)

LR is one of the most basic and widely used statistical learning methods. It assumes a linear relationship between the independent and dependent variables and fits the model parameters by least squares to minimise the sum of squares of errors between predicted and actual values. In PV power forecasting, linear regression can help to quickly build a model that provides an initial estimate of the underlying trend

Despite the theoretical simplicity and computational efficiency of linear regression methods, they exhibit certain limitations when dealing with complex PV power prediction problems. PV power generation is highly nonlinearly affected by weather conditions, and relying solely on linear models cannot effectively capture the complex effects of temperature, radiation intensity, and other variables on power generation[10]. Therefore, although linear regression can be used as a baseline model, it is often necessary to incorporate nonlinear methods or more complex models to improve prediction accuracy in practical applications.

3.2. Support Vector Machines (SVMs)

SVM is a supervised learning model originally used for binary classification problems. It maximises the spacing between classes by finding an optimal hyperplane. For regression tasks, SVMs adapt to noisy data by defining a "soft interval" and can handle nonlinear relationships by mapping the data to a high-dimensional space through a kernel function.

The advantage of SVM in PV power prediction lies in its ability to handle small-sample, high-dimensional datasets, especially excelling in cases where data features are complex and non-linear relationships are evident. For example, SVM can effectively handle weather data in PV power generation. SVM maps meteorological data from a low-dimensional space to a high-dimensional space using kernel functions, such as radial basis functions. This approach better captures the complex effects of factors like temperature, humidity, and light intensity on PV power. However, SVMs have high computational complexity and may require more computational resources and time especially when dealing with large-scale datasets.

3.3. Random Forest (RF)

RF is an integrated learning method based on decision trees. It improves the stability and accuracy of predictions by constructing multiple decision tree models and averaging (regression problem) or voting (classification problem) the predictions of these models. The key feature of Random Forest is its random sampling of training data and features for each tree, which reduces the risk of model overfitting.

In PV power prediction, RF is able to handle non-linear relationships between input variables and high-dimensional features, and is robust to noisy data. Another advantage of Random Forest is its built-in feature importance assessment mechanism, which identifies the meteorological factors that have the greatest impact on power prediction. The Random Forest model automatically handles missing values and is less susceptible to overfitting. It is especially suitable for complex PV power prediction problems with many input features. For example, when multiple meteorological variables like temperature, humidity, wind speed, and solar radiation are considered simultaneously[11].

4. Deep Learning Approaches in PV Power Prediction

4.1. Convolutional Neural Network (CNN)

CNN is an important model for deep learning, which is initially used in the field of image processing. CNN is capable of automatically extracting hierarchical features of the input data through the stacking of convolutional, pooling and fully connected layers. The convolutional layer effectively reduces the number of parameters of the model and improves the training efficiency through local perception and parameter sharing mechanism.

In recent years, CNNs have been applied to PV power prediction, especially in extracting spatial features from satellite images and meteorological data. CNNs can identify local patterns in meteorological data, such as the movement paths and distribution of clouds, through convolutional operations, and this information is crucial to accurately predict PV power. Despite its advantages in processing spatial data, CNNs have limited ability to process time-series data. Therefore, CNNs usually need to be used in conjunction with other models capable of handling time series (e.g., LSTM) to improve prediction accuracy.

4.2. Long and Short Term Memory Networks (LSTMs)

LSTMs are a type of Recurrent Neural Networks (RNNs) designed to solve the gradient vanishing and exploding problems that are common in traditional RNNs[12]. LSTMs are capable of controlling the flow of information through a neural network by introducing forgetting gates, input gates, and output gates, and thus excel in dealing with long time dependencies.

In PV power prediction, LSTM is particularly suitable for dealing with weather data and power data with long time dependency. PV power generation is usually closely related to weather conditions over the past few hours or even days, and LSTM can effectively capture such long-term dependencies[13]. For example, by using an LSTM model, future PV power generation can be predicted based on temperature, humidity, and light data over the past few hours. However, the training process of LSTM models is complex and usually requires a large amount of computational resources and data to ensure the accuracy and stability of the model.

4.3. Bidirectional Long and Short-Term Memory Network (BiLSTM)

The BiLSTM is developed based on LSTM. It considers both forward information from past to present and inverse information from present to past when processing time series data. Through bidirectional processing, BiLSTM is able to capture the global dependencies in the time series more comprehensively.

BiLSTM has a wide range of applications in PV power prediction, especially in scenarios where both historical data and future trends need to be considered. For example, when predicting PV power, BiLSTM can not only utilise historical meteorological data, but also integrate information about upcoming weather changes to provide more accurate and stable prediction results[14]. Although BiLSTM can significantly improve the accuracy of the prediction, it has a high demand on computational resources and usually requires hardware acceleration devices such as GPUs to increase the training speed.

5. Conclusion

In photovoltaic power prediction, data preprocessing and feature engineering are key aspects to ensure model accuracy. Data preprocessing techniques such as missing value processing, outlier detection and correction, and data normalization are widely adopted in existing studies. These

methods effectively improve data integrity and consistency. Meanwhile, feature engineering methods, such as historical power data analysis and meteorological data feature extraction, help the models capture key seasonal and trend variations. In addition, the application of various machine learning methods and deep learning methods to PV power prediction demonstrates their advantages in handling different types of data and capturing complex relationships.

However, existing studies still have some shortcomings. Firstly, most models have a high dependence on data quality and quantity, and missing and noisy data may significantly affect the prediction performance. In addition, many models often lack sufficient robustness in the face of diverse and dynamically changing meteorological conditions. Second, although deep learning methods excel in handling complex nonlinear relationships and large-scale datasets, their high computational cost and dependence on hardware resources limit their popularity in practical applications. In addition, the "black box" nature of deep learning models also makes them less interpretable, making it difficult to identify and understand the decision-making process and internal mechanisms of the models.

In the future, PV power prediction research can be further developed in the following directions. Firstly, data fusion and the utilisation of multi-source data should be enhanced by integrating more types of data e.g., satellite images, weather forecast data, etc. to improve the prediction accuracy and robustness of the models. Second, more lightweight and efficient deep learning models should be explored to reduce computational costs and improve the ability of models to be applied on edge devices. In addition, improving the interpretability of models is one of the important directions for future research, especially in practical application scenarios that can help understand and optimise the operation of PV power generation systems. Finally, considering the long-term uncertainty caused by climate change, researchers also need to develop more adaptive models to cope with possible future extreme weather events and long-term climate trend changes. Through these efforts, the practicality and prospects for widespread application of PV power prediction models will be further enhanced.

References

- [1] Y. Xu, J. Pei, J. Yuan, and G. Zhao, "Concentrated solar power: technology, economy analysis, and policy implications in China, " *Environmental Science and Pollution Research*, vol. 29, pp. 1324-1337, 2022, doi: <https://doi.org/10.1007/s11356-021-15779-1>.
- [2] D. Gielen, F. Boshell, D. Saygin, M. D. Bazilian, N. Wagner, and R. Gorini, "The role of renewable energy in the global energy transformation, " *Energy Strategy Reviews*, vol. 24, pp. 38-50, 2019/04/01/ 2019, doi: <https://doi.org/10.1016/j.esr.2019.01.006>.
- [3] M. Filonchyk et al., "Greenhouse gas emissions and reduction strategies for the world's largest greenhouse gas emitters, " *Science of The Total Environment*, vol. 944, p. 173895, 2024/09/20/ 2024, doi: <https://doi.org/10.1016/j.scitotenv.2024.173895>.
- [4] J. Shair, H. Li, J. Hu, and X. Xie, "Power system stability issues, classifications and research prospects in the context of high-penetration of renewables and power electronics, " *Renewable and Sustainable Energy Reviews*, vol. 145, p. 111111, 2021/07/01/ 2021, doi: <https://doi.org/10.1016/j.rser.2021.111111>.
- [5] S. Ghosh and R. Yadav, "Future of photovoltaic technologies: A comprehensive review, " *Sustainable Energy Technologies and Assessments*, vol. 47, p. 101410, 2021/10/01/ 2021, doi: <https://doi.org/10.1016/j.seta.2021.101410>.
- [6] T. Qiu et al., "Potential assessment of photovoltaic power generation in China, " *Renewable and Sustainable Energy Reviews*, vol. 154, p. 111900, 2022/02/01/ 2022, doi: <https://doi.org/10.1016/j.rser.2021.111900>.
- [7] C. Xiong, Y. Su, H. Wang, D. Zhang, and B. Xiong, "Optimal distributed energy scheduling for port microgrid system considering the coupling of renewable energy and demand, " *Sustainable Energy, Grids and Networks*, vol. 39, p. 101506, 2024/09/01/ 2024, doi: <https://doi.org/10.1016/j.segan.2024.101506>.
- [8] X. Wang et al., "Characterization and mapping of photovoltaic solar power plants by Landsat imagery and random forest: A case study in Gansu Province, China, " *Journal of Cleaner Production*, vol. 417, p. 138015, 2023/09/10/ 2023, doi: <https://doi.org/10.1016/j.jclepro.2023.138015>.

- [9] Q. Lin, H. Cai, H. Liu, X. Li, and H. Xiao, "A novel ultra-short-term wind power prediction model jointly driven by multiple algorithm optimization and adaptive selection, " *Energy*, vol. 288, p. 129724, 2024/02/01/ 2024, doi: <https://doi.org/10.1016/j.energy.2023.129724>.
- [10] Z. Yang et al., "Investigation of current density and temperature distribution characteristics under transient conditions based on a 108 cm² segmented proton exchange membrane fuel cell, " *International Journal of Hydrogen Energy*, vol. 80, pp. 125-136, 2024/08/28/ 2024, doi: <https://doi.org/10.1016/j.ijhydene.2024.07.043>.
- [11] M. Hostein, M. Musy, B. Moujalled, and M. El Mankibi, "Generating meteorological files of future climates with heatwaves in urban context to evaluate building overheating: An energy-efficient dwelling case study, " *Building and Environment*, vol. 263, p. 111874, 2024/09/01/ 2024, doi: <https://doi.org/10.1016/j.buildenv.2024.111874>.
- [12] A. Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network, " *Physica D: Nonlinear Phenomena*, vol. 404, p. 132306, 2020/03/01/ 2020, doi: <https://doi.org/10.1016/j.physd.2019.132306>.
- [13] E. Sarmas, E. Spiliotis, E. Stamatopoulos, V. Marinakis, and H. Doukas, "Short-term photovoltaic power forecasting using meta-learning and numerical weather prediction independent Long Short-Term Memory models, " *Renewable Energy*, vol. 216, p. 118997, Nov. 2023, doi: [10.1016/j.renene.2023.118997](https://doi.org/10.1016/j.renene.2023.118997).
- [14] Z. Hu, Y. Gao, S. Ji, M. Mae, and T. Imaizumi, "Improved multistep ahead photovoltaic power prediction model based on LSTM and self-attention with weather forecast data, " *Applied Energy*, vol. 359, p. 122709, Apr. 2024, doi: [10.1016/j.apenergy.2024.122709](https://doi.org/10.1016/j.apenergy.2024.122709).