Optimizing Photovoltaic Power Forecasting Through Machine Learning Algorithms

Senye Zhang
Electrical and Computer Engineering
Lakehead University
Thunder Bay, Canada
ORCID:0009-0005-0121-2771

***Abstract*—This paper focuses on optimizing real-time PV power forecasting to support efficient grid dispatch. By leveraging local weather data, including solar irradiance and temperature, along with historical PV power output data, machine learning models such as Long Short-Term Memory (LSTM) and Random Forest were employed to improve the accuracy of PV generation predictions. The results indicate that these models effectively reduce prediction errors, enhancing the reliability of PV power forecasts. This research provides a foundation for optimizing grid dispatch and addressing the complexities of integrating renewable energy into the power grid.**

***Index Terms*—photovoltaic power forecasting, machine learning, solar energy prediction, renewable energy, power grid scheduling optimization, LSTM, random forest, energy management, stochastic optimization, time series forecasting**

I. INTRODUCTION

The increasing deployment of photovoltaic (PV) systems in Yantai, Shandong Province, has significantly reshaped the region's energy landscape. With the rapid expansion of PV capacity, the traditional peak electricity demand patterns, once driven by cooling loads at midday, have shifted. Due to the large influx of solar power, the grid now faces new challenges, including midday power backfeed and negative load conditions in some areas. Accurate PV power forecasting is essential to ensure stable grid operation and optimize dispatch strategies. However, existing forecasting methods often struggle to capture the variability of PV generation under dynamic weather conditions. This paper addresses the gap by utilizing machine learning algorithms to predict real-time PV power output based on local meteorological data, aiming to improve grid stability and inform future dispatch strategies.

II. LITERATURE REVIEW

Current methods for photovoltaic (PV) power forecasting can be categorized into three main types: statistical, physical, and machine learning methods. Statistical methods, such as ARMA, ARIMA, and exponential smoothing, rely on historical time series data and are effective for short-term forecasting, but struggle with nonlinear relationships. Physical methods are based on environmental variables like solar irradiance and temperature to model PV output, offering insights into weather impacts, but often lack flexibility for dynamic environments and require precise data. Machine learning methods, including Support Vector Machines (SVM), Random Forest, Gradient Boosting, and Long Short-Term Memory (LSTM) networks, are widely used for capturing complex and nonlinear relationships in PV power generation, providing high accuracy but requiring substantial amounts of training data. A growing trend is the use of hybrid methods, which combine physical models with machine learning techniques to enhance both accuracy and adaptability, addressing the variability and complexity in PV power forecasting more effectively.

III. METHODOLOGY

Machine learning methods have become increasingly popular in the field of photovoltaic (PV) power forecasting due to their ability to handle large datasets and capture complex, nonlinear relationships between various input features such as weather conditions and power output. Models like Support Vector Machines (SVM), Random Forest, Gradient Boosting, and Long Short-Term Memory (LSTM) networks have shown great potential in accurately predicting PV power generation, especially in dynamic and uncertain environments. These methods can incorporate diverse factors such as solar irradiance, temperature, and time-series data, making them more adaptive to changing conditions compared to traditional methods. However, there are still several limitations in the current research. First, machine learning models often require vast amounts of high-quality historical data for training, which may not always be available. Second, while these models can perform well in controlled environments, their accuracy may decline when faced with extreme weather conditions or in regions with insufficient data. Additionally, many machine learning models act as "black boxes," making it difficult to interpret how predictions are made, which can be problematic for energy operators who require transparency and reliability. Lastly, the computational complexity of training advanced machine learning models can be high, potentially limiting their practical deployment in real-time grid operations.

A. Data Collection and Preprocessing

B. Machine Learning Models

-
-

Identify applicable funding agency here. If none, delete this.

C. Model Evaluation Metrics

$$a + b = \gamma \tag{1}$$

Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

TABLE I
TABLE TYPE STYLES

Table Head	Table Column Head		
	Table column subhead	***Subhead***	***Subhead***
copy	More table copy[a]		

[a]Sample of a Table footnote.

Fig. 1. Example of a figure caption.

IV. RESULTS AND DISCUSSION

A. Forecasting Accuracy Comparison

B. Impact of Weather Variables on Prediction

C. Implications for Power Grid Scheduling Optimization

V. CONCLUSION

VI. FUTURE WORK

ACKNOWLEDGMENT

REFERENCES

[1] G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955.

[2] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.