A Study on Novel Solar Power Forecasting Using an XGB-LiGBM-RF Hybrid Model and the L-BFGS-B Optimization Algorithm

Tuan Anh Nguyen
Electric Power University, Hanoi, 100000, Vietnam | Ministry of Industry and Trade, Hanoi, 100000, Vietnam
anhnt88@epu.edu.vn

Manh Hai Pham
Electric Power University, Hanoi, 100000, Vietnam | Ministry of Industry and Trade, Hanoi, 100000, Vietnam
haipm@epu.edu.vn (corresponding author)

Minh Phap Vu
Institute of Science and Technology for Energy and Environment, Vietnam Academy of Science and Technology, Hanoi, 100000, Vietnam | Electric Power University, Hanoi, 100000, Vietnam | Ministry of Industry and Trade, Hanoi, 100000, Vietnam
vuminhphap@istee.vast.vn

Ngoc Trung Nguyen
Electric Power University, Hanoi, 100000, Vietnam | Ministry of Industry and Trade, Hanoi, 100000, Vietnam
trungnn@epu.edu.vn

Dang Toan Nguyen
Electric Power University, Hanoi, 100000, Vietnam | Ministry of Industry and Trade, Hanoi, 100000, Vietnam
toandn@epu.edu.vn

Thi Anh Tho Vu
Electric Power University, Hanoi, 100000, Vietnam | Ministry of Industry and Trade, Hanoi, 100000, Vietnam
thovta@epu.edu.vn

Trong Tuan Tran
National Power System and Market Operator Company (NSMO), Hanoi, 100000, Vietnam
tuanttcs@gmail.com

Anh Tuan Do
Faculty of Electrical and Electronic Engineering Technology, Dai Nam University, Hanoi, 100000, Vietnam | A Chau Industrial Technology Joint Stock Company, Hanoi, 100000, Vietnam
doanhtuan@dainam.edu.vn

Received: 5 April 2025 | Revised: 7 May 2025 | Accepted: 17 May 2025

 DOI: https://doi.org/10.48084/etasr.11308

ABSTRACT

Accurate forecasting of solar power is essential for enhancing the stability and efficiency of power systems with high Photovoltaic (PV) penetration. This paper proposes a novel hybrid model based on a Stacking Ensemble (SE) of XGBoost, LightGBM, and Random Forest (RF), with optimal weights determined using the Limited Memory Broyden–Fletcher Goldfarb Shanno with Box constraints (L-BFGS-B) algorithm. The model is trained and tested on real-world data from a 49.5 MW solar power plant in Vietnam. The experimental results show that the proposed SE-XGB-LGBM-RF-OW model outperforms individual learners and deep learning baselines in both accuracy and training time. It consistently achieves a Normalized Mean Absolute Percentage Error (NMAPE) below 1.2% across all seasons. Compared to LSTM and GRU models, SE reduces Root Mean Square Error (RMSE) by more than 90% and shortens training time by over 20 times. Additionally, it significantly lowers the MAPE and NMAPE values, with improvements exceeding 90% in most seasonal test cases, highlighting the model's superior accuracy and generalization capability.

Keywords-forecasting; solar power; stacking ensemble model; XGBoost; LightGBM; random forest

I. INTRODUCTION

PV technology development has significantly contributed to clean energy production. However, PV power output is inherently volatile due to its dependence on meteorological factors, such as solar radiation, temperature, humidity, and cloud cover [1, 2] This intermittency poses challenges for grid operation, highlighting the importance of accurate solar power forecasting for planning, dispatching, and maintaining grid stability [3, 4].

Traditional methods, including statistical and shallow machine learning models, have been applied to solar forecasting. Among them, decision tree-based models, such as RF, XGBoost, and LightGBM, have shown strong predictive capabilities, computational efficiency, and robustness against noisy or incomplete data [5, 6]. These models outperform Artificial Neural Networks (ANNs) and deep learning models, such as LSTM and GRU, in many solar forecasting tasks [7, 8].

Recent studies have increasingly turned to hybrid and ensemble models to take advantage of the strengths of multiple algorithms. Among these, SE methods have stood out for their ability to combine several base models and improve forecasting accuracy [9]. However, many existing SE models still use equal or fixed weights when merging base learners, which limits their flexibility under changing conditions [10, 11]. At the same time, optimization-based methods have shown strong potential for improving model performance by fine-tuning key parameters—examples include the use of Dandelion Optimization in fuel cell modeling [12] and improved arithmetic optimization techniques for PV systems [13].

Building on this progress, the present study proposes a novel SE model, SE-XGB-LGBM-RF-OW, which combines XGBoost, LightGBM, and RF as base learners. Unlike traditional ensembles that use static weights, the proposed model leverages L-BFGS-B [14] to optimize the contribution of each base learner. This adaptive approach not only improves forecasting accuracy, but also enhances the model's resilience to seasonal changes. The proposed method is further validated using real-world data from a 49.5 MW solar power plant in Vietnam.

The main contributions of this study are the introduction of the SE-XGB-LGBM-RF-OW model with dynamic weight optimization through the L-BFGS-B algorithm, and the demonstration of its superior forecasting performance compared to state-of-the-art deep learning methods, such as LSTM and GRU.

II. METHODOLOGY

This study proposes a hybrid solar power forecasting model based on a SE of three decision tree-based algorithms: XGBoost, LightGBM, and RF. The model is trained utilizing historical meteorological and PV output data collected from a 49.5 MW solar power plant located in Central Vietnam. The dataset spans one full year (from 01/01/2022 to 31/12/2022). To evaluate forecasting performance under various weather conditions, several representative days were excluded from the training set and used for testing. These days are characterized by high or low power and irradiance variability across seasons as summarized in Table I.

TABLE I. TYPICAL DAYS IN A SEASON

Typical day	Day with the largest power fluctuation	Day with the smallest power fluctuation	Day with the largest radiation fluctuation	Day with the smallest radiation fluctuation
Spring	06/04/2022	04/03/2022	26/03/2022	05/03/2022
Summer	28/08/2022	08/08/2022	30/08/2022	08/08/2022
Autumn	18/09/2022	12/10/2022	18/09/2022	12/10/2022
Winter	18/02/2022	18/12/2022	07/01/2022	18/12/2022

This selection allows the model to be evaluated across both stable and highly variable operating conditions, providing a comprehensive assessment of its accuracy and robustness. Figure 1 displays the overall structure of the proposed SE-XGB-LGBM-RF-OW model, which is built around three main stages. First, input data go through a preprocessing phase, where relevant features are selected and prepared. Next, three different machine learning models XGBoost, LightGBM, and RF are trained separately to generate their own forecasts. Finally, these forecasts are combined using an optimization step based on the L-BFGS-B algorithm. The latter calculates the most suitable weight for each model, ensuring that they contribute optimally to the final prediction. By assigning dynamic weights rather than fixed ones, the SE model remains flexible and reliable, adapting well to seasonal changes and variations in data quality.

The SE architecture integrates three well-established tree-based regression models: XGBoost, LightGBM, and RF.

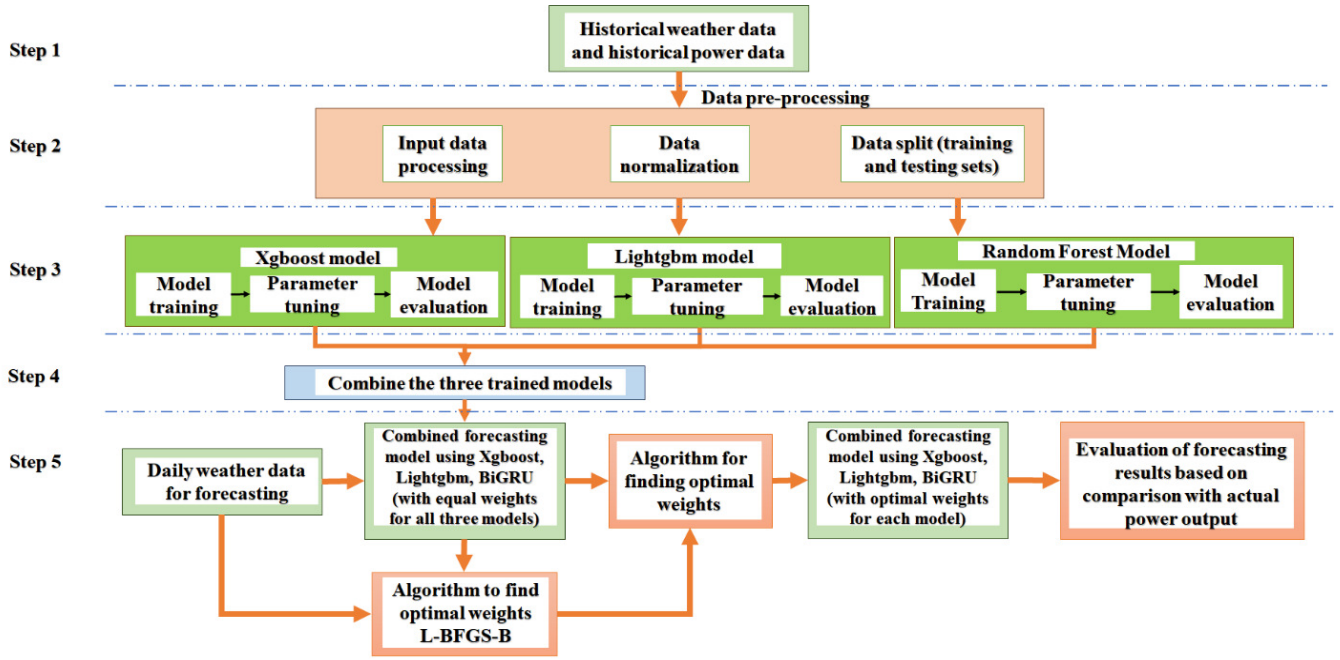

Fig. 1. Architecture of the SE-XGB-LGBM-RF-OW model, showing independent training of base learners and L-BFGS-B-based weight optimization before ensemble output.

A. Base Learners in the Stacking Ensemble

1) XGBoost

Utilizes gradient boosting with regularization and parallel processing, making it effective in handling complex nonlinear patterns [15].

2) LightGBM

Among the selected base learners, LightGBM stands out as a fast, histogram-based algorithm that builds leaf-wise trees and efficiently processes large-scale data [16].

3) Random Forest

RF constructs multiple decision trees using bootstrap aggregation (bagging), providing robust predictions through model averaging. These models were selected due to their strong generalization ability, fast training speed, and complementary strengths in regression tasks [17]. Furthermore, the use of tree-based models, such as XGBoost, LightGBM, and RF offers practical advantages in real world solar forecasting scenarios, particularly when historical data are incomplete or contain irregularities. Unlike time-series deep learning models, like LSTM and GRU, which require continuous and well-structured sequential data for effective learning, decision tree algorithms are inherently more tolerant of missing values and do not depend on strict temporal order. These properties render them well-suited for deployment in PV systems, where issues, such as sensor malfunctions and data transmission losses, frequently lead to gaps in the dataset.

B. Phases of the Proposed Method

The methodology consists of three main phases:

1) Preprocessing and Feature Engineering

All raw input data are cleaned, normalized, and aligned to ensure temporal consistency. Lag features and time-based attributes (hour, day of year) are also extracted to enhance forecasting quality.

2) Model Training and Ensemble Formation

The three base learners XGBoost, LightGBM, and RF are trained independently using 70% of the dataset. This data split strikes a balance between providing enough samples for effective model learning and retaining a sufficient test set for evaluating generalization performance. The hyperparameters for each model are optimized using the Optuna framework, with the objective of minimizing the forecasting error on the validation data, and are depicted in Table II.

3) Ensemble Weight Optimization

Instead of averaging model outputs equally, their contribution weights are optimized using the L-BFGS-B algorithm. This constrained optimization ensures that the sum of the weights equals 1, while each individual weight remains between zero and one. The final forecast is computed as a weighted sum:

$$y_{pred} = w_1 \times y_{xgb} + w_2 \times y_{lgbm} + w_3 \times y_{rf} \quad (1)$$

subject to $w_1 + w_2 + w_3 = 1$ and $0 \leq w_i \leq 1$.

C. Forecast Accuracy Assessment

1) Root Mean Square Error

RMSE measures the average squared difference between predicted and actual values, highlighting large deviations [18]:

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

2) Normalized RMSE (NRMSE)

The NRMSE expresses the RMSE as a percentage of the plant's nominal capacity, allowing for fair performance comparisons across PV systems of different sizes and scales [17]:

$$\text{NRMSE} = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}}{y_{nominal}} \times 100 \quad (3)$$

3) Mean Absolute Percentage Error (MAPE)

The MAPE represents the average absolute error as a percentage of the actual output, providing a scale-independent measure of forecasting accuracy [19]:

$$\text{MAPE} = \frac{100}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{y_i} \quad (4)$$

4) Normalized MAPE (NMAPE)/ Normalized Mean Absolute Percentage Error (NMAPE)

$$\text{NMAPE} = \frac{100}{n}\sum_{i=1}^{n}\frac{|\hat{y}_i - y_i|}{y_{nominal}} \quad (5)$$

All metrics were computed over test datasets using standard formulas, where $\hat{y}_i$ is the predicted power (kW), y_i is the actual power (kW), n is the number of data points, and $y_{nominal}$ is the total installed capacity of the plant (kW).

TABLE II. OPTIMAL HYPERPARAMETERS FOR EACH BASE LEARNER AFTER 100 OPTUNA OPTIMIZATION TRIALS (BASED ON VALIDATION RMSE)

Hyperparameter	**XGBoost**	**LightGBM**	**RF**
Learning rate	0.1	0.05	-
Maximum depth	6	7	20
Number of estimators	300	-	150
Minimum child weight	4	-	-
Gamma	0.2	-	-
Subsample	0.8	-	-
Colsample bytree	0.7	0.8	-
Number of leaves	-	90	-
Minimum child samples	-	20	-
Reg alpha	-	0.2	-
Reg lambda	-	0.5	-
Minimum samples slit	-	-	4
Minimum samples lead	-	-	2

III. RESULTS AND DISCUSSION

A. Results

To evaluate the performance of the individual models before assembling, each base learner was trained and tested independently. The results of these models on both the training and test datasets are shown in Table III. The XGBoost, LightGBM, and RF models consistently outperform LSTM and GRU across all seasons in terms of RMSE and NRMSE. Among them, LightGBM demonstrates the best overall performance, achieving the lowest error metrics in most cases. Furthermore, the tree-based models require significantly less training time, being 20 to 40 times faster than deep learning models. In contrast, LSTM and GRU exhibit higher error rates and longer computation times, indicating that the tree-based approaches are more effective for this task on the test set.

To obtain the optimal weight vector for combining the base models, the L-BFGS-B optimization algorithm was employed. This algorithm efficiently minimizes the ensemble's loss function while respecting the bound constraints, thereby determining the optimal set of weights that yields the best overall prediction accuracy. The results of finding the optimal weights for each model are presented in Table IV. After obtaining the optimal weight vector using the L-BFGS-B algorithm, the SE model was executed to forecast the PV output on the previously selected typical days. These days, which represent seasonal characteristics, were excluded from the training process and serve as independent test samples to evaluate the model's generalization ability under varying weather conditions. Table V provides a comparative evaluation of the proposed SE-XGB-LGBM-RF-OW model against two deep learning models, LSTM and GRU, across four seasonal typical days. The metrics include RMSE, NRMSE, MAPE, and NMAPE, which offer a comprehensive view of forecasting accuracy. As observed, the SE model consistently outperforms LSTM and GRU in all seasons, achieving significantly lower error rates and better generalization capability.

The results in Table VI demonstrate that the proposed SE-XGB-LGBM-RF-OW model outperforms the LSTM and GRU models across all seasons. It achieves significantly lower RMSE, NRMSE, MAPE, and NMAPE values, indicating higher forecasting accuracy and robustness. In contrast, LSTM and GRU models exhibit large errors, particularly in MAPE, which consistently exceeds 99%, reflecting limited generalization capability on seasonal test data.

To better illustrate the effectiveness of the proposed SE-XGB-LGBM-RF-OW model, Figure 2 presents the predicted and actual PV power output curves for one representative day from each season. These specific days were selected based on their high power output variability, representing some of the most challenging forecasting scenarios due to rapidly changing weather conditions. None of these days were included in the training data, allowing for an unbiased evaluation of the model's generalization ability. Despite the complexity of these conditions, the model consistently tracks the actual power output with high accuracy, capturing both gradual trends and sudden fluctuations throughout the day. This strong agreement between the predicted and actual values highlights the robustness of the proposed model under real-world variability. It also reinforces its practical value for short-term solar forecasting, especially in operational settings where accurate predictions are needed to support grid stability and decision-making.

TABLE III. COMPARISON OF TRAINING RESULTS FOR EACH MODEL WITH THE HISTORICAL DATASET DIVIDED BY SEASON

Seasons	Comparisons	XGBoost model	LightGBM model	RF model	LSTM	GRU
Spring	Completion time (s)	0.768	0.311	5.191	151.015	109.434
	RMSE (kW)	1085.779	1012.222	1042.355	1199.625	1179.281
	NRMSE (%)	2.193	2.044	2.105	2.423	2.382
	MAPE (%)	74.068	214.691	41.438	18285.84	5338.227
	NMAPE (%)	0.812	0.766	0.785	1.232	1.013
Summer	Completion time (s)	0.781	0.318	5.179	148.341	104.264
	RMSE (kW)	1251.062	1196.135	1241.774	1396.369	1395.076
	NRMSE (%)	2.527	2.416	2.508	2.821	2.818
	MAPE (%)	246.264	53.964	23.587	1527.623	1360.625
	NMAPE (%)	1.103	1.068	1.100	1.458	1.405
Autumn	Completion time (s)	0.745	0.324	4.794	147.146	99.923
	RMSE (kW)	2177.740	2095.553	2154.299	2305.123	2312.86
	NRMSE (%)	4.399	4.233	4.352	4.656	4.672
	MAPE	46.468	33.376	29.802	222.647	224.971
	NMAPE	1.520	1.466	1.485	1.821	1.854
Winter	Completion time (s)	0.695	0.333	4.356	140.802	103.623
	RMSE (kW)	1264.509	1194.748	1209.349	1308.065	1318.722
	NRMSE (%)	2.554	2.417	2.443	2.642	2.664
	MAPE (%)	94.415	120.76	14.323	2017.471	280.037
	NMAPE (%)	0.752	0.733	0.711	0.865	0.873

TABLE IV. OPTIMAL WEIGHTS FOR EACH MODEL

Seasons	XGBoost model	LightGBM model	RF model
Spring	0.14	0.36	0.5
Summer	-0.45	1.05	0.4
Autumn	-0.13	0.16	0.97
Winter	-0.35	0.85	0.5

TABLE V. PREDICTION RESULTS FOR TYPICAL DAYS OF A SEASON WITH OPTIMAL WEIGHTS FOR THE SE-XGB-LIGBM-RF-OW MODEL

Power		Pmax (kW)	Pmin (kW)	Paverage (kW)	RMSE (kW)	NRMSE (%)	MAPE (%)	NMAPE (%)
Spring	Actual	41541.56	0	10554.4	1183.33	2.390	8.572	1.112
	Predicted	40841.49	1.673	10363.85				
	$\Delta P_{Actual\text{-}Predicted}$	700.07	1.673	190.55				
Summer	Actual	40287.79	0	8556.395	939.332	1.897	12.616	0.744
	Predicted	39869.53	0	8684.206				
	$\Delta P_{Actual\text{-}Predicted}$	418.26	0	127.811				
Autumn	Actual	40710.07	0	6245.002	635.920	1.284	16.721	0.539
	Predicted	39194.77	0	6160.578				
	$\Delta P_{Actual\text{-}Predicted}$	1515.3	0	84.424				
Winter	Actual	39052.66	0	7034.479	975.106	1.969	17.797	0.874
	Predicted	38945.49	0	6858.854				
	$\Delta P_{Actual\text{-}Predicted}$	107.17	0	175.625				

TABLE VI. COMPARISON OF FORECASTING ACCURACY BETWEEN SE-XGB-LGBM-RF-OW AND DEEP LEARNING MODELS (LSTM, GRU) ACROSS SEASONS

Season	Model	RMSE (kW)	NRMSE (%)	MAPE (%)	NMAPE (%)
Spring	SE-XGB-LGBM-RF-OW	1183.33	2.39	8.572	1.112
Summer	SE-XGB-LGBM-RF-OW	939.33	1.897	12.616	0.744
Autumn	SE-XGB-LGBM-RF-OW	635.92	1.284	16.721	0.539
Winter	SE-XGB-LGBM-RF-OW	975.11	1.969	17.797	0.874
Spring	LSTM	17896.97	36.155	100.064	21.32
Summer	LSTM	17896.51	36.155	99.97	21.319
Autumn	LSTM	17897.43	31.156	99.85	21.32
Winter	LSTM	17897.66	36.157	99.68	21.32
Spring	GRU	17897.26	36.156	99.96	21.32
Summer	GRU	17895.85	36.153	100.35	21.319
Autumn	GRU	17896.51	31.155	99.7	21.32
Winter	GRU	17898.03	36.158	99.72	21.32

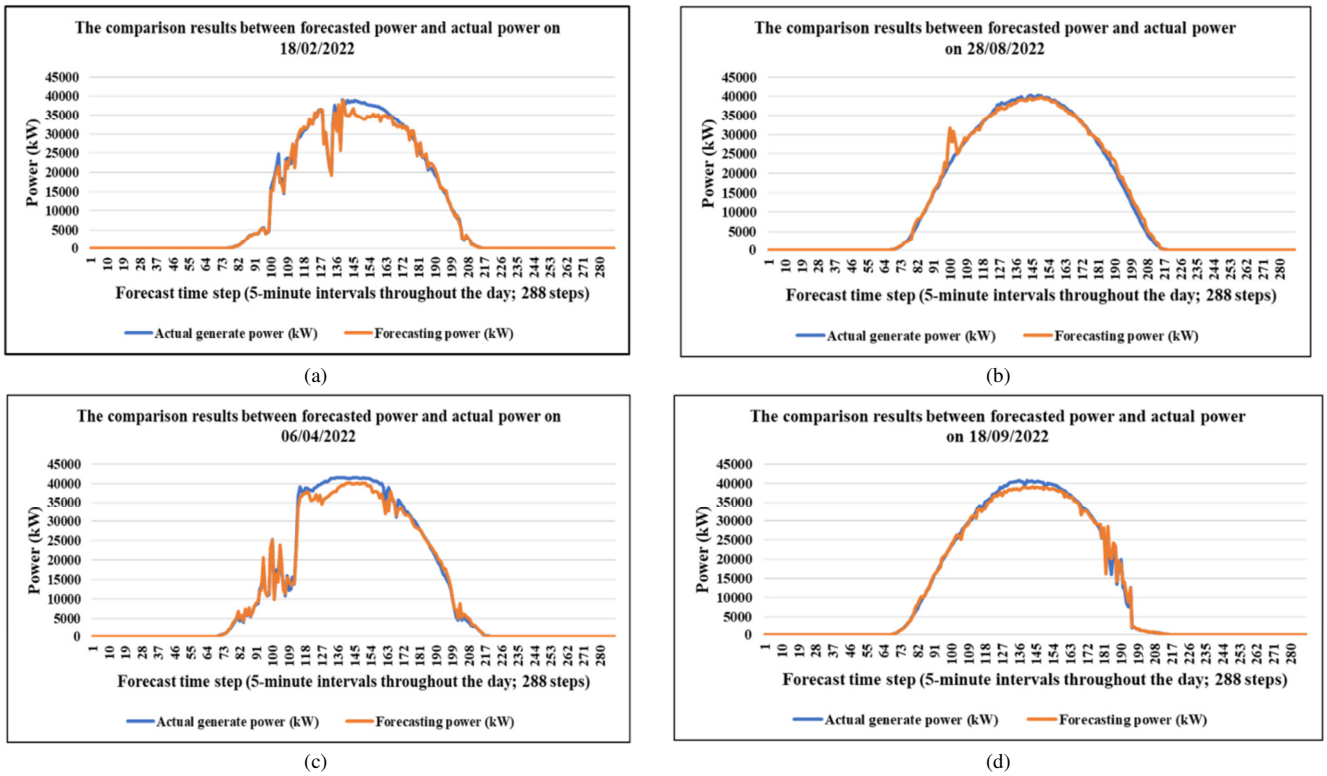

Fig. 2. Predicted versus actual PV power output on representative days from each season: (a) Spring – 06/04/2022, (b) Summer – 28/08/2022, (c) Autumn – 18/09/2022, and (d) Winter – 18/02/2022. The x-axis represents 288 forecast steps, corresponding to 5-minute intervals throughout a 24-hour period.

B. Discussion

The results demonstrate that the proposed SE-XGB-LGBM-RF-OW model achieves superior forecasting accuracy across all seasons, as reflected by the consistently lower RMSE, NRMSE, MAPE, and NMAPE values compared to the deep learning models. In addition, the ensemble model maintains stable performance regardless of seasonal variations, highlighting its robustness and adaptability. These advantages, combined with significantly reduced training time compared to LSTM and GRU, make the proposed approach a more reliable and efficient solution for short-term solar power forecasting. Specifically, the model reduces RMSE and NMAPE by more than 90% compared to LSTM and GRU, while maintaining NMAPE below 1.2% across all typical seasonal days.

IV. CONCLUSION

This study proposed a novel ensemble model, SE-XGB-LGBM-RF-OW, for short-term solar power forecasting. The results demonstrate that the proposed model achieves superior forecasting accuracy across all seasons, as demonstrated by the consistently lower Root Mean Square Error (RMSE), Normalized RMSE (NRMSE), Mean Absolute Percentage Error (MAPE), and Normalized Mean Absolute Percentage Error (NMAPE) values compared to the deep learning models. In addition, the ensemble model maintains stable performance regardless of seasonal variations, highlighting its robustness and adaptability.

A key advantage of the proposed model is its robustness to missing or inconsistent data, which frequently occurs in solar power plants due to sensor or communication issues. Unlike LSTM and GRU, which require complete sequential data, the tree-based structure of the ensemble allows for accurate forecasting even under data irregularities. This makes the model well-suited for real world applications, where data quality cannot always be ensured.

These advantages, combined with the significantly reduced training time compared to LSTM and GRU, make the proposed approach a more reliable and efficient solution for short-term solar power forecasting.

In particular, the proposed model achieved an average RMSE of less than 1200 kW and maintained NMAPE values below 1.2% across all test scenarios. It consistently outperformed LSTM and GRU models, with improvements exceeding 90% in most error metrics. The predicted power curves closely followed the actual output, even on highly volatile days, further demonstrating the model's ability to generalize under challenging conditions.

This level of accuracy meets the practical requirements for integration into real-time power system dispatch, especially in regions experiencing rapid expansion of PV capacity, such as Vietnam. Additionally, the proposed model offers tangible benefits for energy planners by reducing curtailment risks and enhancing confidence in short-term grid balancing decisions

REFERENCES

[1] K. Wang, X. Qi, and H. Liu, "A comparison of day-ahead photovoltaic power forecasting models based on deep learning neural network,"

Applied Energy, vol. 251, Oct. 2019, Art. no. 113315, https://doi.org/10.1016/j.apenergy.2019.113315.

[2] L. Liu *et al.*, "Prediction of short-term PV power output and uncertainty analysis," *Applied Energy*, vol. 228, pp. 700–711, Oct. 2018, https://doi.org/10.1016/j.apenergy.2018.06.112.

[3] M. Aslam, S.-J. Lee, S.-H. Khang, and S. Hong, "Two-Stage Attention Over LSTM With Bayesian Optimization for Day-Ahead Solar Power Forecasting," *IEEE Access*, vol. 9, pp. 107387–107398, Jul. 2021, https://doi.org/10.1109/ACCESS.2021.3100105.

[4] D. W. van der Meer, M. Shepero, A. Svensson, J. Widén, and J. Munkhammar, "Probabilistic forecasting of electricity consumption, photovoltaic power generation and net demand of an individual building using Gaussian Processes," *Applied Energy*, vol. 213, pp. 195–207, Mar. 2018, https://doi.org/10.1016/j.apenergy.2017.12.104.

[5] H. Sangrody, N. Zhou, and Z. Zhang, "Similarity-Based Models for Day-Ahead Solar PV Generation Forecasting," *IEEE Access*, vol. 8, pp. 104469–104478, Jun. 2020, https://doi.org/10.1109/ACCESS.2020. 2999903.

[6] J. Gaboitaolelwe, A. M. Zungeru, A. Yahya, C. K. Lebekwe, D. N. Vinod, and A. O. Salau, "Machine Learning Based Solar Photovoltaic Power Forecasting: A Review and Comparison," *IEEE Access*, vol. 11, pp. 40820–40845, Apr. 2023, https://doi.org/10.1109/ACCESS.2023. 3270041.

[7] .Wang, P. Li, R. Ran, Y. Che, and Y. Zhou, "A Short-Term Photovoltaic Power Prediction Model Based on the Gradient Boost Decision Tree," *Applied Sciences*, vol. 8, no. 5, May. 2018, Art. no. 689, https://doi.org/10.3390/app8050689.

[8] C. Saigustia and P. Pijarski, "Time Series Analysis and Forecasting of Solar Generation in Spain Using eXtreme Gradient Boosting: A Machine Learning Approach," *Energies*, vol. 16, no. 22, Jan. 2023, Art. no. 7618, https://doi.org/10.3390/en16227618.

[9] A. Abdellatif *et al.*, "Forecasting Photovoltaic Power Generation with a Stacking Ensemble Model," *Sustainability*, vol. 14, no. 17, Jan. 2022, Art. no 11083, https://doi.org/10.3390/su141711083.

[10] W. Khan, S. Walker, and W. Zeiler, "Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach," *Energy*, vol. 240, Feb. 2022, Art. no. 122812, https://doi.org/10.1016/j.energy.2021.122812.

[11] N. Fraccanabbia, R. G. da Silva, M. Henrique Dal Molin Ribeiro, S. R. Moreno, L. dos Santos Coelho, and V. C. Mariani, "Solar Power Forecasting Based on Ensemble Learning Methods," in *2020 International Joint Conference on Neural Networks (IJCNN).*, Glasgow, UK, 2020, pp. 1-7, https://doi.org/10.1109/IJCNN48605.2020.9206777.

[12] R. Abbassi, S. Saidi, A. Abbassi, H. Jerbi, M. Kchaou, and B. N. Alhasnawi, "Accurate Key Parameters Estimation of PEMFCs' Models Based on Dandelion Optimization Algorithm, " *Mathematics*, vol. 11, no. 6, Jan. 2023, Art. no. 1298, https://doi.org/10.3390/math11061298.

[13] A. Abbassi *et al.*, "Improved Arithmetic Optimization Algorithm for Parameters Extraction of Photovoltaic Solar Cell Single-Diode Model," *Arabian Journal for Science and Engineering*, vol. 47, no. 8, pp. 10435–10451, Aug. 2022, https://doi.org/10.1007/s13369-022-06605-y.

[14] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Transactions on Mathematical Software*, vol. 23, no. 4, pp. 550–560, Sep. 1997, https://doi.org/10.1145/279232.279236.

[15] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.*, New York, NY, USA, 2016, pp. 785–794, https://doi.org/10.1145/2939672.2939785.

[16] G. Ke *et al*, "LightGBM: a highly efficient gradient boosting decision tree," in *Proceedings of the 31st International Conference on Neural Information Processing Systems.*, Red Hook, NY, USA, 2017, pp. 3149–3157.

[17] L. Breiman, "Random Forests,"*Machine Learning*, vol. 45, no. 1, pp. 5–32, Jul. 2001, https://doi.org/10.1023/A:1010933404324.

[18] Y. Kassem, H. Camur, M. T. Adamu, T. Chikowero, and T. Apreala, "Prediction of Solar Irradiation in Africa using Linear-Nonlinear Hybrid Models," *Engineering, Technology & Applied Science Research*, vol. 13, no. 4, pp. 11472–11483, Aug. 2023, https://doi.org/10.48084/etasr.6131.

[19] *Standard Process of Renewable Energy Forecasting*, Decision 67/QĐ-ĐTĐL, Ministry of Industry and Trade (MOIT), Vietnam, 2020.