Deployment-Aware Evaluation of Short-Term PV Forecasting Models Using a Composite Efficiency-Accuracy Index for Smart Grids

Özge Çelik
İbn-i Sina Vocational and Technical Anatolian High School
Manisa, Türkiye
ozgephdcelik@gmail.com

Kadriye Filiz Balbal
Department of Computer Science
Dokuz Eylül University
İzmir, Türkiye
kadriyefiliz.balbal@deu.edu.tr

Sebahattin İkikardeş
Faculty of Arts and Sciences
Balıkesir University
Balıkesir, Türkiye
skardes@balikesir.edu.tr

***Abstract*— Conventional photovoltaic (PV) forecasting studies primarily evaluate model performance using global statistical accuracy metrics such as RMSE and MAE. However, practical smart-grid deployment also requires consideration of computational efficiency, execution latency, and operational robustness under highly variable weather conditions. Motivated by these limitations, this study proposes a deployment-aware evaluation framework based on a composite Efficiency Accuracy Index (EAI) for short-term PV forecasting systems. Experimental analyses were conducted using real-world hourly production data collected from a 100 kW rooftop PV power plant located in Manisa, Türkiye. Although the Vanilla Transformer architecture achieved the lowest global RMSE, the proposed EAI framework identified the 1D-CNN Hybrid architecture as the most deployment-oriented candidate, due to its substantially lower CPU utilization and nearly three times the execution speed. Furthermore, the Value of Forecasting (VoF)-based operational interpretation suggested that lightweight, computationally efficient forecasting architectures may reduce imbalance-related operational costs during highly unstable meteorological conditions. The findings demonstrate that the statistically most accurate forecasting model may not necessarily represent the most operationally suitable solution for real-world smart-grid deployment environments.**

Keywords— photovoltaic forecasting, smart grids, edge computing, deployment-aware evaluation, efficiency–accuracy index, deep learning

I. INTRODUCTION

The increasing integration of photovoltaic (PV) energy systems into modern smart grids has introduced significant operational challenges due to the intermittent and highly variable nature of solar power generation. Maintaining grid stability, managing reserve activation mechanisms, and minimizing imbalance-related operational costs require highly reliable short-term forecasting frameworks. Consequently, one-hour-ahead PV power forecasting has become a critical component of real-time grid management, energy scheduling, and distributed energy resource coordination.

Recent advances in deep learning, particularly Transformer-based architectures, have significantly improved photovoltaic forecasting performance in complex multivariate time-series environments. However, existing studies still primarily evaluate forecasting models using conventional statistical metrics such as RMSE, MAE, and R^2, whereas deployment-oriented factors including execution latency, hardware efficiency, and edge-deployment feasibility remain insufficiently investigated.

Existing PV forecasting literature has evolved across several methodological paradigms including statistical learning [1, 2], deep learning, Transformer-based forecasting, and explainable artificial intelligence (XAI) frameworks [3, 4]. The major operational limitations of these approaches are summarized in TABLE I.

TABLE I. TAXONOMY OF PV FORECASTING LITERATURE AND MISSING OPERATIONAL DIMENSIONS

Study Category	**Primary Focus**	**Common Evaluation Metrics**	**Missing Operational Dimension**
Statistical & classical ML models [1, 2]	Forecast accuracy	RMSE, MAE, MAPE	Operational robustness under rapid weather transitions
Deep learning forecasting [4]	Sequence learning & error minimization	RMSE, MAE, R^2	Explainability and physical interpretability
Transformer-based forecasting [5, 6]	Representation learning & architectural complexity	RMSE, MSE, MAE	Latency, hardware efficiency, and deployability
Explainable AI (XAI) studies [3, 4]	Feature importance & interpretability	SHAP, attention weights	Economic interpretation and operational value
Existing PV forecasting literature	***Statistical forecasting performance***	***Conventional accuracy metrics***	***Unified deployment-aware evaluation***

As summarized in TABLE I, existing PV forecasting studies have predominantly focused on improving statistical prediction accuracy, whereas deployment-oriented operational dimensions such as latency, hardware efficiency, economic interpretation, and edge-deployment suitability have received relatively limited attention. These limitations motivate the need for a unified deployment-aware evaluation framework for smart-grid forecasting applications. The main contributions of this study can be summarized as follows:

- A deployment-aware evaluation framework based on a composite EAI is proposed for short-term PV forecasting systems.
- Forecasting architectures are jointly evaluated using statistical accuracy, execution latency, and computational compactness metrics.
- Operational smart-grid scenarios including edge deployment and real-time forecasting constraints are incorporated into the evaluation perspective.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

- Experimental analyses demonstrate that lightweight forecasting architectures may offer improved deployment suitability for edge-oriented smart-grid applications.

II. Proposed Methodology

The overall workflow of the proposed deployment-aware evaluation methodology is illustrated in Fig. 1. Unlike conventional approaches that focus solely on algorithmic error reduction, the proposed framework evaluates candidate forecasting models using a multidimensional profiling structure. As shown in Fig. 1, after acquiring real-world smart grid data, candidate architectures are profiled across three distinct operational dimensions: statistical accuracy, inference latency, and hardware compactness. These profiles are then synthesized within the proposed Efficiency-Accuracy Index (EAI) engine to make the optimal edge deployment decision.

A. Candidate Forecasting Architectures

Four previously developed forecasting architectures from the author's prior doctoral study [5] were selected as candidate deployment models within the proposed evaluation framework, including Vanilla Transformer, LSTM Hybrid, GRU Hybrid, and 1D-CNN Hybrid encoders. The baseline Transformer architecture consists of multi-head self-attention and position-wise feed-forward layers, designed for modeling temporal dependencies in multivariate time-series forecasting tasks [6, 7].

To improve local temporal feature extraction capability [8], the conventional feed-forward structure was replaced with alternative temporal encoding modules, including recurrent (LSTM and GRU) and convolutional (1D-CNN) components. To ensure a fair and reproducible comparison, all forecasting architectures were optimized using Bayesian Optimization under identical search conditions. The hyperparameter search space included embedding dimensions (8, 16, and 32), the number of encoder blocks ($1-10$), dropout rates ($0.0-0.5$), and activation functions (ReLU and Linear). Model training was performed using Adam optimization with a maximum of 100 epochs and early stopping based on validation loss.

Among these architectures, the convolution-based 1D-CNN encoder utilizes local connectivity and weight-sharing mechanisms to efficiently capture short-term temporal fluctuations while maintaining relatively low computational complexity [9, 10]. This property makes convolution-based hybrid structures particularly suitable for edge-oriented smart-grid forecasting applications.

All experiments were implemented in Python using TensorFlow/Keras. The dataset was divided into training, validation, and test subsets using a 70% - 15% - 15% partitioning strategy. Execution time and CPU utilization measurements were recorded during model inference to evaluate deployment suitability in addition to forecasting accuracy. The final model configurations were selected according to their validation performance to avoid architecture-specific tuning bias.

B. Efficiency-Accuracy Index (EAI)

Conventional forecasting evaluation approaches generally assess model performance using independent metrics with different physical units, including forecasting error, execution latency, and hardware utilization. To enable deployment-oriented comparison under edge-computing constraints, these heterogeneous metrics were integrated into a unified EAI framework.

Prior to integration, all cost-related metrics were normalized using a reversed Min-Max transformation, where lower metric values correspond to higher operational suitability scores, as defined in (1):

$$Score(X) = 1 - \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

The proposed EAI formulation combines normalized forecasting accuracy, latency, and computational compactness into a single deployment-oriented evaluation metric, as expressed in (2):

$$EAI = \alpha(Acc_{norm}) + \beta(Lat_{norm}) + \gamma(Comp_{norm}) \quad (2)$$

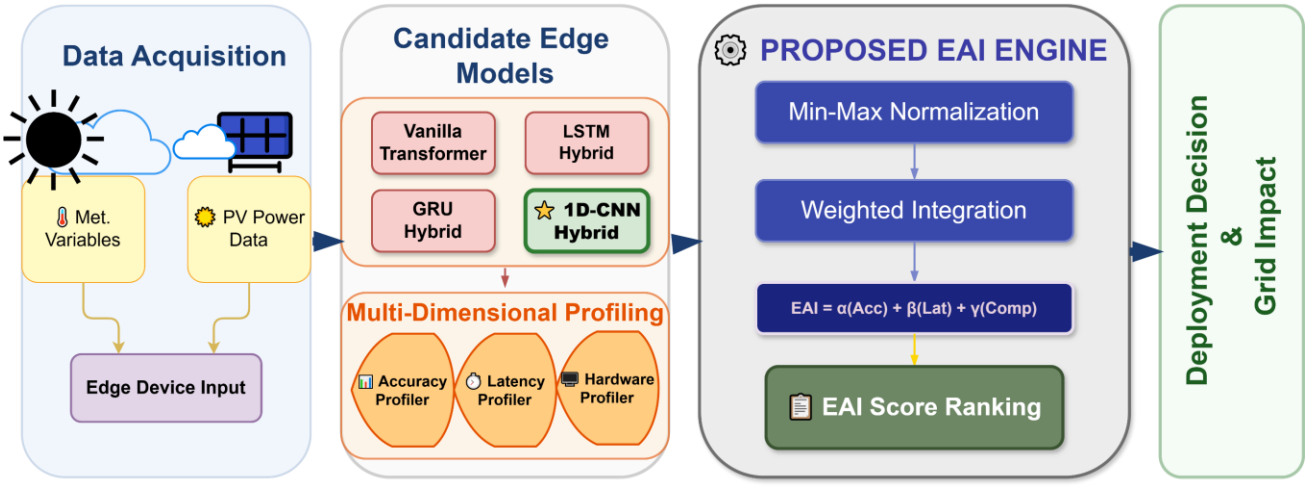

Fig. 1. Overall block diagram of the proposed deployment-aware evaluation framework.

where Acc_{norm} represents normalized forecasting accuracy, Lat_{norm} denotes normalized execution latency, and $Comp_{norm}$ corresponds to normalized computational compactness based on CPU utilization.

The weighting coefficients satisfy the normalization constraint given in (3):

$$\alpha + \beta + \gamma = 1 \quad (3)$$

and can be adaptively adjusted according to specific smart-grid operational scenarios including energy market optimization, real-time nowcasting, and edge deployment environments.

C. Experimental Setup

Experimental analyses were conducted using real-world hourly PV production data collected from a $100\ kW$ rooftop photovoltaic power plant located in Akhisar, Manisa, Türkiye. The dataset covers the period between December 2022 and May 2024 and includes inverter-level generation records together with 24 meteorological variables obtained from the Turkish State Meteorological Service.

To improve numerical stability and maintain temporal consistency, the dataset was normalized, temporal features were encoded, and a chronological train-validation-test split was applied. All candidate forecasting architectures were implemented under identical optimization settings to ensure fair comparative evaluation.

All experiments were implemented in Python using TensorFlow/Keras within the Google Colab environment. Model training and evaluation were performed using a Google Colab Free runtime equipped with NVIDIA T4 GPU acceleration and approximately 12 GB RAM.

In addition to conventional statistical forecasting metrics, execution latency and CPU utilization measurements were directly incorporated into the proposed deployment-aware EAI framework.

III. Results and Discussion

A. Conventional Forecasting Performance

The forecasting performances of the candidate architectures were first evaluated using conventional statistical error metrics. Aggregated test results demonstrated that the Vanilla Transformer architecture achieved the best overall forecasting accuracy with the lowest global $RMSE$ value ($8248.90\ W$) and the highest coefficient of determination ($R^2 = 0.8944$). When normalized by the average PV production level, the Vanilla model also produced the lowest normalized RMSE.

Despite its lightweight convolution-based structure, the 1D-CNN Hybrid architecture closely followed the Vanilla Transformer with only a marginal reduction in forecasting accuracy ($RMSE = 8430.25\ W,\ R^2 = 0.8897$). These findings indicate that convolution-based temporal encoding structures can preserve competitive forecasting capability while significantly reducing computational complexity.

B. Computational Cost Analysis

Execution latency and hardware utilization analyses revealed substantial differences among the candidate architectures in terms of deployment suitability for edge-oriented smart-grid environments. The Vanilla Transformer required approximately $510\ s$ to complete forecasting operations and produced average CPU utilization levels exceeding 14.5%, indicating relatively high computational overhead for embedded deployment scenarios.

In contrast, the 1D-CNN Hybrid architecture significantly reduced inference latency to nearly $160\ s$ while maintaining CPU utilization around 11%, representing the most computationally efficient architecture among all evaluated models. The local connectivity and weight-sharing mechanisms of convolutional encoding structures reduced computational burden without introducing substantial forecasting degradation. The comparative statistical and deployment-oriented performances of the candidate forecasting architectures are presented in TABLE II.

TABLE II. Comparative Performance and Deployment Metrics of Candidate Architectures

Model	RMSE (W)	R^2	Execution Time (s)	CPU Usage (%)	EAI Score
Vanilla Transformer	**8248.90**	**0.8944**	510	14.5	0.510
1D-CNN Hybrid	8430.25	0.8897	**160**	**11.0**	**0.963**
LSTM Hybrid	8506.71	0.8877	525	12.0	0.627
GRU Hybrid	10743.82	0.8290	325	14.5	0.137

Although the Vanilla Transformer achieved the lowest global RMSE, the 1D-CNN Hybrid architecture demonstrated substantially improved deployment suitability, with lower execution latency and reduced CPU utilization. These findings suggest that lightweight convolution-based forecasting architectures may provide improved scalability and operational feasibility for real-time smart-grid applications operating under limited hardware resources.

C. Evaluation via Proposed EAI Framework

Conventional forecasting evaluation approaches based solely on statistical error minimization may not accurately reflect operational deployment suitability under real-world smart-grid constraints. Within the proposed EAI framework, normalized forecasting accuracy, execution latency, and computational compactness were integrated into a unified, deployment-oriented evaluation framework. The proposed EAI framework supports adaptive weighting strategies according to different smart-grid operational scenarios. Representative deployment configurations are summarized in TABLE III.

TABLE III. SCENARIO-BASED EAI WEIGHTING STRATEGIES FOR SMART-GRID APPLICATIONS

Deployment Scenario	α (Accuracy)	β (Latency)	γ (Compactness)	Operational Priority
Energy market operation	0.70	0.15	0.15	Forecast accuracy
Real-time nowcasting	0.40	0.50	0.10	Low latency
Edge deployment	0.40	0.20	0.40	Hardware efficiency

These deployment-specific weighting strategies enable flexible operational prioritization between forecasting accuracy, computational efficiency, and hardware compactness. In addition to the representative scenarios listed in TABLE III, a balanced deployment configuration was also evaluated to provide a neutral reference case for model comparison. Under this balanced smart-grid deployment scenario where forecasting accuracy, latency, and hardware efficiency were weighted as $\alpha = 0.50$, $\beta = 0.25$, and $\gamma = 0.25$ respectively, the resulting deployment rankings differed substantially from conventional RMSE-centered evaluation results.

Although the Vanilla Transformer achieved the best global forecasting accuracy, its relatively high computational cost reduced its EAI score to 0.510. Conversely, the 1D-CNN Hybrid architecture achieved the highest deployment-oriented ranking with an EAI score of 0.963 by combining competitive forecasting performance with substantially improved computational efficiency and reduced latency.

To evaluate the robustness of the proposed EAI framework, a fine-grained sensitivity analysis was performed over $5{,}151$ feasible weighting configurations under the constraint (3) with a step size of 0.01. The distribution of model selections across the entire weighting space is illustrated in Fig. 2.

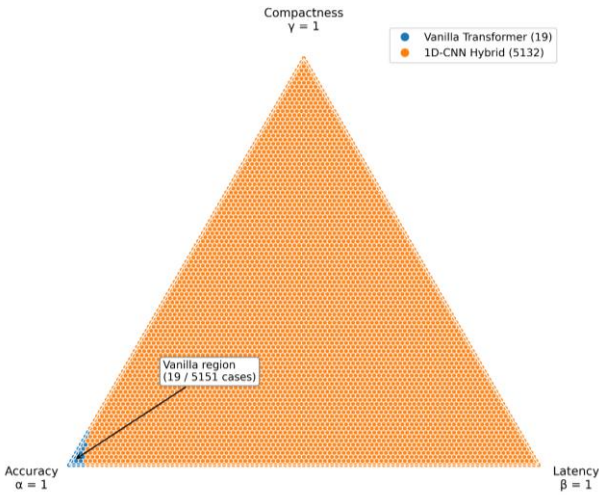

Fig. 2 Sensitivity analysis of EAI weighting space.

The results show that the 1D-CNN Hybrid achieved the highest EAI score in $5{,}132$ configurations, corresponding to 99.63% of the weighting space. The Vanilla Transformer was preferred only in 19 configurations, all of which were concentrated within a narrow accuracy-dominant region corresponding to extremely high accuracy priorities (approximately $\alpha > 0.94$). Neither the LSTM Hybrid nor the GRU Hybrid architecture achieved the highest EAI score under any feasible weighting configuration. This indicates that the deployment ranking is not strongly dependent on a single manually selected weighting configuration, and that the 1D-CNN Hybrid consistently provides a favorable balance between forecasting accuracy, execution latency, and computational compactness. The sensitivity analysis further demonstrates the stability of the proposed EAI framework under varying weighting preferences, supporting its applicability across diverse smart-grid operational scenarios.

Consequently, deployment-oriented evaluation metrics such as the proposed EAI can provide a more realistic assessment of forecasting architectures intended for practical smart-grid operation than accuracy-based metrics alone.

D. Operational and Economic Implications

Operational analyses further revealed that forecasting architectures responded differently under highly unstable meteorological conditions characterized by rapid fluctuations in irradiance and severe cloud transitions. Daily performance analyses indicated that the Vanilla Transformer occasionally experienced substantial forecasting degradation under extreme weather variability, whereas the 1D-CNN Hybrid architecture maintained comparatively stable forecasting behavior with narrower error fluctuations.

From a smart-grid operational perspective, reduced forecasting variance may directly contribute to lower reserve activation requirements and reduced imbalance-related operational costs [11, 12]. To quantify the practical impact of forecasting improvements, a Value of Forecasting (VoF) analysis was conducted using representative imbalance-cost assumptions reported in the literature [13, 14]. Based on the observed reductions in forecasting error during highly variable weather conditions, the operational benefits associated with the 1D-CNN Hybrid architecture were estimated to range between approximately $3.6\,USD$ and $36\,USD$ per day for a $100\,kW$ rooftop PV installation, depending on the assumed market penalty and balancing-cost scenario.

Overall, the findings suggest that forecasting architectures should be evaluated not only in terms of statistical accuracy but also with respect to deployment efficiency, operational robustness, and potential economic value. Lightweight forecasting architectures capable of maintaining stable predictive performance under dynamically changing weather conditions may therefore provide improved deployment suitability for modern edge-oriented smart-grid applications.

IV. CONCLUSION

This study presented a deployment-aware evaluation framework for short-term photovoltaic power forecasting systems using a composite EAI. Unlike conventional forecasting assessment approaches relying solely on statistical error minimization [1, 2], the proposed framework jointly evaluates forecasting accuracy, execution latency, and

computational compactness to support operationally suitable model selection for smart-grid applications.

Experimental analyses conducted using real-world PV production data demonstrated that the Vanilla Transformer architecture achieved the best global forecasting accuracy in terms of $RMSE$ and R^2 metrics [5]. However, deployment-oriented evaluation using the proposed EAI framework revealed that the 1D-CNN Hybrid architecture provided a substantially more balanced tradeoff between forecasting performance and computational efficiency. In particular, the lightweight convolution-based architecture achieved significantly lower execution latency and reduced CPU utilization while maintaining competitive forecasting accuracy under highly variable meteorological conditions.

The findings further indicate that forecasting architectures optimized exclusively for statistical accuracy may become computationally impractical for edge-oriented smart-grid deployment scenarios. In contrast, lightweight and computationally efficient forecasting structures may offer improved operational robustness, scalability, and reduced imbalance-related operational costs under real-world grid conditions [14].

Generally, the proposed EAI framework introduces a multidimensional deployment-oriented evaluation perspective that extends beyond conventional accuracy-centered forecasting assessment methodologies. Future research will investigate the applicability of the proposed framework to recently developed forecasting architectures, including Autoformer, FEDformer, and PatchTST, while also incorporating additional deployment-oriented metrics such as memory utilization, energy consumption, GPU efficiency, and distributed edge-device scalability. Such extensions may further enhance the practical relevance of deployment-aware forecasting evaluation for next-generation smart-grid systems.

REFERENCES

[1] D. W. Van der Meer, J. Widen, and J. Munkhammar, "Review on probabilistic forecasting of photovoltaic power production and electricity consumption," Renewable and Sustainable Energy Reviews, vol. 81, pp. 1484-1512, 2018. https://doi.org/10.1016/j.rser.2017.05.212.

[2] D. Chicco, M. J. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation," PeerJ Computer Science, vol. 7, e623, 2021.

[3] B. Lim, S. Ö. Arık, N. Loeff, and T. Pfister, "Temporal fusion transformers for interpretable multi-horizon time series forecasting," International Journal of Forecasting, vol. 37, no. 4, pp. 1748-1764, 2021. https://doi.org/10.1016/j.ijforecast.2021.03.012.

[4] M. Husein, E. J. Gago, B. Hasan, and M. C. Pegalajar, "Towards energy efficiency: A comprehensive review of deep learning-based photovoltaic power forecasting strategies," Heliyon, vol. 10, no. 13, 2024. https://doi.org/10.1016/j.heliyon.2024.e33419.

[5] Ö. Çelik, "Development of Explainable Hybrid Transformer Architectures for Photovoltaic Power Forecasting Using Feature Extraction Based on Multivariate Time Series," Ph.D. dissertation, Graduate School of Natural and Applied Sciences, Balıkesir University, Balıkesir, Türkiye, 2026. [Online]. Available: Balıkesir University Open Access Repository

This work was supported by the Balıkesir University Scientific Research Projects Coordination Unit (BAP) under Project No. 2024/007.

[6] Y. Liu et al., "iTransformer: Inverted transformers are effective for time series forecasting," in International Conference on Learning Representations (ICLR), 2023. *arXiv e-prints*, arXiv-2310. https://doi.org/10.48550/arXiv.2310.06625.

[7] A. Vaswani et al., "Attention is all you need," in Advances in Neural Information Processing Systems (NourIPS), vol. 30, pp. 5998-6008, 2017. https://doi.org/10.48550/arXiv.1706.03762.

[8] M. Geva, R. Schuster, J. Berant, and O. Levy, "Transformer feed-forward layers are key-value memories," in Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing, pp. 5484-5495, 2021. https://doi.org/10.48550/arXiv.2012.14913.

[9] J. P. Rala Cordeiro et al., "A machine learning based methodology for very short-term prediction of global horizontal irradiance," IEEE Latin America Transactions, vol. 19, no. 11, pp. 1827-1834, 2021.

[10] S. Yalçın and M. S. Herdem, "Prediction and analysis of weather parameters with global horizontal solar irradiance using LSTM-CNN based deep learning technique," BSEU Journal of Science, vol. 9, no. 1, pp. 340-356, 2022. https://doi.org/10.35193/bseufbd.1037563.

[11] M. Hodge, E. Ela, and M. Milligan, "The distribution of wind power forecast errors from operational systems," National Renewable Energy Laboratory (NREL), Tech. Rep. NREL/CP-5500-52568, 2015. Available online: https://www.nrel.gov/docs/fy11osti/52568.pdf (accessed on June 12, 2024).

[12] E. Ela, M. Milligan, and B. Kirby, "Operating reserves and variable generation," National Renewable Energy Laboratory (NREL), Tech. Rep. NREL/TP-5500-51978, 2011. https://doi.org/10.2172/1023095.

[13] Zhang et al., "Assessment of the economic impact of forecasting errors in Peer-to-Peer energy trading," Applied Energy, vol. 374, p. 123750, 2024. https://doi.org/10.1016/j.apenergy.2024.123750.

[14] M. Sage, J. Campbell, and Y. F. Zhao, "Enhancing Battery Storage Energy Arbitrage with Deep Reinforcement Learning and Time-Series Forecasting," in Proc. of the ASME 2024 18th International Conference on Energy Sustainability (ES2024), Paper No. ES2024-130538, 2024. https://doi.org/10.1115/es2024-130538.