

Article

Embedded Deep Learning for Short-Term PV Forecasting Under Export Constraints

Aymen Mnassri ^{1,2,*} , Nouha Mansouri ² , Sihem Nasri ¹, Abderezak Lashab ³ , Juan C. Vasquez ³ 
and Adnane Cherif ¹

¹ Analyze and Process Electrical and Energy Signals (ATSSEE) Research Laboratory, Faculty of Sciences of Tunis (FST), University of Tunis El Manar, Farhat Hached University Campus, Tunis 1068, Tunisia; sihem.nasri@fst.utm.tn (S.N.); adnen2fr@yahoo.fr (A.C.)

² Esprit School of Engineering, Ariana 2083, Tunisia; nouha.mansouri@esprit.tn

³ Center for Research on Microgrids (CROM), Department of Energy Technology, Aalborg University, 9220 Aalborg, Denmark; abl@energy.aau.dk (A.L.); juq@energy.aau.dk (J.C.V.)

* Correspondence: aymen.mnassri@esprit.tn

Abstract

The increasing penetration of photovoltaic (PV) systems requires accurate and stable short-term forecasting to ensure reliable grid operation under operational constraints. This paper investigates short-horizon multi-step PV power forecasting using one full year of high-resolution (5 min) real-world data from a 111-kW grid-connected rooftop installation. The forecasting problem is formulated as a direct multi-output supervised learning task with a 30 min prediction horizon. A comprehensive comparative evaluation is conducted across baseline (persistence), tree-based (XGBoost), and deep learning architectures (LSTM, GRU, and Temporal Convolutional Networks—TCN). Results show that deep learning models significantly outperform conventional baselines, with LSTM achieving the lowest normalized RMSE ($\approx 10.3\%$), while TCN provides a competitive trade-off between predictive accuracy, temporal stability, and computational efficiency. The direct multi-step formulation was adopted to reduce potential error propagation effects commonly observed in recursive forecasting approaches. Beyond forecasting accuracy, the study evaluates computational complexity and inference latency to assess practical deployability in resource-constrained environments. The proposed framework demonstrates that high-resolution real-world PV forecasting can achieve both strong predictive performance and operational feasibility. These findings contribute to the development of robust short-term forecasting strategies for distributed renewable energy systems operating under regulatory export constraints.

Keywords: photovoltaic forecasting; embedded AI; energy management system; export limitation; Raspberry Pi; multi-step prediction; LSTM; GRU; temporal convolutional network; smart building



Academic Editor: Gian
Giuseppe Soma

Received: 8 April 2026

Revised: 22 June 2026

Accepted: 25 June 2026

Published: 28 June 2026

Copyright: © 2026 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 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The rapid expansion of PV power has become a central pillar in global clean energy transition pathways, driven by the high reliability of PV plants and the vast availability of solar resources [1]. As worldwide PV installations continue to grow, accurate forecasting of PV power generation is increasingly essential for improving electrical grid management and ensuring the efficient integration of renewable energy sources [2]. In recent years, the global PV market has witnessed exceptional growth: cumulative installed capacity increased from 1186 GW in 2022 [3] to approximately 2260 GW by the end of 2024 and is

projected to approach 2.9 TW by the end of 2025, reflecting the unprecedented acceleration of solar deployment worldwide according to the latest IEA PVPS Trends Report [4].

This rapid development is particularly evident in emerging markets such as Brazil, where installed PV capacity surged from 2.46 GW in 2018 to over 47 GW by August 2024, reaching approximately 55 GW of operational installed capacity by February 2025 [5]. However, the large-scale integration of PV generation into electrical networks introduces significant operational challenges due to the inherent variability and uncertainty of solar irradiance. PV output is strongly affected by rapidly changing meteorological conditions, including cloud cover, temperature, and humidity, which can lead to sudden fluctuations in power production [6]. Such variability may disrupt the supply–demand balance, cause voltage and frequency deviations, and increase reliance on ancillary services, thereby threatening grid stability and reliability [7].

Consequently, PV power forecasting has emerged as a key solution to enhance solar energy integration. By providing reliable predictions across multiple temporal and spatial scales, forecasting models support improved planning, scheduling, and real-time grid operation. Grid operators can exploit accurate PV forecasts to optimize dispatch strategies [8], reduce reserve margins, and minimize balancing costs. Furthermore, forecasting plays a vital role in electricity market participation, financial risk mitigation, revenue maximization for PV plant owners, and long-term investment planning. Over the past decade, PV forecasting methodologies have evolved from traditional statistical approaches to advanced machine learning and hybrid physical–data-driven models [9]. Despite significant progress, forecasting accuracy remains limited by the intermittent nature of solar power, complex nonlinear relationships among atmospheric variables, and the persistent scarcity of large-scale high-quality training datasets [10].

Accurate prediction of PV power generation and solar irradiance across multiple time scales, ranging from very short-term nowcasting to day-ahead forecasting, is essential for reliable grid operation and energy planning [11]. Conventional statistical approaches, including persistence models, Autoregressive Integrated Moving Average (ARIMA), and Auto Regressive Moving Average with eXogenous inputs (ARMAX), are often considered baseline methods. However, their forecasting accuracy tends to deteriorate significantly under highly variable weather conditions and over longer prediction horizons, leading to substantial errors [12,13].

In contrast, artificial intelligence-driven techniques have demonstrated superior predictive performance. Machine learning models such as Support Vector Machines (SVM), Random Forests (RF), Extreme Learning Machines (ELM), as well as deep learning architectures, generally achieve lower Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), while providing higher coefficients of determination (R^2) compared to traditional methods [14].

Several comparative investigations indicate that LSTM networks frequently outperform conventional Recurrent Neural Networks (RNN) and GRU in forecasting PV output and irradiance. Reported results include RMSE values around 67.5 MWh and prediction error rates near 2.3% [15,16]. For more complex spatiotemporal forecasting tasks, hybrid deep learning frameworks such as CNN–LSTM models, especially when enhanced with attention mechanisms, have been shown to further improve forecasting accuracy [17].

Beyond generation forecasting, AI techniques also play a key role in energy dispatch and operational decision-making. Reinforcement Learning (RL) and Deep Reinforcement Learning (DRL) strategies have been increasingly applied to optimize the coordination of PV systems coupled with energy storage in competitive electricity markets, enhancing profitability while reducing curtailment risks [18]. For instance, an Attentive Convolutional Deep Reinforcement Learning approach integrating convolutional and attention layers

achieved revenue gains between 11% and 23%, alongside a curtailment reduction of up to 76% relative to benchmark strategies [17,18].

Moreover, AI contributes to broader PV system advancements, including optimal sizing, Maximum Power Point Tracking (MPPT) control, fault diagnosis, and predictive maintenance [19]. Metaheuristic optimization algorithms such as Particle Swarm Optimization (PSO), Genetic Algorithms (GA), Artificial Bee Colony (ABC), and hybrid variants have been widely adopted to improve PV/storage design and control parameters, often surpassing standalone techniques in terms of economic performance and convergence efficiency.

Overall, intelligent solar energy systems increasingly rely on AI-based tools to enhance forecasting accuracy, detect operational anomalies, and optimize storage management under real-world conditions [12,13].

Despite considerable progress in Home Energy Management Systems (HEMS), important difficulties remain in optimizing energy consumption while preserving user comfort and ensuring efficient renewable-energy utilization [20]. The intermittent and uncertain nature of renewable energy sources further complicates energy management and motivates the development of robust and adaptive control strategies capable of operating under highly dynamic conditions [21].

Several optimization-based approaches have been investigated to improve energy dispatch and renewable integration in smart-grid and microgrid environments. Advanced control strategies, including model predictive control and evolutionary optimization techniques, have demonstrated promising capabilities for reducing operational costs, improving power scheduling efficiency, and enhancing renewable-energy utilization [22–25]. However, many of these approaches generally involve increased computational complexity and may become difficult to deploy on low-cost embedded platforms operating under strict real-time constraints.

Consequently, practical deployment increasingly requires lightweight and computationally efficient energy management solutions capable of balancing forecasting accuracy, operational performance, and embedded implementation feasibility.

To enable the practical deployment of these data-driven and AI-based techniques, recent studies have increasingly relied on embedded-system-based architectures for real-time data acquisition, monitoring, and control of PV systems. Recent research has also investigated privacy-preserving renewable-energy forecasting frameworks using distributed and federated learning approaches. For example, federated deep reinforcement learning strategies have been proposed to enable collaborative renewable forecasting while preserving data privacy across distributed energy systems. Such approaches represent an emerging research direction for secure and decentralized smart-grid forecasting applications [26]. In large-scale PV installations, embedded platforms integrated with atmospheric sensors have been employed to continuously monitor PV power output alongside environmental variables such as temperature, humidity, and irradiance, providing high-resolution input data for energy management and forecasting applications [27]. Other works have extended this approach by combining embedded devices with cloud-based infrastructures, enabling AI-driven prediction and anomaly detection through the integration of PV generation data, atmospheric measurements, and device status information for power optimization in large PV plants [28]. Low-cost IoT-based monitoring solutions have also been developed using microcontroller platforms to improve the accessibility and scalability of PV supervision systems [29]. In microgrid environments, embedded controllers combined with wireless communication modules and distributed sensors have been utilized to monitor environmental conditions and appliance-level power consumption, contributing to energy usage reduction and enhanced grid stability [30]. Furthermore, smart PV-based microgrids deploying embedded platforms such as Raspberry Pi and Arduino Mega, along with com-

munication switches and energy storage units, have demonstrated effective real-time data exchange and improved operational control under diverse operating conditions [31].

In this context, this work focuses on the development and experimental validation of an integrated PV power forecasting and embedded energy management framework for prosumer-scale systems operating under grid export constraints. The proposed approach aims to bridge the gap between high-accuracy short-term PV forecasting methods and their practical deployment within real-time, resource-constrained energy management systems.

The main contributions of this study are summarized as follows:

- The exploitation of one full year of high-resolution (5 min) real-world PV power data collected from a 111-kW grid-connected rooftop installation, enabling the evaluation of short-horizon (30 min) direct multi-step forecasting under realistic operating conditions, including measurement noise and data gaps;
- A comprehensive and fair comparison of baseline, tree-based, and deep learning forecasting models, with explicit consideration of both predictive accuracy and computational efficiency, highlighting the trade-offs between forecasting performance and suitability for embedded deployment.
- The design, implementation, and real-time deployment of a lightweight and regulation-aware EMS on a Raspberry Pi 4 platform, demonstrating significant improvements in photovoltaic self-consumption, grid import reduction, and curtailment mitigation while strictly complying with grid export limitations.

The remainder of this paper is organized as follows. Section 2 describes the photovoltaic system and the real-world dataset used in this study. Section 3 presents the proposed AI-based PV power forecasting methodology. Section 4 reports and analyzes the forecasting results. Section 5 details the embedded implementation and performance evaluation on a Raspberry Pi platform. Section 6 presents the integration of the forecasting framework into the Energy Management System (EMS). Finally, Section 7 concludes the paper and outlines future research directions.

2. System Description and Dataset

2.1. Photovoltaic System and Data Acquisition

This study is conducted on a grid-connected rooftop photovoltaic (PV) installation with a rated capacity of 111 kW, deployed on a university campus under real outdoor operating conditions. As illustrated in Figure 1, the PV system is embedded within a campus-scale energy architecture comprising local electrical loads, a utility grid subject to export constraints, and a real-time embedded Energy Management System (EMS) implemented on a Raspberry Pi platform. The experimental installation is located at the National Engineering School of Sfax, Tunisia, which is characterized by a Mediterranean climate with high solar irradiation and marked seasonal variability. The PV arrays are south-oriented with a fixed tilt angle adapted to local climatic conditions and interfaced through grid-connected string inverters. The PV plant represents a typical prosumer-scale installation, where production variability and short-term power fluctuations can directly affect local energy usage and interactions with the utility grid. From an operational perspective, such systems require reliable monitoring and short-term forecasting capabilities to support informed real-time decision-making, particularly in contexts where grid export is limited to a fixed fraction of the energy generated.

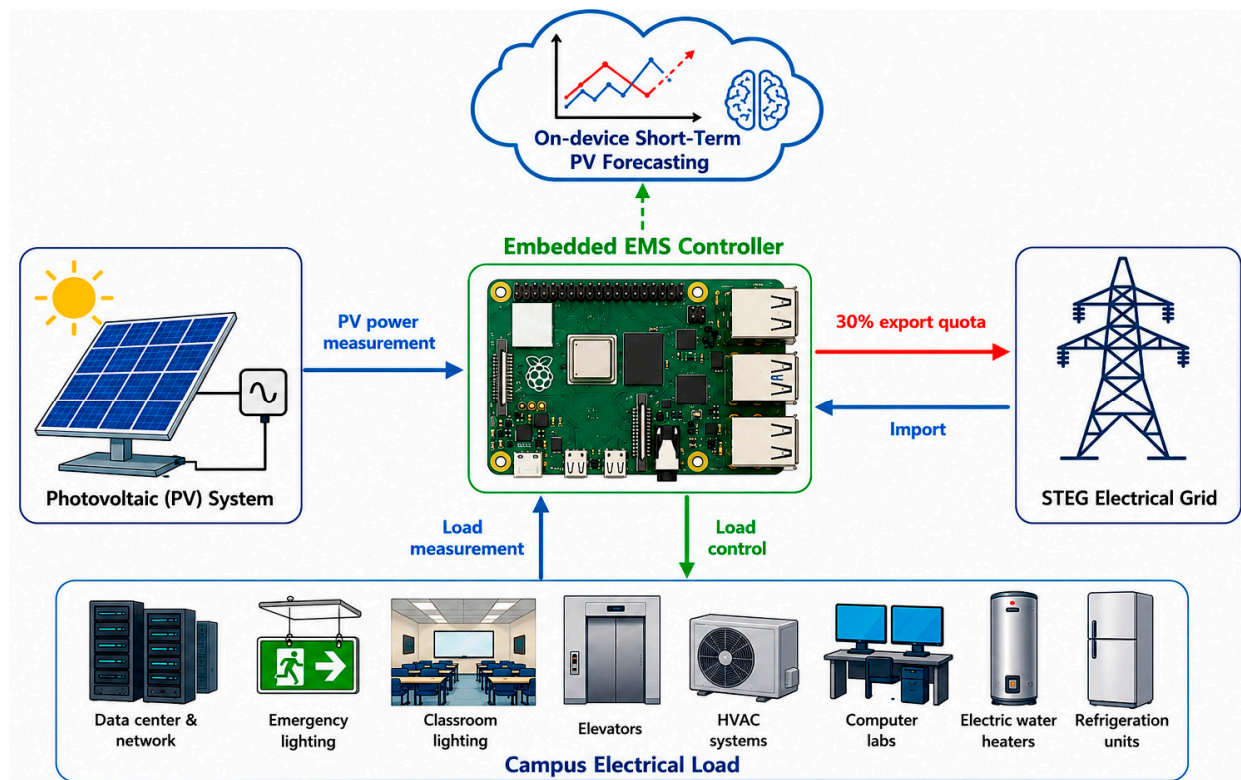


Figure 1. Global architecture of the forecasting-assisted real-time embedded energy management system for a grid-connected university campus under a 30% export constraint.

The PV plant is equipped with a monitoring platform based on a data logger/SCADA export, which records plant-level PV output power at a fixed temporal resolution of 5 min. The measured signal corresponds to the electrical power delivered at the inverter AC output. Solar irradiance measurements are acquired using on-site plane-of-array irradiance sensors. Unlike synthetic or simulated datasets, real monitoring data inherently include realistic measurement noise, operational irregularities, and occasional data gaps, making the resulting dataset representative of practical field deployment conditions.

A 5-min sampling resolution is particularly well suited for short-term operational studies, as it enables the capture of intra-hour PV dynamics such as cloud-induced fluctuations and rapid power ramps more accurately than coarser resolutions. At the same time, this resolution remains fully compatible with quasi-real-time EMS operation, where forecasting, control decisions, and monitoring updates are executed at short and regular intervals. Consequently, the selected temporal resolution provides a solid basis for developing short-term PV forecasting models intended to support actionable energy management decisions within an embedded control framework.

Figure 2 provides a detailed overview of the PV system and the dataset construction pipeline adopted in this work, from on-site monitoring to the generation of the final yearly time-series dataset. The figure highlights the real monitoring origin of the data (5-min resolution), the daily spreadsheet-based export structure, and the subsequent integration, cleaning, and harmonization steps required to obtain an analysis-ready dataset suitable for both forecasting model development and EMS integration.

The main characteristics of the PV system and data acquisition setup are summarized in Table 1.

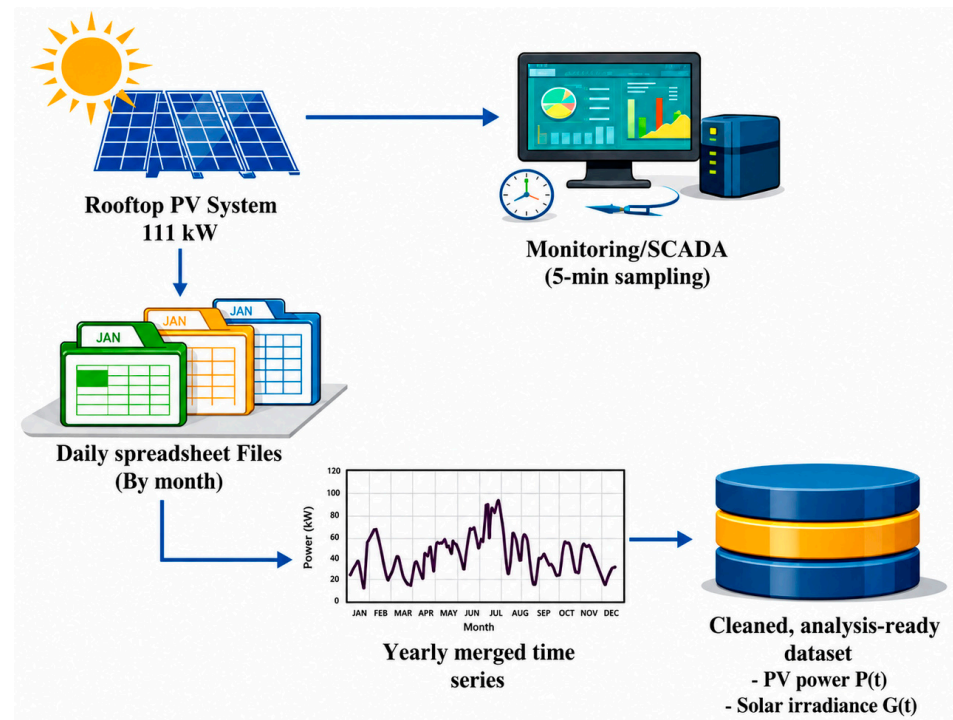


Figure 2. Photovoltaic data acquisition and dataset construction pipeline.

Table 1. PV system and acquisition characteristics.

Item	Value/Description
PV plant type	Grid-connected rooftop PV system
Rated capacity	111 kW
Measurement source	Monitoring export (data logger/SCADA)
Measured quantity	Plant-level PV output power $P_{PV}(t)$
Sampling interval	5 min
Study period	One full year

2.2. Dataset Structure and Integration

A real-world dataset covering one full year is used in this work. The raw data are provided as daily spreadsheet files, organized into 12 monthly folders. This structure is typical of industrial PV monitoring systems, which often generate daily operational reports rather than research-oriented continuous datasets. As a result, constructing an analysis-ready yearly time series requires a careful integration and harmonization process.

Each daily spreadsheet is parsed to extract timestamps and measurement columns, and all files are subsequently concatenated in chronological order. During this integration step, timestamp formats are standardized into a unified datetime representation to ensure temporal consistency across the entire dataset. Duplicate timestamps when present due to repeated exports or monitoring system resynchronization are handled by enforcing uniqueness at the time index level. This operation is essential to avoid ambiguities in temporal ordering and to enable reliable time-based analysis.

At a sampling interval of 5 min, the nominal number of samples over one year is given by:

$$N_{\text{nominal}} \approx 365 \times 24 \times 12 = 105120 \quad (1)$$

After integration and re-indexing onto a uniform 5-min time grid, the dataset spans from 1 January 2025 00:00:00 to 1 January 2026 00:00:00, resulting in 105,121 timestamps.

The dataset includes the plant-level PV output power $P(t)$, solar irradiance $G(t)$ when available in monitoring exports, and timestamps used for temporal indexing.

For clarity and reproducibility, the variables and units used throughout the paper are summarized in Table 2.

Table 2. Dataset variables and units.

Variable	Symbol	Unit	Description
PV output power	$P(t)$	kW (after conversion)	Plant-level PV power at 5-min resolution
Solar irradiance (if available)	$G(t)$	W/m^2	Irradiance measurement aligned with PV timestamps
Timestamp	t	Min	Date and time index (5-min grid)

2.3. Data Cleaning and Final Dataset Characteristics

Field monitoring data may contain duplicated timestamps, short monitoring interruptions, and anomalous measurements caused by communication failures or export artifacts. A preprocessing procedure was therefore applied to ensure physical consistency while preserving realistic operational variability. PV power measurements were expressed in kilowatts, and a physical plausibility filter based on the rated PV capacity ($P_{rated} = 111$ kW) was applied to remove extreme values associated with scaling or conversion errors. After filtering, the maximum observed PV power was 109.12 kW, consistent with the expected operating limits of the installation.

All measurement files were integrated and re-indexed onto a uniform 5-min time grid. Duplicate timestamps were removed, while missing measurements remained below 1.5% of the full dataset (less than five days over the study period). Missing PV values were not interpolated. Instead, only complete input–output windows containing valid measurements over both the historical input sequence (L samples) and prediction horizon (H samples) were retained during supervised sample generation.

Using the selected sliding-window configuration ($L = 24$, $H = 6$), a total of 48,666 valid supervised samples were generated from the original dataset containing 105,121 timestamps. The final supervised dataset was divided chronologically into 38,932 training samples (80%) and 9734 testing samples (20%). The main dataset characteristics and supervised dataset statistics are summarized in Tables 3 and 4.

Table 3. Yearly dataset summary and quality indicators.

Indicator	Value
Time span	1 January 2025 00:00:00 → 1 January 2026 00:00:00
Sampling interval	5 min
Total timestamps (uniform grid)	105,121
Missing PV measurements	≤5 days (≤1.5%)
PV rated capacity	111 kW
Maximum PV after filtering	109.12 kW

Such real-world characteristics combining high temporal resolution with missing intervals and measurement imperfections make the dataset representative of practical PV monitoring deployments and provide a realistic basis for evaluating forecasting models intended for embedded energy management.

Table 4. Supervised dataset generation statistics.

Processing Step	Number of Samples
Total timestamps (uniform grid)	105,121
Valid supervised samples	48,666
Training samples (80%)	38,932
Testing samples (20%)	9734

3. Methodology: AI-Based PV Power Forecasting

This section presents the proposed methodology for short-term PV power forecasting. The objective is to generate accurate, stable, and computationally efficient short-horizon predictions at high temporal resolutions to support near real-time energy management and embedded deployment. The overall forecasting workflow adopted in this study is summarized in Figure 3, which outlines the complete pipeline from raw measurements to multi-step PV power predictions used as actionable inputs for an EMS layer.

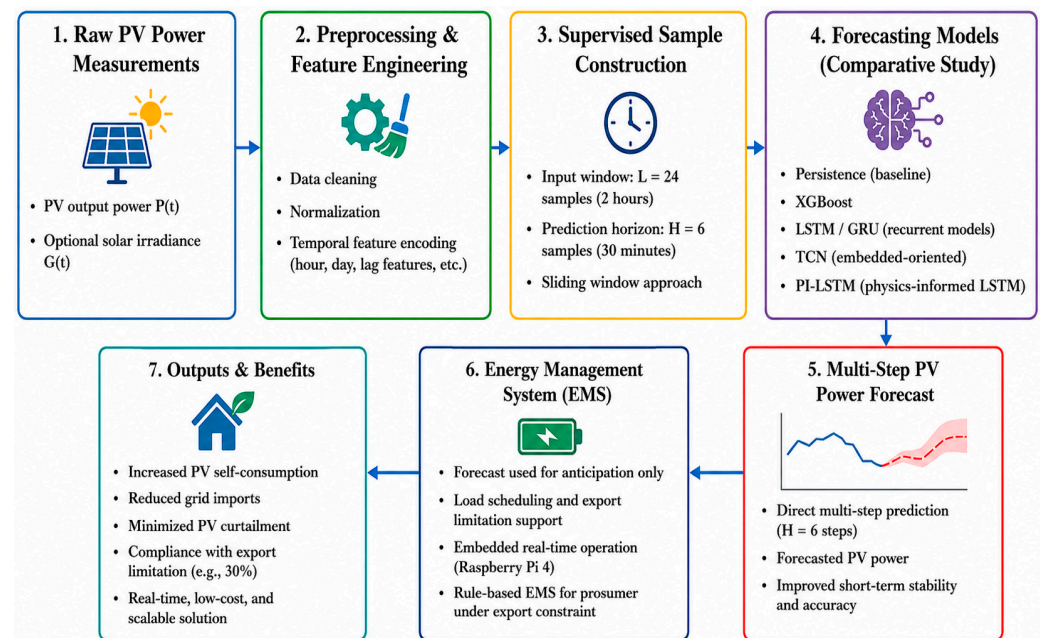


Figure 3. Overview of the short-term photovoltaic power forecasting methodology and its integration with the embedded energy management system for anticipative control.

3.1. Forecasting Formulation, Strategy, and Input Features

3.1.1. Forecasting Problem Formulation

Let $P(t)$ denote the PV output power measured at discrete time instant t , sampled at a constant interval

$$\Delta t = 5 \text{ min.} \quad (2)$$

The short-term PV forecasting task is formulated as a supervised multi-step regression problem. Given a historical observation window of length L , the objective is to predict the PV power over the next H future time steps [32].

The supervised learning formulation is expressed as:

$$X_t \longrightarrow Y_t \quad (3)$$

where the input sequence is defined as:

$$X_t = \{x_{(t-L+1)}, \dots, x_t\}, \quad x_t = [P(t), Z(t)] \quad (4)$$

and the corresponding multi-step output vector is:

$$Y_t = [P(t+1), P(t+2), \dots, P(t+H)] \quad (5)$$

Here, $Z(t)$ denotes the vector of exogenous and temporal features associated with time step t .

In this study, the forecasting parameters are fixed as follows:

$$L = 24 \text{ samples (2 h)}, H = 6 \text{ samples (30 min)}, \Delta t = 5 \text{ min} \quad (6)$$

3.1.2. Direct Multi-Step Forecasting Strategy

A direct multi-step forecasting strategy is adopted, in which a single model predicts the entire forecast horizon simultaneously:

$$Y_t = f_\theta(X_t) \quad (7)$$

where $f_\theta(\cdot)$ represents the forecasting model parameterized by θ .

The proposed forecasting framework intentionally relies on a compact set of locally available input features to preserve its lightweight embedded implementation characteristics. Although additional meteorological variables may further improve forecasting accuracy, their integration increases sensing, communication, and computational requirements. The selected feature configuration therefore represents a compromise between forecasting performance and real-time embedded deployability.

3.1.3. Temporal Feature Encoding and Exogenous Inputs

To capture the strong temporal regularities inherent to PV generation, cyclic time features are incorporated using sinusoidal encoding. For the hour-of-day $h(t) \in [0, 23]$ and the day-of-year $d(t) \in [1, 365]$, the temporal feature vector is defined as:

$$\left[\sin\left(\frac{2\pi h(t)}{24}\right), \cos\left(\frac{2\pi h(t)}{24}\right), \sin\left(\frac{2\pi d(t)}{365}\right), \cos\left(\frac{2\pi d(t)}{365}\right) \right] \quad (8)$$

This representation preserves the periodic nature of daily and seasonal PV patterns while avoiding artificial discontinuities at cycle boundaries.

When available, solar irradiance measurements $G(t)$ are included as an exogenous input variable. The inclusion of irradiance information improves the model's ability to capture rapid PV power variations caused by transient cloud cover and atmospheric dynamics.

3.2. Forecasting Models and Training Configuration

Several forecasting models are evaluated under the same problem formulation and data splitting strategy in order to ensure a fair and consistent comparison. The selected models range from simple reference baselines to advanced deep learning architectures, allowing the assessment of both forecasting accuracy and suitability for embedded deployment.

3.2.1. Baseline and Tree-Based Machine Learning Models

As a reference benchmark, a persistence model is implemented, defined as:

$$\hat{P}_{PV}(t+k) = P(t), \quad k = 1, \dots, H \quad (9)$$

This model assumes that future PV power remains equal to the most recent measurement and provides a minimum performance baseline commonly used in PV forecasting studies.

In addition, two tree-based machine learning models are considered due to their ability to capture nonlinear relationships while maintaining moderate computational complexity.

A Random Forest model is trained as a multi-output regressor, directly predicting the full forecast horizon H . The ensemble structure improves robustness against noise and overfitting, which is particularly relevant when dealing with real monitoring data.

XGBoost (version 1.7.6.) is implemented using a one-regressor-per-horizon-step strategy, where an independent regressor is trained for each prediction step k . This approach allows each forecast horizon to be modeled with different nonlinear mappings and has shown strong performance in short-term PV forecasting applications.

3.2.2. Training Setup, Normalization, and Physical Constraints

All input features are normalized using Min–Max scaling, and the models are trained using a chronological data split to preserve temporal causality.

To ensure physically meaningful predictions, PV power outputs are constrained by the rated capacity of the system:

$$0 \leq \hat{P}_{PV}(t+k) \leq P_{rated}, \quad \forall k \quad (10)$$

where $P_{rated} = 111$ kW. Predictions outside this range are clipped to maintain physical consistency.

To address the growing interest in physics-informed forecasting approaches and provide a comprehensive benchmark, a Physics-Informed LSTM (PI-LSTM) model is also included in the comparative study. Unlike the conventional LSTM, which applies physical constraints only during post-processing, the PI-LSTM incorporates the rated-capacity constraint directly into the training objective:

$$Loss = MSE + \lambda L_{phys} \quad (11)$$

where λ is a weighting coefficient controlling the contribution of the physical constraint term. To balance forecasting accuracy and physical consistency, the weighting coefficient was empirically fixed at $\lambda = 0.01$ throughout all PI-LSTM experiments.

$$L_{phys} = \max(0, \hat{P}_{PV} - P_{rated})^2 \quad (12)$$

This additional penalty encourages physically consistent predictions while preserving the same architecture, input variables, and training conditions as the standard LSTM model. Deep learning models are trained using the Adam optimizer with early stopping and learning-rate reduction to prevent overfitting and ensure convergence. The complete set of model architecture and training hyperparameters is summarized in Table 5.

Table 5. Forecasting models and training configuration.

Model	Input Variables	Architecture/Formulation	Training Setup
Persistence	PV	$\hat{P}_{PV}(t+k) = P_{PV}(t)$	No training
Random Forest	PV + G + time	Multi-output regressor	Depth/trees tuned; chronological split
XGBoost	PV + G + time	One regressor per horizon step	Moderate depth; learning-rate tuned

Table 5. Cont.

Model	Input Variables	Architecture/Formulation	Training Setup
LSTM	PV + time/PV + G + time	$1 \times \text{LSTM} (64) \rightarrow \text{Dense} (H)$	epochs = 80, batch = 64, Adam, Early Stopping (pat = 8), ReduceLR (pat = 4)
GRU	PV + time/PV + G + time	$1 \times \text{GRU} (64) \rightarrow \text{Dense} (H)$	same as LSTM
TCN-fixed	PV + G + time	4 dilated blocks (1, 2, 4, 8), filters = 32, kernel = 3, Flatten $\rightarrow \text{Dense} (128) \rightarrow \text{Dense} (H)$	Epochs = 120, batch = 64, Adam (1×10^{-3}), dropout 0.1–0.2
PI-LSTM	PV + time/PV + G + time	$1 \times \text{LSTM} (64) \rightarrow \text{Dense} (H)$; Loss = MSE + λL_{phys}	epochs = 80, batch = 64, Adam, Early Stopping (pat = 8), ReduceLR (pat = 4), $\lambda = 0.01$

3.3. Evaluation Protocol and Performance Metrics

The forecasting performance was evaluated using a chronological validation protocol and standard error metrics commonly adopted in short-term PV forecasting studies. A chronological train–test split was employed to preserve temporal causality and avoid information leakage from future observations. The supervised dataset generated using sliding windows was divided into 80% training samples and 20% testing samples. No random shuffling was applied. For deep learning models, an additional validation subset was extracted from the training data to support early stopping and adaptive learning-rate control.

Although a single train–test split was adopted in this study, this evaluation strategy reflects realistic operational conditions of the proposed EMS, where future PV generation must be predicted exclusively from historical observations. Future work will investigate rolling-origin and walk-forward validation strategies to further assess long-term temporal generalization.

Forecasting accuracy was assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and normalized RMSE (nRMSE), defined as [33]:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (P_i - \hat{P}_i)^2} \quad (13)$$

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |P_i - \hat{P}_i| \quad (14)$$

$$\text{nRMSE}(\%) = \frac{\text{RMSE}}{P_{\text{rated}}} \times 100 \quad (15)$$

where P_i and $\hat{P}_i$ denote the measured and predicted PV power values, respectively, and N is the total number of evaluated samples and $P_{\text{rated}} = 111 \text{ kW}$ corresponds to the rated PV capacity. All metrics were computed after inverse scaling and reported in kilowatts.

Since a direct multi-step forecasting strategy was adopted, errors were aggregated over the complete prediction horizon H , providing an overall assessment of the capability of each model to anticipate short-term PV power evolution for embedded EMS applications.

4. Forecasting Results and Analysis

This section presents the results obtained with the proposed short-term PV power forecasting framework and discusses their implications. Quantitative metrics are first analyzed to compare the forecasting accuracy of the different models. This is followed by a

qualitative analysis based on time-series plots and statistical error distributions in order to assess robustness under realistic operating conditions.

4.1. Quantitative Forecasting Results

The quantitative performance of the evaluated forecasting models for a 30-min prediction horizon ($H = 6$) is reported in Table 5. The comparison includes a persistence baseline, a tree-based machine learning model (XGBoost), and four deep learning architectures (LSTM, GRU, TCN-fixed, and PI-LSTM).

Repeated experiments under different initialization conditions exhibited low variability and preserved consistent model ranking behavior, confirming the stability and reproducibility of the obtained forecasting results.

As shown in Table 6, all learning-based models outperform the persistence baseline, confirming the effectiveness of data-driven approaches for short-term PV forecasting. Among the evaluated models, LSTM and PI-LSTM achieve the best overall performance, with LSTM slightly outperforming in purely data-driven accuracy, while PI-LSTM improves physical consistency. GRU follows closely, while TCN-fixed provides competitive accuracy with favorable characteristics for embedded deployment.

Table 6. Quantitative forecasting performance for a 30-min horizon ($H = 6$).

Model	RMSE (kW)	MAE (kW)	nRMSE (%)
Persistence	13.18	6.99	11.88
XGBoost	12.38	7.02	11.15
LSTM (PV + G + time)	11.38	6.07	10.26
GRU	11.54	6.45	10.40
TCN-fixed	12.09	7.42	10.89
PI-LSTM	11.25	11.25	10.18

To assess the variability of the deep learning results, multiple training runs were performed using different random seed initializations while maintaining the same chronological data split and hyperparameter configuration. The mean and standard deviation of the forecasting metrics are reported in Table 7.

Table 7. Forecasting variability across repeated runs.

Model	RMSE (kW) Mean $\pm$ Std	MAE (kW) Mean $\pm$ Std	nRMSE (%) Mean $\pm$ Std
LSTM	11.38 $\pm$ 0.12	6.07 $\pm$ 0.08	10.26 $\pm$ 0.11
GRU	11.54 $\pm$ 0.15	6.45 $\pm$ 0.10	10.40 $\pm$ 0.13
TCN-fixed	12.09 $\pm$ 0.10	7.42 $\pm$ 0.09	10.89 $\pm$ 0.09
PI-LSTM	11.25 $\pm$ 0.10	6.05 $\pm$ 0.07	10.18 $\pm$ 0.10

The results indicate consistent forecasting performance across repeated runs for all deep learning architectures. The standard deviation of nRMSE remains below 0.15% for all models, demonstrating stable forecasting behavior under different random initializations.

Among the evaluated models, LSTM and PI-LSTM achieve the most stable and accurate performance, with PI-LSTM showing slightly improved robustness due to the inclusion of physics-informed constraints in the training objective. GRU follows closely, while TCN-fixed maintains competitive and stable performance with lower computational complexity, making it particularly attractive for embedded deployment scenarios.

A sensitivity analysis was conducted to evaluate the influence of the historical input window length on forecasting performance. Different input window sizes were investigated

while maintaining the same forecasting horizon and training configuration. The results obtained are summarized in Table 8.

Table 8. Sensitivity of forecasting performance to historical input window length.

Input Window Length	nRMSE (%)
12 samples (1 h)	11.42
24 samples (2 h)	10.26
48 samples (4 h)	10.31

The results indicate that very short input windows provide insufficient temporal context for accurately capturing short-term PV dynamics. Conversely, larger windows provide only marginal accuracy improvements while increasing computational complexity and memory requirements. The selected configuration ($L = 24$ samples, corresponding to 2 h) therefore provides the best compromise between forecasting accuracy and computational efficiency.

4.2. Time-Series Prediction and Robustness Analysis

The qualitative behavior of the forecasting models is illustrated using representative figures based on the TCN-fixed model, which offers a favorable trade-off between forecasting accuracy and computational efficiency for embedded energy management applications. As expected in short-term PV forecasting, prediction uncertainty progressively increases with longer forecasting horizons. Although the present study focuses on a 30-min horizon to satisfy real-time EMS operational requirements, the proposed framework can be extended to longer horizons at the cost of reduced forecasting accuracy and increased uncertainty.

Figure 4 shows the comparison between measured and predicted PV power over a representative 24-h period. The predicted curve closely follows the measured PV output, accurately reproducing the main diurnal production profile, including the morning ramp-up, midday peak, and evening ramp-down. These results demonstrate that the TCN model effectively captures the dominant temporal dynamics of photovoltaic generation despite real-world variability and measurement noise.

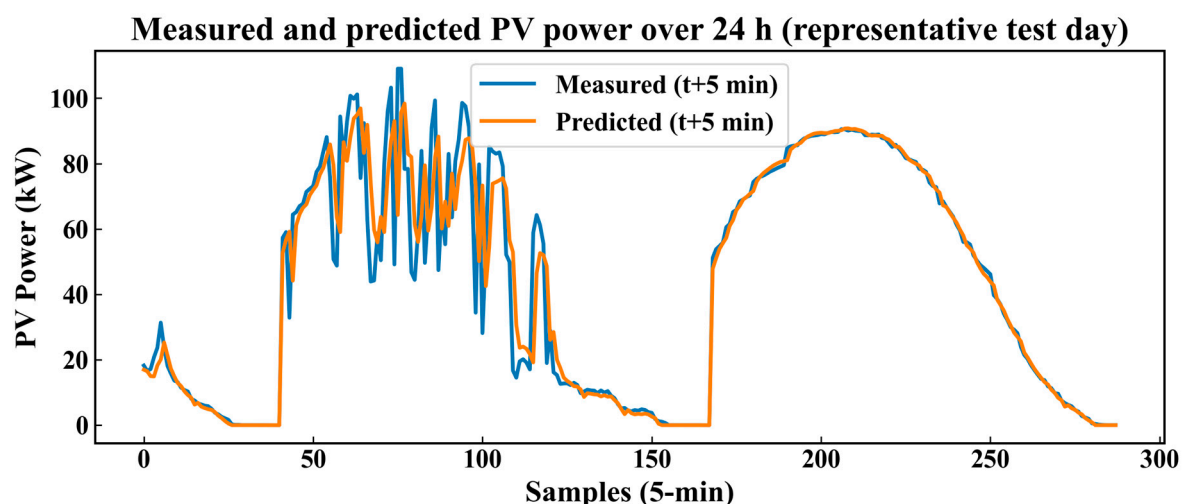


Figure 4. Measured and TCN-predicted PV power over a representative 24-hour period (5-min resolution).

To evaluate robustness under challenging operating conditions, Figure 5 presents a zoomed view during periods of rapid irradiance-induced PV power fluctuations. Despite abrupt changes in the measured signal, the TCN predictions remain stable and responsive,

with limited lag and no spurious oscillations. This behavior is particularly important for short-term operational use, as unstable predictions during fast power ramps may negatively affect downstream energy management and control actions.

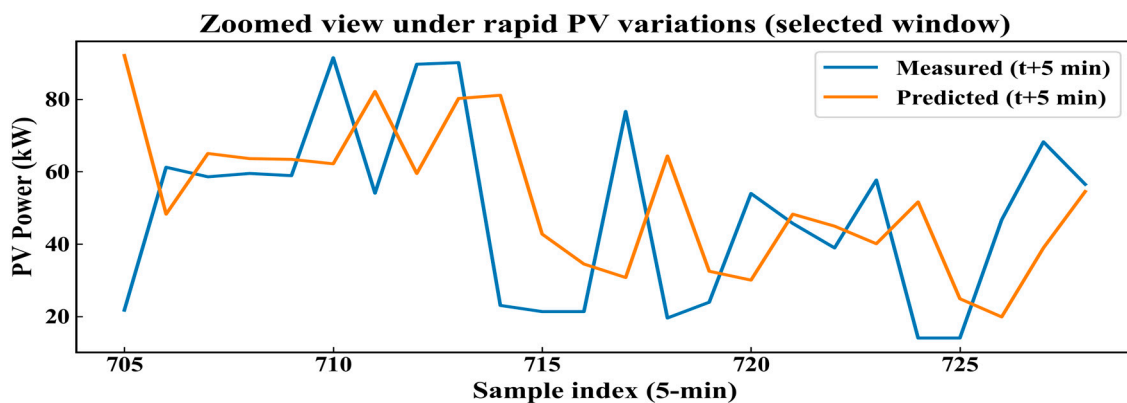


Figure 5. Zoomed-in comparison between measured and TCN-predicted PV power during rapid irradiance-induced fluctuations.

4.3. Error Distribution and Forecast Horizon Analysis

The statistical properties of the forecasting errors obtained with the TCN model are illustrated in Figure 6, which shows the distribution of prediction errors for the 30-min forecasting horizon. The error distribution is centered around zero, indicating the absence of systematic bias. Most errors remain within a relatively narrow range, while extreme deviations occur infrequently. This confirms that the forecasting performance remains stable over time, even in the presence of real-world data imperfections and rapid PV variability.

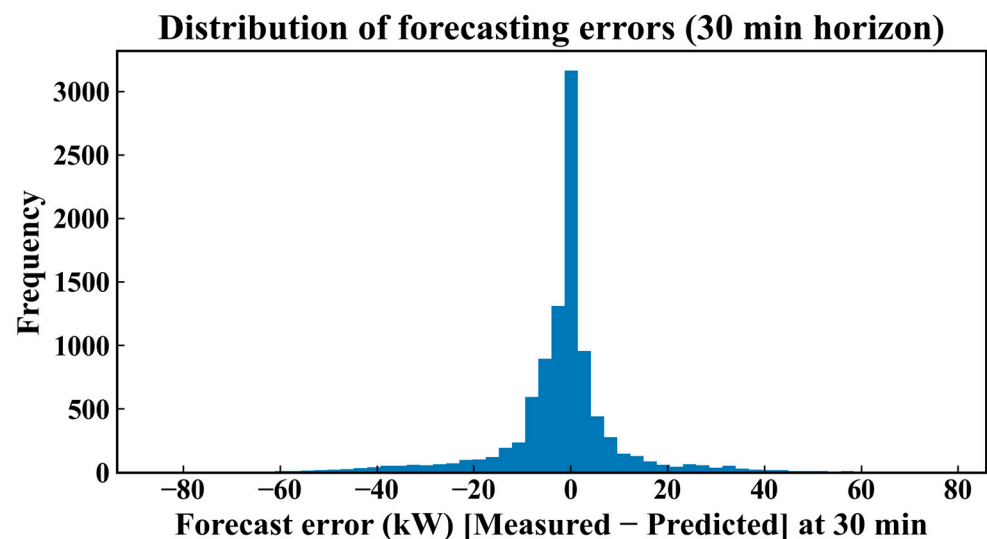


Figure 6. Distribution of TCN forecasting errors for a 30-min prediction horizon.

The effect of the prediction horizon on forecasting accuracy is analyzed in Figure 7, which reports the evolution of RMSE as a function of the forecast step. As expected, the prediction error increases gradually with the forecast horizon due to growing uncertainty. However, the degradation remains moderate over the full 30-min horizon considered in this study, indicating that the proposed forecasting framework provides a suitable compromise between anticipation capability and reliability for short-term energy management applications.

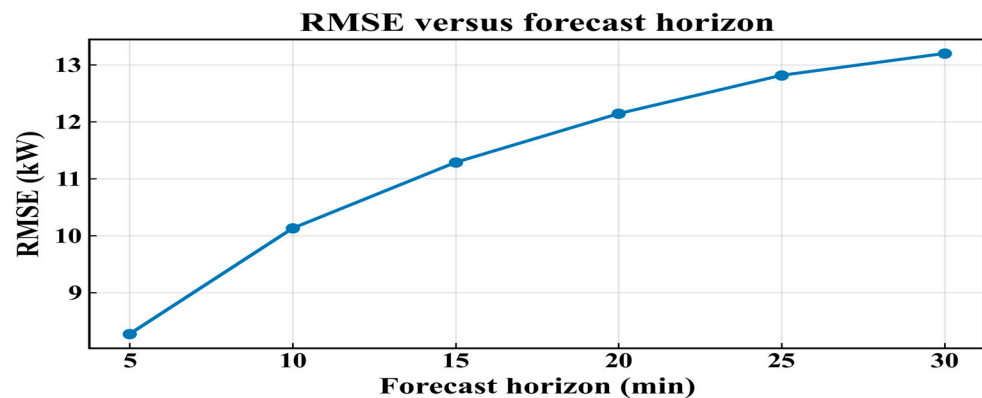


Figure 7. Evolution of RMSE as a function of the forecast horizon for the TCN model.

To further evaluate the robustness of the proposed forecasting framework, additional stress tests were conducted under noisy and incomplete input conditions. The selected perturbation levels were not intended to reproduce a specific failure scenario but rather to represent realistic monitoring uncertainties commonly encountered in practical PV installations, including sensor inaccuracies, communication disturbances, and temporary data losses. Gaussian noise with a standard deviation corresponding to 5% of the normalized PV signal amplitude was artificially added to the input sequences, while a separate experiment simulated incomplete monitoring conditions by randomly removing 10% of the input windows. These values were selected as representative moderate disturbances frequently adopted in robustness evaluations to assess model sensitivity under degraded measurement conditions. The corresponding forecast performance is summarized in Table 9.

Table 9. Robustness evaluation under noisy and incomplete input conditions.

Condition	nRMSE (%)
Original test data	10.89
+5% Gaussian noise	11.34
10% missing windows	11.52

The results indicate that the forecasting framework remains relatively stable under both perturbation scenarios. Although forecasting accuracy slightly degrades in the presence of noisy or incomplete inputs, the increase in nRMSE remains limited, confirming the robustness of the proposed approach under realistic measurement imperfections and monitoring disturbances commonly encountered in practical PV installations.

Overall, these results demonstrate that accurate and robust short-term PV power forecasting can be achieved using high-resolution real-world data. While recurrent neural network models achieve slightly lower error metrics, the TCN architecture offers a competitive alternative with reduced architectural complexity, smoother temporal behavior, and favorable inference characteristics. These properties make the TCN model particularly attractive for deployment in constrained computational environments and for integration within embedded energy management systems, as discussed in the following section.

5. Embedded Implementation and Performance Evaluation on Raspberry Pi

5.1. Embedded Platform and Deployment Strategy

The embedded implementation of the proposed EMS was carried out on Raspberry Pi 4, selected as a representative low-cost edge-computing platform widely adopted in building energy management and industrial Internet of Things (IoT) applications. This

platform offers a favorable compromise between computational capability, energy efficiency, and deployment cost, making it well suited for real-time EMS operation at the building and campus scale.

The system runs on a Yocto-based optimized Linux operating system, specifically configured to minimize background services and improve execution determinism. Such a lightweight operating environment is essential to ensuring stable long-term performance and predictable timing behavior in real-time embedded EMS applications.

The EMS software stack is implemented in Python 3 for system integration and control logic, while on-device forecasting inference is performed using the TensorFlow Lite runtime. All embedded inference experiments were conducted using CPU-only execution on the Raspberry Pi 4 ARM Cortex-A72 processor. No GPU, TPU, NPU, or external hardware acceleration mechanisms were employed during either forecasting inference or EMS execution. This configuration was intentionally selected to evaluate the practical deployability of the proposed forecasting framework under realistic low-cost edge-computing constraints. The EMS operates with a fixed control interval of $\Delta t = 5$ min, which is consistent with both the PV data sampling resolution and the EMS decision-making horizon. Within each control interval, PV forecasting, EMS decision logic, and data logging are executed sequentially in a deterministic manner.

Two forecasting models were deployed and evaluated under identical embedded conditions:

- An LSTM model, retained as a high-accuracy reference model.
- A fixed-parameter TCN-fixed, selected for embedded deployment.

Although LSTM models demonstrated strong offline forecasting accuracy, their recurrent structure and inherently sequential execution resulted in higher computational overhead and increased inference latency on embedded hardware. In contrast, the convolutional architecture of the TCN enables parallel computation on ARM-based processors, leading to reduced inference latency and improved computational efficiency.

To further optimize embedded performance, the trained TCN-fixed model was converted to TensorFlow Lite format using FP16 precision and directly integrated into the EMS execution loop on the Raspberry Pi. This deployment strategy ensures that forecasting, control decisions, and export constraint enforcement can be executed reliably within each 5-min control cycle.

For reproducibility purposes, the implementation environment and embedded deployment settings adopted throughout the experiments are summarized in Table 10. These parameters include the software framework, training configuration, deployment environment, and execution settings used during forecasting model development and embedded implementation.

Table 10. Embedded platform and implementation/reproducibility configuration.

Parameter	Specification
Embedded device	Raspberry Pi 4 (ARM Cortex-A72)
Operating system	Yocto-based optimized Linux
Operating system kernel	Linux kernel 6.1 LTS
Programming language	Python 3
Deep learning framework	TensorFlow 2.18 (Keras API)
Machine learning library	Scikit-learn 1.6
AI inference engine	TensorFlow Lite
Deployment mode	CPU-only execution
Optimizer	Adam

Table 10. Cont.

Parameter	Specification
Initial learning rate	1×10^{-3}
Batch size	64
Maximum epochs	120
Early stopping	Patience = 8
Learning-rate reduction	Patience = 4
Random seed	42
Chronological train–test split	80% training/20% testing

5.2. Embedded Performance Analysis

The embedded performance of the EMS was evaluated in terms of CPU utilization, RAM usage, and inference latency. All measurements were collected on the Raspberry Pi 4 during real-time EMS operation under CPU-only execution conditions. CPU utilization and RAM consumption were monitored continuously during 24 h experimental runs, and the values reported in Table 11 correspond to average resource utilization over the monitoring period. Inference latency was measured as the execution time required to perform a single forecasting inference, excluding data acquisition, communication, and input/output operations. The reported latency values correspond to average inference latency, while the worst-case latency values represent the maximum observed latency during experiments. Figure 8 presents the CPU utilization profiles over a continuous 24 h period for both summer and winter operating conditions, comparing the TCN-fixed and LSTM implementations.

Table 11. Accuracy–resource trade-off for embedded deployment on Raspberry Pi 4.

Model	nRMSE (%)	CPU Usage (%)	RAM Usage (MB)	Inference Latency (ms)	Worst-Case Latency (ms)
LSTM	10.26	41.3	512	184	258
GRU	10.40	37.8	468	151	208
TCN-fixed	10.89	24.5	286	73	101
PI-LSTM	10.18	46.8	540	201	289

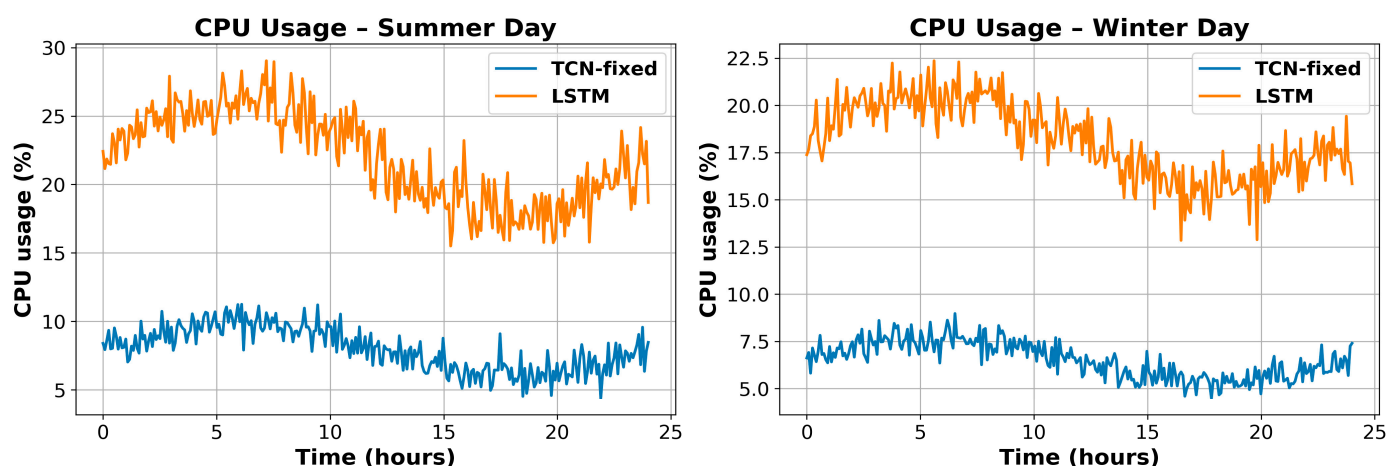


Figure 8. CPU utilization over 24 h for summer and winter scenarios, comparing TCN-fixed and LSTM models executed on a Raspberry Pi 4.

The results show stable CPU usage throughout the day for both models, with short periodic peaks corresponding to forecasting and EMS decision-making tasks. The TCN-based implementation consistently exhibits lower CPU utilization than the LSTM-based model, providing a larger computational safety margin and reducing the risk of processor saturation during continuous operation.

Memory usage profiles for both seasonal scenarios are illustrated in Figure 9.

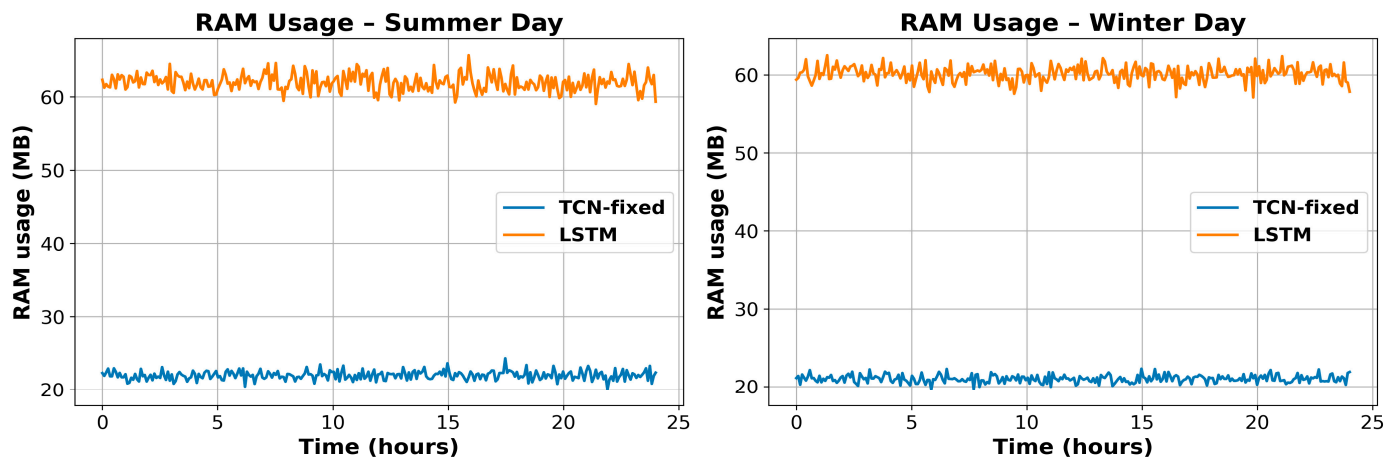


Figure 9. RAM usage over 24 h for summer and winter scenarios, comparing TCN-fixed and LSTM models.

RAM usage remains stable over time with no evidence of memory leaks or progressive growth. The TCN-fixed model requires significantly less memory than the LSTM model, confirming its suitability for long-term embedded deployment on resource-constrained hardware.

Inference latency is a critical performance metric for real-time EMS applications, as delayed predictions may compromise system responsiveness. Figure 10 compares the inference latency distributions of the TCN-fixed and LSTM models under summer and winter operating conditions.

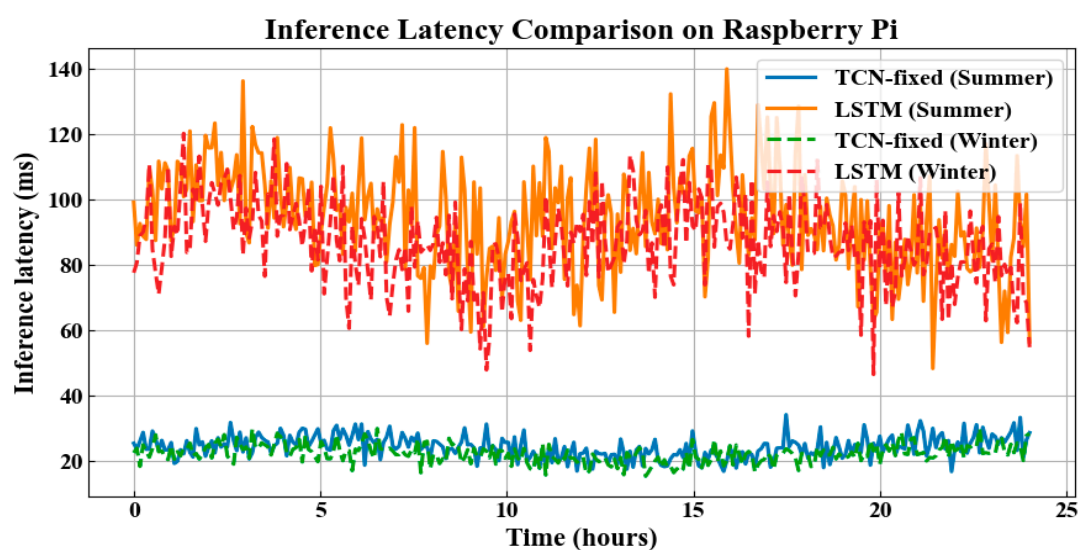


Figure 10. Inference latency comparison for TCN-fixed and LSTM models under summer and winter scenarios.

Both models satisfy the real-time constraints imposed by the EMS control interval ($\Delta t = 5$ min). However, the TCN-fixed model consistently achieves lower and more predictable inference latency than the LSTM model across both seasonal scenarios. This behavior is primarily attributed to the parallelizable convolutional operations of TCN architecture, in contrast to the inherently sequential execution of recurrent neural networks.

Although the LSTM model achieves the best forecasting accuracy, the TCN-fixed model was selected for embedded implementation because it provides a more favorable trade-off between prediction accuracy and computational resource consumption. To quantify this trade-off, Table 11 compares the forecasting accuracy and embedded resource utilization of the LSTM, GRU, TCN-fixed, and PI-LSTM models on the Raspberry Pi 4 platform.

As shown in Table 12, the TCN-fixed model provides the most efficient embedded performance, achieving the lowest CPU usage, memory consumption, and inference latency among all evaluated models. In contrast, PI-LSTM introduces additional computational overhead due to the physics-informed loss formulation, resulting in higher resource consumption compared with both LSTM and TCN-fixed.

Table 12. Model complexity comparison.

Model	Trainable Parameters	Model Size (MB)
LSTM	18,310	0.070
GRU	14,022	0.053
TCN-fixed	25,190	0.096
PI-LSTM	18,310	0.071

Although PI-LSTM shares the same base architecture as LSTM, the inclusion of the physics-informed loss does not significantly affect the model size. However, it introduces additional computational overhead during both training and inference due to the evaluation of the physics-based constraint.

6. Integration into Energy Management System

6.1. EMS Architecture, Decision Logic, and Mathematical Formulation

The proposed EMS is designed for real-time operation in a grid-connected university campus equipped with a photovoltaic (PV) generation system [34]. Its main objectives are to maximize local PV self-consumption, minimize grid imports, and ensure compliance with grid-imposed export constraints. The EMS follows a lightweight rule-based control architecture suitable for embedded implementation and operates with a fixed control interval of $\Delta t = 5$ min. At each time step $t = k\Delta t$, the EMS performs discrete control actions based on instantaneous measurements, short-term PV forecasts, and cumulative energy constraints.

The choice of a lightweight rule-based EMS was motivated by the real-time and resource-constrained nature of the targeted embedded platform. Although optimization-based approaches such as MPC, MILP, and Reinforcement Learning can provide advanced energy management capabilities, they generally involve higher computational complexity and more demanding tuning procedures. In contrast, the proposed strategy ensures deterministic execution, low computational overhead, and reliable operation within the fixed 5-min EMS control interval implemented on the Raspberry Pi platform.

The EMS evaluation focuses on February and July as representative extreme seasonal conditions. February reflects a low-irradiance winter scenario, while July represents a high-irradiance summer scenario with frequent export constraint activation. These operating conditions were selected to stress-test the EMS under both deficit- and surplus-dominated

regimes. Future work will extend the evaluation to full-year operation, including transitional seasons and resilience under communication disturbances.

At each control step k , the EMS acquires instantaneous measurements of PV generation and campus electrical demand, denoted by $P_{PV}(k)$ and $P_{Load}(k)$, respectively. The total campus load is decomposed into critical and flexible components as:

$$P_{Load}(k) = P_{Crit}(k) + P_{Flex}(k) \quad (16)$$

where $P_{Crit}(k)$ represents non-deferable loads and $P_{Flex}(k)$ corresponds to controllable loads managed by the EMS.

The instantaneous power balance between generation and demand is evaluated as:

$$P_{Surplus}(k) = P_{PV}(k) - P_{Load}(k) \quad (17)$$

This variable constitutes the core decision criterion of the EMS and allows distinguishing between surplus operation ($P_{Surplus}(k) > 0$) and deficit operation ($P_{Surplus}(k) \leq 0$).

Under surplus conditions, the EMS prioritizes local PV self-consumption by activating flexible loads within their operational limits. The remaining surplus power after flexible load activation is computed as:

$$P_{Surplus,rest}(k) = P_{Surplus}(k) - P_{Flex}(k) \quad (18)$$

The cumulative PV energy production since the beginning of the year is updated at each control step as:

$$E_{PV,y}(k) = E_{PV,y}(k-1) + P_{PV}(k) \frac{\Delta t}{60} \quad (19)$$

The maximum authorized exported energy is defined as a fraction of the cumulative PV energy:

$$E_{Export,max}(k) = \alpha E_{PV,y}(k), \quad \alpha = 0.3 \quad (20)$$

The cumulative exported energy is updated as:

$$E_{Export,y}(k) = E_{Export,y}(k-1) + P_{Export}(k) \frac{\Delta t}{60} \quad (21)$$

The remaining export margin available at each control step is therefore:

$$E_{Quota,rest}(k) = \max(0, E_{Export,max}(k) - E_{Export,y}(k)) \quad (22)$$

If the remaining export quota is positive, grid export is enabled and limited by the available surplus, the technical grid constraint, and the remaining annual quota:

$$P_{Export}(k) = \min(P_{Surplus,rest}(k), P_{Grid}^{max} \frac{60}{\Delta t} E_{Quota,rest}(k)) \quad (23)$$

Otherwise, surplus curtailment (non-remunerated injection) is applied.

When the system operates in deficit mode, grid import ensures supply continuity:

$$P_{Import}(k) = \max(P_{Load}(k) - P_{PV}(k), 0) \quad (24)$$

In deficit operation, the EMS exploits short-term PV power forecasts that are recomputed at every 5-min control interval. At each control step k , a new multi-step forecast over a prediction horizon H (e.g., $H = 6$, corresponding to 30 min) is generated and used in a rolling-horizon manner to anticipate near-future operating conditions. An imminent

surplus condition is identified when at least one forecasted step within the prediction horizon satisfies:

$$\exists h \in \{1, \dots, H\} \text{ such that } \hat{P}_{PV}(k+h) > P_{Load}(k+h) \quad (25)$$

When this condition is satisfied, selected non-critical loads are deferred to the next control interval. Importantly, this forecast-based mechanism does not alter the instantaneous power balance or grid import at time step k ; it only affects future load scheduling decisions, thereby increasing PV self-consumption without compromising critical load supply.

For supervision and regulatory compliance, the EMS continuously monitors the export ratio defined as:

$$E_{Export,ratio}(k) = \frac{E_{Export,y}(k)}{\max(E_{PV,y}(k), \varepsilon)} \quad (26)$$

where ε is a small positive constant introduced to avoid division by zero.

6.2. Campus Electrical Equipment and Load Classification

The university campus electrical demand is composed of heterogeneous equipment with different operational constraints and flexibility levels. Table 13 summarizes the main equipment categories considered by the EMS, together with their rated power, operating duration, and daily energy consumption.

Table 13. Electrical equipment classification and daily energy consumption.

Category	Equipment	Rated Power (kW)	Operating Time (h/Day)	Energy (kWh/Day)	Flexibility
Critical loads	Data center & network	20	24	480	None
	Security systems (CCTV)	3	24	72	None
	Emergency lighting	2	24	48	None
Semi-critical loads	Classroom lighting	10	8	80	Low
	Elevators	6	4	24	Low
Flexible loads	HVAC systems	20	6	120	Medium
	Computer labs	10	4	40	High
	Electric water heaters	5	2	10	Medium
	Refrigeration units	4	6	24	Medium
Total		≈ 898 kWh/day			

The EMS flowchart (Figure 11) provides an intuitive representation of the real-time decision logic, explicitly distinguishing instantaneous power management from cumulative annual energy constraints.

6.3. EMS Operational Performance and Impact Assessment

The operational performance of the proposed EMS is evaluated through monthly simulations conducted over 30 consecutive days under two contrasted seasonal conditions: February (winter) and July (summer). The assessment relies on four key performance indicators commonly used to quantify PV integration efficiency at the local level: the self-consumption rate (SC), the self-sufficiency rate (SS), the export ratio (ER), and the curtailment ratio (CR).

$$SC = \frac{E_{PV}^{cons}}{E_{PV}^{gen}} \quad (27)$$

$$SS = \frac{E_{PV}^{cons}}{E_{load}} \quad (28)$$

$$ER = \frac{E_{PV}^{exp}}{E_{PV}^{gen}} \quad (29)$$

$$CR = \frac{E_{PV}^{curt}}{E_{PV}^{gen}} \quad (30)$$

where E_{PV}^{gen} denotes the total PV energy generated, E_{PV}^{cons} the fraction of PV energy directly consumed on-site, E_{PV}^{exp} the energy exported to the grid, E_{PV}^{curt} the curtailed PV energy, and E_{load} the total campus electrical demand over the considered period.

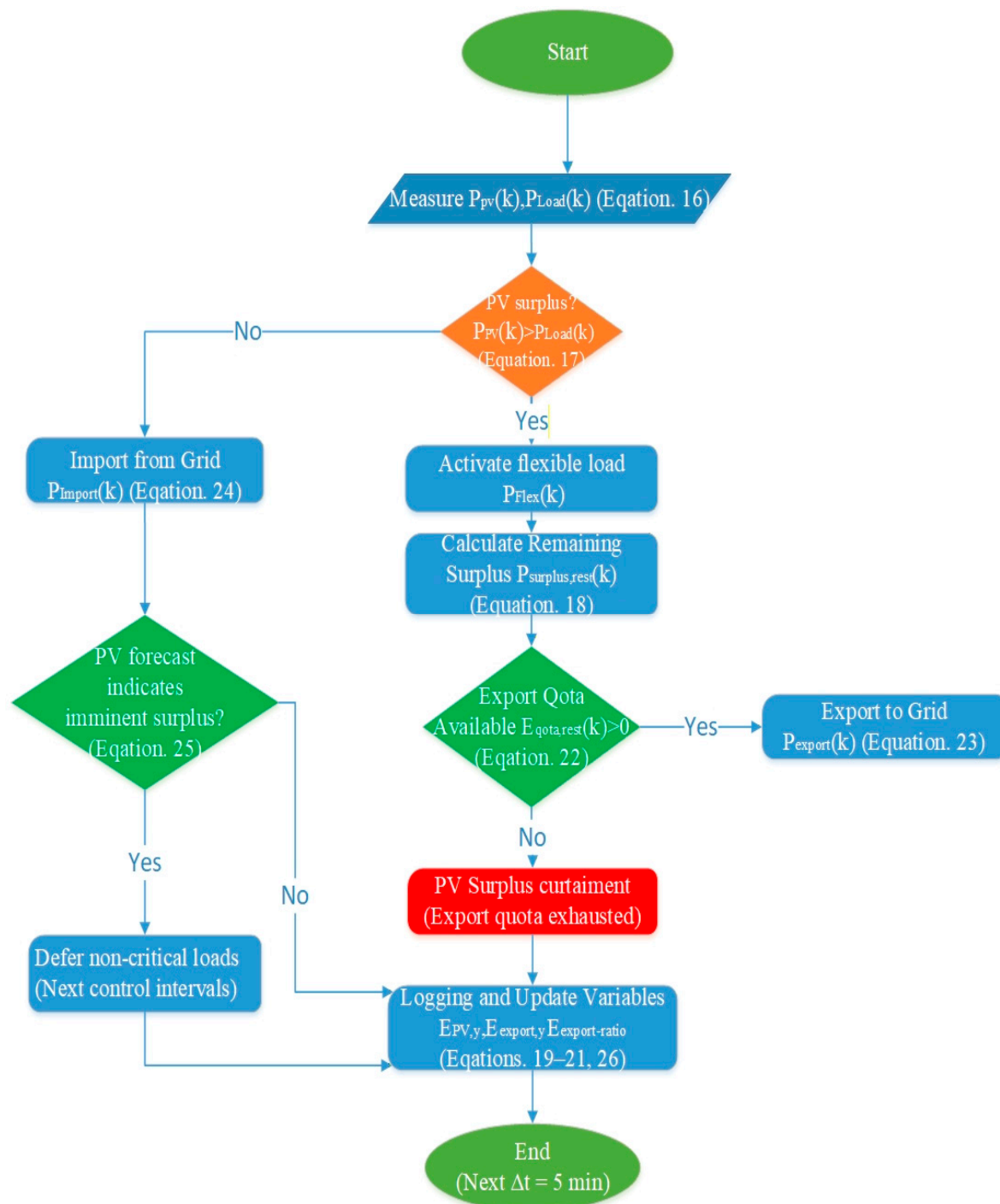


Figure 11. Flowchart of the proposed embedded energy management system (EMS).

The resulting quantitative performance indicators are summarized in Table 10, while the corresponding demand reshaping behavior and PV power allocation mechanisms are illustrated in Figures 12–15. All results explicitly account for the regulatory constraint limiting PV exports to 30% of the cumulative PV energy generated (annual export quota), reflecting realistic grid-integration conditions.

Mean Daily PV Generation and Load Profiles (July)

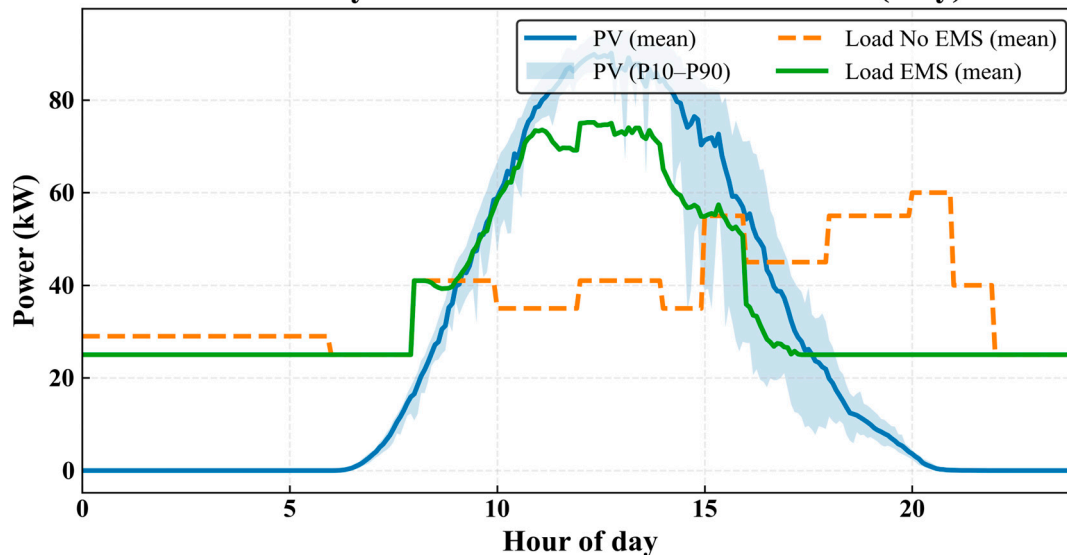


Figure 12. Average campus load profiles with and without EMS compared to mean PV generation in July. The shaded area represents the P10–P90 range of PV production over 30 days.

Mean Daily PV Generation and Load Profiles (February)

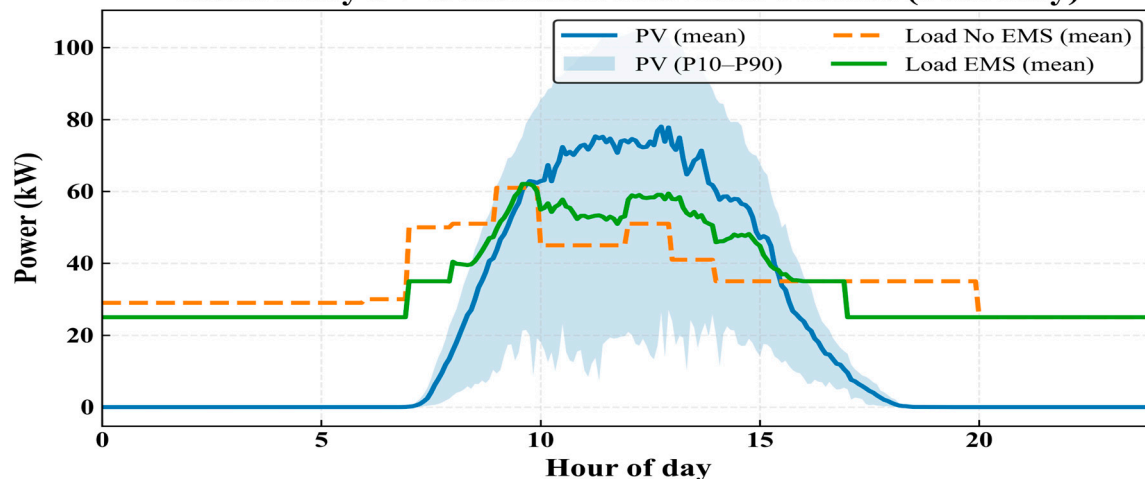


Figure 13. Average campus load profiles with and without EMS compared to mean PV generation in February, including PV production variability over the same period.

As reported in Table 14, the introduction of the EMS significantly improves photovoltaic self-consumption (SC) in both seasons. In July, characterized by high solar availability, SC increases from $61.21\% \pm 2.37\%$ without EMS to $84.56\% \pm 3.68\%$ with EMS operation, representing an absolute gain of more than 23% points. In February, despite lower PV production and higher variability, SC is also enhanced from $69.50\% \pm 13.29\%$ to $81.50\% \pm 10.33\%$.

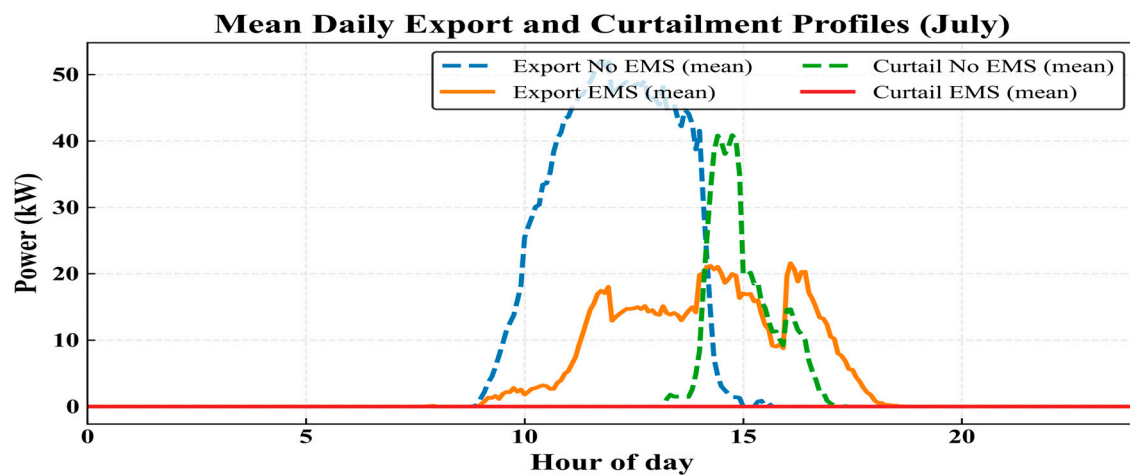


Figure 14. Mean PV export and curtailment profiles in July with and without EMS under a 30% export quota, averaged over 30 days.

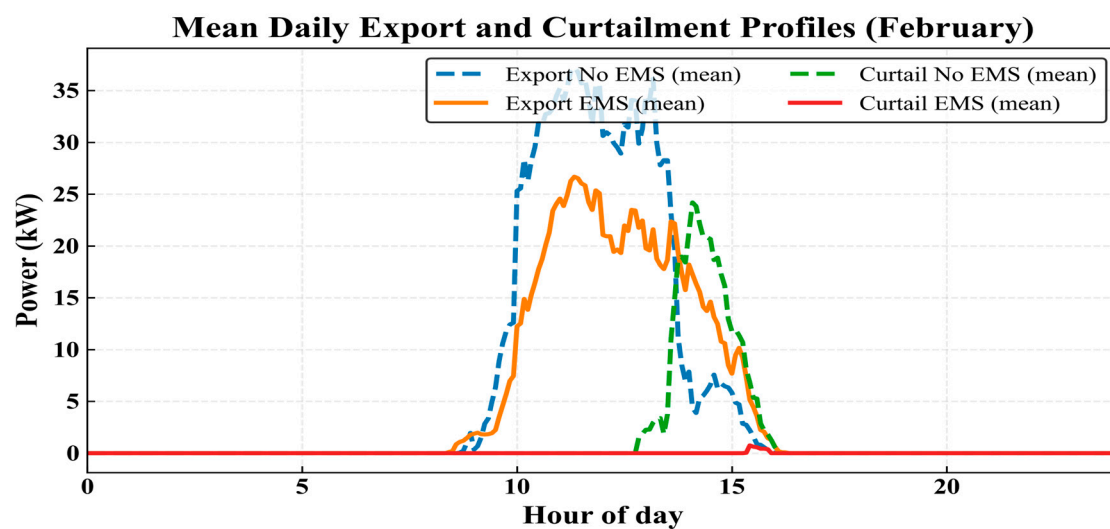


Figure 15. Mean PV export and curtailment profiles in February with and without EMS under the same export constraint.

Table 14. Monthly performance indicators of the campus PV system with and without EMS (mean $\pm$ std over 30 days).

Month	Case	SC (%)	SS (%)	ER (%)	CR (%)	Grid Import (kWh/Day)	PV Export (kWh/Day)	Curtailment (kWh/Day)
February	Without EMS	69.50 $\pm$ 13.29	37.00 $\pm$ 11.13	24.67 $\pm$ 9.31	5.83 $\pm$ 5.07	540.53 $\pm$ 95.51	135.56 $\pm$ 66.95	35.27 $\pm$ 31.14
	Proposed EMS (30% quota)	81.50 $\pm$ 10.33	45.53 $\pm$ 13.34	18.45 $\pm$ 10.29	0.04 $\pm$ 0.21	444.15 $\pm$ 78.20	106.62 $\pm$ 68.69	0.27 $\pm$ 1.42
July	Without EMS	61.21 $\pm$ 2.37	44.19 $\pm$ 3.80	30.00 $\pm$ 0.00	8.79 $\pm$ 2.37	501.19 $\pm$ 34.12	194.80 $\pm$ 18.79	57.73 $\pm$ 17.74
	Proposed EMS (30% quota)	84.56 $\pm$ 3.68	61.37 $\pm$ 3.02	15.41 $\pm$ 3.68	0.00 $\pm$ 0.00	343.70 $\pm$ 22.17	102.13 $\pm$ 31.72	0.00 $\pm$ 0.00

Impact of Forecast-Assisted EMS Operation

To isolate the direct contribution of the PV forecasting module, an additional EMS operating scenario is introduced. This scenario, referred to as reactive EMS, uses the same

rule-based decision logic and export limitation mechanism as the proposed EMS, but it does not use short-term PV forecasts. Decisions are therefore based only on instantaneous PV generation and load measurements. In contrast, the proposed forecast-assisted EMS uses the 30-min PV power forecast in a rolling-horizon manner to anticipate upcoming surplus conditions and schedule flexible loads accordingly.

This comparison isolates the direct contribution of the forecasting module from the general benefits associated with rule-based EMS operation. The results show that the forecast-assisted EMS provides additional gains compared with the reactive EMS, particularly during high-PV periods. By anticipating future surplus generation, the proposed approach improves the temporal alignment between PV production and flexible demand, leading to higher self-consumption, lower grid import, and reduced curtailment.

Table 15 summarizes the comparative performance of the reactive EMS and the proposed forecast-assisted EMS under the imposed export constraints.

Table 15. Comparison between reactive EMS and proposed EMS operation under export constraints.

Month	EMS Strategy	SC (%)	Grid Import (kWh/Day)	Curtailment (kWh/Day)
February	Reactive EMS	76.8 ± 11.2	476.3 ± 84.5	8.9 ± 6.3
	Forecast-assisted EMS	81.50 ± 10.33	444.15 ± 78.20	0.27 ± 1.42
July	Reactive EMS	78.9 ± 4.1	389.4 ± 28.7	12.8 ± 6.2
	Forecast-assisted EMS	84.56 ± 3.68	343.70 ± 22.17	0.00 ± 0.00

The results demonstrate that integrating short-term PV forecasting into the EMS provides measurable operational improvements compared with purely reactive control. In February, forecast-assisted EMS increases PV self-consumption by approximately 4.7% points while reducing grid import by about 32 kWh/day. In July, characterized by higher PV surplus conditions, the improvement reaches approximately 5.6% points, with a reduction in grid import of nearly 46 kWh/day. These results indicate that the benefits of forecast-assisted operation are more pronounced during periods of high PV generation, whereas smaller differences are observed under lower PV production conditions.

These improvements are primarily driven by the load-reshaping strategies implemented by the EMS. As shown in Figures 12 and 13, flexible loads are systematically shifted toward periods of higher PV availability around midday. This redistribution improves the temporal alignment between local generation and demand without increasing total energy consumption, thereby maximizing direct PV utilization.

The improved PV-load alignment under EMS operation leads to a substantial reduction in grid energy imports. According to Table 15, average daily grid imports in July decrease from 501.19 ± 34.12 kWh/day to 343.70 ± 22.17 kWh/day, corresponding to a relative reduction of approximately 31%. In February, grid imports are reduced by about 18%, from 540.53 ± 95.51 kWh/day to 444.15 ± 78.20 kWh/day.

In parallel, the EMS significantly reduces PV export volumes. In July, average PV exports drops from 194.80 ± 18.79 kWh/day without EMS to 102.13 ± 31.72 kWh/day with EMS activation. This confirms that a larger fraction of the PV generation is consumed locally rather than injected into the grid. This behavior is clearly illustrated in Figures 14 and 15, where export peaks are attenuated and temporally smoothed when the EMS is active.

Under the 30% export quota, PV curtailment becomes significant in the absence of coordinated energy management, particularly during summer. As shown in Table 11, average curtailment in July reaches 57.73 ± 17.74 kWh/day without EMS. With EMS operation, curtailment is eliminated throughout the simulated period.

In February, curtailment levels are inherently lower due to reduced PV generation; however, the EMS still reduces curtailment from 35.27 ± 31.14 kWh/day to nearly zero (0.27 ± 1.42 kWh/day). These results demonstrate the robustness of the proposed EMS across contrasting seasonal conditions and confirm its ability to maintain compliance with export constraints while minimizing renewable energy losses.

The combined analysis of Table 15 and Figures 12–15 highlights the clear seasonal dependency of EMS benefits. During summer months, the EMS delivers pronounced gains by mitigating export saturation, eliminating curtailment, and substantially reducing grid imports. In winter, although PV surplus is less frequent, the EMS still provides consistent improvements through enhanced PV self-consumption and demand alignment.

Overall, these results confirm that a lightweight, rule-based EMS driven by short-term PV forecasting can effectively enhance campus-level energy performance while respecting regulatory constraints. The observed improvements are achieved without relying on complex optimization solvers, reinforcing the practical feasibility and scalability of the proposed approach for real-world deployment.

7. Conclusions

This study presents an integrated framework for short-term PV power forecasting and embedded energy management tailored to prosumer-scale systems operating under grid export constraints. The proposed approach was validated using one year of high-resolution (5-min) real-world data collected from a 111 kW rooftop PV installation.

A direct multi-step forecasting strategy with a 30-min prediction horizon was evaluated using baseline, tree-based, and deep learning models. Among them, the LSTM achieved the highest forecasting accuracy in terms of nRMSE ($\approx 10.26\%$), while the PI-LSTM further improved physically consistent predictions by embedding physics-based constraints directly into the learning process.

Although the TCN-fixed model exhibited slightly lower predictive accuracy, it provided significantly reduced CPU utilization, memory consumption, and inference latency, making it particularly suitable for real-time embedded deployment.

The forecasting models were integrated into a lightweight rule-based Energy Management System (EMS) and deployed on a Raspberry Pi 4 using TensorFlow Lite. Comparative EMS results showed that the proposed forecast-assisted strategy increased PV self-consumption by more than 20 percentage points, reduced grid imports by up to 31%, and virtually eliminated PV curtailment while respecting a 30% export limitation.

Although the framework was validated on a single university-campus PV installation, the results may be influenced by site-specific operational and climatic conditions. Nevertheless, the proposed architecture remains transferable to other grid-connected PV systems through appropriate retraining and parameter adaptation.

Overall, the proposed framework effectively bridges accurate short-term PV forecasting and practical embedded energy management, offering a scalable, low-cost, and regulation-compliant solution for smart buildings, prosumer energy communities, and campus-scale energy systems under increasing renewable penetration and grid export constraints.

Author Contributions: Conceptualization, A.M. and N.M.; methodology, A.M.; software, A.M.; validation, A.M., N.M. and S.N.; formal analysis, A.M., N.M. and S.N.; investigation, A.M. and N.M.; resources, N.M.; data curation, A.M.; writing—original draft preparation, A.M. and N.M.; writing—review and editing, A.M., N.M. and S.N.; visualization, A.M. and A.L.; supervision, J.C.V. and A.C.; project administration, A.M., N.M. and A.C.; funding acquisition, A.M. and N.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: Data available and source code in a publicly accessible repository. The source code presented in the study is openly available in GitHub at <https://github.com/aymenmnassri1/energy-forecast-lstm> (accessed on 24 June 2026). The experiments were conducted using version v1.0 (release ID: af5de03), ensuring reproducibility of the results.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

ABC	Artificial Bee Colony
ARIMA	Autoregressive Integrated Moving Average
AI	Artificial Intelligence
CNN	Convolutional Neural Network
CPU	Central Processing Unit
CR	Curtailement Ratio
DRL	Deep Reinforcement Learning
ELM	Extreme Learning Machines
EMS	Energy management system
ER	Export Ratio
FP16	16-bit Floating Point
GA	Genetic Algorithm
HVAC	Heating, Ventilation and Air Conditioning
HEMS	Home energy management system
IoT	Internet of Things
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MPPT	Maximum Power Point Tracking
nRMSE	normalized Root Mean Square Error
PSO	Particle Swarm Optimization
PV	Photovoltaic
RAM	Random Access Memory
RF	Random Forest
RL	Reinforcement Learning
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SC	Self-Consumption
SCADA	Supervisory Control And Data Acquisition
SS	Self-Sufficiency
SVM	Support Vector Machines
TCN	Temporal Convolutional Network
TFLite	TensorFlow Lite
VSC	Voltage Source Converter
XGBoost	Extreme Gradient Boosting

References

1. Mansouri, N.; Lashab, A.; Ali, M.; Bouchoucha, C.; Guerrero, J.; Vasquez, J. Reserve Technique in Integrating Large Sustainable Energy Sources: A Case Study of the Tunisian Grid. *Sustainability* **2024**, *16*, 10791. [CrossRef]
2. Gu, Z.; Li, B.; Zhang, G.; Li, B. Optimizing photovoltaic integration in grid management via a deep learning-based scenario analysis. *Sci. Rep.* **2025**, *15*, 14851. [CrossRef] [PubMed]
3. Di Leo, P.; Ciocia, A.; Malgaroli, G.; Spertino, F. Advancements and Challenges in Photovoltaic Power Forecasting: A Comprehensive Review. *Energies* **2025**, *18*, 2108. [CrossRef]
4. IEA PVPS. *Trends in Photovoltaic Applications 2024: Survey Report of Selected IEA Countries Between 1992 and 2023*; Report IEA-PVPS T1-15:2006, IEA PVPS Task 1; IEA Press: Paris, France, 2024.

5. Araujo, M.; Aguilar, S.; Souza, R.; Cyrino Oliveira, F. Global Horizontal Irradiance in Brazil: A Comparative Study of Reanalysis Datasets with Ground-Based Data. *Energies* **2024**, *17*, 5063. [\[CrossRef\]](#)
6. Al-Dahidi, S.; Madhilarasan, M.; Al-Ghussain, L.; Abubaker, A.; Ahmad, A.; Alrbai, M.; Aghaei, M.; Alahmer, H.; Alahmer, A.; Baraldi, P.; et al. Forecasting Solar Photovoltaic Power Production: A Comprehensive Review and Innovative Data-Driven Modeling Framework. *Energies* **2024**, *17*, 4145. [\[CrossRef\]](#)
7. Ejub Che, E.; Roland Abeng, K.; Iweh, C.D.; Tsekouras, G.J.; Fopah-Lele, A. The impact of integrating variable renewable energy sources into grid-connected power systems: Challenges, mitigation strategies, and prospects. *Energies* **2025**, *18*, 689. [\[CrossRef\]](#)
8. Liu, X.; Gou, Z. Hybrid forecasting and optimization framework for residential photovoltaic-battery systems: Integrating data-driven prediction with multi-strategy scenario analysis. *Build. Simul.* **2025**, *18*, 1587–1609. [\[CrossRef\]](#)
9. Howlader, A.S. Machine learning applications in renewable energy: Predictive analytics for solar cell performance optimization and energy yield forecasting. *Am. J. Sch. Res. Innov.* **2025**, *4*, 392–427. [\[CrossRef\]](#)
10. Dave, R.; Vajpai, J. A Review on Solar Power Generation Forecasting Methods. *Trans. Indian Natl. Acad. Eng.* **2025**, *11*, 1–24. [\[CrossRef\]](#)
11. Gallardo, I.; Amor, D.; Gutiérrez, Á. Recent Trends in Real-Time Photovoltaic Prediction Systems. *Energies* **2023**, *16*, 5693. [\[CrossRef\]](#)
12. Mahmoud, A.A.; Albadry, O.A.; Mohamed, M.I.; El-Khozondar, H.; Nassar, Y.; Hafez, A.A. Charging systems/techniques of electric vehicle. *Sol. Energy Sustain. Dev. J.* **2024**, *13*, 18–44. [\[CrossRef\]](#)
13. Qazi, S. Fundamentals of standalone photovoltaic systems. In *Standalone Photovoltaic (PV) Systems for Disaster Relief and Remote Areas*; Qazi, S., Ed.; Elsevier: Amsterdam, The Netherlands, 2017; pp. 31–82.
14. Nayar, C.V.; Islam, S.M.; Dehbonei, H.; Tan, K.; Sharma, H. Power electronics for renewable energy sources. In *Alternative Energy in Power Electronics*; Rashid, M.H., Ed.; Butterworth-Heinemann: Boston, MA, USA, 2011; pp. 1–79.
15. Yousef, L.A.; Yousef, H.; Rocha-Meneses, L. Artificial intelligence for management of variable renewable energy systems: A review of current status and future directions. *Energies* **2023**, *16*, 8057. [\[CrossRef\]](#)
16. Szczepaniuk, H.; Szczepaniuk, E.K. Applications of artificial intelligence algorithms in the energy sector. *Energies* **2023**, *16*, 347.
17. Raghavendra, K.V.G.; Zeb, K.; Muthusamy, A.; Krishna, T.N.V.; Kumar, S.V.S.V.P.; Kim, D.-H.; Kim, M.-S.; Cho, H.-G.; Kim, H.-J. A comprehensive review of DC–DC converter topologies and modulation strategies with recent advances in solar photovoltaic systems. *Electronics* **2019**, *9*, 31. [\[CrossRef\]](#)
18. Yousri, D.; Abd Elaziz, M.; Oliva, D.; Abualigah, L.; Al-qaness, M.A.A.; Ewees, A.A. Reliable applied objective for identifying simple and detailed photovoltaic models using modern metaheuristics: Comparative study. *Energy Convers. Manag.* **2020**, *223*, 113279. [\[CrossRef\]](#)
19. Shoaie, M.; Noorollahi, Y.; Hajinezhad, A.; Moosavian, S.F. A review of the applications of artificial intelligence in renewable energy systems: An approach-based study. *Energy Convers. Manag.* **2024**, *306*, 118207. [\[CrossRef\]](#)
20. Sankhwar, P. Energy reduction in residential housing units. *Int. J. Adv. Res.* **2024**, *12*, 667–672. [\[CrossRef\]](#)
21. Olatunde, T.M.; Okwandu, A.C.; Akande, D.O.; Sikhakhane, Z.Q. The impact of smart grids on energy efficiency: A comprehensive review. *Eng. Sci. Technol. J.* **2024**, *5*, 1257–1269. [\[CrossRef\]](#)
22. Ghiasi, M.; Niknam, T.; Dehghani, M.; Baghaee, H.R.; Wang, Z.; Ghanbarian, M.M.; Blaabjerg, F.; Dragicevic, T. Multipurpose FCS model predictive control of VSC-based microgrids for islanded and grid-connected operation modes. *IEEE Syst. J.* **2022**, *17*, 2558–2569.
23. Ghiasi, M.; Niknam, T.; Dehghani, M.; Siano, P.; Haes Alhelou, H.; Al-Hinai, A. Optimal multi-operation energy management in smart microgrids in the presence of res based on multi-objective improved de algorithm: Cost-emission based optimization. *Appl. Sci.* **2021**, *11*, 3661. [\[CrossRef\]](#)
24. Qayyum, N.; Jamil, U.; Mahmood, A. Optimizing energy cost in the residential sector through home energy management systems in a smart grid environment. *Energy Explor. Exploit.* **2025**, *43*, 1647–1664. [\[CrossRef\]](#)
25. Mnassri, A.; Mansouri, N.; Nasri, S.; Lashab, A.; Vasquez, J.C.; Cherif, A. An Intelligent Framework for Energy Forecasting and Management in Photovoltaic-Integrated Smart Homes in Tunisia with V2H Support Using LSTM Optimized by the Harris Hawks Algorithm. *Energies* **2025**, *18*, 5635. [\[CrossRef\]](#)
26. Li, Y.; Wang, R.; Li, Y.; Zhang, M.; Long, C. Wind power forecasting considering data privacy protection: A federated deep reinforcement learning approach. *Appl. Energy* **2023**, *329*, 120291. [\[CrossRef\]](#)
27. Mostofa, K.Z.; Islam, M.A. Creation of an Internet of Things (IoT) system for the live and remote monitoring of solar photovoltaic facilities. *Energy Rep.* **2023**, *9*, 422–427. [\[CrossRef\]](#)
28. Zikria, Y.B.; Ali, R.; Afzal, M.K.; Kim, S.W. Next-Generation Internet of Things (IoT): Opportunities, Challenges, and Solutions. *Sensors* **2021**, *21*, 1174. [\[CrossRef\]](#) [\[PubMed\]](#)
29. López-Vargas, A.; Fuentes, M.; Vivar, M. Current challenges for the advanced mass scale monitoring of Solar Home Systems: A review. *Renew. Energy* **2021**, *163*, 2098–2114. [\[CrossRef\]](#)

30. Pawar, P.; TarunKumar, M.; Vittal K., P. An IoT based Intelligent Smart Energy Management System with accurate forecasting and load strategy for renewable generation. *Measurement* **2020**, *152*, 107187. [[CrossRef](#)]
31. González, I.; Calderón, A.J.; Portalo, J.M. Innovative Multi-Layered Architecture for Heterogeneous Automation and Monitoring Systems: Application Case of a Photovoltaic Smart Microgrid. *Sustainability* **2021**, *13*, 2234. [[CrossRef](#)]
32. Sumorek, M.; Idzkowski, A. Time Series Forecasting for Energy Production in Stand-Alone and Tracking Photovoltaic Systems Based on Historical Measurement Data. *Energies* **2023**, *16*, 6367. [[CrossRef](#)]
33. Radzi, P.N.L.M.; Mekhilef, S.; Shah, N.M.; Akhter, M.N.; Seyedmahmoudian, M.; Almalaq, Y.; Wahyudie, A. Enhancing short-term solar power forecasting using a hybrid deep learning approach and a comparative analysis of machine learning models. *Energy Convers. Manag. X* **2025**, *29*, 101431. [[CrossRef](#)]
34. Sethu, S.S. SCADA-Based EMS Simulation for Minimising Grid Export and Maximising PV Self-Consumption for the HelioS Energy Hub. Master's Thesis, University of Skövde, Skövde, Sweden, 2025.

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.