Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 6, 2026

BLOCKCHAIN-READY LEAKAGE-AWARE MACHINE LEARNING FRAMEWORK FOR SHORT-TERM SOLAR AC POWER FORECASTING AND ENERGY DATA INTEGRITY VERIFICATION

Fahad Soomro[*1], Syeda Tehreem Naqvi[2], Abdul Wahid Memon[3], Bilal Ahmed Shaikh[4]

[*1,2,4]*Department of Computer Science, DHA Suffa University, Karachi, Pakistan*
[3]*Computer Systems Engineering Department, Q.U.E.S.T, Nawabshah*

[*1]fahad.soormo@dsu.edu.pk, [2]tehreem.naqvi@dsu.edu.pk, [3]awam@quest.edu.pk, [4]bilal.shaikh@dsu.edu.pk

DOI:https://doi.org/10.5281/zenodo.20700249

Keywords
Solar power forecasting; photovoltaic systems; leakage-aware machine learning; Extra Trees; smart energy management; blockchain-ready verification; SHA-256; energy data integrity; tamper detection; IoT-enabled energy systems

Article History
Received: 16 April 2026
Accepted: 28 May 2026
Published: 15 June 2026

Corresponding Author: *
Fahad Soomro

Abstract
For reliable smart-energy management, short-term (ST) photovoltaic (PV) power forecasting is crucial, as is the exchange of energy data with accurate trustworthiness. The output of PV ACs is influenced by irradiation, environment and module temperature, and short-term generation behavior, and distributed energy records require integrity verification against unauthorized modification. This paper presents a blockchain-ready and leakage-aware framework, for solar AC power forecasting of the next step and provides a verification of the integrity over the energy records. The forecasting part forecasts the next-step AC power based on the weather, temporal and lag features derived from the PV generation data and the weather sensor data. Following the pre-processing and feature engineering steps, the final data set consisted of 68,708 records from 22 inverter/source units with 54,966 records split into a training set and 13,742 records split into a test set through a chronological split. The tested regression models were: Linear Regression, Ridge Regression, Random Forest, Extra Trees, Gradient Boosting and XGBoost. To reduce direct inverter-side data leakage, in the main forecasting experiment, the power from the DC side was omitted. Extra Trees achieved the best performance with MAE = 12.8138, RMSE = 37.1821, MAPE = 3.8496%, and R^2 = 0.991146. A separate inverter-aware estimation experiment with DC power was retained only to demonstrate the strong electrical dependency between DC-side and AC-side PV power. For integrity verification, the best forecasting outputs were converted into hash-secured records containing plant ID, source key, timestamp, actual AC power, predicted AC power, error value, and SHA-256 hash. A total of 2,000 records were stored in the verification layer, and all 100 intentionally modified records were detected, achieving a 100% tamper detection rate. The results show that leakage-aware solar AC forecasting can be coupled with lightweight, blockchain-ready record verification in a reproducible workflow.

1. INTRODUCTION

The increasing integration of solar photovoltaic (PV) systems into smart energy infrastructure has made accurate power forecasting essential for reliable energy monitoring, planning, and operational control. Solar PV generation is intermittent because power output is affected by irradiation, ambient temperature, module temperature, and time-dependent operating conditions [1]. The growing use of PV systems has increased the need for forecasting methods that support data-driven energy management.

Data-driven and deep learning-based forecasting methods have shown strong potential for PV power prediction because they can handle complex relationships between weather inputs and PV output [2]. Comprehensive studies further note that forecasti ng reliability depends on preprocessing quality, th e selected prediction model, and the evaluation fr amework used [3]. In real-world energy systems, short-term forecasting helps operators respond to changes in weather and power generation more effectively [4].

Machine learning methods have been widely used for PV forecasting by using weather-related inputs and past generation data. Comparative studies indicate that prediction accuracy mainly depends on the chosen algorithm and the quality of input features. [5]. Research on hybrid physical and machine learning forecasting suggests that PV prediction can be improved when electrical behavior is considered along with data-driven patterns [6]. This forms the foundation for a controlled design which distinguishes between practical forecasting and direct electrical conversion estimation.

It is essential to deal with the time variability of the solar generation data to ensure a proper short-term PV forecasting. Previous works have demonstrated that performing feature selection and enhancing the capacity of recurrent learning can help in probabilistic short-term PV forecasting [7]. The importance of learning temporal dependencies in solar power generation is further shown using hybrid CNN-LSTM approaches [8]. But a number of forecasting studies are still concerned with accuracy measures and fail to pay much attention to the reliability of energy records following prediction.

In the case of Distributed Smart Energy, the actual and predicted Solar data is exchanged between monitoring systems, Grid Operators and Energy Management Platforms. Previous research indicates that blockchain and AI can be combined in smart grids to enable trustworthy energy services and secure information exchange [9]. Transparent management of energy records and automated energy operations with the use of smart contracts have also been investigated [10]. Data security, integrity and trusted communication are also emphasized as requirements for smart grid systems in the literature [11].

Due to these issues, this paper presents a machine learning-based forecasting system with a blockchain framework for short-term solar AC power forecasting and energy records verification. The forecasting part forecasts the next-step AC power based on weather, temporal and lag-based features derived from PV generation and weather sensor data. The verification element records the actual and forecasted energy values in a SHA-256 hash-secured record so that unauthorized modifications can be detected during integrity verification.

The experimental dataset used in this study contains PV generation and weather sensor records. After preprocessing and feature engineering, the final dataset contained 68,708 records from 22 inverter/source units. A chronological split was used, with 54,966 records for training and 13,742 records for testing. The main forecasting experiment excluded DC power to avoid direct inverter-side leakage, while an additional inverter-aware experiment included DC power only to analyze the electrical relationship between DC-side PV generation and AC-side grid output.

Six regression models were evaluated: Linear Regression, Ridge Regression, Random Forest, Extra Trees, Gradient Boosting, and XGBoost. In the main forecasting experiment, Extra Trees achieved the best performance with MAE = 12.8138, RMSE = 37.1821, MAPE = 3.8496%, and R^2 = 0.991146. For energy record verification, 2,000 hash-secured records were generated, and all 100 intentionally modified records were detected, achieving a 100% tamper detection rate.

The main contribution of this work is a reproducible forecasting and verification workflow that combines next-step solar AC power prediction with lightweight record-level integrity checking. Unlike approaches that only emphasize forecasting accuracy, this study separates realistic forecasting without DC power from inverter-aware DC-to-AC estimation. The proposed workflow also demonstrates that actual and predicted solar energy records can be verified without storing the complete sensor dataset in the blockchain-style layer.

2. Related Work

Solar PV forecasting research has evolved from conventional statistical techniques toward machine learning, deep learning, and hybrid prediction approaches. Spatio-temporal graph neural networks have been used for multi-site PV forecasting by modeling spatial and temporal dependencies across solar plants [12]. Meta-learning and numerical-weather-prediction-independent LSTM models have also been explored for short-term PV forecasting under varying weather conditions [13].

Attention-based learning has become an important direction in PV forecasting because it can identify relevant temporal and feature patterns from complex input streams. A dual-stream attention network has been proposed for PV power forecasting to improve representation learning from solar-related inputs [14]. Double-layer decomposition with WOA-BiLSTM-Attention has also been used for point-interval PV forecasting while considering weather classification [15].

Meteorological inputs remain important in solar forecasting because irradiation and weather conditions directly affect PV output. Comparative studies of global and regional downscaled numerical weather prediction models show that weather input quality influences irradiance and PV power forecasting performance [16]. Ensemble numerical weather prediction combined with physical model chains has also been applied for probabilistic PV forecasting [17].

Transfer learning and deep neural networks have been used to improve PV forecasting where plant-specific data may be limited. Day-ahead PV prediction using transfer learning has shown the potential of reusing learned patterns across forecasting scenarios [18]. Dual-stream CNN-LSTM architectures have also been applied to combine spatial and temporal information for solar power prediction [19]. These studies support the broader trend of using data-driven models for solar forecasting, while also showing that model complexity should be matched with the available data and research objective.

Comparative machine learning studies are important because no single model performs best across all solar datasets. A recent evaluation of multiple machine learning models for solar power generation prediction showed that dataset-specific model comparison is necessary for reliable selection [20]. Hybrid convolutional, LSTM, and attention frameworks have further shown that short-term PV forecasting can benefit from feature interaction and temporal sequence learning [21]. Self-attention mechanisms have also been used to improve the generalizability of time-series PV forecasting in distributed urban PV systems [22].

Transformer-based models represent another recent direction in solar forecasting. Operational day-ahead PV forecasting using a transformer variant has shown the ability of attention-based architectures to model complex temporal dependencies [23]. Multi-step solar power prediction using transformer structures further demonstrates the growing role of advanced deep learning in renewable energy forecasting [24]. Multi-scale transformer designs such as CGA former have also been introduced for short-term PV forecasting through enhanced feature representation [25].

Solar-related forecasting is also connected with irradiance prediction because solar irradiance directly affects PV generation. Artificial neural network models have been developed for direct normal irradiance forecasting using operational numerical weather prediction data [26]. Broader reviews of solar forecasting techniques confirm that artificial intelligence has become central to modern solar energy prediction [27]. Near real-time PV forecasting using recurrent neural networks and open-access data further supports the feasibility of reproducible forecasting pipelines based on available measurement sources [28].

Blockchain and AI integration is also receiving attention in smart energy systems alongside forecasting research. Existing reviews suggest that combining AI with blockchain can support secure data exchange, intelligent operational support, and trusted services in the energy domain [29]. A bibliometric assessment further confirms that AI, IoT, blockchain, and big data are increasingly being studied together in renewable energy-focused power systems [30].

Compared with the reviewed studies, this paper focuses on a compact and implementation-oriented workflow rather than a high-complexity deep learning architecture or a full-scale blockchain deployment. The main forecasting experiment excludes DC power to reduce direct electrical leakage, while the additional inverter-aware experiment uses DC power only to analyze the physical DC-to-AC relationship. The blockchain-style verification layer is intentionally lightweight and uses SHA-256 hashing for record-level integrity checking of actual and predicted solar energy values.

Table 1. Critical comparison of recent PV forecasting and blockchain-enabled energy studies.

Ref.	Study	Dataset	Method & Algorithm	Limitation
[1]	Iheanetu, 2022 — Solar PV forecasting review	Published PV forecasting literature	Review of forecasting methods	Identifies meteorological and temporal variability as major drivers of PV forecasting error; does not implement secure verification of generated energy records.
[2]	Yu et al., 2024 — Deep learning for PV forecasting	Recent deep-learning PV forecasting studies	Review of DL-based PV models	Shows that deep models capture nonlinear PV patterns; blockchain-based integrity of actual and predicted records remains outside the scope.
[3]	Al-Dahidi et al., 2024 — Data-driven PV production forecasting	PV forecasting datasets and published modeling frameworks	Comprehensive review and data-driven framework	Emphasizes preprocessing, model selection, and evaluation metrics; security and tamper detection of forecasting records are not directly addressed.
[5]	Markovics and Mayer, 2022 — ML comparison for PV forecasting	PV forecasting data supported by numerical weather prediction inputs	Comparative machine learning models	Confirms that forecasting performance depends on model and input selection; does not include record-level blockchain verification.
[6]	Mayer, 2022 — Physical and ML hybridization	PV forecasting datasets with physical-model information	Hybrid physical and machine learning forecasting	Highlights the benefit of combining PV-domain knowledge with ML; predicted records are not secured against modification.
[7]	Liu et al., 2022 — Short-term probabilistic PV prediction	PV generation and weather-related variables	Feature selection with improved LSTM	Improves short-term probabilistic forecasting; focuses on uncertainty rather than integrity verification of forecasting outputs.
[8]	Agga et al., 2022 — Short-term PV production prediction	PV production time-series data	CNN-LSTM hybrid model	Captures temporal behavior for short-term PV prediction; no blockchain or hash-based verification layer is included.

[12]	Simeunovic et al., 2022 — Multi-site PV forecasting	Multi-site PV systems	Spatio-temporal graph neural network	Models spatial and temporal dependencies across PV sites; requires more complex graph modeling and does not secure energy records.
[14]	Khan et al., 2023 — Attention-based PV forecasting	PV forecasting dataset	Dual-stream attention network	Improves feature representation for PV forecasting; focuses on prediction accuracy without energy-record integrity checking.
[20]	Aldosari et al., 2024 — Solar farm power prediction	Grid-connected 300 MW solar farm data	Multiple machine learning models	Supports dataset-specific comparison of ML models; blockchain-style verification of actual and predicted values is not included.
[22]	Yu et al., 2024 — Distributed PV time-series prediction	Urban distributed PV system data	Self-attention mechanism	Improves generalizability in distributed PV forecasting; trusted exchange of forecast records is not addressed.
[23]	Tao et al., 2024 — Operational day-ahead PV forecasting	Operational PV forecasting data	Transformer variant	Captures long-range temporal dependencies; model complexity is higher and secure record verification is not considered.
[28]	La Fata et al., 2025 — Near real-time PV forecasting	Open-access meteorological and PV measurement data	Recurrent neural network	Supports reproducible near real-time forecasting; remains forecasting-focused without tamper detection of records.
[9]	Hua et al., 2022 — AI and blockchain for smart grids	Smart grid and prosumer literature	Review of AI-blockchain applications	Identifies AI and blockchain as enabling smart-grid technologies; does not implement PV forecasting with record-level verification.
[10]	Kirli et al., 2022 — Smart contracts in energy systems	Energy-system smart contract literature	Systematic review of smart contracts	Shows potential of smart contracts for transparent energy operations; PV forecasting outputs are not integrated with data integrity checking.
[29]	Al Shareef et al., 2024 — AI-blockchain integration in energy	Energy-sector AI-blockchain literature	Review-based analysis	Highlights secure data exchange and intelligent decision support; lacks an implementation-level PV forecasting and hash-verification workflow.
[30]	Jaramillo et al., 2025 — AI, IoT, blockchain, and big data in renewable energy	Bibliometric data on renewable energy-oriented power systems	Bibliometric assessment	Confirms convergence of AI, IoT, blockchain, and big data; does not provide a compact experimental PV forecasting and verification pipeline.

Gap synthesis. The reviewed literature shows that PV forecasting studies mainly focus on improving predictive performance through machine learning, recurrent networks, attention mechanisms, transformers, or hybrid forecasting models. In contrast, blockchain-related energy studies mainly focus on smart contracts, secure exchange, energy trading, or broad AI-blockchain integration. The present study addresses the intersection of these directions by combining next-step solar AC power forecasting with SHA-256-based record-level verification of actual and predicted energy values. It also separates the realistic forecasting experiment without DC power from an inverter-aware DC-to-AC estimation experiment, which reduces overstatement of forecasting performance.

3. Proposed Methodology

This study proposes a lightweight AI and blockchain-ready verification framework for short-term solar AC power forecasting and secure energy record verification. The framework uses solar PV generation and weather sensor data to predict next-step AC power. The predicted and actual energy values are then converted into hash-secured records to verify data integrity and detect tampering. The overall workflow of the proposed framework is shown in Fig. 1.

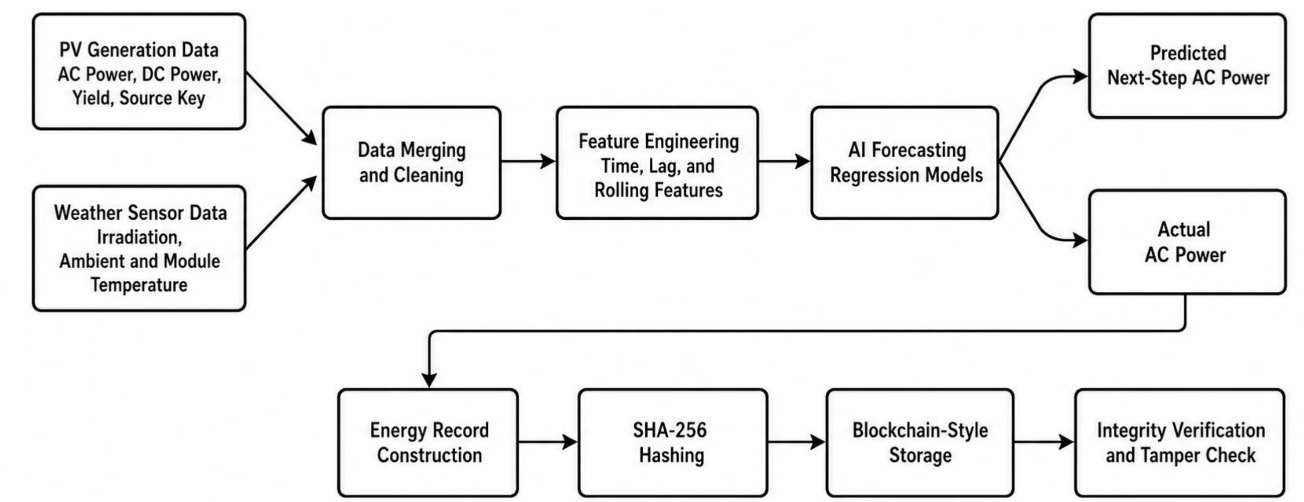

Fig. 1. Proposed AI-blockchain framework for solar AC power forecasting and secure energy record verification.

The proposed framework consists of four main stages: data preprocessing, feature engineering, AI-based forecasting, and hash-based integrity verification. First, PV generation data and weather sensor data are merged using timestamp and plant identifier. The merged data are cleaned, sorted chronologically, and transformed into a forecasting dataset using lag and rolling features. Second, machine learning models are trained to predict next-step AC power using weather, temporal, and historical generation features. Third, the best forecasting output is combined with the actual AC power value to generate energy records. Finally, each energy record is converted into a SHA-256 hash and stored in a verification layer for integrity checking and tamper detection.

3.1 Dataset and Experimental Design

The dataset contains PV generation records and weather sensor measurements. The generation data include inverter/source-level power parameters, while the weather data include irradiation, ambient temperature, and module temperature. After preprocessing and feature engineering, the final dataset contained 68,708 records collected from 22 inverter/source units. The data covered the period from 15 May 2020 00:30 to 17 June 2020 23:30. Because solar power generation is time-dependent, a chronological train-test split was used instead of a random split. The training set contained 54,966 records, while the testing set contained 13,742 records.

Table 2. Dataset summary and experimental settings.

Item	Value
Total records after preprocessing	68,708
Training records	54,966
Testing records	13,742
Inverter/source units	22
Data period	15 May 2020 00:30 to 17 June 2020 23:30
Main target variable	Next-step AC power
Main experiment	Forecasting without DC power
Additional experiment	Inverter-aware estimation with DC power
Blockchain records used	2,000
Tampered records tested	100

3.2 Feature Engineering and Forecasting Formulation

The main forecasting experiment used 14 input features: ambient temperature, module temperature, irradiation, temperature difference, hour, day, month, weekday, AC power lag-1, AC power lag-2, AC power rolling mean, irradiation lag-1, irradiation rolling mean, and module temperature lag-1. DC power was excluded from the main forecasting experiment because it has a direct electrical relationship with AC power and may lead to overly optimistic results. Therefore, the main experiment was formulated as a leakage-aware next-step AC power forecasting task, where features available at time t are used to predict AC power at time t+1.

The forecasting target was defined in (1):

(1) $y(t+1) = AC\_POWER(t+1)$

where y(t+1) denotes the next-step AC power predicted using feature values available at time t.

Six regression models were trained and compared: Linear Regression, Ridge Regression, Random Forest, Extra Trees, Gradient Boosting, and XGBoost. The prediction error for each testing sample was calculated in (2):

(2) $e_i = y_i - \hat{y}_i$

where y_i is the actual AC power, $\hat{y}_i$ is the predicted AC power, and e_i is the prediction error.

The forecasting models were evaluated using MAE, RMSE, MAPE, and R^2, as defined in (3)-(6):

(3) $MAE = (1/n) \Sigma |y_i - \hat{y}_i|$

(4) $RMSE = \sqrt{[(1/n) \Sigma (y_i - \hat{y}_i)^2]}$

(5) $MAPE = (100/n) \Sigma |(y_i - \hat{y}_i) / y_i|$

(6) $R^2 = 1 - [\Sigma (y_i - \hat{y}_i)^2 / \Sigma (y_i - \bar{y})^2]$

where n is the number of testing samples and $\bar{y}$ is the mean value of the actual AC power.

3.3 Hash-Based Blockchain-Ready Energy Record Verification

After forecasting, the predicted AC power values were converted into hash-secured energy records. Each record contained the plant identifier, source key, timestamp, actual AC power, predicted AC power, prediction error, and SHA-256 hash. The purpose of this layer was not to store the complete sensor dataset or to claim full blockchain deployment, but to provide a lightweight record-level integrity mechanism that can be integrated with a blockchain or smart-contract platform in a future implementation.

Table 3. Blockchain energy record structure.

Field	Description
Plant ID	Solar plant identifier
Source key	Inverter/source unit identifier
Timestamp	Date and time of the energy record
Actual AC power	Observed AC power value
Predicted AC power	Forecasted AC power value
Error value	Difference between actual and predicted AC power
Hash	SHA-256 hash of the complete energy record

The hash value for each record was generated in (7):

(7) $H_i = \text{SHA-256}(R_i)$

where R_i is the complete energy record and H_i is its generated hash value.

During verification, the hash of the received record was recalculated and compared with the stored hash. The verification rule was defined in (8):

(8) V_i = Valid, if $H_i = H'_i$; otherwise Tampered, if $H_i \neq H'_i$

where H_i is the stored hash, H'_i is the recalculated hash, and V_i is the verification status of the record. If any field in the energy record is modified, the recalculated hash changes and the record is identified as tampered. In this study, 2,000 blockchain-style records were stored, and 100 modified records were used to evaluate tamper detection.

4. Results and Discussion

In this section, the experimental results of the proposed leakage-aware forecasting and hash-based energy record verification workflow are reported. Results are presented in four components: a) forecasting without DC power, b) estimating using inverter aware DC power methods, c) feature contribution analysis, and d) record-level integrity verification. To prevent direct inverter-side leakage and to evaluate the next-step solar AC power forecast based on weather, time-related and lag-based features, the main forecasting setup excludes the direct feed of DC power.

4.1 Solar AC Power Forecasting Performance

Table 4 shows the forecasting performance of the six regression model groups in the base case (DC-excluded) experiment and the three groups in the inverter-aware estimation experiment. In the main forecasting experiment, Extra Trees achieved the lowest error with MAE = 12.8138, RMSE = 37.1821, MAPE = 3.8496%, and R^2 = 0.991146. Random Forest produced comparable performance with MAE = 12.8994, RMSE = 37.6003, MAPE = 3.8886%, and R^2 = 0.990945. Extra Trees resulted in a 36.46% reduction in RMSE when compared with XGBoost, and a 67.96% reduction in RMSE when compared with Linear Regression. The results show that ensemble tree-based learning performance is better in the selected forecasting setup for the nonlinear behavior of solar AC power than for linear baselines.

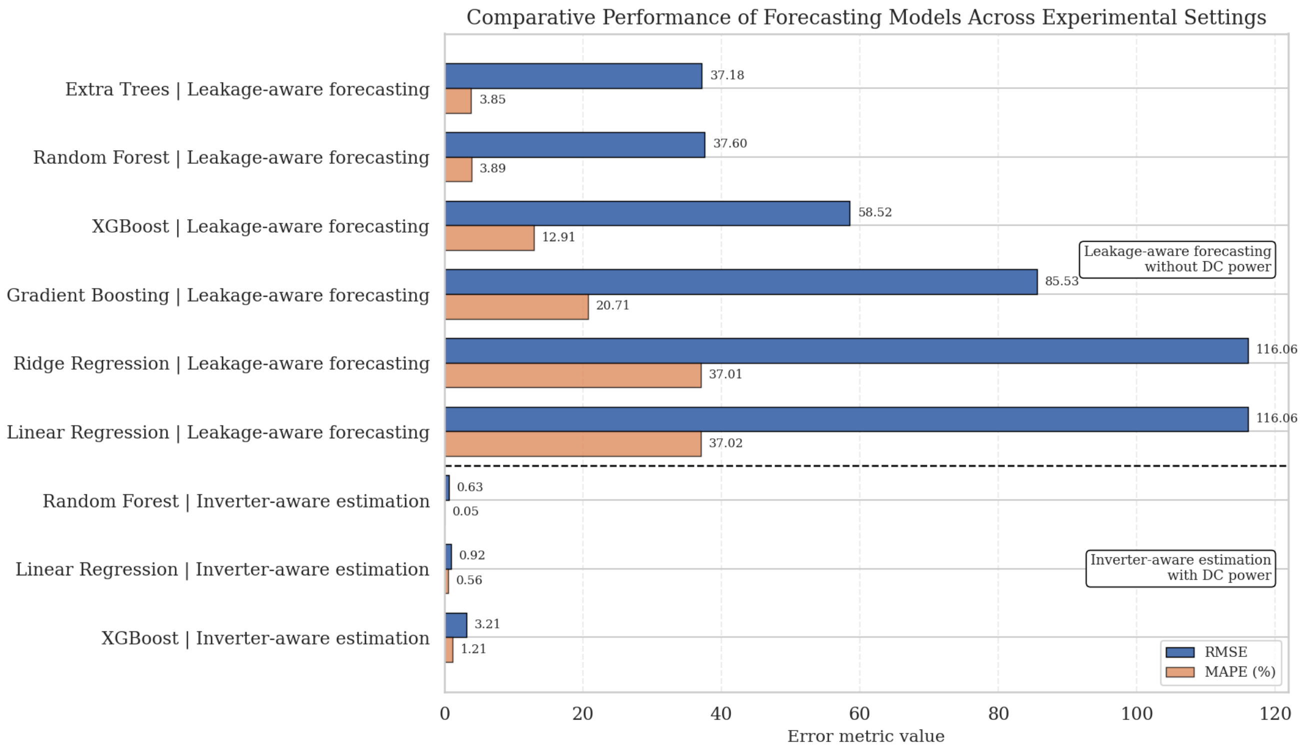

Fig. 2. Comparative RMSE and MAPE analysis of leakage-aware forecasting and inverter-aware estimation models.

Fig. 2 summarizes the error behavior of both experimental settings in a single visual form. In the leakage-aware forecasting setting, Extra Trees and Random Forest show the lowest RMSE and MAPE values, confirming that ensemble tree-based models captured the nonlinear relationship among irradiation, temperature, temporal indicators, and lagged generation features more effectively than the linear baselines. The inverter-aware estimation models show much smaller errors because DC power is directly related to AC power; therefore, this setting is interpreted only as a DC-to-AC electrical estimation analysis, while the DC-excluded setting remains the main forecasting result.

Table 4. Performance comparison of forecasting and inverter-aware estimation models.

Experiment	Model	MAE	RMSE	MAPE (%)	R^2
Leakage-aware forecasting (without DC power)	Extra Trees	12.8138	37.1821	3.8496	0.991146
Leakage-aware forecasting (without DC power)	Random Forest	12.8994	37.6003	3.8886	0.990945
Leakage-aware forecasting (without DC power)	XGBoost	29.2319	58.5204	12.9099	0.978806
Leakage-aware forecasting (without DC power)	Gradient Boosting	43.3229	85.5322	20.7053	0.953145

Leakage-aware forecasting (without DC power)	Ridge Regression	67.3251	116.0553	37.0109	0.913737
Leakage-aware forecasting (without DC power)	Linear Regression	67.3279	116.0570	37.0219	0.913735
Inverter-aware estimation (with DC power)	Random Forest	0.1277	0.6267	0.0513	0.999997
Inverter-aware estimation (with DC power)	Linear Regression	0.6450	0.9177	0.5632	0.999995
Inverter-aware estimation (with DC power)	XGBoost	1.5777	3.2104	1.2138	0.999934

The additional inverter-aware estimation experiment produced much lower error values because DC power was included as an input feature. Random Forest achieved RMSE = 0.6267 and R^2 = 0.999997 in this setting, confirming the strong electrical dependency between DC-side PV generation and AC-side grid output. Therefore, this experiment is reported as an electrical estimation analysis, while the main forecasting findings are based on the DC-excluded setting.

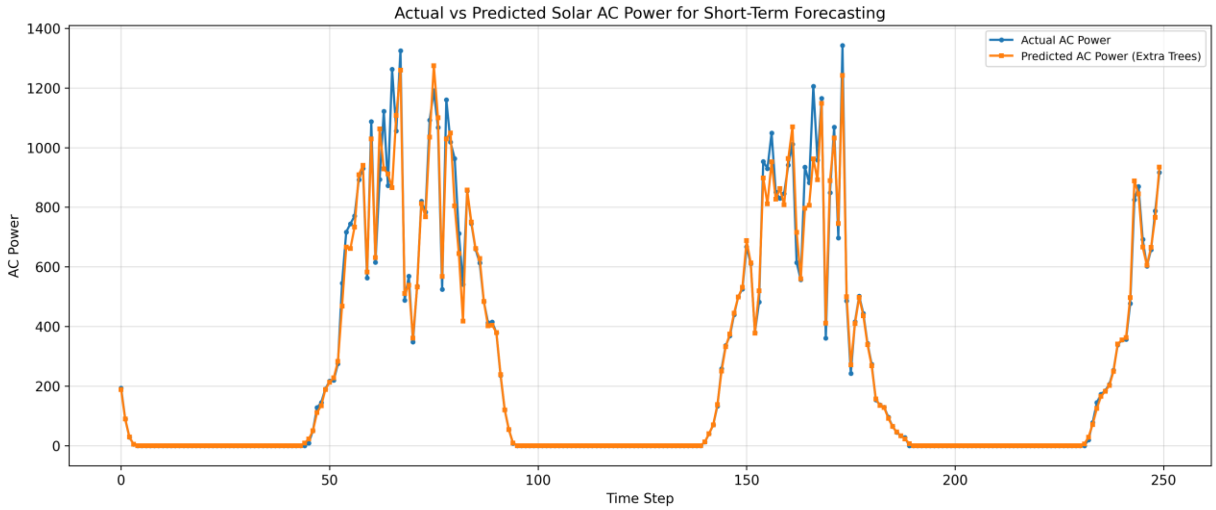

Fig. 3. Actual and predicted solar AC power using the Extra Trees forecasting model.

Fig. 3 shows that the predicted AC power closely follows the actual generation profile over low-generation, ramp-up, high-generation, and ramp-down intervals. The visible deviations mainly occur during sharp transitions, which is expected in short-term PV forecasting because sudden changes in irradiation and module behavior can create rapid power fluctuations.

4.2 Forecasting Error Distribution

To further examine the forecasting behavior of the best model, the residual error distribution was analyzed. The error distribution is concentrated near zero, indicating that most predictions remained close to the actual AC power values. A smaller number of larger positive and negative errors were observed during unstable generation intervals, which are typically associated with rapid solar variability. This diagnostic analysis supports the reliability of the Extra Trees model beyond the aggregate MAE and RMSE values.

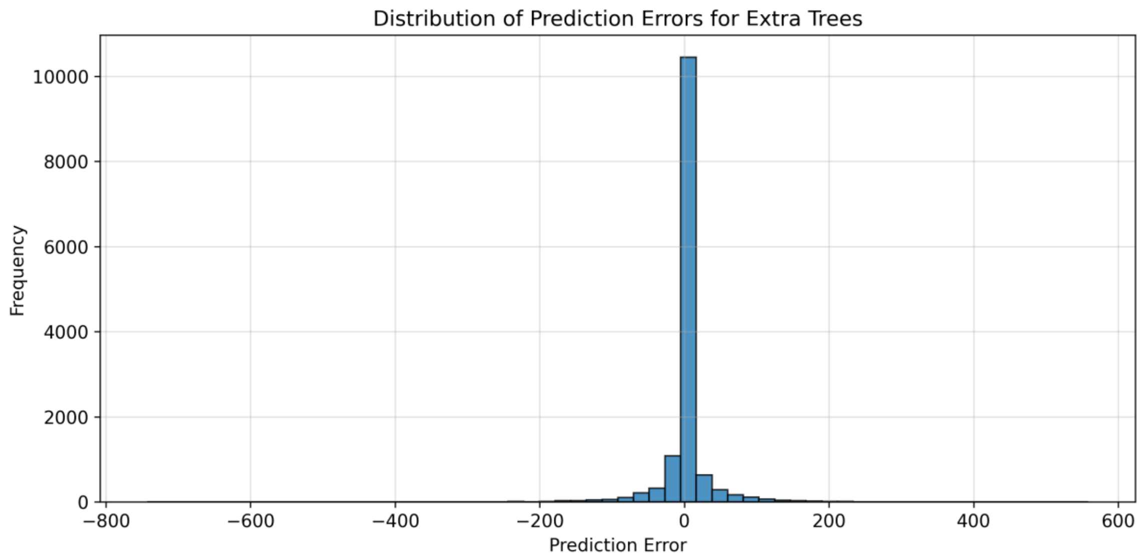

Fig. 4. Distribution of prediction errors for the Extra Trees forecasting model.

4.3 Feature Contribution Analysis

Feature contribution analysis was used to interpret the main forecasting experiment. As shown in Fig. 4, irradiation was the most influential variable with an importance score of 0.5978, followed by the 3-step rolling AC power feature (0.1693) and temperature difference (0.1179). Together, these three features accounted for approximately 88.50% of the total importance. This pattern is consistent with PV behavior because irradiation directly drives solar generation, recent AC power values represent short-term temporal continuity, and temperature difference reflects the thermal condition affecting module output.

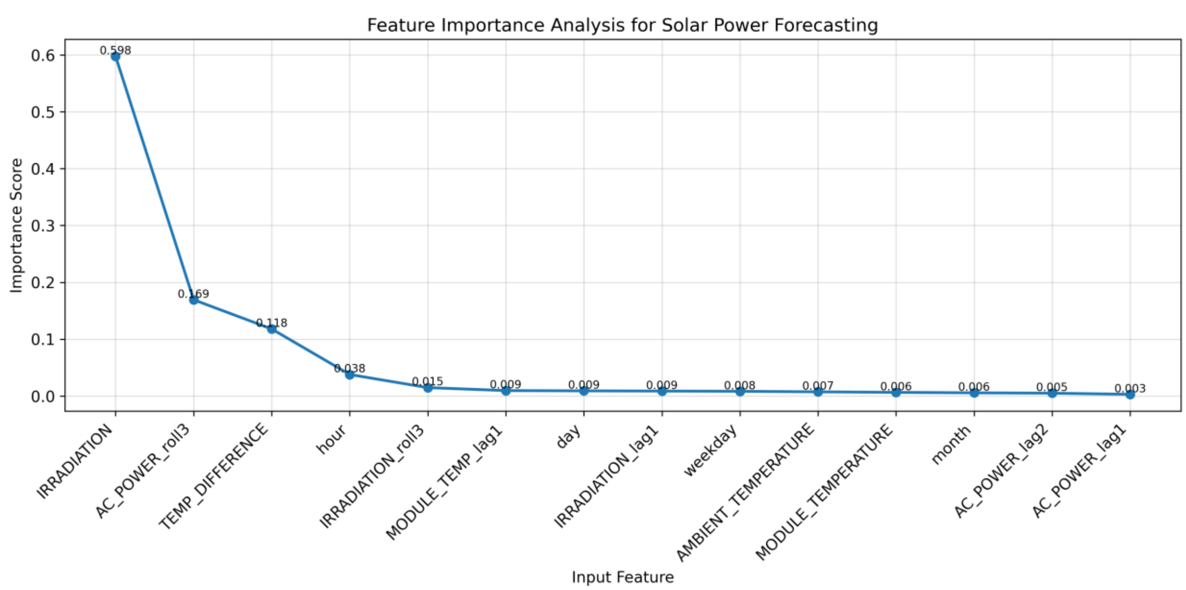

Fig. 5. Feature importance analysis for the main solar AC power forecasting experiment.

4.4 Hash-Based Energy Record Verification Results

After forecasting, the best model outputs were converted into hash-secured energy records containing plant ID, source key, timestamp, actual AC power, predicted AC power, prediction error, and SHA-256 hash. Table 5 summarizes the record storage and verification results. A total of 2,000 records were stored in 0.110651 seconds, resulting in a throughput of 18,074.8842 records per second. During verification, all 2,000 original records were successfully validated in 0.092154 seconds.

Table 5. Hash-based energy record verification performance.

Metric	Value
Stored records	2,000
Storage time in seconds	0.110651
Throughput records/second	18,074.8842
Verified original records	2,000
Verification time in seconds	0.092154
Tampered records tested	100
Tampered records detected	100
Tamper detection rate (%)	100.00

To test tamper detection, 100 records were intentionally modified by changing the actual AC power value. All 100 modified records were detected because the recalculated hash no longer matched the stored hash. Fig. 6 presents the verification and tamper detection outcome. The reported storage and verification values should be interpreted as record-level hash verification performance rather than full blockchain consensus or smart-contract latency.

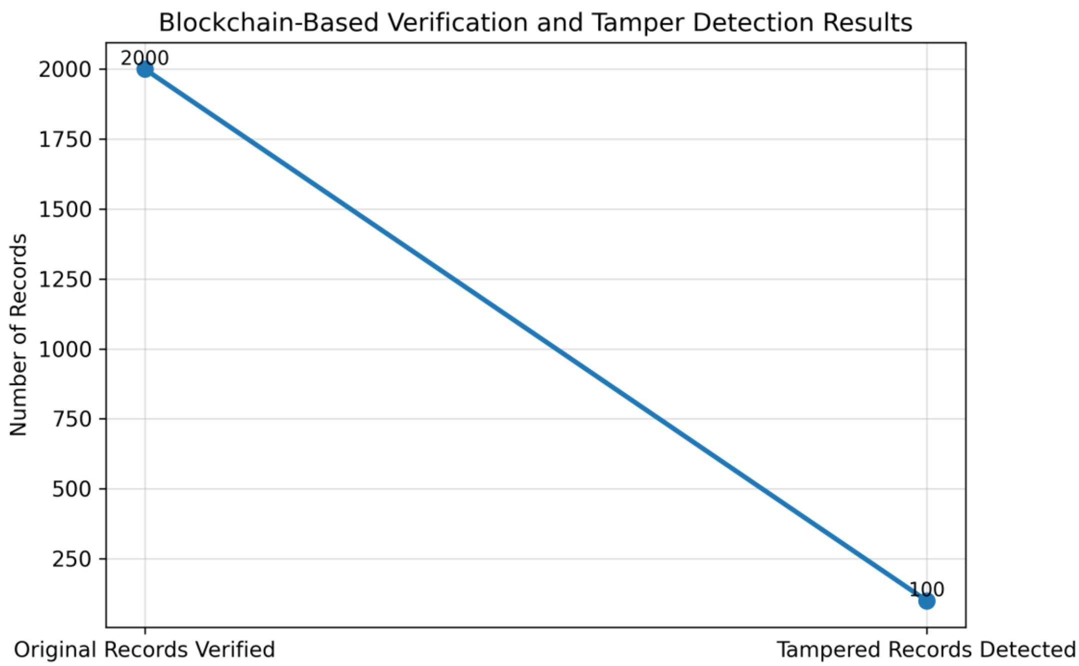

Fig. 6. Hash-based verification and tamper detection results.

4.5 Discussion

The results indicate that that the Extra Trees model yielded the best performance in the primary leakage-aware forecasting experiment. Close performance of Random Forest indicates the appropriateness of ensemble tree models in capturing the nonlinear interaction between features of irradiation, temperature, and time, and features of the generation lag. A weaker performance of linear models suggests that solar AC power prediction problem is not entirely linear, meaning that there are nonlinear patterns in the problem that are not captured exclusively by linear models.

The experiment which uses the inverter shows that the use of DC power in the main forecasting is more of an informative component than a true forecasting component, although the DC power is very useful for forecasting AC power. The separation between these two experimental settings enhances the credibility of the evaluation and mitigates the risk of over-optimistic forecasting performance.

The hash-based verification results show that solar energy records can be verified before transmission to monitoring dashboards or distributed energy management systems to ensure they have not been modified. The existing deployment is designed to be light weight and not to be a full blockchain deployment. It is designed to test a blockchain-ready record integrity layer, which can be deployed with a smart contract platform like Ethereum, Hyperledger Besu or Hyperledger Fabric.

5. Conclusion and Future Work

In this paper, a leakage-aware machine learning-based energy record integrity verification system that is connected to blockchain was presented for short-term solar AC power forecasting. The proposed solution involved both generating powers using PV and forecasting subsequent power generation using weather sensors; once the next-step AC power was predicted, it was then protected using a SHA-256 based record verification system to safeguard both the actual and the predicted energy values. The main forecasting experiment did not include the use of DC power to minimize direct inverter side leakage and also to evaluate the short-term solar forecasting realistically.

The experimental results showed that ensemble tree-based models performed better than linear and boosting-based models for the selected forecasting task. Extra Trees achieved the best forecasting performance with MAE = 12.8138, RMSE = 37.1821, MAPE = 3.8496%, and R^2 = 0.991146. Random Forest also produced comparable results with MAE = 12.8994, RMSE = 37.6003, MAPE = 3.8886%, and R^2 = 0.990945. The inverter-aware experiment confirmed the strong relationship between DC-side PV generation and AC-side grid output, where Random Forest achieved RMSE = 0.6267 and R^2 = 0.999997 when DC power was included.

The hash-based verification experiment demonstrated that predicted and actual energy records can be protected using a lightweight integrity mechanism. A total of 2,000 energy records were stored, all original records were successfully verified, and all 100 intentionally modified records were detected, achieving a 100% tamper detection rate. These findings confirm that forecasting outputs can be coupled with record-level verification to improve the reliability of data-driven solar energy management workflows.

Future work will focus on deploying the verification layer on a real smart-contract platform such as Ethereum, Hyperledger Besu, or Hyperledger Fabric. Further experiments should include longer-duration datasets, multi-site PV plants, real-time IoT data streams, source-wise validation, hyperparameter sensitivity analysis, and advanced deep learning models for improved forecasting under rapidly changing weather conditions.

References

[1] K. J. Iheanetu, "Solar photovoltaic power forecasting: A review," *Sustainability*, vol. 14, no. 24, Art. no. 17005, 2022, doi: 10.3390/su142417005.

[2] J. Yu, X. Li, L. Yang, L. Li, Z. Huang, K. Shen, X. Yang, Z. Xu, D. Zhang, and S. Du, "Deep learning models for PV power forecasting: Review," *Energies*, vol. 17, no. 16, Art. no. 3973, 2024, doi: 10.3390/en17163973.

[3] S. Al-Dahidi, M. Madhiarasan, L. Al-Ghussain, A. M. Abubaker, A. D. Ahmad, M. Alrbai, M. Aghaei, H. Alahmer, A. Alahmer, and P. Baraldi, "Forecasting solar photovoltaic power production: A comprehensive review and innovative data-driven modeling framework," *Energies*, vol. 17, no. 16, Art. no. 4145, 2024, doi: 10.3390/en17164145.

[4] G. Zhang, D. Yang, G. Galanis, and E. Androulakis, "Solar forecasting with hourly updated numerical weather prediction," *Renewable and Sustainable Energy Reviews*, vol. 154, Art. no. 111768, 2022, doi: 10.1016/j.rser.2021.111768.

[5] D. Markovics and M. J. Mayer, "Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction," *Renewable and Sustainable Energy Reviews*, vol. 161, Art. no. 112364, 2022, doi: 10.1016/j.rser.2022.112364.

[6] M. J. Mayer, "Benefits of physical and machine learning hybridization for photovoltaic power forecasting," *Renewable and Sustainable Energy Reviews*, vol. 168, Art. no. 112772, 2022, doi: 10.1016/j.rser.2022.112772.

[7] R. Liu, J. Wei, G. Sun, S. M. Muyeen, S. Lin, and F. Li, "A short-term probabilistic photovoltaic power prediction method based on feature selection and improved LSTM neural network," *Electric Power Systems Research*, vol. 210, Art. no. 108069, 2022, doi: 10.1016/j.epsr.2022.108069.

[8] A. Agga, A. Abbou, M. Labbadi, Y. El Houm, and I. H. O. Ali, "CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production," *Electric Power Systems Research*, vol. 208, Art. no. 107908, 2022, doi: 10.1016/j.epsr.2022.107908.

[9] W. Hua, Y. Chen, M. Qadrdan, J. Jiang, H. Sun, and J. Wu, "Applications of blockchain and artificial intelligence technologies for enabling prosumers in smart grids: A review," *Renewable and Sustainable Energy Reviews*, vol. 161, Art. no. 112308, 2022, doi: 10.1016/j.rser.2022.112308.

[10] D. Kirli, B. Couraud, V. Robu, M. Salgado-Bravo, S. Norbu, M. Andoni, I. Antonopoulos, M. Negrete-Pincetic, D. Flynn, and A. Kiprakis, "Smart contracts in energy systems: A systematic review of fundamental approaches and implementations," *Renewable and Sustainable Energy Reviews*, vol. 158, Art. no. 112013, 2022, doi: 10.1016/j.rser.2021.112013.

[11] M. K. Hasan, A. Alkhalifah, S. Islam, N. B. M. Babiker, A. H. Habib, A. H. M. Aman, and M. A. Hossain, "Blockchain technology on smart grid, energy trading, and big data: Security issues, challenges, and recommendations," *Wireless Communications and Mobile Computing*, vol. 2022, Art. no. 9065768, 2022, doi: 10.1155/2022/9065768.

[12] J. Simeunović, B. Schubnel, P.-J. Alet, and R. E. Carrillo, "Spatio-temporal graph neural networks for multi-site PV power forecasting," *IEEE Transactions on Sustainable Energy*, vol. 13, no. 2, pp. 1210–1220, 2022.

[13] E. Sarmas, E. Spiliotis, E. Stamatopoulos, V. Marinakis, and H. Doukas, "Short-term photovoltaic power forecasting using meta-learning and numerical weather prediction independent long short-term memory models," *Renewable Energy*, vol. 216, Art. no. 118997, 2023.

[14] Z. A. Khan, T. Hussain, and S. W. Baik, "Dual stream network with attention mechanism for photovoltaic power forecasting," *Applied Energy*, vol. 338, Art. no. 120916, 2023.

[15] M. Yu, D. Niu, K. Wang, R. Du, X. Yu, L. Sun, and F. Wang, "Short-term photovoltaic power point-interval forecasting based on double-layer decomposition and WOA-BiLSTM-Attention and considering weather classification," *Energy*, vol. 275, Art. no. 127348, 2023.
[16] M. J. Mayer, D. Yang, and B. Szintai, "Comparing global and regional downscaled NWP models for irradiance and photovoltaic power forecasting: ECMWF versus AROME," *Applied Energy*, vol. 352, Art. no. 121958, 2023.
[17] M. J. Mayer and D. Yang, "Pairing ensemble numerical weather prediction with ensemble physical model chain for probabilistic photovoltaic power forecasting," *Renewable and Sustainable Energy Reviews*, vol. 175, Art. no. 113171, 2023.
[18] S. M. Miraftabzadeh, C. G. Colombo, M. Longo, and F. Foiadelli, "A day-ahead photovoltaic power prediction via transfer learning and deep neural networks," *Forecasting*, vol. 5, no. 1, pp. 213–228, 2023.
[19] H. Alharkan, S. Habib, and M. Islam, "Solar power prediction using dual stream CNN-LSTM architecture," *Sensors*, vol. 23, no. 2, Art. no. 945, 2023, doi: 10.3390/s23020945.
[20] O. Aldosari, S. Batiyah, M. Elbashir, W. Alhosaini, and K. Nallaiyagounder, "Performance evaluation of multiple machine learning models in predicting power generation for a grid-connected 300 MW solar farm," *Energies*, vol. 17, no. 2, Art. no. 525, 2024, doi: 10.3390/en17020525.
[21] F. Hu, L. Zhang, and J. Wang, "A hybrid convolutional–long short-term memory–attention framework for short-term photovoltaic power forecasting, incorporating data from neighboring stations," *Applied Sciences*, vol. 14, no. 12, Art. no. 5189, 2024.
[22] H. Yu, S. Chen, Y. Chu, M. Li, Y. Ding, R. Cui, and X. Zhao, "Self-attention mechanism to enhance the generalizability of data-driven time-series prediction: A case study of intra-hour power forecasting of urban distributed photovoltaic systems," *Applied Energy*, vol. 374, Art. no. 124007, 2024.
[23] K. Tao, J. Zhao, Y. Tao, Q. Qi, and Y. Tian, "Operational day-ahead photovoltaic power forecasting based on transformer variant," *Applied Energy*, vol. 373, Art. no. 123825, 2024.
[24] F. Mo, X. Jiao, X. Li, Y. Du, Y. Yao, Y. Meng, and S. Ding, "A novel multi-step ahead solar power prediction scheme by deep learning on transformer structure," *Renewable Energy*, vol. 230, Art. no. 120780, 2024.
[25] R. Chen, G. Liu, Y. Cao, G. Xiao, and J. Tang, "CGAformer: Multi-scale feature transformer with MLP architecture for short-term photovoltaic power forecasting," *Energy*, vol. 312, Art. no. 133495, 2024, doi: 10.1016/j.energy.2024.133495.
[26] S. Pereira, P. Canhoto, and R. Salgado, "Development and assessment of artificial neural network models for direct normal solar irradiance forecasting using operational numerical weather prediction data," *Energy and AI*, vol. 15, Art. no. 100314, 2024.
[27] O. Barhmi, B. H. Tahrani, and N. Anani, "A review of solar forecasting techniques and the role of artificial intelligence," *Solar*, vol. 4, no. 1, pp. 99–135, 2024.
[28] A. La Fata, G. Moser, R. A. R. de Moura, and R. Procopio, "Near real-time photovoltaic power forecasting through recurrent neural network using timely open-access data," *Energies*, vol. 18, no. 23, Art. no. 6220, 2025, doi: 10.3390/en18236220.
[29] A. M. Al Shareef, R. Alkhudary, and M. Alkhudary, "Integration of blockchain with artificial intelligence in the energy sector: A review," *Frontiers in Energy Research*, vol. 12, Art. no. 1377950, 2024, doi: 10.3389/fenrg.2024.1377950.

[30] M. Jaramillo, A. P. Delgado, J. C. Rojas, and R. Villalba, "A bibliometric assessment of AI, IoT, blockchain, and big data in renewable energy-oriented power systems," *Energies*, vol. 18, no. 12, Art. no. 3067, 2025, doi: 10.3390/en18123067