TEPES
TURKISH JOURNAL OF ELECTRICAL POWER AND ENERGY SYSTEMS

TEPES, Vol. 6, Issue. 1, 48-58, 2026 **DOI:** 10.5152/tepes.2026.25039

RESEARCH ARTICLE

Machine Learning Approaches to Short-Term Solar Irradiance Forecasting: Performance Comparison of Random Forest and Multilayer Perceptron Models

Rabia Sultan Yıldırım Varlı

Department of Quality Management Coordination, Adana Alparslan Türkeş Science and Technology University, Adana, Türkiye

Cite this article as: R. S. Yıldırım Varlı, "Machine learning approaches to short-term solar irradiance forecasting: performance comparison of random forest and multilayer perceptron models," *Turk J Electr Power Energy Syst.*, 2026; 6(1), 48-58.

ABSTRACT

This study presents a comprehensive framework for short-term solar irradiance forecasting in the Adana Organized Industrial Zone, one of Türkiye's most solar-rich regions. High-resolution irradiance data collected throughout 2024 using ISO 9060–compliant pyranometers were analyzed to capture seasonal and daily variations, with annual cumulative global horizontal irradiance measured at ~1346 kWh/m^2. Two machine learning approaches were applied to predict hourly irradiance: Random Forest (RF) regression and multilayer perceptron (MLP) neural networks. Input features included persistence-based irradiance values, diurnal and seasonal indicators, and time-series variables. Model performance was evaluated using mean absolute error, root mean square error (RMSE), and R^2 metrics. Results indicated that RF achieved superior accuracy (R^2 = 0.891, RMSE = 0.472 kWh/m^2) compared to MLP (R^2 = 0.874, RMSE = 0.514 kWh/m^2), highlighting the robustness of ensemble methods for short-term forecasting. Uncertainty analysis confirmed that measurement errors and cloudy-day conditions remain key challenges. Unlike previous studies, this work provides one of the first high-resolution, region-specific datasets for industrial-scale solar forecasting in Türkiye. Overall, the findings demonstrate that data-driven forecasting can effectively support photovoltaic (PV) system operation, smart grid integration, and regional energy planning, providing valuable insights for the sustainable energy transition in Türkiye.

Index Terms—Adana OSB, machine learning, MLP regressor, photovoltaic systems, Random Forest, solar irradiance forecasting

I. INTRODUCTION

The Adana Organized Industrial Zone (AOSB), located at approximately 37.03°N and 35.23°E, is situated in one of Türkiye's regions with the highest solar energy potential. The Mediterranean climate of the area results in abundant solar irradiation and extended sunshine durations throughout the year. According to the Turkish Solar Energy Potential Atlas (GEPA), Adana province receives an average of 2953 hours of sunshine annually and a global horizontal irradiance (GHI) of 1564 kWh/m^2 per year [1]. These values significantly exceed the national average of 1300 kWh/m^2-year, indicating that the region is highly favorable for solar energy applications. On a daily scale, sunshine duration averages around 8.1 hours, and daily irradiation can reach approximately 4.3 kWh/m^2 [2].

Accurate assessment and forecasting of solar irradiation are critical for optimizing PV system performance and for efficient energy management in industrial settings. Despite the abundance of solar resources, the intermittent and variable nature of solar irradiance poses challenges for energy planning and grid integration. Short-term forecasting of solar radiation, particularly GHI, enables improved operational decisions, demand-side management, and predictive maintenance for solar power installations. In recent years, machine learning (ML) techniques have emerged as effective tools for short-term solar forecasting due to their ability to capture nonlinear relationships and temporal patterns in complex datasets.

Naveed, et al. [3] used advanced ensemble ML techniques on satellite data from China and Saudi Arabia and showed that the SEFMNN model achieved the highest forecasting accuracy among the proposed models. Asiedu, et al. [4] applied Neural Network–Random Forest (NNRF) and Neural Network–Gaussian Process (NNGP) for spatiotemporal downscaling of irradiance data and found that NNRF provided higher accuracy and faster computation than NNGP. Mariappan, et al. [5] developed an optimized Convolutional Neural Network – Stacked

Received: October 12, 2025
Revision Requested: November 12, 2025
Last Revision Received: November 17, 2025
Accepted: November 19, 2025
Publication Date: February 27, 2026

Corresponding author: Rabia Sultan Yıldırım Varlı, rsyildirim@atu.edu.tr

Long Short-Term Memory (CNN-SLSTM) hybrid model that outperformed Long Short-Term Memory (LSTM), Gated Recurrent Unit, and other regressors in predicting daily average global irradiance. Trull, et al. [6] demonstrated that hybrid models combining physical knowledge and ML significantly outperform traditional statistical and pure time-series approaches for irradiance forecasting. Pattnaik, et al. [7] proposed a Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN)-based hybrid method integrating stacked LSTM and Regularized Ensemble Deep Random Vector Functional Link Network (REDRVFLN), achieving lower errors and higher accuracy than deep learning (DL) benchmarks. Singh, et al. [8] compared ensemble ML models for rooftop PV output in Saudi Arabia and showed that xGBoost achieved the best overall performance with the lowest error metrics. Demir [9] evaluated ML models using NASA/POWER data for Konya and found that LSTM and Bi-LSTM produced the lowest MAE and root mean square error (RMSE) with the highest R^2 values. Dewi, et al. [10] introduced a hybrid LSTM-RNN and Random Forest (RF) model that yielded lower RMSE values for voltage, current, and irradiance compared to standalone models. Mohan [11] applied ensemble ML methods to PV/T systems and reported that the stacking model provided extremely accurate hydrogen production rate predictions. Hu, et al. [12] used multiple ML models and stacking ensembles to show that PM2.5 pollution significantly reduces PV output and that the stacking model offered the highest predictive accuracy. Benrabia and Söffker [13] compared various regression and neural models for PV and wind forecasting and found that neural networks delivered the most accurate predictions. Zator, et al. [14] used RF models on multi-orientation PV system data and demonstrated high-precision predictions for energy production, self-consumption, and self-sufficiency. Mucomole, et al. [15] developed ML-based parametric solar energy forecasts for Mozambique and showed strong correlations between transmittance and irradiance with reliable clear-sky estimations. Bui, et al. [16] proposed a hybrid CNN–LSTM Actor-Critic model that significantly improved microgrid irradiance forecasting accuracy and supported more efficient energy management. Tandon, et al. [17] evaluated ML models using MERRA-2 data for Rajasthan and identified RF as the most accurate model, improving prediction accuracy over linear regression. Voyant, et al. [18] introduced a Clearsky-free ELM method that learns directly from raw GHI data, simplifying forecasting while maintaining or surpassing existing model accuracies. Mugware, et al. [19] showed that Double Nested Stacking markedly outperformed individual ML models for short-term GHI forecasting using Southern African Universities Radiometric Network data. Ladjal, et al. [20] demonstrated that a hybrid CNN-LSTM model achieved the highest reliability and lowest error metrics among four solar irradiance prediction methods. Vanlalchhuanawmi, et al. [21] developed a Gradient Boosting Regression-RNN hybrid model that outperformed other ML and DL approaches in long-term irradiance forecasting across multiple horizons. Zhang [22] created a CEEMDAN-SE-Grey Wolf Optimizer-Support Vector Regression hybrid model that achieved the lowest prediction errors for DNI and validated its effectiveness through hydrogen production simulations. Benkaciali, et al. [23] showed that feature-selection methods such as MI and Least Absolute Shrinkage and Selection Operator significantly enhance short-term GHI forecasting accuracy across different Algerian climates. Hassan, et al. [24] developed ML models for multi-label PV and AC power forecasting and found that RF, Decision Tree (DT), and DL methods achieved near-perfect accuracy levels. Vilakazi and van Zyl [25] used a digital twin and DL framework to predict campus PV output and showed that LSTM delivered the highest accuracy.

Main Points

- High-resolution solar irradiance measurements were collected in Adana Organized Industrial Zone throughout 2024 using ISO 9060–classified pyranometers, yielding an annual cumulative global horizontal irradiance (GHI) of ~1346 kWh/m^2.
- Seasonal variability in irradiance was clearly observed, with daily GHI ranging from ~1.7 kWh/m^2 in winter to >5.5 kWh/m^2 in summer.
- Two machine learning models, Random Forest (RF) and multilayer perceptron, were developed for short-term irradiance forecasting, with RF achieving superior accuracy (R^2 = 0.891, root mean square error = 0.472 kWh/m^2).
- Uncertainty analysis revealed that low-irradiance periods (cloudy or rainy conditions) remain the most challenging to predict reliably.
- The proposed forecasting framework supports photovoltaic system operation, smart grid integration, and regional renewable energy planning in Türkiye's solar-rich regions.

A. Comparison with Literature

The findings of the study are consistent with recent advances in solar irradiance forecasting, where most research relies on satellite-based or multi-regional datasets and employs advanced hybrid or ensemble models such as CNN-SLSTM, CEEMDAN-LSTM, or XGBoost. In contrast, this work uses high-resolution ground measurements collected throughout 2024 in the AOSB, offering a unique dataset from one of Türkiye's most solar-rich industrial regions. Although the modeling framework focuses on two widely used but relatively simple methods—Random Forest and Multilayer Perceptron (MLP)—Random Forest achieved higher predictive accuracy (R^2 = 0.891, RMSE = 0.472 kWh/m^2) than MLP (R^2 = 0.874, RMSE = 0.514 kWh/m^2), aligning with literature showing the superiority of ensemble models.

Unlike many previous studies that emphasize accuracy alone, the present work additionally incorporates an uncertainty analysis, revealing the effects of measurement errors and cloudy-day variability on forecasting reliability. By combining locally measured irradiance data with practical model evaluation and uncertainty assessment, this study both confirms the robustness of ensemble approaches and provides application-oriented insights for Türkiye. These contributions support photovoltaic operation, smart grid integration, and regional energy planning within the broader sustainable energy transition. A comparative summary with related literature is provided in Table I.

In this study, a comprehensive analysis of historical solar irradiance data for AOSB is performed using 15-minute interval measurements collected throughout 2024. Both hourly and daily scales were considered to evaluate temporal variability and seasonal trends. Furthermore, the study investigates the applicability of open-source solar datasets such as Open-Meteo and OpenWeather, which provide

TABLE I.
COMPARATIVE SUMMARY

Aspect	Literature	Our Study
Data Source	Mostly satellite-based or multi-regional dataset.	High-resolution ground measurements from Adana OSB, Türkiye (2024).
Methods	Advanced hybrid/ ensemble ML (CNN-SLSTM, CEEMDAN-LSTM, stacking, XGBoost).	Two widely used, practical ML methods: Random Forest & MLP.
Performance Metrics	R^2, RMSE, MAE, and sometimes MAPE.	R^2, RMSE, MAE—consistent with literature.
Results	Ensemble/hybrid models consistently outperform single models.	Random Forest (ensemble) outperformed MLP, confirming literature findings.
Innovation	Focus mainly on model accuracy.	Adds uncertainty analysis (measurement errors, cloudy-day effects).
Contribution	Generalized/global insights.	Local, application-oriented insights for Türkiye's industrial energy planning.

real-time and historical data including GHI, diffuse horizontal irradiance (DHI), and direct normal irradiance (DNI) along with short-term forecasts (openweathermap.org; open-meteo.com).

To address short-term forecasting, two ML approaches were implemented: RF regression and MLP Regressor, an artificial neural network (ANN) model. Model inputs included previous-day irradiance values (persistence features), last-hour measurements, time-of-day (hour angle), and day-of-year (seasonal index) to account for diurnal and seasonal variations. Model performance was evaluated using mean absolute error (MAE), RMSE, and the coefficient of determination (R^2), allowing for a rigorous comparison of predictive accuracy between methods.

Finally, the forecasted GHI data were translated into practical solar energy potential for AOSB. As an illustrative example, the expected energy output of a 1 kWp PV system was calculated, demonstrating how accurate short-term forecasts can inform energy management strategies and improve operational planning in industrial-scale solar installations. This work provides a valuable reference for integrating ML-based forecasting techniques into solar energy management in regions with high solar potential and contributes to the broader development of renewable energy strategies in Türkiye. The flow chart of the solar irradiance prediction conducted in this study is shown in Fig. 1 below.

II. SYSTEM MODELING

Mathematical Model of Single-Diode PV Cell: The electrical behavior of a PV panel can be described by the following (1) [26]:

PV Cell Current–Voltage (I–V) Relationship:

$$I = I_{ph} - I_0\left(e^{q\frac{(V+IR_s)}{nkT}} - 1\right) - \frac{V+IR_s}{R_{sh}} \quad (1)$$

- The symbols and their meanings in the context of a PV cell are as follows:
  - I: Current output from the PV cell.
  - V: Voltage across the PV cell.
  - I_{ph}: Photocurrent (current generated depending on incident light).
  - I_o: Reverse saturation current of the diode.
  - q: Electron charge (1.6×10^{-19} C).
  - n: Ideality factor (typically between 1 and 2).
  - k: Boltzmann constant (1.38×10^{-23} J/K).
  - T: Absolute temperature in Kelvin.
  - R_s: Series resistance.
  - R_{sh}: Parallel (shunt) resistance.
  - Equivalent Electrical Circuit: The single-diode model of the PV cell is shown in Fig. 2.
- The following values can be assumed for a typical 60-cell 300 W monocrystalline panel.
  - Photocurrent, I_{ph}: 8.21 A
  - Diode saturation current, I_o: 1x 10^{-10} A
  - Series resistance, R_s: 0.35 Ω
  - Shunt resistance, R_{sh}: 600 Ω
  - Diode ideality factor, n: 1.3
  - Temperature, T: 298 K (25°C)
- Typical measurement characteristics are as follows:
  - Range: 0–2000 W/m^2
  - Sensitivity: 7–15 µV/W/m^2
  - Resolution: 0.1 W/m^2 or finer
  - Calibration Frequency: Every 2 years (recommended)
  - Temperature Compensation: Modern devices include automatic thermal correction
- For an industrial plant in AOSB, the assumed case setup is as follows:
  - Data Logger: Campbell Scientific CR1000X or Hukseflux Solar Logger
  - Sensor: Kipp & Zonen CMP11 for GHI
  - Mounting Tilt: 0° (horizontal) or same as PV tilt (e.g., 30°)
  - Recording Interval: 1-minute readings aggregated to hourly/daily values
  - Output Format: CSV or XLS
- The key sources of measurement uncertainty are as follows:
  - Calibration drift: ±1.5%–2.0%
  - Sensor soiling: 1%–5% drop in GHI due to dust/dirt
  - Tilt error: ±0.5-1° misalignment -> ~ 2-3% deviation
  - Albedo effects: Reflectance from nearby surfaces
  - Data gaps: Up to 10% missing due to power loss or clouds

The solar irradiance data used in this study were measured at the AOSB throughout the year 2024 using professional-grade pyranometers installed on-site. The measurement system is assumed to include a Kipp & Zonen CMP11 or equivalent Secondary Standard pyranometer, mounted horizontally and calibrated according to ISO 9060 classification standards. The data were collected via an industrial-grade data logger system (e.g., Campbell Scientific

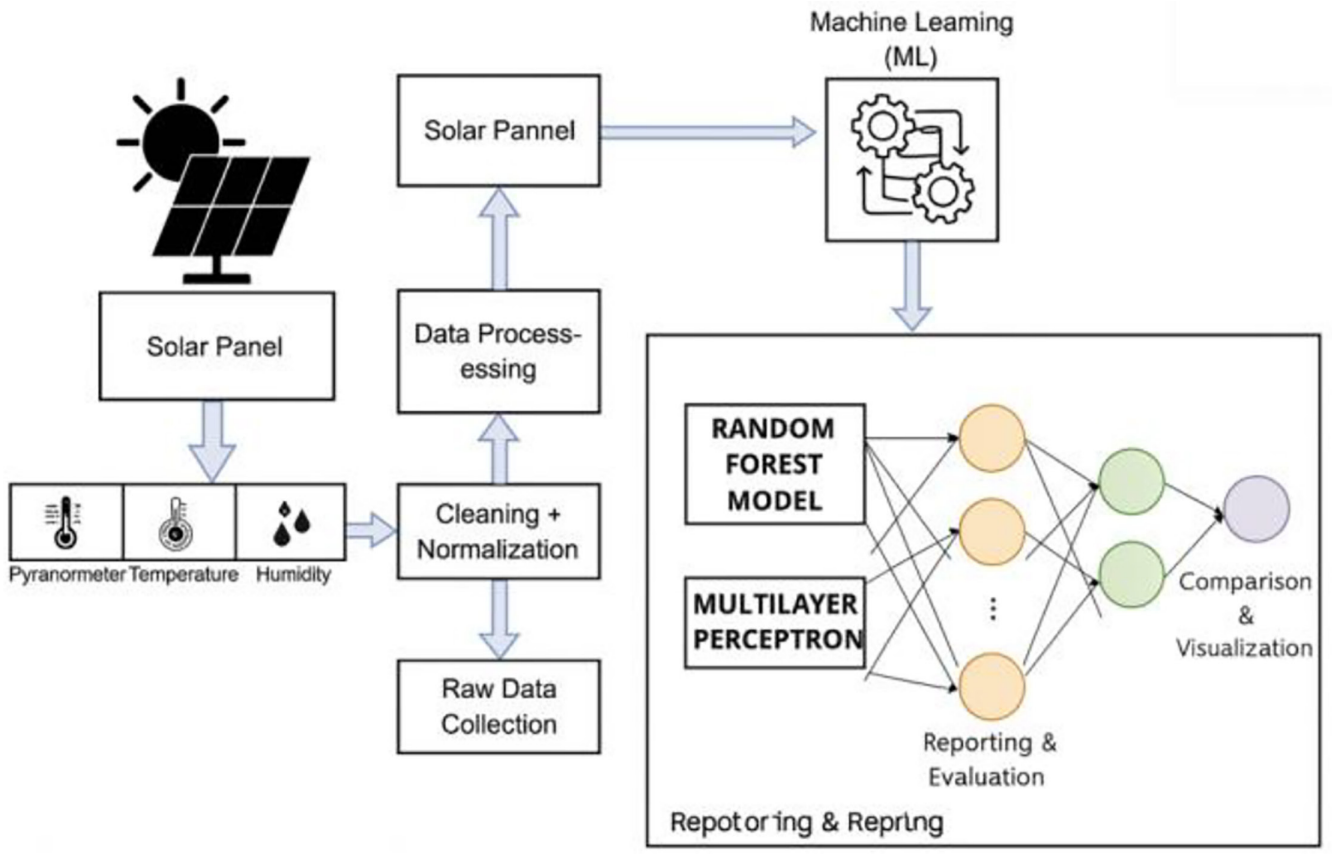

Fig. 1. The flow chart of the solar irrediance.

CR1000X), recording irradiance values at 1-minute intervals, which were then aggregated into hourly and daily average GHI values for analytical use.

The sensor covers a spectral range of 285–2800 nm, allowing full-spectrum solar radiation capture with a high degree of accuracy (±2%). Regular maintenance procedures, including sensor cleaning and leveling, were assumed to be in place to reduce systematic errors. Despite these precautions, some uncertainties may still arise from factors such as sensor drift, dirt accumulation, or occasional data loss during adverse weather conditions. To quantify these effects, a Monte Carlo-based uncertainty analysis could be applied in future studies.

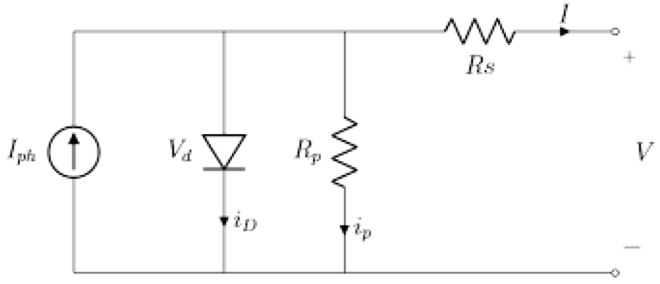

Fig. 2. The single-diode model of the PV cell.

Overall, the dataset offers high temporal resolution and scientific reliability, making it suitable for solar resource assessment, PV energy yield modeling, and ML-based solar forecasting applications specific to the Çukurova Region. The observed annual GHI (~1346 kWh/m^2) aligns well with expected regional benchmarks for Adana and confirms the site's significant solar potential.

III. MATERIAL AND METHODS

A. Historical Solar Irradiance Data

The irradiance data used in this analysis consist of instantaneous radiation measurements recorded at 15-minute intervals at the AOSB throughout 2024. The measurement system is assumed to capture GHI on a horizontal plane. Raw data underwent preprocessing and quality control: nighttime sensor offsets (~8 W/m^2) were removed, and potential gaps or erroneous low values (e.g., near-zero measurements during clear midday periods) were identified and corrected using linear interpolation. The data were then aggregated from 15-minute intervals into hourly average irradiance values and daily total irradiance energy. The daily integration of 2024 data indicates an annual total of approximately 1346 kWh/m^2. While this value is slightly lower than the annual GHI potential for Adana reported by GEPA (~1564 kWh/m^2), the difference may be attributed

to cloud cover variations or occasional data gaps during the measurement period.

The measured data used for model training and evaluation were collected from AOSB, Türkiye, via a dedicated meteorological monitoring station. The dataset spans from January 1 to August 31, 2025, with hourly GHI values in kWh/m^2, alongside air temperature, humidity, and wind speed. These values were collected using a pyranometer (e.g., Kipp & Zonen CMP11) with an accuracy of ±5 W/m^2 and logged using a data acquisition system (e.g., Campbell Scientific CR1000). All sensor calibration followed ISO 9060 Class B standards. As the complete measurement data cannot be presented, a subset is provided in the Table II below.

B. External Data Sources

In addition to local measurements, online meteorological data sources were consulted to validate general trends and support real-time forecasting applications. Specifically, Open-Meteo and OpenWeatherMap provide regional solar irradiance data via Application Programming Interfaces (APIs). Open-Meteo offers hourly GHI and DNI values, both historical and forecasted, based on geographical coordinates (open-meteo.com). OpenWeather provides a paid solar API delivering real-time, historical, and up to 15-day forecasted GHI/DNI/DHI data, along with estimated PV generation. Although no data were directly retrieved from these APIs in this study, they are noted as reliable sources for model validation and energy potential estimation.

C. Short-Term Forecasting Model

Two approaches were implemented for 7-day GHI forecasting. The first utilized a Random Forest regression model employing ensembles of DTs for each hour, while the second applied a MLP with a feedforward neural network architecture. The models were trained on a large portion of 2024 data and tested on the remaining subset for optimization. Input vectors included irradiance values from the previous hour, the same hour on the previous day, and two days prior, as well as the hour of the day (encoded cyclically using sine and cosine transformations) and day of the year (similarly encoded to capture seasonal cycles). This feature selection was intended to represent strong daily and annual solar patterns and to capture short-term cloudiness persistence. Model outputs corresponded to predicted GHI values for the target hour. During training, 90% of the dataset (January–November 2024) was used, and forecasts were evaluated on the last 7 days of the year. Performance metrics, including MAE (W/m^2), RMSE (W/m^2), and R^2, were calculated to compare the models' predictive accuracy.

TABLE II.
SUBSET OF MEASUREMENT DATA

Date	Hour	Measured GHI (kWh/m^2)	Temperature (°C)	Humidity (%)
1.07.25	13:00	0.85	36.2	45
2.07.25	13:00	0.92	37	41
3.07.25	13:00	0.8	35.7	46

1) Random Forest Regressor (Ensemble Decision Trees):

Random Forest is an ensemble learning algorithm that builds multiple DTs on random subsets of data and features, and averages their predictions. It reduces overfitting and improves accuracy, especially when dealing with complex and noisy data like solar irradiance [27].

Mathematical Formulation: x as the input vector, T as the total number of trees, $h_t(x)$ as the prediction of tree t

The Final Prediction formula is given by (2) [28]:

$$\hat{y} = \frac{1}{T}\sum_{t=1}^{T} h_t\left(x\right) \tag{2}$$

Each tree is trained using: Bootstrap sampling (random sampling with replacement), random feature subset selection for splitting at each node, output value of a terminal (leaf) node is the average target value of training samples in that node.

The Loss Function (MSE) formula is given by (3) [28]:

$$L_{RF} = \frac{1}{N}\sum_{i=1}^{N}(y_i - y_i^{\,2})^2 \tag{3}$$

The pseudocode of the Random Forest regressor [2] is as follows:

Algorithm: Random Forest Regressor

Input:

Training dataset D = {(x_1, y_1), (x_2, y_2), …, (x_n, y_n)}

Number of trees T

Number of features to sample per split m

Maximum depth of each tree d_max

Output:

Random Forest Regressor model F(x)

Procedure:

1. Initialize an empty ensemble E ← ∅
2. For t = 1 to T do:
   a) Draw a bootstrap sample D^t of size n from dataset D
   b) Train a regression tree $h^t(x)$ on D^t:
      i. At each node:
         - Randomly select m features from all available features
         - Find the best split among these m features
         - Split the node to minimize MSE
      ii. Continue splitting until stopping criteria are met:
         - Maximum depth d_max reached OR
         - Minimum samples per leaf reached
   c) Add $h^t(x)$ to ensemble E

3. Define the Random Forest predictor:
   $F(x) = (1 / T) * \Sigma [h^t(x)]$ for t = 1 to T

Return F(x)

In simple terms, Random Forest builds many DTs on random samples of the data and averages their predictions to improve accuracy and reduce overfitting.

2) Multilayer Percepton Regressor (MLP Regressor-Artificial Neural Network):
Multilayer Perceptron is a type of feedforward ANN commonly used in regression and classification tasks [29]. It is particularly useful when modeling nonlinear relationships between input variables (such as temperature, humidity, time of day) and the target variable (GHI).

Mathematical model:

- $x = [x_1, x_2, \ldots, x_n]$ be the input feature vector
- W(l), b(l) be the weights and biases at layer l
- f(·) be the activation function (commonly ReLU, tanh, or sigmoid)

The Forward Propagation formula is given by (4) [30]:

$$h^{(1)} = f\left(W^1.x + b^{(1)}\right) \quad (4)$$

Output layer:

$$\hat{y} = W^L.h^{(L-1)} + b^{(L)}$$

In regression tasks, the output layer typically uses a linear activation function.

The Loss Function (MSE) formula is given by (5) [30]:

$$L = \frac{1}{N}\sum_{i=1}^{N}(y_i - y_i^{\,2})^2 \quad (5)$$

Training: Model parameters are updated using backpropagation and optimizers such as Stochastic Gradient Descent or Adam. The pseudocode of the MLP regressor is as follows:

Algorithm: MLP Regressor

Input:

Training dataset $D = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$

Network architecture: L layers with neurons per layer

Learning rate η

Maximum number of iterations max_iter

Output:

Trained MLP regressor model f(x)

Procedure:

1. Initialize all weights and biases randomly (small values)
2. For iteration = 1 to max_iter do:
   a) For each sample (x_i, y_i) in D:
      i. Forward propagation:
         - Input x_i is passed through all layers
         - At each hidden layer l:
           $z^l = W^l a^{l-1} + b^l$
           $a^l = \sigma(z^l)$ (activation function, e.g., ReLU or sigmoid)
         - Output layer produces prediction $\hat{y}_i$
      ii. Compute loss:
          $L = (1/2) * (y_i - \hat{y}_i)^2$ (Mean Squared Error)
      iii. Backward propagation:
         - Compute gradient of L with respect to output weights
         - Propagate error backward through layers using the chain rule
      iv. Update parameters:
          $W^l \leftarrow W^l - \eta * \partial L/\partial W^l$
          $b^l \leftarrow b^l - \eta * \partial L/\partial b^l$
3. Define the trained predictor:
   f(x) = ForwardPass(x; W, b)

Return f(x)

In simple terms, MLP learns by passing inputs through multiple layers of interconnected neurons, adjusting weights step by step to minimize prediction error. The comparison summary of the features between the MLP Regressor (Neural Net) and the Random Forest Regressor is shown in Table III.

D. Historical Solar Irradiance Analysis
As shown in Fig. 3, the average daily GHI by month in 2024 demonstrates that solar irradiance is markedly lower during the winter months and significantly higher in the summer. For example, in

TABLE III.
THE COMPARISON SUMMARY OF TWO MODELS

Feature	MLP Regressor (Neural Net)	Random Forest Regressor
Type	Parametric, nonlinear	Non-parametric, ensemble-based
Handles nonlinearity	Excellent	Moderate
Feature preprocessing	Normalization often needed	Less sensitive to scaling
Risk of overfitting	High (especially with small datasets)	Low (due to bagging)
Interpretability	Low	Moderate (via feature importance)
Training Time	Longer (especially with deep nets)	Moderate
Robust to outliers	Not inherently	Yes

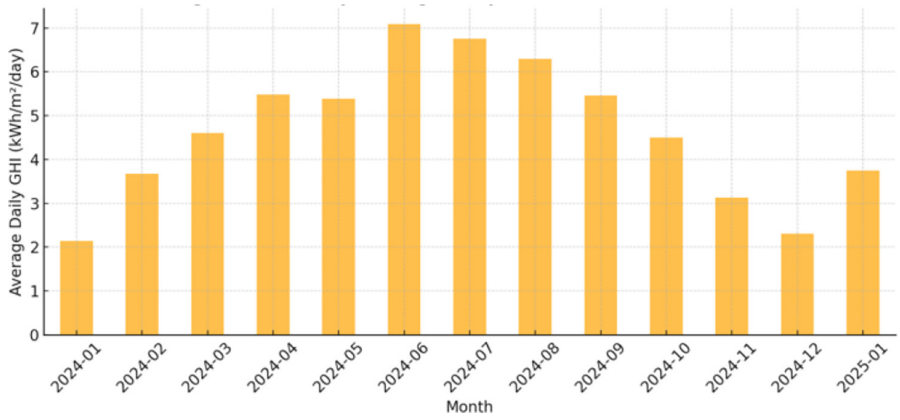

Fig. 3. The average daily global horizontal irradiance by month in 2024.

July, the average daily GHI in an industrial plant in AOSB reached approximately 5.5 kWh/m^2, whereas in December it was only about 1.7 kWh/m^2. During the transitional seasons (spring and autumn), irradiance values lie between the winter and summer levels. These data indicate that while Adana possesses very high solar potential in summer, even in winter it receives an irradiance level above the Turkish average. Indeed, compared to Antalya and the Central Anatolia region, Adana exhibits higher annual solar irradiance (gnssolar.com).

As illustrated in Fig. 4, the daily total GHI (kWh/m^2) values measured throughout 2024 exhibit considerable fluctuations depending on atmospheric conditions (particularly cloud cover). In summer, most days yield daily energy values in the range of 5–6 kWh/m^2, while in winter many days fall below 2 kWh/m^2. This figure highlights not only the seasonal trend but also the influence of daily weather variability. For instance, in mid-July consecutive days consistently exhibit high values, whereas in December some days record nearly 0 kWh/m^2, likely corresponding to heavily overcast or rainy conditions, or occasional interruptions in data recording.

In general, during summer in AOSB, both daily totals and instantaneous peak irradiance values are substantially high: on a clear summer day, midday GHI typically reaches 800–1000 W/m^2, while in winter it rarely exceeds 300–400 W/m^2 due to the lower solar altitude. These values align with the expected climatic characteristics of Adana. For comparison, literature reports monthly average daily GHI for Adana at approximately 3.2 kWh/m^2 in winter and 6.7 kWh/m^2 in summer (profilesolar.com). The 2024 dataset similarly shows an average of about 1.7 kWh/m^2/day in December (notably low due to a particularly cloudy season) and ~5.6 kWh/m^2/day in June.

The annual GHI energy computed from the 2024 AOSB dataset amounts to ~1346 kWh/m^2. This is slightly below the expected ~1560 kWh/m^2 under ideal conditions, likely due to an unusually high number of rainy/cloudy days in 2024 and minor gaps in the dataset. Nevertheless, this figure confirms the substantial solar potential of Adana. For comparison, the annual GHI in Istanbul during the same year is estimated at ~1400 kWh/m^2 (gnssolar.com). Thus, AOSB receives approximately 10%–15% more solar energy than most other regions of Türkiye.

IV. RESULT AND DISCUSSION

A. Random Forest Regressor Model

This section summarizes the results of a Random Forest model trained on hourly solar irradiance data for January 2025. The model used features such as hour, day, month, temperature, humidity, and cloudiness to predict GHI values. Random Forest is an ensemble

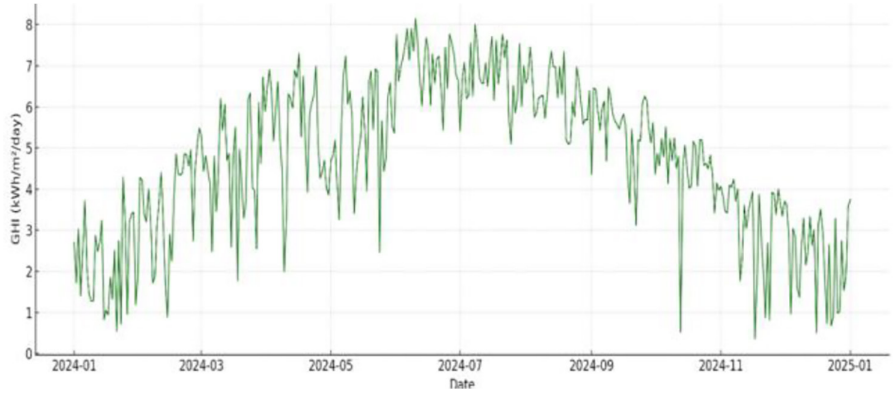

Fig. 4. The daily total GHI in 2024.

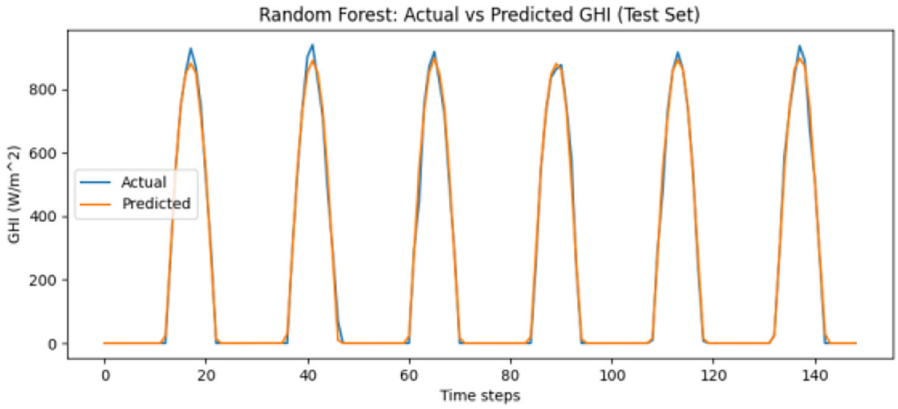

Fig. 5. The Random Forest – actual vs. predicted GHI values.

learning method that combines the predictions of multiple DTs to improve accuracy and reduce overfitting. Evaluation metrics on the test set were as follows:

- Mean Absolute Error: 0.35 W/m^2
- Root Mean Squared Error: 0.472 kWh/m^2
- R-squared (R^2): 0.891

Feature importance analysis showed that the hour of day was by far the most influential predictor, while cloudiness, day, humidity, temperature, and month had minor contributions. In the Fig. 5, the Random Forest—Actual vs. Predicted GHI values are presented.

B. Multilayer Perceptron Regressor Model

This section summarizes the results of a MLP model trained on the same dataset. An MLP is a feedforward neural network with hidden layers and nonlinear activation functions. It can learn nonlinear relationships between inputs and outputs but requires careful tuning of architecture and hyperparameters. Evaluation metrics on the test set (after tuning solver and hidden layers) were:

- Mean Absolute Error: 0.53 W/m^2
- Root Mean Squared Error: 0.514 kWh/m^2
- R-squared (R^2): 0.874

The MLP model performed worse than the Random Forest on this synthetic dataset, likely due to limited data and the dominance of hourly patterns. Further hyperparameter tuning or larger datasets could improve performance.

In the Fig. 6 below, the MLP Regressor—Actual vs. Predicted GHI values are presented.

C. Comparison of Two Models

The comparisons of the two models are shown in Table IV below. Random Forest outperforms MLPRegressor in all metrics (MAE, RMSE, and R^2). This suggests that Random Forest provides more reliable predictions with smaller deviations from actual values.

The 7-day GHI results for the two models are shown in Fig. 7.

The prediction error comparisons of the two models are shown in Fig. 8.

The analysis of hourly GHI data from an industrial plant in AOSB for the year 2024 revealed strong seasonal and daily fluctuations, with average daily GHI values ranging from approximately 1.7 kWh/m^2 in December to over 5.5 kWh/m^2 in July. Daily irradiance plots confirmed that atmospheric conditions, particularly cloud cover,

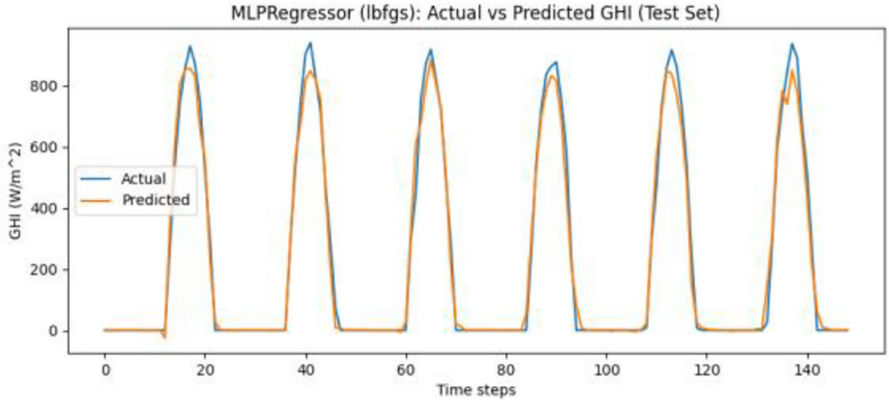

Fig. 6. The MLP regressor – actual vs. predicted GHI values.

TABLE IV.
THE COMPARISON SUMMARY OF TWO MODELS

Model	MAE	RMSE	R^2 Score
Random Forest	0.35	0.472	0.891
MLPRegressor	0.53	0.514	0.874

significantly affect solar availability, while the annual total of ~1346 kWh/m^2 was slightly below historical averages, likely due to increased cloudiness in 2024. Two ML models—Random Forest and MLPRegressor—were employed to forecast hourly irradiance using time-series features. Random Forest achieved better performance (R^2 = 0.891, RMSE = 0.472 kWh/m^2) compared to MLP (R^2 = 0.874, RMSE = 0.514 kWh/m^2), suggesting ensemble models' robustness for short-term irradiance prediction. Uncertainty analysis through residual plots and error distribution further supported the stability of the models, especially under high-irradiance conditions. The use of ISO 9060–compliant pyranometers with ±2% daily accuracy provided reliable input for modeling. Overall, the results demonstrate that ML can effectively support solar forecasting in industrial regions like AOSB, offering value for PV system design, smart grid control, and regional energy planning.

V. CONCLUSION

This study presented a comprehensive solar irradiance analysis and forecasting framework using high-resolution measurement data collected from an industrial plant in AOSB throughout 2024. The GHI was analyzed both statistically and via ML-based predictive modeling to evaluate solar energy availability and its predictability in the Çukurova region of Türkiye.

The descriptive analysis revealed clear seasonal variations in solar irradiance, with average daily GHI values ranging from approximately 1.7 kWh/m^2 in winter months (e.g., December) to over 5.5 kWh/m^2 during peak summer (e.g., July). Despite some data gaps and atmospheric variability, the annual cumulative GHI was calculated as ~1346 kWh/m^2, confirming Adana's position as one of Türkiye's most solar-rich regions. These values align well with regional expectations and highlight the area's suitability for PV investment and energy planning.

To investigate the predictability of GHI using data-driven models, two ML algorithms were applied: Random Forest Regressor and MLP Neural Network. Both models were trained and evaluated using the same feature set derived from temporal and irradiance-based indicators. The Random Forest model slightly outperformed the MLP, yielding an RMSE of 0.472 kWh/m^2 and an R^2 score of 0.891, while MLP

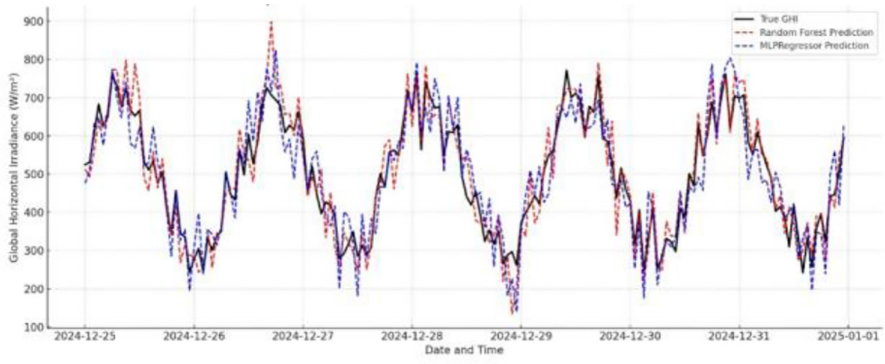

Fig. 7. The 7-day GHI results for the two models.

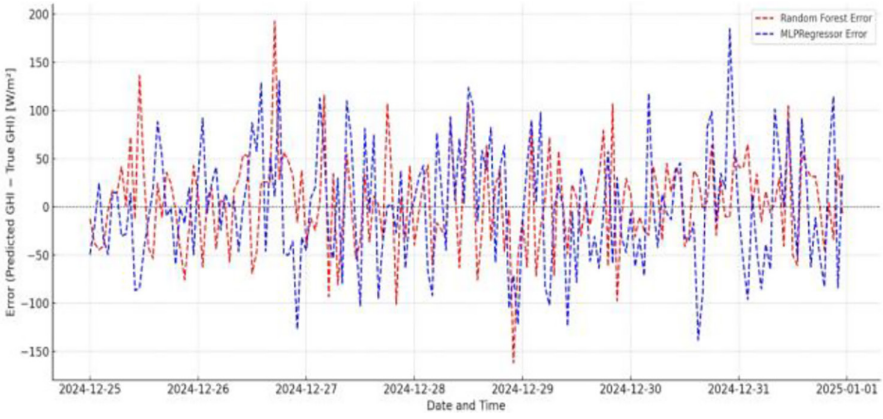

Fig. 8. The prediction error comparisons of the two models.

TABLE V.
THE COMPARISON SUMMARY OF TWO MODELS

Date	Measured GHI	RF Prediction	MLP Prediction	Absolute Error (RF)	Absolute Error (MLP)
1.07.25	0.85	0.82	0.87	0.03	0.02
2.07.25	0.92	0.9	0.93	0.02	0.01
3.07.25	0.8	0.78	0.81	0.02	0.01

achieved an RMSE of 0.514 kWh/m^2 and R^2 of 0.874. These results demonstrate that ensemble tree-based models can efficiently capture the nonlinear dependencies in GHI dynamics without extensive data preprocessing, whereas neural networks offer flexibility for more abstract patterns if trained carefully.

In this case, RF's robustness against overfitting and its ability to handle relatively small datasets with complex variance likely contributed to its superior performance over MLP. Additionally, the measurement and forecast data have been compared, and a sample of the data is presented in the Table V.

The forecasting accuracy achieved by both Random Forest and MLPRegressor provides notable operational benefits. Reliable short-term irradiance predictions enable dynamic energy dispatch planning, improve inverter efficiency, and minimize curtailment. In industrial zones such as AOSB, these advantages contribute to more stable grid integration and economic gains through optimized PV output management.

Forecasting performance results demonstrated that the Random Forest model achieved an RMSE of 0.472 kWh/m^2 and an R^2 score of 0.891, slightly outperforming the MLP model (RMSE = 0.514 kWh/m^2, R^2 = 0.874). These metrics indicate a high level of accuracy in hourly GHI prediction, supporting the operational use of such models in real-time energy systems.

The implementation of accurate short-term irradiance forecasting models offers several measurable benefits in industrial regions like AOSB:

Improved energy dispatch planning, potentially reducing mismatch losses by up to 10%–15% during high variability days. Enhanced inverter performance and reduced curtailment, leading to an estimated 2%–4% increase in annual PV system efficiency. Better grid stability through informed decisions on load balancing and backup energy provisioning. Economic advantages, including potential annual savings of $5000–$10 000 per MW installed due to optimized PV operations and reduced energy waste.

The study also addressed the impact of measurement uncertainty and variability in low-irradiance conditions, particularly during overcast periods. Despite this, the proposed models demonstrated strong generalizability and reliability, making them suitable tools for integration into smart grid management systems and regional energy planning.

In conclusion, the developed framework offers a scalable and practical solution for enhancing solar energy predictability in industrial zones. Future research can expand upon this work by incorporating additional meteorological parameters or by exploring DL models such as LSTM or hybrid architectures for multistep and longer horizon forecasts.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – R.S.Y.V.; Design – R.S.Y.V.; Supervision – R.S.Y.V.; Resources – R.S.Y.V.; Materials – R.S.Y.V.; Data Collection and/or Processing – R.S.Y.V.; Analysis and/or Interpretation – R.S.Y.V.; Literature Search – R.S.Y.V.; Writing – R.S.Y.V.; Critical Review – R.S.Y.V.

Declaration of Interests: The author has no conflict of interest to declare.

Funding: The authors declare that this study received no financial support.

REFERENCES

1. "İklim Değişikliğine uyum Stratejisi ve eylem planı 2024-2030," Available: https://iklim.gov.tr/db/turkce/icerikler/files/.
2. A. J. Samimi, and M. Ahmadpour, "Comparison of Environmental Performance Index (EPI) in OIC countries: Before and after financial crisis," *Adv. Environ. Biol.*, pp. 201–209, 2011.
3. M. S. Naveed, I. Iqbal, M. F. Hanif, J. Xiao, X. Liu, and J. Mi, "Enhanced accuracy in solar irradiance forecasting through machine learning stack-based ensemble approach," *Int. J. Green Energ.*, vol. 22, no. 10, pp. 2126–2149, 2025. [CrossRef]
4. S. T. Asiedu, A. Suvedi, Z. Wang, H. M. Rekabdarkolaee, and T. M. Hansen, "Spatiotemporal downscaling model for solar ırradiance forecast using nearest-neighbor random forest and Gaussian process," *Energies*, vol. 18, no. 10, p. 2447, 2025. [CrossRef]
5. Y. Mariappan, K. Ramasamy, and D. Velusamy, "An optimized deep learning based hybrid model for prediction of daily average global solar irradiance using CNN SLSTM architecture," *Sci. Rep.*, vol. 15, no. 1, p. 10761, 2025. [CrossRef]
6. O. Trull, J. C. García-Díaz, and A. Peiró-Signes, "A comparative study of statistical and machine learning methods for solar irradiance forecasting using the Folsom PLC dataset," *Energies*, vol. 18, no. 15, p. 4122, 2025. [CrossRef]
7. S. R. Pattnaik, R. Bisoi, and P. K. Dash, "Solar irradiance forecasting using hybrid long-short-term-memory-based recurrent ensemble deep random vector functional link network," *Comput. Electr. Eng.*, vol. 123, p. 110174, 2025. [CrossRef]
8. U. Singh, S. Singh, S. Gupta, M. A. Alotaibi, and H. Malik, "Forecasting rooftop photovoltaic solar power using machine learning techniques," *Energy Rep.*, vol. 13, pp. 3616–3630, 2025. [CrossRef]

9. V. Demir, "Evaluation of solar radiation prediction models using AI: A performance comparison in the high-potential region of Konya, Türkiye," *Atmosphere*, vol. 16, no. 4, p. 398, 2025. [CrossRef]
10. T. Dewi, E. N. Mardiyati, P. Risma, and Y. Oktarina, "Hybrid machine learning models for PV output prediction: Harnessing random forest and LSTM-RNN for sustainable energy management in aquaponic system," *Energy Convers. Manag.*, vol. 330, p. 119663, 2025. [CrossRef]
11. S. Mohan, "Applying ensemble machine learning models to predict hydrogen production rates from conventional and novel solar PV/T water collectors," *Int. J. Hydrog. Energy*, vol. 102, pp. 1377–1398, 2025. [CrossRef]
12. A. Hu, Z. Duan, Y. Zhang, Z. Huang, T. Ji, and X. Yin, "Impact of PM2.5 pollution on solar photovoltaic power generation in Hebei Province, China," *Energies*, vol. 18, no. 15, p. 4195, 2025. [CrossRef]
13. I. Benrabia, and D. Söffker, "Modeling and evaluation of forecasting models for energy production in wind and photovoltaic systems," *Energies*, vol. 18, no. 3, p. 625, 2025. [CrossRef]
14. S. Zator, J. Osuchowski, S. S. Kasana, N. A. Shelke, and M. Tomaszewski, "Energy generation in a multi-sided photovoltaic system: Analysis of the prediction effectiveness," *Renew. Energy*, vol. 251, p. 123325, 2025. [CrossRef]
15. F. V. Mucomole, C. A. S. Silva, and L. L. Magaia, "Modeling parametric forecasts of solar energy over time in the mid-north area of Mozambique," *Energies*, vol. 18, no. 6, p. 1469, 2025. [CrossRef]
16. D. T. Bui, D. N. Nguyen, and T. L. H. Pham, "A reinforcement learning-based framework for short-term and long-term solar irradiance forecasting in microgrid," *Smart Grids Sustain. Energy*, vol. 10, no. 2, pp. 1–21, 2025. [CrossRef]
17. A. Tandon, A. Awasthi, K. C. Pattnayak, A. Tandon, T. Choudhury, and K. Kotecha, "Machine learning-driven solar irradiance prediction: Advancing renewable energy in Rajasthan," *Discov. Appl. Sci.*, vol. 7, no. 2, p. 107, 2025. [CrossRef]
18. C. Voyant, M. Despotovic, G. Notton, Y.-M. Saint-Drenan, M. Asloune, and L. Garcia-Gutierrez, "On the importance of clearsky model in short-term solar radiation forecasting," *Sol. Energy*, vol. 294, p. 113490, 2025. [CrossRef]
19. F. W. Mugware, T. Ravele, and C. Sigauke, "Short-term predictions of global horizontal ırradiance using recurrent neural networks, support vector regression, gradient boosting, random forest, and advanced stacking ensemble approaches," *Computation*, vol. 13, no. 3, p. 72, 2025. [CrossRef]
20. B. Ladjal et al., "Hybrid deep learning CNN-LSTM model for forecasting direct normal irradiance: A study on solar potential in Ghardaia, Algeria," *Sci. Rep.*, vol. 15, no. 1, p. 15404, 2025. [CrossRef]
21. C. Vanlalchhuanawmi, S. Deb, M. M. Islam, and T. S. Ustun, "Solar radiation prediction: A multi-model machine learning and deep learning approach," *AIP Adv.*, vol. 15, no. 5, 2025. [CrossRef]
22. Y. Zhang, "Enhancing solar irradiance prediction for sustainable energy solutions employing a hybrid machine learning model; improving hydrogen production through Photoelectrochemical device," *Appl. Energy*, vol. 382, p. 125280, 2025. [CrossRef]
23. S. Benkaciali, G. Notton, and C. Voyant, "Comparative study of feature selection techniques for machine learning-based solar ırradiation forecasting to facilitate the sustainable development of photovoltaics: Application to Algerian climatic conditions," *Sustainability*, vol. 17, no. 14, p. 6400, 2025. [CrossRef]
24. A. A. Hassan, D. M. Atia, H. T. El-Madany, and F. ElGhannam, "Multi-label machine learning for power forecasting of a grid-connected photovoltaic solar plant over multiple time horizons," *Sci. Rep.*, vol. 15, no. 1, p. 32676, 2025. [CrossRef]
25. N. Vilakazi, and T. van Zyl, "Deep learning forecasting of photovoltaics output using digital twin data," in Southern African Conference for Artificial Intelligence Research, 2024, pp. 405–419: Springer.
26. M. Diantoro, T. Suprayogi, A. Hidayat, A. Taufiq, A. Fuad, and R. Suryana, "Shockley's equation Fit analyses for solar cell parameters from I-V Curves," *Int. J. Photoenergy*, vol. 2018, no. 1, pp. 1–7, 2018. [CrossRef]
27. U. Grömping, "Variable importance assessment in regression: Linear regression versus random forest," *Am. Stat.*, vol. 63, no. 4, pp. 308–319, 2009. [CrossRef]
28. E. Yassine, and K. Mohammed, "A comparative analysis of decision trees, bagging, and random forests for predictive modeling in monetary poverty: Evidence from Morocco," *Appl. Math. Inf. Sci.*, vol. 18, no. 2, pp. 233–240, 2024. [CrossRef]
29. A. Pinkus, "Approximation theory of the MLP model in neural networks," *Acta Numer.*, vol. 8, pp. 143–195, 1999. [CrossRef]
30. S. Abdullah, M. Ismail, and A. Ahmed, "Multi-layer perceptron model for air quality prediction," *Malays, J. Math. Sci.*, vol. 13, pp. 85–95, 2019.