





## Article

# Advancing Solar Power Forecasting: Integrating Boosting Cascade Forest and Multi-Class-Grained Scanning for Enhanced Precision

Mohamed Khalifa Boutahir <sup>1,\*</sup>, Yousef Farhaoui <sup>1</sup>, Mourade Azroul <sup>1</sup>, Ahmed Sedik <sup>2,3</sup>  
and Moustafa M. Nasralla <sup>2,\*</sup>

- <sup>1</sup> STI Laboratory, T-IDMS Faculty of Sciences and Techniques of Errachidia, Moulay Ismail University of Meknès, Errachidia 52000, Morocco; y.farhaoui@fste.umi.ac.ma (Y.F.); mo.azroul@umi.ac.ma (M.A.)  
<sup>2</sup> Smart Systems Engineering Laboratory, Communications and Networks Engineering Department, College of Engineering, Prince Sultan University, Riyadh 11586, Saudi Arabia; asedik@psu.edu.sa  
<sup>3</sup> Department of the Robotics and Intelligent Machines, Faculty of Artificial Intelligence, Kafrelsheikh University, Kafrelsheikh 33511, Egypt  
\* Correspondence: moha.boutahir@edu.umi.ac.ma (M.K.B.); mnasralla@psu.edu.sa (M.M.N.)

**Abstract:** Accurate solar power generation forecasting is paramount for optimizing renewable energy systems and ensuring sustainability in our evolving energy landscape. This study introduces a pioneering approach that synergistically integrates Boosting Cascade Forest and multi-class-grained scanning techniques to enhance the precision of solar farm power output predictions significantly. While Boosting Cascade Forest excels in capturing intricate, nonlinear variable interactions through ensemble decision tree learning, multi-class-grained scanning reveals fine-grained patterns within time-series data. Evaluation with real-world solar farm data demonstrates exceptional performance, reflected in low error metrics (mean absolute error, 0.0016; root mean square error 0.0036) and an impressive R-squared score of 99.6% on testing data. This research represents the inaugural application of these advanced techniques to solar generation forecasting, highlighting their potential to revolutionize renewable energy integration, streamline maintenance, and reduce costs. Opportunities for further refinement of ensemble models and exploration of probabilistic forecasting methods are also discussed, underscoring the significance of this work in advancing solar forecasting techniques for a sustainable energy future.

**Keywords:** sustainable solar energy; machine learning; Boosting Cascade Forest; multi-class-grained scanning; forecasting



**Citation:** Boutahir, M.K.; Farhaoui, Y.; Azroul, M.; Sedik, A.; Nasralla, M.M. Advancing Solar Power Forecasting: Integrating Boosting Cascade Forest and Multi-Class-Grained Scanning for Enhanced Precision. *Sustainability* **2024**, *16*, 7462. <https://doi.org/10.3390/su16177462>

Academic Editor: Nuria Novas Castellano

Received: 3 August 2024  
Revised: 21 August 2024  
Accepted: 25 August 2024  
Published: 29 August 2024



**Copyright:** © 2024 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) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The rapid shift towards cleaner and more sustainable energy production is driving the adoption of solar photovoltaic (PV) generation as a key renewable energy source. However, the inherent intermittency of solar irradiation presents significant challenges for maintaining a consistent and stable PV output. Fluctuations in solar irradiation occur on multiple timescales, from momentary variations caused by passing clouds to seasonal shifts due to changes in sun angles [1]. This variability directly impacts PV output power, posing risks to power quality and grid reliability if not accurately forecast and managed. To fully realize the potential of solar PV as a reliable and renewable energy source, it is crucial to develop innovative forecasting strategies that can effectively predict and mitigate these fluctuations [2].

The infrastructure for solar power generation and distribution, as depicted in Figure 1, encompasses several integral components operating collaboratively to capture solar energy and deliver it to end users. Photovoltaic (PV) panels' pivotal role is converting solar irradiation into direct current (DC) electricity. This DC is subsequently transformed into

alternating current (AC) through inverters, facilitating integration with the electrical grid. The generated AC power is then connected to the grid through substations and any necessary voltage transformations. Given our research’s focus on advancing solar forecasting capabilities, a comprehensive understanding of this interconnected system becomes imperative. Failure to accurately predict and manage PV output variability, influenced by fluctuating solar resources, can significantly impact grid stability and reliability. By refining prediction methodologies to accommodate the intricacies of this generation–distribution network, we can actively contribute to optimizing renewable energy system performance and fostering a sustainable energy future with reduced reliance on conventional generation.

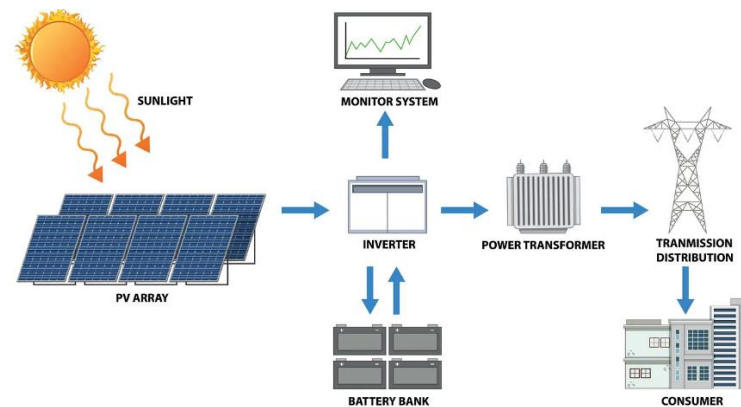


Figure 1. Solar power generation and distribution process.

Precise forecasting of solar power generation has emerged as a pivotal factor in advancing the objectives of cleaner energy production and sustainability, a significance underscored by recent research conducted by the National Renewable Energy Laboratory (NREL). The NREL study [3] delves into prospective scenarios targeting substantial grid decarbonization by 2035 and 2050. As depicted in Figure 2, the visual representation of key findings from the NREL study highlights the consequential impact of accurate solar forecasting on grid decarbonization. The scenario yielding the most substantial reduction in carbon emissions envisions the electrification of additional building, transportation, and industrial energy loads. Solar capacity is anticipated to grow from 3% of the U.S. electricity supply in 2020 to 40% by 2035 and 45% by 2050 [3]. To achieve 95% grid decarbonization by 2035, considerable annual installations of solar photovoltaics (PVs) are imperative, as illustrated in Figure 2.



Figure 2. Solar power generation and grid decarbonization pathway.

In an evolving energy landscape focused on cleaner production and sustainability, the National Renewable Energy Laboratory's (NREL) recent study [3] signifies the critical role of accurate solar power generation forecasting. Technological advances could enable a 95% decarbonized grid by 2035 without affecting electricity prices, as illustrated in Figure 2. The cost for achieving a fully decarbonized grid by 2050, coupled with increased electrification, is projected to be approximately USD 210 billion, resulting in overall savings of USD 1.7 trillion. These findings underscore the economic and environmental benefits of precise solar forecasting, highlighting its essential role in transitioning to cleaner energy production.

Accurate solar power generation forecasting is integral to sustainable energy systems. Improved prediction methodologies enhance reliability in renewable power supply, enabling better integration of solar energy into electricity grids [4]. Our research focuses on advancing forecast accuracy, addressing the intricacies of solar variability, and contributing to the broader adoption of renewable solar power. This work is positioned at the intersection of economic, environmental, and technical dimensions, aiming to promote sustainability across these realms [5].

The primary challenge addressed in this research is the accurate forecasting of solar power generation to ensure a reliable power supply and mitigate the impacts of solar intermittency on the electrical grid. Solar irradiance's unpredictable nature leads to discrepancies between actual generation and scheduled supply, necessitating costly auxiliary reserves for utility-scale PV plants [6–14]. Our research aims to enhance short- and long-term solar forecasts, address these challenges, and facilitate the broader adoption of renewable energy sources [15].

In addressing these limitations, our research proposes a novel approach combining Boosting Cascade Forest and multi-class-grained scanning, tailored for solar PV forecasting across various timescales. Boosting Cascade Forest excels in capturing complex, nonlinear effects, while multi-class-grained scanning enables the detection of fine-grained patterns. This integrated approach promises significant improvements in short- and long-term solar forecasts, addressing the challenges of solar intermittency and promoting the comprehensive adoption of renewable energy sources.

The potential transformative outcomes of this research offer numerous advantages to renewable energy systems and sustainability efforts. Enhanced solar forecasting can lead to better integration of solar PV into the power grid, reducing reliance on supplementary reserves, stabilizing the electricity supply, and minimizing power mismatches. Improved predictions for distributed residential PV systems could mitigate voltage fluctuations and equipment degradation. Additionally, understanding solar intermittent trends may optimize PV capacity planning and siting based on geographical and meteorological factors, fostering increased adoption of renewable solar energy and contributing to a more sustainable and resilient future.

The paper is structured as follows: Section 2 provides a comprehensive review of prior solar forecasting work, emphasizing the application of machine learning techniques. Section 3 delves into the dataset and data preprocessing tasks, laying the foundation for the subsequent methodology. Section 4 outlines our methodology, presenting tested machine learning models and evaluation metrics. Section 5 gives and discusses experimental results, including comparative analyses and evaluations of our ability to capture short- and long-term patterns. Section 6 explores the interpretation and implications of our findings, reflecting on the study's limitations and suggesting future research avenues. Finally, we summarize our contributions and emphasize the importance of our research in advancing sustainability in solar power generation, paving the way for a greener and more sustainable energy future.

## 2. Related Works

Solar power generation forecasting is critical in integrating renewable energy, ensuring grid stability, and promoting environmental sustainability. The accurate prediction of solar power output is fundamental for effectively managing energy resources. Over time, diverse

methodologies have been proposed to tackle the challenges posed by the intermittent nature of solar irradiance, with the overarching goal of ensuring the reliability of power supply and advancing sustainability objectives. This section comprehensively reviews pertinent research focusing on studies that underscore the intricate connection between solar power forecasting and sustainability. Our examination emphasizes works that quantitatively assess the environmental and economic advantages of precise solar power forecasting. By synthesizing insights from these studies, we aim to contribute a nuanced understanding of the pivotal role that accurate solar power forecasting plays in advancing sustainability within the broader landscape of renewable energy integration.

- **Statistical Models and Machine Learning Algorithms**

Traditionally, solar power generation forecasting heavily relied on statistical models that utilized historical weather and solar power output data, often employing regression analysis for predicting future power output. However, as the field of solar forecasting evolved, machine learning algorithms emerged as a transformative approach, leveraging historical data to train models capable of accurately predicting solar power output. These machine learning techniques have significantly improved the precision of solar forecasting.

Hybrid models combining statistical and machine learning techniques have been proposed to enhance forecasting precision further. By synergizing the strengths of both approaches, these hybrid models hold the potential to improve the reliability of solar power generation forecasts.

In the dynamic landscape of solar forecasting, researchers have explored innovative techniques. Zhang Q et al. [16] introduced the Transform Graph model for electricity net load forecasting, seamlessly integrating Transformer and graph convolutional networks (GCNs) to achieve superior forecasting accuracy and stability, setting a new standard for solar forecasting models. Shifting the focus to environmental predictions, Gupta et al. [17] delved into forecasting harmful algal blooms, employing random forest (RF) and ensemble average (EA) algorithms to showcase the crucial role of predictive accuracy in addressing ecological concerns. Recognizing the importance of accurate forecasts at varying timescales, Hategan S et al. [18] introduced an ensemble model for intra-hour solar resource forecasting, combining statistical extrapolation, machine learning, and all-sky imagery to emphasize the need for selecting suitable forecasting methods based on the specific forecast horizon for optimal accuracy.

In space-based communications and navigation systems, Yarrakula M et al. [19] emphasized the significance of precision in ionospheric total electron content (TEC) predictions. Using machine learning, the study showcased the superior accuracy of this approach compared to traditional models, highlighting the critical role of accurate space weather predictions. Shifting to the environmental realm, Chen, H et al. [20] utilized empirical statistical and machine learning models to predict the abnormal proliferation of *Phaeocystis globosa*, emphasizing the importance of accurate PC concentration indicators and meteorological data in environmental monitoring and prediction. Lastly, Sedai A et al. [21] evaluated a spectrum of models for long-term solar power production forecasting, reinforcing the critical role of accurate long-term forecasts in the renewable energy sector's drive toward sustainability and resilience. These diverse studies collectively contribute to the evolving landscape of solar power forecasting, showcasing the significance of innovative techniques and the integration of various methodologies in advancing the accuracy and applicability of solar generation predictions.

- **Grid Stability and Environmental Impact**

The imperative transition to cleaner and more sustainable energy systems underscores the paramount importance of grid stability and environmental impact considerations. Recent research endeavors have yielded insightful approaches that address these concerns and leverage the opportunities presented by renewable energy technologies.

Elliott M. et al. [22] scrutinized the effects of managing the charging schedules of electric school buses to mitigate simultaneous high loads on the grid, a potential threat to

grid stability. Their study highlights vehicle-to-grid (V2G) services as a viable solution, providing grid control while contributing to stability. The study, which simulates managed chargers using V2G interactions and DC fast chargers, reveals the potential reduction in peak load periods and consequent avoidance of carbon dioxide emissions, emphasizing the pivotal role of innovative charging strategies in grid stability and environmental impact reduction. Temraz A. et al. [23] present a dynamic process model for integrated solar combined cycle (ISCC) power plants, establishing its reliability in assessing capabilities and control structures. The model's accuracy in predicting actual measurements underscores the importance of ISCC plants for grid stability. It connects reliable grid behavior with environmental impact, showcasing the facilitation of renewable energy source integration. Adewuyi, O.B. [24] explores the potential of natural gas-based distributed generation (GtP-DGs) in Nigeria to enhance electricity infrastructure and promote environmental sustainability. Their optimization identifies optimal locations for GtP-DGs and reactive power compensators, enhancing grid stability and demonstrating significant technical, economic, and ecological sustainability benefits.

Mohsin, S.M. et al. [25] tackle forecasting challenges associated with solar and wind energy, crucial for load management, cost reduction, and grid stability. Introducing a harmony search algorithm (HSA)-optimized artificial neural network (ANN) model, their research enhances energy prediction accuracy, contributing to effective load management and cost reduction, environmental preservation, and grid stability. Weidner T. et al. [26] assess the optimal technology mix for building heating in the European Union within planetary boundaries, elucidating the interplay between technology choices, environmental impacts, and grid stability. Emphasizing the need for policy instruments to mitigate increased consumer costs, their work clarifies the intricate connections between grid stability, environmental impact, and economic considerations. Ding, L. [27] critically examines the role of inverter-based resources (IBRs) in modern power grids and their effect on grid stability, providing insights into the differences between various control modes and their impact on grid stability, thereby serving as a bridge between IBRs, grid stability, and the environmental considerations associated with decarbonization.

These papers offer a comprehensive perspective on the inter-relation of grid stability and environmental impact in the context of renewable energy integration. They underscore that innovative strategies, cutting-edge technologies, and robust models are indispensable for achieving grid stability and minimizing environmental footprints during the global transition to cleaner and more sustainable energy systems.

- **Advances in Forecasting Techniques**

In the pursuit of advancing forecasting techniques for solar and renewable energy, researchers have explored innovative methods to enhance accuracy and reliability, transcending conventional models and incorporating technologies such as machine learning, transfer learning, and dynamic optimization. These cutting-edge contributions provide valuable insights into the evolving landscape of renewable energy forecasting, opening new avenues for forecasting excellence and the creation of sustainable energy environments.

Li et al. [28] introduced automated reinforcement learning techniques for power generation, focusing on predicting and scheduling power generation within isolated microgrids. Their research significantly reduced operating costs in microgrid management. However, adapting and extending these techniques to larger, interconnected grids presents distinct challenges. Cao et al. [29] explored the potential of machine learning in renewable energy, focusing on photovoltaic/thermal efficiency. Their research showcased the power of data-driven optimization in simulating and optimizing efficiency. Yet, questions arise regarding integrating these optimizations into the broader domain of solar energy forecasting. Abu-Salih et al. [30] researched short-term solar energy forecasting using smart meter data, developing a long short-term memory neural network (LSTM). While their research showed remarkable performance in short-term forecasting for residential intelligent meter data, challenges exist in scaling these techniques to large, complex, grid-connected solar farms.



Miraftabzadeh et al. [31] explored day-ahead photovoltaic (PV) power prediction, introducing a framework grounded in transfer learning. This technique leverages deep learning models trained on older PV plants to predict the performance of newly installed PV plants. Lim et al. [32] focused on microalgae as a renewable energy source, introducing an advanced forecasting algorithm predicting daily climate conditions a year ahead. This forecast informs a dynamic optimization framework for identifying optimal microalgae biorefinery process pathways. Jakoplić et al. [33] emphasized the growing impact of photovoltaic (PV) systems on power system operation, highlighting the significance of short-term solar forecasting using ground-based cameras.

This suite of papers collectively underscores the ongoing journey towards more precise and adaptable forecasting solutions for renewable energy. Addressing various scales, contexts, and challenges, this research enhances the accuracy of renewable energy production forecasts and promotes a sustainable energy landscape. The integration of these approaches holds the potential to evolve forecasting techniques to meet the diverse demands of the renewable energy sector.

- **Uncertainty Bounds and Error Distributions**

In renewable energy forecasting, the imperative focus on addressing uncertainty has led to diverse approaches aimed at quantifying and managing uncertainties in the context of predictions. This subsection, “Uncertainty Bounds and Error Distributions”, explores a spectrum of methods to navigate the intricate challenge of uncertainty within the domain of renewable energy predictions.

Rodríguez et al. [34] delve into deep learning and probability distributions, presenting a novel method to generate prediction intervals tailored to photovoltaic systems. While promising, this method faces the complex task of fully encapsulating nonlinear error distributions with neural networks. Wang et al. [35] contribute to this discourse by introducing a gray model optimized for long-term energy forecasting in China, offering a unique perspective on addressing uncertainties in renewable energy forecasting. Sobri et al. [36] comprehensively review hybrid approaches for handling uncertainties in renewable energy forecasting, highlighting the delicate balance between interpretability, accuracy, and adaptability.

Shafi et al. [36] propose an innovative artificial neural network (ANN)-based approach to predict power estimates from hybrid wind–solar renewable energy sources. Considering the intermittent and day–night variations in wind and solar intensity, the ANN model is an invaluable tool for real-time assessment and power estimation. The results underscore its effectiveness in enhancing the accuracy and efficiency of renewable energy generation. Omer et al. [37] present a comparative study exploring five ensemble machine learning methods for forecasting photovoltaic (PV) maximum current, addressing the intermittent nature of PV systems. CatBoost stands out, displaying remarkable accuracy, particularly under fast-varying environmental conditions, emphasizing its efficacy in enhancing the reliability and precision of PV power forecasting.

In conclusion, while existing studies have made significant strides in addressing uncertainty bounds and error distributions in renewable energy forecasting, they often fall short in fully capturing the intricate, nonlinear interactions that characterize solar power data. Additionally, many traditional methods struggle to recognize fine-grained patterns within time-series data, limiting their effectiveness in real-time predictions. Our research advances the field by introducing a novel combination of Boosting Cascade Forest and multi-class-grained scanning techniques, which synergistically address these limitations. By integrating ensemble decision tree learning with sophisticated pattern recognition, our approach offers enhanced adaptability, reliability, and precision across various temporal scales. This innovation represents a critical step forward in the ongoing effort to develop more accurate and dependable renewable energy forecasts, contributing to the broader goal of a sustainable energy future.

### 3. Dataset Overview

This section provides a comprehensive overview of the foundational element of our research—the solar power generation dataset. Derived from two utility-scale solar farms in Nasik, Maharashtra, and Gandikotta, Andhra Pradesh, this dataset forms the bedrock of our in-depth analysis of solar energy generation [38]. Publicly accessible on Kaggle, it encompasses measurements from these solar plants, totaling over 8000 examples spanning multiple years [38]. Beyond providing raw data, this dataset plays a pivotal role in bridging the gap between theoretical models and real-world applications, offering a valuable resource for developing and evaluating our forecasting models.

This dataset takes center stage as a beacon of sustainable energy in an era marked by a significant shift towards renewable energy. It transcends being a mere collection of numbers; it is an essential resource for researchers, practitioners, policymakers, and stakeholders striving to harness solar energy effectively for a more sustainable future. Capturing crucial insights about solar energy generation, environmental conditions, and temporal dynamics, this dataset holds significance for the scientific community and individuals aiming to unlock the potential of solar energy, reduce greenhouse gas emissions, and secure a reliable energy supply. The detailed exposition in this section lays the foundation for leveraging these insights in designing, developing, and evaluating robust forecasting models for solar power generation.

The central variable of interest in the dataset is DC\_POWER, representing the direct current power output of the solar panels in kilowatts, serving as the target variable for our prediction models. Table 1 summarizes several input variables capturing factors influencing solar generation, providing environmental context such as site temperature and solar irradiation measurements. Temporal variables like DATE\_TIME enable the analysis of output variations over different timescales.

**Table 1.** Features of the solar energy generation dataset.

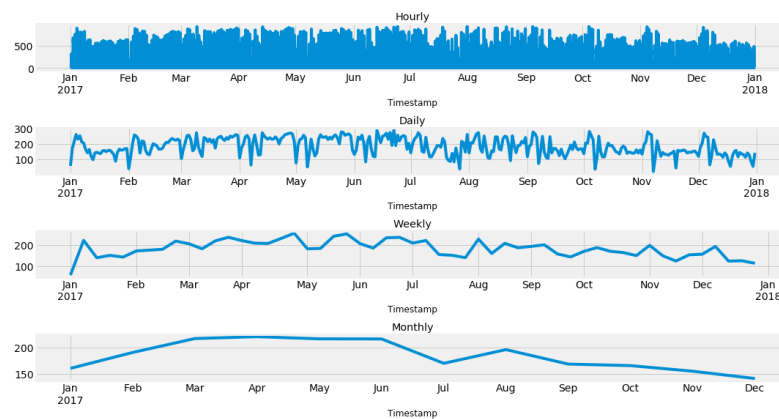
Variable	Description
<b>Target Feature</b>	
DC_POWER	The power output of the solar panels in direct current (DC) units.
<b>Input Feature</b>	
SOURCE_KEY	Unique identifier for the inverter or set of solar panels being monitored.
AMBIENT_TEMPERATURE	The temperature of the air surrounding the solar panels.
MODULE_TEMPERATURE	The temperature of the solar panels themselves.
DAILY_YIELD	This is the total power output of the solar panels for a single day.
TOTAL_YIELD	This is the cumulative power output of the solar panels since they were first installed.
IRRADIATION	Amount of solar radiation received by the solar panels.
DATE_TIME	Date and time of the data readings.

Preprocessing steps encompass handling missing values through interpolation, feature normalization for consistent scales, and temporal alignment of data from multiple sites into a unified data frame. For model training and evaluation, the dataset was chronologically split into 70% training, 15% validation, and 15% test sets to simulate operational forecasting.

While this dataset offers real-world solar farm data, limitations exist concerning the number of sites and sensor measurements available. Enhancements, such as additional geographical locations and meteorological variables, could bolster the robustness of our models.

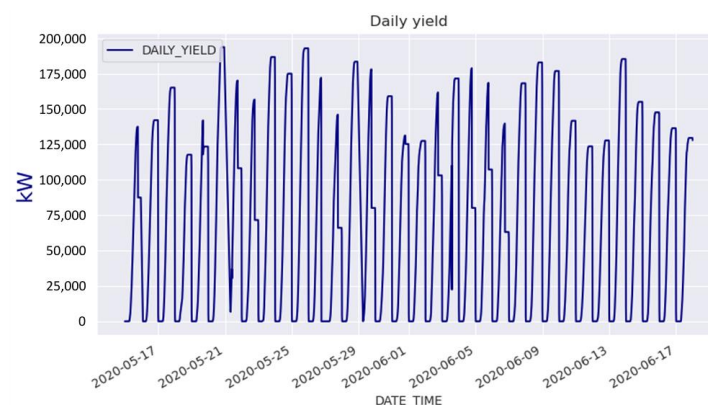
Despite these constraints, the solar power measurements provided by this dataset form a valuable foundation for developing and accessing our proposed forecasting approach.

To understand key characteristics of the solar generation data, Figure 3 provides a sample time-series visualization of solar irradiation measurements from the dataset across hourly, daily, weekly, and monthly timescales. Solar irradiation exhibits significant variability across seasons and times of day. Irradiation is higher in summer around midday, with lower levels in winter and overnight. Irradiation shows much greater hourly and daily aggregate variability than weekly or monthly. This highlights the need to account for short-term and long-term temporal patterns when training models to predict the target DC power output. The multi-class-grained scanning and ensemble techniques proposed in this study are well suited to capture variables and fluctuations across these diverse timescales. However, the model training must cover adequate seasonal and diurnal patterns in the irradiation data to generate accurate forecasts.



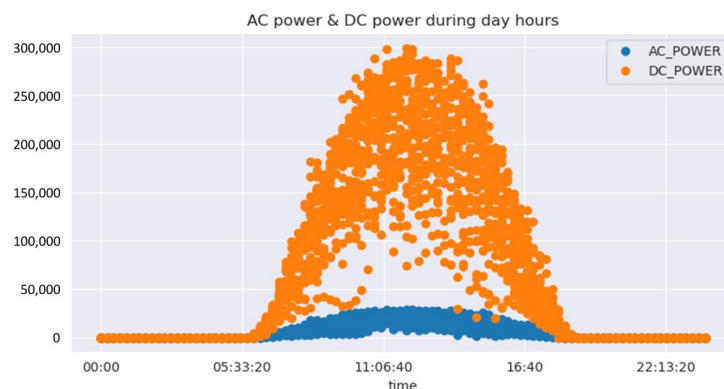
**Figure 3.** Solar irradiation by time: hourly, daily, weekly, and monthly.

Figure 4 presents the daily yield over time, plotted against the date time feature to characterize the solar generation data further. The daily yield in kW hours exhibits high variability, fluctuating between 0 and 20,000 kW hours. This significant volatility highlights the importance of accounting for seasonal and weather-driven effects in predicting solar output. Meanwhile, Figure 5 plots the AC and DC power variables over a sample day. The AC power remains near zero except for during morning hours between 9 am and 4 pm, peaking at around 25,000 kW. The DC power follows a similar daily pattern but reaches much higher levels, up to 300,000 kW in the midday peak-generation hours. The sharp ramp up and down emphasizes the need to capture intraday variability in solar power modeling. The multiple data visualizations provide vital context on distribution and temporal effects, which informed the design choices of our forecasting methodology.



**Figure 4.** Daily solar energy yield over time.





**Figure 5.** AC and DC power output over sample day.

This dataset from two solar farms provides real-world measurements relevant to developing solar forecasting models, including power output, environmental data, and time-based variables. While limited in scope, visual and statistical analyses enabled the characterization of critical data properties like variability, correlations, and temporal patterns. These insights inform design choices for modeling using our proposed machine learning techniques. As such, this section has described the solar generation dataset that supports methodology development, covered next, utilizing multi-class-grained scanning and ensemble learning to predict power output based on these data characteristics.

#### 4. Methodology

Our study uniquely combines Boosting Cascade Forest and multi-class-grained scanning to address the challenges identified in previous research. While Boosting Cascade Forest excels in capturing complex, nonlinear interactions within large datasets through ensemble decision tree learning, it often requires complementary techniques to uncover more subtle patterns in time-series data. This is where multi-class-grained scanning plays a crucial role by detecting fine-grained temporal patterns that are typically overlooked by other models. The synergy between these techniques enables our model to deliver superior accuracy and reliability in solar power forecasting, surpassing the performance of existing methods that typically rely on either ensemble learning or time-series analysis in isolation. This novel integration not only enhances forecasting precision but also offers a more robust framework for managing the inherent variability in solar energy generation.

##### 4.1. Multi-Class-Grained Scanning Technique

The multi-class-grained scanning technology divides time-series data into smaller “grains”. It analyses each grain independently, enabling more granular analysis and pattern detection to improve forecasting accuracy. This technique allows analysts to spot patterns or trends in the data that are not immediately apparent when viewing the data as a whole. The method has been employed in various fields, including finance, health, and weather forecasting, where it has been shown to improve prediction accuracy by including more precise data in the analysis [39].

The multi-class-grained scanning approach, for instance, has been utilized in finance to examine stock market data and pinpoint price patterns connected to certain economic factors. The method has been used in medicine to discover heart rate variability patterns linked to specific medical disorders using ECG data [40]. The technique has been applied to weather forecasting to analyze meteorological data and discover temperature, wind, and humidity patterns linked to certain weather phenomena.

By gathering more specific data about the data and seeing patterns or trends that might not be obvious when analyzing the data, researchers can increase the accuracy of predictions by using the multi-class-grained scanning approach on time-series data. It is crucial to remember that this strategy’s success relies on precise data and analysis goals and that it can call for rigorous testing and fine-tuning to provide the best outcomes [39].

The multi-class-grained scanning approach has been utilized to forecast solar power to spot patterns or trends in weather and solar generation data. This represents the first application of multi-class-grained scanning specifically to solar forecasting. The accuracy of predictions for solar power generation can then be increased by using these patterns or trends as features in machine learning models. For instance, in a recent study, researchers analyzed meteorological and solar production data and found patterns in the data connected to cloud cover and temperature changes. They achieved this by using the multi-class-grained scanning approach. Then, using these patterns as features in a machine learning model, it was possible to estimate solar power generation more accurately than with earlier techniques [39,41].

We utilize a collection of solar generation features, such as SOURCE\_KEY, AMBIENT\_TEMPERATURE, MODULE\_TEMPERATURE, DAILY\_YIELD, TOTAL\_YIELD, IRRADIATION, and DATE\_TIME, as input to the algorithm to predict solar power generation using the multi-class-grained scanning approach. The multi-class-grained scanning technique produces a list of grains, each of which is a fixed-interval subset of the input dataset.

In the Algorithm 1, the user sets the length of each grain ( $L$ ) and the overlap between grains ( $O$ ). The algorithm then creates a set of  $m$  grains by dividing the solar generation features into subsets of length  $L$  with an overlap of  $O$  between adjacent subsets. The output of the multi-class-grained scanning algorithm is a list of grains, which can be used as input to a machine learning model, such as the Boosting Cascade Forest approach, to predict DC\_POWER for each grain.

Below is the algorithm for the multi-class-grained scanning technique applied to solar power generation forecasting using the input and output described above.

---

Algorithm 1: Multi-Class-Grained Scanning for Solar Generation Forecasting

---

**Input:** Solar generation features (SOURCE\_KEY, AMBIENT\_TEMPERATURE, MODULE\_TEMPERATURE, DAILY\_YIELD, TOTAL\_YIELD, IRRADIATION, DATE\_TIME)  
**Output:** A list of grains, where each grain is a subset of the input dataset with a fixed time interval.

Set the length of each grain to  $L$ .  
 Set the overlap between grains  $O$ .  
 Set the start time  $t_0$  to the first timestamp in the dataset.  
 Set the end time  $t_N$  to the last timestamp in the dataset.  
 Initialize an empty list of grains.  
 For each grain  $i$  from 1 to  $m$ :  
 a. Set the start time to  $t_0 + (i - 1) * (L - O)$ .  
 b. Set the end time  $t_i + L - 1$  to  $t_i + L - 1$ .  
 c. Create a dataset subset containing all data points with timestamps between  $t_i$  and  $t_i + L - 1$ .  
 d. Add the subset to the list of grains.  
**Return** the list of grains.

---

The predicted DC\_POWER values for each grain can be combined to generate a final prediction for solar power generation. One way to connect the predicted DC\_POWER values is to calculate the mean value across all grains, which provides a single expected DC\_POWER value for the entire dataset.

The multi-class-grained scanning technique can be combined with other machine learning approaches, such as the Boosting Cascade Forest approach, to improve the accuracy of solar power generation predictions. By dividing the input dataset into multiple grains with a fixed time interval, the multi-class-grained scanning technique allows for more granular and accurate forecasting compared to traditional forecasting methods that use entire datasets.

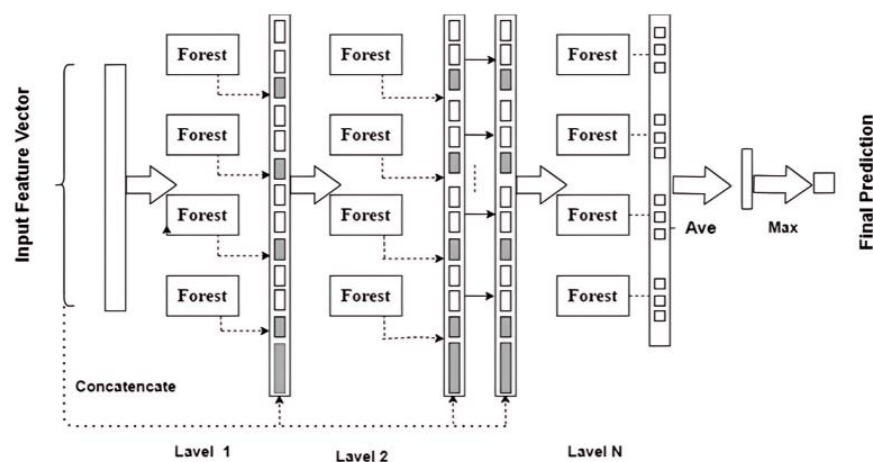
#### 4.2. Boosting Cascade Forest

Boosting Cascade Forest (BCF) is a machine learning algorithm that combines multiple decision tree models in an ensemble to improve overall predictive performance. BCF

leverages the boosting technique to train models on errors from prior models sequentially. The models are arranged in a cascade for multi-stage learning, which is presented in Algorithm 2. BCF is adequate for classification and regression tasks, including power generation forecasting. Power generation forecasting is essential in the energy industry, as accurate forecasts can help utilities and grid operators plan and manage their operations more effectively [41,42].

To use BCF for power generation forecasting, the algorithm needs to be trained on historical data of power generation and weather conditions. The input features could include temperature, humidity, wind speed, solar radiation, and time of day. The algorithm's output forecasts the expected power generation at a given time and location. The training data must be carefully selected and preprocessed to ensure they represent the conditions the algorithm will encounter during deployment.

One potential advantage of BCF for power generation forecasting is its ability to capture complex interactions and dependencies between input features and output [42]. Classification, which involves using decision tree forests to break down the decision-making process into stages, can help capture different aspects of the data. Boosting can also improve the algorithm's accuracy by focusing on examples that are difficult to predict. Figure 6 illustrates the different layers of the BCF algorithm, which combines multiple random forests in a cascade to achieve high performance. The algorithm can be trained using various hyperparameters, such as the number of layers and trees in each layer and the learning rate of the boosting algorithm. These hyperparameters can be tuned using cross-validation or other techniques to improve the algorithm's performance in power generation forecasting [42].



**Figure 6.** Illustration of a typical deep-boosting cascade forest.

However, the performance of BCF for power generation prediction depends on the quality and representativeness of the training data, as well as the specific conditions of the forecast. Factors such as changes in weather patterns, equipment failures, or maintenance schedules can all affect the accuracy of the estimates. The algorithm can be updated and retrained periodically, using new information to address these challenges. Ensemble methods such as BCF can also improve the forecast's robustness by reducing the impact of individual errors or anomalies in the data. With a careful selection of input features, training data, and hyperparameters, BCF has the potential to be a powerful tool for power generation forecasting in the energy industry.

Suppose we have a dataset of solar generation forecasting, with  $n$  examples of solar generation and corresponding weather data. Each example  $x_i$  has  $d$  weather features and a corresponding output  $y_i$  representing the solar generation.

The BCF approach is a machine learning algorithm that can predict solar generation based on weather data. The BCF approach uses the multi-class-grained scanning technique to divide the input dataset into multiple grains with a fixed time interval. For each grain,

the BCF approach trains a cascade of decision trees to predict the DC\_POWER value for that grain. The final prediction is then generated by combining the predicted DC\_POWER values for all grains. With the capabilities to model nonlinear variable interactions and capture complex temporal patterns, BCF is well suited for developing an accurate solar forecasting model using our dataset.

---

#### Algorithm 2: Boosting Cascade Forest (BCF) for Solar Generation Forecasting

---

**Input:** A list of grains, where each grain is a subset of the input dataset with a fixed time interval.

**Output:** The predicted DC\_POWER for each grain.

Initialize the prediction function  $f_0(x) = 0$ .

For each grain  $g$  in the list of grains:

a. Extract the solar generation features from the grain.

b. For each stage  $j$  in the cascade:

- i. Compute the residual error  $r_j(x_i) = y_i - f_{j-1}(x_i)$ , where  $f_{j-1}(x_i)$  is the prediction of the previous stage.
- ii. Train a new Decision Tree  $T_j(x)$  on the residual error  $r_j(x_i)$  using a random subset of the Solar generation features.
- iii. Compute the prediction of the current stage as  $f_j(x) = f_{j-1}(x) + \alpha_j * T_j(x)$ , where  $\alpha_j$  is the learning rate of the boosting algorithm.
- iv. Update the weights of the examples based on their residual errors using the formula  $w_i = \exp(-\gamma * r_j(x_i))$ , where  $\gamma$  is the boosting parameter.
- v. Normalize the weights so that they sum to 1.
- vi. Update the prediction function as  $f(x) = f(x) + f_j(x)$ .

c. Compute the final prediction for the current grain as  $f(g) = 1/J * f(x)$ , where  $J$  is the total number of decision trees in the cascade.

d. Add the predicted DC\_POWER value for the current grain to the list of predicted DC\_POWER values.

**Return** the list of predicted DC\_POWER values.

---

The hyperparameters of the BCF approach include the number of decision trees  $J$ , the depth of each decision tree, the learning rate  $\alpha_j$ , and the boosting parameter  $\gamma$ . These hyperparameters can be tuned using cross-validation or other techniques to optimize the algorithm's performance. By adjusting these hyperparameters, the BCF approach can achieve better accuracy in predicting solar generation, which can help optimize solar power generation and improve grid stability.

#### 4.3. Evaluation Metrics

We employ a set of comprehensive evaluation metrics to evaluate the performance of the Boosting Cascade Forest (BCF) and multi-class-grained scanning techniques in predicting solar power generation. These metrics include the R-squared ( $R^2$ ) score, mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). Collectively, these metrics provide a rigorous assessment of our predictive model accuracy, precision, and robustness in the context of solar power generation forecasting. The  $R^2$  score measures the goodness of fit, while the MAE, MSE, and RMSE offer insights into the magnitude and distribution of prediction errors, ensuring a thorough evaluation of our models.

##### ■ R-squared ( $R^2$ ) score

The  $R^2$  score quantifies the proportion of variance in the target variable, DC\_POWER, that the regression model explains. It indicates how well the model captures the variability in solar power generation. A higher  $R^2$  score signifies a better alignment between the predicted values and the actual observations.

The formula for calculating the  $R^2$  score is presented in Equation (1):

$$R^2 = 1 - \frac{SSE}{TSS} \quad (1)$$

where:

- SSE (sum of squared errors) represents the sum of the squared differences between the actual DC\_POWER values and the predicted DC\_POWER values.
- TSS (total sum of squares) represents the sum of squared differences between the actual DC\_POWER values and the mean DC\_POWER value.

The  $R^2$  score ranges from 0 to 1, where:

- An  $R^2$  score of 0 indicates that the model does not explain any of the variance in the data.
- An  $R^2$  score of 1 implies that the model predicts the target variable perfectly.

We selected the  $R^2$  score as our evaluation metric due to its clarity and interpretability. It provides a straightforward measure of the model's predictive performance and ability to account for variations in solar power generation.

The advantage of using the  $R^2$  score is its ability to gauge the proportion of variability captured by the model, making it an informative metric for assessing the accuracy of solar power generation forecasts. In the context of our study, where precision in predicting solar power output is crucial for grid planning and management, the  $R^2$  score provides valuable insights into the efficacy of our forecasting models.

#### ■ Mean absolute error (MAE)

Mean absolute error (MAE) measures the average magnitude of errors between predicted and actual values. It provides a straightforward way to quantify the model's prediction accuracy. MAE is calculated as presented in Equation (2):

$$MAE = \frac{1}{n} \sum_{i=1}^n |Actual\ i - Predicted\ i| \quad (2)$$

where:

- $n$  represents the number of data points or observations.
- $\frac{1}{n} \sum_{i=1}^n$  indicates the summation of all individual data points.
- *Actual i* represents the actual (observed) value for the  $i$ -th data point.
- *Predicted i* represents the corresponding expected (forecast) value for the  $i$ -th data point.

MAE is chosen because it represents the average prediction error. It is beneficial when we want to understand the magnitude of errors in our forecasts without emphasizing outliers, providing a balanced assessment of prediction accuracy.

#### ■ Mean squared error (MSE)

Mean squared error (MSE) quantifies the average of the squared differences between predicted and actual values. MSE is calculated as presented in Equation (3):

$$MSE = \frac{1}{n} \sum_{i=1}^n (Actual\ i - Predicted\ i)^2 \quad (3)$$

where:

- $n$  represents the number of data points or observations.
- $\frac{1}{n} \sum_{i=1}^n$  Indicates the summation of all individual data points.
- *Actual i* represents the actual (observed) value for the  $i$ -th data point.



- *Predicted  $i$*  represents the corresponding expected (forecast) value for the  $i$ -th data point.

MSE is selected as it penalizes more significant errors due to the squaring operation. It is valuable when we want to emphasize and understand the impact of more significant errors on the overall prediction performance.

- **Root mean squared error (RMSE)**

Root mean squared error (RMSE) is the square root of the MSE, measuring the standard deviation of the prediction errors. It is calculated as presented in Equation (4):

$$RMSE = \sqrt{MSE} \quad (4)$$

where:

- MSE is the squared error metric, the same as for MSE.

RMSE is chosen as it provides an easily interpretable metric in the same units as the target variable. It is a valuable choice when we want to assess prediction accuracy while retaining the original measurement scale, allowing for a more intuitive understanding of error magnitude.

#### 4.4. Computational Environment and Reproducibility

To ensure the reproducibility of the experiments and analyses conducted in this study, we provide the details of the computational environment and the software configurations employed. All model training, testing, and evaluations were conducted on a DELL Precision 5560 workstation equipped with a 32 GB DDR4 RAM and 1 TB NVMe 2 SSD for high-speed storage. The system was powered by an Nvidia RTX A2000 graphics card, enabling accelerated computations, particularly for tasks requiring significant processing power, such as the training of machine learning models.

The software environment consisted of Python 3.11, utilized within both JetBrains PyCharm and Jupyter Notebook. These development environments facilitated efficient model development, code management, and real-time experimentation.

Additionally, the system operated on Windows 11, providing a stable platform for all experiments. Specific parameters, such as batch size, learning rate, and optimizer settings, were fine-tuned within this environment to achieve optimal performance. The details of these hyperparameter settings are available within the code repository, ensuring that future researchers can replicate and build upon this work with minimal deviation.

## 5. Results and Analysis

This section represents the focal point of our study, where we unveil the outcomes of our efforts in creating a robust machine learning model for solar energy generation forecasting. Our methodology rests on a unique fusion of the Boosting Cascade Forest and multi-class-grained scanning techniques. The choice of these techniques was deliberate, as we aimed to elevate our predictions' precision significantly. We based our work on publicly available data from two solar power facilities, one in Nasik, Maharashtra, and the other in Gandikota, Andhra Pradesh. The dataset we harnessed goes beyond mere solar energy generation patterns; it includes a comprehensive array of environmental parameters and performance monitoring data.

This implementation represents a pioneering effort as the first instance of applying the Boosting Cascade Forest and multi-class-grained scanning techniques in the solar sector. These diverse and rich datasets have emerged as a cornerstone of our research. They offer insight into the multifaceted dynamics of solar power generation and provide the essential metrics required for our predictive model meticulous development and rigorous evaluation. The multidimensional nature of the datasets has significantly contributed to the model's robustness and precision.

Beyond immediate application, this comprehensive analysis offers a profound understanding of solar energy generation dynamics. By delving into the intricate relationships between environmental variables and solar power generation, our study serves as an essential stepping stone for future research endeavors in solar forecasting. It enhances our collective capability to harness the full potential of solar energy as a renewable and sustainable power source.

To evaluate our forecasting model, we examined three key metrics: mean absolute error (MAE), mean squared error (MSE), and root mean squared error (RMSE). The MAE measures the average magnitude of errors by calculating the absolute differences between predicted and actual values. A lower MAE indicates better model performance. The MSE, on the other hand, combines bias and variability by squaring the differences before averaging. The RMSE, derived from MSE, emphasizes more significant errors. Our model performed exceptionally well, showcasing a low MAE of 0.0016, a low MSE of  $1.33 \times 10^{-5}$ , and a low RMSE of 0.0036.

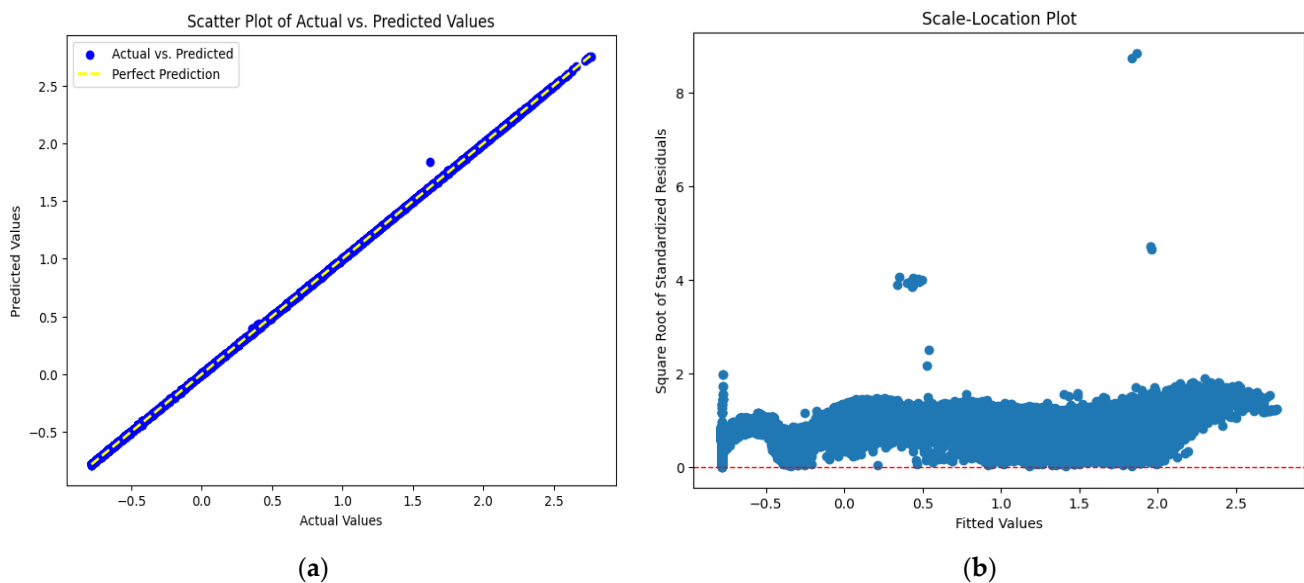
Our solar forecasting model achieved an impressive test set  $R^2$  score of 99.69%. This score indicates that it explained over 99% of the variance in real-world data as presented in Table 2. To provide context, we benchmarked our model's  $R^2$  against prior studies that used the same solar power generation dataset but employed different modeling approaches. Our model outperformed all benchmarks, achieving an  $R^2$  score of more than 0.25, higher than the next best technique. This demonstrates the significant accuracy gains unlocked by our integrated methodology in capturing complex spatiotemporal relationships within the data to predict solar output precisely.

**Table 2.** Benchmark  $R^2$  scores on solar forecasting dataset.

Model	$R^2$ Score	Reference
<i>Our Approach</i>	0.9969	-
SARIMAX	0.986854	[43]
Prophet	0.895611	
Variational Mode Decomposition (VMD) combined with Fuzzy-Twin Support Vector Machine Model	0.9564	[44]
TensorFlow Neural Network	0.9860	[45]
Auto Encoder LSTM	0.8963	[46]

The high accuracy of the developed model has important implications for the practical application of solar energy generation forecasting. Accurate predictions can optimize the operation of solar power systems, ensuring peak efficiency and reducing maintenance costs. Furthermore, precise forecasting can enhance the integration of solar power into the electricity grid, ensuring efficient utilization of solar energy.

In Figure 7a, we present a scatter plot illustrating the relationship between predicted values and actual observations in our solar energy generation forecasting model. Each point on the plot corresponds to a specific observation. The x-axis represents the predicted values, while the y-axis depicts actual observed values. A well-performing model is expected to exhibit a perfect linear correlation, with all points aligned along the 45-degree diagonal line, indicating precise predictions. Our scatter plot demonstrates strong alignment, with a minor scattering of points deviating slightly from the ideal line, implying minor prediction errors or outliers. Overall, the scatter plot underscores the effectiveness of our model in accurately predicting solar energy generation, as evidenced by the tight cluster of data points close to the ideal linear relationship.



**Figure 7.** Scatter Plot (a) and Scale-location Plot (b) for the Actual vs. Predicted data values.

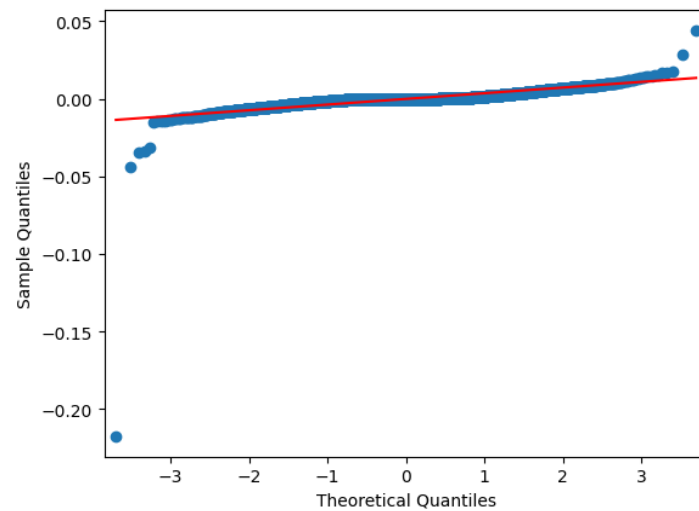
In Figure 7b, we introduce a scale-location plot (spread vs. fitted) to analyze our model's predictive performance better. This plot focuses on the spread or variability in the model's residuals concerning the fitted values. The square root of standardized residuals, a measure of the difference between the actual and predicted values, is plotted on the y-axis, while the x-axis represents the fitted values. Our scale-location plot confirms the model's good performance, with a consistent spread of residuals around zero for most data points. Most residuals are contained within a range between 0 and 2, indicating relatively small prediction errors and a well-behaved model. However, a few outliers with square roots of standardized residuals in the field of 4 to 9 are present, suggesting specific observations where the model may underperform. These outliers are limited in number and call for further investigation to enhance predictive accuracy.

The presence of these outliers suggests potential instances where the model's predictive accuracy significantly deviates from the observed values. To understand their impact on the model's overall performance, we conducted a detailed analysis of these outliers. We found that the majority of these outliers corresponded to extreme weather conditions, such as sudden cloud cover or unexpected shifts in solar irradiance, which were not fully captured by the model's input features. These instances highlight the model's sensitivity to abrupt changes in environmental conditions, which are difficult to predict with the available data. Despite the presence of these outliers, the overall performance of the model remains robust, as evidenced by the high  $R^2$  score and low error metrics. However, future work could explore incorporating additional features, such as real-time satellite data or advanced sensor networks, to better capture these extreme variations and further enhance the model's reliability.

The combined insights from the scatter plot and the scale-location plot offer a comprehensive view of our model's forecasting capabilities. While the scatter plot demonstrates a strong overall alignment between the predicted and actual values, the scale-location plot delves into the behavior of residuals, revealing that most are contained within an acceptable range. The few outliers in the scale-location plot provide valuable pointers for future work, helping us identify areas of improvement and emphasizing the need for more focused analysis. Together, these plots serve as powerful tools for assessing the strengths and weaknesses of our solar energy generation forecasting model, ultimately contributing to more informed and data-driven decision making in renewable energy.

Figure 8 shows the quantile–quantile (QQ) plot, which compares the observed and predicted solar energy generation data distribution. The QQ plot demonstrates that most points closely align with the red line, signifying similar distributions between the observed

and predicted data [47]. While a few points deviate from the red line, indicating discrepancies in specific quantiles, the overall agreement confirms the adequate performance of our methodology, supported by low error metrics such as MAE, MSE, and RMSE.



**Figure 8.** QQ plot comparing observed and predicted solar energy generation.

The QQ plot shows that most points fall close to the red line, indicating that the observed and predicted solar energy generation data have similar distributions. However, a few points fall far from the red line, meaning there may be discrepancies between the observed and predicted data in specific quantiles. Despite these discrepancies, the overall agreement between the observed and forecast data on the QQ plot confirms that the performance of the developed methodology, which utilized Boosting Cascade Forest and multi-class-grained scanning, has been excellent for solar generation prediction. The low values of the error metrics, including MAE, MSE, and RMSE, further support this conclusion.

The excellent agreement between the observed and predicted data on the QQ plot suggests that the developed methodology effectively predicts solar energy generation. The study results have important implications for the practical application of solar energy generation forecasting, as accurately predicting solar energy generation can help optimize the operation of solar power systems and reduce costs associated with maintaining and operating these systems. The study's findings can also inform the development of more sophisticated forecasting models that can capture the complex relationships between environmental factors and solar energy generation and help advance state-of-the-art solar energy generation forecasting.

## 6. Conclusions

This study presents a pioneering machine learning-based solar forecasting model that leverages a unique combination of multi-class-grained scanning and Boosting Cascade Forest techniques. The model's performance evaluation is grounded in real-world data from two utility-scale solar farms, marking a substantial contribution to renewable energy research.

The results of this research underscore the model's exceptional predictive capabilities, as evidenced by its low error metrics, including MAE, and a remarkable  $R^2$  score of 0.99 on unseen testing data. Furthermore, the QQ plot visualizations reveal a compelling alignment between actual and forecast outputs, confirming the model's ability to capture intricate data patterns accurately. These quantitative and graphical validations provide robust evidence of the methodology's effectiveness.

Beyond the realm of academia, our findings hold profound practical implications. Accurate solar power generation forecasting is pivotal in optimizing maintenance sched-

ules, enhancing grid integration and stability, and curbing renewable energy costs and waste. However, this study also reveals opportunities for future enhancement. Subsequent research endeavors may consider the integration of satellite weather feeds, or explore alternative machine learning architectures fine-tuned for solar data, such as convolutional neural networks (CNNs) for spatial data analysis or recurrent neural networks (RNNs) for temporal sequence modeling. Additionally, addressing challenges related to computational efficiency and model interpretability could further improve the practical applicability of the model in real-world scenarios.

In summation, this work not only advances the field by delivering state-of-the-art results but also emphasizes the transformative potential of artificial intelligence in addressing the critical challenge of solar variability and uncertainty. As we look ahead, the research community and industry stakeholders should recognize the promise of AI advances as a catalyst for accelerating the transition toward sustainable energy solutions.

**Author Contributions:** Conceptualization, M.K.B., Y.F. and M.A.; methodology, M.K.B.; software, M.A.; validation, Y.F., M.A. and M.M.N.; formal analysis, Y.F.; investigation, A.S.; resources, M.K.B. and M.M.N.; data curation, A.S.; writing—original draft preparation, M.K.B.; writing—review and editing, M.K.B. and M.A.; visualization, Y.F. and A.S.; supervision, M.M.N. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors would like to acknowledge the support of Prince Sultan University for paying the Article Processing Charges (APC) of this publication.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data are contained within the article.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Mellit, A.; Kalogirou, S.A.; Hontoria, L.; Shaari, S. Artificial intelligence techniques for sizing and power forecasting of grid-connected photovoltaic systems: A review. *Renew. Sustain. Energy Rev.* **2015**, *44*, 376–393.
- Sobri, S.; Koohi-Kamali, S.; Rahim, N.A. Solar photovoltaic generation forecasting methods: A review. *Energy Convers. Manag.* **2018**, *156*, 459–497. [[CrossRef](#)]
- National Renewable Energy Laboratory (NREL). Robert Margolis, Solar Futures Study Energy Analysis. Available online: <https://www.nrel.gov/analysis/solar-futures.html> (accessed on 24 February 2024).
- Antonanzas, J.; Osorio, N.; Escobar, R.; Urraca, R.; Martinez-De-Pison, F.J.; Antonanzas-Torres, F. Review of photovoltaic power forecasting. *Sol. Energy* **2016**, *136*, 78–111. [[CrossRef](#)]
- Wang, W.; Dunford, W.; Pisu, P.; Akyol, K.; Siegert, J. Distributed photovoltaic spatial-temporal forecast: A review. *Renew. Sustain. Energy Rev.* **2020**, *127*, 109897.
- Inman, R.H.; Pedro, H.T.; Coimbra, C.F. Solar forecasting methods for renewable energy integration. *Prog. Energy Combust. Sci.* **2013**, *39*, 535–576. [[CrossRef](#)]
- Vemparala, S.R.; Bhaskar, M.S.; Elmorshedy, M.F.; Almakhlles, D. Performance Enhancement of Renewable System via Hybrid Switched-Inductor-Capacitor Converter. In Proceedings of the 2024 6th Global Power, Energy and Communication Conference (GPECOM), Budapest, Hungary, 4–7 June 2024; pp. 79–84.
- Hissou, H.; Benkirane, S.; Guezzaz, A.; Beni-Hssane, A.; Azrou, M. Advanced Prediction of Solar Radiation Using Machine Learning and Principal Component Analysis. In *Artificial Intelligence, Data Science and Applications*; Farhaoui, Y., Hussain, A., Saba, T., Taherdoost, H., Verma, A., Eds.; Springer Nature: Cham, Switzerland, 2024; pp. 201–207. [[CrossRef](#)]
- Hissou, H.; Benkirane, S.; Guezzaz, A.; Azrou, M.; Beni-Hssane, A. A lightweight time series method for prediction of solar radiation. *Energy Syst.* **2024**, *15*, 1–38. [[CrossRef](#)]
- Ghalib, M.A.; Hamad, S.A.; Elmorshedy, M.F.; Almakhlles, D.; Ali, H.H. Beta Maximum Power Extraction Operation-Based Model Predictive Current Control for Linear Induction Motors. *J. Sens. Actuator Netw.* **2024**, *13*, 37. [[CrossRef](#)]
- Elmorshedy, M.F.; Bhaskar, M.S.; Almakhlles, D.; Kotb, K.M. Relegated Thrust Ripples for Linear Induction Motors Based-Four Voltage Vectors Finite-Set Predictive Control and Model Reference Adaptive System. *Electr. Power Compon. Syst.* **2024**, *10*, 1–12. [[CrossRef](#)]
- Elmorshedy, M.F.; Almakhlles, D.; Allam, S.M. Improved performance of linear induction motors based on optimal duty cycle finite-set model predictive thrust control. *Heliyon* **2024**, *10*, e34169. [[CrossRef](#)]



13. Boutahir, M.K.; Hessane, A.; Farhaoui, Y.; Azrou, M.; Benyeogor, M.S.; Innab, N. Meta-Learning Guided Weight Optimization for Enhanced Solar Radiation Forecasting and Sustainable Energy Management with VotingRegressor. *Sustainability* **2024**, *16*, 5505. [CrossRef]
14. Hissou, H.; Benkirane, S.; Guezzaz, A.; Azrou, M.; Beni-Hssane, A. A Novel Machine Learning Approach for Solar Radiation Estimation. *Sustainability* **2023**, *15*, 10609. [CrossRef]
15. Elsaraiti, M.; Merabet, A. Solar Power Forecasting Using Deep Learning Techniques. *IEEE Access* **2022**, *10*, 31692–31698. [CrossRef]
16. Zhang, Q.; Chen, J.; Xiao, G.; He, S.; Deng, K. TransformGraph: A novel short-term electricity net load forecasting model. *Energy Rep.* **2023**, *9*, 2705–2717. [CrossRef]
17. Gupta, A.; Hantush, M.M.; Govindaraju, R.S. Sub-monthly time scale forecasting of harmful algal blooms intensity in Lake Erie using remote sensing and machine learning. *Sci. Total. Environ.* **2023**, *900*, 165781. [CrossRef]
18. Hategan, S.-M.; Stefu, N.; Paulescu, M. An Ensemble Approach for Intra-Hour Forecasting of Solar Resource. *Energies* **2022**, *16*, 6608. [CrossRef]
19. Yarrakula, M.; Prabakaran, N.; Dabbakuti, J.K. Machine learning based approach for modeling and forecasting of GPS–TEC during diverse solar phase periods. *Acta Astronaut.* **2023**, *206*, 177–186. [CrossRef]
20. Chen, H.; Yao, H.; Liao, P.; Wen, K.; Huang, Y.; Zhong, W. Prediction of abnormal proliferation risk of *Phaeocystis globosa* based on correlation mining of PC concentration indicator and meteorological factors along Qinzhou Bay, Guangxi. *J. Sea Res.* **2023**, *192*, 102365. [CrossRef]
21. Sedai, A.; Dhakal, R.; Gautam, S.; Dhamala, A.; Bilbao, A.; Wang, Q.; Wigington, A.; Pol, S. Performance Analysis of Statistical, Machine Learning and Deep Learning Models in Long-Term Forecasting of Solar Power Production. *Forecasting* **2023**, *5*, 256–284. [CrossRef]
22. Elliott, M.; Kittner, N. Operational grid and environmental impacts for a V2G-enabled electric school bus fleet using DC fast chargers. *Sustain. Prod. Consum.* **2022**, *30*, 316–330. [CrossRef]
23. Temraz, A.; Alobaid, F.; Link, J.; Elweteedy, A.; Epple, B. Development and Validation of a Dynamic Simulation Model for an Integrated Solar Combined Cycle Power Plant. *Energies* **2021**, *14*, 3304. [CrossRef]
24. Adewuyi, O.B.; Kiptoo, M.K.; Adebayo, I.G.; Senjyu, T. Techno-economic analysis of robust gas-to-power distributed generation planning for grid stability and environmental sustainability in Nigeria. *Sustain. Energy Technol. Assess.* **2023**, *55*, 102943. [CrossRef]
25. Mohsin, S.M.; Maqsood, T.; Madani, S.A. Solar and Wind Energy Forecasting for Green and Intelligent Migration of Traditional Energy Sources. *Sustainability* **2022**, *14*, 16317. [CrossRef]
26. Weidner, T.; Guillén-Gosálbez, G. Planetary boundaries assessment of deep decarbonisation options for building heating in the European Union. *Energy Convers. Manag.* **2023**, *278*, 116602. [CrossRef]
27. Ding, L.; Lu, X.; Tan, J. Small-Signal Stability Analysis of Low-Inertia Power Grids with Inverter-Based Resources and Synchronous Condensers. In Proceedings of the 2022 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT), New Orleans, LA, USA, 24–28 April 2022; pp. 1–5.
28. Li, Y.; Wang, R.; Yang, Z. Optimal scheduling of isolated microgrids using automated reinforcement learning-based mul-ti-period forecasting. *IEEE Trans. Sustain. Energy* **2021**, *13*, 159–169. [CrossRef]
29. Cao, Y.; Kamrani, E.; Mirzaei, S.; Khandakar, A.; Vaferi, B. Electrical efficiency of the photovoltaic/thermal collectors cooled by nanofluids: Machine learning simulation and optimization by evolutionary algorithm. *Energy Rep.* **2022**, *8*, 24–36. [CrossRef]
30. Abu-Salih, B.; Wongthongtham, P.; Morrison, G.; Coutinho, K.; Al-Okaily, M.; Huneiti, A. Short-term renewable energy consumption and generation forecasting: A case study of Western Australia. *Heliyon* **2022**, *8*, e09152. [CrossRef] [PubMed]
31. Miraftabzadeh, S.M.; Colombo, C.G.; Longo, M.; Foadelli, F. A Day-Ahead Photovoltaic Power Prediction via Transfer Learning and Deep Neural Networks. *Forecasting* **2023**, *5*, 213–228. [CrossRef]
32. Lim, J.Y.; Teng, S.Y.; How, B.S.; Nam, K.; Heo, S.; Máša, V.; Stehlik, P.; Yoo, C.K. From microalgae to bioenergy: Identifying optimally integrated biorefinery pathways and harvest scheduling under uncertainties in predicted climate. *Renew. Sustain. Energy Rev.* **2022**, *168*, 112865. [CrossRef]
33. Jakoplić, A.; Franković, D.; Kirinčić, V.; Plavšić, T. Benefits of short-term photovoltaic power production forecasting to the power system. *Optim. Eng.* **2020**, *22*, 9–27. [CrossRef]
34. Rodríguez, F.; Galarza, A.; Vasquez, J.C.; Guerrero, J.M. Using deep learning and meteorological parameters to forecast the photovoltaic generators intra-hour output power interval for smart grid control. *Energy* **2022**, *239*, 122116. [CrossRef]
35. Wang, Y.; He, X.; Zhang, L.; Ma, X.; Wu, W.; Nie, R.; Chi, P.; Zhang, Y. A novel fractional time-delayed grey Bernoulli forecasting model and its application for the energy production and consumption prediction. *Eng. Appl. Artif. Intell.* **2022**, *110*, 104683. [CrossRef]
36. Shafi, I.; Khan, H.; Farooq, M.S.; Diez, I.d.I.T.; Miró, Y.; Galán, J.C.; Ashraf, I. An Artificial Neural Network-Based Approach for Real-Time Hybrid Wind-Solar Resource Assessment and Power Estimation. *Energies* **2023**, *16*, 4171. [CrossRef]
37. Omer, Z.M.; Shareef, H. Comparison of decision tree based ensemble methods for prediction of photovoltaic maximum current. *Energy Convers. Manag. X* **2022**, *16*, 100333. [CrossRef]
38. Kannal, A. Solar Power Generation Data. Kaggle.com. Available online: <https://www.kaggle.com/anikannal/solar-power-generation-data> (accessed on 26 December 2023).
39. Hu, G.; Li, H.; Xia, Y.; Luo, L. A deep Boltzmann machine and multi-grained scanning forest ensemble collaborative method and its application to industrial fault diagnosis. *Comput. Ind.* **2018**, *100*, 287–296. [CrossRef]

40. Yin, L.; Zhao, L.; Yu, T.; Zhang, X. Deep Forest Reinforcement Learning for Preventive Strategy Considering Automatic Generation Control in Large-Scale Interconnected Power Systems. *Appl. Sci.* **2018**, *8*, 2185. [CrossRef]
41. Guo, Y.; Liu, S.; Li, Z.; Shang, X. BCDForest: A boosting cascade deep forest model towards the classification of cancer subtypes based on gene expression data. *BMC Bioinform.* **2018**, *19*, 118. [CrossRef]
42. Pati, S.K.; Ghosh, A.; Banerjee, A.; Roy, I.; Ghosh, P.; Kakar, C. Data Analysis on Cancer Disease Using Machine Learning Techniques. In *Advanced Machine Learning Approaches in Cancer Prognosis*; Intelligent Systems Reference Library; Nayak, J., Favorskaya, M.N., Jain, S., Naik, B., Mishra, M., Eds.; Springer: Cham, Switzerland, 2021; Volume 204. [CrossRef]
43. Carmine Minichini “How to Manage a Solar Power Plant”. Available online: <https://www.kaggle.com/code/virosky/how-to-manage-a-solar-power-plant> (accessed on 26 December 2023).
44. Balraj, G.; Victoire, A.A.; Victoire, A. Variational mode decomposition combined fuzzy—Twin support vector machine model with deep learning for solar photovoltaic power forecasting. *PLoS ONE* **2022**, *17*, e0273632. [CrossRef]
45. Michael from Old Wick “Solar Generation Predictions—TF Neural Network”. Available online: <https://www.kaggle.com/code/michaelfromoldwick/solar-generation-predictions-tf-neural-network> (accessed on 26 December 2023).
46. Ibrahim, M.; Alsheikh, A.; Awaysheh, F.M.; Alshehri, M.D. Machine Learning Schemes for Anomaly Detection in Solar Power Plants. *Energies* **2022**, *15*, 1082. [CrossRef]
47. Marden, J.I. Positions and QQ Plots. *Stat. Sci.* **2004**, *19*, 606–614. [CrossRef]

**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.