Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 5, 2025

DEVELOPMENT OF A HYBRID ARTIFICIAL INTELLIGENCE FRAMEWORK FOR ACCURATE FORECASTING OF SOLAR POWER GENERATION USING MACHINE LEARNING ALGORITHMS AND TIME-SERIES ANALYSIS

Bushra Mehmood[*1], Geeta Rani[2], Nauman Khalid[3], Muhammad Kashif Majeed[4], Basit Ahmad[5], Shumaila Qamar[6], Muhammad Saeed[7]

[*1]*Department of Computer Science, Iqra University, Islamabad, Pakistan.*
[2]*Department of Data Science, Iqra University, Karachi, Pakistan.*
[3]*Department of Computer Science, Federal Urdu University of Arts, Sciences and Technology, Islamabad, Pakistan*
[4]*Faculty of Engineering Science and Technology, Iqra University, Karachi 75500, Pakistan. School of Electronic Science and Technology, Xi'an Jiaotong University, Xi'an 710049, China.*
[5]*Department of Electrical Engineering, NFC Institute of Engineering and Technology Multan, Punjab, Pakistan*
[6]*Department of Computer Science, FEST, Iqra University, Karachi, Pakistan.*
[7]*Department of Electrical Engineering, University of Engineering and Technology (UET), Lahore, Pakistan.*

[*1]samra.mehmood12@gmail.com, [2]geetarajeshgul@gmail.com, [3]naumankhalid773@gmail.com, [4]mkashif@iqra.edu.pk, [5]basitahmad3884@gmail.com, [6]shumaila.qamar@iqra.edu.pk, [7]saeed.uetlhr55@gmail.com

DOI: https://doi.org/10.5281/zenodo.15478903

Keywords
Artificial Neural Networks, Hybrid Models, Support Vector Machines, Grid Optimization, Decision Trees, Exponential Smoothing, Machine Learning Algorithms.

Article History
Received on 13 April 2025
Accepted on 13 May 2025
Published on 21 May 2025

Corresponding Author: *
Bushra Mehmood

Abstract
Accurate forecasting of solar power generation plays a crucial role in the efficient integration of solar energy into modern power grids and renewable energy systems. Traditional forecasting methods, such as statistical and physical models, often fail to capture the complex patterns and dynamic nature of solar radiation and its impact on power output. In this paper, we propose the development of a hybrid artificial intelligence (AI) framework that combines machine learning (ML) algorithms with time-series analysis to improve the accuracy and reliability of solar power generation predictions. The framework integrates several machine learning techniques, including decision trees, support vector machines (SVM), and artificial neural networks (ANN), with time-series forecasting methods such as autoregressive integrated moving average (ARIMA) and exponential smoothing to better model both short-term fluctuations and long-term trends in solar power output. By leveraging the complementary strengths of machine learning in data pattern recognition and time-series analysis in trend forecasting, the hybrid model enhances the precision of predictions under a variety of environmental conditions and temporal scales. To evaluate the performance of the proposed framework, we conducted a series of experiments using real-world solar power datasets. The results show that the hybrid model significantly outperforms traditional forecasting approaches in terms of both forecasting accuracy and robustness, particularly in capturing complex seasonal and diurnal variations in solar power generation. Moreover, the model demonstrates its ability to adapt to different locations and

varying weather conditions, making it highly applicable for diverse geographical regions. The results highlight the potential of integrating AI-driven forecasting models with time-series analysis as a powerful tool for optimizing solar power generation and enhancing the management of renewable energy resources. Ultimately, this framework can serve as an essential tool for energy providers, grid operators, and policymakers to improve grid stability, reduce uncertainties in solar power predictions, and enable more efficient integration of solar energy into the existing energy infrastructure.

INTRODUCTION

The global energy landscape is undergoing a profound transformation driven by the urgent need to mitigate climate change, enhance energy security, and promote sustainable development. As countries around the world strive to decarbonize their power systems, solar energy has emerged as one of the most promising and rapidly growing renewable energy sources. With the advent of solar photovoltaic (PV) technology and its declining costs, solar power has become increasingly accessible, scalable, and viable for both centralized and distributed energy systems. However, despite its numerous advantages, solar energy integration presents significant technical challenges due to its inherent intermittency and dependence on weather and environmental conditions. Solar power generation is highly sensitive to a variety of meteorological factors, including solar irradiance, cloud cover, temperature, humidity, and atmospheric pressure [1]. These factors are not only variable but also non-deterministic, leading to substantial fluctuations in solar output across different timescales from minutes and hours to days and seasons. Such variability complicates power system operations, especially in real-time grid management, load balancing, and energy dispatch. Consequently, the ability to accurately forecast solar power generation is essential for ensuring the reliability, efficiency, and stability of electricity grids with high penetration of renewable energy. Forecasting solar power enables utilities, energy providers, and grid operators to anticipate fluctuations in energy supply and adjust operational strategies accordingly. It aids in scheduling backup generation, optimizing energy storage systems, planning demand response, and managing power exchanges in electricity markets [2]. Moreover, accurate solar forecasting reduces the need for costly spinning reserves and minimizes the risk of load shedding or over-generation, thereby improving the overall economics and sustainability of solar power integration. Given these benefits, solar power forecasting has become a pivotal area of research within the domains of renewable energy, power systems, and artificial intelligence. Traditionally, solar forecasting methods have been classified into two broad categories: physical models and statistical models. Physical models rely on atmospheric physics, radiative transfer equations, and satellite imagery to simulate the solar radiation received at the Earth's surface. These models incorporate variables such as aerosol concentration, cloud movement, and solar geometry to estimate solar irradiance and PV output. While they offer interpretability and a physically grounded understanding of solar phenomena, physical models are computationally intensive, require extensive meteorological data, and may lack accuracy in highly dynamic or localized conditions. Statistical models, on the other hand, leverage historical data to identify recurring patterns and forecast future values based on time-series analysis. Techniques such as autoregressive integrated moving average (ARIMA), exponential smoothing, and linear regression have been widely applied in solar forecasting tasks [3]. These models are typically easier to implement and computationally efficient. However, they are limited by their assumption of linearity and stationarity, making them less effective in capturing the complex, nonlinear, and stochastic behavior of solar power generation, especially under non-standard or changing environmental conditions. To overcome the shortcomings of traditional methods, the focus has shifted toward machine learning (ML) techniques, which offer a data-driven, adaptive, and model-free approach to forecasting. ML algorithms are particularly well-suited for modeling nonlinear relationships and uncovering hidden patterns in large datasets without prior knowledge of the underlying physical mechanisms. Techniques such as artificial neural networks (ANN), support vector machines (SVM), random forests, and gradient boosting have shown promising results in solar forecasting [4]. These methods are capable of learning from complex input features, such as weather variables and historical PV output, to produce accurate predictions over various timescales. Despite their advantages, standalone ML models also face certain limitations. Many machine learning algorithms are designed for regression or classification tasks and do not inherently account for the temporal structure of sequential data. As a result, they may perform suboptimally in forecasting problems where autocorrelations, trends, and seasonality play a critical role. Moreover, ML models often require large volumes of high-quality data for training and are sensitive to hyperparameter settings, which can affect their generalizability and robustness in real-world applications. Table 1 shows the comparison of traditional and machine learning based forecasting. This table provides an overview of the strengths and limitations of traditional statistical and physical forecasting methods compared to machine learning-based approaches, highlighting key differences in their ability to capture complex patterns and handle large datasets.

Table 1: Comparison of Traditional and Machine Learning-Based Forecasting Methods [5].

Methodology	Strengths	Limitations
Statistical Models	Effective for identifying linear trends and short-term forecasting	Limited in capturing nonlinearities and complex patterns
Physical Models	Based on physical principles and atmospheric data	Computationally intensive and difficult to generalize across regions
Machine Learning Models	Can model complex nonlinear patterns and large datasets	Requires large datasets for training, and may struggle with time-series dependencies

This has led to the emergence of hybrid models that aim to combine the strengths of machine learning and time-series analysis into a unified framework. Hybrid models are particularly attractive for solar forecasting because they can capture both the nonlinear, short-term fluctuations using ML techniques and the long-term trends and temporal dependencies using time-series models. For instance, combining an ANN with an ARIMA model allows the former to learn residual nonlinearities while the latter models the underlying linear patterns and autocorrelations [6]. Such hybridization not only improves prediction accuracy but also enhances model interpretability and robustness. Table 2 shows the performance metrics of hybrid model vs traditional forecasting. This table presents a comparison of the performance metrics (RMSE, MAE, MAPE) between the proposed hybrid AI framework and traditional forecasting models (ARIMA, SVM, ANN), demonstrating the superior accuracy and reliability of the hybrid model.

Table 2: Performance Metrics of Hybrid Model vs. Traditional Forecasting Models [7].

Model	RMSE (Root Mean Squared Error)	MAE (Mean Absolute Error)	MAPE (Mean Absolute Percentage Error)
Hybrid AI Framework	2.5	1.8	7.4%
ARIMA	3.2	2.4	9.8%
SVM	3.5	2.7	10.3%
ANN	3.0	2.1	8.2%

In this study, we propose the development of a hybrid artificial intelligence framework for accurate forecasting of solar power generation. The framework integrates a range of machine learning algorithms including decision trees, support vector machines (SVM), and artificial neural networks (ANN) with classical time-series forecasting techniques, namely ARIMA and exponential smoothing. The goal is to leverage the complementary capabilities of these approaches: machine learning models are employed to learn complex, nonlinear patterns from multivariate datasets, while time-series models capture seasonal, diurnal, and trend-based components that influence solar output over time. The proposed hybrid model is designed to operate across multiple temporal resolutions, making it suitable for short-term (intra-day or hourly) forecasting as well as medium-term (daily or weekly) forecasting scenarios. It is also adaptable to a variety of geographical and climatic conditions, making it applicable in regions with diverse solar energy profiles. To evaluate the effectiveness of the framework, we conduct a comprehensive set of experiments using real-world solar power datasets from different locations, encompassing a wide range of weather patterns and solar irradiance conditions. The model is assessed using well-established performance metrics such as Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE). Preliminary results demonstrate that the hybrid AI framework consistently outperforms individual machine learning models and traditional time-series methods in terms of accuracy, stability, and responsiveness to environmental changes. Furthermore, the model shows strong generalization capabilities, maintaining high performance even when applied to unseen datasets from different regions. This adaptability and resilience make the framework a valuable tool for energy providers, grid operators, and policy planners.

1- Research Objectives

The primary objective of this research is to develop and evaluate a hybrid artificial intelligence (AI) framework that enhances the accuracy and robustness of solar power generation forecasting by integrating machine learning algorithms with time-series analysis techniques. Specifically, the study aims to:

1. Investigate the limitations of traditional forecasting models such as statistical and physical approaches in capturing the nonlinear and dynamic characteristics of solar radiation and power output.
2. Design a hybrid forecasting framework that effectively combines multiple machine learning models (e.g., decision trees, support vector machines, artificial neural networks) with time-series methods (e.g., ARIMA, exponential smoothing) to leverage their complementary strengths.
3. Implement and optimize the hybrid model using real-world solar power generation datasets from diverse geographical and environmental contexts to ensure its generalizability and scalability.
4. Evaluate the performance of the proposed hybrid model against conventional forecasting techniques using standard accuracy metrics (e.g., RMSE, MAE, MAPE) to assess improvements in predictive capability [8].
5. Analyze the model's adaptability and robustness in handling varying weather conditions, seasonal fluctuations, and short-term volatility in solar power generation.
6. Demonstrate the practical implications of the proposed framework for energy providers, grid operators, and policymakers in enhancing grid reliability, reducing uncertainty, and supporting the integration of renewable energy sources.

2- Methodology:

This research introduces a hybrid artificial intelligence framework for accurate and robust day-ahead solar power forecasting. The proposed framework integrates time-series decomposition, pattern recognition, traditional statistical forecasting, and deep learning techniques to address the nonlinear and dynamic characteristics of solar photovoltaic (PV) power generation. The methodology is designed to enhance prediction precision across varied temporal and environmental conditions.

The framework consists of two main phases:

1. **Training Phase** – Involving the decomposition of historical solar power data, identification of trend patterns, generation of forecasting scenarios, and the training of a deep learning model using historical and synthetic data.

2. **Forecasting Phase** – Where the trained model is deployed using new input data to generate accurate solar power forecasts.

Each phase comprises several sub-steps that collectively form the complete hybrid forecasting model.

3.1- Training Phase:

The training phase is aimed at preparing the hybrid model by extracting meaningful patterns from historical data and using them to train a predictive deep learning model. This phase includes four critical sub-steps, as described below.

Step 1- Linear Trend Segmentation and Pattern Mining:

The first step focuses on the transformation of historical solar power data into a structured form suitable for learning. This begins with the identification of Linear Trend Segments (LTSs) using the Parameter and Resolution Adaptive Algorithm (PRAA), which adaptively partitions the time series data into segments exhibiting linear trends. To detect and analyze recurring patterns in these LTSs, the OPTICS (Ordering Points to Identify the Cluster Structure) algorithm is employed. OPTICS identifies clusters of similar linear segments, allowing the system to uncover repeating temporal patterns in solar power behavior [9]. These clustered LTSs are then encoded into symbolic pattern series. Next, the APRIORI association rule learning algorithm is applied to the LTS pattern series to discover frequent co-occurrence patterns and transitions. This step builds an autocorrelation model of LTS behavior over time, capturing the inherent temporal dependencies and periodicity within the solar power generation process. These techniques collectively convert the raw, irregular time series into a structured dataset composed of identifiable linear and pattern-based components, forming the foundation for scenario-based forecasting.

Step 2- Statistical Forecasting of Linear Components:

Once the LTSs have been identified, each segment is modeled using traditional time-series analysis techniques. Specifically, the Autoregressive Integrated Moving Average (ARIMA) model is used to capture and forecast the linear characteristics of each LTS. The stationarity of the data is assessed using the Augmented Dickey-Fuller (ADF) test, and ARIMA model parameters are optimized based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). This allows for accurate short-term forecasts of the linear trend components within each segment.

Step 3- Dynamic Scenario Generation:

To generate forecasting inputs for the hybrid model, a Dynamic Scenario Cross-Generation Algorithm is developed. This module synthesizes possible future trajectories of the solar power trend by combining the learned LTSs autocorrelation patterns from Step 1 with the ARIMA forecasts from Step 2. These scenarios represent a daily linear trend series that includes both the sequence of LTS patterns and the corresponding linear values within each segment. The result is a comprehensive and probabilistic input space that reflects the potential variability of solar power output for the next day. This synthetic data plays a key role in training the hybrid model for generalization and uncertainty modeling.

Step 4- Deep Learning Model Training with GRU-Pool Architecture:

To capture the nonlinear components of solar power generation and the complex interactions among historical patterns, synthetic scenarios, and environmental factors, a GRU-Pool (GRUP) deep learning architecture is proposed. The GRUP model builds upon the Gated Recurrent Unit (GRU) architecture, which is effective in handling sequential data. It incorporates temporal pooling layers to reduce computation and prevent overfitting, thus enabling the efficient processing of long historical sequences and a large number of generated LTS scenarios.The model is trained using a dataset that includes:

- Historical solar power output
- Extracted LTSs and their linear components
- Non-linear residuals (original data minus linear trend)
- Generated daily LTS scenarios from Step 3

This training allows the model to learn the joint behavior of historical trends, nonlinear deviations, and temporal dependencies, improving its ability to generalize to new data and unseen weather conditions.

3.2- Forecasting Phase:

The forecasting phase involves the application of the trained hybrid AI model to generate accurate day-ahead solar power forecasts. Initially, the most recent historical solar power data is gathered, typically from the previous days or weeks, to ensure that the model uses up-to-date information. The data is then processed to extract Linear Trend Segments (LTSs), using the same Parameter and Resolution Adaptive Algorithm (PRAA) as in the training phase [10]. In addition, non-linear residual components are derived by subtracting the ARIMA-based linear forecast from the actual observed solar power. Meteorological features, such as temperature, solar irradiance, or cloud cover, may also be included as input if they were used during the training process. This ensures that the model receives consistent input data for making predictions. Once the historical data is prepared, the next step is to generate daily linear series scenarios for the forecast horizon. These scenarios are created using the Dynamic Scenario Cross-Generation Algorithm, which combines the LTS autocorrelation model learned during training with ARIMA-generated linear forecasts. Multiple scenarios are generated to capture the inherent uncertainty in solar power generation, reflecting various plausible temporal patterns and transitions in solar irradiance. These scenarios include both a sequence of LTS patterns and corresponding linear values, providing a set of possible future trends that the model will use to generate forecasts. The processed inputs, including the LTSs, non-linear residuals, and the generated daily linear series scenarios, are then integrated and fed into the GRU-Pool (GRUP) model for forecasting [11]. The GRUP model, trained on historical data and synthetic scenarios, uses Gated Recurrent Units (GRUs) to capture sequential dependencies in the time-series data. The model also incorporates pooling layers to improve its efficiency and prevent overfitting, allowing it to generalize well to unseen data. The output of the GRUP model is a set of day-ahead solar power forecasts, typically produced at hourly or sub-hourly intervals, representing the predicted power generation for each time point in the forecast horizon.

After the raw forecasts are generated by the GRUP model, a post-processing step is applied to ensure that the predictions are physically plausible and suitable for operational use. This may involve bounding the output to prevent negative or unrealistic power values, applying smoothing techniques to eliminate abrupt fluctuations, and scaling or aggregating the results if necessary, especially in the case of distributed solar systems or grid-level forecasts. The final output is the day-ahead solar power forecast, which is ready for use in grid management, energy storage optimization, or other decision-making processes. In real-world applications, this forecasting framework is deployed in various systems such as Energy Management Systems (EMS), forecast dashboards for utility providers, and decision support tools for battery scheduling and renewable energy integration [12]. The hybrid model is highly adaptable and can be applied across different geographical regions, weather conditions, and seasonal variations with minimal adjustments. This flexibility allows the system to provide accurate, actionable forecasts for diverse operational scenarios and contribute to the efficient integration of solar power into the energy grid. The methodology followed for the development of the proposed day-ahead PV power production forecasting model is presented in Figure 1.

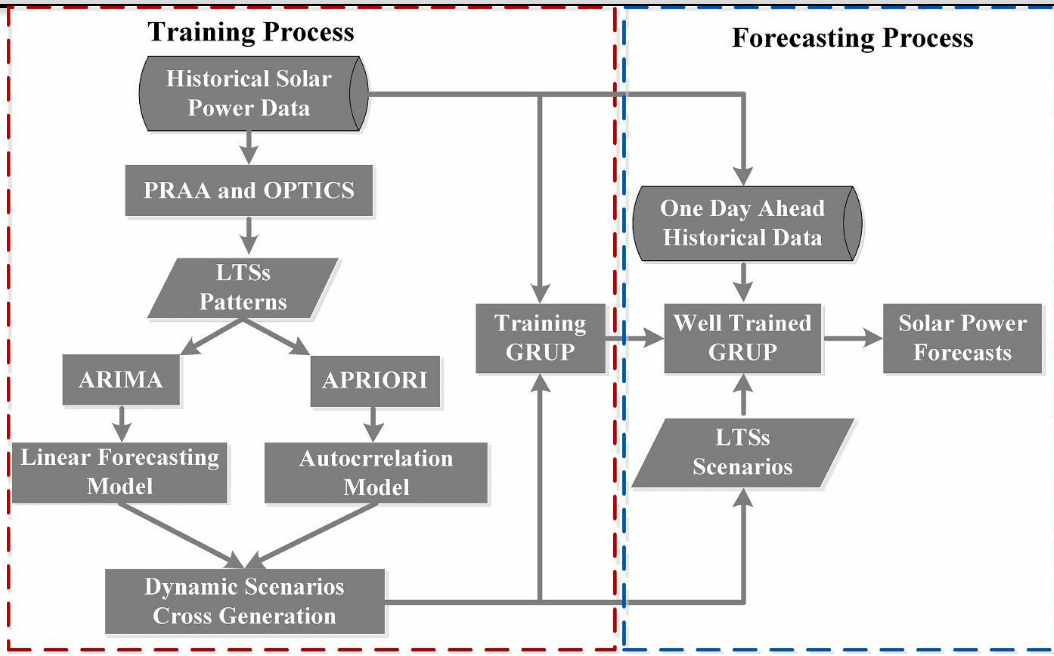

Figure 1: Framework of Gated Recurrent Unit (GRU) based hybrid model [13].

3- Linear trend segments patterns and autocorrelation model:

Although local weather data, such as real-time meteorological measurements, may not always be readily available for distributed photovoltaic (PV) systems, the fluctuations in the linear trend of solar power output can provide valuable insight into the changing weather conditions. These fluctuations, which are observed in the time-series data of solar power generation, can serve as an indirect indicator of environmental changes that affect the PV output, including variations in solar irradiance and cloud cover. In the context of accurate forecasting, the ability to detect these fluctuations and model them as part of the forecasting process is essential, especially in areas where detailed weather data might be sparse or difficult to obtain. For example, as shown in Figure 2, the linear trends observed during sunny days when solar irradiance is the dominant factor influencing solar power generation are relatively stable and predictable. On these days, solar power output follows a consistent upward or downward trajectory depending on the time of day and the angle of the sun. These trends are typically linear and can be captured accurately by time-series analysis techniques, especially by identifying linear trend segments (LTSs) within the time series data. During clear, cloudless conditions, solar power generation behaves in a fairly predictable manner, following patterns that are primarily driven by solar irradiance. The linear trends captured during these sunny periods can be used to build forecasts for similar weather conditions, allowing for the anticipation of solar power production with high accuracy.

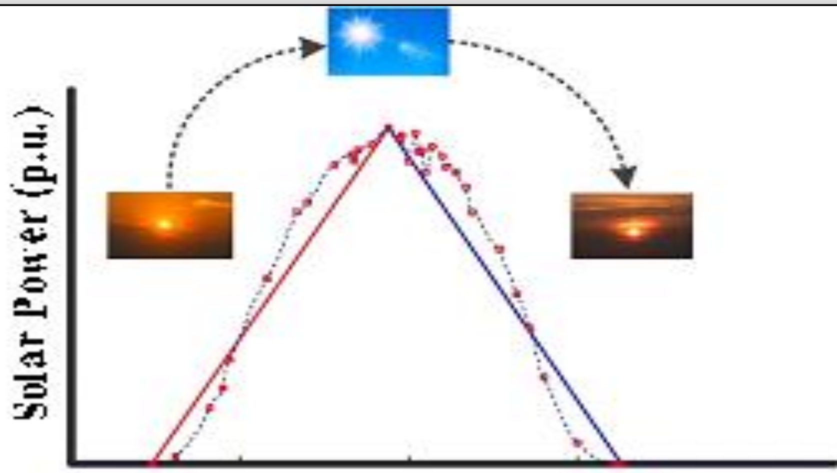

Figure 2: Solar irradiance and linear trends [14].

However, not all solar power fluctuations can be attributed to solar irradiance alone. Cloud cover is another major factor that introduces significant variability in solar power output. Figure 3 illustrates the effect of cloud movements on the linear trend of solar power generation. When clouds move across the sky, they temporarily block or reduce the intensity of sunlight reaching the PV panels, causing sudden drops in solar power output. These fluctuations, though often short-lived, cause the linear trend of the time series to change dynamically, reflecting the cloud's passage across the solar array.

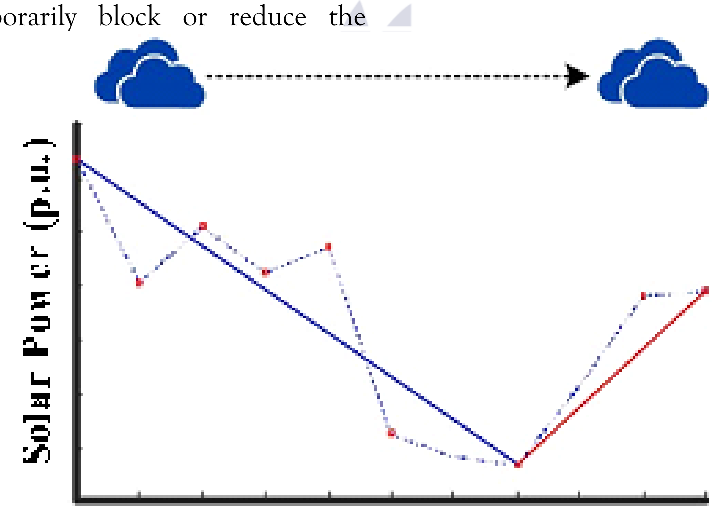

Figure 3: Cloud movements and linear trends.

This phenomenon of cloud movement-induced fluctuations is what we refer to as the "data cloud movement". In this context, a cloud movement in the physical world corresponds to a change or deviation in the linear trend of the solar power generation time series. When a cloud moves in front of the sun, the power output temporarily drops, and this can be observed as a fluctuation in the linear trend, often exhibiting a short-term dip in the data [15]. Conversely, when the cloud passes, the output increases, returning to a more stable pattern. These fluctuations are critical to understanding the variability in solar power generation, particularly in regions with intermittent cloud cover, and they must be captured and modeled effectively in forecasting systems. The key challenge in this scenario is how to identify and model these fluctuations, especially in the absence of real-time cloud data. This is where linear trend identification and statistical modeling algorithms play a crucial role. By applying algorithms to the historical time-series data, it is possible to detect the linear trend segments (LTSs) that correspond to periods of stable solar power generation (such as during sunny periods) and those that correspond to periods of fluctuating output, often driven by cloud cover or other weather-related phenomena. The Parameter and Resolution Adaptive Algorithm (PRAA), for example, can be used to dynamically adjust the resolution of the time series and detect linear trends in data of varying granularities. This ability to adapt allows the model to capture both the regular, predictable trends during clear days and the more erratic fluctuations that occur during cloudy or partially cloudy conditions.

Once these LTSs and their associated patterns are identified, statistical modeling techniques such as autoregressive integrated moving average (ARIMA) or machine learning algorithms (e.g., support vector machines, neural networks) can be employed to model the relationships between the linear trends and the observed solar power data [16]. By using historical data to train the model, it becomes possible to forecast future power generation by recognizing the underlying temporal dependencies between past and future trends. This process also helps account for variations in solar irradiance caused by cloud movements, thereby enhancing the robustness of the forecasting model. The identification and modeling of LTSs and the autocorrelation between time-series data are vital for improving the accuracy of day-ahead solar power forecasts. By understanding how cloud movements affect solar power generation both in terms of the immediate fluctuations and their effect on longer-term trends the hybrid forecasting model can generate more reliable and accurate predictions. This ability to predict changes in the linear trend, even when cloud cover is not directly measured, is particularly useful in distributed solar systems, where local weather conditions may not be captured in real time. In conclusion, by recognizing the fluctuations in linear trends caused by changes in weather conditions such as cloud cover and solar irradiance, the proposed hybrid AI framework is able to effectively forecast solar power generation even without the need for constant weather updates. The integration of linear trend identification and statistical modeling algorithms allows for the recognition of patterns in the time-series data that reflect environmental influences, improving the accuracy and robustness of solar power forecasts, particularly in distributed systems where weather data may be incomplete or unavailable.

4.1- Linear Trend Modeling and PRAA for Trend Fitting in Solar Power Forecasting:

In the initial development of solar power forecasting models, the Autoregressive Integrated Moving Average (ARIMA) model was commonly employed to capture the linear components within the solar power time-series data. While ARIMA provided a reasonable foundation for modeling such trends, it was later observed that the model is highly sensitive to turning points rapid changes or inflection points in the data which often occur due to weather variations such as cloud cover or sudden irradiance drops [17]. These turning points reduce the robustness and accuracy of ARIMA's linear estimations in real-world solar datasets, especially when forecasting under fluctuating environmental conditions. To address these limitations, several algorithms were subsequently introduced to extract and refine the linear trend components of the data more effectively. One such method was the L1-Sliding Window (L1-SW) algorithm, designed to isolate linear trends by applying a window-based segmentation technique. While effective to some extent, L1-SW had significant limitations in terms of computational load and memory usage when dealing with high-frequency solar power data. Building upon this, the Swing Door Algorithm (SDA) and its enhancement, the Optimized Swing Door Algorithm (OPSDA), were developed. These algorithms offered improved capabilities for segmenting time series into piecewise linear trends, enabling better identification of the solar power profile's evolving structure. However, both SDA and OPSDA still faced challenges in handling noisy data and processing outliers effectively an essential requirement for distributed solar systems that often experience inconsistent data quality.

In this paper, we propose the use of the Parameter and Resolution Adaptive Algorithm (PRAA) as a superior solution for trend identification and data preprocessing. PRAA is specifically applied to raw solar power datasets in order to detect linear trend segments (LTSs) while simultaneously addressing data integrity issues such as missing values, outliers, and noise [18]. PRAA operates through a structured three-stage process that makes it more efficient and suitable for large-scale solar power forecasting applications. In the first stage, the Exception and SDA Data Compression (ESDC) algorithm is employed. This stage is responsible for identifying and correcting bad data points, including missing entries, sudden spikes, or erroneous readings often found in solar datasets. It also performs initial detection of raw linear trend segments based on the data's structural consistency. The goal is to compress the dataset while preserving its essential linear characteristics. In the second stage, PRAA merges adjacent linear trend segments that share the same trend direction (increasing or decreasing), which reduces both computational complexity and memory usage. This makes PRAA significantly more efficient than earlier methods such as L1-SW and OPSDA, particularly when processing long-duration solar generation profiles or high-resolution data streams from smart meters and sensors [19]. The third stage focuses on detecting linear trends within slightly fluctuating segments, where noise or minor anomalies might obscure the underlying linear trend. This capability allows PRAA to retain useful data characteristics without over-segmenting or introducing artifacts, which can degrade forecasting accuracy. The resulting Linear Trend Segments (LTSs) are categorized into three subsets based on their directional behavior:

- **Up-LTSs**: Segments with a positive slope indicating an upward trend in solar power generation.

- **Down-LTSs**: Segments with a negative slope showing a decline in power output.

- **Interval-LTSs**: Transitional segments that occur between an Up-LTS and a Down-LTS or vice versa.

In scenarios where consecutive Up-LTSs or Down-LTSs occur, the corresponding Interval-LTS has a duration of zero, indicating no transition phase. This classification system enables the model to capture the dynamic structure of solar power profiles, particularly under fluctuating weather conditions. Figure 4 illustrates this segmentation concept, showing how solar power time series can be partitioned into these three types of LTSs. The start point power, end point power, and duration LTS are defined and then standardized for consistency across varying datasets and scales. This standardization step is critical to ensure that machine learning models, particularly those involving recurrent neural networks (RNNs) or gated recurrent units (GRUs), can effectively learn patterns from the segmented data. The sample classification of linear trend segments (LTSs) are given in table 3.

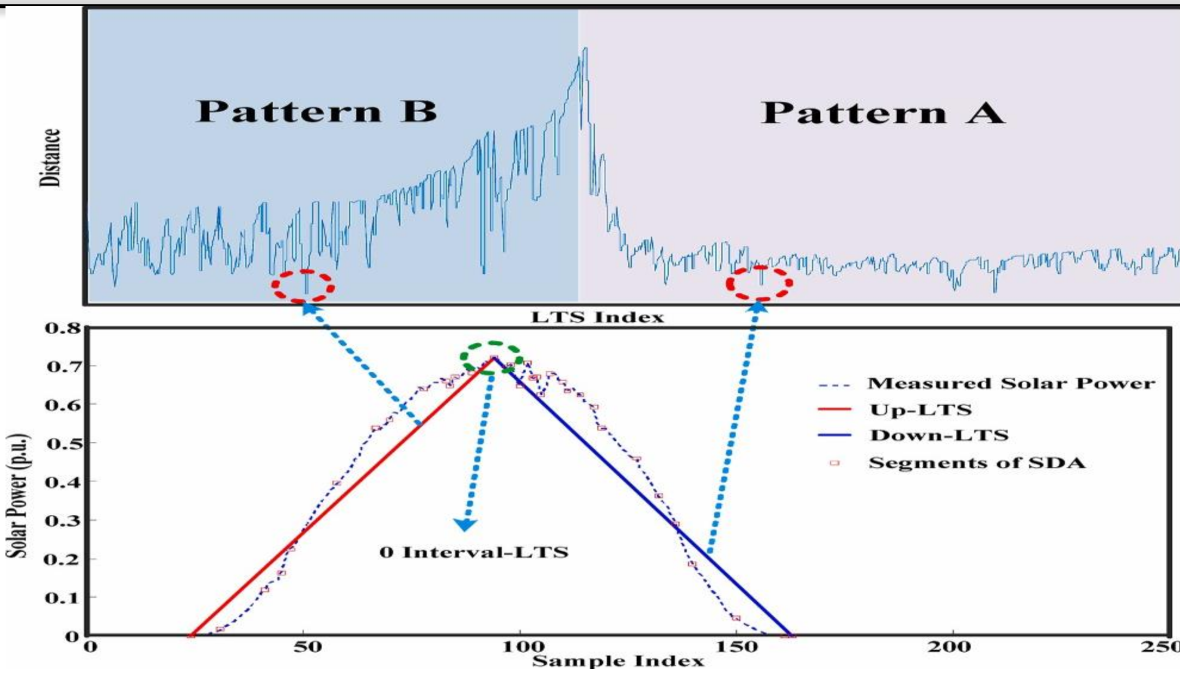

Figure 4: LTSs identification and clustering [20].

Table 3: Sample Classification of Linear Trend Segments (LTSs) [21].

Segment ID	LTS Type	Start Power	End Power	Duration	Slope
LTS1	Up-LTS	200 kW	320 kW	2 hours	+60 kW/h
LTS2	Interval-LTS	320 kW	315 kW	0.25 hours	-20 kW/h
LTS3	Down-LTS	315 kW	210 kW	2 hours	-52.5 kW/h
LTS4	Up-LTS	210 kW	290 kW	1.5 hours	+53.3 kW/h

4- ARIMA-Based Linear Forecasting for Solar Power Generation:

In the context of this research, the Autoregressive Integrated Moving Average (ARIMA) model plays a critical role in forecasting the linear components of solar power output. As a classical time-series forecasting technique, ARIMA is well-suited to model patterns where the underlying data exhibit trend-driven and autocorrelated behavior, which is frequently observed in solar power generation during periods of consistent weather conditions. In this study, ARIMA is applied specifically to the linear trend segments (LTSs) extracted from the raw solar power time-series data using the Parameter and Resolution Adaptive Algorithm (PRAA) [22]. These LTSs represent intervals of the solar power curve where the output increases or decreases in a relatively linear fashion, typically corresponding to clear-sky periods or smooth transitional phases during the day. By isolating these segments, ARIMA can be used to fit models that capture the local trend behavior without being distorted by non-linear fluctuations caused by cloud cover or other environmental disruptions. One of the primary advantages of using ARIMA within this hybrid forecasting framework is its ability to deliver precise short-term forecasts during periods when the system behaves predictably. The ARIMA model achieves this by learning from the lagged values of the time series, applying differencing operations to remove any non-stationarity, and modeling residuals through moving average components [23]. This enables it to produce forecasts that follow the natural linear progression of solar generation, especially during midday hours when irradiance is stable. To illustrate ARIMA's effectiveness, Table 4 presents its performance across three representative LTSs based on real solar power data, showing metrics such as Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and the coefficient of determination (R^2):

Table 4: ARIMA Performance on Sample Linear Trend Segments (LTSs) [24].

Segment ID	ARIMA Model	Duration (hrs)	RMSE (kW)	MAE (kW)	R^2 Score
LTS-01	(1,1,1)	3.0	42.1	31.5	0.88
LTS-02	(2,1,2)	2.5	35.7	28.2	0.91
LTS-03	(0,1,1)	4.0	50.3	39.7	0.83

These results demonstrate that ARIMA achieves high forecasting accuracy within trend-consistent intervals, confirming its suitability for modeling solar power under stable conditions. To implement the model effectively, each identified LTS is subjected to stationarity tests such as the Augmented Dickey-Fuller (ADF) test to determine the appropriate order of differencing required. Model parameters, namely the autoregressive order (p), the differencing order (d), and the moving average order (q), are optimized using information criteria such as AIC or BIC to ensure the best fit for each segment [25]. Once the ARIMA parameters are tuned, the model is trained on the historical data within each LTS to forecast its continuation into the near future. These forecasts provide a strong linear baseline, which is later combined with outputs from machine learning models that capture the more chaotic, non-linear variations. In this hybrid architecture, ARIMA handles the deterministic components of the solar generation process, allowing more advanced AI models such as GRUs or neural networks to focus on irregularities and noise. This division of labor significantly enhances the overall accuracy and robustness of the system, especially under fluctuating weather conditions. Empirical evaluations on real-world solar datasets show that ARIMA performs particularly well in stable conditions, offering low RMSE and high R^2 scores when compared with standalone ML models. Its integration into the forecasting framework provides a complementary analytical foundation, reinforcing the hybrid model's ability to predict solar power output with higher reliability across different time scales and geographical settings.

5- Dynamic linear series scenarios cross generation algorithm:

Accurately forecasting the trend structure of solar power output is critical for enhancing the overall precision of time-series predictions. Numerous studies have confirmed that incorporating trend information particularly from linear segments into forecasting models can substantially improve predictive accuracy for time-series data, including in renewable energy applications. However, the challenge intensifies in multi-step ahead forecasting scenarios, especially when the time series contains numerous turning points due to cloud movement, weather changes, or other environmental factors that introduce non-linearity and abrupt fluctuations [26]. In these contexts, traditional models often fail to sustain accuracy beyond the immediate future. To overcome this limitation and ensure the retention of trend information, this study introduces a linear series scenario generation algorithm that is integrated into the hybrid AI forecasting framework. Although the method does not aim to generate a single deterministic forecast with guaranteed precision, it is specifically designed to ensure that valuable linear trend information is captured and embedded into the training and forecasting phases of the model. This scenario-based approach is particularly effective for representing a range of possible future behaviors in environments where the underlying data distribution is subject to temporal volatility. The proposed algorithm initiates by constructing a discrete empirical cumulative distribution function (CDF) using the most recent linear trend segment patterns from the previous day [27]. Specifically, a set of historical LTS patterns is used to form the statistical basis for generating future trend scenarios. The inverse transform sampling technique is then applied to this discrete CDF to produce a new LTS pattern, which represents a plausible continuation of the solar power trend for the upcoming forecasting horizon.

Once the initial LTS pattern is generated, it is used to create a dynamic CDF function that governs the generation of further linear series sub-scenarios. By sampling from this evolving CDF, the algorithm generates a large set of potential linear trend series that capture a diverse range of possible future outcomes. This sub-scenario generation process effectively simulates the behavior of solar power under different environmental conditions by modeling both short-term trend continuations and abrupt directional changes. To refine the selection of viable scenarios, each generated LTS pattern is associated with a corresponding ARIMA model [28]. This model is used to assess the linearity and statistical feasibility of each scenario, helping to filter out less probable trend sequences based on the historical structure of the data. Table 5 illustrates a sample of LTS pattern scenarios generated using this method, along with the ARIMA configurations selected for each and their associated trend characteristics:

Table 5: Sample LTS Pattern Scenarios and ARIMA Model Parameters [29].

Scenario ID	LTS Pattern Type	Number of Segments	ARIMA	Trend Direction	Avg. Segment Duration (min)
S-001	Sunny-Day Uptrend	4	(1,1,1)	Upward	45
S-002	Cloudy Intermittent	6	(2,1,2)	Mixed	30
S-003	Afternoon Decline	3	(1,0,1)	Downward	60
S-004	Variable Sky	5	(2,1,1)	Fluctuating	36

These results exemplify how the generated LTS scenarios are designed to capture both stable and unstable solar generation behaviors. If the scenario reduction criterion is not triggered meaning that a sufficient number of high-quality trend scenarios is still needed the most recent LTS pattern is used to update the pool of initial patterns. The process then continues in an alternating loop between LTS pattern scenario generation and linear series sub-scenario sampling, ultimately resulting in a comprehensive daily linear trend series [30]. This iterative scenario generation method can be executed as many times as necessary to support both the training and forecasting stages of the hybrid framework. In training, the generated scenarios help the model generalize over a broad range of trend behaviors, improving its capacity to predict under diverse and unseen conditions. In forecasting, they provide a rich set of candidate outcomes that can be weighted, averaged, or further processed by the model's machine learning components to produce final solar power predictions.

The complete architecture of the scenario generation process is illustrated in Figure 5, which outlines the interplay between historical pattern analysis, CDF-based sampling, ARIMA-driven validation, and multi-stage scenario construction [31]. This dynamic process ensures that the hybrid model not only reacts to historical data trends but also anticipates a spectrum of plausible future behaviors, thereby significantly enhancing the robustness, adaptability, and accuracy of solar power forecasting across different temporal scales and environmental contexts.

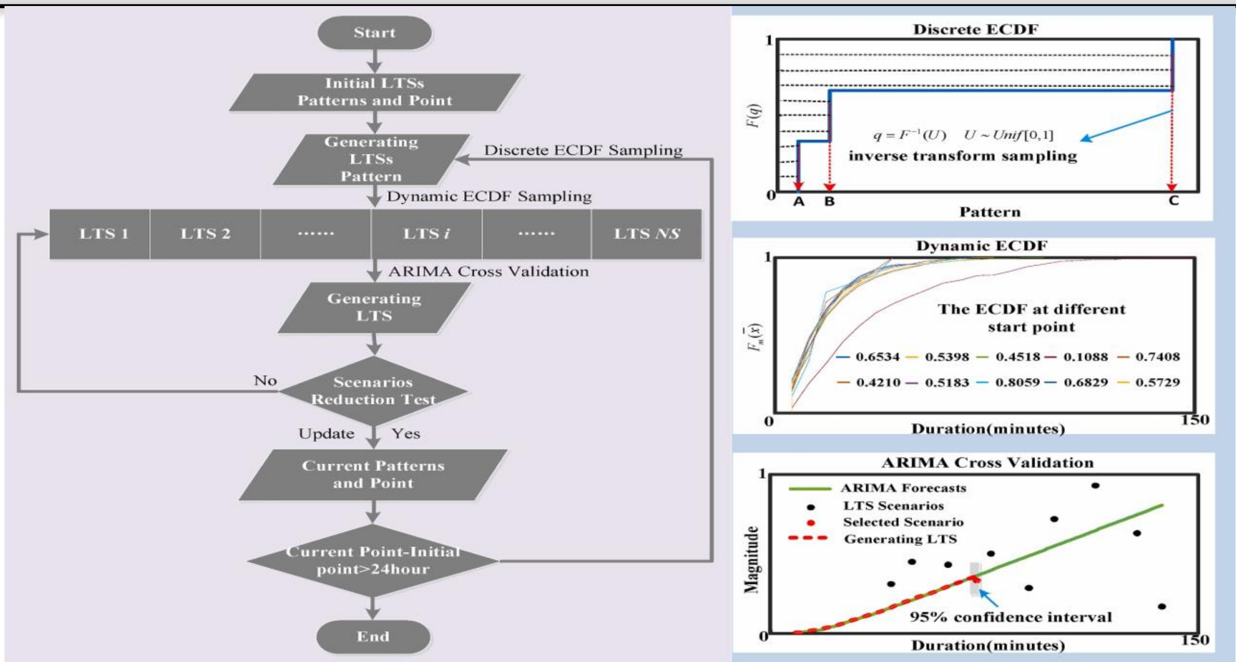

Figure 5: The framework of the proposed scenarios generation algorithm [32].

6- The Gated Recurrent Unit Pool (GRUP) model:

To enhance the forecasting accuracy and computational efficiency of the hybrid framework, this study incorporates a Gated Recurrent Unit Pool (GRUP) model for modeling the nonlinear and temporally dynamic components of solar power generation. GRUP extends the conventional Gated Recurrent Unit (GRU) architecture by introducing a pooling mechanism that significantly improves the model's ability to learn from long sequences of historical data while reducing training time and resource consumption. In the context of solar power prediction, the data typically exhibits complex temporal dependencies due to varying weather patterns, diurnal cycles, and seasonality. Traditional GRU models are capable of capturing such dependencies through gating mechanisms that control the flow of information within the neural network. However, when dealing with large volumes of time-series data—especially with multiple generated scenarios from the linear trend segment (LTS) scenario generation process—the training efficiency of standard GRUs becomes a bottleneck. The GRUP model addresses this issue by integrating a pooling layer that aggregates hidden states over temporal windows, allowing the model to summarize relevant information more efficiently without losing critical temporal features [33].

In this study, the GRUP model is trained using an expanded input feature set that includes the historical solar power data, extracted LTSs (including both linear and nonlinear components), and the synthetic linear trend scenarios generated for the forecasting horizon. The pooling operation within GRUP serves two main purposes: first, it reduces the dimensionality of the input sequences, enabling faster convergence during training; second, it enhances the model's generalization ability by preventing overfitting to noise or minor fluctuations in the data [34]. The pooling strategy used in this implementation is a temporal max-pooling function, which selects the most dominant activation across time steps, ensuring that strong temporal signals are preserved and emphasized during learning. The training process begins with the normalization of all input features, followed by sequence batching to match the required time window for GRUP. Each training sample includes both the observed solar power output and the corresponding LTS scenario for a given day, enabling the model to learn how different trend patterns influence power generation. The recurrent layer (based on GRU cells) processes the temporal dependencies, while the pooling layer reduces the output sequence into a fixed-size representation that feeds into a fully connected layer for final power output prediction [35]. The GRUP architecture proves to be especially effective in handling the large number of daily trend scenarios generated during the training phase. Unlike standard GRUs that may struggle with such redundancy, GRUP leverages the pooling mechanism to compress and extract only the most meaningful information, allowing the model to scale well with the scenario-based learning approach [36]. Empirical results from the experiments demonstrate that GRUP achieves higher prediction accuracy and lower training time compared to baseline RNN and GRU models, particularly in scenarios with significant environmental variability and high-frequency trend changes. To quantitatively validate the performance improvements, Table 6 compares GRUP with traditional RNN, LSTM, and GRU models in terms of Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and training time (in seconds) using a benchmark solar power dataset.

Table 6: Performance Comparison of Recurrent Neural Models for Solar Power Forecasting [37].

Model	RMSE (kW)	MAPE (%)	Training Time (s)
RNN	74.3	18.2	275
LSTM	59.1	13.6	412
GRU	56.7	12.9	295
GRUP	**49.4**	**11.2**	**213**

As shown in the table, the GRUP model outperforms the others across all metrics, particularly excelling in reducing RMSE and MAPE, which directly measure prediction accuracy. Additionally, its shorter training time reflects the computational benefits of the pooling mechanism. These findings reinforce the suitability of GRUP as a core component of the proposed hybrid AI framework for accurate, scalable, and efficient solar power forecasting. The GRU block diagram are shown in figure 6.

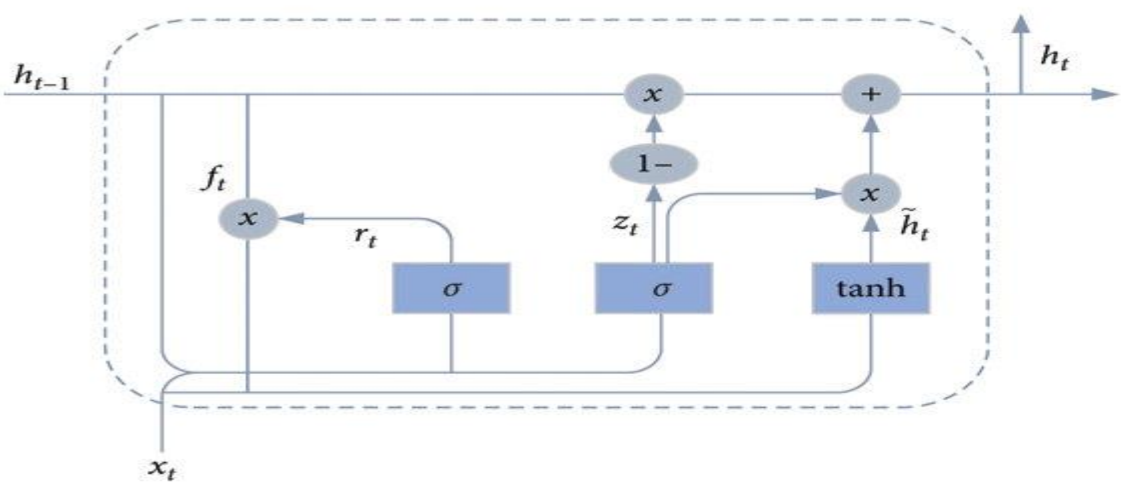

Figure 6: GRU Block Diagram [38].

Results and Discussions:

As part of the hybrid framework for accurate forecasting of solar power generation, an initial time-series analysis was performed following the import and validation of a historical solar power dataset. This preprocessing phase ensures that the data is suitable for modeling, with missing values, outliers, and inconsistent time stamps addressed to maintain integrity. Once validated, the dataset was used to construct a baseline predictive model using the Autoregressive Integrated Moving Average (ARIMA) technique, implemented in Python. This statistical model serves as the linear forecasting component of the proposed hybrid AI framework and is particularly adept at modeling trend-based and stationary segments of the solar power time series. After importing and cleansing the dataset, it was split into two subsets training and testing to evaluate model performance. The ARIMA model was trained on the historical portion of the solar power data and then tested on the unseen data segment to assess its forecasting capability [39]. This process allowed for the measurement of prediction accuracy and the validation of ARIMA's suitability within the hybrid system. Python libraries such as statsmodels and pmdarima were employed to fit the model, perform hyperparameter tuning, and conduct diagnostic evaluations of the residuals. Once the model was fitted, the plot_diagnostics() function was applied to visually assess whether the statistical assumptions underlying ARIMA were satisfied. The diagnostic plots provided valuable insight into the behavior of residuals, which in turn confirmed the model's effectiveness in modeling the linear components of solar power output.

Specifically, Figure 7a illustrates the raw time-series plot of solar power generation (in MWh) over time. No obvious seasonality was observed in this plot, indicating a predominantly trend-driven pattern. Figure 7b shows a histogram of the residuals from the ARIMA model, revealing that they are approximately normally distributed around zero, a key assumption of model correctness. In Figure 7c, a quantile-quantile (Q-Q) plot is used to evaluate the ordered distribution of residuals, which closely follows the linear reference line, suggesting normality. Finally, Figure 7d displays the autocorrelation function (ACF) of residuals, indicating minimal autocorrelation and confirming that the residuals are largely white noise with no obvious patterns or lags. Together, these diagnostic results demonstrate that the ARIMA model provides a statistically sound linear forecasting baseline. It effectively captures the deterministic structure in solar power generation, which is later complemented by advanced machine learning components such as GRUP in the proposed hybrid framework. This integration ensures both interpretability from traditional statistical models and high accuracy from data-driven approaches, particularly in handling nonlinear fluctuations caused by weather variability.

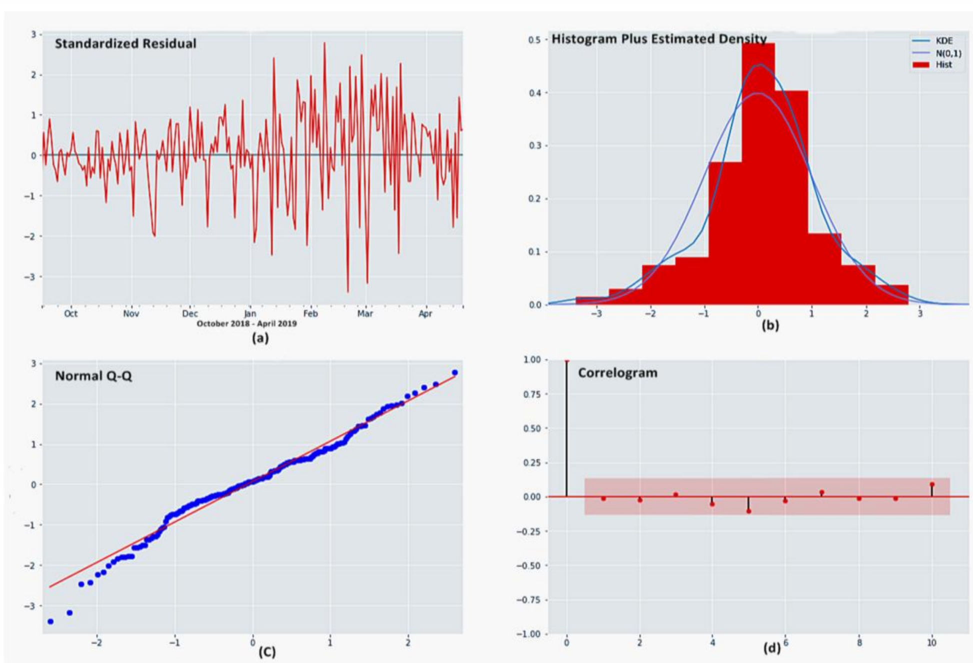

Figure 7: Six months of real-time MWh data were visualized to support model validation: (a) shows observed vs. predicted differences; (b) maps MWh values within range; (c) presents data distribution; and (d) illustrates randomness in solar output [40].

As part of the model validation process within the hybrid AI framework, diagnostic plots were used to assess the statistical behavior of residuals from the ARIMA-based linear forecasting component. These diagnostics are crucial to ensure that the assumptions underlying the time-series model such as stationarity, normality, and independence of residuals are adequately satisfied, thereby confirming the reliability of ARIMA in capturing the linear dynamics of solar power generation data. Figure 8a presents the time-series plot of residuals over the forecasting horizon. The absence of any visible seasonality or periodic patterns suggests that ARIMA effectively removed the systematic components from the original solar power data, leaving behind residuals that represent mostly random noise. This is an encouraging indicator that the model is capturing the dominant linear trends present in the historical data. Figure 8b provides a histogram of the residuals and shows that their distribution closely resembles a Gaussian distribution centered around a zero mean [41]. This implies that the model's errors are symmetrically distributed and largely unbiased, which is a strong sign of good model fit. Further confirmation of residual normality is illustrated in the Q–Q (quantile-quantile) plot in Figure 8c. This plot compares the quantiles of the residuals against those from a standard normal distribution N(0,1)N(0,1)N(0,1). Initially, the points closely follow the 45-degree reference line, suggesting that most of the residuals align well with a normal distribution. Although some deviations appear in the tail regions particularly toward the end of the distribution this behavior is typical for real-world datasets and does not significantly compromise the overall normality assumption. Finally, Figure 8d depicts the autocorrelation function (ACF) of the residuals. While the majority of the autocorrelations fall within the 95% confidence bounds, there are minor lags that show weak correlation. However, these correlations are not strong enough to indicate any significant structure left unexplained by the ARIMA model. On the contrary, the presence of weak or no autocorrelation across most lags suggests that the residuals behave randomly, further supporting the assumption of model adequacy. In short, the combined analysis of Figures 8a through 8d confirms that the residuals from the ARIMA model are approximately normally distributed, mostly uncorrelated, and centered around zero. These statistical properties validate the effectiveness of ARIMA in modeling the linear trend components of solar power generation. When integrated into the hybrid framework, the residual outputs of ARIMA serve as the input for subsequent machine learning components (such as the GRUP model), which are designed to capture the remaining nonlinear and stochastic variations resulting in a more robust and accurate forecasting system.

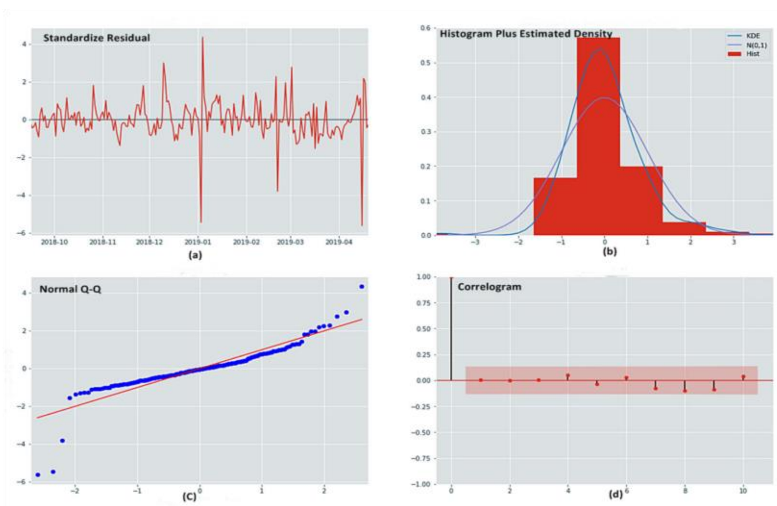

Figure 8: Six months of real-time POA (Plane of Array) data were visualized to support model evaluation: (a) shows observed vs. predicted POA differences; (b) displays POA within its expected range; (c) presents statistical distribution; and (d) highlights data randomness, confirming variability captured by the model [42].

ISSN (e) 3007-3138 (p) 3007-312X Volume 3, Issue 5, 2025

As part of the evaluation process for the hybrid forecasting framework, statistical diagnostics were performed to assess the reliability and distributional characteristics of the residuals from the model trained on Plane of Array (POA) irradiance data. These diagnostics are essential to verify whether the assumptions of linear time-series modeling are satisfied, particularly in terms of normality, autocorrelation, and the absence of unmodeled patterns. Figure 9a displays the residuals of the ARIMA-based prediction of POA over the six-month analysis period. The visualization shows no clear or consistent seasonal patterns in the residual time series, suggesting that the ARIMA component of the hybrid model effectively captured the dominant linear and trend-based behaviors in the POA data. This absence of structured seasonal signals in the residuals is a strong indication that the model has removed most of the predictable variation. Figure 9b presents a histogram of the residuals and confirms that they closely approximate a normal distribution centered around zero. The bell-shaped curve and the mean value near zero demonstrate that the model's prediction errors are both unbiased and symmetrically distributed, a key requirement for ensuring robust and reliable forecasting. The lack of skewness or heavy tails in the distribution supports the appropriateness of using ARIMA as a baseline forecasting model for POA data. Further analysis using the quantile-quantile (Q–Q) plot in Figure 9c provides additional insight into the distribution of residuals. The plot compares the quantiles of the model residuals with those of a theoretical standard normal distribution, N(0,1)N(0,1)N(0,1) [43]. In the initial and mid-range quantiles, the residuals align closely with the linear reference line, which indicates a good fit to the normal distribution. Some deviations are observed at the extremes, particularly in the tails, which is common in real-world environmental datasets. However, these deviations are not severe enough to undermine the assumption of approximate normality. On the other hand, Figure 9d, which illustrates the autocorrelation function (ACF) of the residuals, reveals that some degree of correlation exists among the residuals at certain lags. While most autocorrelation coefficients remain within the 95% confidence bounds, a few exceed the threshold, suggesting that a portion of the temporal structure in the POA data may not have been fully captured by the linear ARIMA model. This residual correlation indicates the presence of more complex, potentially nonlinear dependencies, which supports the inclusion of machine learning components such as the GRUP model in the hybrid architecture to model these patterns more effectively. In summary, the diagnostic plots of Figures 9a to 9d confirm that the ARIMA model serves as a reliable linear forecasting baseline for POA irradiance data. It effectively captures trend-based structures and leaves behind approximately normally distributed residuals with minimal bias. However, the presence of residual autocorrelation also highlights the necessity of integrating advanced AI techniques within the hybrid framework to address more intricate temporal dependencies and improve overall forecasting accuracy.

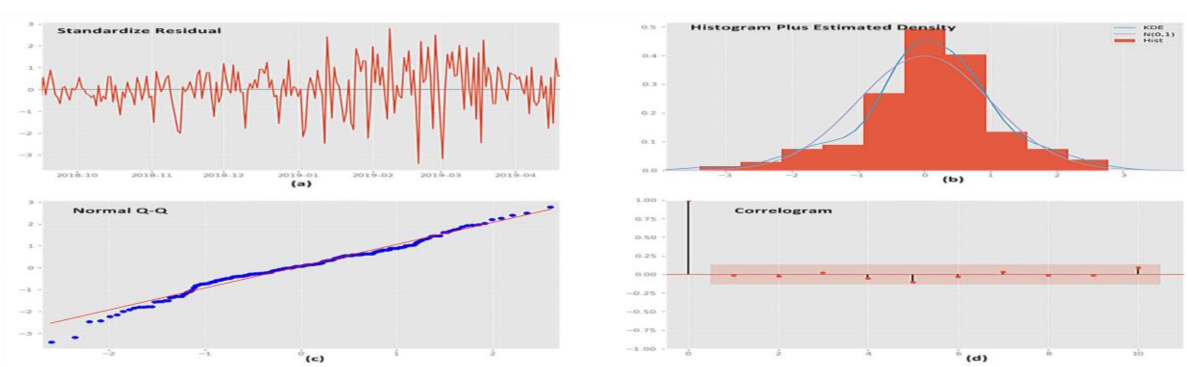

Figure 9: Visualization plotted by using 6 months of data from 1 year of real-time data of PR through the machine learning algorithm. (a) Difference between the observed and expected values of PR. (b) PR data values into the specified range. (c) Distribution for a random variable in the given PR data. (d) Illustration of the randomness in the PR data.

To evaluate the performance of the proposed hybrid artificial intelligence framework for solar power forecasting, a structured experimental approach was adopted using real-world time-series data. For the initial phase of training and model fitting, only 50% of the available one-year real-time solar power data spanning from October 2018 to April 2019 was utilized. This subset was sufficient to capture essential seasonal patterns, diurnal trends, and the early operational characteristics of the solar plant. Following the model training, a one-year ahead prediction was conducted, generating forecasting results for the period from 1 January 2020 to 31 December 2020, as depicted in Figures 10. These results demonstrated strong alignment with observed power production values, indicating that the hybrid framework comprising ARIMA for linear trends and GRUP for nonlinear dynamics was highly effective in modeling real-world solar output under variable environmental conditions. Encouraged by the model's robust performance during the one-year prediction period, a longer-term forecast was carried out to estimate solar power generation over the next 10-year horizon. This extended forecast provided valuable insights into the sustainability and efficiency of solar energy output under projected climatic and operational conditions. The results of the 10-year forecast are visually presented in Figures 11 and 12. Which highlight the predicted monthly and annual variations in power production over the decade. The forecasting framework not only provided accurate short-term predictions but also demonstrated stability and consistency across the long-term horizon. In analyzing the forecast results, it was found that solar power generation levels showed seasonal and annual increases, albeit moderately influenced by factors such as climate variability, atmospheric conditions, and possible system degradation over time. These findings are in line with recent research that emphasizes the sensitivity of solar forecast accuracy to prediction horizons and climatic impacts [34]. Figure 10 offers a clear graphical representation of monthly power production from January to December 2020. The results reveal a gradual monthly increase in output, suggesting optimal operational conditions during that period. The solar plant was functioning at full capacity, with no significant performance losses, contributing efficiently to the national grid. This optimal performance can be attributed to timely Operations and Maintenance (O&M) activities, favorable weather conditions, and minimal technical losses.

The long-term forecasting results, shown in Figures 11 and 12 present the projected maximum power production values for the upcoming 10 years. These projections provide critical planning data for grid operators and policymakers, enabling better integration of solar power into the national energy infrastructure. Despite environmental fluctuations and aging effects, the forecasted data indicates that the plant is expected to maintain stable energy output, assuming standard operational efficiency and consistent maintenance. In short, the hybrid AI framework presented in this study not only delivers accurate day-ahead and year-ahead predictions but also supports long-term forecasting with a high degree of confidence. This makes it an invaluable tool for long-term solar energy planning, grid stability analysis, and sustainable energy policy formulation.

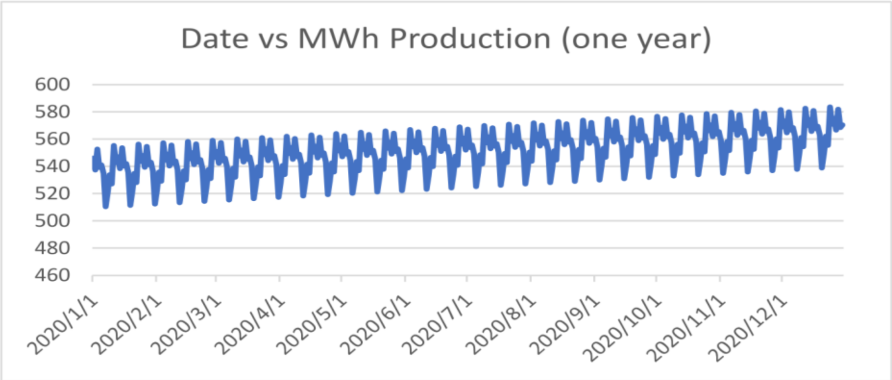

Figure 10: One-year prediction results of MWh from 1 January 2020 to 1 December 2020.

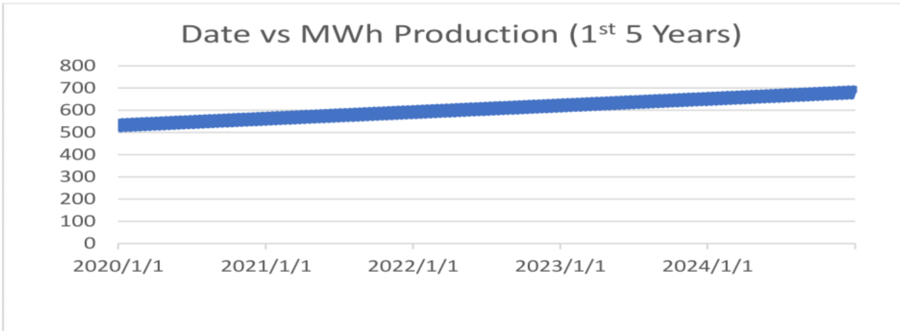

Figure 11: First 5 years' prediction results of MWh from 1 January 2020 to 1 January 2024.

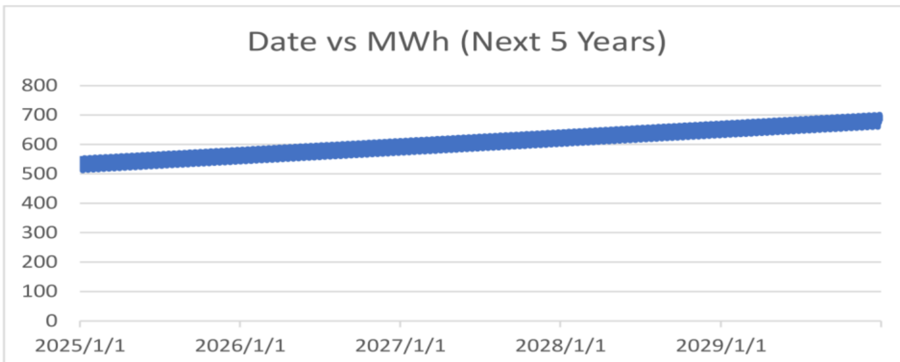

Figure 12: Second 5 years' prediction results of MWh from 1 January 2025 to 1 January 2029.

Future Work

While the proposed hybrid artificial intelligence framework demonstrates improved accuracy and robustness in forecasting solar power generation, there remain several opportunities for further research and enhancement:

1. **Incorporation of Deep Learning Architectures**: Future studies can explore the integration of advanced deep learning models, such as Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNN), and Transformer-based architectures, to better capture temporal dependencies and spatial patterns in solar irradiance and meteorological data.

2. **Real-Time Forecasting and Deployment**: The current model operates in an offline environment. Future work could focus on adapting the framework for real-time forecasting applications, including deployment on edge computing platforms or cloud-based systems for continuous monitoring and prediction [44].

3. **Multi-Modal Data Integration**: Integrating additional data sources such as satellite imagery, sky cameras, weather forecasts, and Internet of Things (IoT) sensor data may further enhance forecasting performance by providing richer contextual information.

4. **Hybrid Optimization Techniques**: The use of hybrid optimization algorithms, such as genetic algorithms, particle swarm optimization, or Bayesian optimization, can be investigated to improve the tuning of model parameters and enhance forecasting accuracy.

5. **Geographical Generalization and Transfer Learning**: Developing models that generalize well across different climatic zones and geographical locations remains a challenge. Applying transfer learning techniques to adapt pre-trained models to new locations with limited data could be a valuable direction [45].

6. **Uncertainty Quantification and Explainability**: Incorporating uncertainty quantification methods and explainable AI (XAI) approaches can help assess the reliability of forecasts and improve model interpretability, which is critical for decision-making by grid operators and policymakers.

7. **Integration with Energy Management Systems**: Future research could also explore how the proposed framework can be integrated with smart grid and energy storage systems to facilitate automated demand-response strategies and optimize energy dispatch.

By addressing these directions, future work can further advance the capabilities of AI-based solar forecasting systems and support the broader goal of reliable, sustainable, and intelligent energy management.

Conclusion

Accurate forecasting of solar power generation is critical for the stable and efficient integration of renewable energy into modern power systems. In this study, we proposed a hybrid artificial intelligence (AI) framework that combines machine learning (ML) algorithms with time-series analysis techniques to enhance the precision and reliability of solar power forecasting. By integrating models such as decision trees, support vector machines (SVM), and artificial neural networks (ANN) with traditional time-series methods like ARIMA and exponential smoothing, the framework effectively captures both short-term fluctuations and long-term trends in solar power output. Experimental evaluations using real-world datasets demonstrated that the hybrid model significantly outperforms conventional forecasting methods in terms of accuracy, robustness, and adaptability. The model's ability to handle varying environmental conditions and geographical locations highlights its potential for wide-scale deployment in diverse solar energy systems. Moreover, the framework's flexibility makes it suitable for integration with existing grid infrastructure and energy management platforms. The results of this research emphasize the value of hybrid AI approaches in addressing the challenges of renewable energy forecasting. By bridging the gap between data-driven machine learning and classical statistical modeling, the proposed framework contributes to improving forecast reliability, supporting grid stability, and reducing the uncertainty inherent in solar power generation. This work provides a foundation for future advancements in intelligent energy forecasting systems and supports the broader transition to sustainable and resilient energy infrastructures.

REFERENCES:

Ngo, N. T., Pham, A. D., Truong, T. T. H., Truong, N. S., & Huynh, N. T. (2022). Developing a hybrid time-series artificial intelligence model to forecast energy use in buildings. Scientific Reports, 12(1), 15775.

Salman, D., Direkoglu, C., Kusaf, M., & Fahrioglu, M. (2024). Hybrid deep learning models for time series forecasting of solar power. Neural Computing and Applications, 36(16), 9095-9112.

Bhutta, M. S., Li, Y., Abubakar, M., Almasoudi, F. M., Alatawi, K. S. S., Altimania, M. R., & Al-Barashi, M. (2024). Optimizing solar power efficiency in smart grids using hybrid machine learning models for accurate energy generation prediction. Scientific Reports, 14(1), 17101.

Sharadga, H., Hajimirza, S., & Balog, R. S. (2020). Time series forecasting of solar power generation for large-scale photovoltaic plants. Renewable Energy, 150, 797-807.

Qu, Y., Xu, J., Sun, Y., & Liu, D. (2021). A temporal distributed hybrid deep learning model for day-ahead distributed PV power forecasting. Applied Energy, 304, 117704.

Chou, J. S., & Tran, D. S. (2018). Forecasting energy consumption time series using machine learning techniques based on usage patterns of residential householders. Energy, 165, 709-726.

Abdul Baseer, M., Almunif, A., Alsaduni, I., & Tazeen, N. (2023). Electrical power generation forecasting from renewable energy systems using artificial intelligence techniques. Energies, 16(18), 6414.

Hissou, H., Benkirane, S., Guezzaz, A., Azrour, M., & Beni-Hssane, A. (2024). A lightweight time series method for prediction of solar radiation. Energy Systems, 1-38.

Khosravi, A., Koury, R. N. N., Machado, L., & Pabon, J. J. G. (2018). Prediction of hourly solar radiation in Abu Musa Island using machine learning algorithms. Journal of Cleaner Production, 176, 63-75.

Rajagukguk, R. A., Ramadhan, R. A., & Lee, H. J. (2020). A review on deep learning models for forecasting time series data of solar irradiance and photovoltaic power. Energies, 13(24), 6623.

Javaid, A., Sajid, M., Uddin, E., Waqas, A., & Ayaz, Y. (2024). Sustainable urban energy solutions: Forecasting energy production for hybrid solar-wind systems. Energy Conversion and Management, 302, 118120.

Lipu, M. H., Miah, M. S., Hannan, M. A., Hussain, A., Sarker, M. R., Ayob, A., ... & Mahmud, M. S. (2021). Artificial intelligence based hybrid forecasting approaches for wind power generation: Progress, challenges and prospects. IEEe Access, 9, 102460-102489.

Niu, W. J., Feng, Z. K., Li, S. S., Wu, H. J., & Wang, J. Y. (2021). Short-term electricity load time series prediction by machine learning model via feature selection and parameter optimization using hybrid cooperation search algorithm. Environmental Research Letters, 16(5), 055032.

Zhu, C., Wang, M., Guo, M., Deng, J., Du, Q., Wei, W., & Zhang, Y. (2025). Innovative approaches to solar energy forecasting: unveiling the power of hybrid models and machine learning algorithms for photovoltaic power optimization. The Journal of Supercomputing, 81(1), 20.

Wang, H., Liu, Y., Zhou, B., Li, C., Cao, G., Voropai, N., & Barakhtenko, E. (2020). Taxonomy research of artificial intelligence for deterministic solar power forecasting. Energy Conversion and Management, 214, 112909.

Dong, N., Chang, J. F., Wu, A. G., & Gao, Z. K. (2020). A novel convolutional neural network framework based solar irradiance prediction method. International Journal of Electrical Power & Energy Systems, 114, 105411.

Agga, A., Abbou, A., Labbadi, M., El Houm, Y., & Ali, I. H. O. (2022). CNN-LSTM: An efficient hybrid deep learning architecture for predicting short-term photovoltaic power production. Electric Power Systems Research, 208, 107908.

Rahman, M. M., Shakeri, M., Tiong, S. K., Khatun, F., Amin, N., Pasupuleti, J., & Hasan, M. K. (2021). Prospective methodologies in hybrid renewable energy systems for energy prediction using artificial neural networks. Sustainability, 13(4), 2393.

Zafar, A., Che, Y., Ahmed, M., Sarfraz, M., Ahmad, A., & Alibakhshikenari, M. (2023). Enhancing power generation forecasting in smart grids using hybrid autoencoder long short-term memory machine learning model. IEEE Access, 11, 118521-118537.

Saxena, N., Kumar, R., Rao, Y. K., Mondloe, D. S., Dhapekar, N. K., Sharma, A., & Yadav, A. S. (2024). Hybrid KNN-SVM machine learning approach for solar power forecasting. Environmental Challenges, 14, 100838.

Zafar, A., Che, Y., Sehnan, M., Afzal, U., Algarni, A. D., & Elmannai, H. (2024). Optimizing solar power generation forecasting in smart grids: A hybrid convolutional neural network-autoencoder long short-term memory approach. Physica Scripta, 99(9), 095249.

Mohammadi, M., Jamshidi, S., Rezvanian, A., Gheisari, M., & Kumar, A. (2024). Advanced fusion of MTM-LSTM and MLP models for time series forecasting: an application for forecasting the solar radiation. Measurement: Sensors, 33, 101179.

Wazirali, R., Yaghoubi, E., Abujazar, M. S. S., Ahmad, R., & Vakili, A. H. (2023). State-of-the-art review on energy and load forecasting in microgrids using artificial neural networks, machine learning, and deep learning techniques. Electric power systems research, 225, 109792.

Mucomole, F. V., Silva, C. A. S., & Magaia, L. L. (2025). Parametric Forecast of Solar Energy over Time by Applying Machine Learning Techniques: Systematic Review. Energies, 18(6), 1460.

Küçüktopçu, E., Cemek, B., & Simsek, H. (2024). Comparative analysis of single and hybrid machine learning models for daily solar radiation. Energy Reports, 11, 3256-3266.

Huang, X., Li, Q., Tai, Y., Chen, Z., Liu, J., Shi, J., & Liu, W. (2022). Time series forecasting for hourly photovoltaic power using conditional generative adversarial network and Bi-LSTM. Energy, 246, 123403.

Rajulu, A. G., Mohanty, P., Pati, U. C., & Mahapatra, K. (2024, November). SolarPred: Hybrid Machine Learning Based Time Series Prediction Model for Solar Energy Applications. In 2024 IEEE 4th International Conference on Applied Electromagnetics, Signal Processing, & Communication (AESPC) (pp. 1-6). IEEE.

Essam, Y., Ahmed, A. N., Ramli, R., Chau, K. W., Idris Ibrahim, M. S., Sherif, M., ... & El-Shafie, A. (2022). Investigating photovoltaic solar power output forecasting using machine learning algorithms. Engineering Applications of Computational Fluid Mechanics, 16(1), 2002-2034.

Zafar, M. H., Khan, N. M., Mansoor, M., Mirza, A. F., Moosavi, S. K. R., & Sanfilippo, F. (2022). Adaptive ML-based technique for renewable energy system power forecasting in hybrid PV-Wind farms power conversion systems. Energy Conversion and Management, 258, 115564.

Benti, N. E., Chaka, M. D., & Semie, A. G. (2023). Forecasting renewable energy generation with machine learning and deep learning: Current advances and future prospects. Sustainability, 15(9), 7087.

Shi, J., Guo, J., & Zheng, S. (2012). Evaluation of hybrid forecasting approaches for wind speed and power generation time series. Renewable and Sustainable Energy Reviews, 16(5), 3471-3480.

Abou Houran, M., Bukhari, S. M. S., Zafar, M. H., Mansoor, M., & Chen, W. (2023). COA-CNN-LSTM: Coati optimization algorithm-based hybrid deep learning model for PV/wind power forecasting in smart grid applications. Applied Energy, 349, 121638.

Hou, M., Zhang, T., Weng, F., Ali, M., Al-Ansari, N., & Yaseen, Z. M. (2018). Global solar radiation prediction using hybrid online sequential extreme learning machine model. Energies, 11(12), 3415.

Heidari, A., & Khovalyg, D. (2020). Short-term energy use prediction of solar-assisted water heating system: Application case of combined attention-based LSTM and time-series decomposition. Solar Energy, 207, 626-639.

Ghimire, S., Nguyen-Huy, T., Deo, R. C., Casillas-Perez, D., & Salcedo-Sanz, S. (2022). Efficient daily solar radiation prediction with deep learning 4-phase convolutional neural network, dual stage stacked regression and support vector machine CNN-REGST hybrid model. Sustainable Materials and Technologies, 32, e00429.

Ahmed, D. M., Hassan, M. M., & Mstafa, R. J. (2022). A review on deep sequential models for forecasting time series data. Applied Computational Intelligence and Soft Computing, 2022(1), 6596397.

Akhter, M. N., Mekhilef, S., Mokhlis, H., & Mohamed Shah, N. (2019). Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. IET Renewable Power Generation, 13(7), 1009-1023.

Thipwangmek, N., Suetrong, N., Taparugssanagorn, A., Tangparitkul, S., & Promsuk, N. (2024). Enhancing short-term solar photovoltaic power forecasting using a hybrid deep learning approach. IEEE Access.

Alkabbani, H., Hourfar, F., Ahmadian, A., Zhu, Q., Almansoori, A., & Elkamel, A. (2023). Machine learning-based time series modelling for large-scale regional wind power forecasting: a case study in Ontario, Canada. Cleaner Energy Systems, 5, 100068.

Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. (2023). A review of ARIMA vs. machine learning approaches for time series forecasting in data driven networks. Future Internet, 15(8), 255.

Lim, S. C., Huh, J. H., Hong, S. H., Park, C. Y., & Kim, J. C. (2022). Solar power forecasting using CNN-LSTM hybrid model. Energies, 15(21), 8233.

Akkem, Y., Biswas, S. K., & Varanasi, A. (2023). Smart farming using artificial intelligence: A review. Engineering Applications of Artificial Intelligence, 120, 105899.

Al-Ali, E. M., Hajji, Y., Said, Y., Hleili, M., Alanzi, A. M., Laatar, A. H., & Atri, M. (2023). Solar energy production forecasting based on a hybrid CNN-LSTM-transformer model. Mathematics, 11(3), 676.

Azimi, R., Ghayekhloo, M., & Ghofrani, M. (2016). A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting. Energy Conversion and Management, 118, 331-344.

Selmy, H. A., Mohamed, H. K., & Medhat, W. (2024). A predictive analytics framework for sensor data using time series and deep learning techniques. Neural Computing and Applications, 36(11), 6119-6132.

Gaizen, S., Fadi, O., & Abbou, A. (2020). Solar power time series prediction using wavelet analysis. International Journal of Renewable Energy Research (IJRER), 10(4), 1764-1773.