Machine Learning for Smart Photovoltaic Applications- Forecasting to Welcome the Solar Era

By

Sahar Abdul- Munim Daebes

Submitted in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

Department of Electronic and Computer Engineering

College of Engineering, Design and Physical Science

Brunel University London

UK

December 2024

DEDICATION

I would like to dedicate this thesis to my late mother and father Rabab and AbdulMonum who helped me in all things great and small. Special thanks to my lovely husband Kamel and my sons Yehia, Ali, Yasser, and Mostafa for their emotional support and patience. A special thanks especially to my lovely daughter-in-law Malak and my baby granddaughter Dalia for all the moments of happiness that helped me throughout my work on the thesis.

ACKNOWLEDGMENTS

All praise is to Almighty Allah for guiding me at every stage of my life. Thanks to Allah for the strengths and guidance in completing this thesis.

I would like to express my sincere gratitude and appreciation to my great supervisor, Professor Mohamed Darwish, for his continuous support throughout my PhD studies. I appreciate his patience, continuous motivation, and immense knowledge. His support and guidance helped me throughout my research and writing of this thesis.

I would also like to thank the faculty members Dr Chun Sing, Prof. Hamed AL Rewashedy, and Prof. Maisam Abod for their insightful comments and encouragement and for the valuable discussions and research questions that widened my study from various perspectives.

Author’s Declaration

I hereby declare and confirm that this thesis is my work and effort. Also, it has not been submitted anywhere for any award. Where other sources of information have been used, they have been acknowledged.

Sahar Abd Al-Minium Daebes

December 2024

Contents

List of Abbreviations used

GHI	Global Horizontal Irradiance.
DHI	Diffuse Horizontal Irradiance.
L S TM	Long Short-Term Memory.
Tmax	Max Air Temperature.
Tmean	Mean Air Temperature.
Tmean	Average Air Temperature.
ANFIS	Adaptive Network Fuzzy Inference System.
A R	Linear Auto-Regressive Model.
A RD	Automatic Relevance Determination.
A RIMA	Autoregressive Integrated Moving Average.
A RMA	Autoregressive Moving Average.
B DT	Boosted Decision Tree.
B PNN	Back-Propagation Neural Network.
C NN	Convolutional Neural Network.
DL	Deep Learning.
DNI	Direct Normal Irradiance.
DP	Dew Point.
F F NN	Feed Forward Neural Network.
F L	Federated Learning.
G BRT	Gradient Boosted Regression Trees.
G CPV	Grid-Connected Photovoltaic.
GFM	Generalized Fuzzy model.
I SOs	Independent System Operators.
L R	Linear Regression.
MAE	Maximum Absolute Error.
MAPE	Mean Absolute Percentage Errors.
MBE	Mean Bias Error.
MLP	Multi-Layer Perception.

MLR Multiple Linear Regression.
MTSF Multivariate Time-Series Forecasting.

N AR Nonlinear Auto-Regressive Model.

NGA Niching Genetic Algorithm.

nMAE Normalized Mean Absolute Error.

NWP Numerical Weather Prediction.

P Pressure.

P HANN Physical Hybrid Artificial Neural Network.

P W Precipitable water.

R BFNN Radial Basis Function Neural Network.

R ENN Recurrent Elman Neural Network.

R F R Random Forest Regression.

R H Relative Humidity.

R MSE Root Mean Square Error.

R NN Recurrent Neural Networks.

S A Surface Albedo.

S CFT Seasonal Clustering Forecasting Technique.

S VM Support Vector Machine.

S ZA Solar Zenith Angle.

T Temperature.

UTSF Univariate Time-Series Forecasting.

Table of figures & tables

Abstract

The inherent uncertainty in photovoltaic (PV) power generation remains a significant challenge to the seamless integration of solar energy into modern power systems. This study addresses this issue by employing advanced machine learning (ML) techniques, with a particular focus on Long Short-Term Memory (LSTM) networks—a class of recurrent neural networks (RNNs) to improve the forecasting accuracy of solar power output. The methodology combines deep learning models with Maximum Power Point (MPP) tracking to enhance both predictive performance and operational efficiency of PV systems.

The research integrates time-series data collected from real-world PV installations, capturing key variables such as solar irradiance, temperature, voltage, and current. The LSTM architecture is trained to model the temporal dependencies inherent in these sequences, allowing for accurate forecasting of solar power production. An ensemble LSTM approach is implemented to further enhance the robustness of predictions and reduce mean squared error (MSE). Moreover, the integration of MPP data enables real-time adaptation to changing environmental conditions, thereby improving energy capture efficiency and enabling early detection of system faults.

Complementary to the LSTM framework, traditional time-series models such as SARIMA and ARIMA are also applied to analyze the temporal variation in solar power production. These statistical models provide valuable baseline comparisons and offer additional insights into the volatility of PV output, which is known to cause operational issues such as frequency instability, dispatch challenges, and voltage/current surges within the grid.

The research demonstrates a hybrid methodology that leverages deep learning and statistical modeling, coupled with MPP analysis, to enhance the reliability, efficiency, and fault tolerance of solar energy systems paving the way for more stable and scalable integration of PV power into the energy mix.

Keywords— Machine Learning algorithms ML, Long Short-Term Memory LSTM, transition energy system, PV forecasting algorithms, SARIMA / ARIMA.

Chapter 1
1.1 Introduction

Solar power production forecasting is a crucial aspect of renewable energy management, as it helps to optimise the integration of solar power into the existing power grid. Accurate forecasting of solar power production is essential for grid operators, energy traders, and renewable energy plant managers to make informed decisions regarding energy generation and distribution [1].

In recent years, there has been a significant increase in the installation of solar power plants, leading to a growing demand for accurate solar power production forecasts. However, forecasting solar power production is a challenging task due to the inherent uncertainty associated with weather conditions, such as cloud cover, wind speed, and atmospheric conditions [2].

To address these challenges, researchers and scientists have developed various methodologies and models to improve the accuracy of solar power production forecasting. This subchapter explores the background of solar power production forecasting and provides an overview of the different techniques and approaches used to reduce uncertainty in forecasting [3].

The accuracy of solar power forecasts is influenced by several factors, including geographic location, local weather conditions, and the time horizon of the forecast. In regions with highly variable weather patterns, such as areas prone to frequent cloud cover or storms, forecasting becomes more complex and uncertain. Nevertheless, continuous improvements in forecasting models and data acquisition technologies have led to more reliable predictions, allowing solar energy to play a more significant role in the energy mix [4].

1.2 Background of Integrated Artificial Intelligence AI in Solar Power Generation Prediction

The convergence of artificial intelligence with photovoltaic technology represents a significant step forward in advancing renewable energy, contributing to technological innovation and environmental sustainability [5]. Photovoltaic (PV) technology has emerged as a key solution for sustainable energy production, driven by the need to reduce carbon emissions and reliance on fossil fuels. However, optimising the performance and efficiency of PV systems remains a challenge due to factors like weather variability, system ageing, and environmental conditions. In recent years, machine learning (ML) and deep learning (DL) techniques have gained attention for their ability to address these complexities. ML and DL algorithms can process vast amounts of data collected from PV systems, enabling more accurate forecasting, fault detection, and system optimisation [6]. These techniques enhance predictive maintenance, power output predictions, and energy yield optimisation, improving overall system performance. By integrating ML and DL models with real-time data, PV systems can adapt dynamically, making renewable energy generation more efficient and reliable. The best machine learning (ML) algorithm for photovoltaic (PV) applications depends on the specific problem being addressed, such as power output forecasting, fault detection, or performance optimisation [7]. Here is a breakdown of commonly used algorithms for different PV-related tasks, such as

Power Output Forecasting: Predicting the power output of a PV system involves dealing with time-series data like solar irradiance, temperature, and historical power output. In this field, the best algorithms using ML are:

- Gradient Boosting Machines (GBM), like XGBoost, that highly effective for structured time-series data. They can handle nonlinear relationships between weather parameters and power output and provide high accuracy [7].
- Recurrent Neural Networks (RNN) or Long Short-Term Memory (LSTM): These deep learning models are designed for sequential data and are particularly effective at capturing patterns over time, making them ideal for long-term power output forecasting. Using LSTMs can model long-term dependencies and trends, making them ideal for PV output that fluctuates over time due to weather conditions [6], [7]
- Support Vector Regression (SVR): For smaller datasets or when more interpretability is needed, SVR can be an excellent choice. It is effective in handling nonlinear relationships between input features and power output [5].
- Fault Detection and Diagnosis: Detecting faults or anomalies in PV systems is another critical application, where classification algorithms excel [5], [8]
- Random Forest: This ensemble learning algorithm is a popular choice for fault detection in PV systems due to its robustness and ability to handle high dimensional data with

many features. Utilising Random Forest provides a reliable classification of faults by combining multiple decision trees and can identify patterns related to diverse types of PV system faults [6], [7].

- Support Vector Machine (SVM), particularly useful for binary fault detection SVM can create clear decision boundaries in high-dimensional spaces. The reason for using SVM is effective when there is a clear distinction between normal and fault conditions, especially with smaller datasets [8], [9].
- k-Nearest Neighbours (k-NN), k-NN is simple and effective for real-time fault detection. It works by classifying a fault based on the closest examples in the data. It is easy to implement and works well with labelled historical data, though it may struggle with large datasets.[8], [9]

Performance Optimisation and Maintenance, for optimising system performance or scheduling predictive maintenance, regression models and anomaly detection algorithms can be applied [8], [9].

Artificial Neural Networks (ANNs), particularly shallow networks, are commonly used for performance optimisation by predicting system efficiency under different conditions. It can model complex relationships in the PV system and provide insights into optimal settings for maximising [5], [6].

Autoencoders, these unsupervised neural networks are used for anomaly detection in PV performance data. They work by learning the normal operating behaviour of the system and flagging deviations as potential issues. They are excellent for identifying subtle faults or performance drops before they become severe [7], [8].

Energy Yield Prediction and Forecasting, predicting overall energy yield based on weather forecasts and system parameters, is another essential task.

Random Forest and GBM, these ensemble models are again useful here because they can model nonlinear relationships between weather variables e.g., temperature, cloud cover, irradiance, and energy production [10].

LSTM or CNN-LSTM Hybrid Models, for more advanced forecasting, LSTM networks can be combined with Convolutional Neural Networks (CNNs) to capture both spatial (CNN) and temporal (LSTM) features from weather data or satellite imagery. This combination is effective when the input data includes spatial-temporal patterns, like cloud movement, which impacts PV performance [11].

Hybrid and Ensemble Models, in some cases, use a hybrid approach or combine several models as ensemble learning to yield better results. For example, it can combine a Random Forest for feature selection and a gradient boosting machine GBM for prediction, or use a stacking method to merge predictions from multiple algorithms like support vector machine SVM, artificial neural network ANNs, and gradient boosting machine GBM [11].

In recent years, solar power has grown significantly in importance as a source of energy in many nations; nevertheless, the intermittent nature of photovoltaic (PV) power generation has a major effect on power networks already in place. Accurate solar power forecasting techniques are needed to reduce this uncertainty and preserve system security [12]. Forecasting solar power involves projecting how much electricity a photovoltaic (PV) system will produce in the future [13]. The process of obtaining and evaluating data to estimate solar power generation over different time horizons to lessen the effects of solar intermittency is known as solar power forecasting [14]. Forecasts of solar power are used to trade power and manage the electric grid more effectively [12]

Issues of intermittency and reliability arise as significant obstacles to the application of solar energy, such as low conversion efficiency and material cost, which continue to decline. Solar forecasting has often helped address and mitigate the intermittency issue [14]. Solar power converts sunlight into electrical energy, either directly through photovoltaics (PV) or indirectly through concentrated solar power CSP. Solar panels use the photovoltaic PV effect to convert light into an electric current, which is defined as the net rate of flow of electric charge through a surface [15].

Recent advances in solar power generation forecasting algorithms have made it possible to use solar energy resources in microgrid systems more effectively. Most researchers have investigated a range of techniques for predicting solar power generation, including systems based on machine learning and statistics. Statistical methods, such as Autoregressive Integrated Moving Average (ARIMA), exponential smoothing, and linear regression, have been commonly employed [16]. These methods may have trouble reflecting the complex and nonlinear interactions present in the production of solar electricity since they rely on mathematical models and historical data to create predictions. Accurate short-term solar power projections can be obtained by combining historical solar irradiance data with statistical models, such as ARIMA. These techniques use solar radiation levels and past weather patterns to forecast the amount of solar power produced in the future [17].

The advent of machine learning techniques has drawn interest in the forecasting of solar power generation because of their capacity to identify complex patterns and manage copious amounts of data. Out of all these methods, artificial neural networks ANN [16]. The unpredictable nature of weather makes it difficult to forecast solar energy generation. The efficiency of solar panels and the quantity of energy they produce can be impacted by variables including gloomy days, wind patterns, and the season. These dynamic components increase complexity and inject uncertainty into accurate prediction-making. Therefore, accurate ML/DL models are crucial for predicting solar power with precision [18].

Typically, the solar power forecast considers the Sun's course and the atmospheric conditions. the solar forecasting techniques depend on the forecasting horizon. Three types of forecasting are available [19]

- Short-term (up to seven days ahead)
- Long-term (up to four hours ahead)
- Nowcasting (weeks, months, years)

Since 1970, various forecast horizons have necessitated diverse approaches for solar resource forecasting, as most writers have agreed [20]. Forecast horizons shorter than an hour are usually needed for complex time series and machine-learning models based on ground-based sky photography. Intra-day horizons often anticipate irradiance values up to four or six hours in advance and therefore require irradiance models and satellite pictures. For forecast horizons longer than six hours, numerical weather prediction (NWP) model outputs are typically used [21].

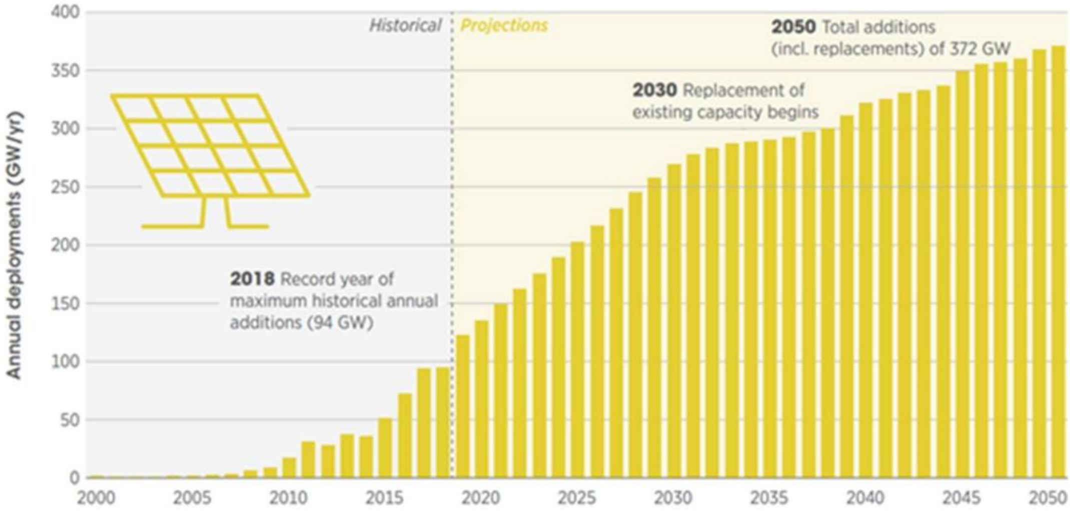
The availability of sunshine can be predicted using a variety of strategies and approaches for effective energy optimisation in various situations. Over time, these techniques for estimating solar power have undergone substantial changes. Occasionally, combining these approaches yields the most precise and successful solar power forecasts. All methods can be traced back to the basic methodologies [22]. Table 1.1 represents the methodologies of forecasting techniques:

Forecasting model	Characteristic model
Physical methods	A predictive approach that relies on numerical weather prediction (NWP) models and sky imaging to simulate the behaviour of the atmosphere, sunlight, and cloud cover

Statistical methods	Persistence forecasting, a straightforward statistical technique, assumes that current solar generation levels, which leverages historical and real-time generation data to statistically adjust predictions
Artificial intelligent	Machine learning ML, deep learning (DL random forest RF and long short-term memory LSTM
Holistic and integrated methods	A forecast-based energy management system combines various forecasting strategies. the integration empowers the energy management system. interaction to continuously optimise the energy flows between all assets.

Table 1.1 The methodologies of solar power forecasting techniques

The exponential growth of PV systems in recent times has made it necessary for PV power generation to be more fully integrated into main grid systems worldwide [23]. Grid-connected photovoltaic (GCPV) systems to reduce the cost of electricity generation [24]. To fully realise the benefits of renewable power generation, scheduling and coordination between the grid system and the renewable source are required [23]. Solar energy poses several challenges due to its inherent unpredictability and irregular nature [25]. Typically, solar sources have variable outputs and peak at random. According to IEA estimates, photovoltaic (PV) installed capacity worldwide was expected to reach 760.4 GW by the end of 2050 [25]. PV plants have been built in millions across the world, but they require constant management, protection, and oversight to remain effective for the rest of their life. Numerous attempts have been made to do this using both conventional and innovative methods in PV technologies [26].



Figur 1.0.1 Annual Global solar PV additions are expected to reach 270 GW in 2030 and 372 GW in 2050 under the REmap scenario, compared with 94 GW in 2018 [27]

Energy technologies are currently moving toward the usage of renewable energy sources to mitigate the consequences of global warming [28]. Alternative energy sources that are most widely used are solar photovoltaic (PV) systems, and these technologies will draw the power operators' attention to electricity generation. In many countries, the use of solar PV energy-producing systems for large-scale electricity production is growing more prevalent [29]. Advanced energy technologies are currently moving toward the usage of renewable energy sources to mitigate the consequences of global warming. Alternative energy sources that are most widely used are solar photovoltaic (PV) systems [30]. [31], [32]. The solar azimuth, solar elevation, and solar zenith angles can all be accurately calculated using the Solar Position Algorithm (SPA) technique [33]. The SPA technique is a method used in solar power forecasting to accurately determine the position of the sun in the sky at any given time and location [34]. The time, date, and coordinates of the location of interest are only a few of the elements that the SPA technique considers [28]. The purpose of SPA calculate the solar zenith and azimuth angles with high precision. These angles are crucial for determining the amount of solar radiation reaching a specific location on Earth.

The sun's position is calculated using trigonometric formulas and astronomical models [35]. The solar azimuth, solar elevation, and solar zenith angles can all be accurately calculated using the SPA technique. The SPA is known for its high accuracy, with uncertainties sun's position [36]. Because of that, it can be used in various solar energy applications, including:

- Optimising the orientation of solar panels

- Calculating potential shading effects
- Improving the accuracy of solar irradiance models [37].

The SPA is particularly useful in clear-sky irradiance, short-term forecasting for accurate sun position data is crucial for predicting solar power output, and long-term planning when assessing the potential solar resource of a location over extended periods. Real-time control tracking to maximise energy capture [38]. SPA is often used in conjunction with other forecasting techniques, such as cloud cover predictions from satellite imagery or numerical weather prediction NWP models to provide more accurate solar power forecasts [[39]. solar power forecasting also depends on other factors like atmospheric conditions, cloud cover, and specific characteristics of the solar energy plant. The Solar Position Algorithm (SPA), created by the National Renewable Energy Laboratory (NREL) in the United States, is a popular solar position algorithm [23].

Accurately forecasting PV power generation can improve system reliability, maintain power quality, reduce the effect of PV power uncertainty on the grid, and increase the penetration of PV systems [37]. Programs known as solar position algorithms pinpoint the sun's location in the sky at a given time, place, and date. The SPA technique can be used to properly determine the solar zenith angle, solar azimuth angle, and solar elevation angle. The time, date, and coordinates of the location of interest are among the many variables that the SPA technique considers [40]. The sun's position is ascertained using trigonometric calculations and astronomical models since solar energy is erratic and intermittent by nature. the solar zenith angle, which is the angle that indicates how directly the solar irradiance is coming in, as shown in Figure 1.2.

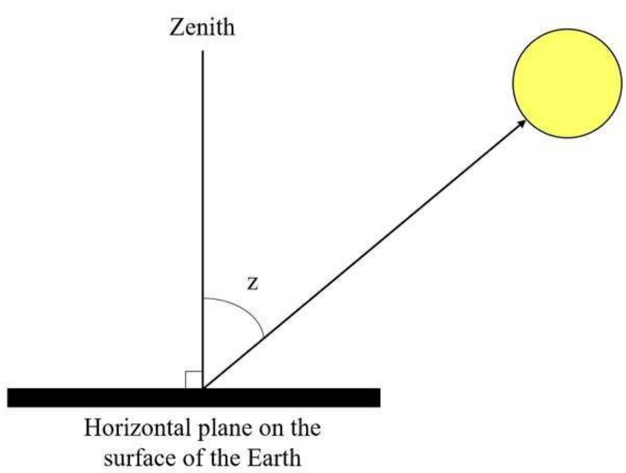


Figure 1.0.2 Zenith angle based on the timestamps and coordinates of the solar PV panel installation location [41], [42].

Solar sources typically have unpredictable peaks and varied outputs. One of the biggest obstacles for electricity companies and challenges include:

- managing the grid more harshly
- maintaining a continuous balance between production and consumption.
- improving power quality, like voltage and frequency stability, and being difficult to schedule and regulate [43].

Keeping sizeable reserves on hand to guarantee steady and reliable operations may be able to mitigate some of the issues that arise with incorporating PV systems into microgrids. Large reserves, however, can result in increased fuel consumption, and operating, transmission, and maintenance costs, which raise carbon emissions for no apparent reason [44]. Accurate energy estimates are essential for lowering risk, gaining a competitive advantage in the energy markets, and effectively managing investments and wind and solar assets. presents a novel ensemble solar irradiance forecasting model that gives an overview of the different learning approaches used for irradiance prediction and includes eight machine learning models to ensure model variety [39].

More research is being done to develop more accurate PV power output forecasting models considering these problems. It is thus possible to determine the anticipated PV system's output power generation based on local solar radiation and meteorological data. Sun radiation and ambient temperature have a significant impact on three important PV panel characteristics [44], [45]. GHI prediction is a crucial first step for PV power forecast systems. GHI forecasting often belongs to one of three categories:

- Forecasting horizons: "the time frame used to generate projections." The plane predicts the horizon using the selected algorithm and historical data. The plane determines which time intervals apply to the forecast horizon using previous data.
- Predictive techniques refer to methods that use previous data as inputs to create precise forecasts about the future direction of trends. Businesses utilise forecasting to allocate their budgets or plan for anticipated spending in the future.
- The variable parameters, data exogenous and endogenous, are input features. Such as air temperature, air speed, humidity, and global horizontal irradiation GHI.

However, weather data is the fundamental pillar in solar power forecasting; if the researchers or operators lack the resources to purchase solar irradiance sensors or do not have access to sun irradiance measurement stations, they may not be able to obtain the necessary data [33]. Many stakeholders in the renewable energy sector, such as utilities, independent system operators independent system operator (ISOs), and solar power producers, want very precise solar power estimates. In certain authorities, power companies are legally obligated to offer their customers accurate power forecasts as part of their power purchase agreement [46]. ISO oversees all system operations, including scheduling generation, transmission, and reserves, acquiring extra auxiliary services, and managing the system in real time. in addition to system architecture [47]. As shown in Figure 1.4, which illustrates the hierarchical structure of the electrical power system and the stepwise reduction in voltage from generation to consumption. Such a structure ensures that electricity is transmitted efficiently over long distances and delivered safely and reliably to a wide range of customers. The figure elegantly captures the technical and logistical coordination inherent in modern power systems, highlighting the essential role of transformers and network segmentation in maintaining operational efficiency and stability.

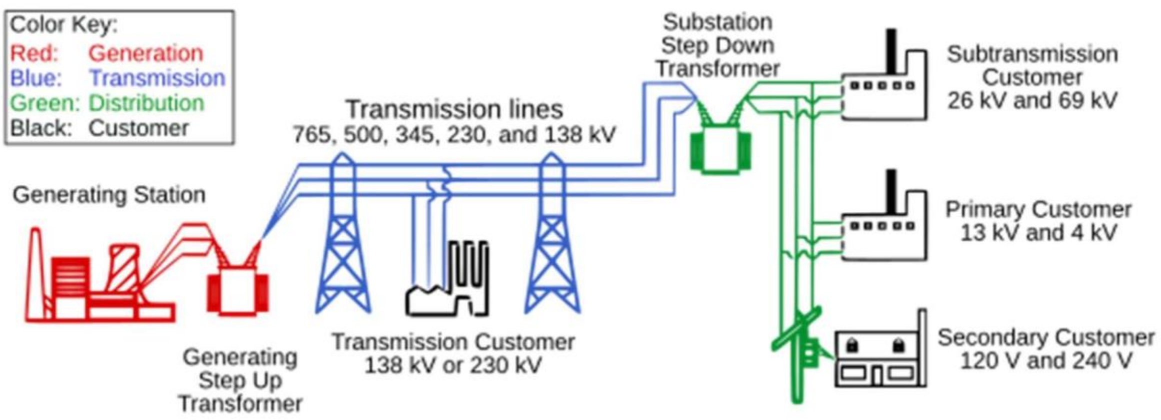


Figure 1. 0.3 Diagram of an electrical grid (generation system in red, transmission system in blue, distribution system in green) [42], [48]

- Generation (Red section) is the process that begins at a generating station, where electricity is produced from various sources (e.g., fossil fuels, nuclear, hydro, or renewables). This step is crucial because higher voltage levels reduce energy loss during transmission.
- Transmission (Blue Section), represent that electricity is transported over long distances through transmission lines, typically at very high voltages such as 765, 500, 345, 230, or 138 kV. High voltage transmission is essential for minimising energy losses and ensuring economic delivery over vast geographic areas.
- Distribution (Green Section) is the upon reaching a substation, a Step-Down Transformer reduces the voltage to levels suitable for distribution. Electricity is then delivered via the distribution network to various categories of customers.
- Customers (Black), these end-users span across industrial, commercial, and residential sectors, each matched with an appropriate voltage level according to their load requirements and usage patterns.

Solar energy forecasting is crucial for the operational integration of solar power into grids and the optimal management of renewable energy resources. For claims that are connected to the electrical grid, PV systems range in size from a few kilowatts (kW) for residential to a few megawatts (MW) for large commercial PV systems [42], [49]. Forecasting solar power makes it hard to predict output when integrated into the grid, and its less reliable, volatile nature can lead to unstable voltage fluctuations, frequency fluctuations, and system outages [49].
One of the proposed forecasts for solar energy is to help balance supply and demand through a combination with electricity storage. As a result, solar forecasting has seen an increase in interest from researchers, grid operators, and other parties involved in the electricity market [42].

1.3 Deep Learning (DL) techniques for solar power forecasting data

The worldwide power grid now uses a lot of solar energy [50]. Improving solar energy prediction accuracy is crucial for effective power system planning, management, and operation. It is critically necessary to adopt a highly precise and sophisticated forecasting strategy to reduce the detrimental effects of photovoltaics on electricity and energy systems [51]. Deep Learning (DL) techniques have become increasingly popular and effective for solar power prediction. By leveraging these DL techniques, practitioners can develop more accurate and robust solar power forecasting models, contributing to better integration into power grids and more efficient energy management systems [52].
The use of Deep Learning techniques for solar energy prediction, in particular Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU), can capture long-term dependencies in solar power data. it is like LSTMs but with a simpler structure [53]. Convolutional Neural Networks (CNNs) are operative in processing image data like satellite imagery or sky images. A key feature of using DL for solar power is like Input Data, extracted spatial features from cloud cover patterns to predict solar irradiance [40], [53] [54].

- Feature Extraction: DL models possess the ability to automatically extract pertinent features from unprocessed data, eliminating the requirement for human feature engineering.
- Temporal Modelling: RNN-based models like LSTM and GRU are particularly good at capturing temporal patterns in solar power data.

- Multi-step Forecasting: This can be made to predict multiple steps, which is important for different time horizons in solar forecasting.
- Handling Non-linearity: It is possible to capture intricate non-linear correlations between input factors and solar power output.
- Scalability: They effectively handle large datasets, making them suitable for processing big data in solar forecasting applications

Probabilistic connections between the past and future data are found by applying Deep Learning (DL) techniques to historical data. It has been shown that these models outperform statistical models in predicting solar irradiation [31]. The design workflow for the forecasting application consists of four main components [34], [54].

- Cleaning, inspecting, merging, and analysing data sets is known as data pre-processing.
- Feature selection comprises deciding which features to combine to achieve the optimum data use performance.
- method selection, which uses predicted performance to determine which DL algorithm is best.
- Model design, which modifies the hyperparameter settings to try to maximise the performance of the model. The design process has four steps, as shown in Figure 1.4. A substantial portion of the literature now in publication applies ML approaches to solar power and solar irradiance forecasting and modelling; the most widely used model for solar irradiance forecasting is the LSTM model [31], [46]. The following figure shows the metrological data sources for choosing forecasting modelling.

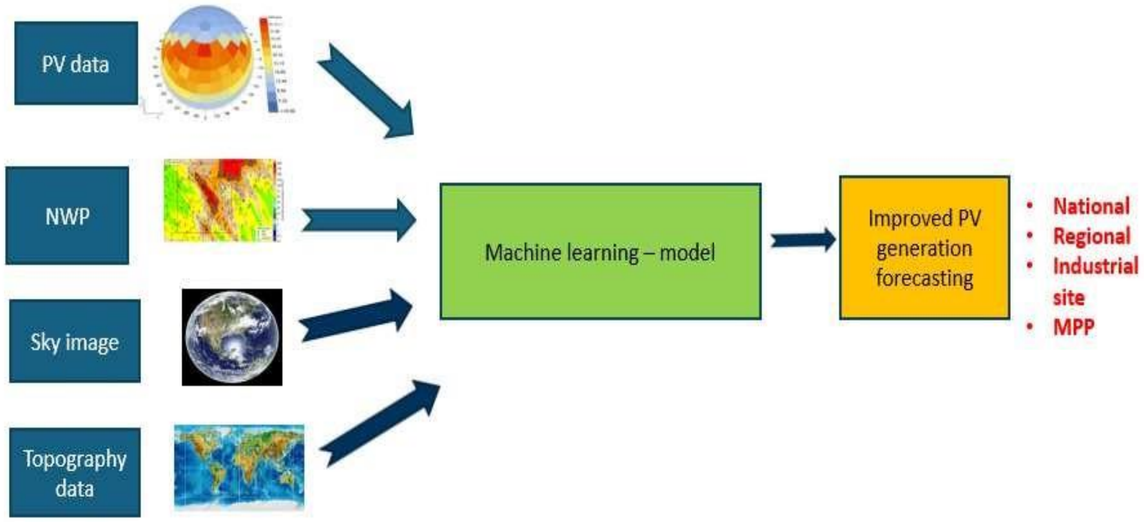


Figure 1.0.4 Machine learning forecasting techniques

A widely explored solar radiation forecasting technique based on DL draws attention to the fascinating interest in recently found processes involving the assessment of forecasting methods. DL technique for a special time horizon, spatial-temporal resolution, examination of input variables and location-based accuracy assessments. Future recommendations focus on establishing a benchmarking framework and creating publicly available standardised datasets. a block diagram summarising the contents of different techniques is presented in Figure 1.3 [53], [54], [55].

1.4 Problem statement

The increasing reliance on solar energy as a key component of the global energy mix presents significant challenges due to its inherent variability and intermittency, which is primarily driven by changing weather conditions and atmospheric dynamics. This variability can lead to imbalances in power supply and demand, creating operational challenges for grid operators, energy market participants, and energy storage systems. Therefore, the central problem in solar power forecasting is the need to accurately predict solar energy generation over various time horizons to ensure the reliable integration of solar power into the grid, optimise energy markets, and improve the efficiency of energy storage and dispatch systems. The integration of variable solar power into the grid requires balancing supply and demand in real-time for grid stability and Energy management. Without accurate forecasting, grid operators may face overgeneration or shortages, leading to costly adjustments such as energy curtailment. These issues can cause inefficiencies and undermine the economic and environmental benefits of solar power.

1.5 Objectives

The number of PV panel installations has increased recently due to factors such as the growing competitiveness of solar PV panels as a renewable energy source. The main objectives of predicting solar power are:

- Solar forecasting methods with ML algorithms to understand the Significance of Accurate Solar Output Predictions.
- Optimising energy production, helping operators of solar power plants and assessing forecasting methods.
- For the 2050 scenario of net-zero emission, all utility companies and investors to make informed judgments about resource planning and investment, they need reliable projections of solar generation.
- Future recommendations and consistency of the training data and stability, dependability of the grid, and Solar energy integration to enhance careful supply and demand balancing.

1.6 Research Questions

To investigate these objectives by using machine learning and deep learning techniques to answer the research questions related to the research objectives, by the following:

1- Does the solution deliver a flexible level of forecasting to meet the need?
2- Is the solar forecasting solution built on a framework and methodology that is proven in a commercial environment?
3- Which are the most widely used ML-based techniques to forecast solar power output?
4- What are the main operational, economic, and managerial grid challenges addressed by forecasting solar power outputs?

1.7 Research Gap

A research gap in solar power forecasting arises from the current dichotomy between two major forecasting approaches: (1) forecasting solar irradiance, which predicts the amount of sunlight reaching the Earth's surface, and (2) forecasting solar power output, which directly estimates the energy generated by photovoltaic (PV) systems. While both approaches are crucial for optimising solar energy systems, existing studies often focus on one aspect without integrating both. This separation limits the accuracy and utility of predictions, especially in real-world applications where both irradiance and power output fluctuate due to dynamic environmental factors.

In contrast, my research addresses this gap by combining solar irradiance forecasting with solar power output forecasting into a unified model. This integrated approach will enhance the accuracy of solar energy predictions, providing a more comprehensive solution for energy management systems. Additionally, this research employs Long Short-Term Memory (LSTM) networks with stacked layers, which can capture temporal dependencies in both irradiance and power output. LSTMs are well-suited for this task due to their ability to handle complex time-series data, and using multiple layers allows the model to capture more intricate patterns in solar energy variations. By bridging this gap, the research aims to offer a more robust and reliable forecasting model with all forecasting approaches that benefit both grid operators and solar energy producers, optimising energy allocation and improving sustainability efforts.

1.8 Methodological approach to Close the uncertainty of solar energy production with the LSTM approach

The use of machine learning techniques for solar power forecasting. Meanwhile, more data availability due to machine learning techniques has enabled improved prediction performance. For many stakeholders in the energy sector, forecasting solar PV energy output is essential. Time series models and machine learning can be used to do this. This study compares various time series models and machine learning techniques. Machine learning algorithms were easier to implement. Artificial neural networks, gradient-boosting regression trees, and long short-term memory (LSTM).

The purpose of this study is to examine the effectiveness of the LSTM model, which has grown in popularity as a forecaster of solar irradiation. The goal of the study is to provide reliable models for solar irradiation hourly estimates. In the first section of this work, the behaviour of LSTM models will be analysed under geographical and climatic conditions and in compliance with historical data on sun irradiation. employing exogenous and endogenous variables, respectively, as input features for hourly solar irradiance forecasting models. The research includes a comparison of Multivariate Time-Series Forecasting and Univariate Time-Series

Forecasting (UTSF) (MTSF). Second, to better understand the interactions between model components, forecasting models often incorporate many high-dimensional variable inputs. The evaluation of solar power predicting methods based on artificial intelligence (AI) techniques for spatial-temporal resolution consideration, specific time horizon, resolution, examination of input variables accessibility, location-based accuracy estimations, and utilisation of assessment metrics for planned uses. The creation of publicly accessible standardised datasets and the establishment of a benchmarking framework are the main proposals for the future. Predictive modelling is characterised as a multivariate problem where one variable can affect other input and output variables in a multitude of straightforward or intricate ways. Most people agree that predictive modelling is a multidimensional problem in which a single variable can have a wide range of simple or complex effects on other input and output variables. Figure 1.5 shows the endogenous and exogenous data in applying AI and machine learning to predict solar power.

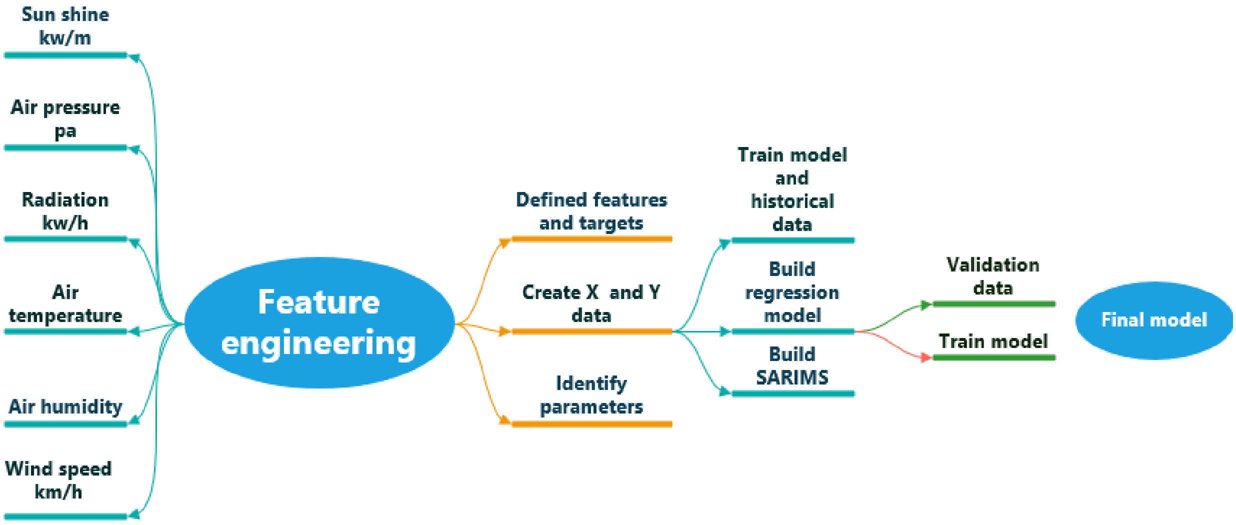


Figure 1.0.5 An AI-feature engineering block diagram based on a solar forecasting application.

The creation of publicly accessible standardised datasets and the establishment of a benchmarking framework are the main proposals for the future. From this viewpoint, the forecasting process can be reduced to three important parts:

- The data collection,
- The data pre-processing and post-processing,
- The testing and training data model through model assessment and output evaluation [57].

As a result, several queries are raised, which the current study will try to address:
These types of associations that occur between exogenous and endogenous variables deliver a flexible level of prediction that could enhance the output result. In solar power forecasting algorithms, redundancy and relevancy are crucial ideas, especially when it comes to feature selection and enhancing model performance. The relevance of solar power forecasting refers to the input features with high relevance that have a strong correlation or predictive power for the target variables such as solar irradiance, temperature, sky cover cloud, time of the day, season of the year and historical data [58], [59], [60].

This study will present a feature selection method based on correlation analysis for redundancy and relevancy measures to decide which qualities are redundant or irrelevant. Pearson's correlation coefficient is typically used to quantify redundant qualities to identify linear relationships between the exogenous variables. It may be countered that determining duplicate variables requires more information than just a linear connection.
It is noted that the predictive performance is impacted and dependent on spatial and meteorological factors when a univariate LSTM forecasting model is converted to a multivariate forecasting model [61]. For this reason, the LSTM model must take seasonality into account. Additional research reveals that variations in seasonality necessitate a deeper investigation of seasonality patterns in meteorological data using LSTM models, which can extract nonlinearity patterns originating from external and endogenous variables [61], [62].
The effectiveness of the LSTM model was demonstrated when it was compared to DL approaches in the field of solar irradiance forecasting. This demonstrated the model's capacity to learn from nonlinearity behaviour in solar irradiance data with a wide variety of temporal dependencies [61], [63], [64].
Due to the various weather conditions, most of the research indicated that the LSTM model should be a viable method for addressing the computational complexity [65]. However, problems with data quality can also affect high-dimensional heterogeneous data. To overcome these, the data mining procedure may be used to search through big datasets for correlations, anomalies, patterns, and trends. To reduce uncertainty in data, a clustering method like k-means could be specifically used to classify data points as sunny, overcast, or wet clusters. solar irradiance forecasting [57], [58].

1.9 Increasing Solar Power Forecast Accuracy depends on exogenous and endogenous variables

Increasing solar power forecast accuracy depends on both exogenous and endogenous variables. Exogenous Variables include weather Conditions, cloud cover, the most crucial factor affecting solar irradiance, Aerosols like Smoke, dust, temperature, wind speed, and humidity. As well as the geographic Factors like location, latitude, and longitude. However, shading, solar angles and daytime length vary seasonally [66]. Figure 1.6 shows the exogenous variables as external factors that can influence solar power generation.

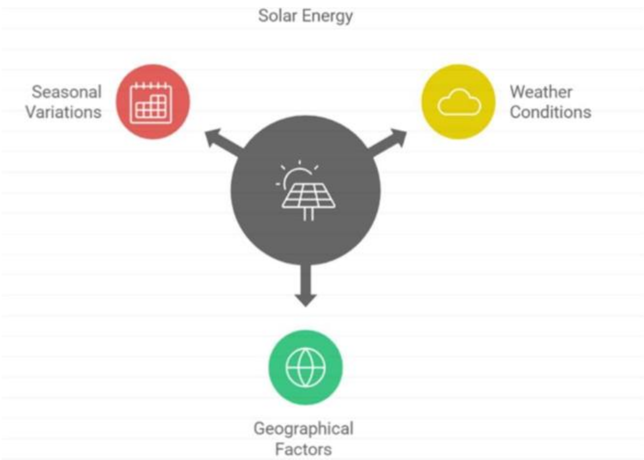


Figure 1.0.6: The relationship of exogenous variables influencing solar power forecasting

Endogenous Variables are the second part of gathering data for more efficient output results, and that depends on the PV system characteristics, panel type and efficiency, system size configuration, inverter specifications and historical performance data for the last power output measurements [66]. Table 1.2 shows that the system itself influences endogenous variables in the context of solar power plant forecasting.

Aspect	Details
Variable Type	Endogenous (Dependent Variable)
Name	Solar Power Output
Unit	Kilowatts (kW) or Megawatts (MW
Time Resolution	Typically hourly, sub-hourly, or daily
Time Resolution	Typically hourly, sub-hourly, or daily
Solar Power Output	The amount of electrical power generated by a photovoltaic (PV) system at a given time. kW or MW
PV System Efficiency	The real-time efficiency of the solar conversion process, often affected by internal system losses. % (percentage
DC Power Output	Power output measured before conversion by the inverter (direct current side). kW or MW
AC Power Output	Power is delivered to the grid or load after conversion by the inverter (alternating current side).
State of Energy Storage	In systems with battery integration, the energy level in storage impacts and is influenced by the forecast kWh
Net Power Exported	The power exported to the grid (after local consumption and losses). kW or MW
Inverter Performance Ratio	The efficiency or effectiveness of the inverter in converting DC to AC power. Unite as % (percentage)

Table 1.2 endogenous variables in solar forecasting

PV power production forecasting is a two-step process. First, air temperature, solar radiation, and other meteorological factors like wind speed and relative humidity are projected. The second phase involves converting the expected variables to power output while taking the power plant's technical parameters into account. When defining a PV system, far more precision and detail can be obtained than when defining a weather system. The uncertainty surrounding the forecast of solar radiation is over ten times greater than the uncertainty around the conversion of solar radiation to PV power generation [67].

The interactions and nonlinearities that may occur between variables have not yet been fully explored in the literature [26], [68]. Still, these are necessary ingredients for building robust prediction models. To create more accurate forecast models, the second stage will concentrate on the metrics and characteristics of redundancy and relevancy, and explore ways to improve and reduce them, respectively. Interestingly, the initial research findings showed that the LSTM performance decreased with the addition of more variables. Many questions are thus brought up, which the current study will attempt to answer [47], [68].

The accuracy of solar power predictions can be increased by using advanced forecasting techniques to account for both system-specific and external influences. Better grid integration and financial benefits from solar energy systems are made possible by exogenous and endogenous variables [69]. PV power production forecasting is a two-step process. The first predicts solar radiation, air temperature, and other meteorological factors like wind speed and relative humidity. The second phase involves converting the expected variables to power output while considering the power plant's technical parameters [50]. the correlations between exogenous and endogenous variables are vital for improving forecasting accuracy. When defining a PV system, far more precision and detail can be obtained than when defining a weather system. The uncertainty around the forecast of solar radiation is over ten times more than the uncertainty surrounding the conversion of solar radiation to PV power generation. As a result, a major determinant of the accuracy of the PV power forecast is the forecasted solar radiation [50], [69].

1.10 Considerations of correlations exist between the exogenous and endogenous variables to improve the output results.

Numerous meteorological factors influence solar irradiation, creating a complicated structure that makes forecasts of it difficult and inaccurate. To simplify the model, the prediction techniques now in use concentrate on examining the relationship between features and radiation; nevertheless, they fail to take into consideration duplicated analysis within the feature subset [47], [68]. To reduce the information redundancy in the feature set and improve prediction accuracy. Precise forecasting is essential for solar energy since it enables sources to be optimised and efficiency raised. Accurately forecasting outcomes is essential to strategic planning and decision-making procedures [32].

Most correlation analyses that have been done in the literature try to determine whether there is a linear relationship between the input features [70]. This study will propose a feature selection method based on correlation analysis for redundancy and relevancy measurements, which helps determine which traits are redundant or unimportant. To find linear correlations between the exogenous variables, redundant features are usually quantified using Pearson's correlation coefficient. One could argue that choosing which redundant variables to use requires more than just understanding linear connections.

The monotonic associations between each external variable and the endogenous variable are calculated to obtain the Spearman rank correlation coefficient, which is used to quantify irrelevant features. The Spearman rank correlation coefficient is used to quantify irrelevant attributes by calculating the monotonic relationships between each exogenous variable and the endogenous variable.

It is discovered that when a univariate LSTM forecasting model is changed to a multivariate forecasting model, the predictive performance is affected and depends on geographical and meteorological aspects [29], [39]. Seasonality must thus be considered in the LSTM model. According to more research, seasonality patterns in meteorological data need to be further investigated using LSTM models, which can generate nonlinearity patterns depending on both exogenous and endogenous variables, to comprehend seasonality shifts [29], [39], [71], [72]. Efficiency-wise, the LSTM model's capacity to learn from nonlinearity behaviour in solar irradiance data with a wide variety of temporal dependencies was demonstrated when it was compared to DL approaches in the field of solar irradiance forecasting [72]. Second, the research revealed that many weather events should make the LSTM model a viable solution for the computational complexity of multivariate prediction. Furthermore, the literature notes that the use of huge datasets, which have been demonstrated to improve decision-making capacities, may improve the accuracy of LSTM-based forecasting. Figure1.6. represents the drop layers of LSTM, and the output results depend on the variables.

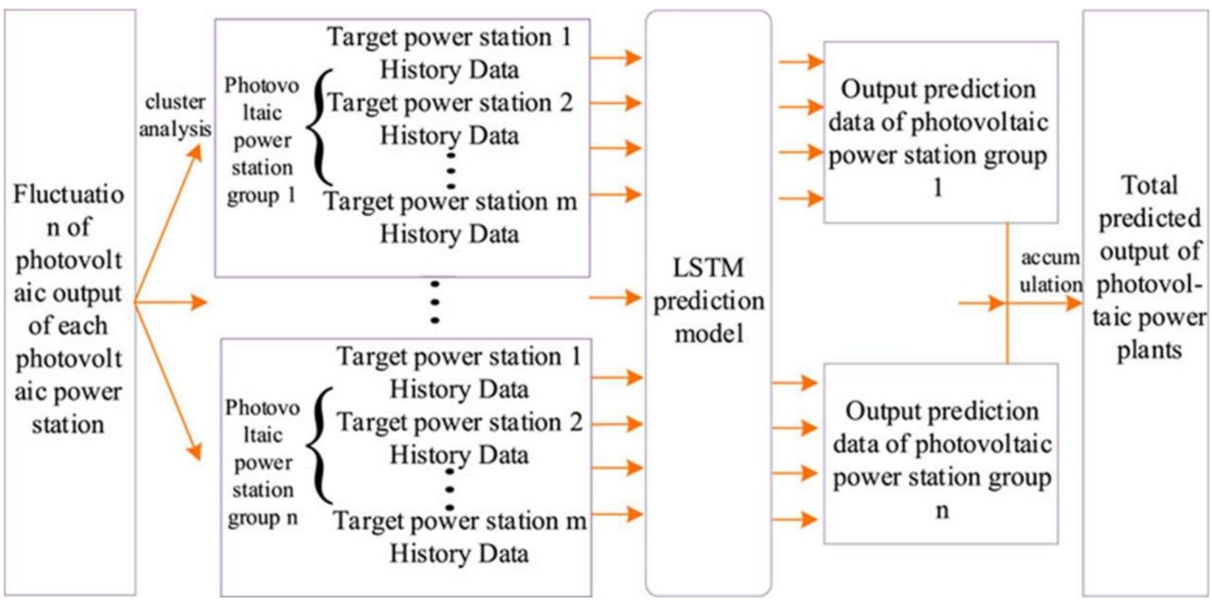


Figure 1 7 input gate, Hayden & drop layers, and the output gate[73]

Data quality problems, however, can affect high-dimensional heterogeneous data. To get around this, data mining can be used to comb through large datasets looking for trends, patterns, anomalies, and correlations. Using a clustering technique such as k-means, etc., data points could be categorised as belonging to sunny, overcast, or wet clusters to reduce uncertainty in solar irradiance estimates [29], [39], [70], [74].

1.11 Technical Solar Power Penetration Forecasting and Challenging

Because solar and wind energy systems depend on the weather, their power generation is very variable. For management and operation strategies to be implemented effectively, accurate prediction data for these variable energy sources is necessary. Physical, regressive, machine learning and time series models are used to forecast solar photovoltaic power. These models can produce probabilistic or point forecasts. Furthermore, methods for gathering data on solar irradiance, cloud movement, and weather forecasts are covered. These methods include satellite imaging, all-sky imaging, on-site sensor network measurements, and numerical weather prediction-based models. a summary of the error measures frequently used to assess the forecast models' performance in solar power forecasting [70], [74]. Solar power forecasting depends on several factors, including:

- Weather conditions include temperature, humidity, precipitation, and cloud cover. sun radiation amounts. Variations by Season and Time of Day.
- Forecasting models can be trained using historical data, which can also be used to spot patterns and trends that could increase forecast accuracy. Among these elements are the date, time, latitude, and longitude of the observer's position on Earth, which are necessary to pinpoint the sun's exact location. By considering these factors, the method provides two essential pieces of information: the zenith angle and the azimuth angle [75].
- Angle of tilt for both location and amount of irradiance. The sun's position concerning the observer's horizontal axis is shown by the azimuth angle. In terms of expression, it is given in degrees, with 0° denoting north, 90° east, 180° south, and 270° west. This angle can be used to align solar panels or solar tracking systems for maximum exposure to sunlight [76].

- Forecasting models often use a combination of these factors and employ various techniques such as statistical methods, machine learning algorithms, and numerical weather prediction models to provide accurate solar power forecasts.

The solar panel recent focus on artificial intelligence (AI) approaches, such as machine learning (ML) and deep learning (DL), has improved the efficiency of performance in several disciplines and increased the application of these techniques to solve various challenges. Artificial intelligence (AI) techniques have been primarily employed in the PV field to address the efficiency and reliability of the entire system. Power operations will be alerted by the energy derived from solar irradiation [67]. For the operation of the electrical system, solar energy penetration and solar power forecasting will become critical [77]. The process of forecasting solar radiation is difficult, and the generation of solar electricity poses distinct difficulties for the grid's transmission and distribution networks. Centralised solar plants are a non-dispatchable part of the generation pool that provides solar power on the transmission side. Solar electricity is produced on the distribution side by many dispersed panels that are mounted on building rooftops, which modifies the load. Electric grid managers must constantly balance supply and demand to ensure the grid's stability therefore, forecast errors related to solar power generation can create significant financial losses and reliability problems. A sophisticated statistical approach based on artificial intelligence approaches for forecasting solar power [78]. The main challenge facing solar power prediction is the data set daily fluctuations, making accurate predictions difficult, grid integration, energy storage, land & cost and technological limitations in photovoltaic cell efficiency, figure 1.7. illustrate the importance of PV power forecasting in various types of grid connection PV system.

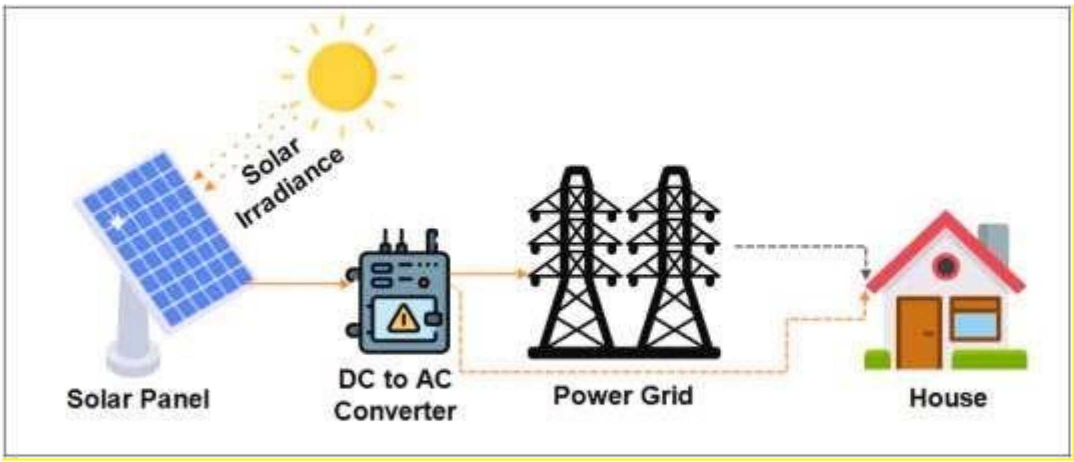


Figure 1.0.7 Grid connecting solar power plant [78]

Monitoring and maintenance for all components of the solar system depends on utilising different forecasting algorithms by using machine learning.

1.6 Publications Associated with the Thesis

The work explained in this chapter contains materials that were published at the following conferences and journal publications that are associated with this thesis:

- Design a model of ensemble LSTM to close the uncertainty prediction of solar power output
- Deep learning MPP tracking based on perturbing & observation with partial shading conditions.
- Close uncertainty of solar power production using deep machine learning

1.7 Thesis Organisation

To lay a solid theoretical foundation, the next chapters:

- Chapter 2: describes the basic principles of machine learning and photovoltaics as well as the research that has already been done on solar irradiance forecasting systems. Moreover, the improvements made at each step of the feature and technique selection process, together with the suggestions made for the use of solar irradiance predictions.
- Chapter 3: outlines the study process, provides sources and data gathering details, and conducts a thorough exploratory data analysis. outlines the sequence of actions required to finish this project, such as gathering and preparing data, designing and developing LSTM models, and creating an appropriate experiment structure.
- Chapter 4: the behaviour of long short-term memory (LSTM) is examined concerning the input feature construction. The design and implementation of the SVO-LSTM and ED-LSTM models are presented, and their outcomes are discussed. Additionally, confirmatory data analysis is carried out for validation.
- Chapter 5: concludes with a detailed discussion of the findings and relevant research, as well as its limits and future work directions.
- Chapter 6: conclusion and future works recommendation

Chapter 2
Literature Review

2.1 Introduction

The broad use of solar power systems is hampered by certain issues. Overcoming solar power's non-dispatchable nature would be one such issue [75], [79] Unexpected weather phenomena can have a substantial impact on power systems that rely heavily on solar power for generation, as [80], highlights. Quick variations in the weather, such as more clouds covering the sky, could significantly reduce the power production of PV systems. In the absence of appropriate energy dispatching techniques, such as battery energy storage utilised in conjunction with grid-connected PV systems, unanticipated weather changes could render electrical grids unstable [76]. Furthermore, Chen, Kwan, and Tan [81] contend that sophisticated information processing techniques could be applied to estimate solar power generation accurately, lowering the amount of battery capacity needed for PV systems that are battery companioned.

One of the most popular alternative energy sources is solar power. However, there is a great deal of unpredictability in solar irradiation, which leads to significant fluctuations in the generation of wind and solar electricity. It is challenging to integrate solar energy technologies into electrical networks [70]. For the safe and dependable functioning of electrical systems, the management of solar power plants, and the provision of high-quality electric power at the lowest feasible cost, more precise solar irradiation and one-day-ahead forecasting is therefore essential. Because the patterns of sun irradiation per month throughout the year are similar, clouds have an impact on solar irradiation forecasting and data categorisation for subsequent years. A subset of machine learning (ML) called deep learning is motivated by the architecture of the human brain. DL attempts to emulate the functions of human brain neurons by imitating them [81].

The goal is to reduce the scope of the research that has been done in terms of forecasting horizon and forecasting approaches, while also exploring the key elements of the body of existing literature. To understand the stochastic connection between the past and the future, DL algorithms employ historical data. Numerous studies have indicated that in solar irradiance predictions, these models perform better than statistical models. Thus, the focus of my research is on using DL approaches for modelling and forecasting solar irradiance one day ahead of schedule. It is challenging to compare DL forecasting methodologies to existing models because of the differences in the methods by which they have been examined, as well as the various assumptions and variety of inputs involved [39]. Numerous studies have indicated that

in solar irradiance predictions, these models perform better than statistical models. Thus, the focus of my research is on using DL approaches for modelling and forecasting solar irradiance one day ahead of schedule. It is challenging to compare DL forecasting methodologies to existing models because of the differences in the methods by which they have been examined, as well as the various assumptions and variety of inputs involved [29], [82]. These usually make use of the weather, data quality, and geographic location as references. Because of this, they may be simply divided into two groups: statistical and traditional machine learning, or physical approaches like ANN, ARMA, ARIMA, FL, and NWP (numerical weather prediction). Each of these groups uses a different forecasting strategy, as shown in Figure 2.1. Weather forecasting aids in the prediction of power generation from PV solar plants, as radiation and temperature are the primary determinants of power generation in PV systems [[82]. Predicting solar energy can help increase the stability of a power system by providing an approximation of the amount of solar power that can be generated at a given place in the future [70]. There are various techniques for forecasting solar power, and machine learning neural networks are one of them. In this study, machine learning techniques are used to estimate the power generation from a solar plant by projecting future weather conditions [72]. The efficiency of converting solar energy into electricity is significantly impacted by accurate solar photovoltaic (PV) system model parameters. Swarm and evolutionary optimisation algorithms have been frequently used in this regard to solve real-world issues because of their simpler conceptual frameworks, effectiveness, adaptability, and ease of implementation [[39], [72]
Numerous GHI forecasting methodologies have been put out in the pertinent literature. As references, these usually consider the weather, data quality, and geographic location. Because of this, they are readily divided into two categories: statistical and classical machine learning, or physical approaches like ANN, ARMA, ARIMA, FL, and NWP (numerical weather prediction), each of which uses a different forecasting strategy, as illustrated in Figure 2.1.

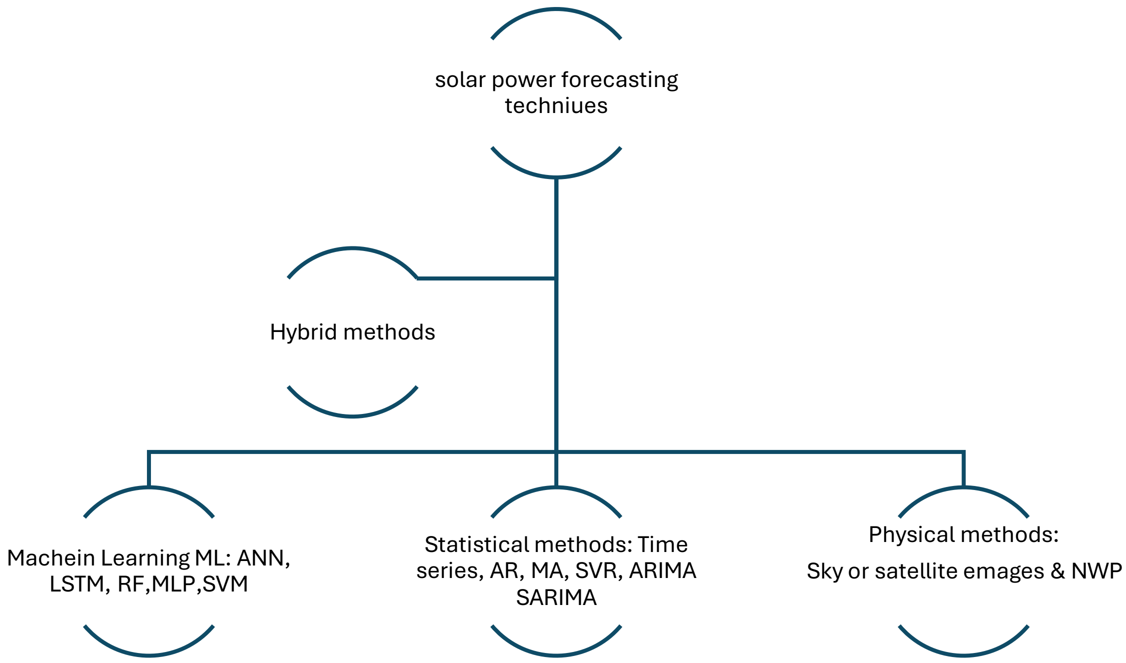


Figure 2.1 Forecasting Techniques

2.2 Deep learning models to enhance the performance of solar irradiance prediction

Enhancing the performance of solar irradiance prediction using Deep Learning DL and data mining techniques has been the focus of several recent studies. These studies explore various deep-learning models and data mining techniques to improve the accuracy and reliability of solar irradiance forecasting. Although CNNs are mostly used for image processing, they have been modified to forecast time-series data, such as solar irradiance. They record the relationships between space and time in meteorological data, including cloud cover and atmospheric conditions [83]. Studies by [83] established that by successfully identifying pertinent features from input data, CNNs can beat standard models in short-term solar irradiance forecasts.

In terms of their working principles, advantages, and disadvantages, convolutional neural network (CNN), gated recurrent unit (GRU), recurrent neural network (RNN), deep neural network (DNN), and a hybrid model (CNN–LSTM) have been utilised recently to forecast solar irradiance [84], [85].

In terms of their working principles, advantages, and disadvantages, convolutional neural network (CNN), gated recurrent unit (GRU), recurrent neural network (RNN), deep neural

network (DNN), and a hybrid model (CNN–LSTM) have been utilised recently to forecast solar irradiance [85].

Based on their simulation results, the low root mean square error of LSTM indicates that it works better than the other standalone models (RMSE). Nevertheless, by including LSTM in a hybrid model (i.e., CNN–LSTM), the accuracy of solar irradiance forecasts can be further enhanced; this hybrid model outperforms independent models in terms of error reduction. It was suggested by [85] to contrast CNN-LSTM with other hybrid models. In [86]. When compared to traditional methods, solar power generation prediction accuracy can be enhanced by 27% using Deep Learning techniques like help vector machine. In [87] grid-connected solar systems, a deep learning model can optimise energy management by lowering energy prediction errors by more than 0.3 per cent. Recommended by [88] Their study of the "Deep Learning Model on Energy Management in Grid-Connected Solar Systems". In [89], although deep learning techniques perform better with hybrid networks in wind and solar energy prediction in terms of accuracy, robustness, and generalisation ability, it is preferable to focus on using and developing hybrid DL techniques with other optimisation techniques to enhance and optimise their structure.

Below is a summary of key research findings and methodologies of Deep Learning Models for Solar Irradiance Forecasting. Recurrent Neural Networks (RNNs), This model integrates an RNN with an adaptive neural network model (ANNM) imputation module to handle missing data, which is crucial for accurate predictions. The model shows superior performance in low-light conditions and when dealing with incomplete datasets [90]. Table 2.2 illustrates that CNN's deep learning methodology depends on dataset collection, data selection variables, choosing the proper model, testing, and training data output prediction.

Conventional neural network CNNs are particularly effective in processing grid-like data, such as images. In the context of solar irradiance prediction, CNNs can analyse satellite imagery to extract features related to cloud cover and atmospheric conditions. By training on historical irradiance data alongside corresponding satellite images, CNNs can learn to predict future solar irradiance levels with high precision.

Model Type	**Key Features**	**Strengths**	**Reported Accuracy (R^2)**

CNN	Feature extraction from time-series data	High accuracy, efficient, good for short-term	0.930–0.971
CNN-LSTM (Hybrid)	Combines CNN (spatial) + LSTM (temporal)	Best for sequential dependencies, robust	Up to 0.99925
TCNN/HTCNN	Dilated, causal convolutions, long-range memory	Outperforms RNNs, handles long sequences	Higher than LSTM/GRU
DNN	Deep, fully connected layers	Good, but less effective than CNN/CNN-LSTM	Lower than CNN

Table 2.1: Deep learning model for enhancing accurate prediction.

Numerous important techniques, including statistical, physical, machine learning, and ensemble methods, have been used in predicting investigations [23], [91]. For solar energy forecasting, such as PV power generation forecasts, more research into spatial and temporal correlation is necessary [92], [93]. The input parameters and forecast horizons affect how well any technique performs. Large-scale datasets are necessary for forecasting analytical characteristics such as spatial-temporal correlation, which improves accuracy by referring to space and time, respectively [94], [95]. Long short-term memory (LSTM) strategies have been shown in numerous studies to increase the forecasting accuracy of time-series statistical methodologies.[96]. LSTMs are a type of recurrent neural network (RNN) that can remember information for much longer periods refer to deep learning DL analyses of the commonly utilized RNN models for time series data predictions, making them suitable for photovoltaic solar power production forecasting problems. Recently, LSTMs have also been widely adopted for short-term wind speed predictions [97].

2.2.1 Discussion

Deep learning models, particularly LSTM and CNN-LSTM, are more accurate than conventional machine learning models in predicting solar irradiance and PV power. The most significant finding is that the deep learning models of interest are more suitable for predicting solar irradiance and PV power than other conventional machine learning models. Additionally, we recommend using the relative RMSE as the representative evaluation metric to facilitate accuracy comparison between studies. The classification of solar power forecasting in numerous research projects has been carried out to create forecasting models that are suitable for precisely predicting the electricity generation of solar photovoltaic systems while minimising complexity and expense. Predicting PV output power typically requires three steps. The initial step is to extract the energy characteristics and examine the influences on them, selecting the prediction technique and refining the forecasting model constitutes the second step. PV output power forecasting techniques can be broadly classified into three categories: physical, statistical and hybrid, based on the forecasting method used. Predicting PV output power typically requires three steps. The initial step is to extract the energy characteristics and examine the influences on them. Selecting the prediction technique and refining the forecasting model constitute the second step.

2.3.1 Multi-layer Perception (MLP) Model-based Forecasting of Solar Irradiance

Multi-layer Perceptron (MLP) is an effective technique for forecasting solar irradiance, with studies showing it can achieve high accuracy in predicting daily global solar irradiance [98]. MLP has been shown to outperform other deep learning techniques, such as Long-Short Term Memory (LSTM) and Gated Recurrent Unit (GRU) in terms of computational efficiency, with MLP using significantly fewer parameters [99]. Time series analysis using MLP has also been effective in mapping temporal patterns in solar irradiance data, including hourly, monthly, yearly, and seasonal variations [100]. Embedding temporal information, such as timestamp data, into the MLP model has been found to significantly improve the accuracy of solar irradiance forecasting compared to using only historical irradiance values [101].

According to [72], the early model used a Back-Propagation Neural Network (BPNN) to forecast solar radiation, depending on the available meteorological data and other environmental factors. Four locations in the southeast United States provided a total of twenty-three years' worth of weather data, which were divided into two sets: an 11-year training set and a 12-year testing set. These meteorological data comprised day length, daily cloudless sky radiation, precipitation, and minimum air temperature. For these 12 years, the observed daily sun radiation values were compared with the anticipated daily solar radiation values.

To accurately forecast solar power, [96], used machine learning algorithms to present intriguing solutions that will help ensure that renewable energy resources are used efficiently. According to [26], deep learning techniques like deep recurrent neural networks, deep belief networks, and stack auto-encoders have the potential to increase the accuracy of renewable energy forecasts. To increase the accuracy of solar irradiance forecasts, a variety of approaches, including hybrids of physical models, statistical approaches, and artificial intelligence systems, have been developed [75]. In contrast to all other models that were evaluated, the author of [22]. The proposed DNN-LSTM model significantly improved predicting accuracy and obtained the lowest validation values for MSE and MAE. To increase the accuracy of renewable energy forecasts, a variety of approaches, including hybrids of physical models, statistical approaches, and artificial intelligence algorithms, have been developed [102], [103]. The primary topic of the study article [81]. Solar energy forecasting through the application of multiple methodologies that might be utilised for the prediction of global solar radiation (GSR). The forecasting of solar resources has shifted from traditional mathematical methods to the application of intelligent techniques in the realm of artificial intelligence (AI). The study [81]. investigates the efficacy of fuzzy time collection in hyperspectral photo classification. It aims to elucidate intricate temporal patterns in the optimisation process by identifying the optimal FTS parameters that produce superior type accuracy [82].

A BPNN for modelling monthly mean daily values of global solar radiation from 41 data collection stations in the Kingdom of Saudi Arabia was proposed in another study on BPNN by [79]. The stations were split into 31 neural network training locations and 10 testing locations [104]. To predict solar radiation values, the suggested model made use of latitude, longitude, altitude, and the length of sunshine. In terms of the Mean Absolute Percentage Errors (MAPE), which are approximately 19.1 percent, the suggested model's results showed quite superior performance between the observed and projected values. Additionally, research on BPNN has been expanded by [105]. They calculated the values of global radiation for several Sultanate of Oman locations. From 1987 to 1992, meteorological information was gathered from six weather stations located within the Sultanate of Oman. The location, month, mean temperature, mean pressure, mean relative humidity, mean wind speed, mean sunshine duration, and mean evaporation are among the eight input features that were input into the BPNN. According to the findings, the suggested BPNN-based model can determine the global radiation value for the provided dataset with a 93 percent accuracy rate. They created a model to estimate global radiation based on historical data from the previous 12 months for a location in the Sultanate of Oman that has global radiation measurement equipment to further assess the

generalisation capability of the suggested model. The model achieved a prediction accuracy of 95%. After that, the model was applied to forecast global radiation data for another Sultanate of Oman location without direct global radiation measuring equipment. Even while the LMNN model seemed to be helpful, more studies on BPNN produced inconsistent findings [106]. For instance [26]. describes DL methods for solar radiation estimation that start with estimating the clearness index, which is the product of the daily maximum radiation and the average daily sun radiation. Using long-term data from eight sites in Oman over ten years (1986–1998), the RBFNN and BPNN models were examined. Latitude, longitude, altitude, sunshine ratio, and month of the year were the input parameters. The clearness index is the output parameter [107]. While earlier studies have provided some evidence for the utility of BPNN, subsequent research by. included a comparison of BPNN with the Adaptive Network Fuzzy Inference System (ANFIS) and Levenberg Marquardt Neural Network (LMNN),
Recurrent Elman Neural Network (RENN), and Radial Basis Function Neural Network (RBFNN) for the forecasting of mean hourly global solar radiation [108]. Mean hourly solar radiation values on a horizontal level, obtained on the French island of Corsica, in W/m2, were the data used throughout the investigation [109]. When it comes to predicting hourly global solar radiation, the LMNN model performs better than other methods when compared in terms of RMSE and training time [108], [109].

2.3.2 Discussion of using MLP techniques

Multi-layer Perceptron (MLP) has emerged as a highly effective tool for forecasting solar irradiance. Studies have demonstrated its ability to predict daily global solar irradiance with remarkable accuracy, making it a strong candidate for renewable energy forecasting applications. Compared to other deep learning models like Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU), MLP offers superior computational efficiency. This is mainly because MLP models typically require fewer parameters, reducing computational complexity without sacrificing performance. Time series analysis using MLP has proven particularly effective in identifying and mapping temporal patterns in solar irradiance data across various time scales, including hourly, monthly, yearly, and seasonal variations. By embedding temporal information, such as timestamps, into the MLP model, its accuracy in forecasting solar irradiance improves further, enabling more precise predictions of solar energy availability. This makes MLP a valuable tool for optimising the integration of solar power into energy systems and enhancing overall energy management.

2.4.1 Practical Swarm Optimisation PSO and Bach Progression Neural Network BPNN for more accurate solar power forecasting

The Particle Swarm Optimisation (PSO) model is a computational technique inspired by the social behaviour of birds flocking or fish schooling. It is used to optimise problems by iteratively improving a candidate solution based on its previous experiences and those of its neighbours. In the context of predicting daily diffuse solar radiation, the PSO algorithm optimises the parameters of a predictive model, such as a regression or neural network, to minimise prediction errors (like root mean square error or RMSE).

Seven characteristics are chosen as the evaluation indices in this study: month of the year, sunshine length, mean temperature, rainfall, wind speed, relative humidity, and daily global solar radiation [110]. Compared to the BPNN and BPNN optimised by GA models, the BPNN optimised by the PSO model forecasts daily diffuse solar radiation more correctly and [111]some studies depend on PSO to identify optimal parameters for solar radiation models, making it more efficient than traditional methods like grid search or manual tuning, effectively trains an artificial neural network (PSO-ANN) to estimate global solar radiation at locations without direct measurement stations, outperforming the standard backpropagation algorithm [112], [113], [114], [115]. Figure 2.3 shows the flowchart of PSO [116].

To increase the backpropagation neural network (BPNN) model's efficiency and capacity for generalisation in the prediction of daily diffuse solar radiation, the paper provides two optimisation techniques: genetic algorithms (GA) and particle swarm optimisation (PSO). Two models, BPNN and BPNN optimised by GA, were compared with the predictions made by the BPNN optimised by the PSO model. The findings indicate that there is potential for the suggested BPNN optimised by the PSO model to estimate daily diffuse sun energy accurately [102], [113], [114]. Using only daily maximum and lowest air temperature and precipitation data, a neural network model can more precisely estimate daily solar radiation, which improves crop growth and development predictions [105]. According to study results, the neural network model was created and tested for a small number of sites; nevertheless, when measurements of only daily maximum and minimum air temperature and precipitation are available, the model may be used to estimate daily solar radiation [117].

The study uses Artificial Neural Network-Based Diffuse Solar Radiation Prediction for Each Day [118]. The PSO model-optimised BPNN forecasts daily diffuse solar radiation more effectively than the genetic algorithm GA-optimised BPNN and BPNN models. To increase the backpropagation neural network (BPNN) model's efficiency and capacity for generalisation in the prediction of daily diffuse solar radiation, the paper provides two optimisation techniques:

genetic algorithms (GA) and particle swarm optimisations (PSO) [119]. The evaluation indices are based on seven parameters: the month of the year, the amount of sunshine, the mean temperature, the amount of rainfall, the wind speed, the relative humidity, and the daily global solar radiation. Two models, BPNN and BPNN optimised by GA, were compared with the predictions made by the BPNN optimised by the PSO model. The findings indicate that there is potential for the suggested BPNN optimised by the PSO model to estimate daily diffuse sun energy accurately [120].

In the case study of [121], Prediction of horizontal diffuse solar radiation using clearness index-based empirical models, to produce solar hydrogen, the diffuse coefficient model with linear form offers the most accurate monthly mean daily prediction for diffuse solar radiation in Kerman, Iran. The diffuse fraction model in linear form is the best fit for the daily-based prediction, with a root mean square error (RMSE) and correlation coefficient (R) of 1.3081 MJ/m2 and 0.8767, respectively. The diffuse coefficient model, also in linear form, offers the highest accuracy for the monthly mean daily prediction, with RMSE and R of 0.5391 MJ/m2 and 0.9258, respectively. Since the clearness index can be determined using just global solar radiation data, it is convenient to use these models exclusively for solar system applications, especially solar hydrogen production. This study has been cited by [92], [122], [123]. Figure 2.3 shows the structure of ANN to predict GSR.

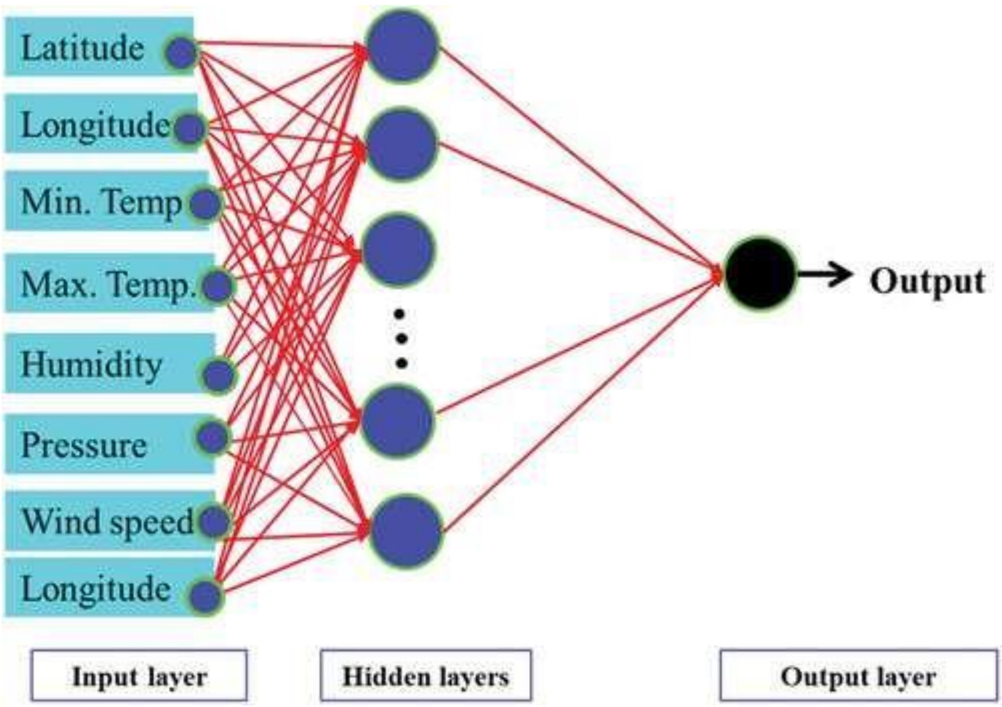


Figure 2.3: Structure of ANN to Predict GSR [124]

connecting multiple variables of linear regression in large volumes of unreliable data. Less computing work is needed for ANNs to create a link between input parameters and targets. A summary of numerous studies on creating solar irradiance forecasts using MLP models by previous studies with various methodologies and different outcomes measured, as shown in Table 2.2.

Ref.	Paper	Abstract	Main findings	Methodology	Outcome measured	Duration /region	Model
[125]	LSTM and RBFNN Based Univariate and Multivariate Forecasting of Day-ahead Solar Irradiance for the Atlantic Region in Canada and the Mediterranean Region in Libya	The proposed univariate time series forecasting method performs better than other models about minimum forecasting errors	The UTSF_LSTM model outperformed other models in terms of minimum forecasting errors. based on R2 and RMSE metrics.	exploring the incorporation of meteorological and geographical variables along with historical data sets using deep learning models. The study compared the performance of the LSTM model to the Radial Basis Function neural network (RBFNN) for both multivariate and univariate time series forecasting, with the UTSF_LSTM method showing superior results in terms of R2 and RMSE.	Performance of forecasting models in terms of coefficient of determination (R2) and Root Mean Square Error (RMSE)	1hour-24 hrs/Canada, Libia	LSTM and RBFNN
[126]	Day-ahead forecasting of solar irradiance using a hybrid improved Cuckoo Search-LSTM approach	The proposed model provides a better dayahead forecasting of global horizontal irradiance.	The novel forecasting approach using LSTM networks and an improved variant of the Cuckoo search algorithm outperformed other models in day-ahead forecasting of global horizontal irradiance.	The methodology involved using historical meteorological data as input factors, combining LSTM networks with an improved Cuckoo search algorithm, and evaluating model performance with RMSE and MAE.	Accuracy of day-ahead forecasting of global horizontal irradiance using RMSE and MAE as statistical parameters	Day-ahead/ California	LSTM, BPNN
[127]	Day-Ahead Hourly Solar Irradiance Forecasting Based on Mult Attributed Spatiotemporal Graph Convolutional Network	The spatial adjacency of the stations, temporal changes in the meteorological variables, and a variety of variables to the forecasting performance are synergistic.	The proposed model outperformed existing models in various evaluation metrics, showed higher and more stable performances across different cloudiness levels and months, and exhibited the best performance when using all 15 variables.	The methodology collects solar irradiance and meteorological variables from ASOS stations, modelling the data as a dynamic attribute network, and developing a forecasting model based on spatiotemporal correlations.	None	Day-Ahead/India	Spatiotemporal Graph Convolutional Network

[128]	Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM	The proposed algorithm outperforms these competitive algorithms for single-output prediction.	using weather forecasting data, which outperforms competitive algorithms and shows significant improvements in accuracy, with the LSTM algorithm demonstrating a substantial decrease in prediction RMSE compared to BPNN.	The methodology involves using weather forecasting data for solar irradiance prediction, structuring the prediction problem to jointly predict multiple outputs, training the prediction model with LSTM networks, and comparing different algorithms for solar irradiance prediction.	accuracy of solar irradiance prediction in terms of root mean square error (RMSE) compared to other algorithms, percentage improvement in RMSE using the proposed LSTM algorithm compared to BPNN	Hourly dayahead solar irradiance/ Saudi Arabia	LSTM, BPNN
[129]	Day-Ahead Solar Irradiance Forecasting Model	A multilayer perceptron model consists of 5 inputs, 64 neurons in a hidden layer with 1 neuron in the output layer.	The model using a multilayer perceptron architecture with specific input variables included total cloudiness, air temperature, relative humidity, atmospheric pressure, and the cosine of the solar incidence angle.	involved developing a day-ahead solar irradiance forecasting model using artificial neural networks with a multilayer perceptron architecture, training the model with ground measurements of solar radiation and meteorological data, optimizing input data combination and hyperparameters, and utilizing specific input variables for predicting solar irradiance.	The main outcome measured in the study is the hourly average solar irradiance on a tilted surface.	Day-Ahead/ France	BPNN
[130]	Day-ahead to week-ahead solar irradiance prediction using convolutional long short-term memory networks	The proposed framework can be trained using a small amount of training data within the duration of only two months.	The proposed mechanism for solar irradiance prediction can be trained with a small amount of data in a brief time frame, making it suitable for initial deployment when data is limited. The system was validated using a challenging dataset from Taiwan with a tropical and subtropical marine island climate.	The methodology involves using a convolutional Long Short-Term Memory (LSTM) model for solar irradiance prediction, extracting features from hourly data using convolutional filters, concatenating features for input into the LSTM network, and combining the LSTM output with original data and utilising a fully connected layer for final predictions. The model can be trained with limited data in a brief period.	accuracy of solar irradiance predictions	Dayahead/Oman	LSTM, BPNN

[131]	Univariate model for hour ahead multistep solar irradiance forecasting	A feature selection algorithm selects the highly correlated lag variables to reduce the complexity of the model.	The study focuses on forecasting GHI using time-series decomposition techniques and machine learning algorithms, optimising parameters for each IMF and evaluating performance across annual and seasonal variations.	The methodology involves comparing two time-series decomposition techniques (EMD and EEMD) for forecasting GHI, transforming the problem into a supervised learning task, using a feature selection algorithm to reduce complexity, and applying machine learning algorithms (SVM and RF) for forecasting.	forecasted Global Horizontal Irradiance (GHI) values	1 hour/ Greece	BPNN, SVM

Table 2.2: Long Short-Term Memory (LSTM) and Radial Basis Function Neural(RBFNN) Based Univariate and Multivariate Forecasting of Day-ahead Solar Irradiance.

2.4.2 Discussion

These studies focus on forecasting daily diffuse solar radiation using Artificial Neural Network (ANN) models, particularly optimising the Backpropagation Neural Network (BPNN) with Particle Swarm Optimisation (PSO) and Genetic Algorithms (GA). The PSO-optimised BPNN outperforms the GA-optimised BPNN and the standard BPNN models in predicting daily diffuse solar radiation. The efficiency and generalisation capability of the BPNN are enhanced using these optimisation techniques. Seven parameters were used to evaluate the model: month of the year, sunshine duration, mean temperature, rainfall, wind speed, relative humidity, and daily global solar radiation.

In comparing models, the PSO-optimised BPNN showed better accuracy, indicating its strong potential for reliable diffuse solar radiation forecasting. Additionally, the case study on predicting horizontal diffuse solar radiation using clearness index-based empirical models demonstrated that the linear form of the diffuse coefficient model produced the most accurate monthly and daily predictions in Kerman, Iran, making it particularly useful for solar hydrogen production. These models are convenient due to their reliance on global solar radiation data alone, making them suitable for solar energy applications.

2.5.1 Solar power prediction using deep learning with LSTM techniques based on Back-Propagation Neural Network. BPNN is a hybrid method

Outperforming current techniques, the suggested deep learning model utilising BPNNLSTM efficiently maximises solar panel output power under a range of load circumstances. In the study [132]. The author creates a deep learning model that assists in obtaining the maximum power point by utilising a back propagation neural network (BPNN), when the solar panels are coupled to a boost converter and the load is changeable, the deep learning model seeks to maximise the output power from the solar grids. By predicting the reference voltage under varied weather circumstances, BPNN-LSTM makes it possible to separate different output powers and guarantee maximum power while maintaining a steady output voltage. The suggested BPNN-DL is put to the test in various scenarios to gauge how resilient the modules are against interference from the outside and within. The simulation's findings demonstrate that, when compared to current approaches for regression analysis on training, testing, and validation, the suggested method extracts the highest output power from each panel. Future operations of photovoltaic power plants can operate more efficiently because of the superior short-term solar power predictions provided by the LSTM network. To sustain a balanced and all-encompassing operation, it will be crucial to create models that enable accurate short-term forecasting of solar PV generation. An analysis is conducted to see how well a deep learning method based on the Long Short-Term Memory (LSTM) algorithm can forecast solar power data. The Multi-layer Perceptron (MLP) network and the LSTM network were evaluated for performance using the following metrics: Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Squared Error (RMSE), and Coefficient of Determination (R2). The [133]. Propose an LSTM development model. The prediction outcome demonstrates that the LSTM network provides the best outcomes for an everyday category. As a result, it offers trustworthy information that makes future photovoltaic power plant operations more effective. In an LSTM (Long Short-Term Memory) network, the feed-forward (FF) process refers to how input data is passed through the network to produce an output. At each time step, the LSTM cell receives the current input and the previous hidden and cell states. The cell uses internal gates (input, forget, and output gates) to control the flow of information, deciding what to keep, update, or discard. The output from each cell at each time step is then passed forward to the next time step, and ultimately to the output layer if present

This process is like other neural networks in that data moves forward through layers (or time steps, in the case of sequences), but LSTMs are unique in their ability to maintain and update a memory cell, allowing them to capture long-term dependencies in sequential data. The backpropagation process in LSTM is an extension of the standard backpropagation algorithm, adapted for the recurrent structure of LSTMs, as shown in Figure 2.4.

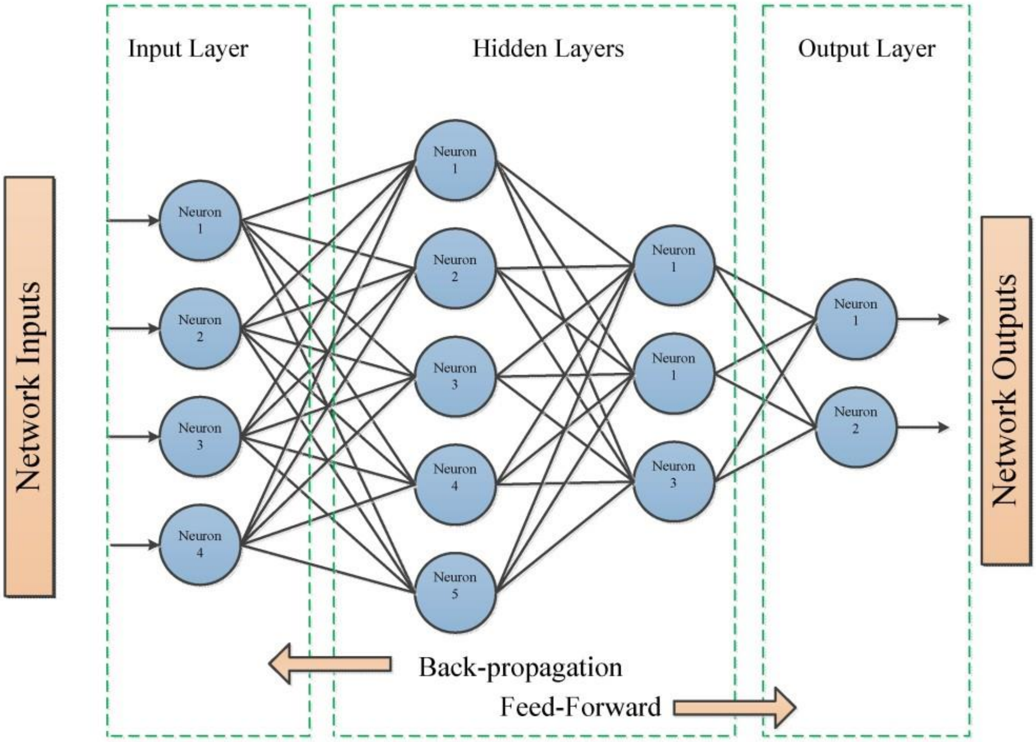


Figure 2.4: Backpropagation and feed forward in Neural Network [133].

Deep learning and energy efficiency together constitute a combination that has a bright future, particularly in terms of advancing energy sustainability, decarbonization, and the digitisation of the electrical industry [133]. When it comes to long-term wind speed prediction, the LSTM approach in [134], outperforms ARIMA, providing a more precise and effective way to integrate renewable energy. Developments in deep learning techniques provide the prospect of creating a multistep prediction model that is more sophisticated than shallow neural networks (SNNs) [134]. By comparing the common autoregressive integrated moving average model with artificial neural networks (ANNs), recurrent neural networks (RNNs), and long short-term memory (LSTM), a special type of RNN model in [134], seeks to determine the most effective predictive model for time series, with fewer errors and higher accuracy in the predictions (ARIMA).

The root mean square error (RMSE) approach is used to measure the outcomes. The comparison's outcome demonstrates that LSTM outperforms ARIMA in terms of accuracy. Long-short-term memory when paired with deep neural networks effectively increases the accuracy of wind speed prediction in wind power prediction for electricity generation, as demonstrated in the Application of Long-Short-Term-Memory Recurrent Neural Networks to Forecast Wind Speed [134]. To solve the issue of gradient information disappearing or exploding during the training process that the recurrent neural network (RNN) encountered when studying time series, LSTM was employed. The outcomes demonstrated that the suggested model's use can significantly raise wind speed forecast accuracy. Table 2.2 shows the different optimisation algorithms using LSTM for more accurate results.

	Paper	Model using	Outcome
[135]	Solar Power Forecasting Using Machine Learning Learning Techniques 2024	Pre-processing methods: feature removal, data imbalance handling using SMOTE and Machine learning regression models: linear, elastic net, random forest	Achieved R2 score of 0.964/ Best-performing a regression model: random forest
[136]	IoT -Based on solar power forecasting using AMA-ELM techniques 2024	LSTM and variational mode decomposition	Essential grid management and distribution network planning
[137]	Forecasting the wind and solar power using deep learning-based techniques	Feedforward neural network (FFNN) & LSTM	Enhance power network reliability and optimisation forecasting
[138]	The Solar Energy Forecasting Using LSTM Deep Learning Technique	Deep learning Techniques &LSTM	RMSE & MSE used for evaluation. Atmospheric factors affect the prediction accuracy
[139]	Medium-term Solar Power Prediction using a stacked-based deep learning Technique	Stacked LSTM with dropout - Optimized kernel Extreme Learning Machine	The proposed SLSTM-OKELM model outperforms other techniques in solar power prediction. - MAPE values of 1.54%, 1.56%, and 1.34% achieved for different seasons

[140]	Solar Power Forecasting Using Deep Learning Techniques	- Deep learning technique based on LSTM algorithm - Comparison with Multi-layer Perceptron (MLP) network	LSTM network outperformed the MLP network in solar power forecasting. - Promotes energy sustainability, decarburization, and digitization in the electricity sector
[80]	PV performance using a modified back propagation neural network based on the different solar power	- Modified Back Propagation Neural Network architecture - Investigation of module temperature, solar irradiance, and AC power	The polycrystalline panel performed best among monocrystalline and thin film. - Modified BPNN used for accurate short-term power prediction.

Table 2.2: LSTM model with different optimisation algorithms for more accurate output

2.5.2 Ensemble Machine Learning Methods

Ensemble approaches have become a potent tool for raising solar power prediction accuracy [141]. Solar energy is stochastic, the power generated by photovoltaic systems varies significantly within short periods. PV forecast models correlate relevant weather variables like solar insolation, wind speed, relative humidity, ambient temperature, etc [142] to comprehend and control these temporal fluctuations in the generation, the system output is crucial [143]. These models may not fully capture the effects of site-dependent parameters. In the case of solar PV power plants with sufficient operational data, machine learning (ML) techniques can effectively correlate weather parameters with PV system performance [144].

Machine learning ensemble methods are successfully used to model energy resources and power systems. ML methods based on non-linear regression, such as support vector machine (SVM), k-nearest neighbour (k-NN), Markov chain, etc., have been presented with consideration to the application of predicting solar PV power production [145]. These models exhibit moderate accuracy when used for solar PV power estimates due to their relative simplicity and low computing burden [146]. Artificial neural networks (ANNs), one type of deep learning algorithm, are frequently utilised for prediction schemes to improve the model's performance further [147].

Solar radiation is the most crucial factor in the design and sizing of solar power systems. However, weather circumstances, which are unpredictable and changeable, have a substantial impact on solar radiation. Therefore, integrating solar energy resources into the electric grid can be aided by anticipating global sun radiation [148]. In [149], the author offers a novel hybrid method for predicting global solar radiation one hour in advance based on the seasonal clustering algorithm and artificial neural networks (ANNs). To do this, three years' worth of monthly average experimental data were clustered into distinct seasons using the fuzzy c-means method (FCM), considering the solar and climatic factors. In [57]. Hourly meteorological data for Évora City have been used for forecasting. Based on statistical indicators, the hybrid strategy outperforms the solo ANN model in the findings. the [150] research proposed a web-based optimal prediction system It uses machine learning approaches to estimate sun radiation based on location and meteorological data. the results demonstrate that the stacking model outperformed all the models with the lowest RMSE, MSE, and MAE.

Furthermore, the model developed enables real-time investigation of solar energy production and potential investment for a specific region [151], [152].

2.5.3 Bagging, Bootstrap and Aggregating with LSTM

Numerous studies that selected the LSTM model for predicting the accurate output. Bagging involves training multiple LSTM models on different subsets of the data, usually created through random sampling with replacement [153]. Each model generates its prediction, and the final output is the average (for regression tasks) or a majority vote for classification tasks [18]. Boosting focuses on training a series of LSTM models, where each model in the sequence corrects the errors of its predecessor. Models are trained iteratively, with more focus on misclassified or poorly predicted data points from previous iterations [154].

Boosting improves accuracy by focusing on difficult-to-predict instances and can be particularly useful in capturing complex patterns in time series data, such as solar irradiance affected by intermittent cloud cover or other environmental factors [18], [153].

In terms of Stacked LSTM networks are a powerful variant of recurrent neural networks (RNNs) that are particularly effective for sequence prediction tasks. By stacking multiple LSTM layers, these models can capture complex patterns and dependencies in sequential data, making them suitable for a variety of applications, including malware detection, programming language modelling, stock price prediction, and more. The research in [155]. The author enhanced the stacked LSTM method with pre-training to improve malware detection accuracy and robustness in safety-critical systems, achieving 99.1% accuracy in detecting Internet of Thing (IoT) malware samples. Hoyden layers suggested [156]. H-LSTMs, using grow-and-prune training, are compact, fast, and accurate in image captioning, speech recognition, and neural machine translation applications. The research employs grow-and-prune (GP) training to iteratively adjust the hidden layers through gradient-based growth and magnitude-based pruning of connections. This learns both the weights and the compact architecture of H-LSTM control gates.

The paper in [157] proposes an effective Photovoltaic (PV) Power Forecasting (PVPF) technique based on hierarchical learning combining Nonlinear Auto-Regressive Neural Networks with exogenous input, the stacked LSTM model, optimized by the search algorithm, uses the residual error correction associated with the original data to produce a point, the model improved prediction accuracy, outperforming the benchmark approaches with an overall normalized Rooted Mean Squared Error (nRMSE) of 1.98% and 1.33% respectively.

The deep residual network with Bidirectional LSTM improves one-hour-ahead wind power forecasting accuracy and parameter efficiency, enabling a general forecasting framework for grid operations. The research presents a deep residual network for improving time-series forecasting models, indispensable to reliable and economical power grid operations, especially with high shares of renewable energy sources [158]. Motivated by the potential performance degradation due to the overfitting of the prevailing stacked bidirectional long short-term memory (Bi-LSTM) layers associated with its linear stacking.

The LASSO and LSTM integrated forecasting model effectively and accurately predicts short-term solar intensity using meteorological data, aiding smart grid energy management suggested by [159] As a special form of the Internet of Things, a smart grid is an Internet of both power and information, in which energy management is critical for making the best use of the power from renewable energy resources, such as solar and wind, while efficient energy management is hinged upon precise forecasting of power generation from renewable energy resources. In this paper, we propose a novel least absolute shrinkage and selection operator (LASSO) and long short-term memory (LSTM) integrated forecasting model for precise short-term prediction of solar intensity based on meteorological data. In [160]. The smart grid (SG) has emerged as an important form of the Internet of Things. Despite the high promises of renewable energy in the SG, it brings about great challenges to the existing power grid due to its nature of intermittent and uncontrollable generation. To fully harvest its potential, accurate forecasting of renewable power generation is indispensable for effective power management. The LASSO-based approach achieves higher accuracy in solar power generation forecasting, using less training data and being robust to anomaly data points, making it a highly competitive solution for smart grid management.

The adaptive learning hybrid model (ALHM) effectively forecasts short-term and long-term solar intensity using meteorological data, outperforming several benchmarks suggested by [161]. It proposes a novel adaptive learning hybrid model (ALHM) for precise solar intensity forecasting based on meteorological data. We first present a time-varying multiple linear model (TMLM) to capture the linear and dynamic properties of the data. The proposed ALHM captures the linear, temporal, and nonlinear relationships in the data, and keeps improving the predicting performance adaptively online as more data are collected. Simulation results show that ALHM outperforms several benchmarks in both short-term and long-term solar intensity forecasting.

The synthesis explores the use of LSTM models, particularly in combination with bagging and other ensemble techniques, for solar power forecasting [162]. Solar power forecasting is crucial for the efficient integration of photovoltaic (PV) systems into the power grid. Accurate predictions help in balancing supply and demand, optimising grid operations, and enhancing the reliability of power systems. LSTM networks, a type of deep learning model, have shown promise in improving the accuracy of solar power forecasts [162]. The simplified LSTM model outperforms the MLP model in forecasting one-day-ahead solar power generation, with an average RMSE of 0.512, making it suitable for short-term solar power forecasting applications [163]. Power demand forecasting with high accuracy is a guarantee to keep the balance between power supply and demand. Due to the strong volatility of industrial power load, ultra-short-term power demand is difficult to forecast accurately and robustly [164]. To solve this problem, this article proposes a Long Short-Term Memory (LSTM) network-based hybrid ensemble learning forecasting model. A hybrid ensemble strategy, which consists of Bagging, Random Subspace, and Boosting with ensemble pruning, is designed to extract the deep features from multivariate data, and a new loss function that integrates peak demand forecasting error is proposed according to the bias-variance trade-off. Experimental results on open datasets and practical datasets show that the proposed model outperforms several state-of-the-art time series forecasting models and obtains higher accuracy and robustness to forecast peak demand [164].

2.5.4 Discussion

Solar Irradiance Prediction: Ensemble methods with LSTM can significantly enhance the accuracy of solar irradiance forecasting by accounting for temporal dependencies and complex interactions between weather parameters. The forecasting approach of ensemble methods with LSTM is particularly valuable. Here is a brief overview of common ensemble techniques used with LSTM:

- Bagging (Bootstrap Aggregating) with LSTM, Bagging involves training multiple LSTM models on different subsets of the data (usually created through random sampling with replacement). Each model generates its prediction, and the final output is the average for regression tasks or a majority vote for classification tasks. The benefits of this approach reduce the variance and help prevent overfitting, making predictions more stable and robust, especially when dealing with noisy or incomplete data, common in solar radiation and weather forecasting.

- Boosting with LSTM, focuses on training a series of LSTM models, where each model in the sequence corrects the errors of its predecessor. Models are trained iteratively, with more focus on misclassified or poorly predicted data points from previous iterations. the benefits of boosting improve accuracy by focusing on difficult-to-predict instances and can be particularly useful in capturing complex patterns in time series data, such as solar irradiance affected by intermittent cloud cover or other environmental factors.
- Stacking with LSTM is a more complex ensemble method where multiple LSTM models are trained on the same data, and their outputs are combined using another model like, a gradient boosting decision tree (GBDT) or a linear regression model might be used as the meta-learner to combine predictions from multiple LSTMs. This approach gives the benefits of Stacking allows leveraging different LSTM architectures or hyperparameters, combining their strengths to improve prediction accuracy. It can capture diverse temporal patterns that might be missed by a single model.
- Hybrid ensemble methods combine LSTM with other machine learning models, such as ARIMA Auto-Regressive Integrated Moving Average or support vector machines (SVM). LSTM captures long-term dependencies in the time series, while traditional models handle short-term trends or seasonality more effectively. The benefits of this hybrid approach often led to superior performance by exploiting the strengths of both deep learning (LSTM) and traditional statistical methods, making it ideal for highly seasonal or volatile solar radiation data.

2.6 Ensemble LSTMs forecasting techniques

ensemble methods that combine Long Short-Term Memory (LSTM) models have been quite effective for solar energy forecasting [165]. These methods leverage the strength of multiple models to enhance predictive performance, capturing both temporal patterns and non-linear relationships in solar irradiance data [166]. LSTMs are excellent for time-series data because they can remember long-term dependencies, which is critical in predicting solar energy generation as weather conditions follow daily and seasonal patterns. ensemble methods can improve generalisation. This reduces the risk of overfitting to specific noise or trends in the training data. Numerous studies utilised LSTM as an ensemble method with different machine learning. A novel hybrid model by [162].

The study presents a hybrid approach for forecasting solar radiation. It combines LSTM with ensemble learning methods, such as Gradient Boosting and Random Forest. The hybrid model shows improved accuracy compared to standalone models. The LSTM handles time dependencies, while the ensemble learning methods help capture non-linearities and enhance overall prediction. According to [167].

The research investigates an ensemble approach that combines multiple deep learning models, including LSTM, CNN, and GRU (Gated Recurrent Units), for solar power forecasting. The ensemble model outperforms individual models by leveraging their strengths to handle complex weather data, improving both short-term and long-term solar power predictions. The study in [166]. The study proposes an ensemble approach that blends LSTM networks with statistical forecasting methods such as ARIMA Auto-Regressive Integrated Moving Average. The ensemble method demonstrated superior performance in forecasting solar power, reducing the forecasting error through the combination of deep learning's flexibility with statistical models' interpretability. The author in [168]. introduces a hybrid ensemble framework that uses an ensemble of LSTM models in combination with decision trees and boosting algorithms like XGBoost. The ensemble effectively captures temporal patterns and complex nonlinear relationships in weather and irradiance data, achieving better forecasting accuracy compared to individual models. A hybrid deep learning model for solar power forecasting based on LSTM and CNN models in [165]. In the paper, the authors propose a hybrid ensemble model combining LSTM and CNN for short-term solar power forecasting. The LSTM models learn the sequential patterns, while CNN captures spatial features from satellite imagery. The ensemble model outperformed traditional machine learning models and standalone deep learning models. The paper in [165] Learning for Solar Energy Prediction introduces an ensemble of LSTM, GRU, and feedforward neural networks for day-ahead solar energy predictions. The ensemble strategy showed a significant improvement in prediction accuracy over individual models by combining multiple architectures that capture various aspects of the data.

2.7 Discussion

These studies demonstrate how ensemble approaches combining LSTM models are effective for solar energy forecasting, as they leverage the strengths of various models to produce more accurate and reliable predictions [167]. The application of ensemble methods that integrate Long Short-Term Memory (LSTM) networks in solar energy forecasting has gained significant attention due to the growing need for accurate and reliable predictions. Solar energy generation is highly dependent on weather patterns, which exhibit both short-term variability and long-term trends [169]. Accurate forecasting helps to optimise energy management, reduce operational costs, and improve grid stability [170]. Ensemble methods, particularly those combining LSTM models with other machine learning techniques, have shown promising results in addressing the inherent complexity of solar power prediction [171]. Ensemble learning techniques, which combine multiple models, are effective in overcoming the limitations of individual models by aggregating their predictions to improve overall accuracy and robustness [172]. By leveraging the strengths of multiple models, ensemble methods can reduce the variance and bias in solar energy forecasts. For instance, hybrid approaches often combine LSTM models with statistical methods like ARIMA, machine learning models like Random Forests or even other deep learning techniques like CNN or GRU. These hybrid models improve prediction accuracy by capturing both temporal patterns through LSTMs and non-linear dependencies through other models [173], [174].

Chapter 3
The methodology of Forecasting Solar power

3.1 Introduction

The current landscape of solar power forecasting research often presents a disjointed approach, with a significant body of work focusing either on solar irradiance forecasting or on solar power output forecasting. However, these two elements are inherently linked in real-world applications, where fluctuations in solar irradiance directly impact the power output of photovoltaic systems. [175]. The separation between these two forecasting approaches introduces a critical gap in the accuracy and practicality of existing models, especially for energy management systems that rely on precise predictions for both variables. [115], [176].

This research addresses the gap by proposing an integrated forecasting model that simultaneously predicts both solar irradiance and power output. The methodology leverages Long Short-Term Memory (LSTM) neural networks, which are highly effective in capturing temporal dependencies in time-series data. By utilising stacked LSTM layers, the model is designed to process complex relationships and interactions between irradiance and power output over time, providing a more comprehensive and accurate forecasting tool. [177]

The methodology consists of several key phases: data pre-processing, model architecture design, training, validation, and evaluation. First, historical time-series data on solar irradiance and corresponding power output is pre-processed to create sequences that capture past trends. [178]. The LSTM model is then constructed with multiple layers to capture both short-term and long-term dependencies in the data. The model is trained using this dataset, and its performance is evaluated based on metrics such as Mean Absolute Error (MAE) and Mean Squared Error (MSE). This integrated approach addresses the limitations of previous studies by combining the prediction of both irradiance and power output into a unified framework, providing more reliable forecasts for practical applications in solar energy management [179].

One of the most important steps in estimating future solar power generation from photovoltaic (PV) systems is solar power forecasting. Solar power forecasting employs several important approaches, including physical, statistical, and artificial neural networks, numerical weather production and machine learning algorithms. The predicting methodology depends on the forecast horizon (very short-term to long term), spatial and temporal resolution requirements and available input data (weather, irradiance, PV system data).

The dataset of PV power output was collected from the database (data for two PV plants located in Iraq). Then, a data pre-processing, which aims to confirm the input form of the dataset to the Long Short-Term Memory (LSTM) model, was carried out. The pre-processing corresponds to data sparsity, interpolation of any missing values, and data scaling and normalization. 80% of the time series data were used as inputs for the generation of the PV power prediction model, whereas the rest of the data was used to verify that the model could predict the PV power output [180] .This chapter has used the ensemble method based on LSTM to give the details of the dataset used, describe the issues that were tackled during the pre-processing phase, and also present the forecasting method applied in this work. By combining accurate forecasting with fault detection, this methodology offers a comprehensive solution for improving the efficiency and reliability of solar power plants. This integrated approach addresses the limitations of previous studies, providing more reliable forecasts and early fault detection, which is crucial for reducing downtime and optimizing energy generation in real-world applications.

3.2 Machine Learning Forecasting Methods

Accurate forecasting of solar power generation is crucial for optimising energy management and grid stability. This document explores two prominent machine learning methods for solar power forecasting: Artificial Neural Networks (ANN) and Support Vector Regression (SVR) [176]. Each method's principles, advantages, and applications will be discussed to provide a comprehensive understanding of their roles in solar energy forecasting. Artificial Neural Networks (ANN) are computational models inspired by the human brain's neural networks. They consist of interconnected nodes (neurons) organised in layers, which can learn complex patterns from data. ANNs are particularly effective in handling non-linear relationships, making them suitable for solar power forecasting [181]. The key Features of ANN:

- Structure: ANNs typically consist of an input layer, one or more hidden layers, and an output layer. Each connection between neurons has an associated weight that is adjusted during the training process [182], [183].
- Learning Process: The training of an ANN involves feeding it historical solar power data along with relevant features like weather conditions, and time of day. The network learns to minimise the error between predicted and actual values using optimisation algorithms like backpropagation [182], [183].
- Flexibility: ANNs can be tailored to various forecasting horizons, from short-term to long-term [182], [183].

The advantage of ANN is high accuracy because it can capture complex patterns in data, leading to high forecasting accuracy as shown in Figure 3.1. Furthermore, adaptability to retraining with new data allows them to adapt to changing environmental conditions. In addition, robustness for handling noisy data and missing values effectively has been successfully applied in various solar power forecasting scenarios, including predicting daily solar energy output and optimising energy storage systems [184].

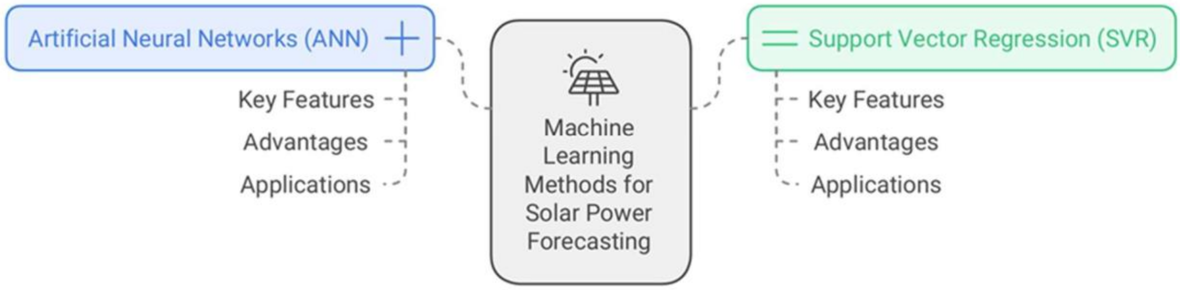


Figure 3.1: Machine Learning Methods for Solar Power Forecasting

On the other hand, ML Support Vector Regression (SVR) is a type of Support Vector Machine (SVM) that is used for regression tasks. SVR aims to find a function that deviates from the actual observed [185]. Key Features of SVR:

- Kernel Trick: SVR employs kernel functions to transform input data into higher-dimensional spaces, enabling it to capture non-linear relationships without explicitly mapping the data [[87], [185].
- Margin of Tolerance: SVR introduces a margin of tolerance, allowing for some errors in predictions, which can be particularly useful in noisy datasets [87], [185].
- Regularisation: SVR includes a regularisation parameter that helps prevent overfitting, ensuring that the model generalises well to unseen data [87], [185]

The advantage of SVR it is efficient and effective for Small Datasets. It has been utilised in numerous studies for short-term solar power forecasting, including predicting hourly solar irradiance and energy output based on meteorological data. [186]

In the realm of machine learning, datasets play a crucial role in training algorithms to make predictions or uncover patterns. This document explores the two primary types of datasets used in machine learning: supervised and unsupervised datasets. By understanding the differences between these two categories, one can better grasp how machine learning models learn from data and how they can be applied to various problems. [187], [188].

3.2.1 Supervised Learning

Supervised learning involves training a model on a labelled dataset, which means that each training example is paired with an output label. The goal of supervised learning is to learn a mapping from inputs to outputs, allowing the model to make predictions on unseen data. In supervised learning, the quality and quantity of the labelled data are crucial, as they directly impact the model's ability to generalise and make accurate predictions. Common algorithms used in supervised learning include linear regression, logistic regression, decision trees, support vector machines, and neural networks, among others, as shown in Figure 3.2.

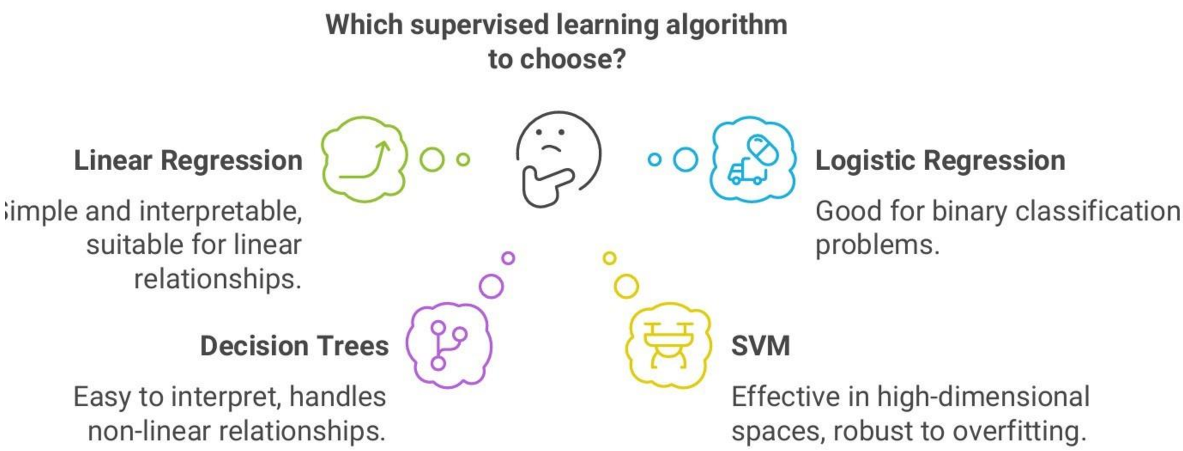


Figure 3.2: Characteristics of Supervised Datasets

The data's characteristics are as follows: first, it is labelled data; each instance in the dataset has a corresponding label or output. Second, Training and Testing data are used to evaluate the model's performance. Third, algorithms like linear regression, Decision Trees, support vector machines, and Neural networks are used. Overall, supervised learning is a powerful tool in various applications, including image recognition, spam detection, and medical diagnosis, where accurate predictions are essential for decision-making processes.

The labelled data of solar power forecasting includes data like sunshine, wind speed, air pressure, and air temperature. Input data show the solar energy output maximum voltages kwh as shown in Figure 3.3 by depending on the features like Windspeed', Sunshine, Air Pressure, Radiation, Air Temperature, Relative Air Humidity. The Y access represents the (System Production Voltages kwh), X access represents (Date in Hours).

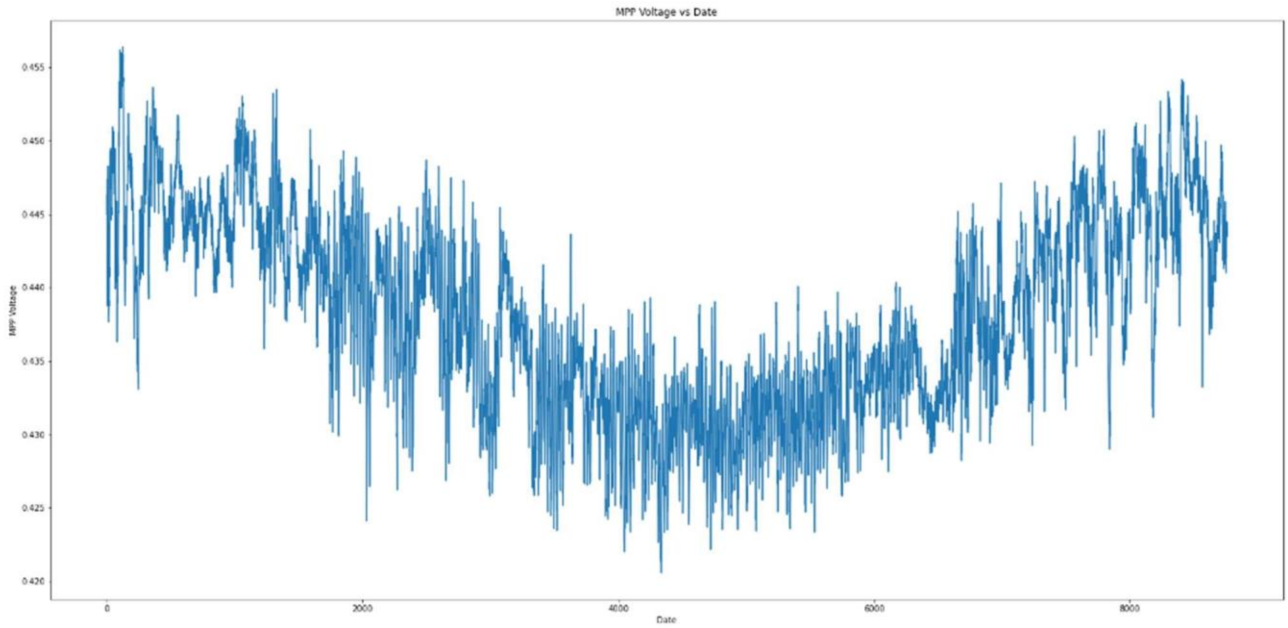


Figure 3.3: The maximum power point voltages of solar energy through time by using labelled data as a result kwh in hourly times and dates.

3.2.2 Unsupervised Datasets

There is unsupervised learning, on the other hand, deals with datasets that do not have labelled outputs. The objective here is to identify patterns or structures within the data without any prior knowledge of the outcomes. The dataset consists of input data without any corresponding output labels. The focus is on discovering hidden patterns or intrinsic structures in the data. Common Algorithms: Algorithms used in unsupervised learning include k-means clustering, hierarchical clustering, and Exploratory Data Analysis (EDA). Figure 3.4 shows the diagram of the unsupervised dataset.

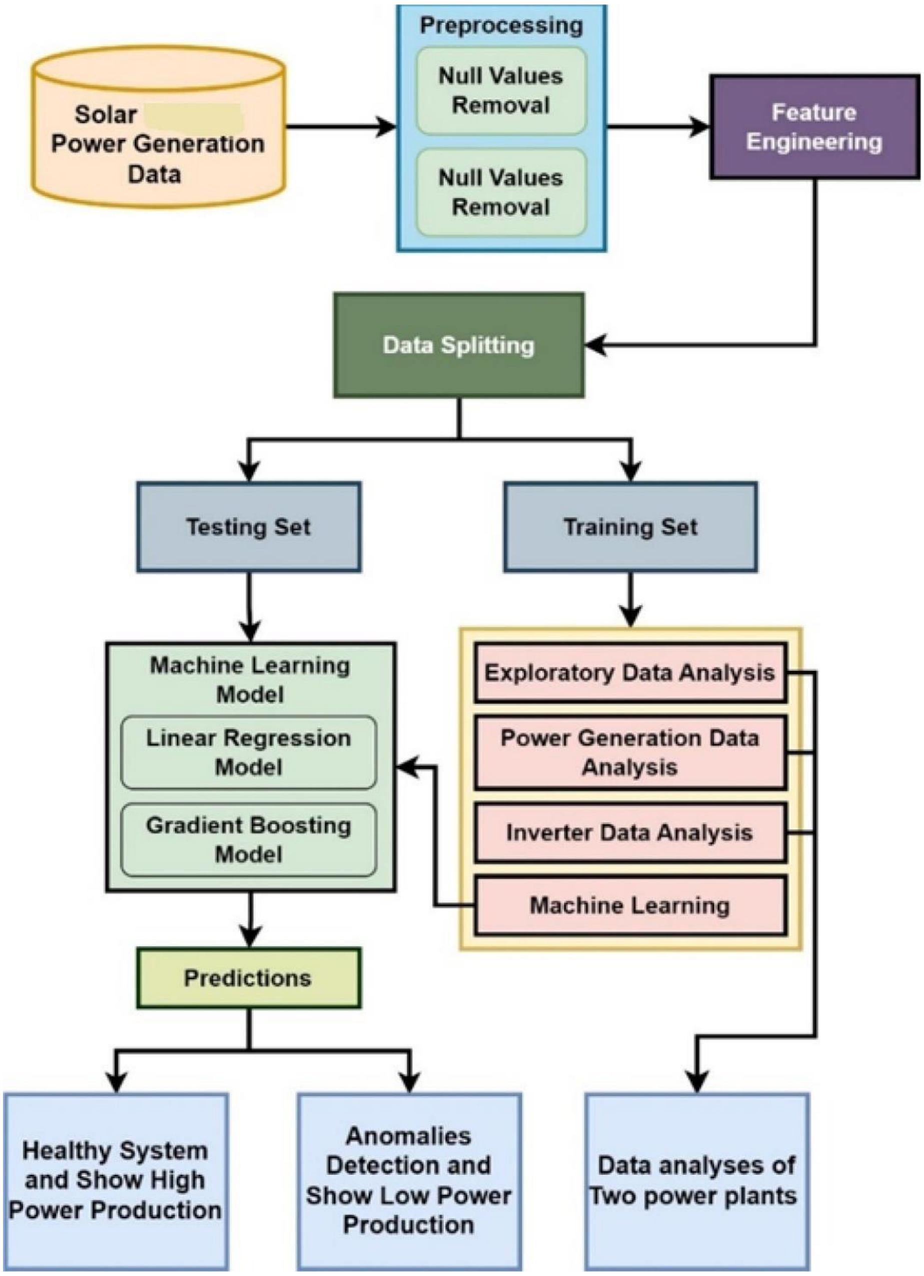


Figure 3.4: Expository Data Analysis supervised learning task

In these two examples of this task, Clustering means grouping similar data points. Dimensionality Reduction means reducing the number of features in a dataset while preserving its essential characteristics. Unsupervised data analysis using Keras is usually an open-source neural network library written in Python. It is designed to facilitate the development of deep learning models and provides a user-friendly interface for building complex neural networks. Here are some key aspects of Keras shown in Table 3.1.

level	Description	Use case
Low-level API	Greater control with custom components	Researchers and advanced users
Intermediate API	Balance of flexibility and user-friendliness	Practitioners building standard models
Highly level API	User-friendly, minimal code for a quick model creation	Beginners and rapid prototyping

Table 3.1: Three primary levels of deep learning by Keras

API stands for Application Programming Interface. It is a set of rules and protocols that allows different software applications to communicate with each other. APIs define the methods and data formats that applications can use to request and exchange information. APIs are essential for modern software development as they enable different applications and services to work together, facilitating integration and functionality across platforms. They allow developers to build applications more efficiently by leveraging existing services and functionalities. Table 3.2. shows both supervised and unsupervised data ML classification features of solar power production.

Aspect	Supervised data	Unsupervised data
Labelled	Required labelled data	None labels available
Goals	Predict outputs based on inputs	Discover patterns and relationships
Example	Classification, regression	Clustering, association
Model evaluation	Can evaluate performance using labels	Evaluation is more subjective
Complexity	More complex due to the labels	Typically, simpler but exploratory

Table 3.2: supervised and unsupervised data classification.

3.2.3 Method discussion

Understanding the distinction between supervised and unsupervised datasets is fundamental for anyone venturing into the field of machine learning. Supervised datasets provide a clear path for training models with known outputs, while unsupervised datasets allow for the exploration of data without predefined labels. Both types of datasets have their unique applications and are essential for developing robust machine learning solutions. In the field of machine learning, data can be categorised into two main types: supervised data and unsupervised data. Understanding the differences between these two types is crucial for selecting the appropriate algorithms and methodologies for specific tasks.

Both supervised and unsupervised data play vital roles in machine learning. Supervised learning is powerful for tasks where the outcome is known, while unsupervised learning is essential for discovering insights and structures in unlabelled data. Understanding the differences and applications of each type is crucial for effective data analysis and model development. As shown in Figure 3.5.

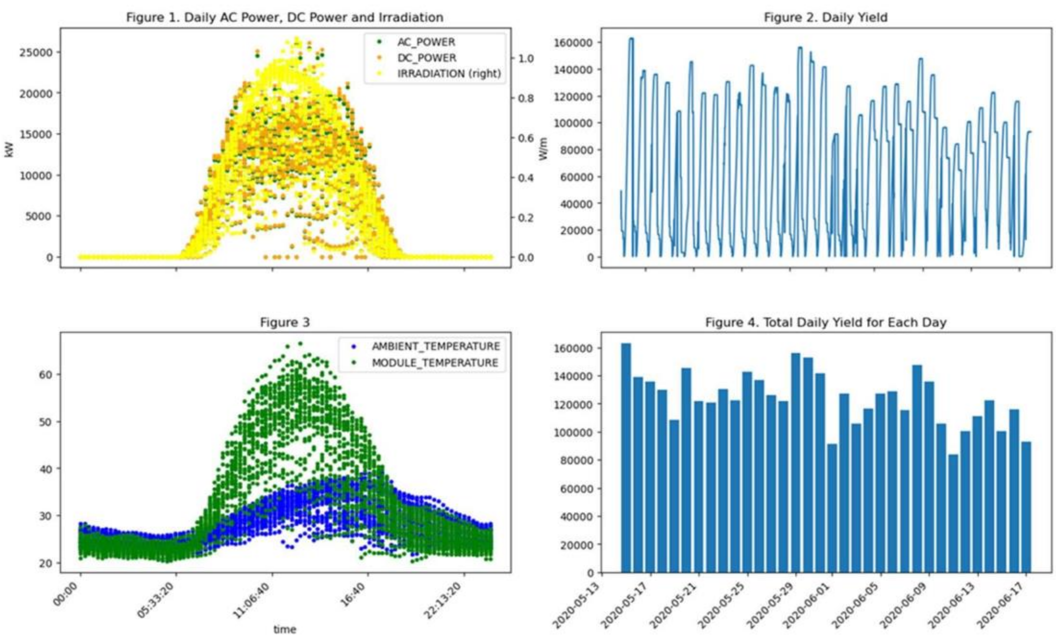


Figure 3.5: Solar power plant DC & AC generation prediction.

3.3.1 Statistical Methods

Statistical solar power forecasting is a vital aspect of optimizing solar energy usage. By utilizing various statistical methods and addressing challenges, accurate predictions can be made, which is essential for the integration of solar power into energy systems. Predictions are made using statistical methods using both historical and current data. Time series models employ historical patterns to predict future values, regression models discover links between variables to produce predictions, and persistence forecasting, which is a key method, assumes current conditions. The statistical method is effective for forecasts that are made soon (minutes to hours ahead). Statistical Applications in Grid Management help balance supply and demand on electrical grids, Investment Decisions assist stakeholders in making informed decisions regarding solar investments and Policymaking to support the development of renewable energy policies and incentives.

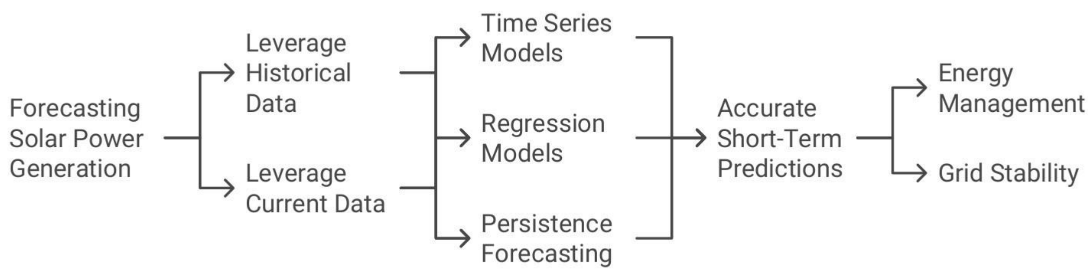


Table 3.3: Statistical Solar Power Forecasting Diagram3.3.2 Time Series

Time series models are a cornerstone of solar power forecasting. They utilize historical data to identify patterns and trends over time, allowing for the prediction of future solar power generation. These models analyse past solar irradiance, temperature, and other meteorological factors to forecast future values. Common time series techniques include Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing, which can capture seasonality and cyclic behaviour in solar energy production. Key Features of Time Series Models, Historical Data Utilization: Time series models analyse extensive historical datasets, which encompass numerous factors that influence solar energy production. Identifying Patterns: By examining past data, these models can detect recurring patterns and trends, which are vital for making informed forecasts about future solar power output.

The models specifically investigate several meteorological elements, including solar irradiance (the amount of solar power received per unit area, which directly impacts energy generation) and temperature (temperature changes can affect the efficiency of solar panels and overall electricity production). However, for other Meteorological Variables, factors such as humidity, cloud cover, and wind speed can also play significant roles in forecasting solar energy.

3.3.3 Regression Models

Regression models are another vital statistical approach in solar power forecasting. These models explore the relationships between solar power generation and various independent variables, such as weather conditions, time of day, and geographical location. By establishing these links, regression models can produce predictions based on current and forecasted values of the influencing variables. Multiple linear regression and polynomial regression are frequently employed to enhance the accuracy of these forecasts.

Regression Model	**Strengths**	**Use Cases**
Linear Regression	Simplicity, interpretability	Baseline forecasting
Lasso/Ridge Regression	Handles multicollinearity, regularization	Feature selection, robust to noise
Elastic Net Regression	Combines Lasso/Ridge benefits	High-dimensional data
Random Forest Regression	Handles nonlinearity, robust to overfitting	High-accuracy forecasting
Gradient Boosting Regressor	High predictive power, adapts to complexity	Real-time, smart grid integration
Hybrid/Ensemble Models	Combines the strengths of multiple models	Enhanced accuracy, reliability

Table 3.4: Enhancing solar power forecasting with regression

Persistence forecasting is a straightforward yet effective method that assumes current conditions will persist into the near future. This approach is particularly useful for short term forecasts, typically ranging from minutes to hours ahead. By using the most recent data points, persistence forecasting can provide quick estimates of solar power generation, making it a valuable tool for immediate decision-making in energy management. As shown in Figure 3.7

3.3.4 Discussion

Statistical methods play a crucial role in solar power forecasting, enabling accurate predictions that support energy planning and grid operations. Time series models, regression models, and persistence forecasting each offer unique advantages for short-term forecasting. As solar energy continues to grow in importance, the refinement and application of these statistical techniques will be essential for optimising solar power utilisation. Using these statistical methods allows for more accurate and reliable solar power forecasting. By leveraging historical data, environmental factors, and advanced algorithms, stakeholders can make informed decisions regarding energy production and consumption.

3.4.1 Physical methods

To anticipate solar power output, physical approaches rely on physical principles and numerical weather prediction (NWP) models. These methods usually include the use of satellite and sky imagery to estimate solar irradiance, along with NWP models to forecast cloud cover and atmospheric conditions. irradiance forecasts into power output projections by modelling the physical properties of PV installations. Physical approaches are helpful for medium- to long-term forecasts (6+ hours ahead). Solar power forecasting using physical methods relies on modelling the physical and environmental processes that affect solar energy production. These methods use meteorological data and knowledge of solar radiation principles to predict the amount of solar energy that will be available at a particular location and time. The key components and techniques involved in physical methods of solar forecasting include Numerical Weather Prediction (NWP) models are used to predict weather conditions that influence solar radiation, such as cloud cover, temperature, humidity, and wind speed. These models are based on the physical laws of atmospheric dynamics and thermodynamics. Most of the physical prediction (Clear Sky) models predict solar irradiance under cloudless conditions, using the physical principles of solar radiation and the interaction of sunlight with the atmosphere. These models consider factors like the position of the sun, atmospheric composition (e.g., aerosols, water vapour), and geographical features like altitude and latitude. This model can be used as a reference for estimating the maximum possible solar power on a sunny day and for correcting forecasts when clouds are present.

3.4.2 Exploratory Data Analysis (EDA) of Solar Power Generation Visualization Analysis

Analyse solar power generation visualisation, which can analyse data and long-term trends in solar energy consumption. Important considerations when producing solar power visualisations include installed capacity versus actual generation, percentage of solar energy used in the grid, solar technology cost, emissions reduction, and seasonal and daily generation patterns. By incorporating these elements, visualisations can provide a comprehensive view of solar power's growing role in the global energy landscape. The techniques of Exploratory Data Analysis (EDA) examine and visualise data to understand its main characteristics, identify patterns, spot anomalies, and test hypotheses. It helps summarise the data and uncover insights before applying more advanced data analysis techniques. Aimed to use this process to understand the data in depth and learn its distinctive characteristics, often using visual means. It allows a machine learning model to predict the dataset better and gives more accurate results [189], [190]. Steps Involved in Exploratory Data Analysis:

- Understand the Data (identify the object of analysis)
- Data Collection (collect the required databases)
- Data Cleaning (handle missing values: Impute or remove missing data)
- Data Transformation (normalise or standardise the data to create features through feature engineering)
- Data Integration (Integrate data from various sources to create a complete data set)
- Data Exploration (Univariate Analysis, Bivariate Analysis and Multivariate Analysis)
- Data Visualization (Visualize data, using visual tools such as bar charts, line charts, scatter plots, heatmaps, and box plots).
- Descriptive Statistics [Calculate central tendency measures (mean, median, mode) and dispersion measures (range, variance, standard deviation)]
- Identify Patterns and Outliers Detect patterns, trends, and outliers in the data using visualizations and statistical methods.
- Hypothesis Testing [Formulate and test hypotheses using statistical tests (e.g., t-tests, chi-square tests) to validate assumptions or relationships in the data]
- Data Summarisation (Summarise findings with descriptive statistics, visualisations, and key insights).
- Documentation and Reporting (iteration data set based on feedback) [191], [189].

3.4.3 univariate and multivariate time data series

The two primary time series data types utilised in forecasting and analysis are univariate and multivariate. A univariate time series involves one variable that depends on time to the highest measurements of one variable over regular time intervals. Each data point represents the value of the variable at a specific time. The analysis focuses on patterns and trends within that single variable. For example, temperature changes through periodic time. On the other hand, A multivariate time series involves two or more variables that depend on time. It records measurements of multiple variables simultaneously over time. Each data point represents values of multiple variables at a specific time. The analysis examines relationships and dependencies between variable patterns. Table 3.5 shows the difference between multi- & univariate [192].

Aspect	Univariate	Multivariate
Variable numbers	One	Two or more
Complexity	Lower	Higher
Relationship	Self-dependencies	Inter-variables dependencies
Modelling	ARIMA, Exponential Smoothing	Vector ARIMA (VARIMA), Neural Networks
Forecasting approach	Based on past values of a single variable	Considers past values and interactions of multiple variables

Table 3.5: multi-variant and union variant dataset techniques

The Analysis Techniques AT for Univariate analysis often employ Autocorrelation functions like Autoregressive Integrated Moving Average ARIMA as a simple forecasting model. However, Multivariate analyses of more complex models using Vector Autoregression VAR and each variable depends on the previous variable as a Cross-Correlation function test. Consideration using this model by choosing appropriate modelling techniques that can capture inter-variable dependencies [191], [193].

3.4.4 Time series in visualisation techniques based on data sets, structure and operational

Time series data refers to a sequence of data points collected or recorded at specific time intervals. In the context of solar power, this data can include metrics such as solar irradiance, energy output, temperature, and weather conditions over time [194]. Analysing this data helps in identifying trends, seasonal variations, and anomalies in solar power generation [195]. Time series data refers to a sequence of data points collected or recorded at specific time intervals, as shown in Table 3.6. In the context of solar power, this data can include metrics such as solar irradiance, energy output, temperature, and weather conditions over time. Analysing this data helps in identifying trends, seasonal variations, and anomalies in solar power generation [194], [195].

Technique	**Purpose/Insight**
Line Chart	Trends, seasonality, and forecast comparison
Box Plot	Variability, outliers, distribution
Decomposition Plot	Trend/seasonal/residual components
Wavelet Decomposition	Frequency analysis, no stationarity
Error/Residual Plot	Model performance, error analysis
Comparative Line Chart	Actual vs. predicted output
Pie/Bar Chart	Categorical/seasonal distribution
Heatmap	Two-dimensional temporal patterns

Table 3.6: Time series visualisation techniques of solar power

A time series is a dataset where each observation corresponds to a specific point in time. the type of time series data that is collected at regular intervals over time. It is useful for analysing patterns, trends, and behaviours across time [196]. In this research, Python has been used in several libraries that can be used in various tools for plotting time series data, including

Matplotlib, Seaborn, and Pandas. These libraries provide diverse types of visualisations such as line graphs, scatter plots, histograms, etc. It is distinct from other kinds of datasets due to the presence of time dependency, indicating that an observation's value at any current time is dependent upon its values in the past. Because of this characteristic, time series analysis is distinct and difficult because it calls for specialised methods and equipment for data processing and interpretation [197]. cleaning time series data to check for missing values, errors, or equipment failure is a common cause of missing values. Interpolation methods like linear interpolation and forward filling can be used to fill in missing values. The next step is outlier variables, which can be detected using statistical methods, such as Auto-regression variables, Augmented-dicky failure, and Z-score method [198], [199], [153].

Overall, Python provides a rich ecosystem for working with time series data and creating compelling visualisations. Effective visualisation of time series data is essential for understanding solar power generation patterns. By employing various visualisation techniques such as line charts, area charts, bar charts, heat maps, scatter plots, and box plots, stakeholders can gain valuable insights into solar energy performance. These visualisations not only aid in decision-making but also enhance communication among stakeholders in the renewable energy sector. As the solar industry continues to grow, mastering these visualisation techniques will be crucial for optimising solar power generation and maximising its potential [92], [95].

Forecasting is a critical aspect of decision-making in various fields, including business, economics, and environmental science. This document explores various forecasting techniques that leverage data sets, structural elements, and operational contexts to enhance predictive accuracy [200]. By understanding the nuances of different methodologies, organizations can better prepare for future trends and make informed decisions. The accuracy of solar forecasting findings is influenced by the choice of data set for a certain geographic area. Several devices must be installed at the target area to collect data on various meteorological and solar irradiation components; however, the instruments' ageing, erroneous behaviour, and shadowing during sunrise and sunset hours, dust, raindrops, cloud coverage, etc., induce the errors in the recorded data. Three primary categories may be distinguished between the forecasting models derived from data sets: those based on:

- Time series data sets dependent on historical data.
- Structural data sets are operated based on meteorological and geographical data sets.
- Hybrid data sets-based models combine the features of both the previously mentioned models.

- Maximum power point tracker which depends on the self-referencing data (manufacturing components data)

The literature has several models that differ only in how they are used, operated, and structured. These models were designed with various intelligent forecasting techniques [18], [200], [201] They can be classified into three categories shown in Figure 3.9.

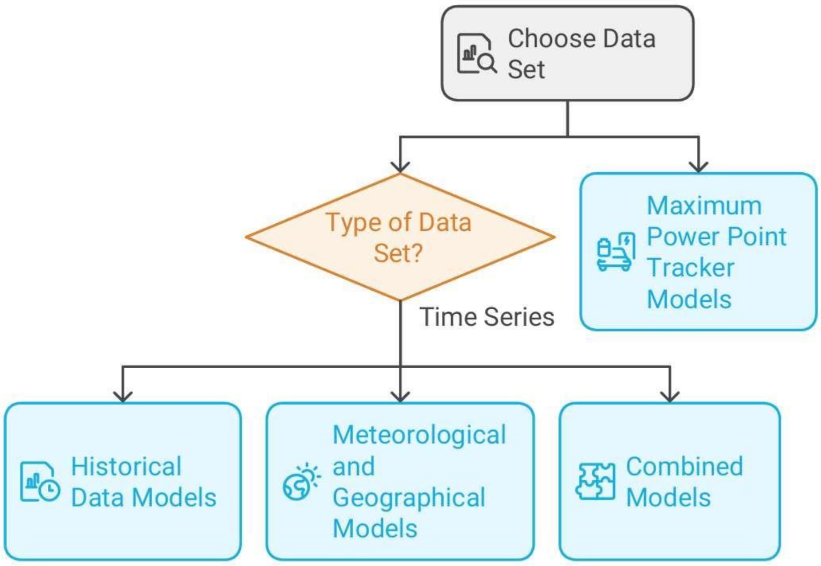


Figure 3.6: classified categories of forecasting techniques

Time series forecasting techniques analyse historical data points collected or recorded at specific time intervals [195]. These methods are effective when past patterns are expected to continue. Key time series techniques include Moving Averages (MA), the averages calculated over a specific number of past periods to smooth out fluctuations. Secondly, Exponential Smoothing (ES) is a technique that applies decreasing weights to older observations, giving more importance to recent data. Thirdly, Autoregressive Integrated Moving Average (ARIMA) is a sophisticated model that combines autoregression, differencing, and moving averages to capture various data patterns [202].

3.5 Enhancing solar power prediction by utilising LSTM & MPPT

LSTM model and associated hybrid models in diverse fields; several research endeavours were unable to furnish adequate particulars regarding the models. However, since the LSTM model can estimate solar power, the LSTM standalone and hybrid models, along with its partially covered and uncovered criteria

The MLP or feed-forward neural network (FFNN), is a core component of deep learning architectures [96]. In Figure 3.1, it is seen the MLP or feed-forward neural network (FFNN), is a core component of deep learning architectures. This image aids in illuminating the MLP's fundamental structure, which is crucial to comprehend since it serves as the foundation for more intricate deep learning models shown in Figure 3.5.1 Power systems commonly use MLPs to safeguard transmission lines, find transformer problems, and keep an eye on online voltage stability [73], [96].

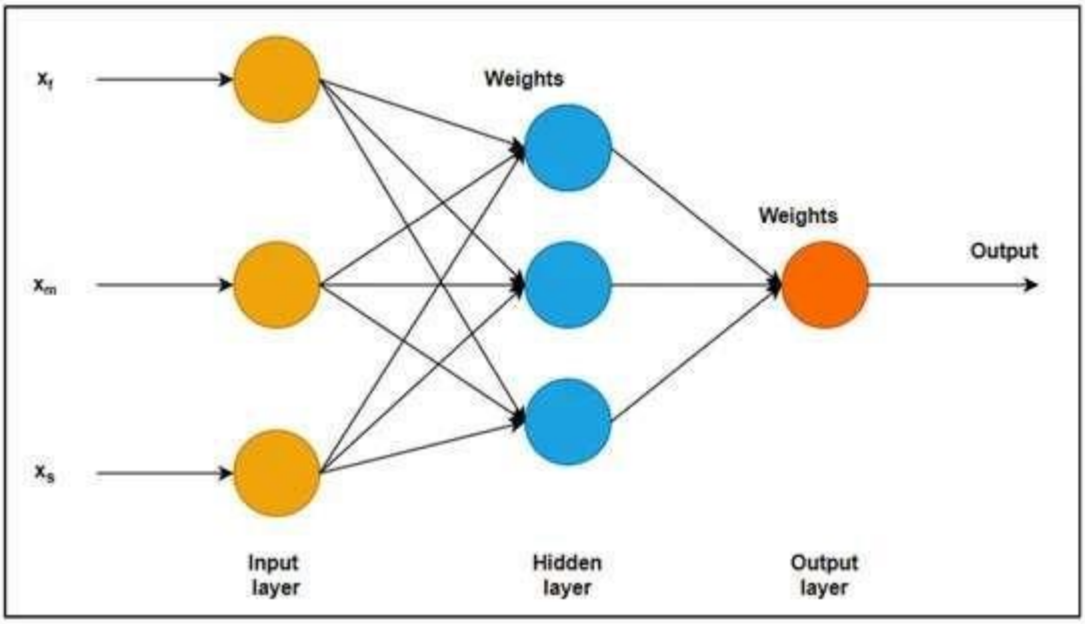


Figure 3.7: feed forward neural network FFNN model of LSTM [73]

Power systems commonly use MLPs to safeguard transmission lines, find transformer problems, and keep an eye on online voltage stability [147]. As depicted in figures 3.10 and 3.11 show how data is transmitted linearly from one side of the diagram to the other, with no lines ever returning to a specific layer or node. Furthermore, input is sent to a certain node just once and never again [203]. This pattern of information sharing suggests that only the most recent input and training instructions are remembered in an FNN, indicating memory loss. Therefore, the FNN-supplied method is not helpful for forecasting or prediction unless previous knowledge is provided [166].

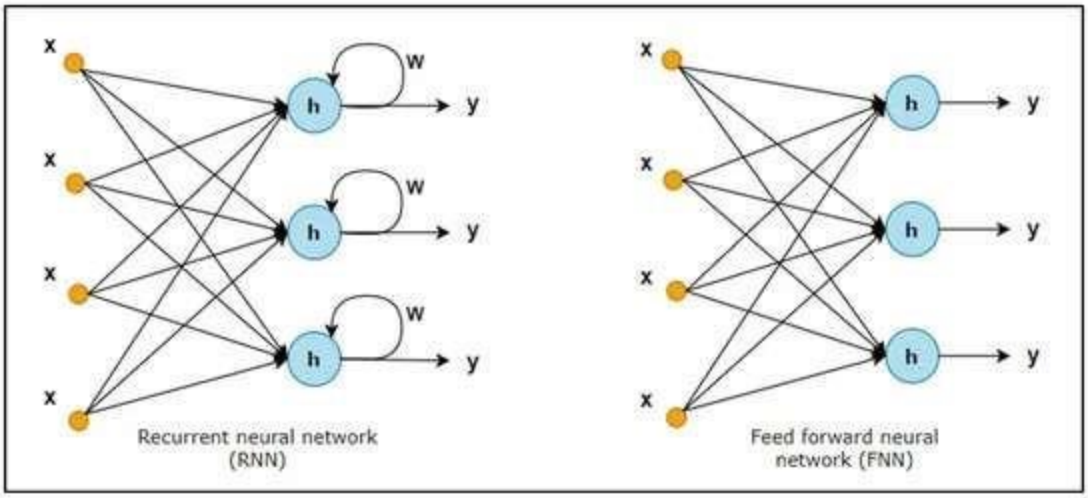


Figure 3.8: A comparison between RNNs and FNNs, with loops shown in the hidden layers (blue circles) of the former [166].

The previous output and the current state input are used to inform the data processing decision. For instance, by incorporating the outcome from the previous time step into the current time step, the irradiation data of a specific date or time can be predicted [1], [201]. Other data types can also be used with this approach. Furthermore, since people often learn in progressive sequences as opposed to random sequences, RNNs resemble human synapses more than FFNNs do. The RNN is therefore the best option for predictive models [38], [204].

The ensemble methods based on BI-LSTM have been chosen for LSTM1 with 80 hidden layers and LSTM2 with 50 hidden layers in this research. BI-LSTM has been selected for enhancing the forecasting accuracy based on the ensemble methods, as shown in Figure 3.3. This model clarifies sequence length and feature count [samples, timesteps, features], the historical meteorological datasets used for model training of PV system power output readings. Data include six features in the figure: wind speed, Irradiation, temperature, sunshine, air pressure, the relative air humidity in the X definition, and system production in the Y definition, as shown in Table 3.7

X parameter	WindSpeed (km/h)	Sunshine (W/m^2)	AirPressure (pa)	Radiation (W/m^2)	Air temp. (°C)	Relative AirHumidity (%RH)
Y parameter	System Production (kWh)	To be predicted	To be predicted	To be predicted	To be predicted	To be predicted

Table 3.7: The definition of the X& Y parameters

The data set characteristics X and Target y are provided, which can be divided into training and testing (Sequential) as (X _test, test, X _train, Y train, and test size typically (20 %). The model is a feedforward neural network (NN) with a linear stack of layers in the (Keras) deep learning toolkit.

- The accuracy of output from two LSTM ensemble algorithms with a short-term forecast horizon was compared [15]
- each followed by a dropout layer for regularisation and a final dense layer for output.
- The model is compiled with the Adam Optimizer and is designed to minimize the mean squared error, making it suitable for regression tasks.
- Adam is often recommended as the default optimizer to use for deep learning DL technique. Dense Layer: as a fully connected layer.
- The expected input shape for the model is sequences of length 1 with 6 features each.
- Dropout layer: to prevent overfitting. Used 20% of neurons dropped out. According to the results, the seven-day forecasting model can predict well, since a visual examination of the results indicates that the predicted power output signal reacts to each fluctuation and follows the trend of the actual power output signal. Furthermore, the Root Mean Squared Error (RMSE) of our model when applied to the test data gives a value of 0.11049111870458088, whereas when applying the k-fold cross-validation, the mean of the resulting Mean Absolute Error (MAE):

0.05512441745205155 with a standard deviation. R2 Score: 0.634048582919565.

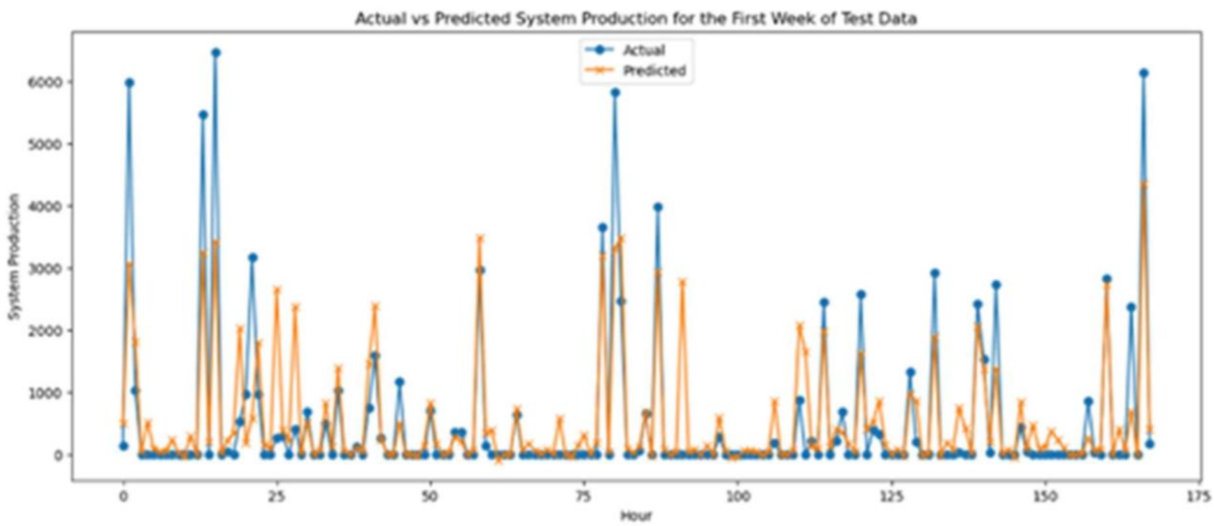


Figure 3.9: The actual and predicted energy production

In recent years, artificial intelligence (AI) techniques have emerged as promising tools for improving the efficiency of renewable energy systems. Long Short-Term Memory (LSTM) neural networks, a type of recurrent neural network (RNN), offer significant potential for enhancing MPPT precision and adaptability in solar panel systems. Unlike traditional MPPT algorithms, LSTM networks can effectively capture temporal dependencies in the input data and learn complex patterns, enabling more accurate and dynamic control of solar panel operation [8, 9]. The application of LSTM neural networks represents a novel approach to addressing the challenges associated with traditional MPPT techniques, thereby unlocking new opportunities for optimising solar energy harvesting.

3.5.1 Discussion

A double-mode maximum power point tracking (MPPT) technique is provided in this publication. To realise PV MPPT in both normal operation and partial shade condition (PSC), a radial basis function neural network (RBFNN) is built and driven by historical temperature and irradiance data; for MPPT under normal conditions, the conventional perturb and observe (P&O) method is used. Long short-term memory (LSTM) based solar power prediction is used to detect the partial shading condition. The deviation between the predicted and measured power is used to trigger the MPPT modes switching smoothly and adaptively, improving MPPT speed and accuracy even in the presence of erratic solar irradiance and ambient temperature. An MPPT method is demonstrated through the construction of a PV system model using MATLAB/Simulink. Several instances, such as PSC and regular operation, are considered. The method is superior to existing classic MPPT methods, as shown by the comparison findings. It can immediately realise MPPT in a variety of work settings and diagnose PSC with accuracy.

3.6 Summary

The chapter describes a detailed study design for forecasting wind energy using the stacked-LSATM model with different layers in both directions of forward and backwards input data and optimises with Grid search algorithm, the method of designing solar power forecasting model to discuss the global energy landscape in terms of integrating with renewable sources to highlight the importance of solar energy in power production and the sustainable environment. The methodology to get the most accurate prediction of solar power production depends on machine learning algorithms in the different regions considering the self-referencing data provided with the solar power plant types of equipment the model considers exogenous and endogenous data sets and how affecting the forecasting result, in the next chapter will be more discussion about model using in predicting the most accurate solar power forecasting techniques like model optimising, validates, visualisation and performance.

Chapter 4
Model Discussion

Integrating solar power into electrical grids has become increasingly critical in the global transition toward renewable energy sources. However, solar power generation's inherent intermittency and weather-dependent nature present significant challenges for grid stability and energy management. Accurate forecasting of solar power production is essential for optimal grid operation, energy trading, and maintaining power quality. While various forecasting methods have been proposed, the complexity of solar power generation patterns, influenced by multiple environmental and technical factors, necessitates more sophisticated approaches

4.1 introduction

This research presents an innovative hybrid methodology that combines Long ShortTerm Memory (LSTM) neural networks with Maximum Power Point Tracking (MPPT) algorithms to enhance solar power forecasting accuracy. The study utilizes operational data from two distinct solar power plants, incorporating both historical production patterns and real-time equipment performance metrics. The LSTM component leverages its capability to capture long-term dependencies in time series data, while the MPPT integration provides crucial system-specific optimization parameters through self-referencing equipment data.

The proposed ensemble approach addresses several key limitations of existing forecasting methods. First, it incorporates the dynamic efficiency characteristics of solar power equipment through MPPT data, which traditional forecasting models often overlook. Second, by utilizing self-referencing data from actual plant equipment, the model adapts to system-specific characteristics and degradation patterns. Third, the dual-plant dataset enables the validation of the model's generalizability across different installation configurations and geographical locations.

The integration of MPPT algorithms with LSTM-based forecasting represents a significant advancement in solar power prediction methodology. Unlike conventional approaches that treat forecasting and power optimization as separate processes, this research demonstrates how their combination can lead to more accurate and applicable predictions. The methodology not only considers historical weather patterns and solar irradiance but also accounts for the real-world performance characteristics of solar power equipment, resulting in a more robust and reliable forecasting system. The research aims to evaluate the effectiveness of this hybrid approach in improving solar power production forecasting accuracy, with particular emphasis on its practical implications for grid operators and power plant managers. The research also explores the scalability of the proposed method and its potential application in diverse solar power installations.

4.2 LSTM ensemble method

photovoltaic (PV) technology is the most immediate way to convert solar radiation into electricity. Nevertheless, PV power output is affected by several factors, such as location, clouds, etc. As PV plants proliferate and represent significant contributors to grid electricity production, it becomes increasingly important to manage their inherent alterability. Therefore, solar PV forecasting is a pivotal factor in supporting reliable and cost-effective grid operation and control. In this paper, a stacked long short-term memory network LSTM, which is a significant component of the deep recurrent neural RNN network, is considered for the prediction of PV power output 1.5 h ahead. Historical data of PV power output from a PV plant in Nicosia, Cyprus, were used as input to the forecasting model

4.3 Solar Power Generation Visualization Analysis

Comparative Analysis of Two Solar Power Plants in Energy Production data set aims to recognise the significance of Solar Power Plants as a crucial investment and energy source for environmental sustainability. This part of the research is to:

- Variability in Production: Depending on the equipment, location, and weather, each plant's capacity to produce energy may vary.
- Examining Energy Production: The objective is to examine the data related to energy production as well as the environmental factors that affect this output from two different solar power plants.
- Investment Direction: Using this comparison will help you allocate funds and maintenance tasks more efficiently.
- Performance Under Conditions: Learn how both plants perform in different weather conditions and use this information to adjust energy production plans. As shown in figure 4.1

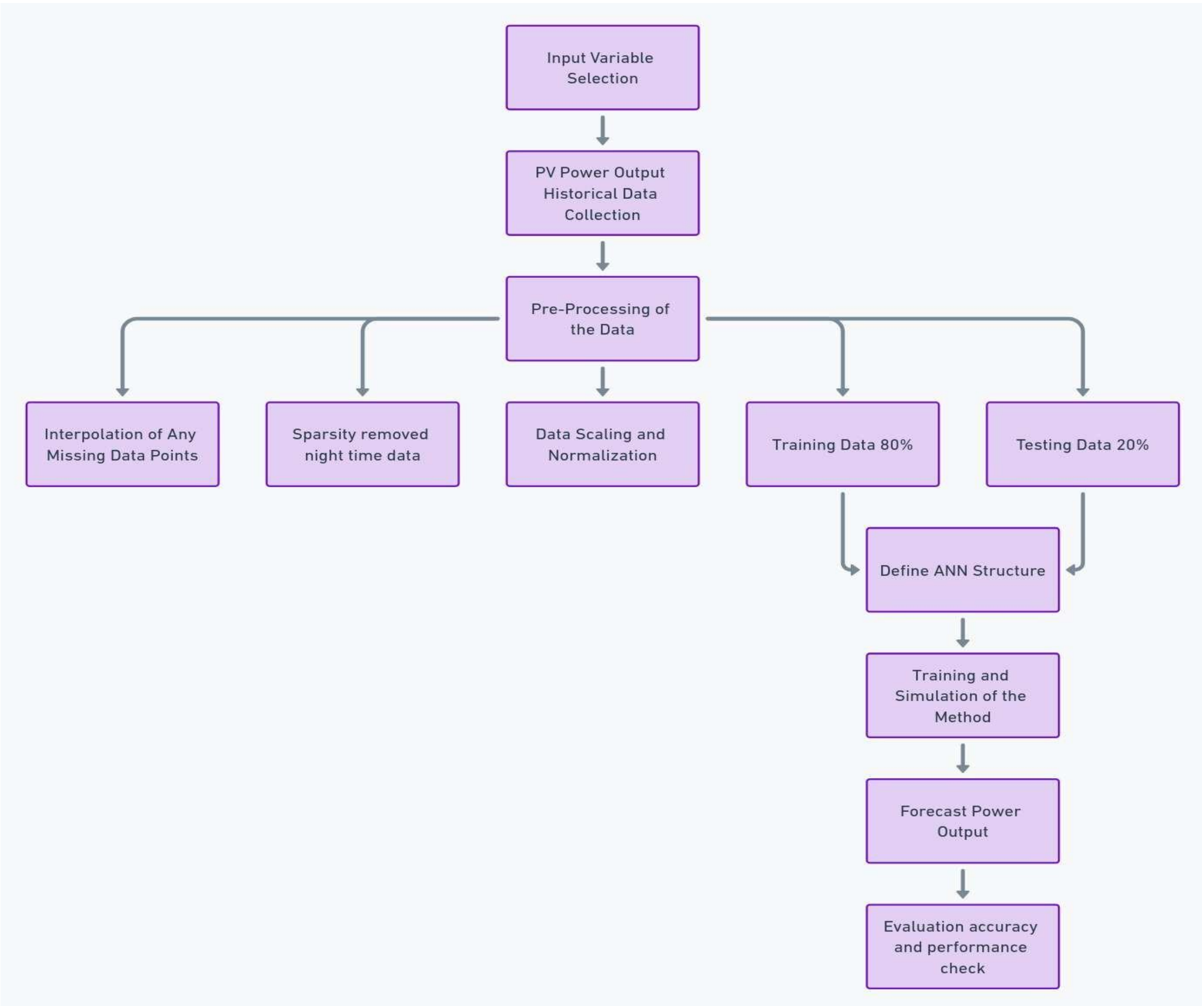


Figure 4.1: Flowchart of the proposed photovoltaic (PV) power forecasting method

4.5 Data Set

Currently, the public can access vast amounts of power output data via several websites, including PVGIS, PVsyst, and Energy Plan an electric utility upon request. Data for the study were gathered via Aurora Vision, a web-based technology that lets users control their PV plants from a distance. The information gathered consists of observations of the Middle Eastern system's solar power production (measured in watts). These data points, which show the value of solar power output for 15 minutes, make up the time series. Since the goal is to use only endogenous data, extra data regarding solar irradiance and other climatic variables like cloud cover, wind speed, and direction were not taken into consideration. The source key identifier for inverters, the date and time of data recording, DC Power & AC Power for Metrics indicating the direct and alternating current outputs, and (Daily Yield & Total Yield) values capturing both daily and cumulative energy generation are all included in the data set used in this section of the research. Most data used Ambient Temperature & Module Temperature Indicators of environmental and equipment conditions. Irradiation: Reflecting the sunlight received by the panels. Shown in Figure 4.2, the table shows solar power plant data, typically used for monitoring and analysis of energy generation performance. A breakdown of each column and what it represents in the context of solar power plant operations. The date and time when the data was recorded. An identifier for the solar power plant. This helps distinguish data from different plants, especially in datasets containing multiple sites. the identifier for each inverter within the plant. Inverters convert DC power from solar panels to AC power for the grid and tracking performance monitoring at the inverter level. The amount of direct current (DC) power generated by the inverter at that time. The amount of alternating current (AC) power output by the inverter is the usable electricity sent to the grid. Daily performance tracking. The total cumulative energy produced by the inverter since installation or a reset point.

Out[6]:

DATE_TIME	PLANT_ID	SOURCE_KEY	DC_POWER	AC_POWER	DAILY_YIELD	TOTAL_YIELD
2020-05-15	4136001	4UPUqMRk7TRMgml	0.0	0.0	9425.000000	2.429011e+06
2020-05-15	4136001	81aHJ1q11NBPMrL	0.0	0.0	0.000000	1.215279e+09
2020-05-15	4136001	9kRcWv60rDACzjR	0.0	0.0	3075.333333	2.247720e+09
2020-05-15	4136001	Et9kgGMDl729KT4	0.0	0.0	269.933333	1.704250e+06
2020-05-15	4136001	IQ2d7wF4YD8zU1Q	0.0	0.0	3177.000000	1.994153e+07

Figure 4.2: The daily production of solar energy depending on the data visualisation

In Figure 4.3 below is a graph of solar power generation over time. The graph is divided into 34 subplots, with each subplot representing a different day. The x-axis of the graph shows the time of day, and the y-axis shows the amount of solar power generated in kilowatts (kW). The graph shows that the solar power generation is highest in the middle of the day, when the sun is at its highest point in the sky. The graph also shows that there is some variability in the solar power generation from day to day. This is likely due to several factors, such as the weather conditions and the amount of cloud cover.

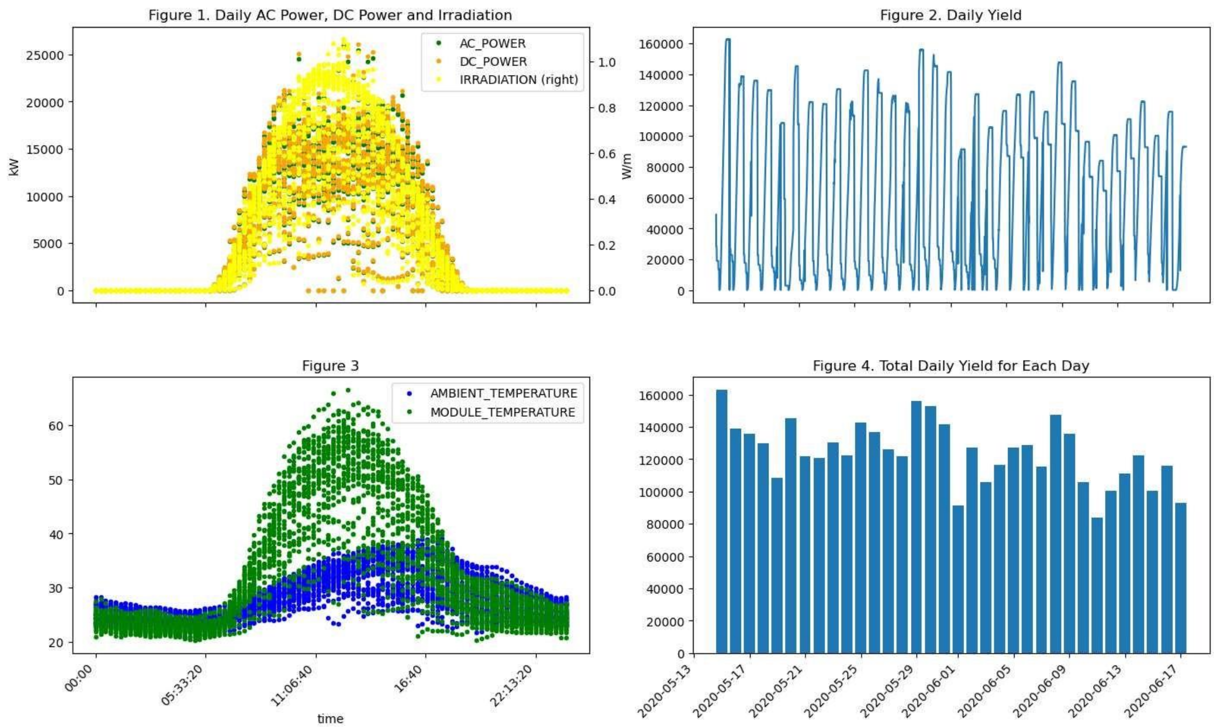


Figure 4.3: the solar power production depending on the irradiance and temperature

4.6 Data Pre-Processing

In the context of a solar power plant, pre-processing data refers to the actions made to get raw data ready for modelling, analysis, or machine learning algorithms. This procedure is necessary since raw data which is obtained from sensors, weather stations, and other monitoring systems often contains noise, missing numbers, and discrepancies. Good pre-processing raises the data's quality, which makes forecasting and optimization more precise.

Because of any spikes or non-stationary components in the forecasting models' input data, the PV power production model is improperly trained, which will result in a significant prediction error. These problems are invariably encountered since most models rely on historical PV-power output data and meteorological data, both of which are subject to change and unpredictability due to weather. Pre-processing the input data can thereby lower the computing cost and inappropriate training problems, significantly increasing the model's accuracy. The training and testing procedures, together with the data pre-processing methods used for sparsity, missing values, and feature scaling, are described in the next subsections. Pre-processing the input data can thereby lower the computing cost and inappropriate training problems, significantly increasing the model's accuracy. The training and testing procedures, together with the data pre-processing methods used for sparsity, missing values, and feature scaling, are described in the next subsections. Pre-processing the input data can thereby lower the computing cost and inappropriate training problems, significantly increasing the model's accuracy. The training and testing procedures, together with the data pre-processing methods used for sparsity, missing values, and feature scaling, are described in the next subsections. Pre-processing the input data can thereby lower the computing cost and inappropriate training problems, significantly increasing the model's accuracy. The training and testing procedures, together with the data pre-processing methods used for sparsity, missing values, and feature scaling, are described in the next subsections. Pre-processing the input data can thereby lower the computing cost and inappropriate training problems, significantly increasing the model's accuracy. The training and testing procedures, together with the data pre-processing methods used for sparsity, missing values, and feature scaling, are described in the next subsections. As shown in Figure 4.1

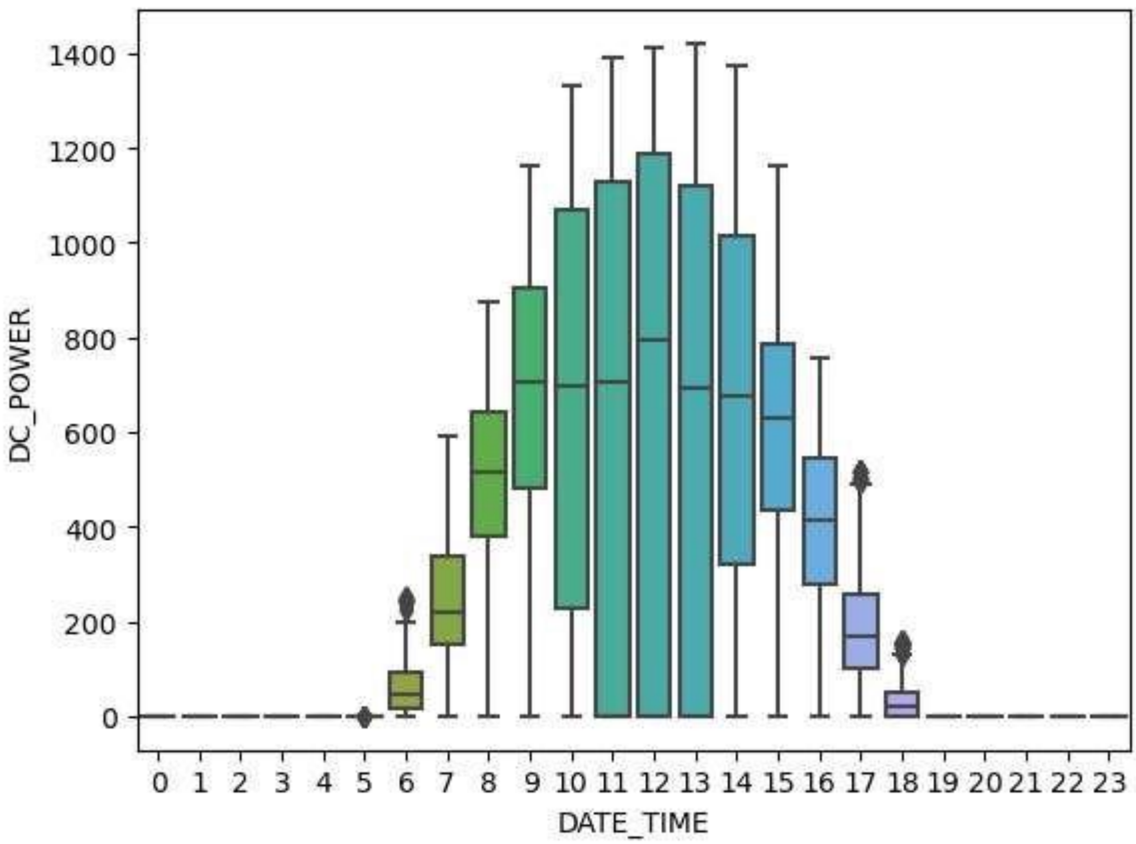


Figure 4.4: DC solar power production (kwh) depending on endogenous and exogenous variables in 24 hours.

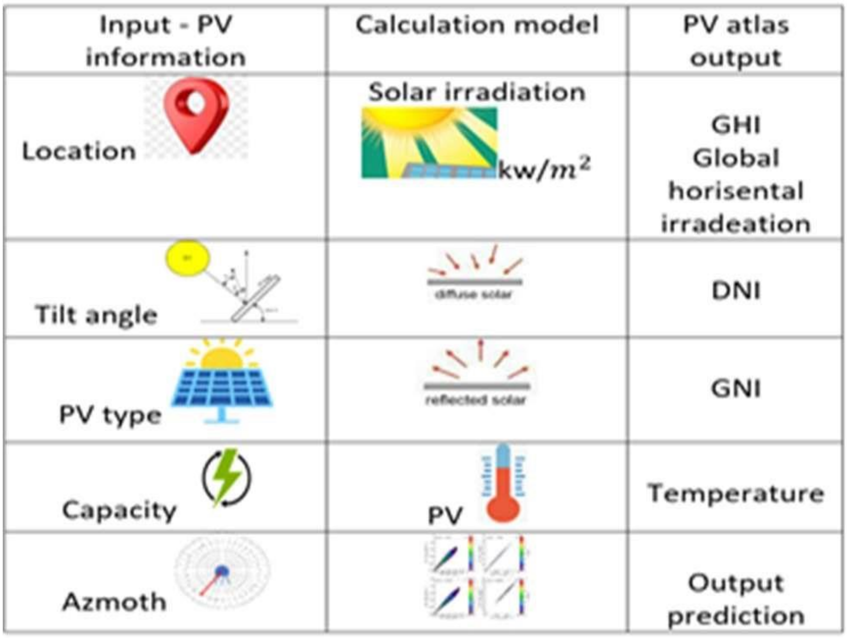
4.7 Training and Testing Groups

Without supposing a particular model relationship, machine learning is capable of learning on its own and producing precise predictions. Therefore, the original time series was divided into two groups: a training group and a testing group, to assess the performance of our model and confirm how well it functions on unseen data. While the testing set was used to test the model, the first one was used to train it. Although the testing set uses the same algorithm as the training group, it is not dependent on the training dataset. Only the testing set is used to calculate the error metrics. The division into these two groups would have often occurred at random. For time series, however, a random subset would not be representative. As a result, the model was tested using the supplemental data from the last study and trained using a predetermined percentage of the initial data in this work. Specifically, the training set included the first 80% of the observations, while the testing set included the remaining data.

4.8 LSTM as ensemble methods (model one)

Transition in energy systems for efficiency improvements, at a higher level with the target of reducing the impact of climate change. Investing in solar energy that is approved by the global scientific community is obligatory. The uncertainty associated with photovoltaic (PV) systems is one of the core obstacles that hinder their seamless integration into power systems. This project focuses on machine learning forecasting algorithms for solar power generation using the Long Short-Term Memory (LSTM) algorithm, a type of recurrent neural network (RNN). To maximize the efficiency and reliability of solar energy systems, accurate prediction models are crucial with high PV penetrations. The LSTM architecture's ability to capture temporal dependencies makes it well-suited for time series forecasting tasks such as solar irradiance prediction. Using ensemble LSTM to improve the output accuracy and reduce mean square error through solar energy plant production. Temporal Variation of Solar Power Production using SARIMA / ARIMA The volatility of PV power generation results in some problems of the grid: Frequency instability, dispatch difficulties and surge in current/voltage.

Using the model of two LSTMs as an ensemble method one model with 80 dropping layers and the second with 50 dropping layers to get the most accurate prediction and prevent data overfitting as shown in Figure 4.5 The validity of the forecasting method is demonstrated by applying it to the power production of a real PV power plant. The time horizon of the prediction, the location, the data that is available, and the required level of accuracy all influence the forecasting method selection. It is a widespread practice to use an ensemble or hybrid technique that combines several methodologies to improve forecasting performance overall. The variables chosen in this model are represented in Figure 4.5.



Input - PV information	Calculation model	PV atlas output
Location	Solar irradiation kw/m^2	GHI Global horisental irradeation
Tilt angle	diffuse solar	DNI
PV type	reflected solar	GNI
Capacity	PV	Temperature
Azmoth		Output prediction

Figure 4.5: the variables that affect solar power production

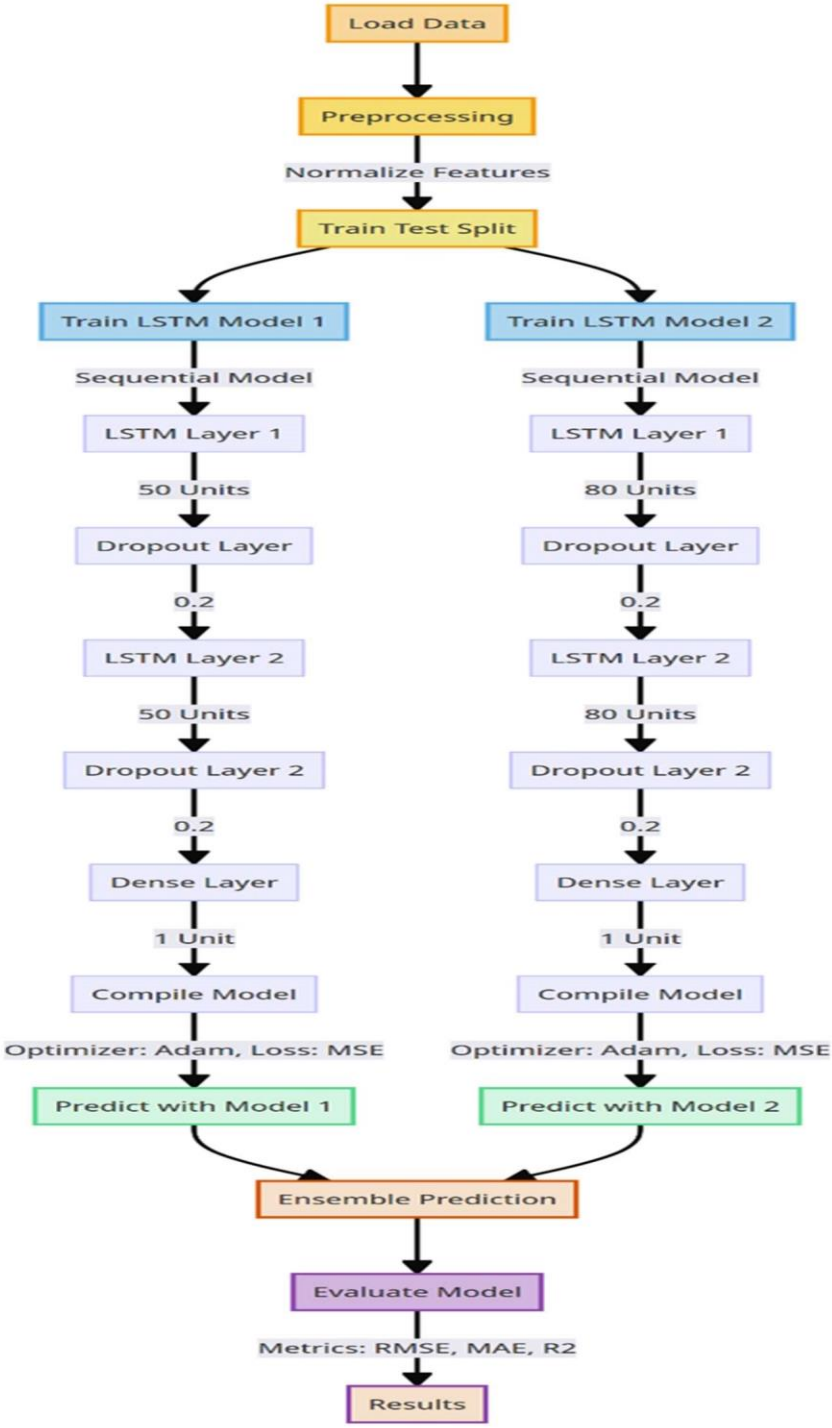


Figure 4.6: Two models of LSTM one with 80 dropping layers and one with 50 dropping layers to prevent overfitting dataset

4.9 Adaptive Moment Estimation ADAM Optimiser with LSTM

A common option for training Long Short-Term Memory (LSTM) networks a kind of recurrent neural network (RNN) intended to simulate sequential data is the ADAM (Adaptive Moment Estimation) optimizer. Because LSTMs can capture long-term dependencies, they are especially useful for tasks involving speech recognition, natural language processing, and time series prediction. Key Features of Using ADAM with LSTMs:

- Adaptive Learning Rates: Because LSTM training is complicated and frequently noisy, ADAM dynamically modifies the learning rate for each parameter based on the first and second moments of the gradients.
- Faster Convergence: ADAM is ideally suited for the usually drawn-out training procedures of LSTMs because of its ability to converge more quickly than classic optimizers thanks to the combination of momentum and adaptive learning rates.
- Bias Correction: For LSTM structures to effectively learn long-term dependencies, bias-correction methods are incorporated into ADAM to increase training stability, particularly in the initial stages.

By utilising these characteristics, ADAM improves the effectiveness of LSTM network training and overall performance across a range of applications, including predictive analytics and language modelling. The accuracy result is shown in Figure 4.7.

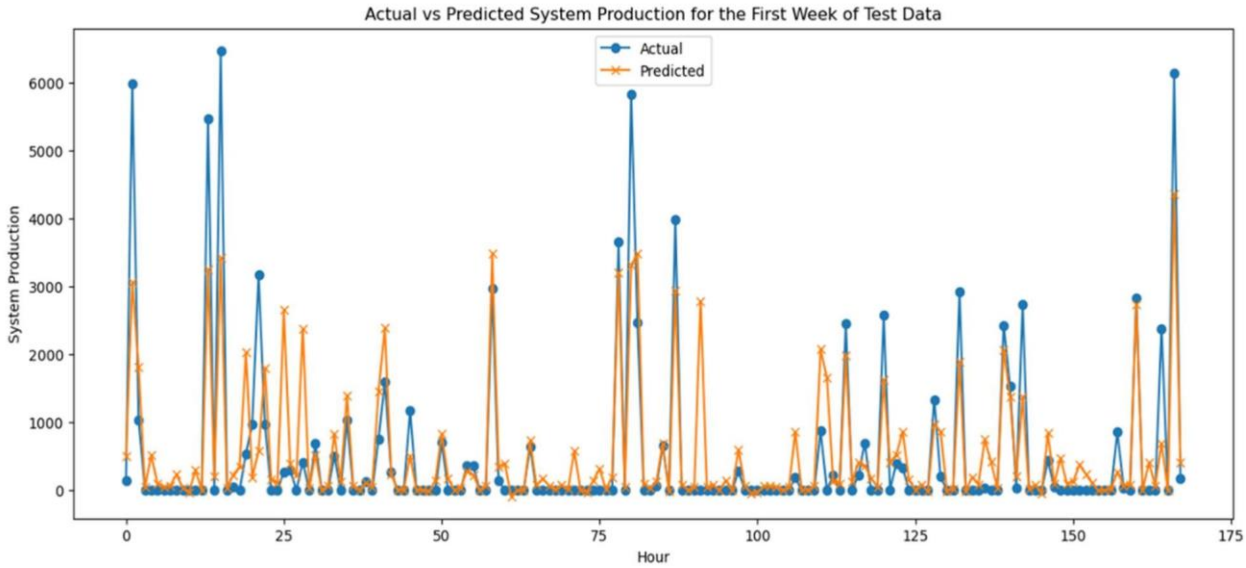


Figure 4.7: solar energy production in actual and predicted for one week.

(Root Mean Squared Error (RMSE): 0.11049111870458088

Mean Absolute Error (MAE): 0.05512441745205155

R2 Score: 0.634048582919565)

4.10 ADAM Mathematical Formulation

The updates in the ADAM optimizer are given by the following equations:

Initialize variables:

$m \quad (first\ moment\ vector)$

$v \quad (second\ moment\ vector)$

t = 0 (time step)

At each time step t

- Increment time step: t = t+1
- Compute gradients: gt (gradient of the loss function concerning parameters) · Update biased first-moment estimate:
- $m \qquad + (\beta - 1)gt$
- Efficiency: Requires little memory and is computationally efficient.
- Robustness: Works well in practice and is robust to noisy gradients.
- Low Tuning: Often performs well with default parameters.

An optimization algorithm ADAM optimizer is used to train both linear and nonlinear machine learning models. By fitting a linear equation to the observed data, the Linear Regression approach can be used to model the connection between a dependent variable and one or more independent variables. The function is of the form:

$$\gamma = \beta + \beta \times$$

To train nonlinear models, such as neural networks including LSTMs, which may reflect intricate, nonlinear connections in data, the ADAM optimizer is frequently utilised in training parameter optimization. This will be a statistical model used to capture the linear interdependencies among multiple time series. It generalizes the univariate autoregressive (AR) model to multivariate time series data. Using many time series variables at once, Vector Auto Regression (VAR) approaches enable the analysis of their correlations. One of VAR's primary features is that every variable in the system is represented as a linear function of both its own and other variables' delayed values. by first analysing and extracting relevant features from a range of time series, and then feeding those features into an LSTM model, VAR may be utilised in feature engineering in conjunction with LSTM. This can aid in comprehending how the variables relate to one another. Figure 4.8 shows the plot result when dealing with univariant and multivariant data sets.

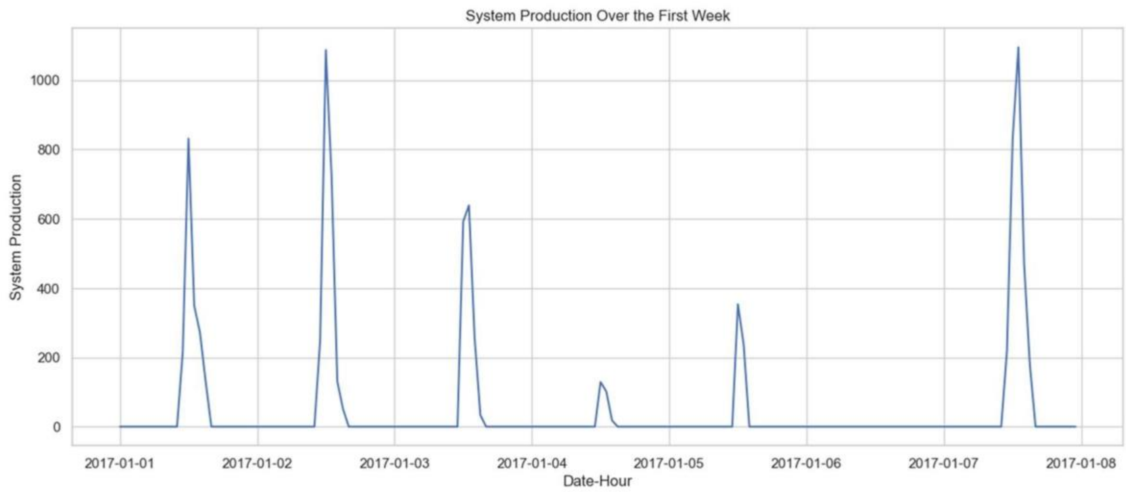


Figure 4.8: system prediction for one week

4.11 Temporal Variation of Solar Power Production using SARIMA / ARIMA

Accurately forecasting solar energy output is crucial to maximising solar power generation. The ARIMA (Autoregressive Integrated Moving Average) approach is used in this study to predict solar power generation by utilising its ability to effectively analyse time series data. As an optimizer, the Augmented Dickey-Fuller (ADF) test is used to verify stationarity in the solar power data, ensuring the model's robustness. It can improve the accuracy of the ARIMA predictions by verifying that the time series is stationary. In addition to helping to comprehend the temporal fluctuations in solar production, this method facilitates improved energy management and planning decisions in the renewable energy industry. Predicting solar energy output accurately is essential to the goal of maximising solar power generation. The ARIMA (Autoregressive Integrated Moving Average) approach is used in this work to anticipate solar power generation, making use of its capacity to efficiently analyse time series data. As an optimizer, the Augmented Dickey-Fuller (ADF) test is used to verify stationarity in the solar power data, ensuring the model's robustness. We improve the accuracy of the ARIMA predictions by verifying that the time series is stationary. In addition to helping to comprehend the temporal fluctuations in solar production, this method facilitates improved energy management and planning decisions in the renewable energy industry.

ARIMA (Autoregressive Integrated Moving Average) is a popular statistical method for analyzing and forecasting time series data. It is particularly useful in understanding temporal variations in solar power production, which can be influenced by numerous factors like weather conditions, seasonal changes, and technological advancements. The plot result of (System Production Over Time) illustrates the temporal variation of solar power production throughout a year shown in Figure 4.9& 4.10, it illustrates the

X-axis (Date-Hour). This represents the time, in hourly increments, spanning from the beginning of 2017 to the start of 2018. Y-axis (System Production). This represents the amount of solar power produced, in kilowatts (kW) or a similar unit.

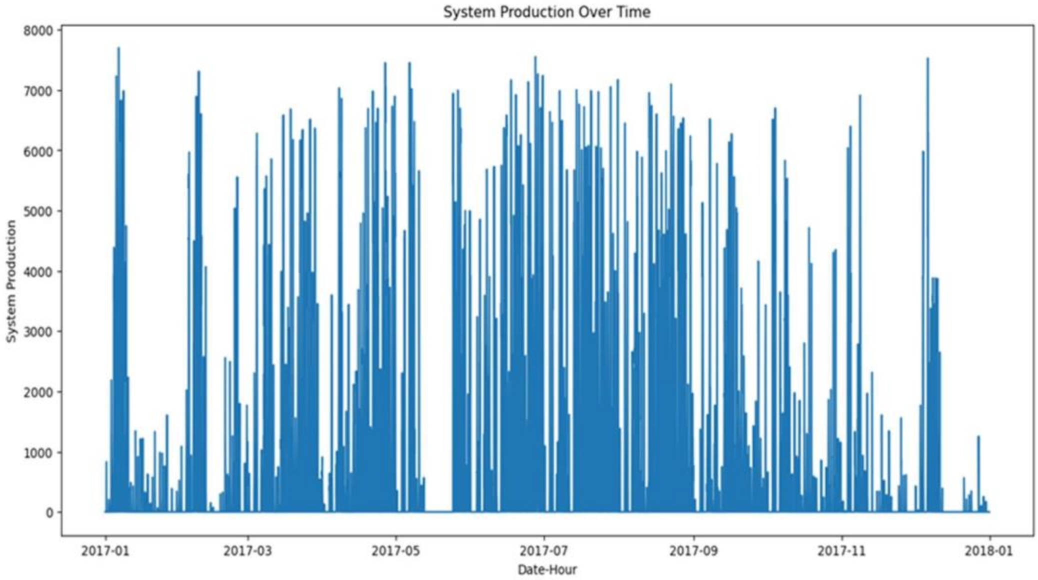


Figure 4.9:Auto regresion intgrated moving average (ARIMA) solar energy output (kwh) over time

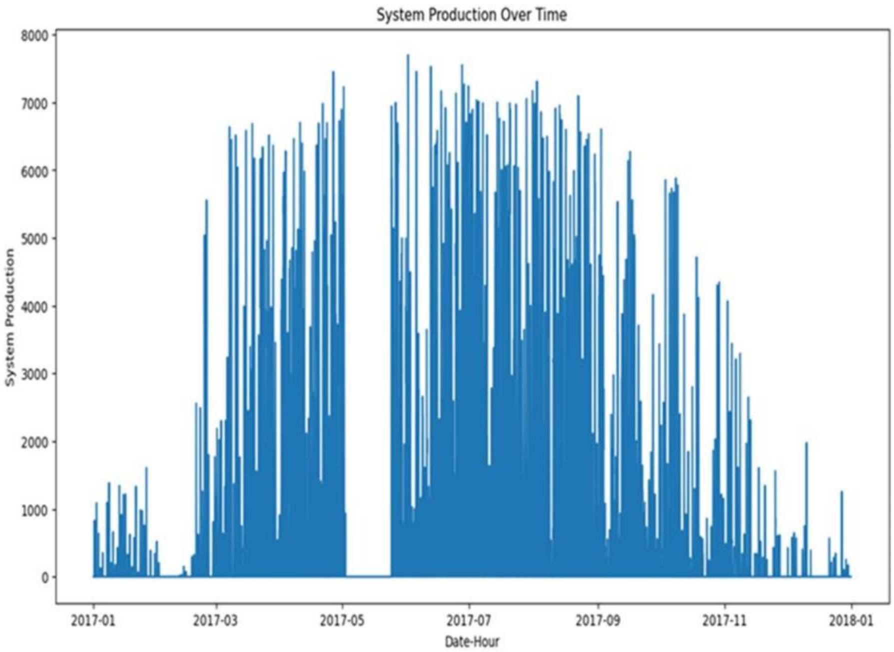


Figure 4.10: Sesonal auto regration intgrated moving averge (SARIMA) solar energy output over time

A key component of ARIMA is Autoregressive (AR) uses the dependency between an observation and several lagged observations. Integrated represents the differencing of raw observations to make the time series stationary. Moving Average (MA) Models the dependency between an observation and a residual error from a moving average model.

A stationarity time series is a statistical property that does not change over time. Important for applying ARIMA, as it assumes stationary data. Steps to Analyse Solar Power Production with ARIMA.

Data Collection by gathering historical solar power production data from reliable sources, including relevant variables like weather data (temperature, cloud cover, etc.). Pre-processing data handles missing data and outliers. Convert the data to a stationary series by differencing if necessary.

Seasonal variations production is higher during summer months with longer days and more direct sunlight, explaining the higher peaks and troughs in the middle of the year. Weather patterns Cloudy days would result in lower production dips even during daylight hours, contributing to the irregular pattern within the daily cycles. Analysing this temporal variation is crucial for understanding the reliability and predictability of solar power. This is where ARIMA or SARIMA models come in.

These statistical models are used to analyse and forecast time series data, like solar power production data, by capturing patterns, trends, and seasonality.

4.12 Model Identification, Fitting and Validation

Use tools like the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) to determine the order of AR and MA components. Identify the differencing order (d) needed for stationarity. Fit the ARIMA model to the data using the identified parameters (p, d, q). Evaluate the model's performance using metrics like the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Validation model by splitting the data into training and testing sets. Validate the model by comparing predicted values to actual observed values in the test set to forecast future solar power production.

By leveraging the temporal dependencies inherent in solar power production, ARIMA/SARIMA models can provide valuable insights for grid management, energy trading, and optimizing solar energy systems. Figure 4.11 shows an autocorrelation plot (ACF) on the left and a partial autocorrelation plot (PACF) on the right. These plots are commonly used in time series analysis to identify patterns and dependencies within the data.

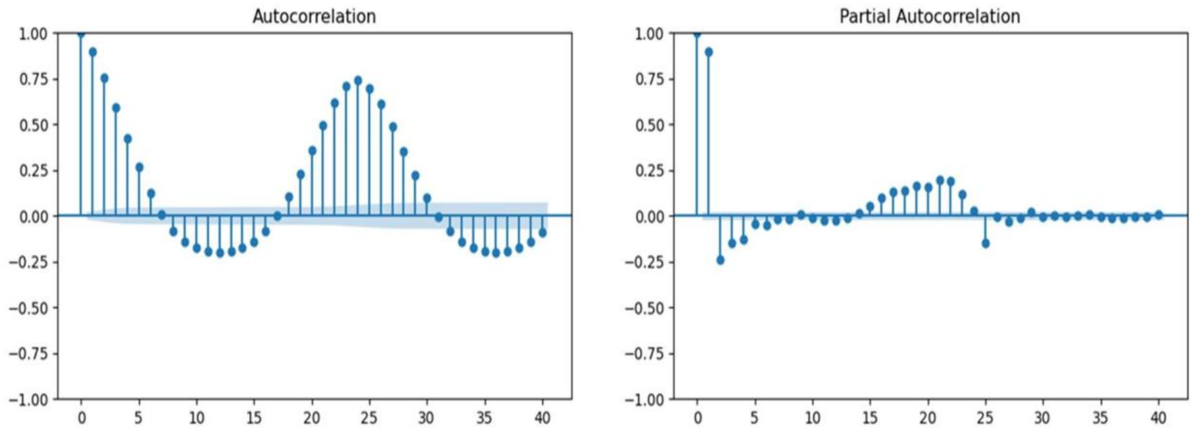


Figure 4.11: shows an autocorrelation plot (ACF) on the left and a partial

X-axis: Represents the time lag, indicating the correlation between data points separated by a specific number of periods.

Y-axis: Shows the correlation coefficient, ranging from -1 to 1. A value of 1 indicates a perfect positive correlation, -1 perfect negative correlation, and 0 no correlation. Blue Stems: Each stem represents the correlation at a specific lag. The height of the stem indicates the strength of the correlation.

Blue Shaded Area: Represents the confidence interval. Stems exceeding this area are considered statistically significant correlations.

4.13 Discussion

Interpretation of the PACF plot shows significant spikes at lags 1, 2, and 20. The spike at lag 1 confirms the strong direct correlation. The smaller spike at lag 2 suggests a possible additional short-term dependency. The spike at lag 20 reinforces the presence of the seasonal pattern observed in the ACF plot.

These plots suggest that the underlying time series data exhibits both short-term and seasonal dependencies. This information is crucial for selecting and fitting appropriate time series models, such as ARIMA or SARIMA, which can capture these patterns and make accurate forecasts.

4.14 The Heatmap Dataset Correlation of LSTM

In data analysis, understanding the relationships between variables is crucial for drawing meaningful insights and making informed decisions. One effective way to visualize these relationships is through a correlation heatmap, highlighting the strength and direction of linear associations between multiple variables in a dataset. correlation heatmap, can identify patterns, uncover underlying relationships, and discover potential multicollinearity issues that may affect predictive modelling.

The following correlation heatmap in figure 4.7 visualizes the relationships between different variables in the dataset. Each cell in the heatmap represents the correlation coefficient between two variables, indicating the strength and direction of their linear relationship. “Correlation Coefficient values range from -1 to . As shown in table 4.1.

Correlation Coefficients values	Relationship with variables
1: Perfect positive correlation	as one variable increases, the other also increases
-1: Perfect negative correlation	as one variable increases, the other decreases
0: No correlation	no linear relationship between the variables

Table 4.1: Correlation coefficient values range from -1 to 1

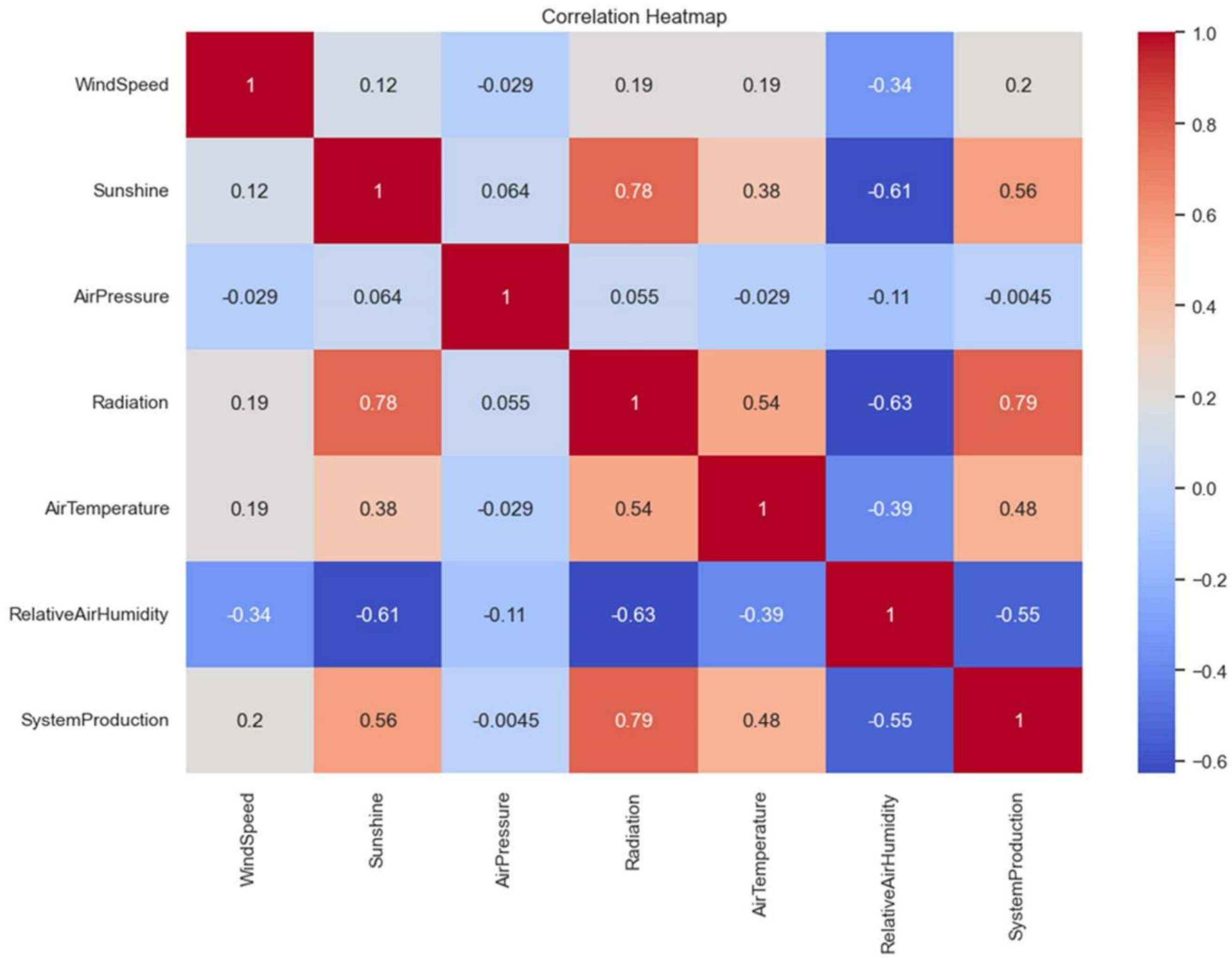


Figure 4.12: The heatmap uses a colour gradient to represent correlation values:

- Dark Blue: Indicates strong negative correlation.
- Light Blue: Indicates weak negative correlation.
- White: Indicates no correlation.
- Light Red: Indicates weak positive correlation.
- Dark Red: Indicates strong positive correlation.

This visualization helps identify which variables are closely related and can inform further analysis or model selection. The correlation heatmap provides valuable insights into the relationships among the variables in the dataset. Key observations include:

- Strong Positive Correlations: Identify pairs of variables with high positive correlation, which may indicate that they influence each other or share common factors.
- Strong Negative Correlations: Note pairs of variables with high negative correlation, suggesting that as one increases, the other tends to decrease.
  This can be important for identifying potential trade-offs.

- Weak or No Correlation: Variables with coefficients close to zero suggest that changes in one variable do not significantly impact the other, which may indicate independent behaviour.

4.14 Discussion

Understanding these correlations is essential for feature selection, multicollinearity checks, and improving predictive modelling efforts. Use this heatmap as a foundational tool for further exploratory data analysis and hypothesis testing. The heatmap will be accompanied by detailed explanations that describe how to interpret the correlation coefficients, as well as key observations regarding strong and weak correlations. Through this analysis, we aim to establish a foundational understanding that will guide subsequent exploratory data analysis and inform model selection processes.

4.15 Conclusion

For model one in chapter 4: In this analysis, we applied two Long Short-Term Memory (LSTM) models to the same set of variables, leveraging the insights gained from the heatmap. The relationships highlighted in the heatmap guided our feature selection process, allowing us to focus on the most influential variables that contribute to predictive performance. Model performance of the comparative results from the two LSTMs revealed how varying model architectures and hyperparameters can impact the ability to capture temporal dependencies and complex patterns within the data. the Insights gained by correlating the performance metrics of both models with the underlying relationships identified in the heatmap can further understand how specific variables influence the predictions. The two LSTM model analyses have enriched our understanding of the dataset, enabling us to make data-driven decisions that enhance our predictive modelling efforts. The next model will involve fine-tuning the models based on these insights and exploring additional features that could improve accuracy.

4.6 Model Tow: Bi-LSTM (Fine Tuning) to Enhancing the Maximum Power Point Tracking

4.6.1 Introduction

Fine-tuning refers to the process of making small adjustments to a model's parameters or architecture to improve its performance on a specific task. In the context of machine learning and deep learning, including models like Long Short-Term Memory (LSTM) networks, fine-tuning is often performed after the initial training phase and can include hyperparameter optimization through adjusting hyperparameters such as learning rate, batch size, number of epochs, and dropout rates to find the most effective configuration for the model. The purpose of Fine-Tuning is to Improve accuracy and enhance the model's performance metrics.

4.6.2 Model Architecture Adjustments

Modifying the structure of the model, such as adding or removing layers, changing the number of units in a layer, or altering activation functions to better capture the underlying data patterns. Specific Data the model to better adapt to the unique characteristics of the dataset being used, leading to more effective predictions. It can also lead to efficiency in training to faster convergence during the training process when the model is better tailored to the data. Transfer learning utilizing a pretrained model and adapting it to a new, but related, task by fine-tuning its weights on a smaller, task-specific dataset. The regularization techniques by implementing techniques like L1/L2 regularization or dropout to prevent overfitting and improve generalization to unseen data. Identifying and selecting the most relevant features based on insights from exploratory data analysis (such as the correlation heatmap) to enhance model performance and produce a critical step in the machine learning workflow, aimed at optimizing model performance through careful adjustments and enhancements.

4.6.3 Maximum power point of solar power plant

Maximizing the power output of solar panels is inherently challenging due to variations in environmental conditions such as sunlight intensity and temperature. Maximum Power Point Tracking (MPPT) algorithms are instrumental in optimizing solar panel performance by continuously adjusting the operating point to extract maximum power from the solar array.

Traditional MPPT techniques, including Perturb and Observe (P&O) [4], Incremental Conductance (Inc Cond) [5], and Hill Climbing (HC) [6], have been widely employed for this purpose. However, these methods have inherent limitations in terms of accuracy, efficiency, and adaptability to changing environmental conditions [7]. This model aims to investigate the efficacy of LSTM-based MPPT controllers for enhancing solar panel efficiency. The model provides a comprehensive review of related works, including traditional MPPT techniques and recent advancements in ML-based approaches. The plotting result of the model outlines the methods and materials employed in the research, including data collection, LSTM architecture, and training procedures to present the experimental results, comparing the performance of LSTM-based MPPT controllers with conventional algorithms. On Figur 4.13 below represents the data that be visualized more clearly in a graph. Graphing the current and voltage creates a curve that is referred to as an I-V curve. The blue line in the graph is an I-V curve. The current is plotted in amps (A) on the left y-axis. The voltage is plotted in volts (V) on the x-axis On the same graph, the power for each current-voltage combination is plotted in pink. The power is plotted in watts (W) on the right y-axis. This power curve clearly shows the maximum power point. A red line identifies the voltage and current associated with the maximum power point.

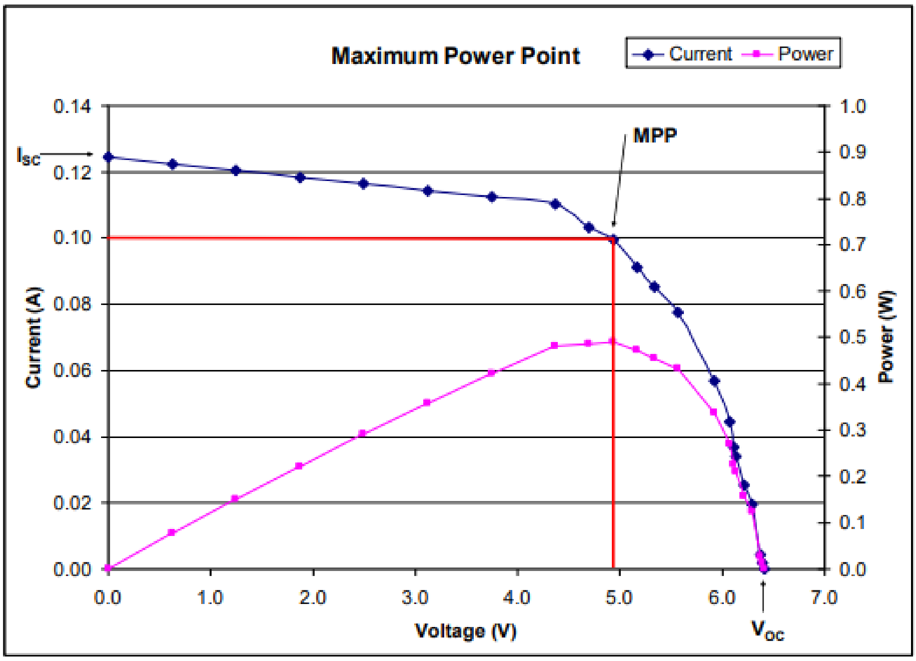


Figure 4.13: The daily solar power generation (AC & DC) based on irradiance, ambient temperature, and model temperature

4.6.4 Visualisation Data

Data collection is a crucial step in training and validating LSTM-based MPPT controllers. To ensure comprehensive coverage of environmental conditions, data is collected from various solar power plants located across different regions. The collected dataset includes measurements of voltage, current, and power see table 4.1 output of solar panels, recorded at regular intervals. Additionally, external factors such as temperature and solar irradiance are incorporated into the dataset to capture their influence on solar panel performance. By gathering data from diverse geographical locations and environmental conditions, the dataset provides a comprehensive basis for training and testing LSTM models for MPPT control. A sample of the collected data is shown in Table 4.2

Trial #	Collected Data		Calculated
	Voltage (V)	Current (A)	Power (W)
1	0.000	0.124	0.000
2	0.624	0.122	0.076
3	1.248	0.120	0.150
4	1.872	0.118	0.222
5	2.497	0.116	0.291
6	3.121	0.114	0.357
7	3.745	0.112	0.421
8	4.369	0.110	0.482
9	4.700	0.103	0.485
10 [MPP]	4.934	0.100	0.491
11	5.165	0.091	0.472
12	5.336	0.085	0.455
13	5.566	0.078	0.432
14	5.927	0.057	0.337
15	6.070	0.045	0.270
16	6.113	0.037	0.225
17	6.137	0.034	0.209
18	6.210	0.025	0.156
19	6.294	0.020	0.123
20	6.373	0.004	0.027
21	6.389	0.002	0.013
22	6.407	0.000	0.000

Table 4.2: Data visualisation by collected voltage and current data from PV panel trials

The values in Table 4.1 were obtained by using a potentiometer to vary the resistance in the PV circuit, which directly affects the voltage and current in the circuit. A potentiometer is a small device that changes the resistance with the turn of a knob. The changing resistance affects the overall power output of the panel. In this example, the short circuit current, Isc = 0.124 A (or current when V = 0), and open circuit voltage, Voc = 6.407 V (or voltage when I = 0). The MPP can also be found as the point at which the product of the current and voltage equal the greatest value. The power calculation shows that the MPP has a voltage of VMPP = 4.934, a current of IMPP = 0.100 A, with the power, P = 0.491 W.

4.6.5 Fine-tuning of the LSTM Model and Potential ApplicationsBased MPPT Controllers.

Fine-tuning the LSTM model is an iterative process to improve its performance in MPPT control [18]. This involves adjusting hyperparameters, retraining the model with additional data, and fine-tuning the network architecture to better capture complex patterns in the input data. Between fine-tuning techniques such as grid search and Bayesian optimization, grid search was employed due to its simplicity to systematically explore the hyperparameter space and identify the optimal configuration for the LSTM model. Finetuning the LSTM model is essential for achieving high accuracy and robustness in real-world applications of solar panel systems.

4.6.6 Key Insights and Anomalies Correlation Patterns

The analysis reveals several significant correlations in the solar plant data DC Power and AC Power strong positive correlation observed indicates efficient power conversion in the inverters, further investigation quantifies this correlation using Pearson's correlation coefficient. Power and Irradiation is the high correlation between power output and solar irradiation. It also demonstrates the direct impact of sunlight on energy production. Further investigation of a scatter diagram plot is needed to visualize this relationship between power and temperature.

Correlation noted between power output and both module and ambient temperatures It suggests that temperature influences solar panel efficiency. The analysis of the temperature coefficient of power for the panels is shown in Figure 4.9. The figure shows the correlation coefficients between different variables related to solar power generation.

Each row and column represents a variable such as DC power, AC power, daily yield, total yield, ambient temperature, module temperature, irradiation, sensor number, hours, minutes, and minutes passed. The colour of each cell represents the strength and direction of the correlation between the two variables. The red colour indicates a positive correlation (as one variable increases, the other tends to increase). The darker the red, the stronger the positive correlation. Blue colour indicates a negative correlation (as one variable increases, the other tends to decrease). The darker the blue, the stronger the negative correlation. White colour indicates a very weak or no correlation, ranging from -1 to +1. The 4.14 figure illustrated the (Strong Positive Correlations): (dark red) between: "DC Power and AC Power", "DC Power and Module Temperature", "AC Power and Module Temperature", "Module Temperature and Irradiation". The (Strong Negative Correlations) between Ambient Temperature and Total Yield. (Other Correlations): weaker correlations between different variables.

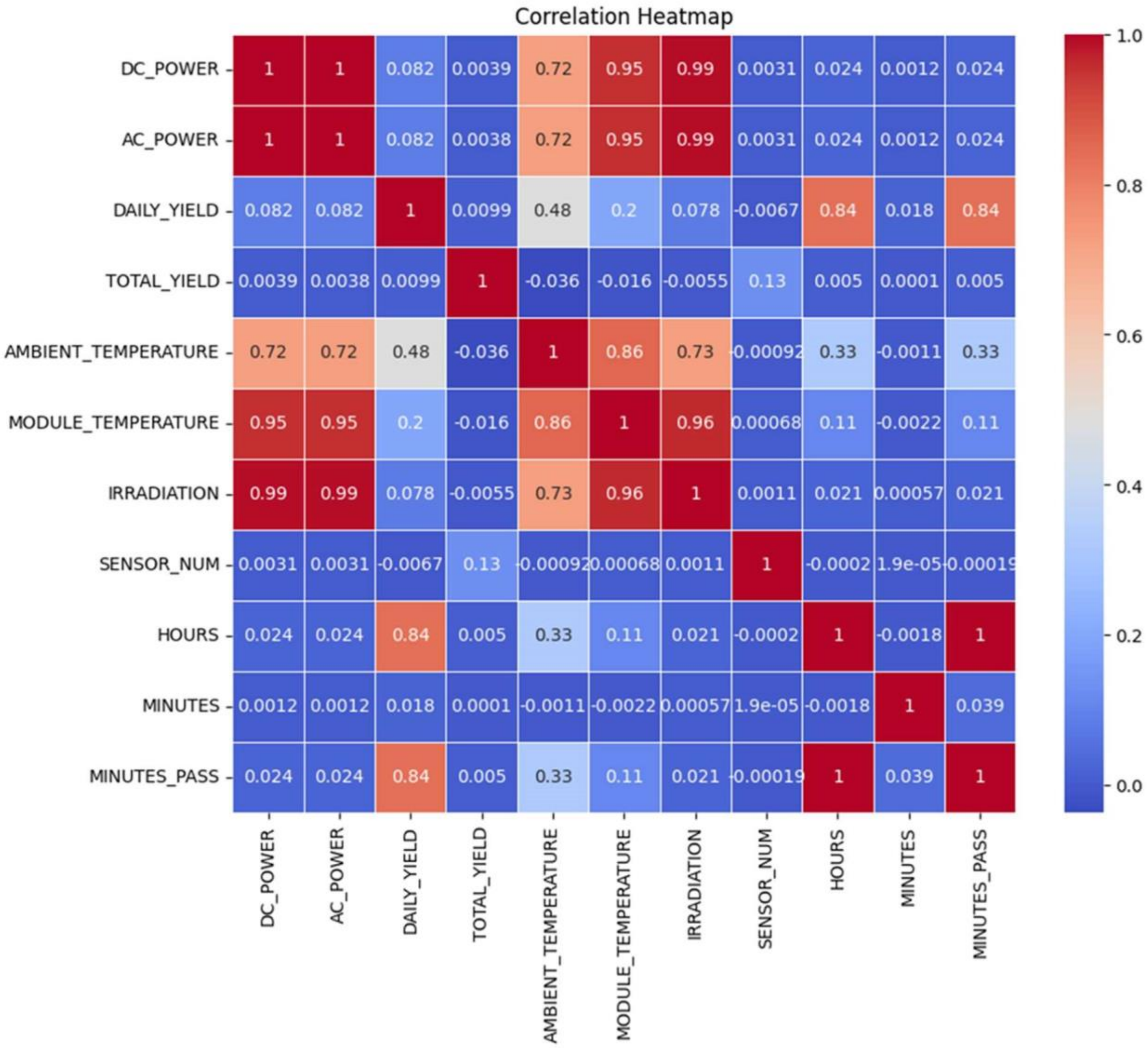


Figure 4.14: correlation function heatmap key observations.

This heatmap is useful for understanding the relationships between varied factors influencing solar power generation. For example, the strong positive correlation between irradiation and module temperature suggests that as sunlight intensity increases, so does the temperature of the solar panels.

Daily Yield and Ambient Temperature, correlation observed, due to increased daylight hours in warmer periods. Further investigation conducted a seasonal analysis to confirm this pattern. Outlier Detection and Implications identified in the data provide valuable insights for maintenance and fault detection.

AC/DC Power vs. Irradiation Outliers:

- indicates potential panel line failures
- Implication: Sunlight present but no power generation suggests faulty photovoltaic cells
- Action item: Implement automated alerts for significant power-irradiation **mismatches**

DC Power vs. AC Power Outliers:

- Rare occurrences point to inverter issues
- Implication: DC power input without expected AC power output suggests inverter malfunction
- Action item: Develop a monitoring system to flag inverters with abnormal DCAC conversion rates
- Equipment Groups and Installation Patterns
- Analysis of TOTAL_YIELD vs. SENSOR_NUM reveals:

4.6.7 Data Quality Concerns

A critical issue has been identified in the DAILY_YIELD data:

- Observation: DAILY_YIELD decreases during some days
- Expected Behaviour: DAILY_YIELD should monotonically increase within a day
- Root Cause: DAILY_YIELD is calculated from measured DC_POWER · Implication: Potential issues in the data generation or collection process **Action Items:**

- Audit the data collection and processing pipeline
- Implement data validation checks to flag impossible DAILY_YIELD decreases
- Investigate the possibility of energy storage or grid feedback affecting measurements

Recommendations for Further Analysis:

- Time Series Decomposition (TSD) analyses trends, seasonality, and residuals in power output data to identify long-term performance degradation and cyclical patterns. Second,

- Predictive Maintenance Model (PMM) develops a machine learning model to predict equipment failures based on identified outliers and patterns.
- Efficiency Analysis:
- Calculate and track conversion efficiency (AC Power / DC Power) over time
- Identify factors contributing to efficiency fluctuations ·

Environmental Impact Study:

- Correlate power output with local weather data (beyond temperature and irradiation)
- Assess the impact of factors like humidity, dust, and wind on plant performance

Comparative Performance Analysis:

If data is available, compare this plant's performance with industry benchmarks or similar plants in the region. By addressing these points and implementing the suggested analyses, it can gain a more comprehensive understanding of the solar plant's performance, improve maintenance strategies, and optimize energy production.

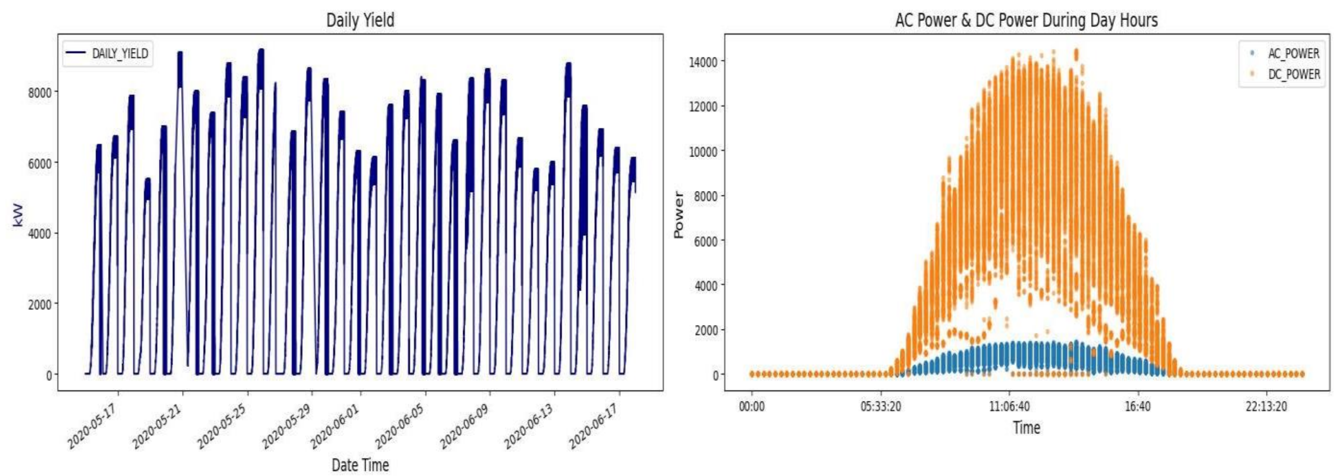


Figure 4.15: Daily Yield and AC Power

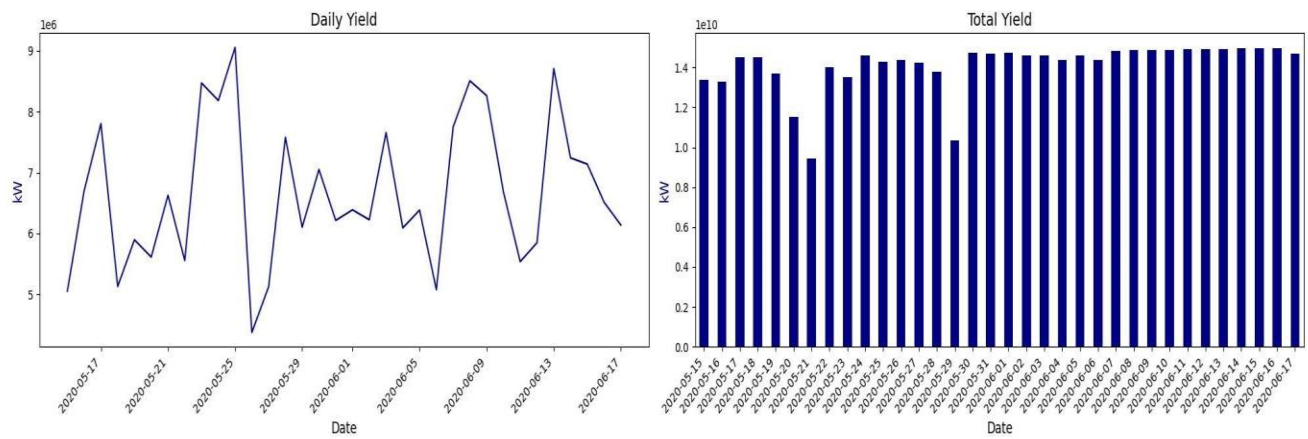


Figure 4.16: Daily and Total Yield

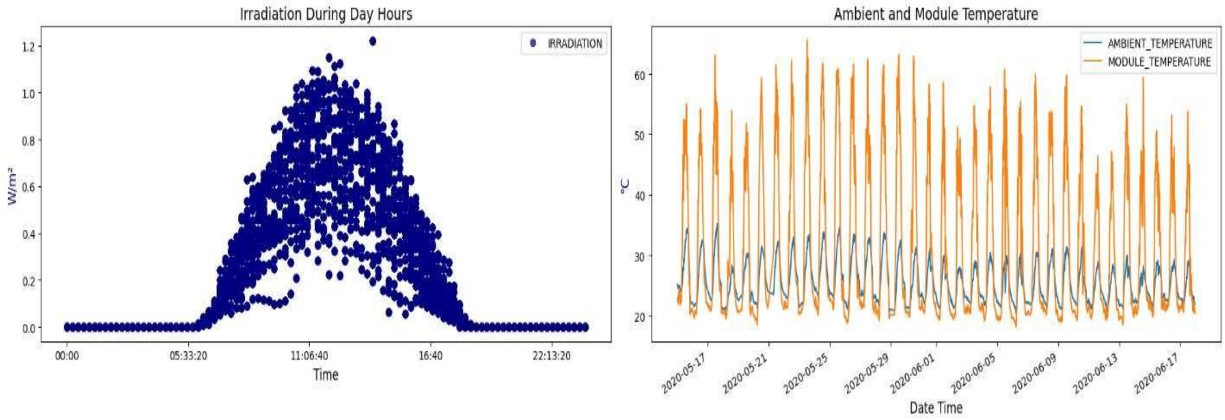


Figure 4.17: Irradiation, Ambient and Module Temperature

It is a Keras Sequential Model, where layers are arranged in a sequence, meaning data flows through each layer one after another. This is common in feedforward networks, where the input moves straight through each layer to the output. As shown in Table 4.3.

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 32)	4,352
dense (Dense)	(None, 8)	264
dropout (Dropout)	(None, 8)	0
dense_1 (Dense)	(None, 1)	9

Total params: 4,625 (18.07 KB)

Trainable params: 4,625 (18.07 KB)

Non-trainable params: 0 (0.00 B)

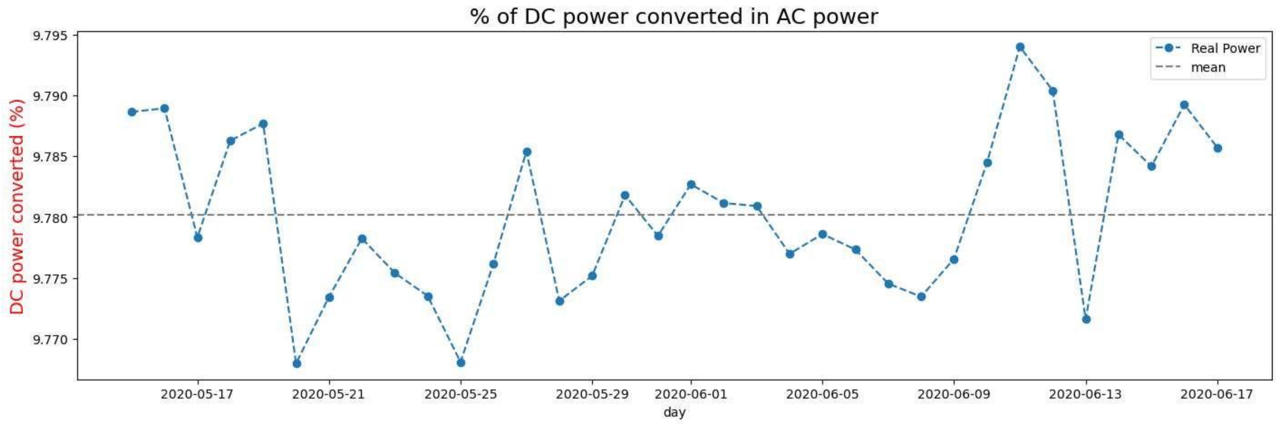
Table 4.3: The Sequential model

Table 4.2 provides a concise overview of the model's architecture and the number of parameters in each layer, to understand the model's complexity and potential performance. The model Structure of the LSTM layer for input and processing data (layer one) as sequential data. It outputs a shape of (None, 32), meaning the batch size is not fixed (None) and each item in the sequence has a 32-dimensional representation. It has 4,352 trainable parameters.

The dense layer (layer 2) is a fully connected dense layer that takes the output of the LSTM layer and maps it to an 8-dimensional space. It has 264 trainable parameters. The dropout layer

(layer 3) is a dropout layer used for regularization to prevent overfitting. It randomly sets a fraction of input units to 0 during training. It has no trainable parameters. Layer 4 (dense_1) is the final dense layer that produces a single output value for each item in the sequence. It has 9 trainable parameters.

Total parameters, the total number of parameters in the model is 4,625, which takes up 18.07 KB of memory. Trainable parameters 4,625 parameters in the model are trainable, meaning their values will be adjusted during the training process. Figure 4.18.



Figur4.18: Recurrent Neural Network (RNN), specifically (LSTM) to show the percentage of energy converted from (DC-AC).

The figure above shows the power electronics and energy systems to monitor and analyse the performance of power conversion equipment over time. The stable efficiency of around 9.78% suggests this is a specific type of power conversion system operating under controlled conditions. The potential application for (Time Series) Prediction. LSTM has been used to predict future conversion efficiency values. This could help in preventive maintenance by forecasting when efficiency might drop below acceptable levels. Clustering algorithms (K-means) could group similar efficiency patterns. Reinforcement Learning could be used to optimize the conversion process to adjust parameters to maximize efficiency under different conditions. The time series data create engineering features for predictive models through (Rolling averages, Peak-to-peak variations, Time-of-day patterns and Seasonal components).

4.6.8 Irradiance and Temperature variables for the most accuracy prediction

The integration of Machine Learning (ML) techniques in photovoltaic (PV) systems has emerged as a promising approach to optimize Maximum Power Point Tracking (MPPT) and enhance overall solar power generation efficiency. Traditional MPPT methods, while effective under stable conditions, often face challenges in rapidly changing environmental conditions and partial shading scenarios. The implementation of ML algorithms offers adaptive and predictive capabilities that can significantly improve system performance by considering multiple variables simultaneously.

key variables affecting solar power generation like environmental parameters as exogenous variables such as solar irradiance (W/m²) and temperature (°C) (model temperature and Ambient temperature) as shown in figures 4.19.

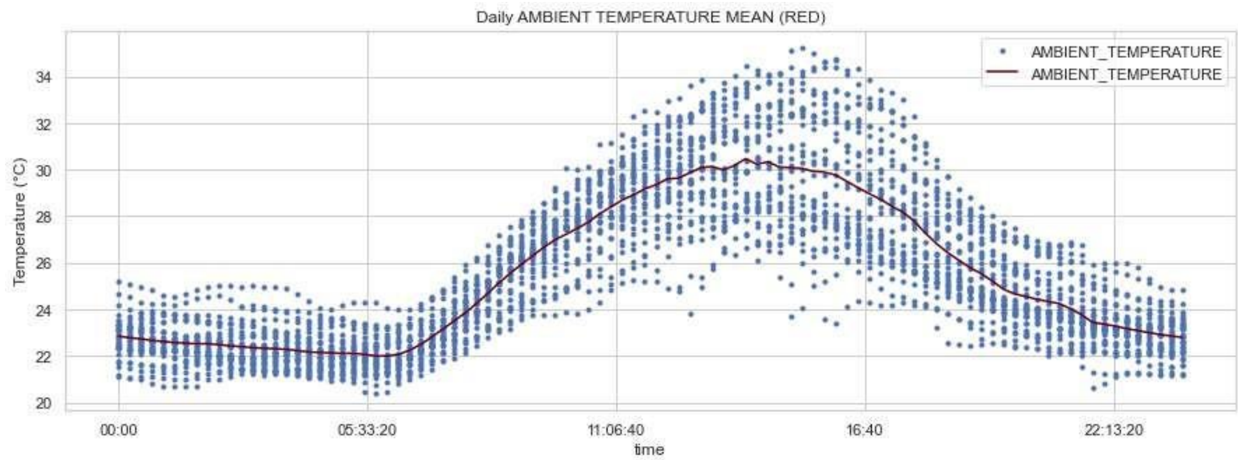


Figure 4.19: solar power forecasting techniques with the mean ambient temperature

The irradiation variables in solar power forecasting are a significant key ML forecasting model, temporal Features like (Time of day, Day of the year, Season and Historical values) features depend on the metrological data such as (temperature, solar irradiance, humidity air pressure and wind speed and metrological data). Considering location-specific (Latitude/Longitude) the forecasting result related to irradiation is shown in Figure 4.20.

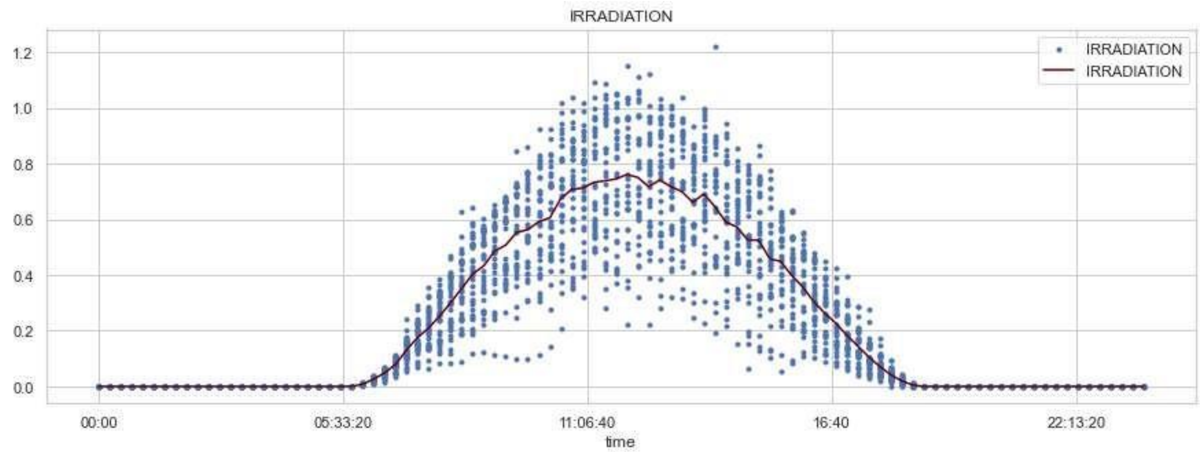


Figure 4.20: Predicted and actual solar irradiance

Other factors can be considered in terms of solar power forecasting and energy generation model temperature combined with weather temperature as shown in the correlation function heatmap figure 4.21. This matrix is useful for understanding system dependencies, identifying key performance indicators, optimising monitoring parameters, and selecting features for predictive models.

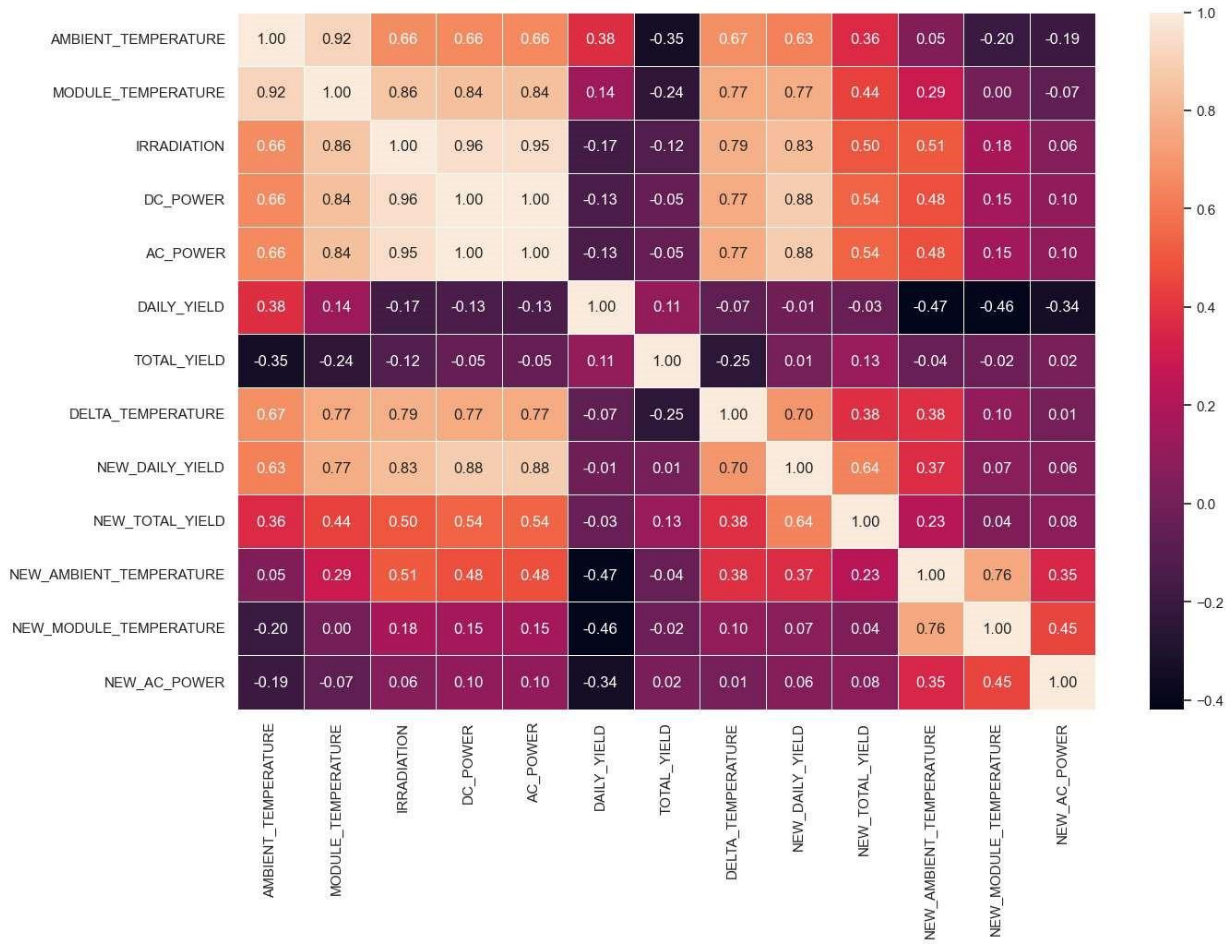


Figure 4.21: Heatmap autocorrelation function for the model focusing on endogenous and exogenous variables

This figure is a correlation matrix heatmap showing the relationships between various parameters in a solar power system. To explain the key insights, strong positive correlations (0.8 to 1.0, light colours) are as follows:

- Ambient and Module temperature (0.92). new temperature and new ambient temperature (0.76)
- Power-related correlations: DC and AC show perfect correlation (1.0). Both strongly correlate with irradiation (0.96 and 0.95)

- Temperature and Power model temperature show strong correlation with irradiation (0.86)
- Weak or Negative Correlations (dark colours), total yield shows weak or negative correlations with most parameters. Daily yield has a negative correlation with many new measurements.

The key observations are that environmental factors (temperature, irradiation) strongly influence power output, and the strong relationship between DC and AC power indicates efficient conversion. New measurements (prefixed with "NEW_") often show different correlation patterns from the original measurements Temperature parameters are key indicators of system performance. Figure 4.22 shows the regression plot of AC &DC. However, the relationship btween X and Y points form a clear straight line pattern, indicating an extremely strong linear relationship between the two variables. The two variables are in a positive direction, as DC power increases, AC power increases proportionally, showing a positive correlation. Near Perfect Fit The data points lie very close to the regression line with minimal scatter, suggesting the relationship has a correlation coefficient very close to 1.0.

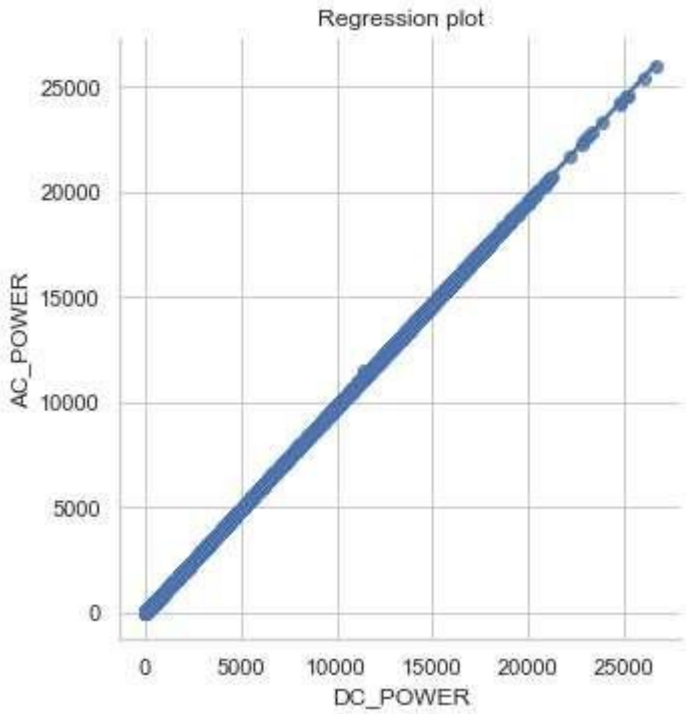


Figure 4.22: scatter regression between AC and DC of proportional relationship appears to pass through or very near the origin (0,0).

X-axis: Represents DC poweY-axis: Represents AC power. The plot visualizes the relationship between DC power and AC power. The data points are clustered tightly along a straight line with a positive slope, indicating a strong positive correlation between the two variables. This suggests that AC power also tends to increase proportionally as DC power increases.

4.7 Summary

The integration of Machine Learning techniques in solar power forecasting and Maximum Power Point Tracking represents a significant advancement in renewable energy optimization. Several keys impact the forecasting techniques such as critical variables Impact, parameters, forecasting horizons (long and short-term) and system integration features. The successful implementation of forecasting model-based approaches in solar power systems demonstrates significant potential for improving overall system efficiency and reliability. The strong correlations observed between environmental factors and system performance highlight the importance of comprehensive monitoring and adaptive control strategies. As technology continues to evolve, the integration of ML in solar power systems will become increasingly crucial for maximizing energy yield (MPPT) and system longevity. The next chapter will illustrate solar power plant management by forecasting the faults that can occur in the system.

Chapter 5

PV power plant management generation

5.1 Introduction

Countries all around the world are getting increasingly interested in solar energy, and boosting its efficacy is turning into a difficult problem. A double mode maximum power point tracking (MPPT) technique is provided in this publication [205]. To realise PV MPPT in both normal operation and partial shade condition (PSC), a radial basis function neural network (RBFNN) is built and driven by historical temperature and irradiance data; for MPPT under normal conditions, the conventional perturb and observe (P&O) method is used. Long short-term memory (LSTM) based solar power prediction is used to detect the partial shading condition [206]. The deviation between the predicted and measured power is used to trigger the MPPT modes switching smoothly and adaptively, improving MPPT speed and accuracy even in the presence of erratic solar irradiance and ambient temperature [205].

5.2 integration AI-ML for optimising managing solar power plant

The artificial intelligence-based MPPT algorithm, which performs better in both transient and steady states, is suggested as a solution to the issues. The fields of neural networks (NN), genetic algorithms (GA), fuzzy logic (FL), and neuro-fuzzy have grown in significance [207]. The FL control is proposed in [206], [208], where it demonstrates its effectiveness in tracking the global MPPT value with less tracking time and, in turn, reduced settling time and minimum oscillation. It does not function well under PSC, though, since it is incapable of self-learning. Under rapidly changing illumination, the ANN-based MPPT approach performs well [209] in terms of efficiency and reaction time. Of them, NN is most used because of its high nonlinear fitting capabilities, which have been shown to make it a more effective tool for implementing MPPT. However, the calibre of the training dataset and model parameters have an impact on NN's performance [210]. To maximise its internal parameters, the neural network needs to be retrained with new training data. It should be mentioned that despite their extreme efficiency, these AI-based approaches are more complicated than conventional ones. Additionally, the model training process which is impacted by the training techniques and dataset that are available determines how well the model functions [210].

Partial shading effect on solar power generation and performance. PV Solar Power has emerged as the best source of green energy in the recent past in a country like the Middle East which gets a good amount of solar insolation. With the continuous development of efficient PV modules, Battery storage, smart grid etc. Power Generation through PV Solar Plant has gained momentum further and has a very promising future [211]. Figure 5.1 shows the result of analysing data.

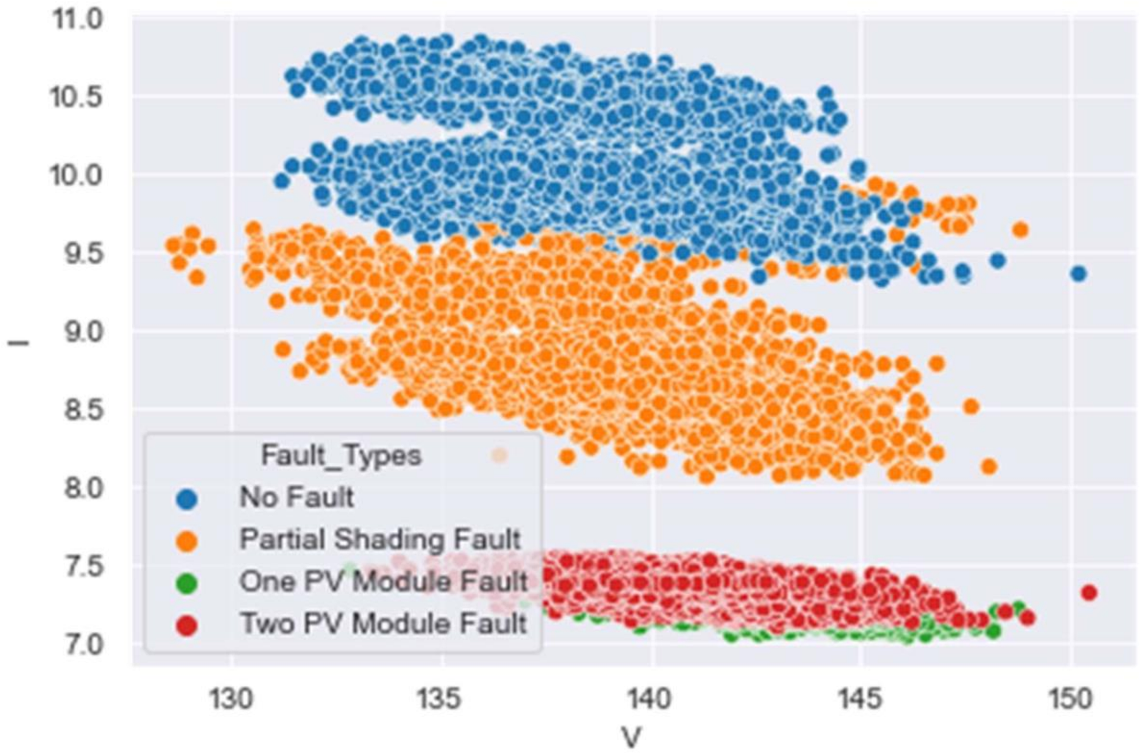


Figure 5.1: Managing and early notification of the fault that happened in the PV plant

The PV system includes various components as a core performance in terms of any fault or shade affecting the system output, figure 5.2 & 5.3 shows the detecting type of fault that occurs in the PV system and type of fault that reduces the maximum output.

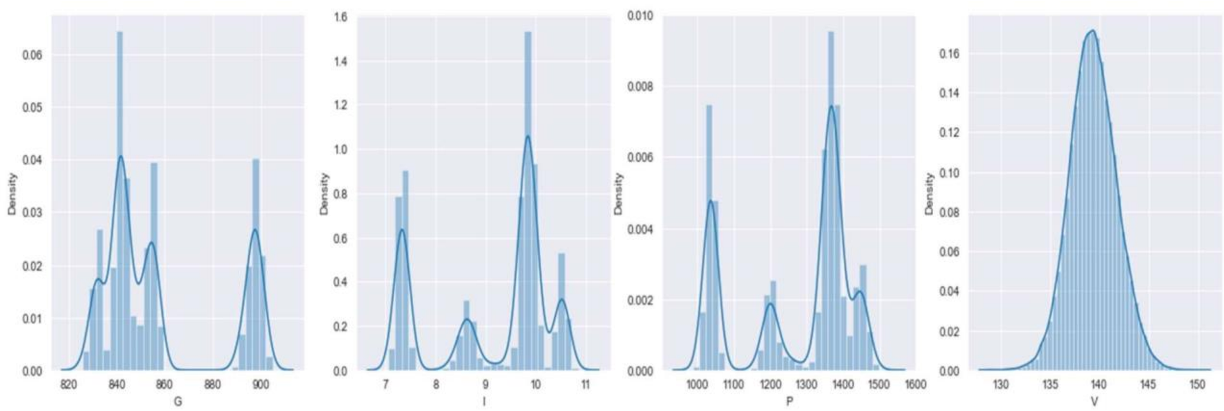


Figure 5.2: PV system ploting of global irradiance G, current I, power P and voltage V

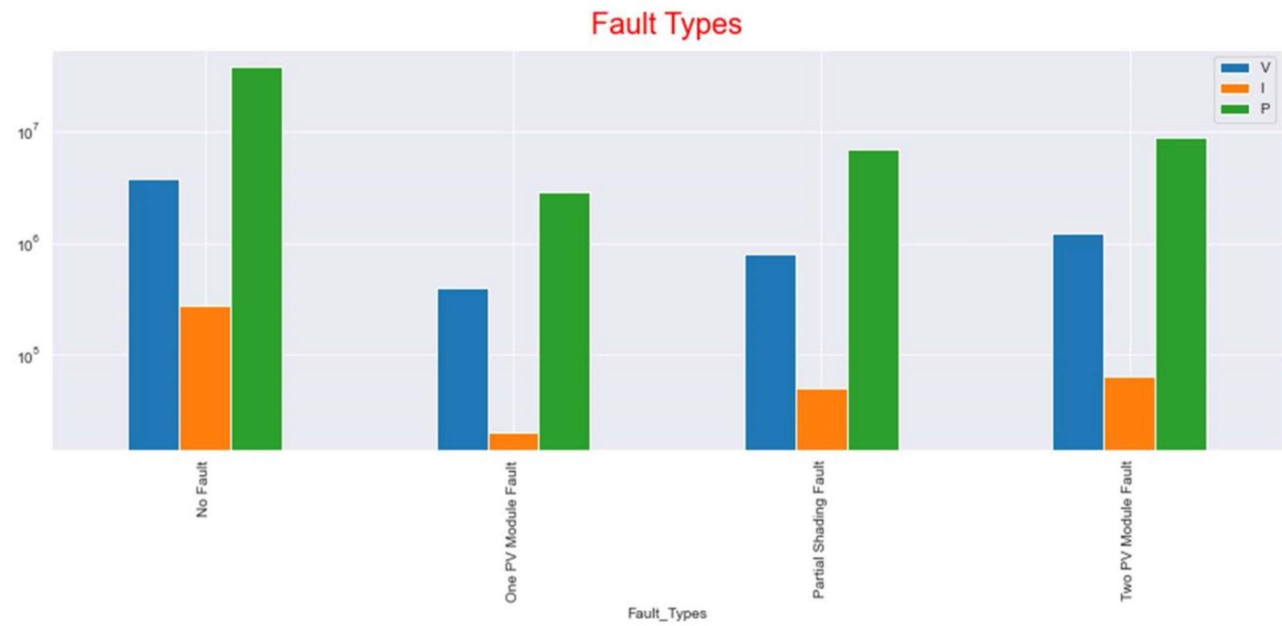


Figure 5.3: The current value we can distinguish between healthy, partially shaded, and faulty PV systems.

To achieve the MPP in the PV system, techniques are used that aim to maximize the power output from solar panels by continuously adjusting their operating point to match the Maximum Power Point (MPP), which varies with changes in sunlight and temperature. The techniques are categorized into two main types (Indirect and Direct).

Indirect MPPT Techniques

Indirect methods rely on pre-determined or approximate values and are less adaptive to real-time environmental changes [212]. These include the following methods

1. Fixed Voltage Method: This method operates the solar panel at a constant, fixed voltage that is close to the estimated MPP voltage. It is straightforward and inexpensive but does not adapt well to changing conditions, which can lead to suboptimal performance in variable weather [213].

2. Fractional Open Circuit Voltage Method: This technique approximates the MPP voltage as a constant fraction (typically around 76-80%) of the opencircuit voltage (Voc) of the panel. The panel's Voc is periodically measured, and the operating voltage is adjusted accordingly. While simple and effective, this approach still lacks precision in rapidly changing irradiance conditions [78], [206].

Direct MPPT Techniques

Direct methods are more adaptive, actively tracking the MPP by continuously monitoring the power output and adjusting based on real-time feedback. These methods include:

1. Perturb and Observe (P&O) Method: In this method, the controller perturbs (alters) the operating voltage or current slightly and observes the resulting change in power. If power increases, the perturbation continues in the same direction; if power decreases, the direction of perturbation is reversed. While effective, P&O can be slow to reach the true MPP under highly fluctuating conditions and may oscillate around the MPP [214].
2. Incremental Conductance (IncCond) Method: This method compares the incremental conductance (change in current divided by change in voltage) with the instantaneous conductance (current divided by voltage) to determine the direction of the MPP. If they are equal, the system is at the MPP. The IncCond method is more accurate than P&O, as it responds better to rapid changes in irradiance, but it also requires more complex calculations [209]. Figure 5.4 shows direct and indirect MPP technique.

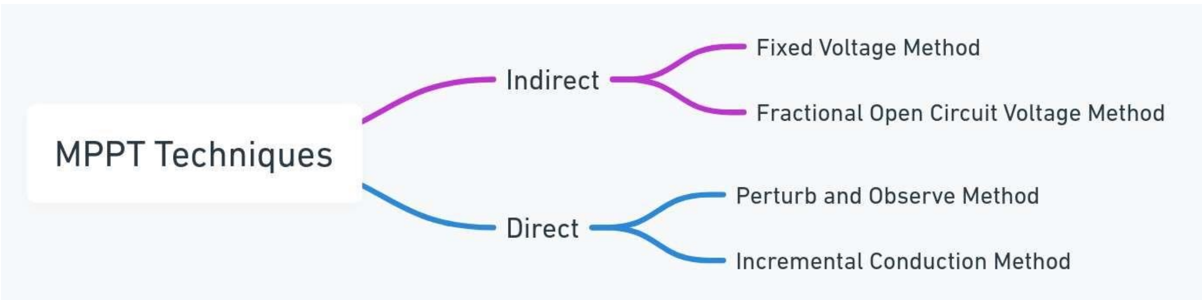


Figure 5.4 MPPT (direct & indirect)

This diagram represents the MPPT techniques. Each method has its advantages and trade-offs. Indirect methods like the Fixed Voltage and Fractional Open Circuit Voltage are simpler and cheaper but may not respond effectively to environmental changes. Direct methods, including (P&O) and Incremental Condition (IC), provide better accuracy and adaptability, particularly in dynamic weather, though they require more sophisticated control systems. The choice of technique often depends on the specific requirements of the solar power system, balancing cost, complexity, and performance [215].

5.3 ML analysis based on Perturb and Observe (P&O) of MPP

A flowchart format is best suited to illustrate the Perturb and Observe (P&O) method [210]. It allows us to represent the decision-making steps involved in adjusting a system's output, such as optimising power in a Maximum PowerPoint Tracking (MPPT) system.

ML create a flowchart for this method. Here is the general structure for the P&O method:

1. Start by measuring the current and voltage.
2. Calculate the power.
3. Check if there is a change in power from the previous measurement.
4. Adjust the operating point (perturb) and observe the change in power.
5. Determine the next action based on the observed result, adjusting further, or reverting if necessary.
6. Continue until the optimal power point is achieved. As shown in Figure 5.5

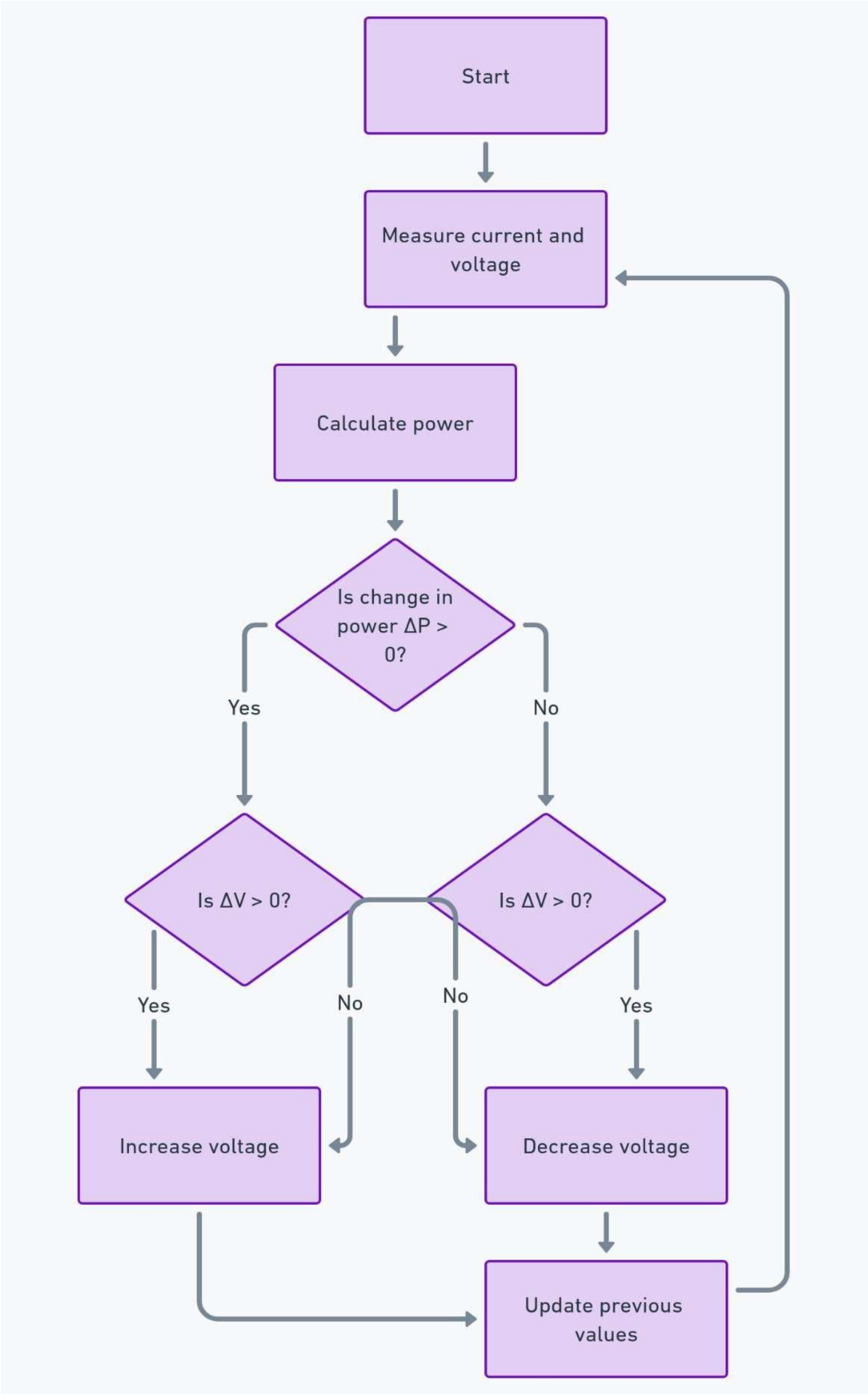


Figure 5.5: The flowchart for the Perturb and Observe (P&O) method

5.4 Equations using for calculation the solar energy output depends on chosing variables:

The Perturb and Observe (P&O) method is a popular algorithm used in Maximum PowerPoint Tracking (MPPT) for photovoltaic (PV) systems. The method adjusts the operating point of the PV system to maximize power output. Here is a breakdown of the mathematical analysis associated with this flowchart [216].

Power Calculation for Perturb and Observe (P&O) method

$P(k) = V(k) \times I(k)$

where:

P(k): Power at the current iteration.
V(k): Voltage at the current iteration.
I(k): Current at the current iteration.

Power Difference:

$\Delta P = P(k) - P(k-1)$

- This is the difference between the power calculated in the current iteration and the previous iteration.

Voltage Difference:

$\Delta V = V(k) - V(k-1)$

This is the change in voltage between the current and the previous iteration. In decision logic, the algorithm observes the direction of changes in power and voltage to determine the correct adjustments [214]. The analysis is as follows:

1. Condition 1: $\Delta P > 0$

- If the power increases ($\Delta P > 0$, it indicates that the perturbation (adjustment) made in the previous step is in the correct direction.

- The algorithm then checks the direction of voltage change (ΔV):
- If $\Delta V > 0$: Increase the duty cycle (D) by a small amount ΔD, moving towards the Maximum Power Point (MPP).
- If $\Delta V < 0$: Decrease the duty cycle (D) by ΔD, moving in the opposite direction to reach the MPP [208].

2. Condition 2: $\Delta P < 0$

- If the power decreases ($\Delta P<0$), it means the perturbation is in the wrong direction, moving the system away from the MPP.
- Again, the algorithm checks ΔV:
- If $\Delta V > 0$: Decrease the duty cycle (D) by ΔD.
- If $\Delta V < 0$: Increase the duty cycle (D) by ΔD.

Goal of Adjustments: The duty cycle, D, adjusts the operating point to achieve maximum power output from the PV system. By continually adjusting D, the algorithm oscillates around the Maximum PowerPoint (MPP), ideally staying as close as possible to it [206], [214] .

Stability and Convergence: Small ΔD, A smaller step size ΔD will lead to more precise convergence at the cost of slower response time. Oscillation, The P&O method can sometimes result in oscillations around the MPP, as it perturbs the voltage back and forth.

The P&O method relies on these fundamental equations:

- Power Calculation: $P(k) = V(k) \times I(k)$
- Change in Power and Voltage: ΔP and ΔV
- Duty Cycle Adjustment: Conditional logic based on the signs of ΔP and ΔV [209], [214].

5.4.1 Summary

This approach effectively performs a gradient-like search by perturbing the voltage and observing the response in power, gradually converging on the Maximum PowerPoint. The success of the P&O algorithm lies in maintaining a balance between response speed (sufficiently large ΔD and accuracy near the MPP.

5.5 Hill Climbing (HC) method in Maximum Power Point Tracking (MPPT)

The Hill Climbing (HC) method in Maximum Power Point Tracking (MPPT) is used to maximise power output in systems like photovoltaic (PV) cells. This method works by adjusting a power converter's duty cycle, D, based on changes in power measurements. It's designed to optimize the power output of solar panels by continuously adjusting the operating point to find and maintain the maximum power point (MPP). A mathematical analysis of the technique is outlined in the flowchart, as shown in Figure 5.6 [217].

A common issue with solar PV panels is the difference between their rated output and the actual power they generate. While we aim to maximize efficiency, the performance of a solar PV system largely hinges on how much sunlight it receives and any shade caused by nearby structures. For instance, if a solar panel is supposed to produce 130 watts, it might find that it's only delivering around 90 watts when checking its output and the charge going to the battery. So, where did that extra 40 watts go? The reality is that it hasn't disappeared; it simply isn't being produced due to those external factors.

The current and voltage generated by a solar panel create a power curve that helps us identify when the panel is performing at its best. This is referred to as the load characteristic. To optimize a solar PV system, it's important to adjust this load characteristic so that the transfer of power from the solar cells to the battery happens at maximum efficiency, especially when there's plenty of sunlight available. The ideal point on this curve is known as the Maximum Power Point (MPP), and monitoring this process is called Maximum Power Point Tracking (MPPT).

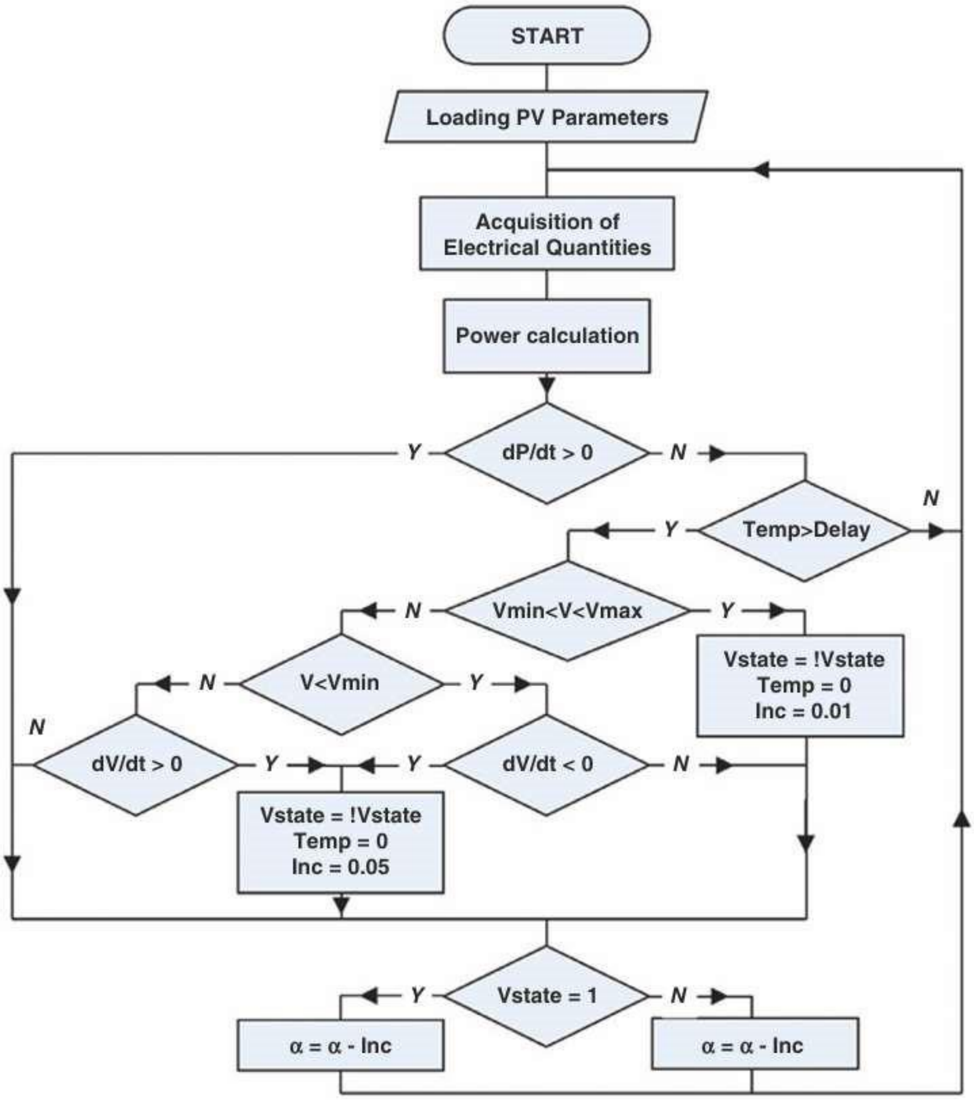


Figure 5.6: Hill Climbing (HC) method in Maximum Power Point Tracking (MPPT) [217].

5.6 Key Variables and Equations for Hill Climbing (HC) Method

The Hill Climbing (HC) method for Maximum Power Point Tracking (MPPT) in photovoltaic systems utilizes essential variables and equations to optimize energy extraction from solar panels. This technique systematically adjusts the operating point to locate the maximum power output, ensuring efficient utilization of solar energy [218].

1. Power Calculation: The output power of a photovoltaic (PV) system is determined by the formula $P = V \times I$, where P represents power, V signifies voltage, and I denotes current[219], [220] $P(n) = V(n) \times I(n)$

   where:

o $P(n)$: Power at the current iteration n. o $V(n)$: Voltage at iteration n. o $I(n)$: Current at iteration n.

2. Power Comparison: The algorithm compares the power $P(n)$ at the current iteration with the power $P(n-1)$ at the previous iteration to determine if the system is moving toward the Maximum Power Point (MPP).

 Power Increase Condition: $\Delta P = P(n) - P(n-1) > 0$

 Power Decrease Condition: $\Delta P = P(n) - P(n-1) < 0$

- Duty Cycle Adjustment: Based on the sign of ΔP, the duty cycle D is adjusted:
- If $\Delta P > 0$ (indicating an increase in power), D is increased by a small increment ΔD. $D = D + \Delta D$.
- If ΔD\ decreases $\Delta P < 0$ (indicating a decrease in power), D. $D = D - \Delta D$.

The Hill Climbing algorithm can be thought of as performing a gradient ascent in terms of power, where the "hill" is the power curve of the PV module. By perturbing the duty cycle and observing the resulting power, the algorithm determines whether it should continue in the same direction or reverse, aiming to stay near the Maximum Power Point (MPP). Stability and Oscillation, small step size ΔD: A small ΔD value leads to smaller oscillations around the MPP but may slow down response time. Oscillatory Behaviour, Due to the perturbation and comparison, this method can exhibit oscillations around the MPP. By fine-tuning ΔD, the oscillation can be minimised for more stability. As shown in Figure 5.7.

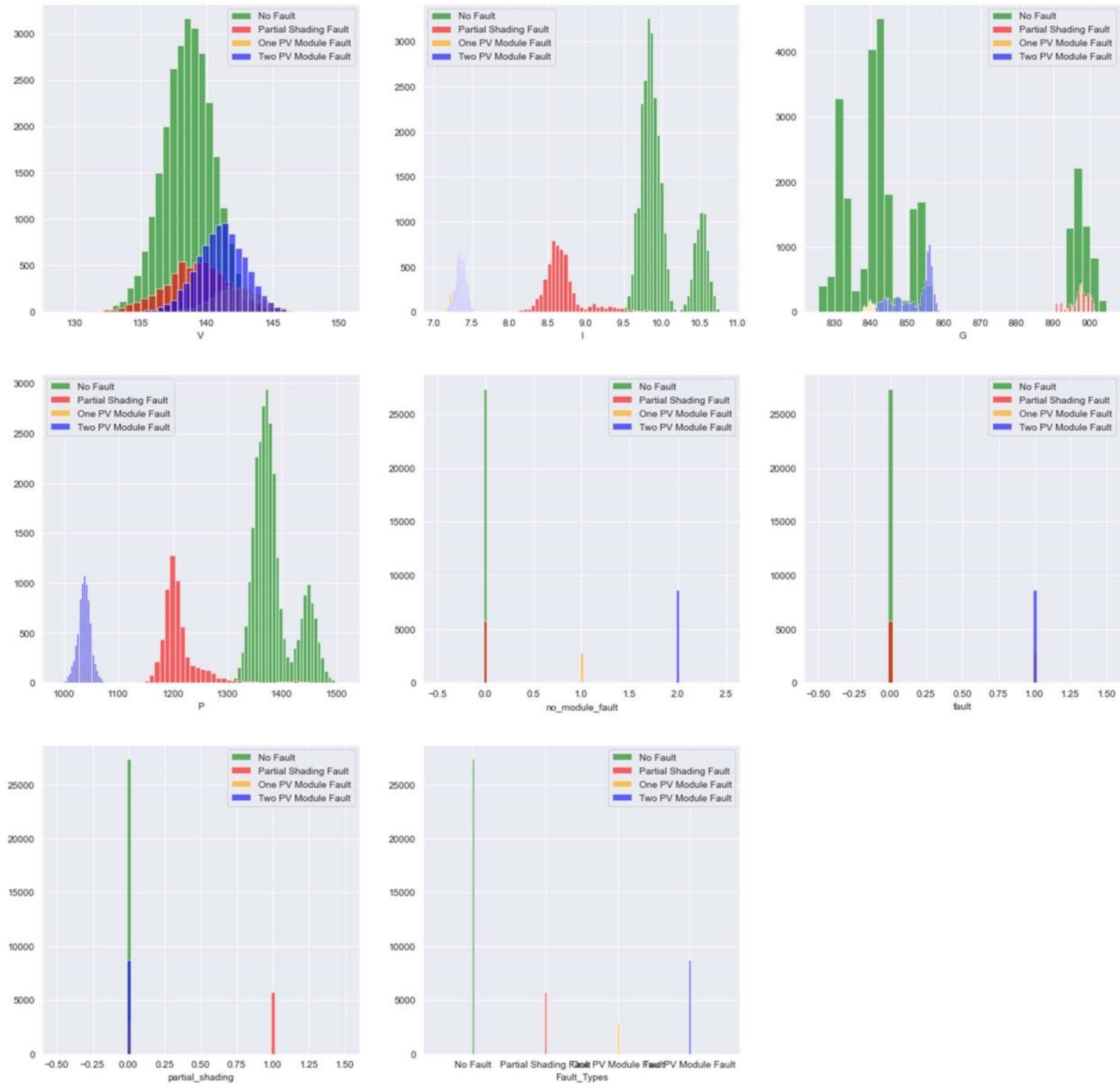


Figure 5.7: Hill Climbing (HC) method in Maximum Power Point Tracking (MPPT)

5.6.1 Summary

The Hill Climbing algorithm is a straightforward, iterative process to track the MPP by gradually adjusting the duty cycle based on power measurements. It works well under stable conditions but can oscillate around the MPP due to constant perturbation, particularly in dynamic environments (e.g., variable sunlight for PV systems). The Hill Climbing algorithm can be thought of as performing a gradient ascent in terms of power, where the "hill" is the power curve of the PV module. By perturbing the duty cycle and observing the resulting power, the algorithm determines whether it should continue in the same direction or reverse, aiming to stay near the Maximum Power Point (MPP): stability and Oscillation.

5.7 Data distribution for more accuracy of PV power forecasting

Accurate photovoltaic (PV) power generation forecasting is paramount in renewable energy for optimising energy management and grid stability. Data distribution plays a crucial role in enhancing the accuracy of these forecasts. By understanding the underlying statistical properties of the data, including its mean, variance, skewness, and kurtosis, we can better model the complexities of solar energy generation. Utilizing advanced data distribution techniques allows us to capture the variability and trends inherent in solar irradiance and temperature data, leading to more reliable predictions. This approach not only improves the efficacy of forecasting models but also aids in mitigating the uncertainties associated with PV power generation. As we delve deeper into the significance of data distribution, we will explore various methods and best practices that can be employed to optimize PV power forecasting accuracy. The plot in Figure 5.8 suggests that the regression model performs well on the training and test data. The points are closely clustered around the perfect fit line, indicating a strong correlation between predicted and actual values. The high R^2 value of 0.981 further supports this conclusion, suggesting that the model explains a substantial portion of the variance in the target variable. However, the RMSE of 55.264 indicates that there is still some error in the model's predictions.

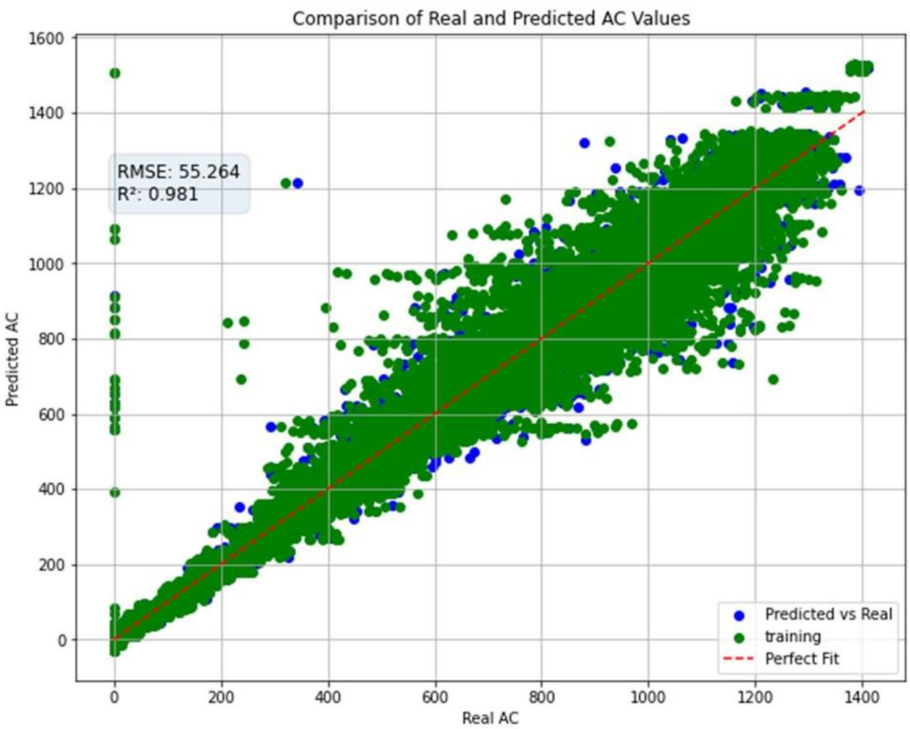


Figure 5.8: Distribution data for real and predicted AC value

The figure is a scatter plot showing the results of a regression model. X-axis (implied): Represents the actual (real) values of the target variable. Y-axis (implied): Represents the predicted values generated by the model.

Data Points of blue dots represent the predicted values versus the real values. This represents a test dataset not used in training the model. Green dots represent the training data used to build the model. The red dashed line represents a perfect fit, where the predicted value perfectly matches the real value for every data point.

Metrics error, RMSE (Root Mean Squared Error): 55.264. This metric measures the average difference between the predicted and actual values. A lower RMSE indicates better model accuracy. Figure 5.6 shows R-squared (R^2): 0.981. This metric represents the proportion of variance in the target variable that is explained by the model. An R^2 value of 1 indicates a perfect fit, while values closer to 0 indicate a poor fit. This process of making predictions using a machine learning model after it has been trained. Involves utilizing a Random Forest Regressor (RFR), which is a popular algorithm in Python's assumes that the model has been successfully fitted to training data and is ready for inference on new data. As Figure 5.9 shows.

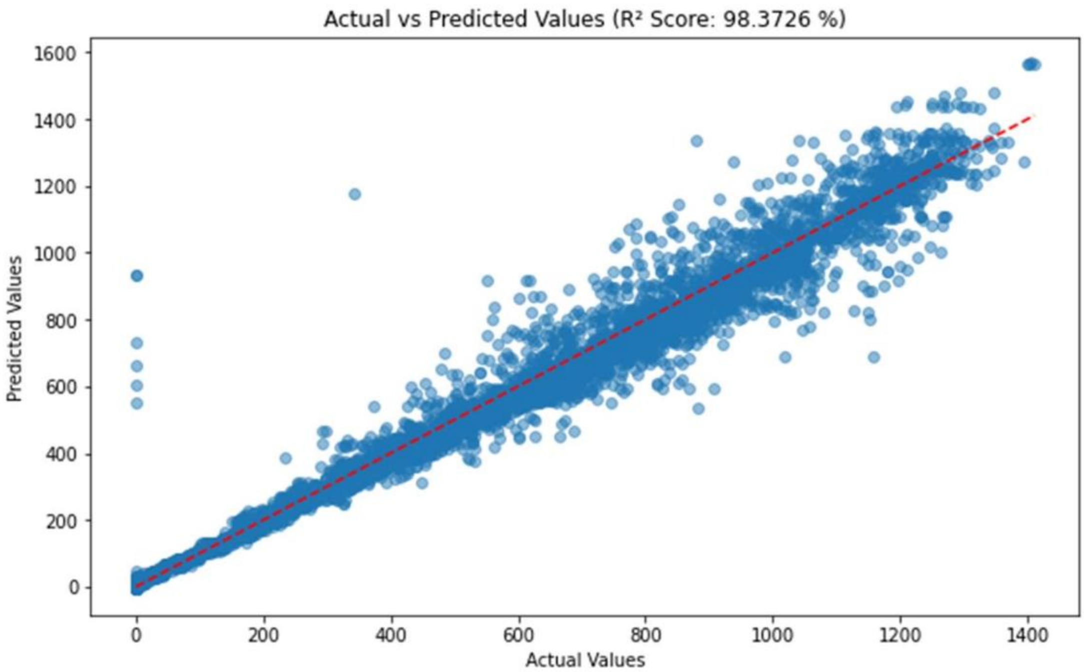


Figure 5.9: R-squared (R^2) of actual vs predicted values5.7 K-Nearest Neighbours

KNN is a simple and widely used machine learning algorithm for classification and regression tasks. Here is a comprehensive overview. The key Features of KNN is a type of lazy learning where the algorithm does not build a model during training. Instead, it stores the training instances and makes decisions based on them at the time of prediction. See figure 5.10.

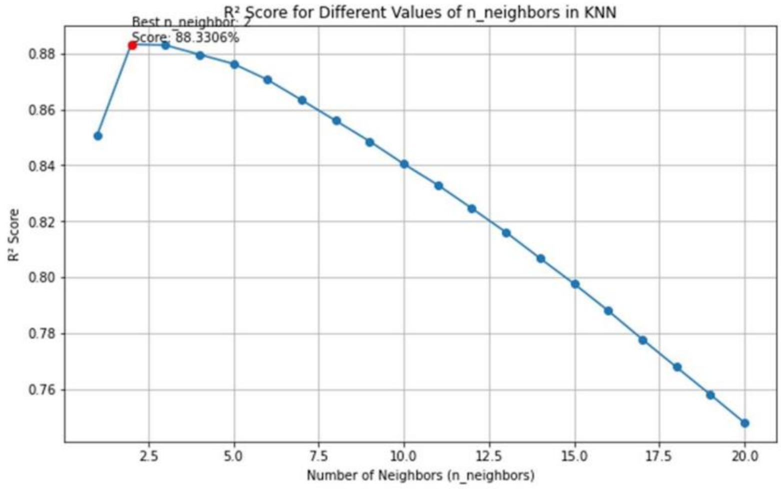


Figure 5.10: KNN based on MPP

The figure representing the (red dot) highlights the optimal K value found through some optimization process (like cross-validation). The X-axis represents K, and the xaxis shows increasing values of K (number of neighbours considered). The Y-axis represents performance [221]. The graph shows a common trend in KNN: performance initially increases with K, reaches a peak, and then gradually decreases. This is because (Low K) Can lead to overfitting and sensitivity to noise. (Optimal K) provides the best balance between bias and variance. (High K) This can lead to underfitting as the model becomes too simplistic [222].

5.7.1 Summary

In summary, the MPP in solar power plants involves a multifaceted approach tailored to the plant's scale and environmental conditions. While the fundamental goal remains the same to maximize power output the specific MPPT mechanisms can range from centralized and simple algorithms in smaller plants to highly adaptive, predictive models and distributed systems in larger installations. As shown in figure 5.11.

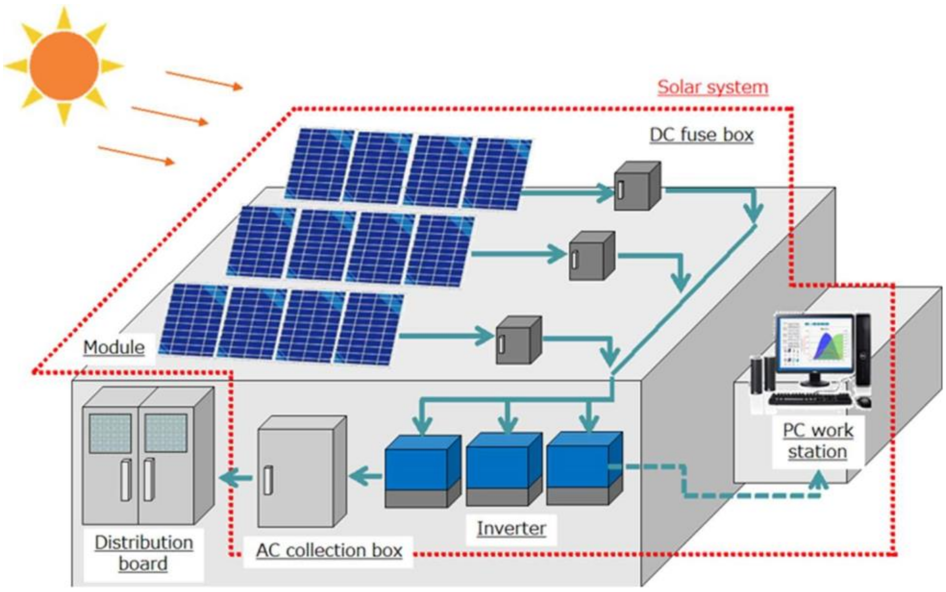


Figure 5.11: MPP involves in the system, depending on the type and scale of the solar power plant [223]

MPP in a solar power plant is a critical element that determines the optimal operating point where the photovoltaic (PV) modules generate the maximum possible power.

Chapter 6

Conclusions and Future Work Recommendations

6.1 Introduction

Solar photovoltaics, or PV, is highly modular technology that may be installed in tiny amounts at a time or produced in big facilities, offering economies of scale. This makes a variety of applications possible, ranging from utility-scale power generation installations to tiny residential roof-top systems. Installing a new grid substation is frequently necessary for large solar farms with an installed capacity of 10 megawatts, typically linked to the high-voltage grid. Solar plants should be situated as near the current road network as feasible to minimise the expense of building new access roads and prevent soil sealing. Areas farther from highways are deemed less favourable than those nearer to roads in the majority of solar plant siting assessments. The environment of solar plants may be negatively impacted by distance from metropolitan centres, such as gas emissions [223], [224].

6.2 Calculation of Area of Solar Power Plant

The position of the city's longitude and latitude, which establish the tilt angle of the solar panel at this location, the lowest sun angle of the year (on December 21), and the solar panel's dimensions all affect the computation of the area needed to build the solar power plant. Calculating the area required for a solar power plant involves several factors that influence the efficiency and total output of the installation. These factors include the target energy generation capacity, solar panel efficiency, local solar irradiance, panel tilt, and other system-related considerations. Below is a structured approach for calculating the area required for a solar power plant:

1. Determine Target Energy Generation Capacity, define the energy output you want from the solar power plant, typically measured in megawatts (MW).

2. Consider Solar Panel Efficiency, solar panel efficiency impacts how much sunlight a panel can convert into electricity. Efficiency typically ranges from 15% to 22% for commercial panels. Higher efficiency panels produce more power per unit area but are generally costlier [224].
3. Calculate Daily Solar Irradiance, the average solar irradiance, or the amount of solar energy received per unit area per day, varies by location and is usually measured in $kWh/m^2/day$. Areas with high solar irradiance require less space to achieve the same energy output as those with lower irradiance [223].

6.3 PV Solar Energy Production

Potential solar panels typically generate, whether it's 100% or just 50% of their household energy needs. Several factors can influence the energy generation capacity of solar panels. Solar panels have been used for decades to convert sunlight into energy [225]. In recent decades, solar technology has rapidly advanced, leading to increased efficiency and reduced costs. With the declining cost of solar energy each year, it has become very feasible for homeowners and businesses to install solar panels on their properties. Solar panels work by allowing particles of light, known as photons, to dislodge electrons from their molecular cores. This process generates a flow of electricity, which is then harnessed and stored as electrical energy. This occurs when light hits a device known as a photovoltaic cell. PV panels consist of hundreds of these cells, and a solar array is made up of multiple solar panels[225] .

These arrays are connected to the existing electrical grid, allowing for the proper distribution of the generated energy. To gauge the amount of electricity a solar panel can produce, it's essential to consider the type of panel technology used. When looking for a solar energy provider. When discussing solar energy generation, we refer to the amount of energy (in watts) obtained from converting sunlight into electricity. This conversion can be achieved through photovoltaic cells, concentrated solar power, or a combination of both. A common question is about how much energy solar panels produce, as many people remain uncertain about the quantities involved [217].

6.4 Adoption of Solar Power Forecasting Using Machine Learning Algorithms

Forecasting solar power is crucial for improving energy output and grid management, particularly as the use of renewable energy sources grows. There is great potential for improving the precision and dependability of solar power forecasts with machine learning techniques. ML models may efficiently understand intricate associations and enhance predicting skills by utilising previous meteorological data, solar irradiance patterns, and other environmental parameters [217], [218]. The following are some major benefits of using ML for solar power forecasting:

- Increased Accuracy: Compared to conventional approaches, machine learning models especially deep learning techniques can better predict outcomes by capturing non-linear correlations in data [220].
- Adaptability: As circumstances change, real-time modifications and better forecasts are made possible by ML algorithms' constant learning from fresh data [219].
- Handling of Large Datasets: As computing power increases, machine learning algorithms are able to process and analyse enormous volumes of data, which improves the resilience of forecasting models [226].

- Integration with Smart Grids: Machine learning makes it easier to combine solar power forecasts with smart grid technology, which improves energy management and demand response tactics [226].

There are still issues, though, such as the requirement for high-quality data, the possibility of model overfitting, and the requirement that ML models be understandable to stakeholders. Future studies should concentrate on enhancing model transparency, utilising hybrid strategies that integrate machine learning with conventional forecasting methods, and investigating how new technologies, such as the Internet of Things, affect data collecting and forecasting precision. In conclusion, using machine learning for solar power forecasting offers a viable way to improve solar energy systems' dependability and efficiency, which will help create a more resilient and sustainable energy future [219], [227].

6.5 Challenges and Future Directions

Despite the advantages, challenges persist in the domain of solar power forecasting using ML. These include:

- Need for High-Quality Data: Accurate forecasts depend on the availability and quality of data.
- Potential for Model Overfitting: Care must be taken to avoid overfitting, where models perform well on training data but poorly on unseen data.
- Interpretability of Models, stakeholders need to understand ML models, necessitating a focus on model transparency.

Future research should aim to address these challenges by improving model interpretability, exploring hybrid approaches that combine ML with traditional forecasting techniques, and examining the potential of emerging technologies, such as the Internet of Things (IoT), in enhancing data collection and forecasting accuracy. In summary, the integration of machine learning in solar power forecasting presents a promising avenue for improving the efficiency and reliability of solar energy systems. This advancement contributes to a more sustainable and resilient energy future, aligning with global efforts to enhance energy production from renewable sources. While the integration of machine learning (ML) in solar power forecasting holds great promise, several limitations persist that can affect the effectiveness and applicability of the research in this field.

6.6 Limitations of The Research in Solar Power Forecasting Using ML

While the research on using machine learning for solar power forecasting has made substantial progress, the limitations outlined above highlight the need for continued investigation and development. Addressing these challenges will be crucial for enhancing the reliability, applicability, and acceptance of ML-driven forecasting solutions in the renewable energy sector but it has a limitation in the flowing points

1- Data Quality and Availability

There is Insufficient Historical Data, many regions lack extensive historical solar radiation and weather data, which limits the ability to train robust ML models. However, data Granularity, and high-resolution data (e.g., minute-by-minute) may not be available for all locations, leading to potential discrepancies in forecasting accuracy. Inconsistent Data Sources, differences in data collection methods across sources can introduce biases and inaccuracies.

2- Model Complexity

Overfitting, as previously mentioned, ML models can become overly complex, fitting noise rather than the underlying pattern in the data. This can lead to poor generalization of new data. Computational Resources, many advanced ML models require significant computational resources and time for training, which may not be feasible for all organizations, especially smaller entities.

3- Interpretability Challenges

Accuracy and Interpretability, despite advances in explainable AI, many ML models remain difficult to interpret, making it challenging for stakeholders to understand and trust the forecasts. Often, more interpretable models (like linear regression) may not achieve the same level of accuracy as complex models (like deep learning), which can deter users from adopting them

5- Dynamic Nature of Solar Power

Variability and Uncertainty: Solar power generation is influenced by numerous factors (cloud cover, atmospheric conditions, geographical variations) that can change rapidly. This inherent variability can make forecasting particularly challenging. In addition, the seasonal variations of different seasons can exhibit drastically different weather patterns, complicating the model training process and potentially leading to seasonally biased predictions.

6-Integration with Existing Energy Systems

Operational constraints integrating ML forecasts with existing energy management systems may pose challenges, particularly if these systems were designed before the advent of advanced forecasting methodologies. However, regulatory and policy barriers to the adoption of ML-based forecasting tools may be hampered by existing regulations and policies that do not accommodate new technologies or methodologies. 7-Lack of Standardization & Limited Field Testing

Diverse Methodologies, the absence of standardized methodologies for data collection, model training, and evaluation can make it difficult to compare results across studies and to establish best practices. In real-world applications, many studies may focus on theoretical models or simulations, with limited validation in real-world conditions. The effectiveness of ML forecasting tools needs to be proven in practical scenarios.

6.7 Recommendations for Building a PV-Powered Data Centre with Accurate Energy Production Forecasting

Building a data centre that relies entirely on photovoltaic (PV) power, while aiming for accurate forecasting of energy production, involves several crucial steps and considerations.

1. Comprehensive Market and Feasibility Analysis, conduct a feasibility study to Analyse the solar resource availability in your location, including seasonal variations, historical weather patterns, and potential shading issues. Investigate the local market analysis for renewable energy, including potential incentives, tariffs, and regulatory frameworks.

2. Data Collection and Management, high-quality data sources access to reliable historical solar radiation, meteorological data, and operational data about potential PV systems. In addition, getting real-time data monitoring to implement systems for monitoring of solar production and environmental conditions using IoT devices and sensors.

3. Development of model selection through comparative analysis of forecasting methods by evaluating different forecasting methods (e.g., statistical models, machine learning models, hybrid approaches) to find the most suitable ones for model training and validation. Train models using historical data and validate them on unseen data to ensure robustness. Utilise cross-validation techniques to avoid overfitting.

4. Hybrid Microgrid Consideration integration of energy storage Consider incorporating battery storage or other energy storage solutions to manage variability and provide backup power during low production periods. Distributed Energy Resources (DERs), by

exploring the integration of other renewable sources (e.g., wind, biomass) to enhance the reliability of the microgrid.

5. Incorporating Advanced Forecasting Techniques through Implementing demand response strategies that allow the data centre to adjust its consumption based on availability, optimizing energy use. Investment and financial modelling with costbenefit analysis, to develop a detailed financial model outlining Capital Expeditor (CAPEX), Operational Expeditor (OPEX), Return on Investment (ROI), and payback periods for investors. Identify risks associated with solar energy production variability, including regulatory, financial, and operational risks.

6. Explainable AI and integrated prioritize models that offer interpretability to build trust with investors and stakeholders. Provide insights into what factors influence production forecasts. Using hybrid models to explore hybrid forecasting approaches that combine statistical and machine learning techniques. For example, ensemble methods can improve prediction accuracy.

7. Simulation and Scenario Analysis, scenario planning uses simulations to model different operational scenarios (e.g., high demand, low sunlight) and their impact on energy production and consumption. Simulation scenario sensitivity Analysis by assessing how changes in key variables (e.g., solar panel efficiency, storage capacity) affect overall energy management and forecasting accuracy.

8. Collaboration and Partnerships by engaging with experts through collaboration with academic institutions or industry experts in renewable energy and forecasting to leverage their expertise. In addition, consider partnerships with local energy providers or technology companies to enhance resources and capabilities
9. Regulatory Compliance and Best Practices by stay updated on regularly review local regulations regarding renewable energy systems, data privacy, and energy management. Follow industry best practices for the design and operation of renewable energy systems and data centres.

10. Continuous Improvement and Adaptation Iterative model improvement

Continuously refine forecasting models based on new data and improved algorithms. Implement feedback mechanisms to monitor performance and adapt strategies based on operational data.

6.8 Conclusion

Solar forecasters would significantly benefit from open data practices. This holds true for researchers' open data policies as well as those at the forefront of public institutions and the scientific community. Policymakers may take the lead in the latter by pressing institutions that get public funding to release solar power forecast data to the public. This comprises meteorological information and measurements of PV power. National and international weather centres' implementation of public data policy, such as the Royal Netherlands Meteorological Institute (KNMI) and European Centre for Medium-Range Weather Forecasts (ECMWF), would lead to better dissemination and accessibility of (past) weather predictions. Since these open data practises will create fertile ground for (rapid) advancements in solar forecasting, this indirectly helps society as a whole. This would thereby improve the power system's ability to schedule supply and demand efficiently, which would lower the amount of balancing capacity needed. This results in lower grid operating costs and perhaps lowers greenhouse gas emissions linked to the energy industry.

In a similar vein, the worth of a large but unobtainable dataset that includes PV power measurements across thousands of PV systems for a minimum of one year is to be revealed. In theory, such a dataset ought to include details on a PV system's tilt, orientation, DC and AC capacity, and location. Such a data platform should ideally include PV power metrics as well. Individual PV systems' power generation is measured and reported in real time by smart inverters, which are frequently linked to online and mobile applications. Thus, intelligent inverters offer real-time information on the local PV power generation and their data are usually collected for commercial and individual use. Given that there are more than two million rooftop photovoltaic systems in the Netherlands, the volume of data that is accessible and its potential applications are astounding. The privacy sensitivity of (real-time) PV power generation data appears to be of limited concern because it depends on local weather conditions and, therefore, measurements do not disclose information on the electricity demand by end users, unlike smart metre data that reports the electricity demand, By establishing an open data platform where PV generation data is gathered in real time, the policy might make it easier to distribute (real-time) PV generation data. Recently, a similar proposal was made. To support the ongoing development of improved and new solar forecasting models for efficient grid operation, energy market trading, and congestion management, the open data platform would be extremely beneficial to researchers, distribution and transmission system operators, utilities, aggregators, and others. Hence, permitting the development of substantial solar PV capacity while maintaining supply security.

References

[1] M. A. Alghamdi, A. S. Al-Malaise Al-Ghamdi, and M. Ragab, "Predicting Energy Consumption Using Stacked LSTM Snapshot Ensemble," Big Data Mining and Analytics, vol. 7, no. 2, pp. 247–270, Jun. 2024, doi: 10.26599/BDMA.2023.9020030.

[2] X. Pang, W. Sun, H. Li, Y. Ma, X. Meng, and W. Liu, "Ultra-short-term power load forecasting method based on stochastic configuration networks and empirical mode decomposition," Front Energy Res, vol. 11, p. 1182287, Jul. 2023, doi: 10.3389/FENRG.2023.1182287/BIBTEX.

[3] L. Zhu, Z. Wu, and S. Huang, "Assessment method of distribution network health level based on multivariate information," Front Energy Res, vol. 11, 2023, doi: 10.3389/FENRG.2023.1178631/FULL.

[4] M. Tan, S. Yuan, S. Li, Y. Su, H. Li, and F. H. He, "Ultra-Short-Term Industrial Power Demand Forecasting Using LSTM Based Hybrid Ensemble Learning," IEEE Transactions on Power Systems, vol. 35, no. 4, pp. 2937–2948, Jul. 2020, doi: 10.1109/TPWRS.2019.2963109.

[5] Haosen Xu, Kaiyi Niu, Tianyi Lu, and Siyang Li, "Leveraging artificial intelligence for enhanced risk management in financial services: Current applications and future prospects," Engineering Science & Technology Journal, vol. 5, no. 8, pp. 2402–2426, Aug. 2024, doi: 10.51594/ESTJ.V5I8.1363.

[6] B. Liu, X. Zhao, H. Hu, Q. Lin, and J. Huang, "Detection of Esophageal Cancer Lesions Based on CBAM Faster R-CNN," Journal of Theory and Practice of Engineering Science, vol. 3, no. 12, pp. 36–42, Dec. 2023, doi: 10.53469/JTPES.2023.03(12).06.

[7] "AI-Driven Solar Energy Generation and Smart Grid Integration A Holistic Approach to Enhancing Renewable Energy Efficiency | International Journal of Innovative Research in Engineering and Management." Accessed: Sep. 30, 2024. [Online]. Available: http://ijirem.irpublications.org/index.php/ijirem/article/view/61

[8] C. Fan, W. Ding, K. Qian, H. Tan, and Z. Li, "Cueing Flight Object Trajectory and Safety Prediction Based on SLAM Technology," Journal of Theory and Practice of Engineering Science, vol. 4, no. 05, pp. 1–8, May 2024, doi: 10.53469/JTPES.2024.04(05).01.

[9] J. T. Dellosa and E. C. Palconit, "Artificial Intelligence (AI) in Renewable Energy Systems: A Condensed Review of its Applications and Techniques," 21st IEEE International Conference on Environment and Electrical Engineering and 2021 5th IEEE Industrial and Commercial Power System Europe, EEEIC / I and CPS Europe 2021 - Proceedings, 2021, doi: 10.1109/EEEIC/51590.2021.9584587.

[10] B. Wu, Y. Gong, H. Zheng, Y. Zhang, J. Huang, and J. Xu, "Enterprise cloud resource optimization and management based on cloud operations," Applied and Computational

Engineering, vol. 76, no. 1, pp. 8–14, May 2024, doi: 10.54254/27552721/76/20240667.
[11] T. Yang, Q. Xin, X. Zhan, S. Zhuang, and H. Li, “Enhancing financial services through big data and AI driven customer insight and risk analyses”, Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), vol. 3, no. 3, pp. 53–62, Jun. 2024, doi: 10.60087/JKLST.VOL3.N3.P53-62.
[12] “What is solar power forecasting? – gridX.” Accessed: Sep. 17, 2024. [Online]. Available: https://www.gridx.ai/knowledge/what-is-solar-power-forecasting
[13] H. Wang et al., “Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network,” Energy Convers Manag, vol. 153, pp. 409–422, Dec. 2017, doi: 10.1016/J.ENCONMAN.2017.10.008.
[14] H. T. Yang, C. M. Huang, Y. C. Huang, and Y. S. Pai, “A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output,” IEEE Trans Sustain Energy, vol. 5, no. 3, pp. 917–926, 2014, doi: 10.1109/TSTE.2014.2313600.
[15] “Installed solar energy capacity.” Accessed: Sep. 17, 2024. [Online]. Available: https://ourworldindata.org/grapher/installed-solar-pv-capacity?country=OWID_WRL~CHN~IND~ESP~BRA~CHL
[16] B. Doha, T. Haidi, D. Mounir, E. Faissal, and T. Ali, “Artificial Intelligence Methods Applied To Wind And Solar Energy Forecasting: A Comparative Study of Current Techniques,” Proceedings - 2023 International Conference on Digital Age and Technological Advances for Sustainable Development, ICDATA 2023, pp. 94–99, 2023, doi: 10.1109/ICDATA58816.2023.00026.
[17] “Forecasting of Energy Production for Photovoltaic Systems Based on ARIMA and ANN Advanced Models.” Accessed: Feb. 13, 2024. [Online]. Available: https://www.hindawi.com/journals/ijp/2021/6777488/
[18] “Ensemble Learning Methods for Deep Learning Neural Networks - MachineLearningMastery.com.” Accessed: Sep. 28, 2024. [Online]. Available: https://machinelearningmastery.com/ensemble-methods-for-deep-learning-neuralnetworks/
[19] A. Boilley, C. Thomas, M. Marchand, E. Wey, and P. Blanc, “The Solar Forecast Similarity Method: A New Method to Compute Solar Radiation Forecasts for the Next Day,” Energy Procedia, vol. 91, pp. 1018–1023, 2016, doi: 10.1016/J.EGYPRO.2016.06.270.
[20] S. Sobri, S. Koohi-Kamali, and N. A. Rahim, “Solar photovoltaic generation forecasting methods: A review,” Energy Convers Manag, vol. 156, pp. 459–497, Jan. 2018, doi: 10.1016/J.ENCONMAN.2017.11.019.
[21] H. T. C. Pedro and C. F. M. Coimbra, “Assessment of forecasting techniques for solar power production with no exogenous inputs,” Solar Energy, vol. 86, no. 7, pp. 2017–2028, Jul. 2012, doi: 10.1016/J.SOLENER.2012.04.004.
[22] H. Bounechba, A. Bouzid, K. Nabti, and H. Benalla, “Comparison of perturb & observe and fuzzy logic in maximum power point tracker for PV systems,” Energy Procedia, vol. 50, pp. 677–684, 2014, doi: 10.1016/J.EGYPRO.2014.06.083.
[23] A. Aghahosseini, D. Bogdanov, and C. Breyer, “Towards sustainable development in the MENA region: Analysing the feasibility of a 100% renewable electricity system in 2030,” Energy Strategy Reviews, vol. 28, Mar. 2020, doi: 10.1016/J.ESR.2020.100466.
[24] A. Mehdi, “Iraqi Oil: industry evolution and short and medium-term prospects Energy Markets Strategist, Livingstone Partners,” 2018, doi: 10.26889/9781784671211.

[25] H. Lund, J. Z. Thellufsen, P. A. Østergaard, P. Sorknæs, I. R. Skov, and B. V. Mathiesen, "EnergyPLAN – Advanced analysis of smart energy systems," Smart Energy, vol. 1, Feb. 2021, doi: 10.1016/J.SEGY.2021.100007.
[26] J. Isabona et al., "Development of a Multilayer Perception Neural Network for Optimal Predictive Modeling in Urban Microcellular Radio Environments," Applied Sciences (Switzerland), vol. 12, no. 11, Jun. 2022, doi: 10.3390/APP12115713.
[27] "Solar Power on the Rise: Global Solar Panel Statistics, Facts, and Trends of 2024." Accessed: Sep. 26, 2024. [Online]. Available: https://www.greenmatch.co.uk/solarenergy/solar-pv-statistics
[28] E. Rodríguez, L. Cornejo-Ponce, J. M. Cardemil, A. R. Starke, and E. L. Droguett, "Estimation of one-minute direct normal irradiance using a deep neural network for five climate zones," Renewable and Sustainable Energy Reviews, vol. 183, Sep. 2023, doi: 10.1016/j.rser.2023.113486.
[29] Y. Lu et al., "Prediction of diffuse solar radiation by integrating radiative transfer model and machine-learning techniques," Science of the Total Environment, vol. 859, Feb. 2023, doi: 10.1016/j.scitotenv.2022.160269.
[30] D. Yang, "Estimating 1-min beam and diffuse irradiance from the global irradiance: A review and an extensive worldwide comparison of latest separation models at 126 stations," Renewable and Sustainable Energy Reviews, vol. 159, May 2022, doi: 10.1016/j.rser.2022.112195.
[31] "Energy Technologies | NRW.GLOBAL Business." Accessed: May 30, 2024. [Online]. Available: https://www.nrwglobalbusiness.com/innovation-topics/energytechnologies?mtm_campaign=sea_ngb-24-004_text_erneuerbareenergien_UK_--_adgapid_023_845011_00021270_103_0_0&gad_source=1&gclid=CjwKCAjwxCyBhAqEiwAeOcTdbAUjEyGNTy4alL4iL21HDbqB9d7wtiNt_q1nP69SEvZMQTd1G0inhoCKcUQAvD_BwE
[32] P. Lyakhov, M. Kiladze, and U. Lyakhova, "System for neural network determination of atrial fibrillation on ecg signals with wavelet-based preprocessing," Applied Sciences (Switzerland), vol. 11, no. 16, Aug. 2021, doi: 10.3390/APP11167213.
[33] M. J. Bush, "Climate change and renewable energy: How to end the climate crisis," Climate Change and Renewable Energy: How to End the Climate Crisis, pp. 1–525, Jan. 2019, doi: 10.1007/978-3-030-15424-0.
[34] "Net Zero Strategy: How To Make Your Business Carbon Neutral." Accessed: May 30, 2024. [Online]. Available: https://www.igniteenergy.co.uk/news/net-zero-forbusinesses/?gad_source=1&gclid=CjwKCAjwx-CyBhAqEiwAeOcTdQNYRlxIe-AIH0TKm3QccW46wiQdGMlhSBNAyAsQ2HQN-dyoMjqapRoCZ70QAvD_BwE
[35] G. Su, S. Zhang, W. Yao, M. Hu, and H. Hao, "Study on the evaluation method of diffuse radiation models——taking 5 typical cities in China as examples," Energy and Built Environment, vol. 4, no. 2, pp. 236–243, Apr. 2023, doi: 10.1016/j.enbenv.2021.12.001.
[36] M. Diagne, M. David, P. Lauret, J. Boland, and N. Schmutz, "Review of solar irradiance forecasting methods and a proposition for small-scale insular grids," Renewable and Sustainable Energy Reviews, vol. 27, pp. 65–76, 2013, doi: 10.1016/J.RSER.2013.06.042.
[37] J. Chen and X. Xue, "A transfer learning-based long short-term memory model for the prediction of river water temperature," Eng Appl Artif Intell, vol. 133, Jul. 2024, doi: 10.1016/j.engappai.2024.108605.
[38] X. Qing and Y. Niu, "Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM," Energy, vol. 148, pp. 461–468, Apr. 2018, doi:

10.1016/J.ENERGY.2018.01.177.
[39] X. Xue, “Prediction of daily diffuse solar radiation using artificial neural networks,” Int J Hydrogen Energy, vol. 42, no. 47, pp. 28214–28221, Nov. 2017, doi: 10.1016/J.IJHYDENE.2017.09.150.
[40] J. Mathe, N. Miolane, N. Sebastien, and J. Lequeux, “PVNet: A LRCN Architecture for Spatio-Temporal Photovoltaic PowerForecasting from Numerical Weather Prediction,” Feb. 2019, Accessed: Sep. 17, 2024. [Online]. Available: http://arxiv.org/abs/1902.01453
[41] L. Benali, G. Notton, A. Fouilloy, C. Voyant, and R. Dizene, “Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components,” Renew Energy, vol. 132, pp. 871–884, Mar. 2019, doi: 10.1016/J.RENENE.2018.08.044.
[42] “Grid-connected PV - European Commission.” Accessed: Sep. 17, 2024. [Online]. Available: https://joint-research-centre.ec.europa.eu/photovoltaic-geographicalinformation-system-pvgis/pvgis-tools/grid-connected-pv_en
[43] “IEEE Xplore Full-Text PDF:” Accessed: Jun. 04, 2024. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=9739682
[44] MTW, H. Liu, and H. Motoda, “Feature extraction, construction and selection: a data mining perspective,” J Am Stat Assoc, vol. 94, no. 448, p. 1390, Dec. 1999, doi: 10.2307/2669967.
[45] P. Vicente, C. Michelle, and P. Pablo, “Position Control Algorithms of Photovoltaic Panels with Respect to the Incidence of Sunlight”, doi: 10.18687/LACCEI2023.1.1.319.
[46] “Energy innovation | Aramco.” Accessed: May 30, 2024. [Online]. Available: https://www.aramco.com/en/what-we-do/energyinnovation?utm_source=&utm_medium=&utm_campaign=&utm_term=&utm_content =&gad_source=1&gclid=CjwKCAjwx-CyBhAqEiwAeOcTdfzdlq_UiWpUloSqFBQe60C7nuVpJYQYwlxJDm3AIVckAgre5dXozxoC7Y0QAvD_BwE
[47] A. Cariow, G. Cariowa, and J. P. Paplinski, “An algorithm for fast multiplication of kaluza numbers,” Applied Sciences (Switzerland), vol. 11, no. 17, Sep. 2021, doi: 10.3390/APP11178203.
[48] “What is solar power forecasting? – gridX.” Accessed: Sep. 17, 2024. [Online]. Available: https://www.gridx.ai/knowledge/what-is-solar-power-forecasting
[49] “Grid Connected PV System Connects PV Panels to the Grid.” Accessed: Sep. 17, 2024. [Online]. Available: https://www.alternative-energy-tutorials.com/solarpower/grid-connected-pv-system.html
[50] K. Barhmi, C. Heynen, S. Golroodbari, and W. van Sark, “A Review of Solar Forecasting Techniques and the Role of Artificial Intelligence,” Solar 2024, Vol. 4, Pages 99-135, vol. 4, no. 1, pp. 99–135, Feb. 2024, doi: 10.3390/SOLAR4010005.
[51] Z. Liu et al., “Artificial intelligence powered large-scale renewable integrations in multi-energy systems for carbon neutrality transition: Challenges and future perspectives,” Energy and AI, vol. 10, Nov. 2022, doi: 10.1016/J.EGYAI.2022.100195.
[52] D. Yang, “SolarData: An R package for easy access of publicly available solar datasets,” Solar Energy, vol. 171, pp. A3–A12, Sep. 2018, doi: 10.1016/J.SOLENER.2018.06.107.
[53] C. Voyant et al., “Machine learning methods for solar radiation forecasting: A review,” Renew Energy, vol. 105, pp. 569–582, 2017, doi:

10.1016/J.RENENE.2016.12.095.
[54] Y. Nie, X. Li, Q. Paletta, M. Aragon, A. Scott, and A. Brandt, "Open-Source Groundbased Sky Image Datasets for Very Short-term Solar Forecasting, Cloud Analysis and Modeling: A Comprehensive Survey," Nov. 2022, Accessed: Sep. 17, 2024. [Online]. Available: http://arxiv.org/abs/2211.14709
[55] "Solar Photovoltaic SPECIFICATION, CHECKLIST AND GUIDE Renewable Energy Ready Home".
[56] K. Barhmi, C. Heynen, S. Golroodbari, and W. van Sark, "A Review of Solar Forecasting Techniques and the Role of Artificial Intelligence," Solar 2024, Vol. 4, Pages 99-135, vol. 4, no. 1, pp. 99–135, Feb. 2024, doi: 10.3390/SOLAR4010005.
[57] A. P. Marugán, F. P. G. Márquez, J. M. P. Perez, and D. Ruiz-Hernández, "A survey of artificial neural network in wind energy systems," Appl Energy, vol. 228, pp. 1822–1836, Oct. 2018, doi: 10.1016/J.APENERGY.2018.07.084.
[58] N. Huang, R. Li, L. Lin, Z. Yu, and G. Cai, "Low Redundancy Feature Selection of Short Term Solar Irradiance Prediction Using Conditional Mutual Information and Gauss Process Regression," Sustainability 2018, Vol. 10, Page 2889, vol. 10, no. 8, p. 2889, Aug. 2018, doi: 10.3390/SU10082889.
[59] B. Jamil and N. Akhtar, "Comparative analysis of diffuse solar radiation models based on sky-clearness index and sunshine period for humid-subtropical climatic region of India: A case study," Renewable and Sustainable Energy Reviews, vol. 78, pp. 329–355, 2017, doi: 10.1016/J.RSER.2017.04.073.
[60] M. Fidan, F. O. Hocaoǀlu, and Ö. N. Gerek, "Harmonic analysis based hourly solar radiation forecasting model," IET Renewable Power Generation, vol. 9, no. 3, pp. 218–227, Apr. 2015, doi: 10.1049/IET-RPG.2014.0057.
[61] "Using LSTM Multivariate and Univariate in order to predict stock market. | by Van Nguyen | Medium." Accessed: Jul. 26, 2024. [Online]. Available: https://vannguyen8073.medium.com/using-lstm-multivariate-and-univariate-in-order-to-predict-stockmarket-39e8f6551c93
[62] A. Inanlouganji, T. A. Reddy, and S. Katipamula, "Evaluation of regression and neural network models for solar forecasting over different short-term horizons," Sci Technol Built Environ, vol. 24, no. 9, pp. 1004–1013, Oct. 2018, doi: 10.1080/23744731.2018.1464348.
[63] N. Bigdeli, M. Salehi Borujeni, and K. Afshar, "Time series analysis and short-term forecasting of solar irradiation, a new hybrid approach," Swarm Evol Comput, vol. 34, pp. 75–88, Jun. 2017, doi: 10.1016/J.SWEVO.2016.12.004.
[64] R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra, "Solar forecasting methods for renewable energy integration," Prog Energy Combust Sci, vol. 39, no. 6, pp. 535–576, Dec. 2013, doi: 10.1016/J.PECS.2013.06.002.
[65] G. Reikard, S. E. Haupt, and T. Jensen, "Forecasting ground-level irradiance over short horizons: Time series, meteorological, and time-varying parameter models," Renew Energy, vol. 112, pp. 474–485, 2017, doi: 10.1016/J.RENENE.2017.05.019.
[66] M. Kudo, A. Takeuchi, Y. Nozaki, H. Endo, and S. Jiro, "Forecasting electric power generation in a photovoltaic power system for an energy network," Electrical Engineering in Japan (English translation of Denki Gakkai Ronbunshi), vol. 167, no. 4, pp. 16–23, Jun. 2009, doi: 10.1002/EEJ.20755.
[67] K. Barhmi, C. Heynen, S. Golroodbari, and W. van Sark, "A Review of Solar Forecasting Techniques and the Role of Artificial Intelligence," Solar 2024, Vol. 4, Pages 99-135, vol. 4, no. 1, pp. 99–135, Feb. 2024, doi: 10.3390/SOLAR4010005.

[68] M. Vahabi, P. Lyakhov, and A. N. Bahar, “Design and implementation of novel efficient full adder/subtractor circuits based on quantum-dot cellular automata technology,” Applied Sciences (Switzerland), vol. 11, no. 18, Sep. 2021, doi: 10.3390/APP11188717.
[69] “How Solargis Optimizes Accuracy in Solar Power Predictions.” Accessed: Sep. 18, 2024. [Online]. Available: https://solargis.com/resources/blog/bestpractices/improving-accuracy-of-solar-power-forecasts
[70] M. Guermoui et al., “An analysis of case studies for advancing photovoltaic power forecasting through multi-scale fusion techniques,” Scientific Reports |, vol. 14, p. 6653, 123AD, doi: 10.1038/s41598-024-57398-z.
[71] Maciej Serda et al., “Synteza i aktywność biologiczna nowych analogów tiosemikarbazonowych chelatorów żelaza,” Uniwersytet śląski, vol. 7, no. 1, pp. 343–354, 2013, doi: 10.2/JQUERY.MIN.JS.
[72] S. Borovkova and I. Tsiamas, “An ensemble of LSTM neural networks for highfrequency stock market classification,” 2019, doi: 10.1002/for.2585.
[73] A. Gozhyj, V. Nechakhin, and I. Kalinina, “Enhancing solar panel efficiency with LSTM-based MPPT controllers,” 2024.
[74] Y. Bai, Z. Yan, and C. Li, “An attention mechanism-based deep regression approach with a sequence decomposition-granularity reconstruction-integration model for urban daily water supply forecasting,” J Hydrol (Amst), vol. 617, Feb. 2023, doi: 10.1016/j.jhydrol.2022.129032.
[75] Maciej Serda et al., “Synteza i aktywność biologiczna nowych analogów tiosemikarbazonowych chelatorów żelaza,” Uniwersytet śląski, vol. 7, no. 1, pp. 343–354, 2013, doi: 10.2/JQUERY.MIN.JS.
[76] R. Ahmed, V. Sreeram, Y. Mishra, and M. D. Arif, “A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization,” Renewable and Sustainable Energy Reviews, vol. 124, p. 109792, May 2020, doi: 10.1016/J.RSER.2020.109792.
[77] Y. Nie, X. Li, Q. Paletta, M. Aragon, A. Scott, and A. Brandt, “Open-Source Groundbased Sky Image Datasets for Very Short-term Solar Forecasting, Cloud Analysis and Modeling: A Comprehensive Survey,” Nov. 2022, Accessed: Sep. 13, 2024. [Online]. Available: http://arxiv.org/abs/2211.14709
[78] R. Srivastava, M. Amir, F. Ahmad, S. K. Agrawal, A. Dwivedi, and A. K. Yadav, “Performance evaluation of grid connected solar powered microgrid: A case study,” Front Energy Res, vol. 10, Nov. 2022, doi: 10.3389/FENRG.2022.1044651/FULL.
[79] S. S. Torbaghan, A. Motamedi, H. Zareipour, and L. A. Tuan, “Medium-term electricity price forecasting,” 2012 North American Power Symposium, NAPS 2012, 2012, doi: 10.1109/NAPS.2012.6336424.
[80] P. N. Liyana, S. Mekhilef, and N. M. Shah, “Photovoltaic Performance Using Modify Back Propagation Neural Network Based On 3 Different Solar Panels,” IOP conference series, vol. 1008, no. 1, pp. 012010–012010, Apr. 2022, doi: 10.1088/1755-1315/1008/1/012010.
[81] A. Nespoli et al., “Day-ahead photovoltaic forecasting: A comparison of the most effective techniques,” Energies (Basel), vol. 12, no. 9, 2019, doi: 10.3390/EN12091621.
[82] S. Monjoly, M. André, R. Calif, and T. Soubdhan, “Forecast horizon and solar variability influences on the performances of multiscale hybrid forecast model,” Energies (Basel), vol. 12, no. 11, Jun. 2019, doi: 10.3390/EN12122264.
[83] S. Chaaraoui et al., “Day-ahead electric load forecast for a ghanaian health facility using different algorithms,” Energies (Basel), vol. 14, no. 2, Jan. 2021, doi:

10.3390/EN14020409.
[84] M. Naveed Akhter et al., "A hybrid deep learning method for an hour ahead power output forecasting of three different photovoltaic systems Hybrid deep learning PV systems SSA-RNN-LSTM PSO-RNN-LSTM GA-RNN-LSTM," Appl Energy, vol. 307, p. 118185, 2022, doi: 10.1016/j.apenergy.2021.118185.
[85] M. N. Akhter et al., "An hour-ahead PV power forecasting method based on an RNNLSTM model for three different PV plants," Energies (Basel), vol. 15, no. 6, p. 2243, Mar. 2022, doi: 10.3390/en15062243.
[86] K. Kawaguchi, Y. Bengio, and L. Kaelbling, "Generalization in Deep Learning," Mathematical Aspects of Deep Learning, pp. 112–148, Nov. 2022, doi: 10.1017/9781009025096.003.
[87] M. Bouzerdoum, A. Mellit, and A. Massi Pavan, "A hybrid model (SARIMA-SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant," Solar Energy, vol. 98, no. PC, pp. 226–235, Dec. 2013, doi: 10.1016/J.SOLENER.2013.10.002.
[88] H. H. Goh, S. W. Lee, Q. S. Chua, K. C. Goh, B. C. Kok, and K. T. K. Teo, "Renewable energy project: Project management, challenges and risk," Renewable and Sustainable Energy Reviews, vol. 38, pp. 917–932, 2014, doi: 10.1016/J.RSER.2014.07.078.
[89] S. Shamshirband, T. Rabczuk, and K. W. Chau, "A Survey of Deep Learning Techniques: Application in Wind and Solar Energy Resources," IEEE Access, vol. 7, pp. 164650–164666, 2019, doi: 10.1109/ACCESS.2019.2951750.
[90] "What are Recurrent Neural Networks?" Accessed: Jul. 22, 2024. [Online]. Available: https://www.analyticsvidhya.com/blog/2022/03/a-brief-overview-of-recurrent-neuralnetworks-rnn/
[91] D. Gielen, F. Boshell, D. Saygin, M. D. Bazilian, N. Wagner, and R. Gorini, "The role of renewable energy in the global energy transformation," Energy Strategy Reviews, vol. 24, pp. 38–50, Apr. 2019, doi: 10.1016/J.ESR.2019.01.006.
[92] M. Rinio, "PVcheck—A Software to Check Your Photovoltaic System," Energies 2021, Vol. 14, Page 6757, vol. 14, no. 20, p. 6757, Oct. 2021, doi: 10.3390/EN14206757.
[93] G. Reikard, "Predicting solar radiation at high resolutions: A comparison of time series forecasts," Solar Energy, vol. 83, no. 3, pp. 342–349, Mar. 2009, doi: 10.1016/J.SOLENER.2008.08.007.
[94] S. Bruno, G. Dellino, M. La Scala, and C. Meloni, "A microforecasting module for energy management in residential and tertiary buildings," Energies (Basel), vol. 12, no. 6, 2019, doi: 10.3390/EN12061006.
[95] D. E. Amos, "Algorithm 644: A Portable Package for Bessel Functions of a Complex Argument and Nonnegative Order," ACM Transactions on Mathematical Software (TOMS), vol. 12, no. 3, pp. 265–273, Sep. 1986, doi: 10.1145/7921.214331.
[96] M. Abdel-Nasser and K. Mahmoud, "Accurate photovoltaic power forecasting models using deep LSTM-RNN," Neural Comput Appl, vol. 31, no. 7, pp. 2727–2740, Jul. 2019, doi: 10.1007/S00521-017-3225-Z.
[97] S. Bouktif, A. Fiaz, A. Ouni, and M. A. Serhani, "Optimal deep learning LSTM model for electric load forecasting using feature selection and genetic algorithm: Comparison with machine learning approaches," Energies (Basel), vol. 11, no. 7, 2018, doi: 10.3390/EN11071636.
[98] S. Tajjour, S. S. Chandel, M. A. Alotaibi, H. Malik, F. P. Garcia Marquez, and A. Afthanorhan, "Short-Term Solar Irradiance Forecasting Using Deep Learning

Techniques: A Comprehensive Case Study," IEEE Access, vol. 11, pp. 119851–119861, 2023, doi: 10.1109/ACCESS.2023.3325292.
[99] "Short-Term Solar Irradiance Forecasting Using Deep Learning Techniques: A Comprehensive Case Study." Accessed: Sep. 23, 2024. [Online]. Available: https://www.openread.academy/en/paper/reading?corpusId=264337422
[100] M. Kumari, C. Kumar, and S. Srivastava, "A study on the estimation of global horizontal irradiance via deep learning technique," 2023 International Conference on Recent Advances in Electrical, Electronics and Digital Healthcare Technologies, REEDCON 2023, pp. 476–481, 2023, doi: 10.1109/REEDCON57544.2023.10151174.
[101] "A study on the estimation of global horizontal irradiance via deep learning technique." Accessed: Sep. 23, 2024. [Online]. Available: https://www.openread.academy/en/paper/reading?corpusId=259179718
[102] Z. Zhen et al., "Pattern Classification and PSO Optimal Weights Based Sky Images Cloud Motion Speed Calculation Method for Solar PV Power Forecasting," IEEE Trans Ind Appl, vol. 55, no. 4, pp. 3331–3342, Jul. 2019, doi: 10.1109/TIA.2019.2904927.
[103] F. Wang, Z. Zhen, B. Wang, and Z. Mi, "Comparative study on KNN and SVM based weather classification models for day ahead short term solar PV power forecasting," Applied Sciences (Switzerland), vol. 8, no. 1, Dec. 2017, doi: 10.3390/APP8010028.
[104] F. Wang, Z. Zhen, C. Liu, Z. Mi, M. Shafie-Khah, and J. P. S. Catalão, "Time-section fusion pattern classification based day-ahead solar irradiance ensemble forecasting model using mutual iterative optimization," Energies (Basel), vol. 11, no. 1, 2018, doi: 10.3390/EN11010184.
[105] D. Alhakeem, P. Mandal, A. U. Haque, A. Yona, T. Senjyu, and T. L. Tseng, "A new strategy to quantify uncertainties of wavelet-GRNN-PSO based solar PV power forecasts using bootstrap confidence intervals," IEEE Power and Energy Society General Meeting, vol. 2015-September, Sep. 2015, doi: 10.1109/PESGM.2015.7286233.
[106] "Applied Sciences | Special Issue : Advanced Information Processing Methods and Their Applications." Accessed: May 24, 2024. [Online]. Available: https://www.mdpi.com/journal/applsci/special_issues/information_processing_methods
[107] P. Lyakhov, "Special Issue on Advanced Information Processing Methods and Their Applications," Applied Sciences (Switzerland), vol. 12, no. 18, Sep. 2022, doi: 10.3390/APP12189090.
[108] A. Ziane et al., "Photovoltaic output power performance assessment and forecasting: Impact of meteorological variables," Solar Energy, vol. 220, pp. 745–757, May 2021, doi: 10.1016/J.SOLENER.2021.04.004.
[109] K. Yan, H. Shen, L. Wang, H. Zhou, M. Xu, and Y. Mo, "Short-term solar irradiance forecasting based on a hybrid deep learning methodology," Information (Switzerland), vol. 11, no. 1, Jan. 2020, doi: 10.3390/INFO11010032.
[110] C. F. Yen, H. Y. Hsieh, K. W. Su, M. C. Yu, and J. S. Leu, "Solar Power Prediction via Support Vector Machine and Random Forest," E3S Web of Conferences, vol. 69, Nov. 2018, doi: 10.1051/E3SCONF/20186901004.
[111] J. D. De Guia, R. S. Concepcion, H. A. Calinao, J. Alejandrino, E. P. Dadios, and E. Sybingco, "Using stacked long short term memory with principal component analysis for short term prediction of solar irradiance based on weather patterns," IEEE Region 10 Annual International Conference, Proceedings/TENCON, vol. 2020-November, pp. 946–951, Nov. 2020, doi: 10.1109/TENCON50793.2020.9293719.

[112] M. A. Hassan, A. Khalil, S. Kaseb, and M. A. Kassem, “Exploring the potential of tree-based ensemble methods in solar radiation modeling,” Appl Energy, vol. 203, pp. 897–916, 2017, doi: 10.1016/j.apenergy.2017.06.104.
[113] H. Yu, Y. Chen, S. G. Hassan, and D. Li, “Prediction of the temperature in a Chinese solar greenhouse based on LSSVM optimized by improved PSO,” Comput Electron Agric, vol. 122, pp. 94–102, Mar. 2016, doi: 10.1016/j.compag.2016.01.019.
[114] S. M. Ebrahimi, E. Salahshour, M. Malekzadeh, and Francisco Gordillo, “Parameters identification of PV solar cells and modules using flexible particle swarm optimization algorithm,” Energy, vol. 179, pp. 358–372, Jul. 2019, doi: 10.1016/j.energy.2019.04.218.
[115] T. T. K. Tran et al., “Improving the prediction of wildfire susceptibility on Hawai‘i Island, Hawai‘i, using explainable hybrid machine learning models,” J Environ Manage, vol. 351, Feb. 2024, doi: 10.1016/j.jenvman.2023.119724.
[116] “Flowchart of the PSO algorithm. | Download Scientific Diagram.” Accessed: Sep. 27, 2024. [Online]. Available: https://www.researchgate.net/figure/Flowchart-of-thePSO-algorithm_fig2_325156684
[117] A. M. Ismail, R. Ramirez-Iniguez, M. Asif, A. B. Munir, and F. Muhammad-Sukki, “Progress of solar photovoltaic in ASEAN countries: A review,” Renewable and Sustainable Energy Reviews, vol. 48, pp. 399–412, Aug. 2015, doi: 10.1016/J.RSER.2015.04.010.
[118] M. A. Munir, A. Khattak, K. Imran, A. Ulasyar, and A. Khan, “Solar PV Generation Forecast Model Based on the Most Effective Weather Parameters,” 1st International Conference on Electrical, Communication and Computer Engineering, ICECCE 2019, Jul. 2019, doi: 10.1109/ICECCE47252.2019.8940664.
[119] K. M. Alawasa and A. I. Al-Odienat, “Power quality characteristics of residential gridconnected inverter ofphotovoltaic solar system,” pp. 1097–1101, Dec. 2017, doi: 10.1109/ICRERA.2017.8191225.
[120] J. Han and W. K. Park, “A Solar Radiation Prediction Model Using Weather Forecast Data and Regional Atmospheric Data,” 2018 IEEE 7th World Conference on Photovoltaic Energy Conversion, WCPEC 2018 - A Joint Conference of 45th IEEE PVSC, 28th PVSEC and 34th EU PVSEC, pp. 2313–2316, Nov. 2018, doi: 10.1109/PVSC.2018.8547750.
[121] H. Khorasanizadeh, K. Mohammadi, and N. Goudarzi, “Prediction of horizontal diffuse solar radiation using clearness index based empirical models; A case study,” Int J Hydrogen Energy, vol. 41, no. 47, pp. 21888–21898, Dec. 2016, doi: 10.1016/J.IJHYDENE.2016.09.198.
[122] N. E. Benti, M. D. Chaka, and A. G. Semie, “Forecasting Renewable Energy Generation with Machine Learning and Deep Learning: Current Advances and Future Prospects,” Sustainability 2023, Vol. 15, Page 7087, vol. 15, no. 9, p. 7087, Apr. 2023, doi: 10.3390/SU15097087.
[123] A. Jäger-Waldau, “Snapshot of photovoltaics - March 2021,” EPJ Photovoltaics, vol. 12, 2021, doi: 10.1051/EPJPV/2021002.
[124] T. A. Woldegiyorgis et al., “Estimating solar radiation using artificial neural networks: a case study of Fiche, Oromia, Ethiopia,” Cogent Eng, vol. 10, 2023, doi: 10.1080/23311916.2023.2220489.
[125] N. Omar, H. Aly, and T. Little, “LSTM and RBFNN Based Univariate and Multivariate Forecasting of Day-ahead Solar Irradiance for Atlantic Region in Canada and Mediterranean Region in Libya,” 2021 4th International Conference on Energy, Electrical and Power Engineering, CEEPE 2021, pp. 1130–1135, Apr. 2021, doi: 10.1109/CEEPE51765.2021.9475652.

[126] V. A. Tikkiwal, S. Vir Singh, and H. O. Gupta, “Day-Ahead forecasting of solar irradiance using hybrid improved Cuckoo Search-LSTM approach,” Proceedings - IEEE 2020 2nd International Conference on Advances in Computing, Communication Control and Networking, ICACCCN 2020, pp. 84–88, Dec. 2020, doi: 10.1109/ICACCCN51052.2020.9362839.
[127] H.-J. ; Jeon et al., “Day-Ahead Hourly Solar Irradiance Forecasting Based on MultiAttributed Spatio-Temporal Graph Convolutional Network,” Sensors 2022, Vol. 22, Page 7179, vol. 22, no. 19, p. 7179, Sep. 2022, doi: 10.3390/S22197179.
[128] X. Qing and Y. Niu, “Hourly day-ahead solar irradiance prediction using weather forecasts by LSTM,” Energy, vol. 148, pp. 461–468, Apr. 2018, doi: 10.1016/J.ENERGY.2018.01.177.
[129] A. F. Narynbaev and A. G. Vaskov, “Day-Ahead Solar Irradiance Forecasting Model,” Proceedings - International Ural Conference on Measurements, UralCon, vol. 2021September, pp. 1–6, 2021, doi: 10.1109/URALCON52005.2021.9559464.
[130] H. Y. Cheng, C. C. Yu, and C. L. Lin, “Day-ahead to week-ahead solar irradiance prediction using convolutional long short-term memory networks,” Renew Energy, vol.
179, pp. 2300–2308, Dec. 2021, doi: 10.1016/J.RENENE.2021.08.038.
[131] P. Gupta and R. Singh, “Univariate model for hour ahead multi-step solar irradiance forecasting,” Conference Record of the IEEE Photovoltaic Specialists Conference, pp. 494–501, Jun. 2021, doi: 10.1109/PVSC43889.2021.9519002.
[132] “Diffuse solar radiation on a horizontal surface: Reviewing and categorizing the empirical models | Request PDF.” Accessed: Jun. 09, 2024. [Online]. Available: https://www.researchgate.net/publication/281069257_Diffuse_solar_radiation_on_a_horizontal_surface_Reviewing_and_categorizing_the_empirical_models
[133] Munoz, MN, Ballantyne, Eef, Stone, and Da, “Development and evaluation of empirical models for the estimation of hourly horizontal diffuse solar irradiance in the United Kingdom”, doi: 10.1016/j.energy.2021.122820.
[134] H. Fraihat, A. A. Almbaideen, A. Al-Odienat, B. Al-Naami, R. De Fazio, and P. Visconti, “Solar Radiation Forecasting by Pearson Correlation Using LSTM Neural Network and ANFIS Method: Application in the West-Central Jordan,” Future Internet 2022, Vol. 14, Page 79, vol. 14, no. 3, p. 79, Mar. 2022, doi: 10.3390/FI14030079.
[135] E. Subramanian, M. M. Karthik, G. P. Krishna, V. S. Kumar, and D. Vaisnav, “SOLAR POWER PREDICTION USING MACHINE LEARNING”.
[136] S. Borgohain, T. H. Kazmi, R. Peesapati, and G. Panda, “IoT based Solar Power Forecasting using AMA-ELM Technique,” 5th International Conference on Energy, Power, and Environment: Towards Flexible Green Energy Technologies, ICEPE 2023, 2023, doi: 10.1109/ICEPE57949.2023.10201641.
[137] V. Bajpai, S. Joshi, and H. G. Yeh, “Forecasting the Wind and Solar Power using Deep Learning based techniques,” 2023 IEEE Green Energy and Smart Systems Conference, IGESSC 2023, 2023, doi: 10.1109/IGESSC59090.2023.10321764.
[138] Jāmiʻat Mu’tah, Jordanian Engineers Association, and Institute of Electrical and Electronics Engineers, “2022 International Conference on Emerging Trends in Computing and Engineering Applications (ETCEA) : proceedings : Mutah University, Karak, Jordan, 23-25 November 2022.”.
[139] A. K. Parida, D. Kumar, R. R. Sahoo, and B. K. Balabantaray, “Medium term Solar Power Prediction using stacked LSTM based deep learning Technique,” Proceedings of 4th International Conference on Cybernetics, Cognition and Machine Learning Applications, ICCCMLA 2022, pp. 19–26, 2022, doi:

10.1109/ICCCMLA56841.2022.9989067.
[140] M. Elsaraiti and A. Merabet, “Solar Power Forecasting Using Deep Learning Techniques,” IEEE Access, vol. 10, pp. 31692–31698, Jan. 2022, doi: 10.1109/ACCESS.2022.3160484.
[141] P. Kumar, P. Sihag, P. Chaturvedi, K. V. Uday, and V. Dutt, “BS-LSTM: An Ensemble Recurrent Approach to Forecasting Soil Movements in the Real World,” Front Earth Sci (Lausanne), vol. 9, p. 696792, Aug. 2021, doi: 10.3389/FEART.2021.696792/BIBTEX.
[142] R. K. Behera, D. Naik, D. Ramesh, and S. K. Rath, “MR-IBC: MapReduce-based incremental betweenness centrality in large-scale complex networks,” Soc Netw Anal Min, vol. 10, no. 1, Dec. 2020, doi: 10.1007/S13278-020-00636-9.
[143] P. Priyanka, P. Kumar, S. Panda, T. Thakur, K. V. Uday, and V. Dutt, “Can machine learning models predict soil moisture evaporation rates? An investigation via novel feature selection techniques and model comparisons,” Front Earth Sci (Lausanne), vol. 12, 2024, doi: 10.3389/FEART.2024.1344690/FULL.
[144] R. K. Behera, D. Naik, S. K. Rath, and R. Dharavath, “Genetic algorithm-based community detection in large-scale social networks,” Neural Comput Appl, vol. 32, no. 13, pp. 9649–9665, Jul. 2020, doi: 10.1007/S00521-019-04487-0.
[145] H. A. H. Al-Najjar and B. Pradhan, “Spatial landslide susceptibility assessment using machine learning techniques assisted by additional data created with generative adversarial networks,” Geoscience Frontiers, vol. 12, no. 2, pp. 625–637, Mar. 2021, doi: 10.1016/J.GSF.2020.09.002.
[146] N. Mali, V. Dutt, and K. V. Uday, “Determining the Geotechnical Slope Failure Factors via Ensemble and Individual Machine Learning Techniques: A Case Study in Mandi, India,” Front Earth Sci (Lausanne), vol. 9, Sep. 2021, doi: 10.3389/FEART.2021.701837/FULL.
[147] N. Pei et al., “Interval Prediction of the Permeability of Granite Bodies in a HighLevel Radioactive Waste Disposal Site Using LSTM-RNNs and Probability Distribution,” Front Earth Sci (Lausanne), vol. 10, Feb. 2022, doi: 10.3389/FEART.2022.835308/FULL.
[148] P. Chawla, J. Z. Gao, T. Gao, C. Luo, H. Li, and Y. We, “An interactive web-based solar energy prediction system using machine learning techniques,” Journal of Management Analytics, vol. 10, no. 2, pp. 308–335, 2023, doi: 10.1080/23270012.2023.2209883.
[149] A. K. Bansal, V. S. Sangtani, P. Dadheech, N. Aneja, and U. Yahya, “Biogeographybased Optimization of Artificial Neural Network (BBO-ANN) for Solar Radiation Forecasting,” Applied Artificial Intelligence, vol. 37, no. 1, 2023, doi: 10.1080/08839514.2023.2166705.
[150] H. Ali-Ou-Salah, B. Oukarfi, and T. Mouhaydine, “Short-term solar radiation forecasting using a new seasonal clustering technique and artificial neural network,” Int J Green Energy, vol. 19, no. 4, pp. 424–434, Mar. 2022, doi: 10.1080/15435075.2021.1946819.
[151] B. Gao, X. Huang, J. Shi, Y. Tai, and J. Zhang, “Hourly forecasting of solar irradiance based on CEEMDAN and multi-strategy CNN-LSTM neural networks,” Renew Energy, vol. 162, pp. 1665–1683, Dec. 2020, doi: 10.1016/j.renene.2020.09.141.
[152] R. Yuan et al., “Daily Runoff Forecasting Using Ensemble Empirical Mode Decomposition and Long Short-Term Memory,” Front Earth Sci (Lausanne), vol. 9, Mar. 2021, doi: 10.3389/FEART.2021.621780/FULL.
[153] “python - How to implying bagging method for LSTM neural network using pyTorch? - Stack Overflow.” Accessed: Sep. 28, 2024. [Online]. Available:

https://stackoverflow.com/questions/59560372/how-to-implying-bagging-method-forlstm-neural-network-using-pytorch
[154] M. P. PERRONE and L. N. COOPER, "When networks disagree: Ensemble methods for hybrid neural networks," pp. 342–358, Sep. 1995, doi: 10.1142/9789812795885_0025.
[155] A. N. Jahromi, S. Hashemi, A. Dehghantanha, R. M. Parizi, and K. K. R. Choo, "An Enhanced Stacked LSTM Method with No Random Initialization for Malware Threat Hunting in Safety and Time-Critical Systems," IEEE Trans Emerg Top Comput Intell, vol. 4, no. 5, pp. 630–640, Oct. 2020, doi: 10.1109/TETCI.2019.2910243.
[156] X. Dai, H. Yin, and N. K. Jha, "Grow and Prune Compact, Fast, and Accurate LSTMs," IEEE Transactions on Computers, vol. 69, no. 3, pp. 441–452, Mar. 2020, doi: 10.1109/TC.2019.2954495.
[157] "(PDF) An Effective Hybrid NARX-LSTM Model for Point and Interval PV Power Forecasting." Accessed: Sep. 28, 2024. [Online]. Available: https://www.researchgate.net/publication/349654976_An_Effective_Hybrid_NARX-LSTM_Model_for_Point_and_Interval_PV_Power_Forecasting
[158] "Deep Concatenated Residual Network with Bidirectional LSTM for One-Hour-Ahead Wind Power Forecasting — Yonsei University." Accessed: Sep. 28, 2024. [Online]. Available: https://yonsei.elsevierpure.com/en/publications/deep-concatenated-residualnetwork-with-bidirectional-lstm-for-on
[159] Y. Wang, Y. Shen, S. Mao, X. Chen, and H. Zou, "LASSO and LSTM integrated temporal model for short-term solar intensity forecasting," IEEE Internet Things J, vol. 6, no. 2, pp. 2933–2944, Apr. 2019, doi: 10.1109/JIOT.2018.2877510.
[160] M. S. Ko, K. Lee, J. K. Kim, C. W. Hong, Z. Y. Dong, and K. Hur, "Deep Concatenated Residual Network with Bidirectional LSTM for One-Hour-Ahead Wind Power Forecasting," IEEE Trans Sustain Energy, vol. 12, no. 2, pp. 1321–1335, Apr. 2021, doi: 10.1109/TSTE.2020.3043884.
[161] Y. Wang, Y. Shen, S. Mao, S. Member, G. Cao, and R. M. Nelms, "Adaptive Learning Hybrid Model for Solar Intensity Forecasting," 2018, doi: 10.1109/TII.2017.2789289.
[162] C. Zhu et al., "Hybrid machine learning and optimization method for solar irradiance forecasting," Engineering Optimization, Sep. 2024, doi: 10.1080/0305215X.2024.2390126.
[163] N. Tang, S. Mao, Y. Wang, and R. M. Nelms, "Solar Power Generation Forecasting With a LASSO-Based Approach," IEEE Internet Things J, vol. 5, no. 2, pp. 1090–1099, Apr. 2018, doi: 10.1109/JIOT.2018.2812155.
[164] G. Huang, Y. Li, G. Pleiss, Z. Liu, J. E. Hopcroft, and K. Q. Weinberger, "Snapshot ensembles: Train 1, get M for free," 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings, 2017.
[165] M. Alizamir et al., "Improving the accuracy of daily solar radiation prediction by climatic data using an efficient hybrid deep learning model: Long short-term memory (LSTM) network coupled with wavelet transform," Eng Appl Artif Intell, vol. 123, p. 106199, Aug. 2023, doi: 10.1016/J.ENGAPPAI.2023.106199.
[166] K. Yao, T. Cohn, K. Vylomova, K. Duh, and C. Dyer, "Depth-Gated LSTM," Aug. 2015, Accessed: Oct. 01, 2024. [Online]. Available: http://arxiv.org/abs/1508.03790
[167] M. Perera, J. De Hoog, K. Bandara, D. Senanayake, and S. Halgamuge, "Day-ahead regional solar power forecasting with hierarchical temporal convolutional neural networks using historical power generation and weather data," Appl Energy, vol. 361, p. 122971, May 2024, doi: 10.1016/J.APENERGY.2024.122971.

[168] M. AlKandari and I. Ahmad, "Solar power generation forecasting using ensemble approach based on deep learning and statistical methods," Applied Computing and Informatics, vol. 20, no. 3–4, pp. 231–250, Jun. 2024, doi: 10.1016/J.ACI.2019.11.002/FULL/PDF.
[169] M. Y. Cheng and M. T. Cao, "Accurately predicting building energy performance using evolutionary multivariate adaptive regression splines," Applied Soft Computing Journal, vol. 22, pp. 178–188, 2014, doi: 10.1016/j.asoc.2014.05.015.
[170] M. Alizamir et al., "Prediction of daily chlorophyll-a concentration in rivers by water quality parameters using an efficient data-driven model: online sequential extreme learning machine," Acta Geophysica, vol. 69, no. 6, pp. 2339–2361, Dec. 2021, doi: 10.1007/S11600-021-00678-3.
[171] H. Citakoglu, B. Babayigit, and N. A. Haktanir, "Solar radiation prediction using multi-gene genetic programming approach," Theor Appl Climatol, vol. 142, no. 3–4, pp. 885–897, Nov. 2020, doi: 10.1007/S00704-020-03356-4.
[172] J. M. Wooldridge, Introductory Econometrics: A Modern Approach, Fourth. Australia: South-Western, 2009.
[173] R. Song, Z. Wang, L. Guo, F. Zhao, and Z. Xu, "Deep Belief Networks (DBN) for Financial Time Series Analysis and Market Trends Prediction," World Journal of Innovation and Modern Technology, vol. 7, no. 4, pp. 1–10, Jul. 2024, doi: 10.53469/WJIMT.2024.07(04).01.
[174] S. Wang, H. Zheng, X. Wen, and S. Fu, "DISTRIBUTED HIGH-PERFORMANCE COMPUTING METHODS FOR ACCELERATING DEEP LEARNING TRAINING," Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (online), vol. 3, no. 3, pp. 108–126, Sep. 2024, doi: 10.60087/JKLST.V3.N4.P22.
[175] F. S. Tymvios, C. P. Jacovides, S. C. Michaelides, and C. Scouteli, "Comparative study of Ångström's and artificial neural networks' methodologies in estimating global solar radiation," Solar Energy, vol. 78, no. 6, pp. 752–762, Jun. 2005, doi: 10.1016/j.solener.2004.09.007.
[176] L. Olatomiwa, S. Mekhilef, S. Shamshirband, K. Mohammadi, D. Petković, and C. Sudheer, "A support vector machine-firefly algorithm-based model for global solar radiation prediction," Solar Energy, vol. 115, pp. 632–644, May 2015, doi: 10.1016/j.solener.2015.03.015.
[177] D. J. Armaghani, E. T. Mohamad, M. S. Narayanasamy, N. Narita, and S. Yagiz, "Development of hybrid intelligent models for predicting TBM penetration rate in hard rock condition," Tunnelling and Underground Space Technology, vol. 63, pp. 29–43, Mar. 2017, doi: 10.1016/j.tust.2016.12.009.
[178] A. Das, J. ki Park, and J. hwa Park, "Estimation of available global solar radiation using sunshine duration over South Korea," J Atmos Sol Terr Phys, vol. 134, pp. 22–29, Nov. 2015, doi: 10.1016/j.jastp.2015.09.001.
[179] D. Dell'Aquila and M. Russo, "Automatic classification of nuclear physics data via a Constrained Evolutionary Clustering approach," Comput Phys Commun, vol. 259, Feb. 2021, doi: 10.1016/j.cpc.2020.107667.
[180] S. Rehman, "Solar radiation over Saudi Arabia and comparisons with empirical models," Energy, vol. 23, no. 12, pp. 1077–1082, 1998, doi: 10.1016/S03605442(98)00057-7.
[181] J. Zhang et al., "A suite of metrics for assessing the performance of solar power forecasting," Solar Energy, vol. 111, pp. 157–175, Jan. 2015, doi: 10.1016/J.SOLENER.2014.10.016.
[182] F. Becattini and T. Uricchio, "Memory Networks," MM 2022 - Proceedings of the 30th ACM International Conference on Multimedia, pp. 7380–7382, Oct. 2022, doi:

10.1145/3503161.3546972.
[183] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” 2nd International Conference on Learning Representations, ICLR 2014 - Conference Track Proceedings, 2014, doi: 10.61603/ceas.v2i1.33.
[184] S. Sukhbaatar, A. Szlam, J. Weston, and R. Fergus, “End-to-end memory networks,” Adv Neural Inf Process Syst, vol. 2015-January, pp. 2440–2448, 2015.
[185] Z. Zhen et al., “SVM based cloud classification model using total sky images for PV power forecasting,” 2015 IEEE Power and Energy Society Innovative Smart Grid Technologies Conference, ISGT 2015, Jun. 2015, doi: 10.1109/ISGT.2015.7131784.
[186] E. Grefenstette, K. M. Hermann, M. Suleyman, and P. Blunsom, “Learning to transduce with unbounded memory,” Adv Neural Inf Process Syst, vol. 2015-January, pp. 1828–1836, 2015.
[187] Y. S. Park and S. Lek, “Artificial Neural Networks: Multilayer Perceptron for Ecological Modeling,” Developments in Environmental Modelling, vol. 28, pp. 123–140, 2016, doi: 10.1016/B978-0-444-63623-2.00007-4.
[188] S. Walczak and N. Cerpa, “Artificial Neural Networks,” Encyclopedia of Physical Science and Technology, pp. 631–645, 2003, doi: 10.1016/B0-12-227410-5/00837-1.
[189] “What is Exploratory Data Analysis? - GeeksforGeeks.” Accessed: Sep. 11, 2024. [Online]. Available: https://www.geeksforgeeks.org/what-is-exploratory-data-analysis/
[190] “7 Exploratory Data Analysis | R for Data Science.” Accessed: Sep. 11, 2024. [Online]. Available: https://r4ds.had.co.nz/exploratory-data-analysis.html
[191] Z. Shao et al., “Exploring Progress in Multivariate Time Series Forecasting: Comprehensive Benchmarking and Heterogeneity Analysis,” Oct. 2023, Accessed: Sep. 11, 2024. [Online]. Available: http://arxiv.org/abs/2310.06119
[192] Z. Shao et al., “Exploring Progress in Multivariate Time Series Forecasting: Comprehensive Benchmarking and Heterogeneity Analysis,” Oct. 2023, Accessed: Sep. 11, 2024. [Online]. Available: https://arxiv.org/abs/2310.06119v1
[193] M. Bozorg, A. Bracale, P. Caramia, G. Carpinelli, M. Carpita, and P. De Falco, “Bayesian bootstrap quantile regression for probabilistic photovoltaic power forecasting,” Protection and Control of Modern Power Systems, vol. 5, no. 1, Dec. 2020, doi: 10.1186/S41601-020-00167-7.
[194] J. Li, J. K. Ward, J. Tong, L. Collins, and G. Platt, “Machine learning for solar irradiance forecasting of photovoltaic system,” Renew Energy, vol. 90, pp. 542–553, May 2016, doi: 10.1016/J.RENENE.2015.12.069.
[195] D. S. Kumar, G. M. Yagli, M. Kashyap, and D. Srinivasan, “Solar irradiance resource and forecasting: a comprehensive review,” IET Renewable Power Generation, vol. 14, no. 10, pp. 1641–1656, Jul. 2020, doi: 10.1049/IET-RPG.2019.1227.
[196] S. Ertekin, “Solar Power Prediction with an Hour-based Ensemble Machine Learning Method,” Hittite Journal of Science & Engineering, vol. 7, no. 1, pp. 35–40, Mar. 2020, doi: 10.17350/HJSE19030000169.
[197] R. P. Masini, M. C. Medeiros, and E. F. Mendes, “Machine learning advances for time series forecasting,” J Econ Surv, vol. 37, no. 1, pp. 76–111, Feb. 2023, doi: 10.1111/JOES.12429.
[198] “Time Series Prediction of Solar Power Generation Using Trend Decomposition - Kavakci - 2024 - Energy Technology - Wiley Online Library.” Accessed: Oct. 11, 2024. [Online]. Available: https://onlinelibrary.wiley.com/doi/full/10.1002/ente.202300914
[199] D. A. Wood, “Trend decomposition aids short-term countrywide wind capacity factor forecasting with machine and deep learning methods,” Energy Convers Manag, vol. 253, Feb. 2022, doi: 10.1016/J.ENCONMAN.2021.115189.

[200] P. Kumari and D. Toshniwal, "Deep learning models for solar irradiance forecasting: A comprehensive review," J Clean Prod, vol. 318, Oct. 2021, doi: 10.1016/J.JCLEPRO.2021.128566.
[201] S. M. J. Jalali, S. Ahmadian, A. Kavousi-Fard, A. Khosravi, and S. Nahavandi, "Automated Deep CNN-LSTM Architecture Design for Solar Irradiance Forecasting," IEEE Trans Syst Man Cybern Syst, vol. 52, no. 1, pp. 54–65, Jan. 2022, doi: 10.1109/TSMC.2021.3093519.
[202] Ü. Ç. Büyükşahin and Ş. Ertekin, "Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition," Neurocomputing, vol. 361, pp. 151–163, Oct. 2019, doi: 10.1016/J.NEUCOM.2019.05.099.
[203] "Grow and Prune Compact, Fast, and Accurate LSTMs | IEEE Journals & Magazine | IEEE Xplore." Accessed: Sep. 28, 2024. [Online]. Available: https://ieeexplore.ieee.org/document/8907435
[204] Y. Yu, J. Cao, and J. Zhu, "An LSTM Short-Term Solar Irradiance Forecasting under Complicated Weather Conditions," IEEE Access, vol. 7, pp. 145651–145666, 2019, doi: 10.1109/ACCESS.2019.2946057.
[205] "What is MPPT?" Accessed: Jan. 11, 2025. [Online]. Available: https://www.redarcelectronics.com/au/resources/solar-faqs/what-is-mppt/
[206] "MPPT vs PWM | Different Types of Solar Charge Controllers." Accessed: Jan. 11, 2025. [Online]. Available: https://www.morningstarcorp.com/faq/what-are-thedifferent-types-of-solar-charge-controllers/
[207] H. Feraoun, M. Fazilat, R. Dermouche, S. Bentouba, M. Tadjine, and N. Zioui, "Quantum maximum power point tracking (QMPPT) for optimal solar energy extraction," Systems and Soft Computing, vol. 6, p. 200118, Dec. 2024, doi: 10.1016/j.sasc.2024.200118.
[208] M. Seyedmahmoudian, B. Horan, R. Rahmani, A. M. T. Oo, and A. Stojcevski, "Efficient Photovoltaic System Maximum Power Point Tracking Using a New Technique," Energies (Basel), vol. 9, no. 3, p. 147, Mar. 2016, doi: 10.3390/en9030147.
[209] "Solar panel MPPT Explained | Caravan Solar Panel Kits & Chargers." Accessed: Jan. 11, 2025. [Online]. Available: https://www.mysolarshop.co.uk/solar-panel-mpptexplained/
[210] "(PDF) Review of maximum power point tracking algorithms of PV system." Accessed: Jan. 11, 2025. [Online]. Available: https://www.researchgate.net/publication/353013414_Review_of_maximum_power_point_tracking_algorithms_of_PV_system
[211] K. S. Hayibo, P. Mayville, and J. M. Pearce, "The greenest solar power? Life cycle assessment of foam-based flexible floatovoltaics," Sustain Energy Fuels, vol. 6, no. 5, pp. 1398–1413, Mar. 2022, doi: 10.1039/D1SE01823J.
[212] "7 Unbelievable Solar Powered Creations | Memolition." Accessed: May 11, 2022. [Online]. Available: https://memolition.com/2016/10/06/7-unbelievable-solarpowered-creations/
[213] J. Mathe, N. Miolane, N. Sebastien, and J. Lequeux, "PVNet: A LRCN Architecture for Spatio-Temporal Photovoltaic PowerForecasting from Numerical Weather Prediction," Feb. 2019, Accessed: Sep. 13, 2024. [Online]. Available: http://arxiv.org/abs/1902.01453
[214] "Solar panel MPPT Explained | Caravan Solar Panel Kits & Chargers." Accessed: Jan. 11, 2025. [Online]. Available: https://www.mysolarshop.co.uk/solar-panel-mpptexplained/

[215] A. N. A. Ali, M. H. Saied, M. Z. Mostafa, and T. M. Abdel- Moneim, 2012 IEEE Energytech. 2012. doi: 10.1109/EnergyTech.2012.6304652.
[216] M. A. Eltawil and Z. Zhao, “MPPT techniques for photovoltaic applications,” Renewable and Sustainable Energy Reviews, vol. 25, pp. 793–813, Sep. 2013, doi: 10.1016/J.RSER.2013.05.022.
[217] P. Manoharan et al., “Improved Perturb and Observation Maximum Power Point Tracking Technique for Solar Photovoltaic Power Generation Systems,” IEEE Syst J, vol. 15, no. 2, pp. 3024–3035, Jun. 2021, doi: 10.1109/JSYST.2020.3003255.
[218] A. S. Oshaba, E. S. Ali, and S. M. Abd Elazim, “PI controller design for MPPT of photovoltaic system supplying SRM via BAT search algorithm,” Neural Comput Appl, vol. 28, no. 4, pp. 651–667, Apr. 2017, doi: 10.1007/S00521-015-2091-9.
[219] L. M. Elobaid, A. K. Abdelsalam, and E. E. Zakzouk, “Artificial neural network-based photovoltaic maximum power point tracking techniques: A survey,” IET Renewable Power Generation, vol. 9, no. 8, pp. 1043–1063, Nov. 2015, doi: 10.1049/IETRPG.2014.0359.
[220] X. Li, Q. Wang, H. Wen, and W. Xiao, “Comprehensive Studies on Operational Principles for Maximum Power Point Tracking in Photovoltaic Systems,” IEEE Access, vol. 7, pp. 121407–121420, 2019, doi: 10.1109/ACCESS.2019.2937100.
[221] P. Gupta and R. Singh, “PV power forecasting based on data-driven models: a review,” International Journal of Sustainable Engineering, vol. 14, no. 6, pp. 1733– 1755, 2021, doi: 10.1080/19397038.2021.1986590.
[222] M. Alizamir, S. Kim, M. Zounemat-Kermani, S. Heddam, A. H. Shahrabadi, and B. Gharabaghi, “Modelling daily soil temperature by hydro-meteorological data at different depths using a novel data-intelligence model: deep echo state network model,” Artif Intell Rev, vol. 54, no. 4, pp. 2863–2890, Apr. 2021, doi: 10.1007/S10462-020-09915-5.
[223] L. Vazquez et al., “Energy System Planning towards Renewable Power System: Energy Matrix Change in Cuba by 2030,” IFAC-PapersOnLine, vol. 51, no. 28, pp. 522–527, Jan. 2018, doi: 10.1016/J.IFACOL.2018.11.756.
[224] H. Abu-Rub, A. Iqbal, and S. M. Ahmed, “Adaptive neuro-fuzzy inference systembased maximum power point tracking of solar PV modules for fast varying solar radiations,” International Journal of Sustainable Energy, vol. 31, no. 6, pp. 383–398, Dec. 2012, doi: 10.1080/1478646X.2011.587517.
[225] “Solar PV – Analysis,” IEA, Accessed: Feb. 13, 2024. [Online]. Available: https://www.iea.org/reports/solar-pv
[226] H. Abouadane, A. Fakkar, D. Sera, A. Lashab, S. Spataru, and T. Kerekes, “MultiplePower-Sample Based P&O MPPT for Fast-Changing Irradiance Conditions for a Simple Implementation,” IEEE J Photovolt, vol. 10, no. 5, pp. 1481–1488, Sep. 2020, doi: 10.1109/JPHOTOV.2020.3009781.
[227] V. Jately, B. Azzopardi, J. Joshi, B. Venkateswaran V, A. Sharma, and S. Arora, “Experimental Analysis of hill-climbing MPPT algorithms under low irradiance levels,” Renewable and Sustainable Energy Reviews, vol. 150, p. 111467, Oct. 2021, doi: 10.1016/J.RSER.2021.111467.