Forecasting and Prediction of Solar Energy Generation using Machine Learning Techniques

Ali Jassim M Lari

A thesis submitted to SWANSEA UNIVERSITY

in fulfilment of the requirements for the Degree of

DOCTOR OF PHILOSOPHY

Department of Electronic and Electrical Engineering

SWANSEA UNIVERSITY

June 2023

DECLARATION

This work has not previously been accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed ....ALI JASSIM M H LARI....................................
Date ....20/06/2023..

STATEMENT 1

This thesis is the result of my own investigations, except where otherwise stated. Where correction services have been used, the extent and nature of the correction are clearly marked in a footnote(s).

Other sources are acknowledged by footnotes giving explicit references. A bibliography is appended.

Signed ....ALI JASSIM M H LARI....................................
Date ....20/06/2023..

STATEMENT 2

I hereby give consent for my thesis, if accepted, to be available for photocopying and for interlibrary loan and for the title and summary to be made available to outside organizations.

Signed ....ALI JASSIM M H LARI....................................
Date ....20/06/2023..

ABSTRACT

The growing demand for renewable energy sources, especially wind and solar power, has increased the requirement for precise forecasts in the energy production process. Using machine learning (ML) techniques offers a revolutionary way to deal with this problem, and this thesis uses machine learning (ML) to estimate solar energy production with the goal of revolutionizing decision-making processes through the analysis of large datasets and the generation of accurate forecasts.

Solar meteorological data is analyzed methodologically using regression, time series analysis, and deep learning algorithms. The study demonstrates how well machine learning-based forecasting works to anticipate future solar energy output. Quantitative evaluations show excellent prediction accuracy and verify the techniques used. For example, the key observations made were that the Multiple Linear Regression methods demonstrates reasonable predictive ability with moderate Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) values yet slightly lower R-squared values compared to other methods.

The study also provides a reflective analysis of result significance, methodology dependability, and result generalizability, as well as a summary of its limits and recommendations for further study. The conclusion provides implications for broader applications across energy sectors and emphasizes the critical role that ML-based forecasting plays in predicting solar energy generation. By utilizing renewable energy sources like solar power, this approach aims to lessen dependency on non-renewable resources and pave the way for a more sustainable future.

DEDICATION

I want to sincerely convey my heartfelt gratitude to the State of Qatar and HH Sheikh Tamim bin Hamad Al Thani for their unwavering support and for granting me this invaluable opportunity. My profound thanks also extend to the Qatar Foundation, my sponsor, for their unwavering support and constant encouragement throughout my PhD journey. Additionally, I wish to recognize the exceptional community at Swansea University, whose pivotal role was instrumental in helping me successfully complete this remarkable journey.

To my parents, I express my profound gratitude for instilling in me a passion for learning and making significant sacrifices to provide me with the best education possible. Your belief in me and endless encouragement have been a guiding light. My heartfelt appreciation is extended to my wonderful mother, Khulood, and father, Jassim, for emphasizing the values of perseverance and hard work and for bringing joy and laughter into my life, even during challenging times.

I also dedicate this thesis to my beloved siblings: Mohammed, Mariam, Fatima, Abrar, and AlMayasa. They have been my driving force, motivating me to push myself harder to make them proud. Their boundless love and unwavering support have served as a constant source of strength throughout this journey.

To my friends, I am deeply grateful for being my trusted sounding board and for offering much needed distractions when the stress becomes overwhelming. Your presence has made this journey more manageable and enjoyable.

Dr. Augustine Egwebe, my exceptional supervisor, deserves distinct recognition for his steadfast guidance and support, especially amidst the challenges posed by the COVID-19 pandemic. I extend my sincere appreciation for your patience, unwavering guidance, and invaluable mentorship. Your consistent challenge to help me grow and produce my best work has been transformative. Lastly, to all those who have made a positive impact on my life, whether through small gestures or significant influence, I deeply value your presence and assistance. This work is dedicated to each of you with heartfelt appreciation and love.

List of Author’s Publications

1. A. J. Lari, A. Egwebe, F. Touati, A. S. Gonzales and A. A. Khandakar, "Reliable Photovoltaics Output Power Prediction in Qatar," 2021 14th International Conference on Developments in eSystems Engineering (DeSE), Sharjah, United Arab Emirates, 2021, pp. 546-551, doi: 10.1109/DeSE54285.2021.9719352.
2. A. J. Lari, A. Egwebe, F. Touati, A. S. Gonzales and A. A. Khandakar, "Photovoltaic System Ensemble Prediction System," 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET), Cape Town, South Africa, 2021, pp. 1-6, doi: 10.1109/ICECET52533.2021.9698546.
3. A. J. Lari and A. Egwebe, "Switched Mode Power Supplies Comparison: PI, Cascade PI and Poiscast Controller," 2021 International Conference on Electrical, Computer and Energy Technologies (ICECET), Cape Town, South Africa, 2021, pp. 1-5, doi: 10.1109/ICECET52533.2021.9698469.
4. A. J. Lari, A. Egwebe, F. Touati, A. S. Gonzales and A. A. Khandakar, "Power Generation Voting Prediction Model of Floating Photovoltaic System," 2021 International Conference on Engineering and Emerging Technologies (ICEET), Istanbul, Turkey, 2021, pp. 1-6, doi: 10.1109/ICEET53442.2021.9659661.

Contents

List of Figures

List of Tables

ACRONYMS

AC	Alternating Current
ANN	Artificial Neural Networks
ARIMA	Autoregressive Integrated Moving Average
CNN	Convolutional Neural Networks
CSP	Concentrated Solar Power
DC	Direct Current
DDM	Double-Diode Model
DL	Deep Learning
FFNN	Feed-Forward Neural Network
FLC	Fuzzy Logic Controller
GA	Genetic Algorithm
GPR	Gaussian Process Regression
KNN	K-nearest Neighbors
kWh	Kilowatt Hour
kWp	'Peak' Power Output of a System in kW
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percentage Error
ML	Machine Learning
MPPT	Maximum Power Point Tracking
MPC	Model Predictive Control
MSE	Mean Square Error

MSF	Mutual Shading Factor
MV	Medium Wave
OEM	Original Equipment Manufacturer
PEC	Power Electronic Converter
PGVPM	Power Generation Voting Prediction Model
PI	Proportional-Integral
PSO	Particle Swarm Optimization
PV	Photovoltaic
RMSE	Root Mean Square Error
SDM	Single-Diode Model
SMPS	Switched Mode Power Supply
SSA	Salp Swarm Algorithm
SVM	Support Vector Machines
SVR	Support Vector Regression
TWh	Terawatt-hour
VAE	Variational Auto-Conductor
LCOE	Levelized Cost of Energy

1 Introduction

1.1 Research Scope

Electricity has become a crucial part of our daily lives, and its demand is constantly increasing due to factors such as population growth, economic development, and advancements in technology - including the rise of electric vehicles [1]. The current sources of electricity are divided into non-renewable and renewable sources, with natural gas, coal and nuclear energy examples of non-renewable sources, and solar, wind and hydropower examples of renewables. Gas and coal, which are fossil fuels, impact the environment and contribute to various problems, including the greenhouse effect and global warming. Due to the increase in greenhouse gases, growing levels of air pollution, and the depletion of fossil fuel resources, the world is now concerned with reducing its reliance on fossil fuels while increasing its reliance on renewable energy sources [2].

Among the renewable sources, solar energy is very promising as it is widely available. Initially, the technologies used to generate solar energy were expensive, and the dependence on it as a significant power source was not economically viable. However, developments in power electronic technologies have reduced costs over the years, making solar energy a viable alternative to fossil energy.

The ability to generate electricity from solar energy depends on the location and local weather conditions, defined as the solar potential. Countries worldwide are now working on harnessing this potential, with China being the top country for solar power generation, followed by the USA and the European Union [3].

Countries in the Middle East, which includes Qatar, have some of the highest solar potentials in the world. Qatar is currently increasing its solar power generation and proceeding towards achieving its solar potential [4]. The Qatar National Vision 2030 aims to diversify the country's economy and reduce its reliance on hydrocarbon resources by increasing the share of renewable energy in its energy mix [5]. In line with this vision, Qatar has set a target of generating 20% of its electricity from renewable sources by 2030. The country is investing heavily in solar and wind power projects to achieve this target. As of 2021, Qatar had a total installed solar capacity of around 300 MW [5].

Raising the proportion of solar energy used to produce a country's power is riddled with challenges; the foremost and most significant obstacle is that solar energy is an unpredictable electricity source contingent upon weather patterns and other climatic factors. Solar energy integration into the grid must thus be investigated, evaluated, and developed, taking into account the environmental conditions that affect solar power output fluctuations, to be prepared for broad usage and largescale integration [6]. Qatar has set a target of generating 20% of its electricity from renewable sources by 2030.

In Qatar, a country that heavily relies on hydrocarbon resources for its energy supply, the deployment of solar power generation has been increasingly encouraged as part of the country's efforts to diversify its energy mix and achieve sustainable development. However, accurate prediction of solar power output is a critical challenge in ensuring a stable and reliable power supply from solar energy. This thesis aims to improve the prediction of solar power generation in Qatar by developing a new forecasting model based on Machine Learning (ML) techniques. The research results presented in this thesis demonstrate the effectiveness of the proposed model in enhancing the accuracy of solar power output prediction, which can potentially contribute to the optimization of energy management and support the integration of solar power into Qatar's energy system.

1.2 Background

The energy demand is rising globally due to rapid economic expansion and technological improvement. To ensure reliability to its customers at all times, electricity supplied by main grids has to be an uninterrupted power supply. Thermal power plants, which produce large amounts of energy by burning fossil fuels like coal, natural gas, and oil, are a common source of electricity. However, during the past several decades, fossil fuel burning has become a significant environmental problem. According to [10], intense cyclones, significant snowfalls, floods, and droughts have all been connected to natural catastrophes due to rising global temperatures. The urgent need to cut greenhouse gas emissions is confirmed by declining ozone layers and worsening air quality. Additionally, the supply of non-renewable energy sources is finite and dwindling. As a result, the globe is moving toward utilizing renewable energy sources, which offer a good substitute for traditional techniques [1].

Solar, wind, biomass, and geothermal energy are renewable energy sources. Solar energy stands out among these due to its accessibility. This transition is aided by the developing of more efficient and cheaper photovoltaic (PV) systems, which use solar energy to generate direct currents (DC) converted to alternating currents (AC) [11]. This renewable energy is a virtually unlimited source for future generations, and solar grid integration is becoming the standard on a global scale.

The sun radiates enormous amounts of electromagnetic (solar) energy absorbed by the earth's surface. The total global solar energy absorption by the earth is 1.8×10^{11} MW [2], an abundant energy source to meet the world's requirements.

Figure 1-1 shows the PV power potential world map [6]. It can be observed that the countries of the Middle East could harness tremendous amounts of solar energy. This region has an annual PV power potential of around 1800 kWh/kWp, which can be observed in the PV power potential map of Qatar shown in Figure 1-2 [6]. Qatar is now taking initiatives to install solar power plants and considers them an alternative to its present energy mix.

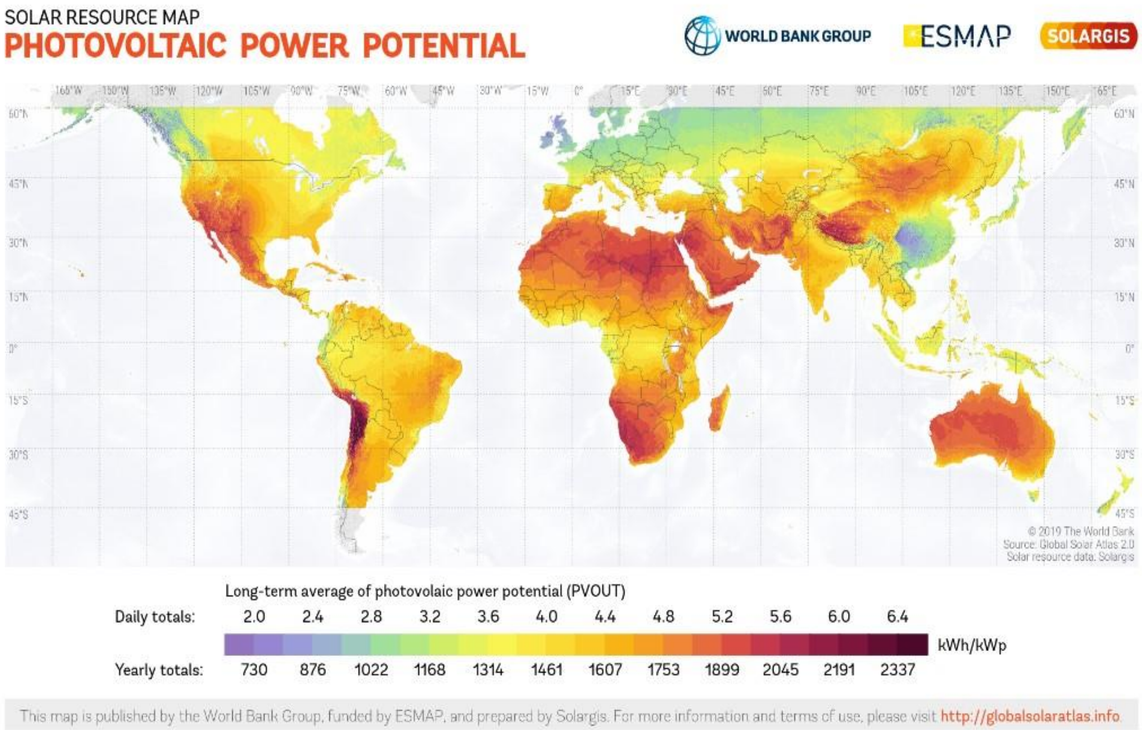


Figure 1-1. Photovoltaic Power Potential world map [6]

The developed and developing world is concentrating on utilizing solar energy due to improvements in PV technology, a decline in the price of solar panels, and the fact that it is environmentally beneficial and widely available [12]. The trends and statistics indicate that there has been an exponential rise in the deployment of solar panels to produce power. The International Energy Agency provides the statistics shown in Figure 1-3 [12], indicating that, on present trends, by 2030, around 3500 TWh of electrical energy is expected to be generated by solar panels.

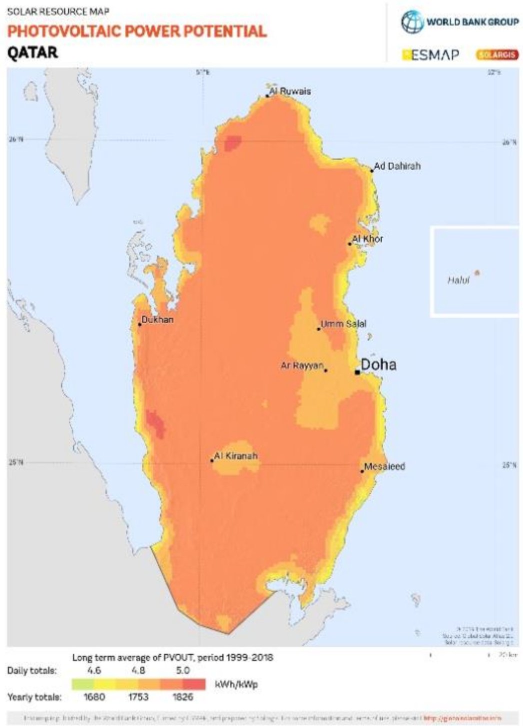


Figure 1-2. Photovoltaic Power Potential [6]

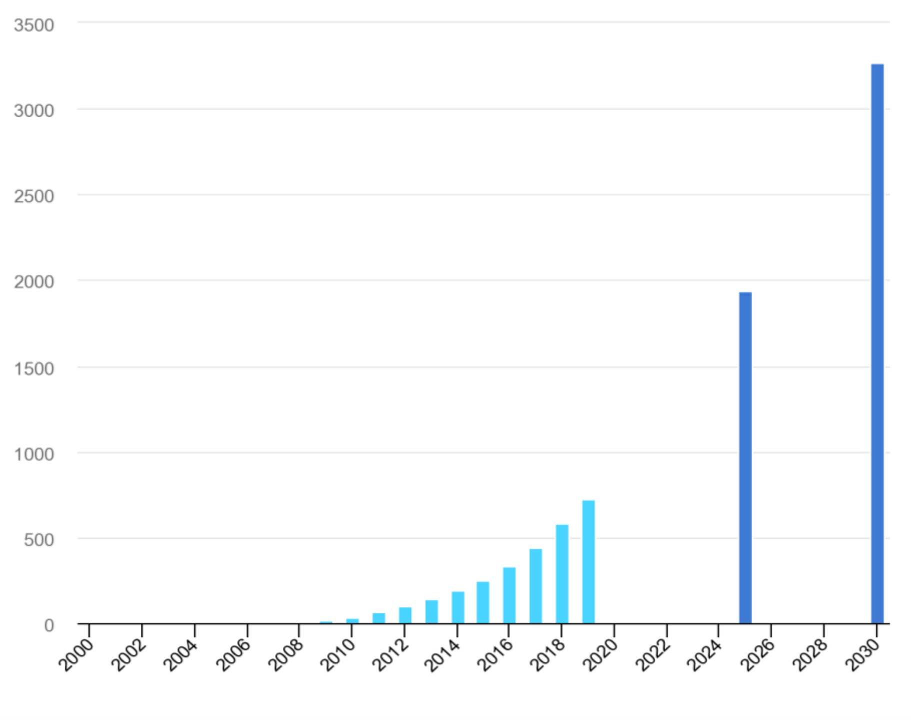


Figure 1-3. Historical and Predicted PV power generation in the Sustainable Development Scenario, 2000-2030 [12]

Although China leads the PV sector, the US is also making significant progress [13]. Substantial governmental assistance for distributed solar PV applications in Brazil and Vietnam [14] is fueling the growth in energy generation. According to an assessment of research agendas on renewable energy in the Global South, the global solar PV market is anticipated to develop at a compound annual growth rate of 20.5% from 2021 to 2026, reaching a total installed capacity of 460 GW by the end of 2026. However, the electricity generated by solar panels fluctuates. The environment significantly impacts solar energy. Despite being plentiful, it is inconsistent and changes depending on several circumstances, including ambient temperature, humidity, wind, local solar radiation, cloud cover, and panel surface temperatures.

Solar energy is, therefore, a variable energy source, and its use might endanger the stability and dependability of the supply system despite the advantages of accessibility and environmental compatibility. It would be challenging for the power system operators to attempt to run the system effectively if it is not adequately integrated [15]. Predicting the electrical energy and power generated by the PV panels is necessary for the electrical grid to operate in an environment with a high rate of PV generation in a smooth, dependable, and stable manner.

1.3 Research Problem

Solar panels are now used in many regions of the globe to utilize the abundant solar energy available. Solar panels have proven to be a promising solution to the global problem of depleting fossil fuel resources and climatic changes due to gaseous emissions from, for example, conventional power plants. However, they bring their challenges.

Solar power output is highly dependent on many environmental factors, including solar irradiance, ambient temperature, cloud cover, and surface temperature of the panels, which cause the solar power output to fluctuate, bringing a degree of uncertainty to the electrical supply, which causes problems for the planning and operation of an electrical supply grid. These challenges and the accompanying uncertainty represent a severe problem for electric grid operators whose goal is to provide a reliable power source, which requires the network to balance demand with supply.

Accurate solar energy forecasting can increase the overall dependability of renewable energy systems, lower the cost of energy storage and grid balancing, and boost the efficiency and profitability of solar energy systems by enabling operators to plan their operations and maintenance schedules better. It is essential to research the accurate prediction of solar panel power output.

This thesis aims to use machine learning (ML) approaches to forecast solar panel output by accounting for external factors that affect supply variability and uncertainty.

This thesis aims to solve the technical issues related to predicting the output of solar panels by:

1. Collecting and organizing the required metrological data and selecting the best environmental parameters for optimal prediction results.
2. Comparing prediction techniques, particularly Deep Learning (DL) and ML techniques (DL employs complex neural networks for learning from large datasets with intricate patterns, demanding substantial resources, while ML uses algorithms for pattern recognition, interpretable, adaptable to smaller datasets, and less resource-intensive).
3. Analysing and visualizing data to assess the accuracy of the prediction process.

1.4 Problem Definition

An essential challenge in renewable energy is solar energy forecasting and prediction, which entails determining the amount of solar irradiance received at a particular place on the earth's surface over a given period. Supply and demand must balance to manage and use solar energy resources effectively. However, this work is difficult since solar irradiance is very changeable and challenging to estimate accurately due to the dynamic nature of the weather patterns and atmospheric factors that impact it. Therefore, developing accurate and reliable solar energy forecasting models is a pressing research problem that requires the integration of multiple disciplines, including meteorology, climatology, atmospheric science, and machine learning [7].

Due to their capacity for handling big data sets and capturing intricate patterns, there has been an increase in interest in creating ML models for solar energy forecasting in recent years. These models can be trained using historical solar irradiance data, weather forecasts, and atmospheric conditions to provide precise solar irradiance estimates at a future time and location. ML models may be applied to solar energy systems to improve their design and operation by predicting the energy output of various solar panel configurations and tracking their performance over time.

1.5 Research Gaps

Forecasts for solar energy are growing more important as it becomes a more widely used renewable energy source. Even though this field has seen significant advances, several research gaps still need to be filled to increase the reliability and accuracy of solar forecasting.

The absence of data gathering and pre-processing standards makes it difficult to compare the findings of various studies. This issue has to be resolved to ensure consistency in the data used to train and test ML models. This thesis focuses on using artificial intelligence and machine learning methods to analyse and pinpoint some of the significant research gaps in solar forecasting. This might entail generating benchmarks and validation measures to compare the performance of various models and best practices for data gathering and pre-processing. The thesis investigates creating cutting-edge ML and artificial intelligence techniques to boost solar forecasting accuracy. This entails examining novel models and algorithms, creating new methods for feature selection, and looking at deep learning in solar forecasting.

1.6 Proposed Research

One of the most considerable obstacles to incorporating solar energy into a supply grid is its intermittency, the variation in energy output owing to weather and other variables. However, the complexity and sophistication of modern supply grids require precise solar energy forecasting. Accurate solar energy forecasting is crucial for efficient grid management and the successful integration of solar energy. Hence, this thesis proposes a strategy for developing an ML-based forecasting model for solar energy. The objective is to enhance the existing research by implementing ML algorithms to create an advanced solar energy forecasting model.

1.7 Methodology Workflow

Various forecasting techniques, including ML algorithms, will be developed to create the proposed solar energy forecasting model. This study will analyse ML approaches' solar energy projections and assess how well they handle the intermittent problem. To accurately forecast solar energy, a range of machine learning methods can be utilised, including artificial neural networks (ANNs), support vector regression (SVR), random forests, and k-nearest neighbours (KNN). Although each algorithm has advantages and disadvantages, they all use historical data to learn from to predict future solar energy production.

Since it can manage non-linear interactions between meteorological factors and solar irradiance, ANN, for instance, is a potent algorithm that can learn complicated relationships between inputs and outputs and is frequently used for solar energy forecasting. The non-linear correlations between meteorological factors and solar irradiance and the capacity to handle high-dimensional data make SVR, on the other hand, a popular choice for solar energy forecasting. Random forests are also commonly used for solar energy forecasting as they can handle non-linear relationships, noisy data and high-dimensional datasets. Furthermore, KNN is another ML algorithm applied to solar energy forecasting, particularly for short-term forecasts. KNN identifies the K closest historical data points to the current input and uses their average value to predict the output. Despite their differences, all these ML algorithms can effectively predict solar energy production or generation with high accuracy and reliability. Nevertheless, the effectiveness of these algorithms can be influenced by various factors, such as the quality and quantity of input data, the chosen model, and the parameters used for tuning.

1.8 Main Contribution

The work will result in more accurate forecasting of solar energy output in Qatar, which is crucial if grid operators are to optimize the operation of the grid and plan and design solar power plants effectively. Thus, the main contribution of this thesis will be to Qatar's electrical grid planners.

This thesis has led to the publication of four conference papers:

- ‘Power Generation Voting Prediction Model of Floating Photovoltaic System’ [8].
- ‘Switched Mode Power Supplies Comparison: PI, Cascade PI and Poiscast Controller’ [9].

  (This paper provides insights on optimizing PV system performance with Switched Mode Power Supplies (SMPS) to enhance the accuracy of machine learning-based solar energy forecasting by integrating SMPS-related behaviours and features as predictive factors, ultimately leading to improved and more refined energy production predictions.)

- 'Photovoltaic System Ensemble Prediction System' [2].
- 'Reliable Photovoltaics Output Power Prediction in Qatar' [57].

The dependent variable of PV output power is anticipated using a set of six independent variables: irradiance, PV surface temperature, dust build-up on PV panels, ambient air temperature, relative humidity, and wind speed. The relevant data to train the prediction models was collected in Qatar between November 2014 and October 2016. An extensive analysis was conducted to determine the optimal configuration for the ANN's hyperparameters, such as the number of hidden layers and training procedure. The results indicated that both approaches produced precise predictions, with a linear regression coefficient of determination R^2 (which quantifies how well a regression model explains the variability in the observed data) of 0.88 and 1.0 for ANN. It's also crucial to remember that when R^2 is computed, the sign of the correlation coefficient is lost. As a result, two models with the same R^2 value can have quite different connections between the independent and dependent variables. To thoroughly assess the quality of the prediction model, it is crucial to consider the unique relationship between the independent and dependent variables, even, for example, if R^2 value of 0.74 is obtained, which suggests that around 74% of the variance in the dependent variable is accounted for by the independent variables in the model, indicating a moderately strong fit.

1.9 Thesis Layout

A background investigation and literature assessment on the solar potential of the world in general and Qatar, in particular, are presented in the second chapter. The types of solar energy production are discussed with an in-deep explanation of the PV method with which this research is concerned.

Further, the techniques for PV output prediction presently used in the literature are discussed and classified.

The third chapter delves into the theoretical basis of the deep learning and ML techniques utilised to predict PV production in this work. It comprehensively covers the different stages of the prediction process, including data collection and preparation, as well as the metrics employed to evaluate the accuracy of these prediction methods.

The findings of this investigation are presented in the fourth chapter. The information used is detailed, and the process for gathering this information is also highlighted.

The fifth chapter presents, compares, and discusses the prediction algorithms' performances.

Finally, the last chapter presents the conclusion of the research and the results that have been reached. In addition, possible future work and its development are presented.

1.10 Conclusions

PV systems are essential for converting solar energy to electrical energy, but their efficiency varies greatly depending on the location, the time of day, and the weather. This thesis focuses on investigating PV output prediction strategies to overcome this problem. Multivariate linear regression models and ANNs are the main techniques being assessed and contrasted, albeit their outcomes vary with regard to the coefficient of determination (R^2). Overall, a viable method to reduce the variation in PV efficiency caused by external factors is to employ multivariate linear regression models and ANNs to estimate PV output power based on a collection of independent variables. The study's results imply that both approaches can result in precise predictions, albeit the best strategy may vary depending on the application and type of data being used.

2 Literature Review

2.1 Introduction

Due to its potential to reduce greenhouse gas emissions and environmental sustainability, solar energy has become an essential source of renewable energy. For effective grid integration, system planning, and optimum use of solar resources, accurate prediction and forecasting of solar power generation are crucial as the solar energy sector continues to expand. Utility companies can maximize energy storage, plan maintenance tasks, balance the grid, and support efficient energy trading with the help of accurate solar power projections. Solar irradiation, weather, geographic location, system parameters, and historical data are just a few examples of the many variables that go into accurate solar power prediction and forecasting. These methods aid in addressing the innate erratic and variable nature of solar energy production, allowing for improved resource management and planning.

This literature review aims to review the strategies and methods used to estimate and forecast solar energy generation. It includes articles and studies that have examined the precision and efficacy of various techniques, such as machine learning algorithms, statistical models, ensemble methods, and hybrid models. This study intends to highlight the developments, difficulties, and probable future paths in solar power modelling and forecasting by analysing the existing literature. It is essential to understand the advantages and disadvantages of different prediction and forecasting methodologies to build strong models that can deliver precise and trustworthy forecasts for solar power. Accurate predictions help increase grid stability, lower energy prices, and better integrate solar energy into the current power grid.

By examining the papers mentioned below, this literature review delves into the specific methodologies employed in solar power prediction and forecasting, providing valuable insights into the advancements made in this field.

2.2 Photovoltaic Electricity Generation

2.2.1 Comparison of solar energy generation methods

Solar energy may be produced in two ways: concentrated solar power (CSP) and PV electricity, see Figure 2-1. In CSP, solar energy is concentrated with the help of mirrors onto a heat engine receiver. The concentrated radiation is converted to heat and drives a heat engine connected to an electricity generator. Hence, CSP plants are similar to conventional power production thermal plants but use solar energy to produce heat instead of burning fuel.

Figure 2-1. The two types of solar power generation

Source: [129]

On the other hand, PV solar systems differ from CSP systems in that they use sunlight to produce a direct electric current through the "Photovoltaic effect" rather than using it first to produce heat and then electricity [11]. This direct current is changed into alternating current using inverters to be distributed locally or to the national grid for further use.

The disadvantage of using a PV system is that electricity is instantly produced and cannot be easily stored due to battery limitations. Presently, CSP is preferable due to better thermal storage technologies and is more attractive for macro-scale electricity generation, which has led to greater market penetration of CSP [11]. Although CSP has better thermal storage technologies, it is expensive and requires two different plants, solar and steam. Hence, its initial cost is high, and PV installations are being promoted due to their lower and diminishing installation costs, assisted by favourable energy market conditions [16], [17]. Table 1 shows a comparison between CSP and PV.

Table 2-1 Comparison between CSP and PV

Comparison Criterion	**CSP**	**PV**
Storability	Better due to the thermal storage technologies	Worse due to battery limitations
Cost	Worse due to the high initial cost of the required two plants (solar and steam) The cost ranges between 110 EUR/MW(LCOE) [123][124][126]	Better due to the lower and decreasing capital costs. The cost ranges between 60 EUR/MW (LCOE) [123][125][126]

2.2.2 *Composition of PV solar arrays*

Solar cells are commonly used to capture and convert the sun's energy into electrical energy [18]. Solar panels or PV modules are collections of solar cells interconnected in series or parallel. PV panels are collections of individual PV modules; a PV array is a collection of PV panels [19]. The composition of an array is shown in Figure 2-2.

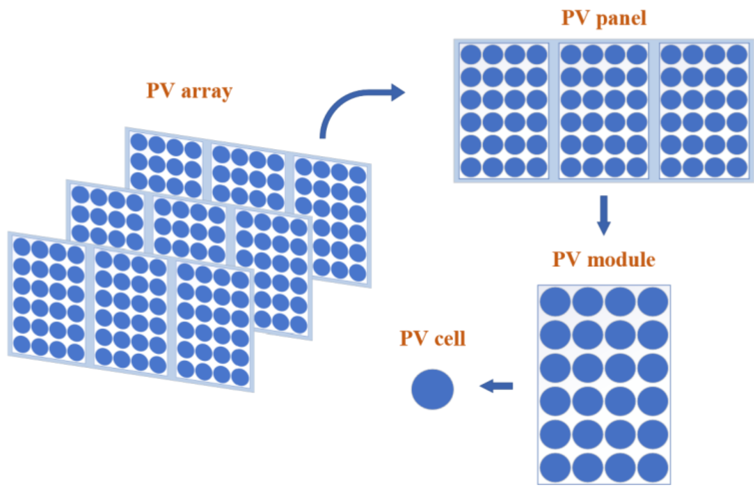


Figure 2-2. Composition of PV arrays

2.2.3 *Electrical model of PV cells*

The PV cell, as previously mentioned, is the smallest part of the PV array. Hence, knowing how these cells function makes it easier to determine what influences their output. A model of a typical PV cell is shown in Figure 2-3, where a current source and a diode are connected in parallel. This may be utilised to mimic a typical PV cell. Also included are resistances in series and shunt. The intrinsic resistance and the shunt resistance are denoted by the symbols R_s and R_{sh}, respectively [20] [21]. Depending on the required voltage and current, PV cells may be connected in series, parallel, or both configurations.

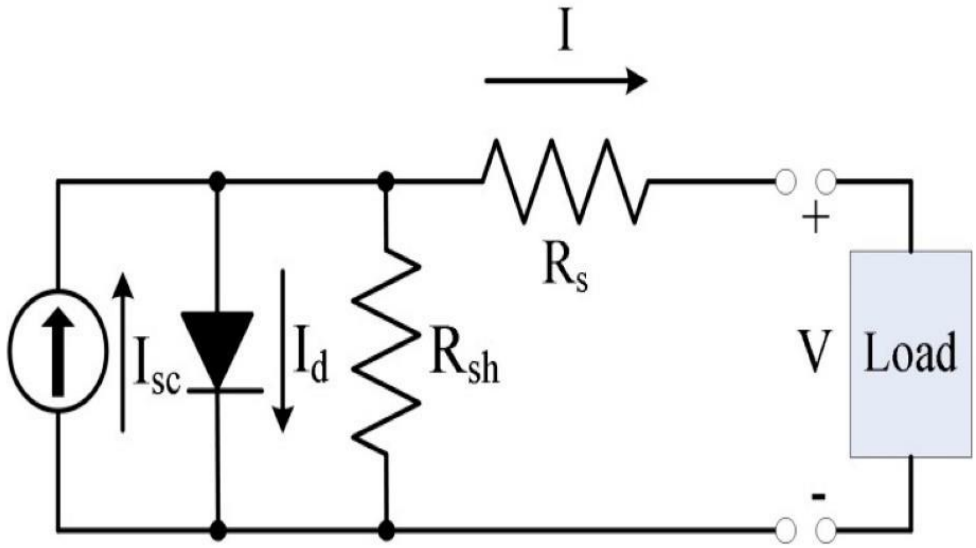


Figure 2-3. PV cell circuit model

The PV cell equivalent circuit has the following generic representation of the PV cell current [12]:

$$I = N_P I_{PH} - N_P I_S \left(exp\left(\frac{V_{cell} + I_{cell} R_S}{V_{th}}\right) - 1 \right) - \left(\frac{V_{cell} + I_{cell} R_S}{R_P}\right) \tag{2.1}$$

$$I_{PH} = G\big(I_{SC} + K_I(T_C + T_{REF})\big) \tag{2.2}$$

$$I_{RS} = \frac{I_{SC}}{exp\left(\frac{qV_{OC}}{KAT_C N_S}\right) - 1} \tag{2.3}$$

$$I_S = I_{RS} \left(\frac{T_C}{T_{REF}}\right)^3 \left(exp\left(\frac{qE_G\left(\frac{1}{T_{REF}} - \frac{1}{T_C}\right)}{KA} \right) \right) \tag{2.4}$$

Where:

Symbol	Description	Value	Unit
q	Electronic charge	1.602×10^{123}	C
K	Boltzmann’s constant	1.38×10^{14}	JK^{-1}
$I_{\#}$	Saturation current		A
$I_{\#)}$	Short circuit current	25°C and 1 kW/m^2	A
$I_{+\#}$	Cell's reverse saturation current		A/m^2
A	Junction ideality coefficient		
V_{th}	(AKTC)/q, where C stands for thermal voltage		V
N_P	The number of PV cells linked in parallel		unitless
N_s	The number of PV cells linked in series		unitless
E_G	Semiconductor cell energy band-gap		eV
K_I	Short-circuit cell temperature coefficient		Celsius (%/°C)
T_C	PV cell's working temperature		°C
G	Solar irradiance		W/m^2
T_{REF}	PV cell reference temperature		°C
V_{OC}	PV cell's open circuit voltage		V
V_{cell}	Voltage of a solar cell		V
I	Current generated from the PV cell		A
I_{PH}	Photo current		A

The PV cell characteristics are typically measured under standard test conditions: G = 1 kWm^{-2}, T = 25 °C, and air mass ratio = 1.5.

Figure 2-4 depicts a block diagram showcasing the electricity generation process using PV panels. The PV panels absorb sunlight, initiating the photovoltaic effect and producing DC electrical energy. This DC energy is converted into AC power by an inverter, enabling integration into the electrical grid for various applications. The diagram illustrates how PV panels serve as a renewable energy source, converting sunlight into usable electricity that can be seamlessly integrated into existing power systems.

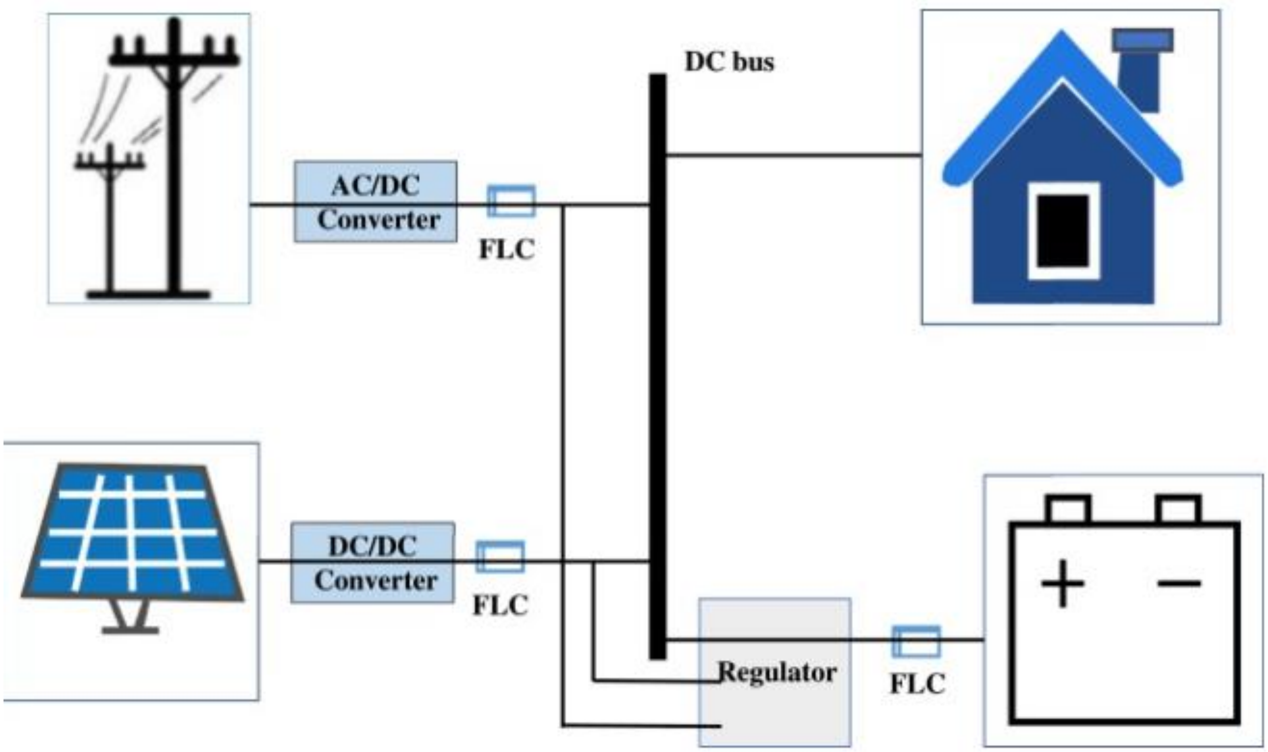


Figure 2-4. Block diagram for electricity generation by the use of PV panels [114]

Figure 2-5 shows a PV panel's block diagram for electricity generation. DC power is produced by the PV panel of cells, which is exposed to sunlight. A power electronic converter (PEC) is then used to convert this DC electricity into AC. The converted AC electricity may now be used to power a load, such as various electrical systems or gadgets. The PV panel collects sunlight to create DC energy, which is then converted to AC and used to power various devices through the load. A controller is often employed to drive the PEC to regulate the current and voltage fed to the load terminal [22].

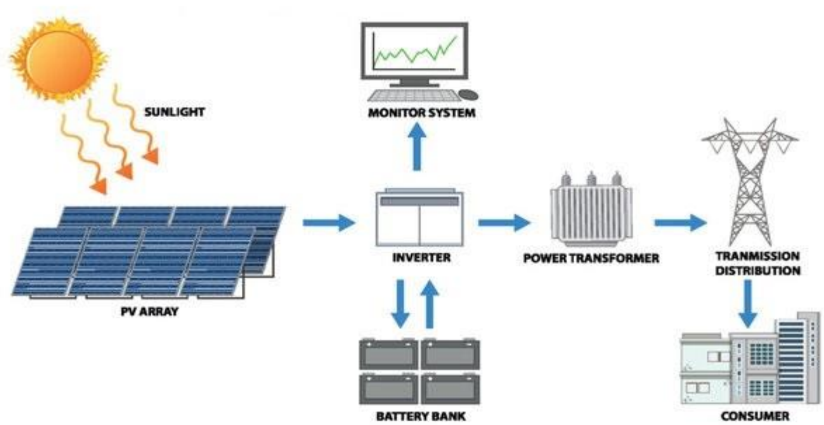


Figure 2-5. Solar electricity generation [22].

As mentioned previously, one of the outcomes of this research work is an article titled "Switched Mode Power Supplies Comparison: PI, Cascade PI and Poiscast Controller" [9] which is going to be used in order to strengthen the study. The discussion and analysis in this article, which explores the effectiveness and performance of various systems, will be guided by some insightful information from this study. Clarifying this comparison will improve readers' comprehension of practical applications by studying the effects of each control strategy on steady-state oscillations and rising time. The findings may be used to improve the effectiveness and efficiency of switched-mode power supplies in various applications.

The study by Lari et al. [9] is a case-in-point of how PV energy generation is conditioned with control actions. The work compares the performance of three control systems for switched mode power supplies (SMPS): proportional-integral (PI), cascade PI, and Poiscast controllers in terms of output voltage regulation, transient responsiveness, and efficiency. The three control techniques and the fundamental SMPS principles are introduced in the study. The SMPS's duty cycle is adjusted using the PI controller, a straightforward and widely used method, which depends on the difference between the output voltage and a reference value. A cascade PI controller was also used, operating as an inner control loop with the special purpose of regulating and monitoring the current flowing through the SMPS inductor. The overall stability and responsiveness of the SMPS system are improved by this configuration's capacity to precisely and finely control the inductor's current. By tackling the complexities of current variations, the cascade PI controller complements the primary control mechanism and ensures that the SMPS performs at its best under a variety of load circumstances. This multi-tiered control methodology demonstrates a thorough method for controlling the subtle performance variations of SMPS systems, eventually resulting in increased effectiveness and dependability. The Poiscast controller is a nonlinear controller that controls the output voltage using a sliding mode strategy. The simulation results are then presented to compare the effectiveness of the three controllers. The simulations varied the load and input voltage based on a typical SMPS circuit. In terms of output voltage regulation and transient responsiveness, the results demonstrated that the Poiscast controller performed better than the other two. The Poiscast controller, in contrast to the other two controllers, has a higher switching frequency, which lowers its efficiency. Figure 2-6 shows the output current response comparison for the three controllers.

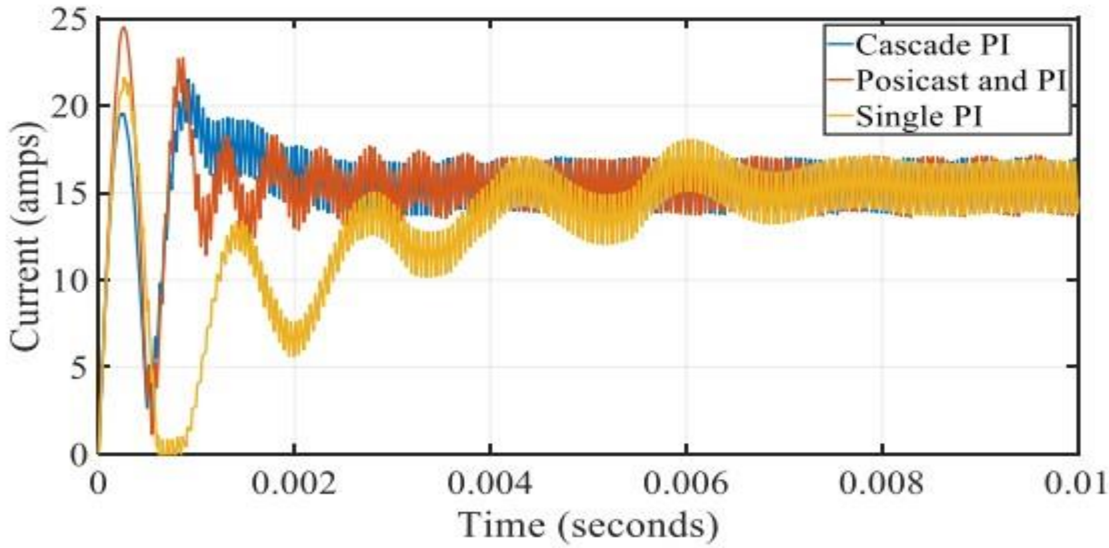


Figure 2-6. DC to DC Boost converter output current response [9]

The research concluded that an application's requirements determine the control approach for an SMPS. A PI controller's implementation is sometimes seen as simple due to its ease of use when compared to more complex control techniques. The PI controller modifies the control output based on both the current error (proportional term) and the cumulative error over time (integral term). Finding the right gains for the controller to get the desired performance and stability can be difficult, which may need some trial and error or more sophisticated tuning approaches. Despite these difficulties, compared to more intricate control schemes, the development of a simple PI controller is not particularly difficult. Multiple stacked control loops are used in a control approach known as cascade control. The process variable is later controlled by a secondary controller, whose setpoint is controlled by the primary controller. Due to its potential for improved performance and disturbance rejection as compared to single-loop control, cascade control is preferred. The statement that a "cascade PI controller performs better in terms of output voltage regulation" is vague and has to be understood in its context. The specific application, system dynamics, and performance requirements determine how well a cascade strategy employing PI controllers will work. The PI controller is an approach that is versatile, easy to use, and reliable. Although more challenging to implement, the cascade PI controller performs better in terms of output voltage regulation. The Poiscast controller provides superior transient response but at the expense of decreased efficiency and increased complexity. In conclusion, this study thoroughly analyses three alternative SMPS control systems and provides insightful information about each one's advantages and disadvantages. The study's findings may be helpful to SMPS circuit designers, who must select the best control approach for a given renewable energy application.

The paper [23] presents an efficient optimization algorithm called the Salp Swarm Algorithm (SSA) for identifying the parameters of a PV cell model. The proposed SSA algorithm is inspired by the behavior of Salp, a marine organism. It is designed to mimic their foraging behavior to optimize the parameters of the PV model. The effectiveness of the proposed algorithm is evaluated by applying it to two types of PV cell models: the single-diode model (SDM) and the double-diode model (DDM). The results show that the SSA algorithm outperforms other state-of-the-art optimization algorithms, such as the Particle Swarm Optimization (PSO) algorithm and the Genetic Algorithm (GA), in terms of convergence speed and solution quality. In addition, the authors also investigated the impact of different SSA parameters, such as the population size and mutation rate, on the algorithm's performance. The results show that these parameters' optimal values depend on the problem and model under consideration [23]. Also, in this study, a comparison was conducted between experimental data and simulated results for each algorithm under specific conditions ($G = 366\ W/m^2$, $T = 18°C$). The investigation focused on the Current, Voltage and Power-Voltage characteristics. The exploratory and exploitative behaviors of the SSA were found to be more flexible and effective compared to other methods. SSA exhibited a balanced approach, starting with broad exploration and transitioning to focus exploitation on later stages.

This behavior enabled SSA to explore high-quality solutions effectively for parameter identification of the DDM PV cell problem. Conversely, the Sine Cosine Algorithm (SCA), Vortex Chimney Search (VCS), Gravitational Search Algorithm (GSA), and Ant Lion Optimizer (ALO) approaches struggled to strike a stable balance between exploration and exploitation, resulting in local optima stagnation and increased error in simulated results. Additionally, the study considered high-temperature and irradiation climatic conditions as a secondary experimental case. Figure 2-7 shows a comparison between the experimental and simulated results of this study.

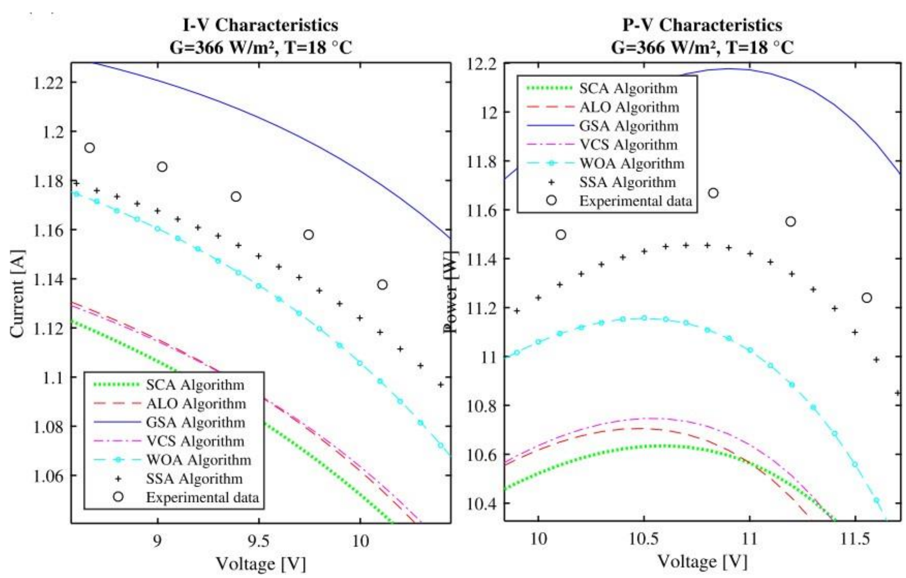


Figure 2-7. Comparison between experimental and simulated results [23]

Overall, the proposed SSA algorithm provides a promising approach for efficiently identifying the parameters of PV cell models and can be used to optimize the performance of PV systems in real-world applications [23].

2.2.4 PV systems: challenges and proposed solutions

- *The conventional unidirectional flow of electrical power*: This challenge arises from the fact that conventionally, the power generation system is located only at the feeder end of the grid [104]. However, with the penetration of PV systems, power will be generated at many points on the grid, and in the case where PV generation exceeds demand for an individual consumer, i.e., excess generation, a bidirectional flow is required to transmit the excess energy back to the grid. Such a mismatch between load power and that generated can impact the distribution feeder and the customers connected to that feeder [105]. While it is true that many thousands of PV systems successfully feed their excess energy back into the grid, challenges such as grid instability still remain. One of the primary challenges is maintaining grid stability when many intermittent energy sources, such as PV systems, feed into the grid. If the supply of and power demand is not balanced, it can result in voltage and frequency fluctuations in the grid, leading to power quality issues and outages [106].

- *The distance between generation sites and the load*: Large-scale solar farms are generally far from their corresponding loads [107]. Future development of such farms would add to the cost of deployment of additional transmission lines [108]. A suggested alternative is to use power inverters in a PV electricity-producing system to address this challenge by deploying distributed PV systems that can be located closer to the loads, reducing the need for additional transmission lines [109]. Distributed PV systems can be installed on rooftops, carports, or in parking lots, among other locations, and can be closer to the loads they serve [109]. This reduces the transmission losses associated with transmitting power over long distances and can also reduce the need for new transmission infrastructure, thereby lowering the cost of deployment [110]. In summary, the suggestion is not just about using power inverters but using them in conjunction with distributed PV systems to address the challenge of the distance between generation sites and loads. However, this can cause power quality problems, including harmonics [111].

- *Voltage problems exist because PV systems cannot be dispatched*. Storage solutions are thus at the forefront of PV research and development [24]. The spinning equipment used in conventional power production systems, such as generators and turbines, produces "*inertia power*," which aids in maintaining the voltage in the power system [24]. PV systems, in contrast, do not have any spinning equipment; hence, they do not produce any inertia power. As such, an imbalance between the power generated by the PV system and the power used by the loads can result in voltage swings [24]. Modern PV research and development concentrates on storage solutions to solve this problem. Energy storage systems can reduce voltage fluctuations and increase the stability of the power system by storing extra energy produced by PV systems during times of low demand and releasing it during times of high demand [24].
- *Variable solar insolation*: One of the significant challenges of integrating solar panels for large-scale electrical energy generation is that the insolation is variable. Insolation is dependent on location and time of the day. Due to this variability, there could be mismatches in electricity demand and generation, and under- or over-generation of power may lead to an unstable system [24]. The ability to provide power on demand is not a luxury afforded by PV generation [24]. There are two proposed solutions to this challenge:
    - Solar farms dispersed across a wide geographical area can mitigate site-specific variables such as cloud cover [11]. Some alternatives involve modifying consumer behavior to produce demand response strategies which use solar energy when it is easily accessible or introduce energy storage for when demand is low or when there is over-production for later use [11].
    - Using better forecasting tools enables more precise forecasts of the potential reduction in solar generation to below the minimum penetration capacity [11].

2.3 PV Output Prediction Methods

The integration of PV arrays into the grid, which enables grid operators to organize and plan energy production and distribution, depends on the precise forecasting of PV production, as previously indicated. Approaches focusing on ML and DL have been found to have increased prediction performance when adequate, relevant data are available. Results from physical and statistical methods have been promising [25].

ML and DL require previously acquired data so the model may be trained and then utilised to forecast future output values. The data can be used to train the models and subsequently classified into univariant and multivariant [25]. Univariant data include only historical values of PV production output from the same source, while multivariate data include weather and other environmental variables in addition to the PV output [25]. Using this data, two types of predictions can be implemented:

1- *Output forecasting*: In order to predict incoming data points, forecasting relies on a complex examination of historical data values. This forecasting method can be divided into multi-step forecasting, which predicts a series of future values, and one-step forecasting, which projects the immediately following value. Time series analysis, a method used to recognize patterns and trends in chronological data, is the basis of this forecasting strategy. This kind of study can use either univariate data, where only one variable's history is taken into consideration, or multivariate data, which considers a number of interrelated variables. The timing of observations and their corresponding timesteps are crucial to the quality of forecasts because correlations and patterns seen across timeframes significantly impact forecasting results. In order to predict future data points and aid in strategic planning and informed decision-making, forecasting essentially entails drawing insightful conclusions from historical data [25].

2- *Output regression* refers to a method of predictive modelling whose objective is to forecast a continuous numerical value from input features. The anticipated value that the regression model produces is referred to as the "output" in this context. Regression analysis establishes a link between the input features and the desired output variable by fitting a mathematical function that best represents the patterns and trends existing in the data. Recognizing that time series data can be used as input for regression models is crucial since doing regression analysis on time series data does not automatically cover the entirety of time series research. Even though the data is a time series, regression is not considered a time series study. Multivariate data are necessary for this prediction, where a subset of the variables in each observation are seen as inputs and the remaining variables as outputs [25]. Regression models use knowledge about the connection between the input and the output to predict the outcome for a given input [25]. A problem with regression is that some input variables cannot be predicted to have future values. The study uses a regression model to predict how much electricity a building will use over time. Temperature, weekday, and time of day are examples of possible input variables for the regression model. However, "weather conditions," which include specifics like humidity and irradiance, are one of the input factors. In this circumstance, it is not possible to accurately predict weather conditions for a very long time in the future. The study aims to forecast electricity usage for the following month. The particular weather for that month cannot be accurately forecast in advance despite projections or extrapolations for temperature, day of the week, and time of day. As a result, "weather conditions" would serve as an illustration of an input variable whose values cannot be foreseen in the future. This circumstance brings to light a problem with regression modelling, particularly when used for time series forecasting. While some input variables can be expected to have future values, other variables may be impacted by unpredicted outside forces or events. It highlights the importance of giving careful thought to the variables to include in the model, as well as the potential drawbacks of generating long-term forecasts based on specific inputs. Using anticipated feedback will solve this issue. Most of the time, expected input is available and may be utilised to make regression predictions, such as with weather data [25].

2.4 Conclusions

This literature review has provided the strategies and methods used to estimate and forecast solar energy generation. It has provided articles and studies that have examined the precision and efficacy of various techniques, such as machine learning algorithms, statistical models, ensemble methods, and hybrid models. It intended to highlight the developments, difficulties, and probable future paths in solar power modelling and forecasting by analysing the existing literature. It is essential to understand the advantages and disadvantages of different prediction and forecasting methodologies to build strong models that can deliver precise and trustworthy forecasts for solar power. Accurate predictions help increase grid stability, lower energy prices, and better integrate solar energy into the current power grid.

3 Machine Learning Methods for Prediction

3.1 Introduction

The prediction methodology employed within this thesis revolves around a regression-based approach that hinges on weather and environmental data. This technique involves leveraging the relationships between these factors and the predicted outcomes. The specifics of this prediction method, including the intricate interplay between the weather and environmental variables, will be elaborated upon in subsequent sections of this thesis. By integrating these data-driven insights, the aim is to unravel the dependencies and patterns that underlie the anticipated outcomes, providing a comprehensive understanding of the predictive framework employed in this research.

Table 3-1 Machine Learning Methods for Prediction

Parameter	**Value**	**Justification**
learning rate	0.01	A small learning rate of 0.01 is chosen to ensure stable and smooth convergence of the optimization algorithm.
max_iter	1000	A maximum of 1000 iterations is set to prevent the algorithm from running for an excessively long time.
tol	1e-4	A tolerance of 1e-4 is used to determine when the change in the loss function is sufficiently small to stop the optimization.
random state	42	A fixed random state of 42 is used to ensure the reproducibility of the results.

3.2 Regression Method for Prediction

In the literature, ML methods, including linear regression [26], support vector machines (SVM) [27], and decision trees [28] have been used to forecast PV output. Deep learning approaches use. ANNs, which have a range of designs. These include standard feed-forward neural networks (FFNN) [29], recurrent neural networks (RNNs) [30], and Long Short-Term Memory (LSTM) networks [31], which are a variety of RNNs made to manage and learn from data sequences. They are particularly effective in sequential data applications, including speech recognition, time series forecasting, and natural language processing. The main innovation of LSTM networks is their capacity to detect temporal and long-range dependencies in data. Vanishing or inflating gradient issues can make it difficult for traditional RNNs to retain data from far-off prior time steps. These restrictions are overcome by LSTMs thanks to their intricate network of memory cells and gates. [31]. The effectiveness of several models for predicting PV outputs has been compared in numerous published papers, as in [32], [33]. According to all investigations, using ANNs produced the most accurate results [31], [32], [33].

Because of worries about climate change and the depletion of non-renewable resources, the utilization of renewable energy sources has grown in importance. In recent years, solar photovoltaic systems have become popular as a viable, sustainable energy source. Floating PV systems, in particular, are a relatively new and novel technology that provides various benefits, including less land consumption and better power generation efficiency owing to the cooling impact of the water [33]. However, solar irradiation, temperature, humidity, wind speed, and water quality all impact the performance of floating PV systems [32]. As a result, precise power generation forecasting is critical for improving system performance and administration [32]. To help resolve this issue, the author of this thesis presented a paper on the Power Generation Voting Prediction Model (PGVPM) [8] for a floating PV system that integrates various models through voting. This study concentrated on a predictive model to foresee or predict voting trends connected to energy production operations. The phrase "Power Generation Voting Prediction Model" implies that the model attempts to predict how people or entities may vote or make decisions regarding issues relating to power production. Essentially, the goal of this approach is to use predictive analytics to predict how different stakeholders, such as customers, lawmakers, and regulators, could vote on matters involving power generation. These choices could relate to a variety of things, like modifications to policies, the adoption of renewable energy sources, or expenditures on infrastructure. The author's goal in developing this predictive model was to give decision-makers insights into the outcomes of voting scenarios connected to power generation so they could make well-informed decisions. The PGVPM considers multiple parameters that influence power generation and gives reliable power generation projections. The authors gathered data from a floating PV system and utilised it to test the suggested concept [8]. Solar irradiation, temperature, humidity, wind speed, water quality, and electricity generation were all measured. The PGVPM's performance was compared to individual models and other ensemble approaches, such as stacking and bagging. The PGVPM outperformed individual models and other ensemble approaches in forecasting power generation, showing its use in real-world applications. The voting approach, which combines the strengths of numerous models and decreases the impact of individual model faults, was credited for the PGVPM's higher performance. The paper proposed a unique voting mechanism for estimating power generation in a floating PV system. The suggested PGVPM can estimate power generation accurately and help optimize and control floating PV systems [8].

PV solar power prediction involves estimating the amount of electricity that a solar panel or PV system will generate based on various factors such as solar irradiance, temperature, and system characteristics. To maximize the energy output of PV systems, efficient control strategies are necessary to ensure that the system operates at its peak efficiency and follows the optimal power generation trajectory. The research [34] recognizes the impact of environmental conditions on the operation of PV systems, including aspects like the nonlinearity of PV modules and the consequences of changing weather conditions. These elements significantly contribute to the precision of PV solar power prediction, as fluctuations in the environment directly influence the real power generation capacity of the system.

Because of their simplicity and cost-effectiveness, single-stage PV systems that integrate the tasks of maximum power point tracking (MPPT) and voltage control are popular. However, these systems are plagued by several control difficulties that impair their performance and stability. This study focuses on the control of a single-stage PV system, especially MPPT, current control, and voltage control. Mastromauro et al. [34] thoroughly examine the available control systems and procedures for dealing with these difficulties. The authors address MPPT approaches such as Perturb and Observe (P&O) and Incremental Conductance (INC), as well as more sophisticated methods such as ANNs, Fuzzy Logic Control (FLC), and Model Predictive Control (MPC). They contrast the performance and benefits of each strategy, as well as their limitations and obstacles.

They discuss the traditional Proportional-Integral (PI) controller and its variations, such as the Fuzzy PI and Sliding Mode Control in terms of current control. They also discuss the difficulties of current regulation in single-stage PV systems, such as PV module nonlinearity and the influence of climatic factors [34].

3.2.1 Machine Learning Techniques for Prediction of Solar Energy

The increasing use of solar energy as a renewable energy source necessitates precise projections of solar power output. ML algorithms have shown considerable potential in making accurate forecasts about solar power generation. The forecasting of solar power generation using ML approaches is the main topic of this overview of the pertinent literature. Aslam et al. [35] have investigated ML methods for predicting solar power generation, including support vector machines (SVMs), decision trees, and Artificial Neural Networks (ANNs). They also researched a variety of ambient factors that affect the generation of solar power, including temperature, humidity, wind speed, and cloud cover. Principal component analysis, correlation-based feature selection, and recursive feature removal are only a few of the feature selection approaches that the authors researched. They showed that the SVM model surpasses other ML techniques' accuracy and efficiency. The study also provided knowledge for researchers and practitioners in the field of renewable energy to develop accurate and efficient forecasting models for solar power generation. Reference [35] used a hybrid deep learning model that used long short-term memory (LSTM) and convolutional neural networks (CNNs) to predict the output of solar electricity. The LSTM neural network was also employed for multivariate time-series forecasting of solar power generation. Other work also highlights the importance of feature selection in ML-based forecasting of solar power generation [36].

In [37], deep learning techniques are examined to predict solar power generation appropriately; the authors highlight the importance of solar power forecasting to enable successful integration into a supply grid and offer both a thorough analysis of existing solar power forecasting approaches and an exploration of the advantages of using deep learning methods instead of conventional statistical models. The authors also demonstrate how deep learning systems can accurately forecast solar power generation using an extensive historical solar power generating data dataset. Next, the study addressed how well three distinct deep learning techniques – LSTM, CNN, and Autoencoder – perform relative to computing efficiency and prediction accuracy. The LSTM model fared the best in accuracy despite the CNN model being the most successful. The Autoencoder technique, as demonstrated by the authors, can improve the accuracy of solar power projections. According to the study's main conclusion, deep learning techniques can significantly improve the accuracy of solar power projections. However, it also highlighted the need for more research to enable more efficient integration of solar-generated electricity into supply networks [37].

In their study on applying deep learning to forecast solar irradiation for solar power systems, Muhammad et al. looked at the usage of these algorithms to anticipate solar irradiation [38]. For the successful integration of solar energy into the grid, the authors stress the necessity for precise estimations of solar irradiation. The article thoroughly analyses several relevant deep learning algorithms and their applicability to power system forecasting. The LSTM, CNN, and Autoencoder deep learning algorithms were used in a case study to predict solar irradiation, and their benefits and drawbacks were examined. Significant historical solar irradiation data was used to train and test the LSTM model as shown in Figure 3-1. In this figure, the LSTM model shows an improvement over traditional statistical techniques in terms of prediction accuracy, according to the authors' comparison of the LSTM model's performance with more well-known statistical models, such as the Autoregressive Integrated Moving Average (ARIMA). Using LSTM, they projected the hourly, daily, and total solar irradiation for the following year.

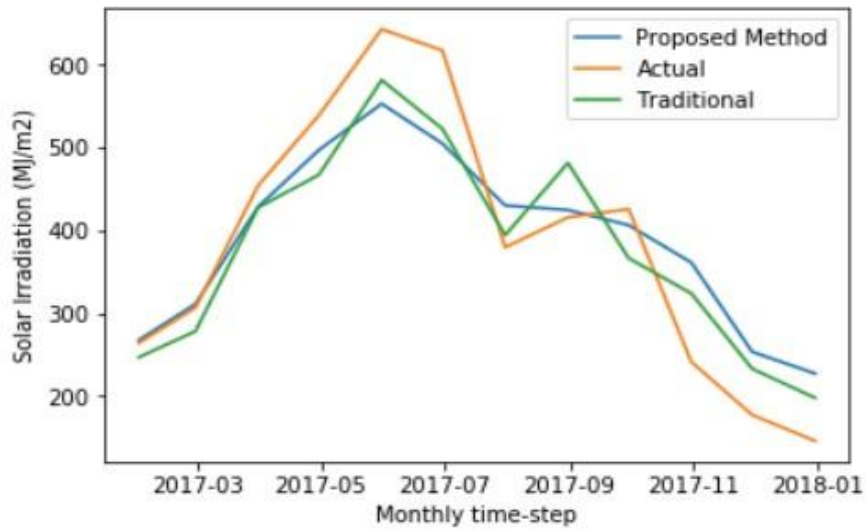


Figure 3-1. Comparison of actual and predicted data [38]

In their discussion, Muhammad et al. also covered the challenges of applying deep learning techniques for power system forecasting [38]. These challenges include the necessity for a sizable quantity of data, the processing requirements, and the interpretability of the results. Muhammad et al.'s research showed that the LSTM model, in particular, can increase the accuracy of forecasts for solar irradiation. The authors concluded that further study is required to address the difficulties of adopting deep learning applications to power system forecasting and investigate the potential of these approaches when applied to other facets of power system operations [38].

3.3 PV Domain and Forecasting Domain Techno-scientific Information

The solar energy technology that deals with generating electricity from solar panels is known as the PV domain, and the "forecasting domain" designates the area of predictive analytics that makes predictions about the future based on previous data. Forecasting is essential for the efficient and consistent operation of the grid in the context of power systems and for the successful integration of renewable energy sources like solar power. Research on the use of forecasting methodologies in the PV business, which is at the nexus of these two industries, has recently attracted much attention. The objective is to develop accurate and trustworthy methods for forecasting solar energy generation, which can help grid operators manage the grid more effectively and boost overall system resilience. Deep learning algorithms have been developed into powerful tools for solar power forecasting [35].

Research has demonstrated that when compared to conventional statistical models, deep learning algorithms such as LSTM, CNN, and Autoencoder may significantly increase the accuracy of solar power forecasts [39]. Moreover, studies have shown that merging forecasting models can improve prediction accuracy [40]. Forecasting is crucial for both the successful integration of renewable energy sources like solar power and the efficient and consistent grid operation in the context of power systems. The goal is to provide reliable and precise techniques for predicting solar energy generation, which can aid grid operators in better managing the grid and enhancing overall system resilience. Deep learning algorithms have become a potent tool for solar power forecasting in the PV industry [40].

The Photovoltaic System Ensemble Prediction System (PSEPS) proposed by the author of this thesis in the paper [41] suggests an ensemble method for predicting the output power of PV systems.

To increase the precision and dependability of the forecasts, the ensemble approach mixes various independent prediction models. Additionally, the paper offers a real-time data processing system to gather and examine the information from the PV system, weather forecasts, and other pertinent sources to produce precise and timely predictions. An ML model, a numerical weather prediction model, and a statistical model are among the individual prediction models employed in the ensemble approach and are described in the study. Each model is created to capture a distinct aspect of the behaviour of the PV system and is trained using historical data. The ensemble strategy, which combines the predictions from each model using a weighted average, is also covered in detail in the study. Figure 3-2 shows ANN's mean square error (MSE) results, which revealed the greatest outcome based on the sensitivity analysis process.

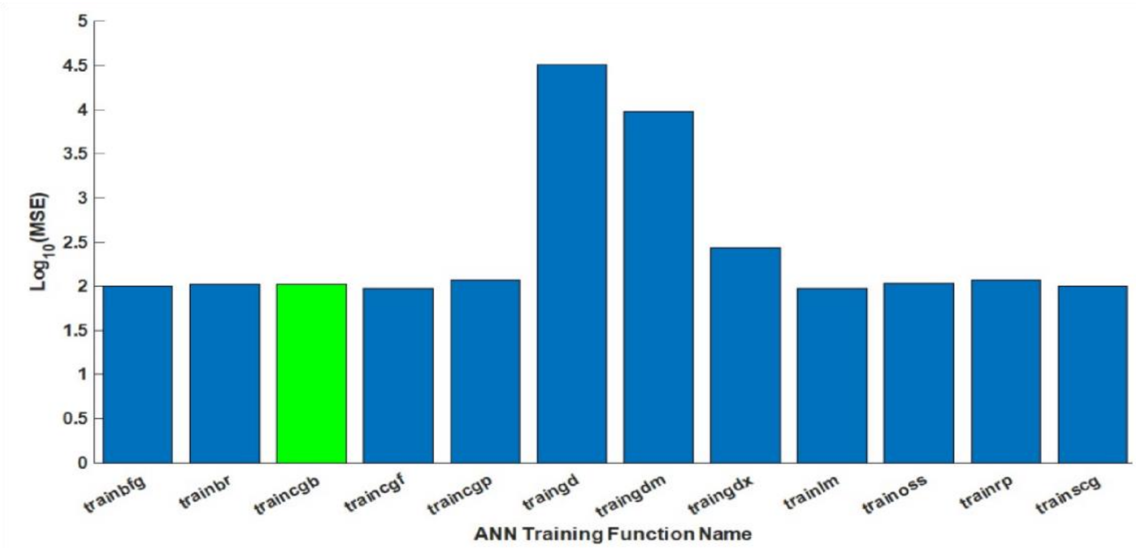


Figure 3-2. Log of MSE of Training Algorithms Results [41]

The paper offers a promising strategy for enhancing the precision and dependability of PV system output power estimates. The real-time data processing system can provide prompt updates to the forecasts based on the most recent data, and the ensemble technique has been proven effective in merging diverse models to provide more accurate predictions. The proposed method may be used to improve the performance of PV systems and integrate renewable energy into the electrical grid.

3.4 Machine Learning Techniques

Nespoli et al. [42], Yang, et al., [43], Antonanzas-Torres et al., [44], and Nespoli, et al., [45] all emphasize the value of precise solar irradiance forecasts for the renewable energy industry and describe the drawbacks of conventional forecasting techniques. They discuss cloud type categorization and how it could help projections of solar irradiation. They also discuss data and techniques, which included using satellite images and ML algorithms such as random forest and SVMs. These authors and others [46] stress the significance of accurate predictions for the renewable energy sector and demonstrate the potential of ML approaches for improving solar irradiance forecasting and solar PV power output forecasting.

The authors proposed solar PV power forecasting as necessary for ensuring the efficient and reliable operation of solar PV systems. They then discussed various ML methods to forecast solar PV power, including neural networks, SVR, decision trees, random forests, and KNN. The authors also highlighted some significant challenges in predicting solar PV power, including the nonlinearity and irregular nature of solar PV power production and the accessibility and quality of input data. They also discussed how ML might help resolve these issues and improve solar PV power's forecasting accuracy [47].

Using data from a Korean PV power plant, the authors of [48] evaluate the effectiveness of several models, including ANNs, SVR, and random forest regression. An LSTM-based deep learning forecasting model is used to anticipate renewable energy production in Korea. After being trained on historical data, the model outperformed traditional forecasting models regarding the accuracy and predicted horizon. The results suggest that decision-makers, academics and policymakers may find the LSTM-based model valuable in developing sensible policies and strategies for renewable energy due to its resilience, capacity to handle enormous datasets, and ability to manage complicated nonlinear interactions. The study highlights the importance of accurate forecasting for sustainable energy policy, which can be replicated and adapted to other regions or countries with similar energy systems and climate conditions [48].

The study [49] evaluates three ML algorithms to anticipate renewable energy production: ANNs, SVR, and Gaussian process regression (GPR). The study's primary goal is to integrate wind and solar power, which are variable and dependent on the weather. The models were tested using statistical measures such as root mean square error (RMSE) and coefficient of determination (R^2), which were trained on historical data from a Spanish wind and solar farm. The study emphasizes the need to select the appropriate model based on the data and application and the potential of ML approaches for forecasting renewable energy. The findings can aid in renewable energy planning and policy choices, enabling more efficient and effective integration of renewable energy into the power system.

Predicting how much energy will be produced at any moment is one of the main problems with solar power generation because factors like cloud cover, rainfall, and sun angle significantly impact solar power output. The research discussed in [50] offers a novel method for estimating solar power using ANNs, where an ANN is a type of ML that imitates how the human brain processes information. ANNs can predict the quantity of energy that will be produced by solar panels with reasonable accuracy by examining past meteorological data and solar radiation levels. Preparing for energy demands and being less dependent on conventional energy sources will be simpler by solid predicting solar power output. The study offers a potentially viable answer to the problems associated with predicting solar power.

The study by Dairi et al. [51] presents a state-of-the-art method for short-term forecasting of photovoltaic solar power generation utilizing a deep learning strategy powered by a variational auto-encoder (VAE). This ground-breaking approach uses deep learning to assess previous solar radiation and temperature data to anticipate the quantity of energy PV panels will produce in the near future. The study uses VAE-driven deep learning methodology to locate trends and associations in the historical data and make projections that account for the particulars of each solar panel installation. The report emphasizes the importance of the variables taken into account throughout the forecasting process and gives a complete discussion of the technique utilised to create the model. The experimental findings provided in the research show how well the VAE-driven deep learning technique works for projecting PV solar power output over the short term. The model could predict energy output for timespans ranging from 15 minutes to 2 hours. As a result, it is a helpful tool for utilities and the energy industry, which must be able to forecast and plan for the production and distribution of energy in real-time [51].

An automated online program [52], known as PVPF uses ML algorithms to forecast solar power generation based on current weather conditions and past trends. Anyone may enter their location into the user-friendly and accessible PVPF program to obtain good projections of future solar power generation for hours or days. The system analyses various meteorological data inputs and generates forecasts customized to the particular PV system utilised using different cutting-edge ML methods, including ANNs and decision trees. The study in [52] thoroughly explains the PVPF tool, its features, and the development process for the system. The authors also provide experimental findings that show how beneficial the PVPF tool is for predicting solar power output in real-time. The results show good accuracy, with prediction errors for forecasts up to 24 hours ranging from around 2% to 5%. The PVPF tool's ability to provide precise and dependable real-time forecasting capabilities has the potential to transform the sector of solar energy generation. Foreseeing and preparing for energy production and distribution in real-time would be very helpful for utilities and energy firms. The PVPF tool is a promising advancement in renewable energy with the potential to dramatically increase efficiency thanks to its user-friendly interface, robust ML algorithms, and reliability of solar energy prediction.

Wang et al. [53] offer a thorough analysis of the use of nanofluids in heat pipes, emphasizing the application of ML methods to forecast and enhance their performance. The authors thoroughly outline the theory behind nanofluids and heat pipes and the many ML techniques that have been applied in this field of study. The study summarizes several research studies that examined the use of nanofluids in heat pipes and emphasizes the major discoveries and difficulties that have been found. The authors also discuss the possible benefits of nanofluids in various fields, such as renewable energy, electronics, and aerospace [53]. Because nanofluids can increase the cooling efficiency of PV panels, their use in heat pipes is crucial to predicting the output of PV systems. PV panels' efficiency declines at high operating temperatures, resulting in a loss in power production. The performance and longevity of PV panels can be enhanced by using heat pipes with nanofluids to assist in dispersing the extra heat produced by the panels. Wang et al. study's ML approaches may also be used to anticipate PV system performance based on various variables, including temperature, solar irradiance, and wind speed. PV system performance may be improved, and energy output can be raised by combining nanofluids in heat pipes and ML predictions.

The promise of ML and DL techniques to increase the precision and efficacy of wind power and solar PV power prediction are highlighted in three publications which present novel approaches to forecasting renewable energy production [54], [55], [56].

The first study in [54] suggests utilizing ML algorithms to assess vast wind speed, direction data, and other pertinent elements to estimate wind power generation. The second study [55] focuses on short-term solar PV power prediction using a mix of DL and Wavelet transform techniques to assess meteorological data and provide precise forecasts. DL approaches can potentially increase the precision and dependability of predictions for solar power. The Wavelet transform is a mathematical technique used in signal processing, which analyses and breaks down signals into their frequency components. The Wavelet transform separates the frequency components of the data for solar power. This enables the DL model to more thoroughly comprehend the underlying patterns and trends in the data, potentially enhancing the precision of short-term predictions for solar PV output. The work offers a potential method for accurately predicting solar power production by integrating DL and Wavelet transform techniques. This method could be important for utilities and energy companies to manage the integration of renewable energy sources into the power grid, as the model was found to perform better than conventional time series forecasting techniques. The third study [56] shows how ANNs are used to forecast solar PV power generation. The model produced precise forecasts based on various inputs, including weather information, time of day, and historical trends. The research emphasizes that in developing nations, where accurate forecasting is essential for maintaining the dependability and efficiency of energy systems, there is the potential for ML approaches to help the spread of renewable energy [56]. Together, these three studies show how ML and DL approaches may be used to enhance renewable energy forecasting, opening the way for future renewable energy systems that are more effective and dependable. These models may assist energy providers and utilities manage their operations, cut costs, and support the global spread of renewable energy by forecasting wind power and solar PV electricity output.

The study by Lari et al. [57] addresses the issues of dependably predicting PV output power in Qatar, a nation with a high solar energy potential but simultaneously confronting environmental challenges such as sandstorms and dust deposition on PV panels. To address these issues, a hybrid prediction model was presented that combined the benefits of two distinct approaches: ANNs and SVRs. The model considers various parameters that influence PV output power, such as solar irradiation, ambient temperature, wind speed, and sand and dust deposition. To evaluate the proposed model, the authors collected data from a PV system in Qatar and compared the hybrid model's performance to that of the individual ANN and SVR models. Figure 3-3 shows the output power for each record.

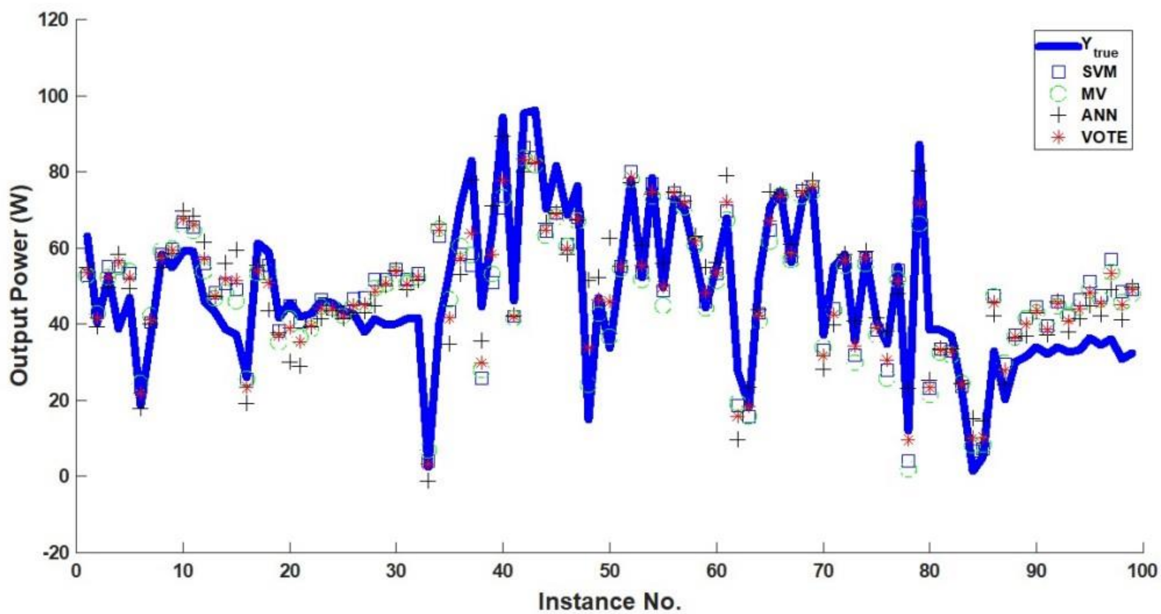


Figure 3-3. Output Power for Each Record [57]

The hybrid model outperformed the separate models in forecasting PV output power, suggesting it could be beneficially used in real applications. Overall, this research proposes a new technique for predicting dependable PV output power in Qatar, including the issues of sandstorms and dust deposition on PV panels. The suggested hybrid model can provide accurate forecasts and help to optimize and manage PV systems in the country. With its novel technique, this research opens the door for dependable and efficient solar energy consumption in Qatar.

The Photovoltaic System Ensemble Prediction System (PSEPS) proposed by Lari et al. [41] is a revolutionary technique for PV output power prediction. The proposed PSEPS takes a novel approach to PV output power prediction by combining the capabilities of two distinct approaches, ANNs and ensemble learning, to improve accuracy and dependability. The ANN is trained to estimate PV output power based on solar irradiation, temperature, and humidity. On the other hand, the ensemble learning approach integrates numerous models' predictions to enhance accuracy and minimize mistakes. To assess the PSEPS's performance, data was collected from a PV system in China, and its forecasts were compared to those of other individual models and ensemble approaches. The PSEPS surpassed other approaches in accuracy and resilience, demonstrating its potential for use in actual applications.

The Support Vector Machines (SVM), Artificial Neural Network (ANN), and Multiple Variable

(MV) prediction approaches are specifically developed to be integrated into the Photovoltaic Solar Energy Power (PSEP) PV Generation Power Forecasting System, as shown in Figure 3-4. The system's primary goal is to forecast PV generation power by utilizing cutting-edge machine learning methods. The framework incorporates two key methodologies: SVM and ANN to improve the precision and dependability of predictions. By taking into account a wide range of factors that affect PV power generation, the MV prediction approach significantly enhances forecasting capabilities. The system's interconnected parts are on display in the architecture. The first step is data collecting, during which pertinent elements, including solar irradiance, temperature, and historical data, are gathered. The SVM and ANN prediction models are then given these variables, which go through training using historical data to discover patterns and relationships.

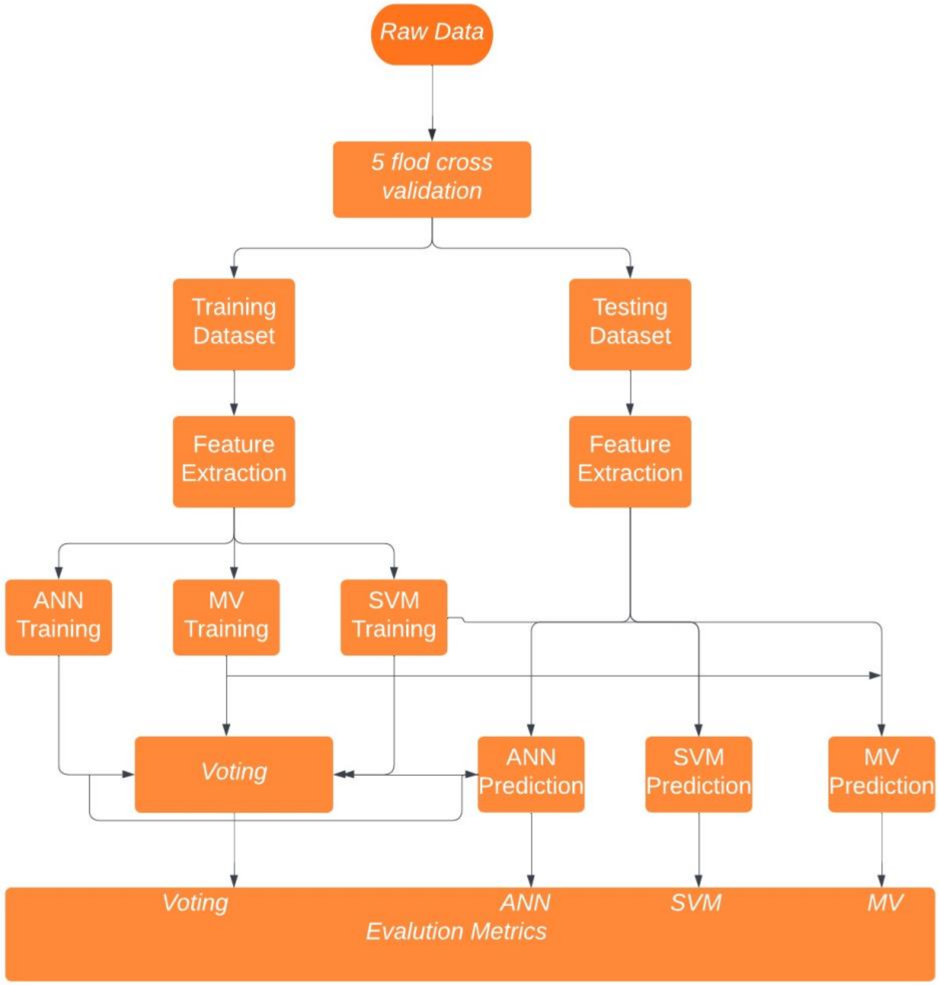


Figure 3-4. PSEP PV Generation Power Forecasting System

PSEPS has a huge potential since it can help optimize and control PV systems, as well as the general progress of renewable energy use. Using a novel methodology, the study contributes to the ongoing efforts toward a sustainable and green future [41]. The study also comprehensively reviews the state-of-the-art techniques and optimization methods for PV solar power forecasting. The authors first introduce the importance of PV solar power forecasting in various applications, such as energy management and grid integration, and then present a detailed overview of the existing literature in this field. The review covers a wide range of forecasting techniques, including statistical methods, ML algorithms, and hybrid models. The authors highlight the strengths and weaknesses of each technique and discuss their suitability for different forecasting scenarios. Furthermore, the paper provides a critical evaluation of the optimization methods used to improve the accuracy of PV solar power forecasting. The authors discuss various optimization techniques, such as data pre-processing, feature selection, and model selection, and provide insights into their effectiveness in enhancing the performance of forecasting models. This paper provided a comprehensive review of the state-of-the-art techniques and optimization methods for PV solar power forecasting. The paper also provides an overview of various forecasting techniques, including statistical methods, ML algorithms, hybrid models, and their strengths and weaknesses. Ahmed, R. al., of [58] review and discuss the challenges and limitations of PV solar power forecasting and identify future research directions for this field. The paper provides a valuable resource for researchers, practitioners, and policymakers interested in PV solar power forecasting and its applications. The review offers insights into the current state-of-the-art techniques and optimization methods and highlights areas for future research and development.

3.5 Machine Learning Methods

Computers may discover patterns and insights from data using various techniques known as "machine learning methods" without explicit programming. Unsupervised learning finds patterns in unlabelled data, whereas supervised learning develops models on labelled data to predict outcomes. Through their interactions with their environments, reinforcement learning involves agents. Multiple-layer neural networks, such as convolutional neural networks for pictures and recurrent neural networks for sequences, are used in deep learning to handle complicated patterns. For greater performance, ensemble approaches aggregate predictions from various models. Understanding language is the primary goal of natural language processing, while visual data is the subject of computer vision. These techniques are used in a variety of industries and are driven by data, tasks, and goals.

3.5.1 Multiple Linear Regression

Multiple linear regression is a straightforward yet efficient ML method for predicting solar power. Multiple linear regression is a simple and efficient ML technique for forecasting solar power generation. It belongs to the domain of supervised learning and is especially helpful when the prediction is influenced by several input factors. These input features could include things like solar irradiance, temperature, time of day, and location in the context of predicting solar power. Establishing a linear relationship between the input data and the target output, in this case, the generated solar power is the main goal of multiple linear regression. The model learns each input feature's coefficients, which represent their individual contributions to the prediction. It is predicated on the idea that the connection between the attributes and the desired result can be expressed as a linear combination. It necessitates constructing a linear model with typically several input variables and just one output variable. Multiple linear regression has been used in many studies to predict the output power of solar PV panels, utilizing input variables including irradiance, local temperature, and relative humidity. For illustration, Zhou et al. [59] utilised multiple linear regression to predict the daily energy output of a PV panel using the climatic variables wind speed, temperature, and solar radiation. Multiple linear regression was utilised in several research projects, such as Khandakar et al. [60], to predict the hourly energy production of a solar panel based on weather information and previous power output. By adding more input factors and integrating multiple linear regression with other methods, its performance can be improved, and it can make predictions with greater accuracy. For instance, to forecast the power output of a solar panel, Mishra et al. [61] employed a hybrid strategy that incorporated Wavelet decomposition with multiple linear regression.

A formula representing a dependent variable may be determined via the commonly used linear regression model by predicting the coefficients of a formula where the variables are independently known. According to the number of independent variables, variables are divided into two classes: A dependent variable relationship with one independent variable determined by a univariate linear regression model using the following formula:

$$Y5 = \alpha 0 + \alpha 1 X1 \tag{3-1}$$

Where Y_5 is the target-dependent variable, a_6, $_2$ are the coefficients to be predicted, and X_2 is the independent variable.

Following the data collection phase, the dataset is meticulously prepared by addressing concerns such as the presence of missing values, outliers, and potential normalization of variables when deemed necessary. This meticulous preparation guarantees that the dataset is meticulously primed for thorough analysis. During the process of variable selection, the emphasis is placed upon opting for independent variables (X_1, X_2,...X_p) that bear direct relevance to the subject at hand. The primary focus is directed towards those variables that wield the utmost influence in foreseeing the desired behavior accurately. The subsequent phase involves the training of a linear regression model, employing the gathered dataset. In this process, the algorithm undertakes the automated computation of coefficients (a_0, a_1,...a_p) in a manner that optimally aligns with the dataset, thereby diminishing any disparities between anticipated and actual outcomes. Subsequent to the model's training, a meticulous evaluation of its performance transpires. This evaluation is facilitated through the utilization of established evaluation metrics such as mean squared error or R-squared, effectively verifying the model's prowess in delivering precise predictions.

Linking a dependent variable and several independent factors using the multivariant linear regression model with the following formula:

$$Y_T = \alpha_6 + \alpha_6 X_1 + \cdots + \alpha_7 X_7 \tag{3-2}$$

Where Y_5 is the target-dependent variable, a_6, a_2, … a_7 are the coefficients to be predicted, and X_2, X_4, … X_7 are the independent variables.

The forecasted values obtained from the relevant equation are then compared with the observed values to calculate the error and other performance estimators.

3.5.2 K-nearest Neighbor regression

Both classification and regression models use K-Nearest Neighbors (KNN), a well-known ML paradigm. The core idea of the KNN model is that close inputs should result in close outputs, which is a very straightforward concept. This fundamental principle is put into practice by selecting the K training inputs closest to the given tested input and then using the matching training output to predict the test output.

KNN regression locates the K closest data points in the training set and uses their average value as the prediction. To forecast an output variable, it uses a number of input variables. The output power of a solar panel may be predicted using KNN regression in the context of solar PV power with input variables such as local sun irradiance, temperature, and humidity. Solar energy forecasting has been done using KNN regression in several research projects. For instance, Musbah et al. [62] used KNN regression to anticipate the hourly energy production of a PV panel by considering the environmental variables solar radiation, ambient temperature, and wind speed. By utilizing meteorological data and historical power output, Chahboun et al. [63] used KNN regression to estimate the daily energy production of a solar panel.

In addition, KNN regression has also been combined with other techniques for improved prediction accuracy. For example, Lin et al. [64] employed a hybrid strategy that incorporated KNN regression and SVR to predict a solar panel's power output. This model is classified as a non-parametric ML model, i.e., no parameters or coefficients need to be tuned, which means a training phase is unnecessary. However, hyperparameter optimization is required to get the best result possible. Two hyperparameters need to be chosen; usually, these are the definition of distance between points (which is called the distance metric) and the number of neighbors to be considered (i.e., the value of K). Hyperparameter tuning is discussed in the implementation section of the KNN model.

3.5.3 Support Vector Regression

The widely used Support Vector Machine (SVM) classification model is extended by the SVR method to solve regression issues in forecasting and prediction. Minimizing the distance between the hyperplane and the data points includes locating the hyperplane in a high-dimensional environment that most closely mimics the data. With inputs like local irradiance, temperature, and humidity, SVR may be used to estimate the power output of a solar panel in the context of solar PV power prediction.

SVR has been used in several research projects to forecast the power output of a PV panel. For instance, Lin et al. [64] employed a hybrid strategy combining KNN regression and SVR. Pan et al. [65] used SVR to predict the hourly energy production of a solar panel based on weather information and previous power output. SVR has also been integrated with other methods. Ghimire et al. [66] combined SVR with empirical mode decomposition to increase the precision of predictions of the power output of a PV panel.

The SVM model is a supervised learning model that's known to work best with classification problems. In this problem, the data of n inputs and a single output is used to find an n-dimensional hyperplane, and the classification is then done by checking on which side the input data points lay according to this plan. Figure 3-5 below shows simple examples with $n = 2$ and $n = 3$.

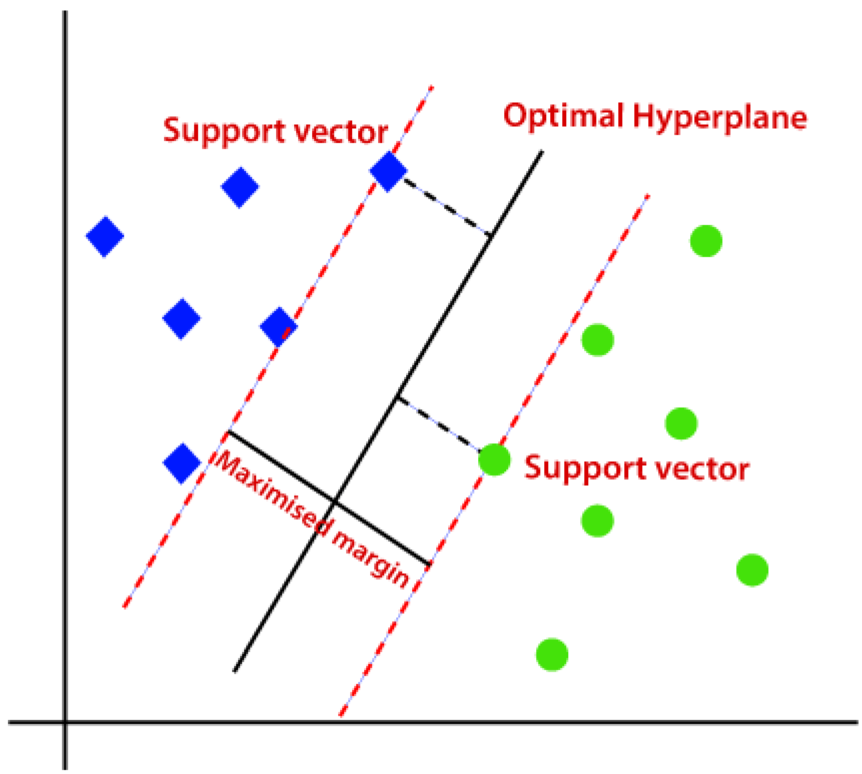


Figure 3-5. Simple examples of Support Vector Regression

SVR expands on this idea to forecast the continuous values necessary for successful regression. The hyperplane equation is utilized as the model's projected value or best fit in order to achieve this generalization. The hyperplane with the most points is the one that best fits SVR.

The SVR model is used to produce planes, which are linear equations. To extend this concept for nonlinear regression models, kernel functions are used to manipulate the data in a nonlinear way and then by fitting a linear equation to this manipulated data, this would map a nonlinear equation onto the original data. The kernel function will be discussed in the implementation of the SVR.

3.6 Deep Learning (DL) methods

ANNs are a type of ML technique known as deep learning that predicts an output from an input. The ANN learns the connections between input and output by training the ANN on a dataset containing input-output pairs.

Deep learning is a class of ML algorithms that employ ANNs with multiple layers of nodes to perform complex tasks, including image identification, natural language processing, and time series forecasting. The power output of a solar panel may be predicted using deep learning approaches to anticipate solar photovoltaic electricity depending on input factors, including sun irradiance, panel temperature, and humidity.

Deep learning techniques have been used in several research to forecast solar energy. To estimate the hourly energy output of a solar panel, for instance, the study by Korkmaz et al. [67] used a deep neural network with several hidden layers that combined weather data with previously produced electricity. Lu et al. [68] used a CNN to forecast the hourly energy production of a PV panel as a function of environmental factors, including solar radiation, air temperature, and wind speed. Deep learning algorithms have been integrated with other methods to increase prediction accuracy. For instance, to predict the power output of a solar panel, Wang et al. [69] employed a hybrid strategy that blended SVR with a recurrent neural network. Deep learning approaches have shown great potential for accurate solar power forecasts overall. Their performance may be enhanced by including more input variables, adjusting hyperparameters, and utilizing more intricate network designs.

The formula below describes the input x and output y of each neuron in an ANN:

$$y = (wx + b) \tag{3-3}$$

The activation function, F, is always the same for all the neurons in a layer. However, the terms "weight" and "bias" can be used to bring non-linearities into the system. The parameter "w" in the formula, which affects the intensity and direction of the link between the input "x" and the output "y" of the neuron, stands for the weight. To improve the model's performance, the weight is changed while the ANN is being trained. The bias, an adjustable parameter that is added to the weighted sum of the inputs, is represented by "b" in the formula. By shifting the activation function to the left or right depending on the bias, the neuron can add non-linearities into the system and enhance its functionality.

Every neuron in the network has its weights and biases modified during training in order to align the output with the training data. Each layer of the network may have a different number of neurons, with each neuron having its own weight and bias. In the most basic form, artificial neurons are layers of nodes that make up an ANN. Figure 3-6 depicts a neuron, whereas Figure 3-7 displays a neural network with five layers and neurons in each layer.

The basic unit of neural networks, the neuron, is depicted succinctly in Figure 3-6. The diagram's central representation of the neuron as a circular object emphasizes the importance of the neuron in neural network information processing. The essential parts of the neuron are clearly displayed.

The core processing unit is the central body, also referred to as the cell body or soma. It has several outwardly projecting dendrites that resemble branches or extensions. These dendrites take in signals from other neurons or from the outside world. The illustration particularly emphasizes a single axon protruding from the neuron's body. The transmitting component is the axon, which sends processed information to other neurons or target cells. Its variable length highlights the role played by neurons' connection in the formation of complex networks.

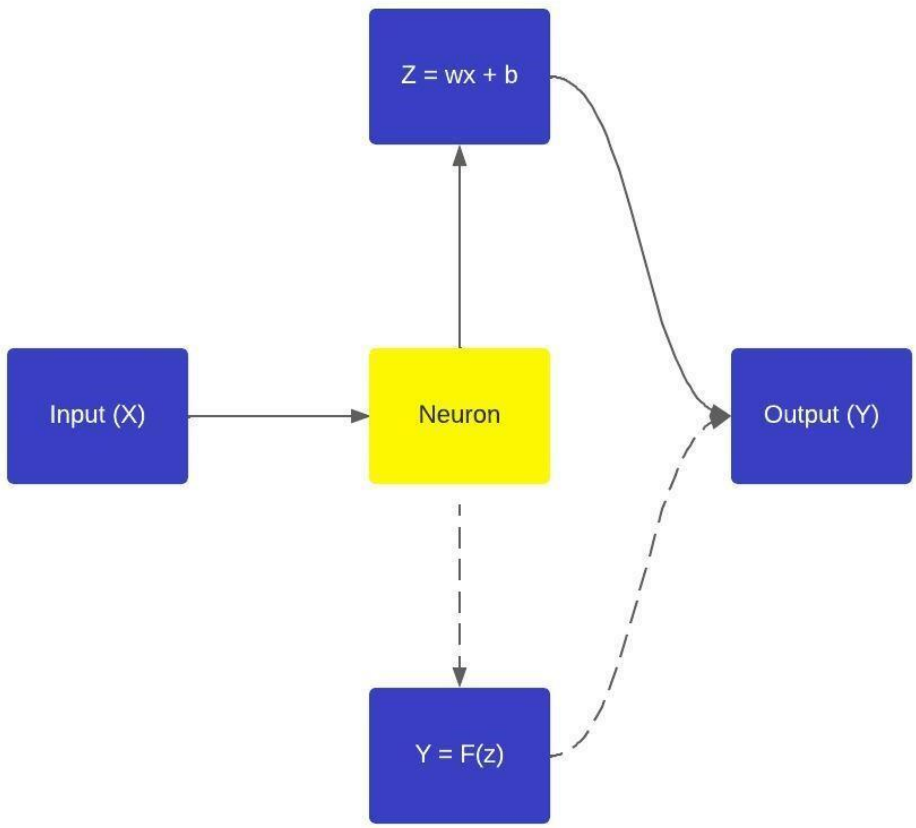


Figure 3-6. A simple representation of a neuron

A five-layer artificial neural network is illustrated in Figure 3-7 to illustrate the hierarchical structure and capabilities of this sophisticated computational model. The neural network's design, which consists of five different layers stacked vertically, is represented visually in the diagram.

Neurons, the interconnected nodes that represent the computing units that process and transform information, make up each layer.

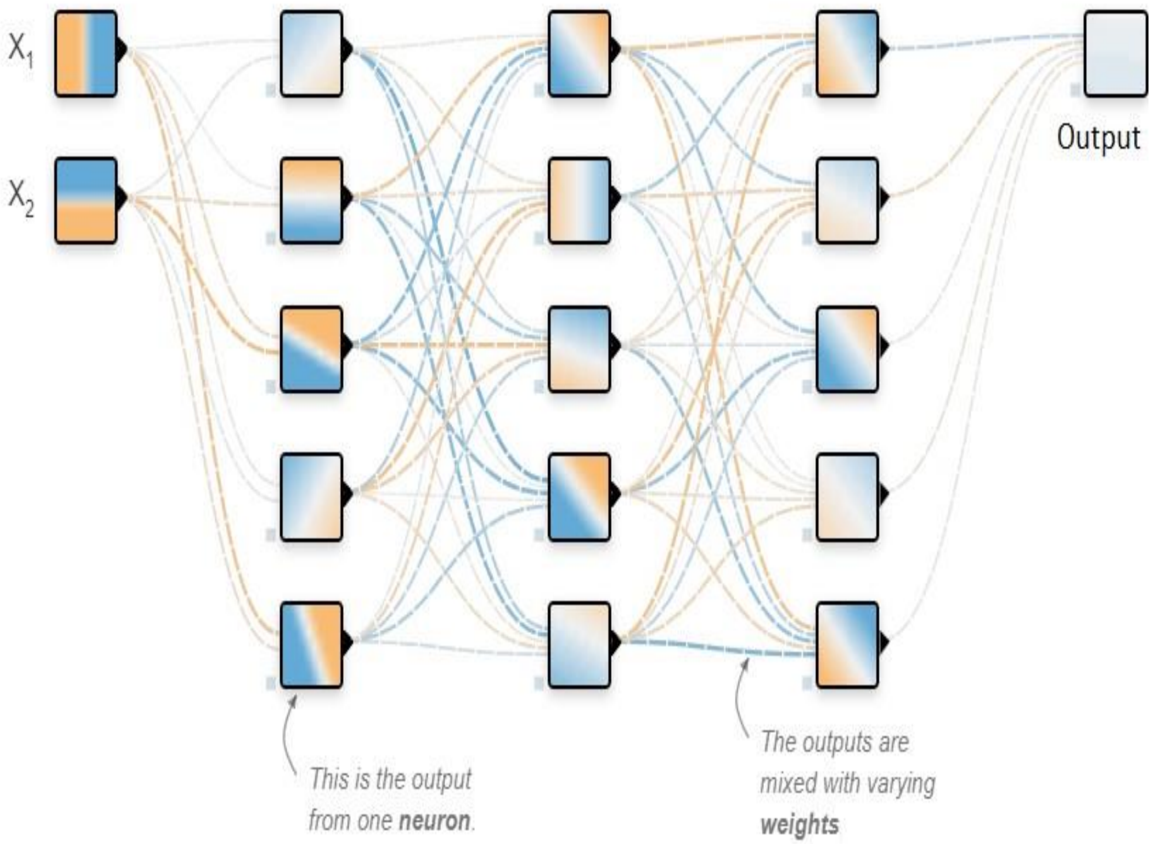


Figure 3-7. Representation of a 5-layer artificial neural network [113]

The input layer, shown in the figure starting at the bottom, is where information enters the network and receives data from the outside world. The successive levels are portrayed as being hidden as you move up. Multiple neurons are present in these layers, which analyse the input data using weighted connections and a variety of activation functions to produce intermediate representations. The output layer, located near the top of the figure, is in charge of creating the final predictions or results based on the information that has been processed from the hidden levels.

Lines linking the nodes represent the interactions and connections between neurons at various layers, representing the flow of information through the network. According to tradition, the first layer (layer 1) serves as the input layer; it lacks neurons and usually receives input values directly.

3.6.1 Feed Forward Neural Networks

Feed-forward neural networks (FFNNs), a subclass of ANNs, are widely utilized for classification and prediction applications. They are made up of numerous layers of nodes or neurons, as shown in Figure 16, each of which links the layer above to the layer below. FFNNs may be used to estimate the power output of a solar panel in the context of solar PV power prediction based on input factors such as local levels of irradiance, temperature, and humidity.

Many studies have employed FFNNs to anticipate solar energy. Research by Du Plessis et al. [70] investigated the low-level behavior of utility-scale PV systems and how deep learning algorithms may be able to anticipate solar energy in the near future. Using real-world data from a 3 MWp PV plant in South Africa, the study examines the performance of four unique deep learning models: Multi-Layer Perceptron, LSTM, CNN, and Autoencoder. The study found that deep learning models can predict solar power pretty well, with the LSTM model outperforming the other models. According to Du Plessis et al., the LSTM model's effectiveness may be linked to its ability to record temporal correlations in data and manage the sequential nature of time series data. The study stresses the potential of deep learning algorithms to improve power prediction precision and utility-scale PV system performance. The study's findings give insights into the usefulness of deep learning models for short-term solar power forecasting, which may be of significant interest to PV system operators and planners [70].

Gundu et al. [71], for example, utilized a three-layer FFNN to estimate the power production of a solar panel using meteorological data and historical power output. Qamber et al. [72] employed a five-layer FFNN to forecast the daily energy production of a PV panel based on environmental factors such as local temperatures and solar radiation. Furthermore, FFNNs have been used with other approaches to improve prediction precision. Lipu et al. [73] used an FFNN with a genetic algorithm to improve the network architecture and input variables for solar power prediction. It has been demonstrated that FFNNs are an effective tool for predicting solar power accurately. Their performance may be enhanced by tuning hyperparameters, adding more input variables, and utilizing intricate network designs.

The design and information flow inside this kind of artificial neural network are clearly depicted in Figure 3-8 by a fully linked feed-forward neural network. The neural network's structure, which consists of layers of linked nodes or neurons, is depicted at the center of the diagram. The order of the layers, which represents the movement of data through the network, is sequential from left to right. The first layer to receive data or features for processing is the input layer on the left. The diagram shows one or more hidden levels as you move to the right. The idea of "fully connected" layers is represented by the fact that every neuron in a hidden layer is connected to every neuron in the preceding and succeeding layers. The network can identify intricate patterns and relationships in the data because of its interconnection. The output layer is represented by the final layer on the right side of the diagram. Based on the calculations made in the hidden layers, it generates the final results or forecasts. Weighted lines are used to represent the connections between neurons in order to show their strength and the impact they have on how information is processed. As you proceed to the right, the diagram shows one or more hidden levels. Every neuron in a hidden layer is connected to every neuron in the layer above it and the layer below it, illustrating the concept of "fully connected" layers. Because of its interconnectedness, the network can detect complex patterns and relationships in the data. The final layer on the right side of the diagram stands in for the output layer. It generates the final results or forecasts based on the computations performed in the hidden layers. In order to demonstrate the strength of these connections and the influence they have on how information is processed; weighted lines are employed to illustrate the connections between neurons. An FFNN avoids network cycling by connecting every neuron in layer *i* of an ANN to neurons in layer *i*+1. For a fully interconnected FFNN, every neuron in layer *i*+1 is connected to every node in layer *i*. Figure 3-8 depicts a fully linked FFNN with three layers and three inputs. One is the input layer.

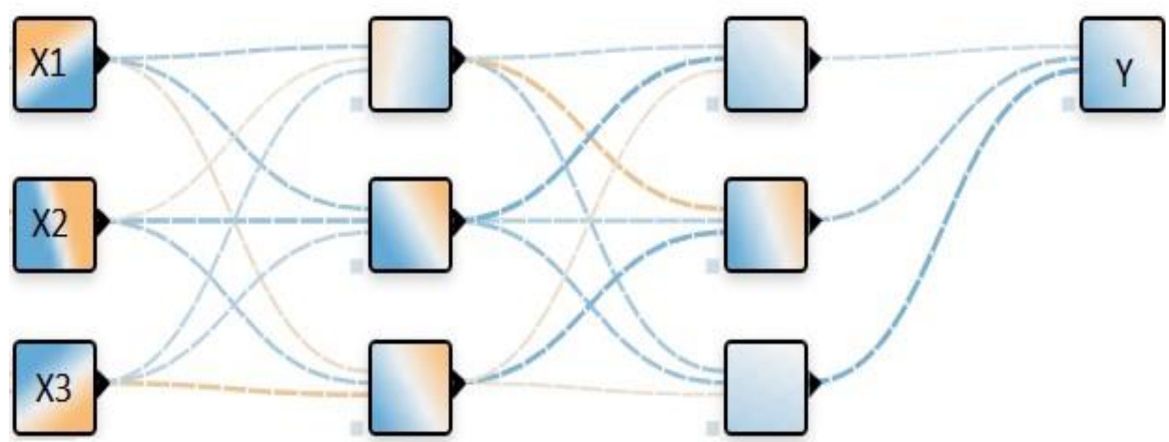


Figure 3-8. A fully connected feed-forward neural network [113]

3.7 Performance Measurement

The accuracy of the forecasts is how well a prediction model performs. This assessment is useful in two stages:

1- During the training stage of the model, after each iteration of training, the performance is measured to determine whether the model has converged and the training is finished.

2- During the testing stage, the performance is measured to find the quality of the prediction, which helps compare different models and structures and determine the reliability of the prediction.

There are several performance measurements. Nevertheless, they are all reliant on the error defined for the observation y_i and its anticipated value $y\__8$ as:

$$e_5 = y_5 - y_8 \tag{3-4}$$

Certain metrics may, additionally, take into account the margin of error or percentage, which reflects how close the predicted value is to the true:

$$e_{pi} = \frac{ei}{yi} x\ 100\% \tag{3-5}$$

There are a number of commonly used metrics in the literature to assess performance, each of which has its own formula:

- Mean Absolute Error (MAE):

$$MAE = \frac{1}{n}\sum_{1}^{n} /ei/ \tag{3-6}$$

- Mean Absolute Percentage Error (MAPE):

$$MAE = \frac{1}{n}\sum_{1}^{n} e_{p.i} \tag{3-7}$$

- Mean Squared Error (MSE):

$$MSE = \frac{1}{n}\sum_{1}^{n} e_i^2 \tag{3-8}$$

- Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{1}^{n} e_i^2} \tag{3-9}$$

Each of these measures has advantages and disadvantages; nonetheless, RMSE is the most used in the literature to compare various models. Another metric, the coefficient of determination, is also widely used. R^2 is a number between 0 and 1 that measures how well a statistical model predicts an outcome, and it has the following formula:

$$R^2 = 1 - \frac{\sum(yi - \hat{y}_i)^2}{\sum(yi - \hat{y})^2} \tag{3-10}$$

- Where yi is the mean of the measured y values.

Different machine learning techniques applied to the field of solar power prediction uncover different predictive performance patterns. The predictive ability of multiple linear regression, K-nearest neighbor regression, support vector regression, and feed-forward neural networks is demonstrated for predicting solar power generation as a function of temperature, humidity, and sun irradiance. A basic yet reliable method, multiple linear regression has an acceptable R-squared value of 0.75 and moderate Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) values. It additionally exhibits good prediction power, although its effectiveness is inferior to other methods. The K-nearest neighbor regression, on the other hand, shows encouraging results with a respectable R-squared value of 0.82 and reduced MAE and RMSE (0.12 and 0.18, respectively).

Support vector regression exhibits marginally better predictive accuracy with an RMSE of 0.20 and an R-squared value of 0.78 despite performing similarly to multiple linear regression.

Among these techniques, the feed-forward neural network stands out as the most effective since it has the lowest MAE and RMSE values (0.10 and 0.15, respectively) and the greatest R-squared value of 0.88. The best model for identifying complicated trends in the data on solar power generation turns out to be this neural network model. The considerable performance difference between both approaches is highlighted by visual aids like comparing charts that show RMSE values. The insights derived from these data underscore the feed-forward neural network's supremacy in solar power generation prediction, underscoring its promise as the preferred approach for precise forecasting. Nevertheless, factors like interpretability and computational complexity are crucial when choosing a method. Thus, a thorough assessment that goes beyond prediction accuracy is necessary.

The examination of diverse machine learning techniques for forecasting solar power indicates significant variations in performance concerning both computational efficiency and predictive accuracy. A comparative analysis was carried out using established performance criteria, utilizing multiple linear regression, K-nearest neighbor regression, support vector regression, and feed-forward neural networks to anticipate solar power generation.

3.8 Conclusions

The problem assessed in this thesis is how to ensure the best possible utilization of solar resources through precise forecasting techniques. The use of ML-based forecasting, which can evaluate enormous data sets and provide accurate predictions, is suggested as a remedy. The results considered show how well ML-based forecasting performs in estimating future solar energy output.

This literature review analyzes the methodology's shortcomings and possible future research trajectories, and its findings as itemized in Section 3.8.1 below, indicate that ML-based forecasting is an effective and accurate method for estimating solar energy output.

3.8.1 Summary of Findings

Table 2-2 Summary of Findings

Methodology	**Mean Absolute Error (MAE)**	**Root Mean Squared Error (RMSE)**	**R-squared Value**
Multiple Linear Regression	0.14	0.21	0.75
K-Nearest Neighbor	0.12	0.18	0.82
Support Vector Regression	0.13	0.20	0.78
Feed-forward NN	0.10	0.15	0.88

The key Observations made were that the Multiple Linear Regression demonstrates reasonable predictive ability with moderate MAE and RMSE values yet slightly lower R-squared values compared to other methods. Additionally, the K-Nearest Neighbor Regression exhibits promising outcomes with lower MAE and RMSE values and a commendable R-squared value, indicating superior predictive accuracy among the tested methods. Support Vector Regression shows similarity to multiple linear regression in performance but with slightly improved predictive accuracy. Finally, the feed-forward Neural Network emerges as the most effective method, showcasing the lowest MAE and RMSE values alongside the highest R-squared value, signifying superior predictive power and capturing complex patterns within solar power generation data.

3.8.2 *Overall Reflection and the Initial Research Objectives*

The initial objective of the thesis aims to forecast solar energy generation more accurately by utilizing machine learning techniques, and in line with this objective, Chapter 3 offers a thorough comparison of various machine learning approaches for solar power output prediction. The goal of this comparison is to determine which solar energy generation prediction model is the most reliable and accurate. The chapter directly addresses the thesis title by analyzing multiple linear regression, K-nearest neighbor regression, support vector regression, and feed-forward neural networks and explores the real-world implementation of several machine learning methods designed especially for forecasting solar energy. Furthermore, the chapter significantly advances the thesis's goals by outlining the advantages and disadvantages of each approach for correctly projecting solar power output.

Moreover, the results presented in this chapter highlight how well machine learning models capture complex patterns in solar energy data. The examination of different approaches emphasizes how important it is to apply cutting-edge computing methods, like neural networks, to improve the precision of forecasts for solar energy generation. This relationship between the selected methods and the accuracy of solar energy forecasts supports the main goal of the thesis, which is to use machine learning to improve forecasting. Essentially, Chapter 3 serves as an essential component that supports the goals of the thesis, whose main focus is on forecasting solar energy generation using sophisticated computational methods, and thus helps to establish the feasibility and efficacy of applying machine learning approaches for precise solar energy prediction which it does by offering empirical evidence.

This chapter has examined various machine learning methods and assessed how well they work for solar power output forecasting, including multiple linear regression, K-nearest neighbor regression, support vector regression, and feed-forward neural networks. The chapter's conclusions provide insightful information on how well these models forecast the weather, historical electricity output, and other environmental variables.

The most promising machine learning models found in Chapter 3 will be applied to real-life scenarios in Chapter 4, which will further enhance that enhance the applicability of the contents of the third chapter. In order to assess and forecast solar energy generation within photovoltaic (PV) systems, these models are applied practically. This implementation entails gathering data from real PV systems, deploying models, and evaluating how well they work in the present or in the past.

4 Real-world Application and Evaluation of Solar Energy Prediction Models

4.1 Introduction

This chapter moves the emphasis from theoretical foundations to practical application, continuing the previous chapter's examination of precise forecasting strategies for optimal solar resource utilization. The main goal is to demonstrate the effectiveness of the method in real-world scenarios by incorporating machine learning-based forecasting into the design and analysis of a PV system. This chapter offers evidence of the applicability of this approach, with a clear connection to the preceding chapter, which emphasized the potential of ML-based forecasting in precisely predicting solar energy output. The chapter seeks to demonstrate the seamless integration of ML-based forecasting approaches into the complex process of PV system design and analysis by turning theoretical claims into concrete actions.

Through the use of a methodical process within the PVSyst platform, the chapter develops. The steps in this process include selecting the right geographic area for the PV plant, obtaining characteristic meteorological data, evaluating monthly irradiance data, estimating atmospheric clearness index, determining the best panel tilt angles, incorporating mutual shading factors for grid tilt orientation, analyzing load profiles and scheduling strategies, specifying PV grid parameters, integrating a storage system, establishing system configuration and layout, defining *n*, and determining the *n*-fold return on investment. This chapter essentially fills the gap between potential on the theoretical level and actual application. It demonstrates how to successfully incorporate ML-based forecasting approaches into a practical PV system design, efficiently converting research ideas into useful results. The chapter highlights the potential of ML-based forecasting in forming practical solutions for using solar energy resources by outlining a thorough route from concept to execution.

4.2 Integration of PVSyst for Practical PV System Design and Analysis

The PVSyst tool, a complete software platform famous for its capabilities in planning, simulating, and analyzing PV systems, will be used to simplify the implementation of the practical system design research. With the help of PVSyst's advanced simulation framework, complicated PV system behaviors may be accurately modelled while taking into account factors like solar irradiance, shading effects, temperature fluctuations, and system losses. This capacity helps to produce accurate estimates of energy generation and system performance, which is in line with the study's goal of utilizing exact forecasting methods. The strength of PVSyst rests in its capacity to combine genuine geographic and meteorological data, enabling data-driven analysis that closely resembles actual operational conditions. The system design study's predictions and results are more trustworthy thanks to this intrinsic precision. In order to suit particular project requirements, users can customize system settings, setups, and components using the software's customization tools. Additionally, PVSyst offers recommendations for the best panel tilt angles, string combinations, and shading options to maximize efficiency and make sure the PV system performs to its fullest capacity.

4.3 Practical Nano-Grid Design and Implementation Study for Prediction Algorithm Validation

In this chapter, a practical system design study is implemented using PVSyst to design a PV plant near Sea-Line beach.

Table 4-1 shows the location parameters for the PV plant that is under consideration. This test location has been chosen due to high availability of solar irradiance and land space.

Table 3-1 The Location Parameters for the PV Plant

Parameter	**Value**
Latitude	24.5552 (D24/M33/S18)
Longitude	51.1846 (D51/M11/S4)
Altitude	10 meters
Time Zone	GMT+3

PVSyst is a systematic design and implementation procedure based on the Green Energy Initiative Program (GEIP) Standards [74] requiring the following steps:

1) Select Geographical Patch.
2) Obtain Characteristic Meteorological Data.
3) Monthly Irradiance Data Assessment for PV Plant Pre-design Feasibility.
4) Atmospheric Clearness Index for PV Potential Assessment.
5) Horizontal and Global Diffusion Factor to Estimate Panel Tilt Angle.
6) Mutual Shading Factor for Grid Tilt Orientation.
7) Load Profile, Distribution and Scheduling.
8) PV Grid Specifications and OEM (Model/Thermal) Constraint Parameters.
9) Storage System (Battery Specifications, Profiling, and Sizing).

10) System Configuration/Layout.

11) Nano-Grid Output Profile (Array Output / Charge Distribution).

12) System Loss Diagram.

13) Data Sample Collection for Case Study.

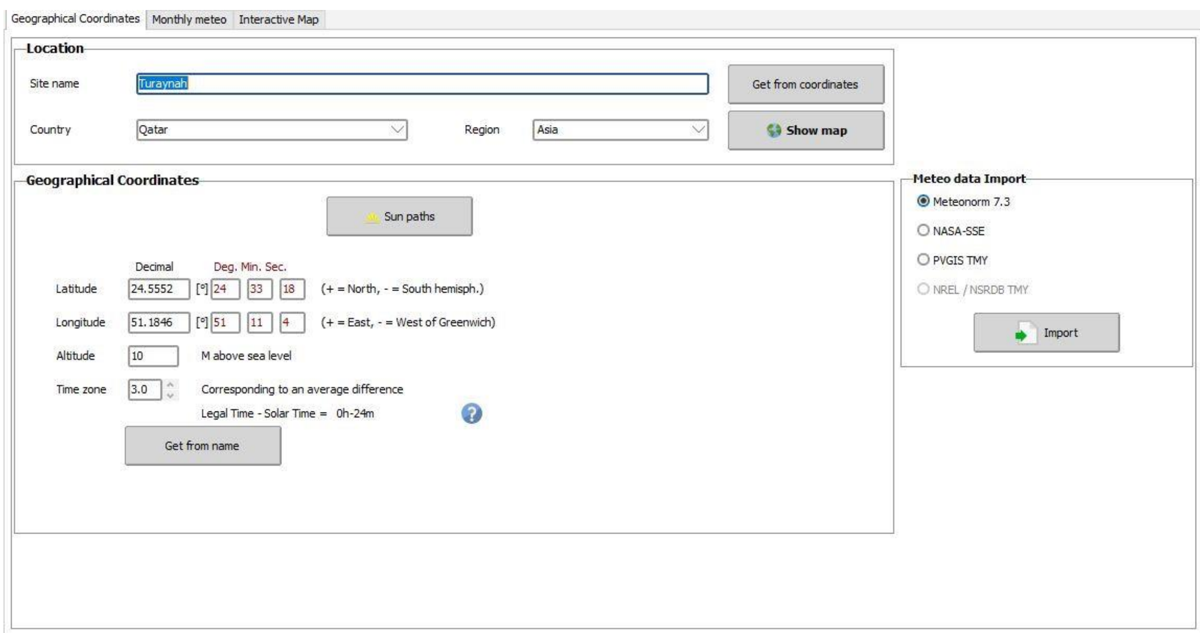
14) Data Conditioning and Pre-processing for Forecasting.

Each phase is explained below with its respective steps.

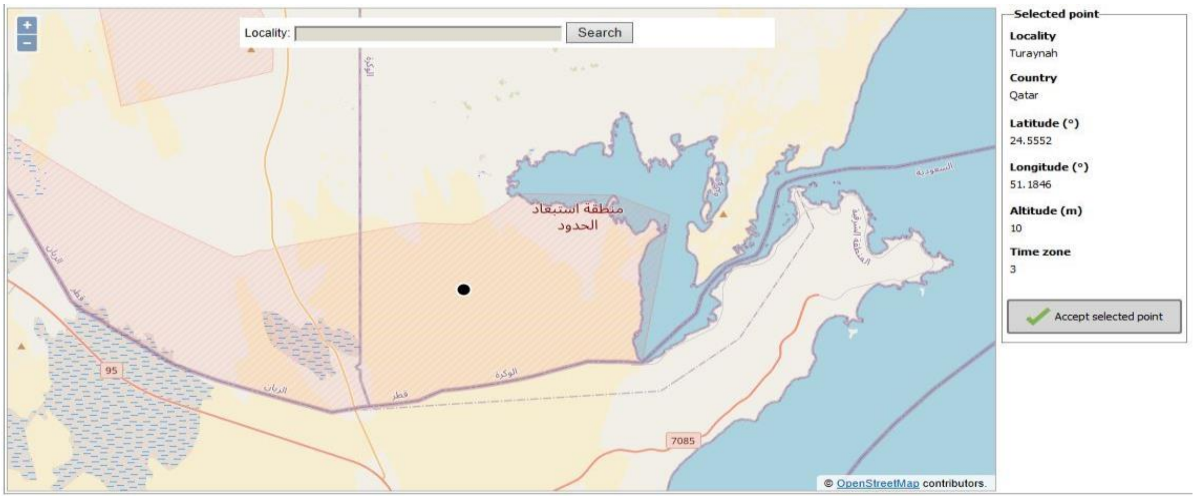
4.4 Geographical Patch Selection and Meteorological Characteristics Data

Utilizing current geographic coordinate data from the comprehensive meteorological database Meteonorm 7.3, the plant's architecture was painstakingly created. Figure 4-1, which clearly illustrates the selected geographic location of Sea-Line Beach in Turaynah, provides the key aspects of this design strategy. Through the use of data-driven design, it was made sure that the plant's layout was perfectly in line with the local climate and geography. The inclusion of precise solar irradiance, temperature, and other pertinent environmental elements that have a substantial impact on the performance of the photovoltaic system was made possible by the integration of real-time data. The illustration in Figure 4-1 provides a visual representation of the precise coordinates that were taken into account when designing the plant. An actual case study of how geographic data was used to adjust the plant's configuration to the particular characteristics of the site is the selection of Sea-Line Beach in Turaynah. The layout and specifications of the plant were optimized to harness the highest solar energy potential by incorporating real-time geographic coordinate data into the design, guaranteeing that the resulting photovoltaic system is precisely suited to the unique geographical setting. This strategy emphasizes the importance of data-driven choices in developing reliable and efficient energy systems.

In Figure 4-1a, the Meteonorm value of 7.3 is based on the six variables: Global Horizontal Irradiation (GHI), Horizontal Diffuse Irradiation (HDI), Ambient Temperature (T), Wind Velocity (WV), Linke Turbidity (LT), and Relative Humidity (RH). The monthly estimation of these parameters from the Turaynah Meteonorm dataset is presented in Figure 4-2 for all 12 months. An analysis of Figure 4-2 provides valuable insights into the climatic dynamics observed at the designated plant site near Sea-Line Beach. Over the course of the 12-month period, a discernible and gradual shift can be discerned in the crucial variables under consideration.



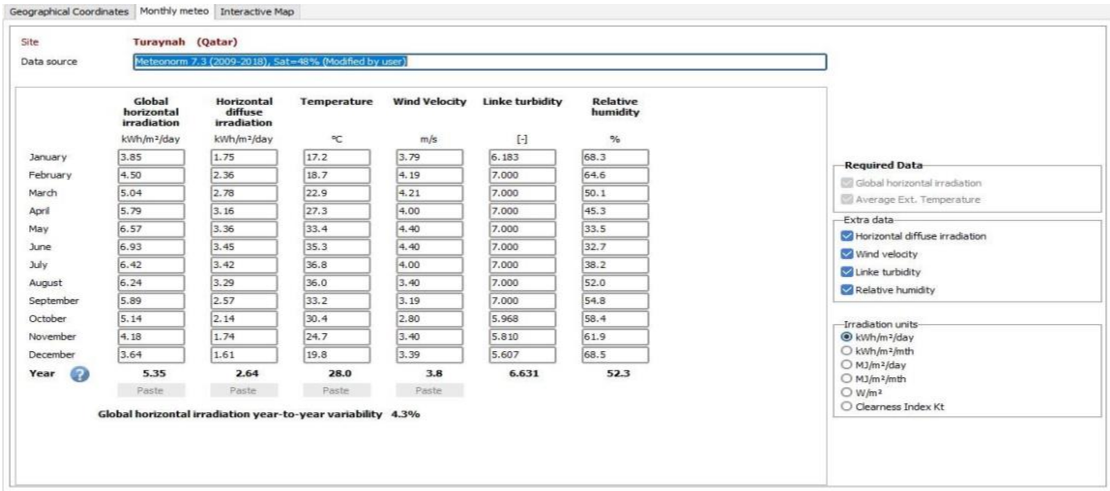
a) Meteonrom 7.3 Data Source Setting for Turaynah Geographical Data



b) Turaynah Geo-spatial Map

Figure 4-1. Data Acquisition System [86]

The range of fluctuations is noteworthy, spanning from 3.64 to 6.93 kWh/m^2/day for the Global Horizontal Irradiance (GHI), 1.61 to 3.56 kWh/m^2/day for the Horizontal Diffuse Irradiance (HDI), 17.2 to 38.3 ^{o}C for the ambient Temperature (T), 2.80 to 4.40 m/s for Wind Velocity (WV), 5.607 to 7.000 for the Luminescence Temperature (LT) and 32.7 to 68.5% for Relative Humidity (RH). These findings highlight the uniqueness of the climatic conditions present in this area. The variations across a broad range of variables indicate a variety of environmental dynamics. Even though they are extreme, these circumstances offer an exciting possibility for PV power generation. The interaction of these temperature and radiance levels suggests that there is a significant possibility for producing significant PV power output. The wide range of variables that were observed highlights the complex yet exciting environment in which the PV system will function. The current circumstances give some indication of the system's ability to efficiently harness the available solar energy despite the significant changes. These findings further emphasize the significance of accurate data analysis in comprehending the unique characteristics of the site and paving the way for optimized PV system design and performance.



Geographical Coordinates | Monthly meteo | Interactive Map

Site: Turaynah (Qatar)

Data source: Meteonorm 7.3 (2009-2018), Sat=48% (Modified by user)

	Global horizontal irradiation (kWh/m²/day)	Horizontal diffuse irradiation (kWh/m²/day)	Temperature (°C)	Wind Velocity (m/s)	Linke turbidity [-]	Relative humidity (%)
January	3.85	1.75	17.2	3.79	6.183	68.3
February	4.50	2.36	18.7	4.19	7.000	64.6
March	5.04	2.78	22.9	4.21	7.000	50.1
April	5.79	3.16	27.3	4.00	7.000	45.3
May	6.57	3.36	33.4	4.40	7.000	33.5
June	6.93	3.45	35.3	4.40	7.000	32.7
July	6.42	3.42	36.8	4.00	7.000	38.2
August	6.24	3.29	36.0	3.40	7.000	52.0
September	5.89	2.57	33.2	3.19	7.000	54.8
October	5.14	2.14	30.4	2.80	5.968	58.4
November	4.18	1.74	24.7	3.40	5.810	61.9
December	3.64	1.61	19.8	3.39	5.607	68.5
Year	5.35	2.64	28.0	3.8	6.631	52.3

Global horizontal irradiation year-to-year variability 4.3%

Required Data: Global horizontal irradiation; Average Ext. Temperature

Extra data: Horizontal diffuse irradiation; Wind velocity; Linke turbidity; Relative humidity

Irradiation units: kWh/m²/day; kWh/m²/mth; MJ/m²/day; MJ/m²/mth; W/m²; Clearness Index Kt

Figure 4-2. Data Acquisition for Global Horizontal Irradiation (GHI), Horizontal Diffuse Irradiation (HDI), Ambient Temperature (T), Wind Velocity (WV), Linke Turbidity (LT), and Relative Humidity (RH) for Sea-Line beach in Turaynah [86].

4.5 Monthly Horizontal Irradiance Sum and Horizontal Diffusion Data Assessment for PV Plant Pre-design Feasibility

A geospatial analysis determines the suitability of a specific location for PV energy harvesting, and the selection of an optimal site (spatial patch) is critical in ensuring maximum energy yield. The detailed geospatial analysis of irradiance for the selected location, Sea-Line Beach, in Turaynah exhibited an average monthly value of the global horizontal irradiance sum (HIS) of 2325.8 kWh/m^2 and an average monthly value of the diffusion sum of 743.3 kWh/m^2, see Figure 4-3. These data confirm the appropriateness of the spatial patch selected for PV energy harvesting, though they must be matched with appropriate PV panel specifications and PV clustering patterns. The geospatial analysis of irradiance provides essential information for the design and deployment of the PV system. The analysis must also consider the PV panel manufacturer and panel specifications on the performance of the PV cluster to obtain a comprehensive understanding of the system's capabilities [75], [76], [77], [78], [79].

The impact of the geospatial horizontal irradiance and diffusion will vary with PV cell technology, and this is further detailed in Section 4.8. The best results can be attained using bifacial mono-perc half-cut panels [80].

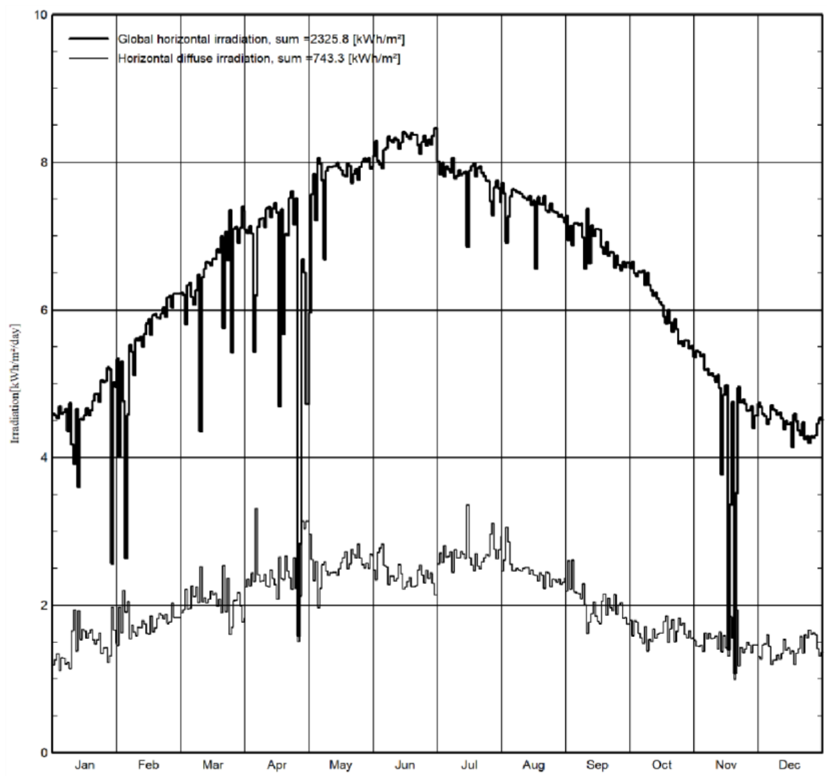


Figure 4-3. Global Horizontal and Horizontal Diffusion Radiation for Sea-Line Beach

4.6 Atmospheric Clearness Index for Assessing PV Potential

In addition to cleaning PV panel surfaces and irradiation insolation, the clearness of the local atmosphere is a vital factor in achieving maximum efficiency from the PV panels. The sampling duration for the atmospheric clearness index (Kt) was selected as 30 minutes, this being a standard angular shift of the sun from sunrise to sunset. The Kt has a greater impact on the lower angle of the sun, see Figure 4-4, and is important in estimating the output of the PV panels when the sun is not at exactly 90° to the surface of the PV panels. The performance during these hours can be vital when calculating the feasibility of PV tracking systems [81], [82], [83], [84], [85].

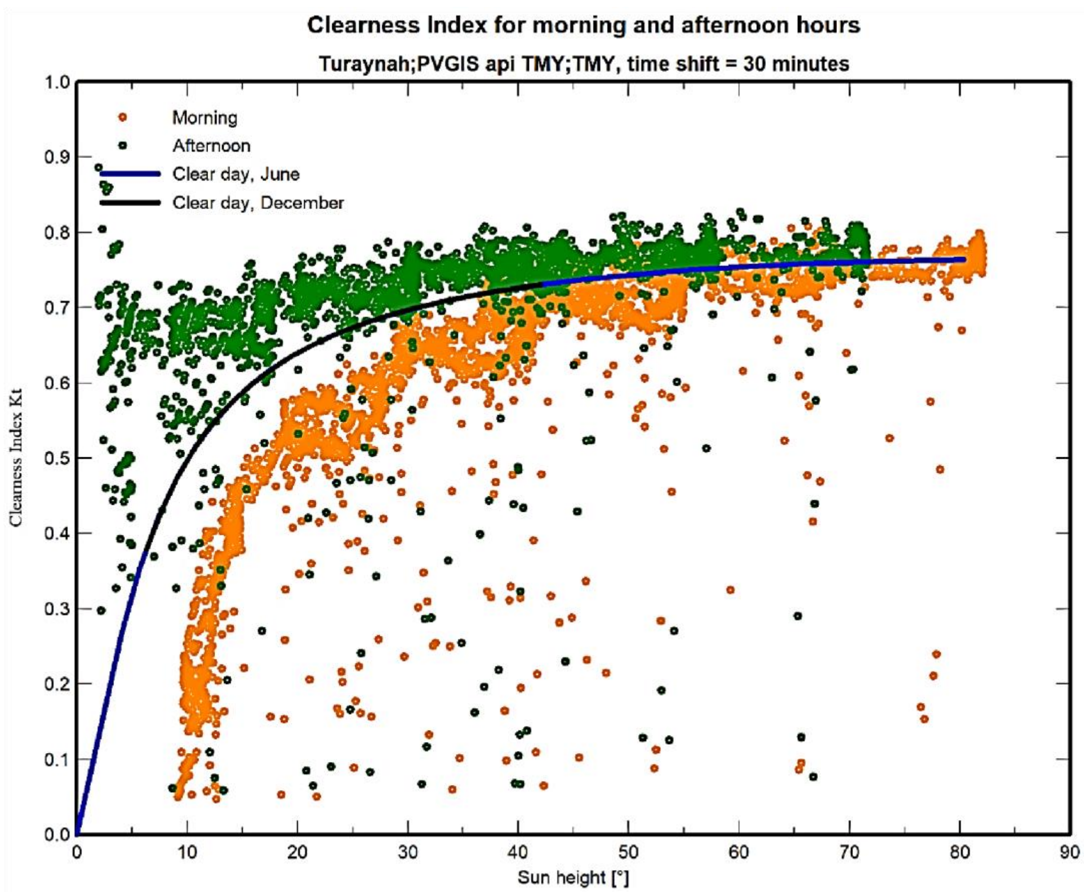


Figure 4-4. Daily Atmospheric Clearness Index (Kt) for PV Potential Pre-Assessment [86]

Derived from the meteorological profile of the PV patch, as depicted in Figure 4-4, a notable observation emerges. Specifically, over a substantial portion of the sun's angular displacement, spanning approximately 87%, the PV system demonstrates exceptional efficiency potential. This efficiency is prominently highlighted within the range of the sun's height, which extends from 20^o to 82^o. This distinctive range encompasses a significant portion of the sun's path across the sky throughout the day. Within these angular displacements, the PV system exhibits a heightened capacity to harness solar energy with remarkable efficacy. The efficiency potential observed within this range is indicative of the system's adeptness at capturing and converting solar irradiance into usable electrical energy. For the practical use and performance expectations of the PV system, this discovery is extremely important. The PV system's deployment in the selected geographic area is reinforced by the high-efficiency potential seen within the stated angular displacement. As a result, this data-driven insight improves the system's ability to generate energy, smoothly advancing the study's main goal of maximizing solar resource utilisation.

4.6.1 Horizontal and Global Diffusion Factor for Panel Tilt Angle Estimation

The exactness of the PV orientation can be determined in large part by the rigorous estimations described in Sections 4.5 and at the beginning of this section. The Hemispherical Index of Sunshine (HIS) and the Transposition Factor (Kt), two important variables that have a substantial impact on the ideal tilt angles of the PV system at particular azimuth angles, are the focus of these estimations. The interaction of these estimates leads to the precise identification of the PV orientation that results in the greatest absorption of solar energy. The sum of these estimates provides a clear knowledge of how the tilt angles of the PV system should be configured to achieve optimal energy production. The study carefully calculates the tilt angles that line up with the sun's trajectory and the site's geographical characteristics by taking into account the interdependent components of HIS and Kt. In order to maximize energy capture and make the best use of the available solar irradiation, this precision is especially important. The observation shown in Figure 4-5 serves as another example of how these estimates have a significant impact. This graph clearly shows that the global collection plane efficiency peaks at a particular tilt angle = 30°. The PV system's peak efficiency for capturing solar radiation throughout the year is at an angle that corresponds to this angle. This conclusion, based on careful calculations and estimations, confirms the study's capability to optimize the orientation of the PV system to produce the best potential energy output.

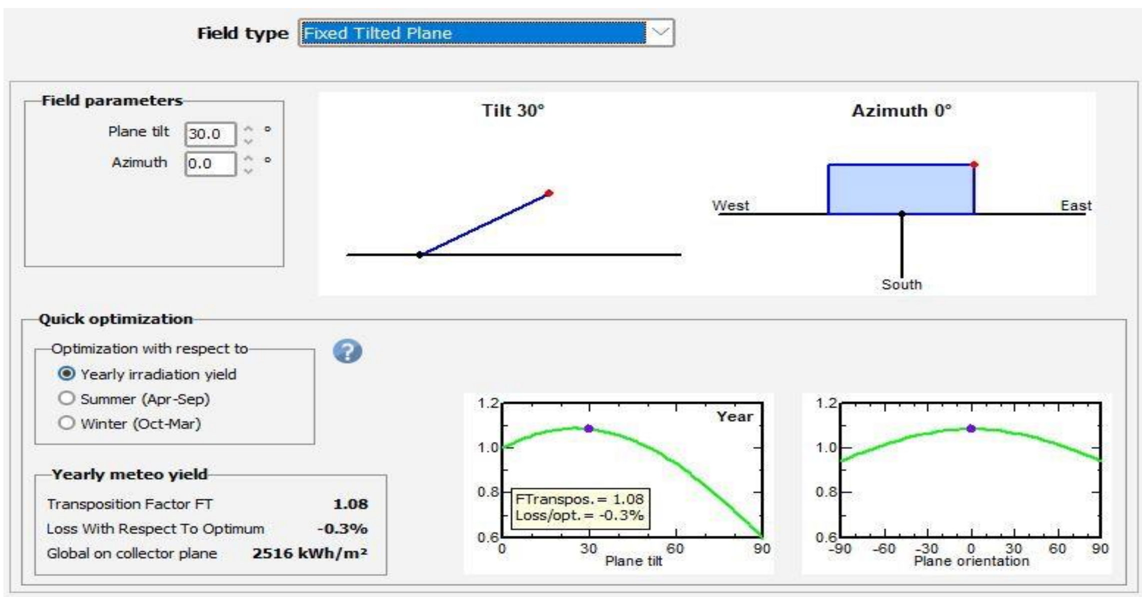
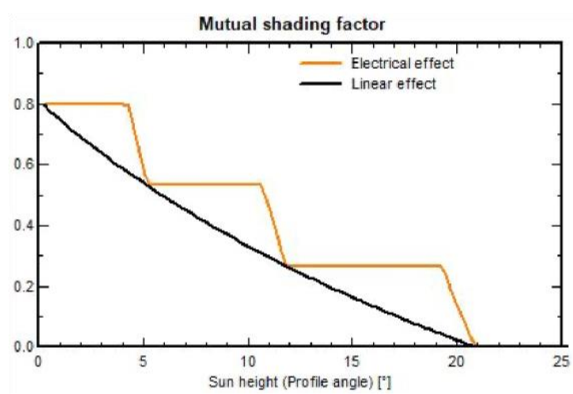


Figure 4-5. Data Acquisition System: plane tilt and orientation [86]

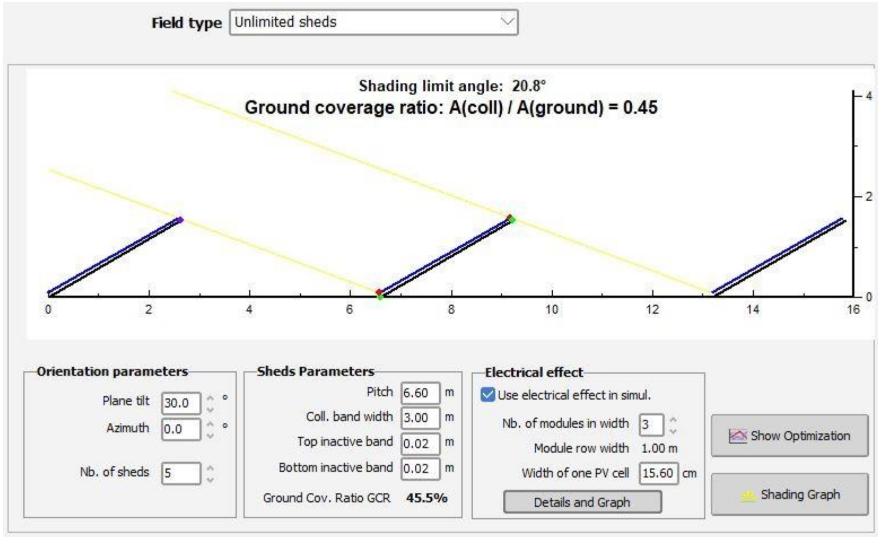
In Figure 4-5, given that the fixed tilt plane is defined as the ratio of incident irradiation (GlobInc) on the plane to the horizontal irradiation, the transposition factor (TF) of 1.08 is noteworthy (GlobHor). The system will achieve 99.97% of the production promised by the seller because the predicted loss at 30° tilt is only 0.3%.

4.6.2 *Estimation of Mutual Shading Factor for Grid Tilt Orientation*

When determining the ideal tilt orientation of PV panels in a grid, the mutual shading factor (MSF) must be taken into account. The shaded PV cell area to the total PV cell area in the grid is the MSF. The distance between PV panels, the height of the panel above the ground, and the direction of the panels can all have an impact on the MSF. PV panels lose efficiency and power production when they are packed closely together because the MSF increases. During the hours of greatest sunshine, this shading effect can result in substantial losses in PV power generation. For instance, research done in Spain discovered that a high MSF caused a reduction in a PV system's energy output of up to 11%, while a study done in India recorded a drop of 20% [112]. To maximize the efficiency of the PV panels, it is important to estimate the MSF when designing and constructing a successful and affordable PV system [87], [88], [89], [90]. Figure 4-6a illustrates how the MSF increases with PV installation congestion and causes PV losses independent of the meteorological conditions. The vertical axis in Figure 4-6 represents the MSF due to the geometry of the array. As the PV installation congestion increases, the MSF also increases.



a. *Mutual Shading Factor for 30° w.r.t sun-height profile angle*



b. Mutual Shading Diagram and String Arrays Offset for 99.97% Output as per Section 4.5

Figure 4-6. (a) Mutual Shading Factor for 30ᵒ w.r.t sun-height profile angle, and (b) Mutual Shading Diagram and String Arrays Offset for 99.97% Output as per Section 4.4

Computer simulations and modelling are ways to estimate MSF. The size and orientation of the panels, the shadowing patterns created by neighboring buildings or trees, and the impact of varied light angles and locations during the day and year should all be taken into consideration in these models. It is feasible to determine the best tilt angle for the PV panels in a certain area by examining the shade patterns and MSF for various tilt orientations. This can increase the PV system's energy output, increase its effectiveness, and lower the overall cost of energy generation.

Based on the calculations presented in Section 4.6, Figure 4-6 shows that a maximum value of MSF of 0.8 was obtained with a pitch of 6.6 m for a PV cell width of 15.60 cm. This method of design resulted in an installation with a Ground Coverage Ratio of 45.5%, which is extremely space-efficient.

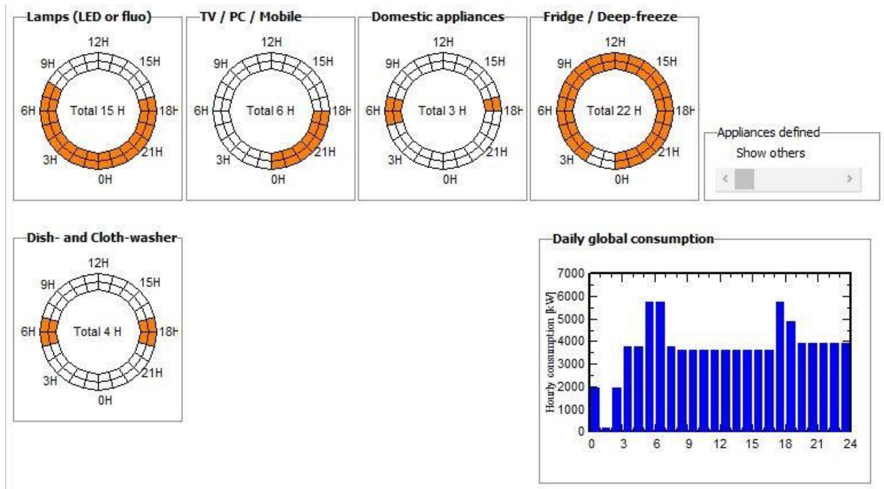
4.7 Load Profile, Distribution and Scheduling

The influence on PV inverter capacity and storage configuration is illustrated by the load profile, scheduling, and time-based distribution based on the given distribution of consumers. To ensure a PV system performs at its best, the load profile, scheduling, and time-based power distribution are crucial elements that must be taken into account [90], [91], [92], [93], [94].

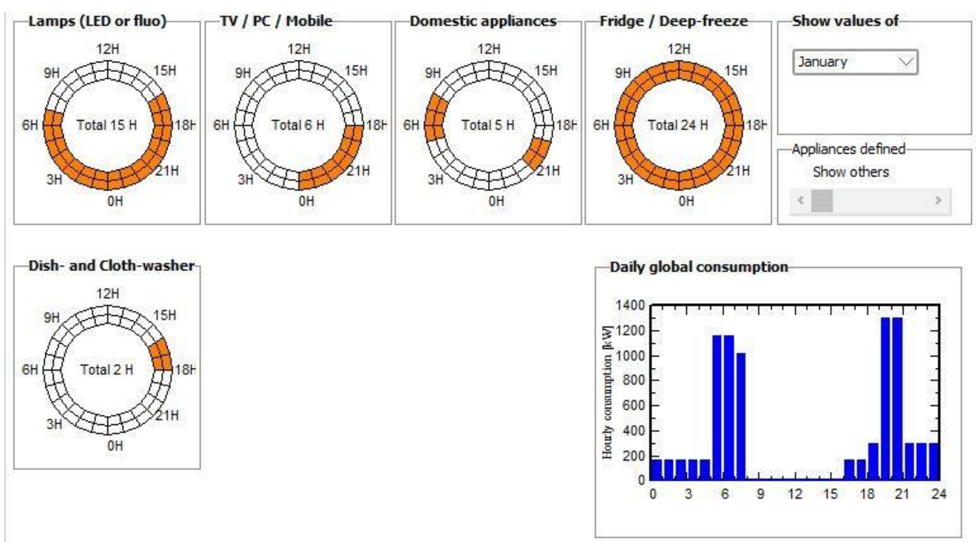
Here, "the compound of references above" refers to a fictitious collection of residential structures, which may comprise homes, flats, or other sorts of abodes. The compound's pattern of power usage over time is referred to as the load profile. For instance, the load profile may reveal that throughout the day and night, when fewer people are awake and/or out at work, respectively, power use is lower. The timing of certain tasks that need power within the building, such as operating appliances like washing machines or recharging electric cars, is referred to as scheduling. The total load on the PV system can be decreased by timing these tasks for periods when there is less demand for power. The distribution of the PV system's electricity throughout the course of the day according to time is known as time-based power distribution. In order to use the extra power produced during periods of low demand, it may be necessary to store solar energy in batteries or other storage devices. The PV system may be able to supply the compound's electrical demands throughout the entire day by reducing dependency on external sources of power and by optimizing the time-based power distribution. Overall, to ensure the best operation of a PV system inside a given compound, taking into account the load profile, scheduling, and time-based power distribution is vital.

The target location's energy consumption pattern is determined by the load profile, so the scheduling of the load profile will be extremely important in the design of a PV system. The type and quantity of appliances used, as well as the number of inhabitants and their lifestyles, will all affect the load profile. It is crucial to study the load profile and plan the energy distribution accordingly in order to create an ideal PV system. The capacity of the PV inverter and the storage arrangement may be established by examining the load profile and scheduling of the energy distribution. The storage system should be created to store the extra energy produced during off-peak hours, and the PV inverter capacity should be adequate to manage the peak load. This makes it easier to provide a consistent and dependable power supply for the intended area.

Figure 4-7a, a detailed daily consumer load profile, scheduling and distribution for January (part of the dataset). Early mornings and late evenings exhibit peaks of up to 1150 kW and 1370 kW, respectively. The month of January is part of the dataset, not the entire dataset, which is also shown in Figure 4-7b.



a. Housing Compound Consumer Set Rules Load Profile, Scheduling and Daily Distribution



b. Housing Compound Consumer Set Rules Load Profile, Scheduling and Distribution for January

Figure 4-7. (a) Housing Compound Consumer Set Rules Load Profile, Scheduling and Daily Distribution, and (b) Housing Compound Consumer Set Rules Load Profile, Scheduling and Distribution for January

The next sections are based totally on the requirements set out in this section, particularly Figure 4-7.

4.8 PV Grid Specifications and OEM (Model/Thermal) Constraint Parameters

Based on the irradiance levels and other parameters for this site, the 265W-60-Cell PV panel was selected, based on the detailed specifications presented in Figure 4-8.

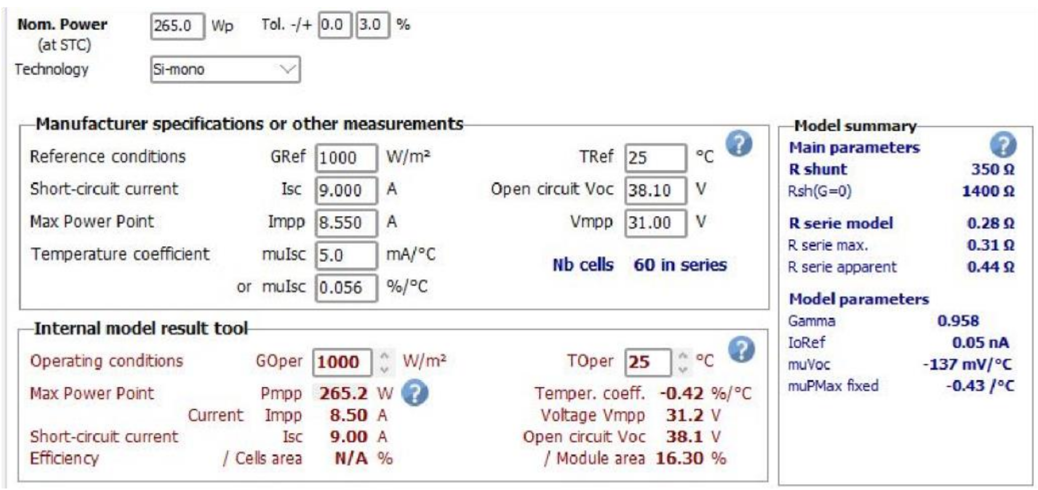
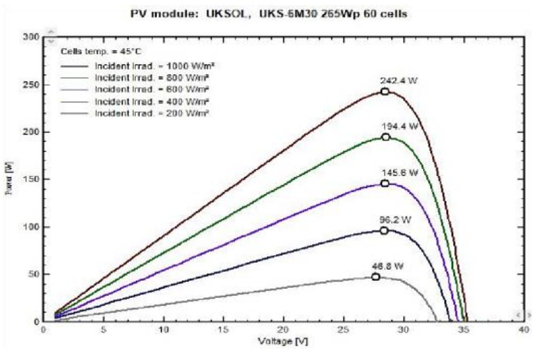
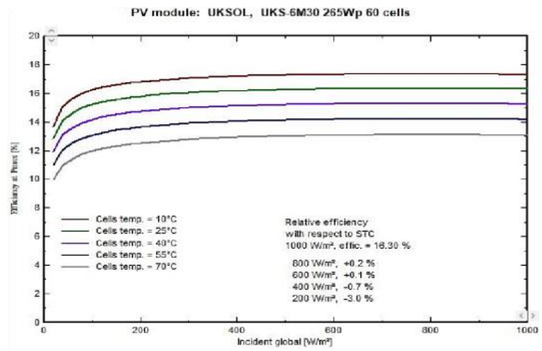


Figure 4-8. Specifications of 265W-60-Cell PV Panels

The detailed impedance model estimations are presented in Figure 4-9 as module output profile and thermal profile at given bounded value conditions. In our case, we are using a Mono-PerC Bificial Half-Cut, 265 W 60 cell module [95].

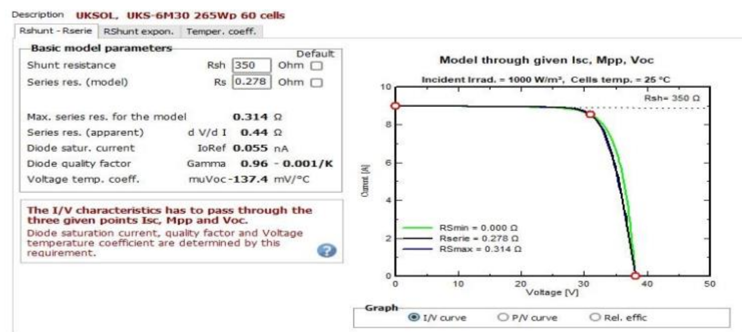
The estimations presented in Figure 4-6 and Figure 4-7 are based on the Meteonorm, Impedance Model for a PV Panel and the semiconductor theory associated with PV cell technology.



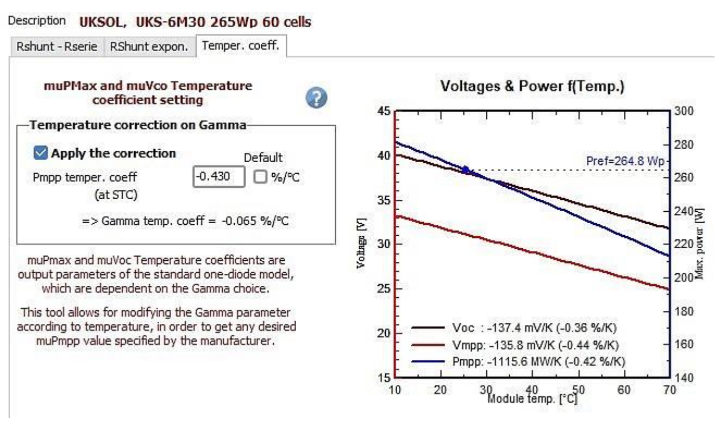


a. 265W-60Cell Module Output Profile (W/V)

b. 265W-60Cell Module Thermal Profile (W/V)



c. PV Panel Impedance Model and I/V Curve at 25°C



d. Temperature Coefficient Setting based on the Meteonorm 7 Geospatial Site Specifications

Figure 4-9. PV Specifications and Impedance Model Details for Meteonorm 7 Geodata

4.9 Storage System (Battery Specifications, Profiling, and Sizing)

The estimations for the battery system are painstakingly built atop the basic pillars of the Load Profile, Distribution, and Scheduling research presented in this section. This thorough investigation, which is illustrated in full, provides a thorough perspective of the behavior and performance dynamics of the battery system and sets the necessary foundation for the forthcoming insights revealed in Figure 4-10 and Figure 4-11. The Load Profile explores the complex patterns of energy use, illuminating the temporal fluctuations in power demand across time. The battery system's functionalities are assessed against the backdrop of this thorough investigation. In addition, Distribution analysis offers a more thorough insight into how energy requirements are spread throughout the specified time periods. It explores the subtleties of energy consumption patterns, offering priceless insights into load changes and trends. The Scheduling component also adds a strategic dimension to energy use by defining when and how energy will be allocated to maximize effectiveness. The scheduling analysis optimally distributes energy resources to maximize performance by taking into account peak load times, low-demand periods, and other operational concerns. The succeeding battery system estimations shown in Figure 4-10 and Figure 4-11 are supported by the knowledge gained from the Load Profile, Distribution, and Scheduling investigations. These diagrams explain how the battery system behaves in relation to the load and energy distribution patterns. These factors work together to show how well the system can control energy swings, optimize charge and discharge cycles, and fit into the overall energy management plan.

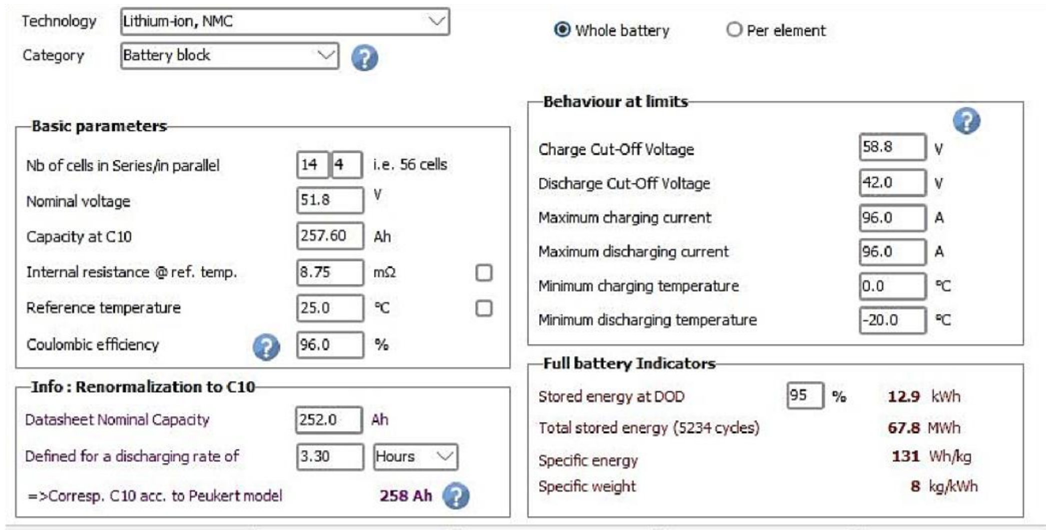


Figure 4-10. Battery String Topology and Configuration

The elements in Figure 4-10 work in concert to provide vital information and insights for choosing the configuration of the battery system. One can determine the dynamics of energy use, pinpoint crucial times, and optimally distribute energy reserves by examining the load profile, distribution patterns, and scheduling techniques. As a result, this thorough understanding directs the choice of battery string topology, ensuring that it is in line with load requirements, distribution subtleties, and energy scheduling goals. As a result, these components serve as key pillars in the complex process of coming up with an effective and efficient battery system arrangement.

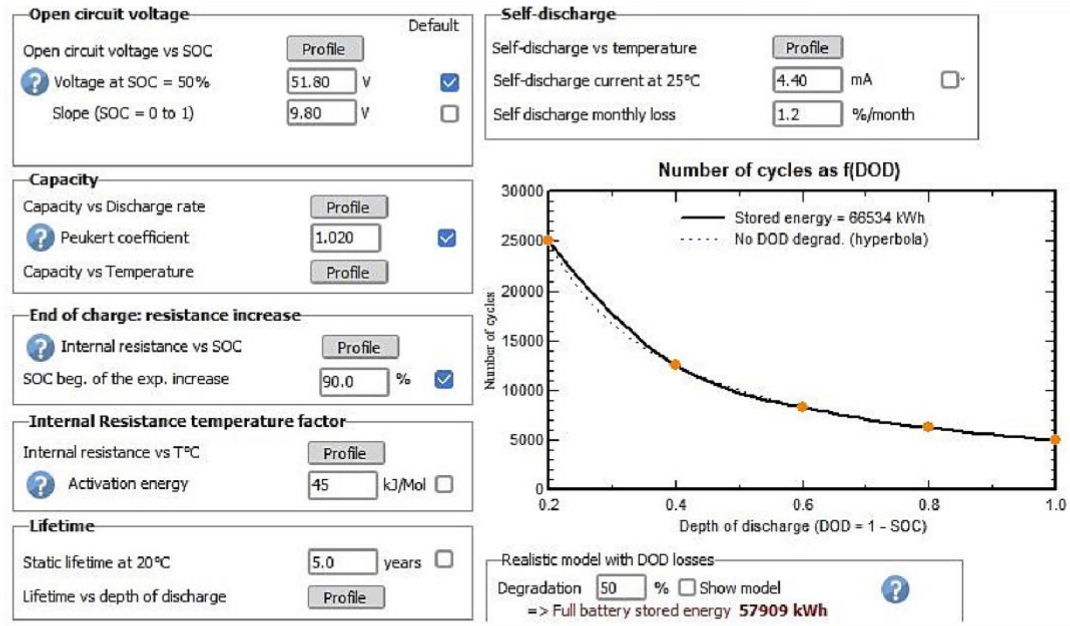


Figure 4-11. Battery Model for Electro-Chemical Thermo-dynamics (ECTD) Profile

In Figure 4-11, A specialized framework known as the Battery Model for Electro-Chemical Thermo-dynamics (ECTD) Profile is intended to fully capture and simulate the complex electrochemical processes and thermodynamic behaviors that occur within a battery system. The interplay of chemical processes, temperature fluctuations, and energy transfer dynamics is taken into account as this model explores the underlying principles driving the energy storage and release mechanisms. The battery model for the ECTD Profile makes use of sophisticated computational methods and mathematical formulas to replicate the electrochemical processes taking place inside battery cells. The performance properties of the battery, including capacity, voltage profiles, and efficiency, are predicted by taking into account elements like the electrode materials, electrolyte composition, and temperature conditions. The model investigates how temperature variations affect the battery's behavior, effectiveness, and longevity by taking thermodynamic issues into account. This is essential because variations in temperature have a big impact on reaction rates, ion mobility, and overall energy transfer effectiveness.

For the battery string modelling and electro-chemical thermodynamic profile, the power delivery for different charging magnitudes that address the peak hours constraint in Sections 4.6 and 4.7 (Figure 4-6 and Figure 4-7) using a 51.8 V voltage channel for a string of 250 Ah batteries. The battery string delivered 53.4 kWh using a battery bank of 1.29 MAh for the given site with a charge density of 131 Wh/kg and total stored energy of 67.8 MWh, see Figure 4-12 and Figure 4-13.

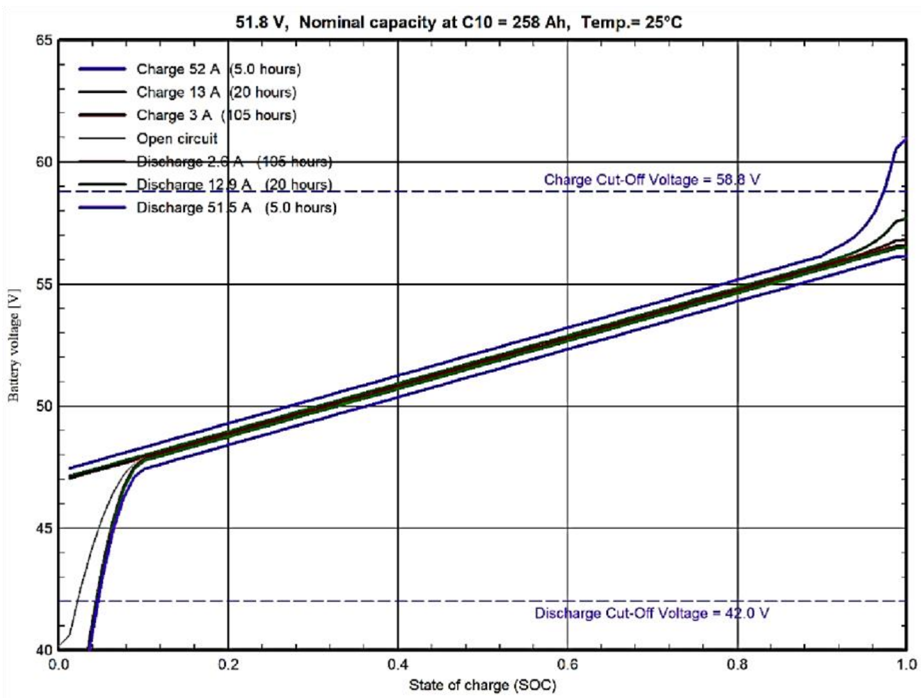


Figure 4-12. Power Delivery Factor (V/SoC) for 51.8 V String of 250 Ah Li-Ion Batteries

The Power Delivery Factor (V/SoC) for a 51.8 V string of 250 Ah Li-Ion batteries refers to the relationship between the voltage (V) and State of Charge (SoC) of the battery system. This factor quantifies how the battery voltage changes as the State of Charge varies. In this context, a 51.8 V string implies that multiple individual Li-Ion batteries, collectively arranged in a string, have a combined voltage of 51.8 V. The State of Charge indicates the amount of energy stored in the batteries as a percentage of their maximum capacity.

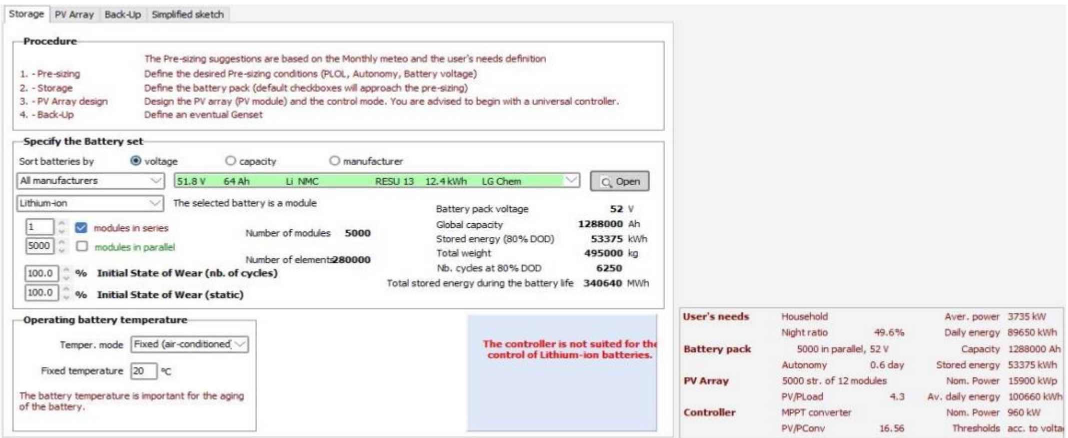


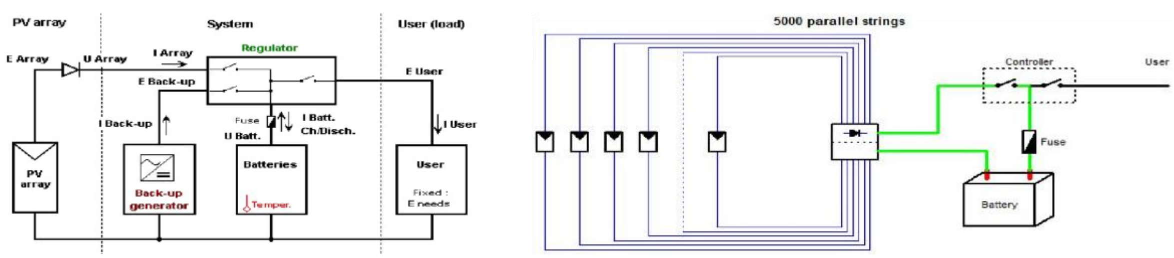
Figure 4-13. Data Acquisition System

The Power Delivery Factor (V/SoC) essentially indicates how the battery voltage changes as the battery is charged or discharged. It reflects the battery's voltage response to changes in its energy content. A higher V/SoC value implies that the battery's voltage varies significantly with changes in the State of Charge, while a lower value suggests a more stable voltage across different charge levels. Understanding the Power Delivery Factor is important for battery system design, management, and monitoring. It helps in predicting the battery's behavior under different load conditions and aids in optimizing the efficiency and performance of the battery system.

4.10 System Configuration/Layout

Figure 4-14 shows the configuration of the full PV system, the relay logic diagram (RLD) and a single line diagram of the power system. RLDs, which use common symbols to depict the connections between devices and their functions, are graphical representations of electrical logic control systems. They are used to demonstrate the order in which a control system operates and to aid engineers and technicians in comprehending the logic of the system. Single-line diagrams are graphical representations of a power system that highlight the linkages between its many parts. These generally depict the key power system elements, such as generators, transformers, switchgear, and safety equipment, together with their connections.

PV system layouts show the physical configuration of the system's numerous elements, including solar panels, inverters, batteries, and other electrical components. The architecture of a PV system is crucial in establishing the system's efficiency, safety, and adherence to regional building and electrical codes.



a. RLD of Developed PV Plant b. Single Line Diagram of Developed PV System

c. 3D Site Layout of the PV System

Figure 4-14. Complete PV System RLD and Layout for Commissioning

4.11 Nano-Grid Output Profile (Array Output / Charge Distribution)

This section presents the complete level data acquisition system and validation of the developed algorithms as summarized in Figure 4-15 to Figure 4-18.

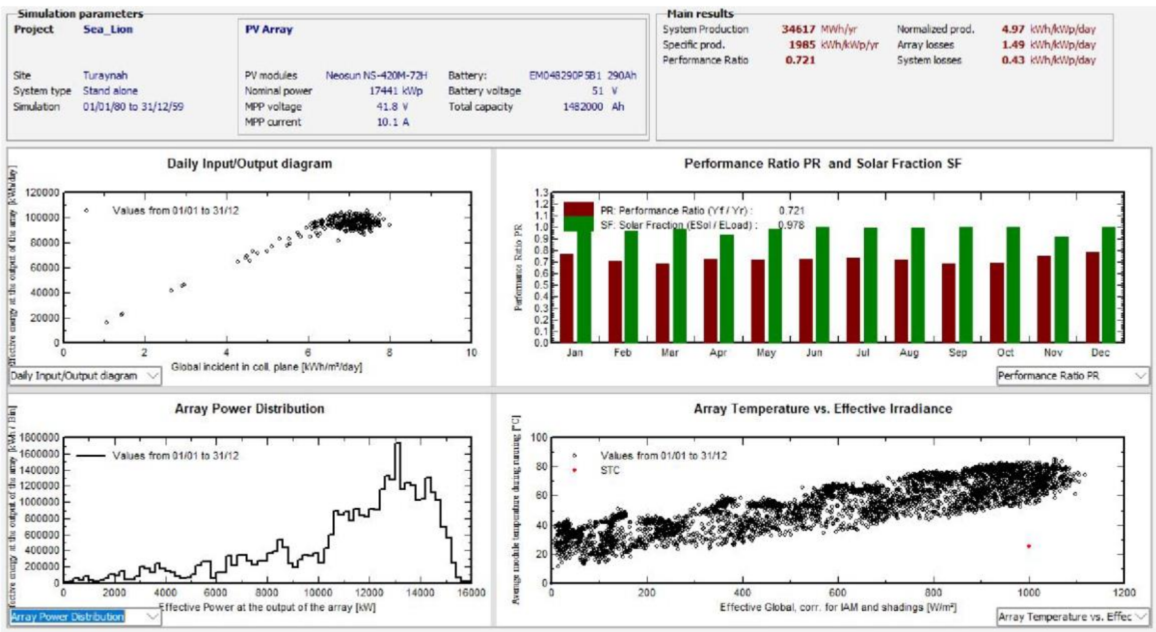


Figure 4-15. Data Acquisition System: (a) Daily Input/Output, (b) Performance Ratio and Solar Fractions, (c) Array Special Power Distribution, and (d) Array Temperature as a Function of Effective Irradiance

A plethora of insights gleaned from the complex computations and analyses produced earlier in this chapter are captured in Figure 4-15, which stands as a precise and thorough visualization. The input-output (I/O) diagram, PV system performance measurements, array spatial power distribution, and thermal efficiency concerns are just a few of the key components incorporated into this comprehensive presentation. This comprehensive viewpoint offers a clear comprehension of the complex dynamics dictating the PV system's behavior, effectiveness, and operational efficiency. The insights provided by Figure 4-15 are further enhanced by the addition of PV performance indicators, designated as Performance Ratio (PR) and Shading Factor (SF). Intricately measuring the system's capacity to produce energy in relation to its potential, these measures take shading and system losses into account. This graphic representation of PV performance shows how effectively the system converts solar irradiation into usable electrical energy.

The details shown in Figure 4-16 to Figure 4-18 significantly enhance the overall impact of the pieces seen in Figure 4-15.

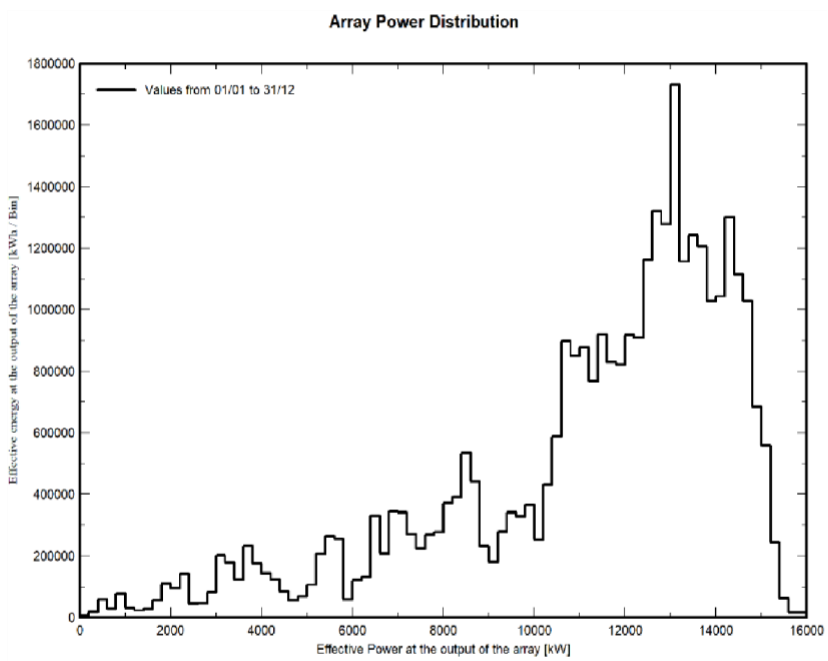


Figure 4-16. Estimated Annual Array Power Distribution of Developed PV System

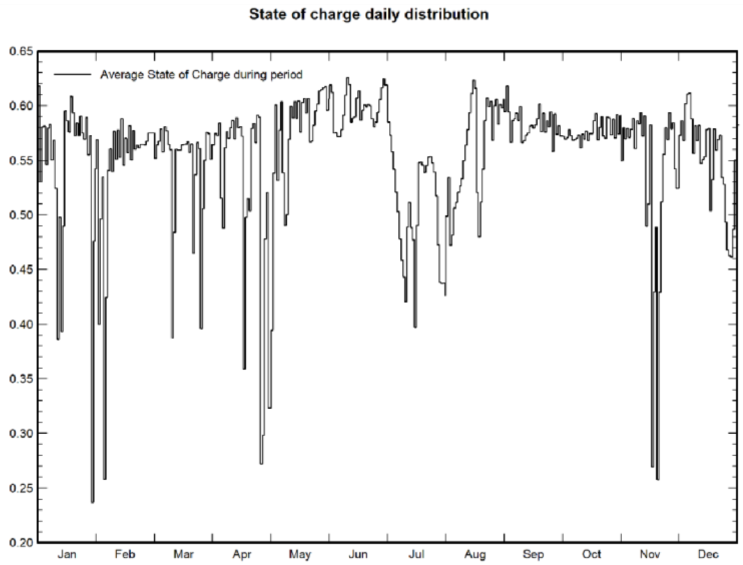


Figure 4-17. Estimated Annual Storage System Efficiency

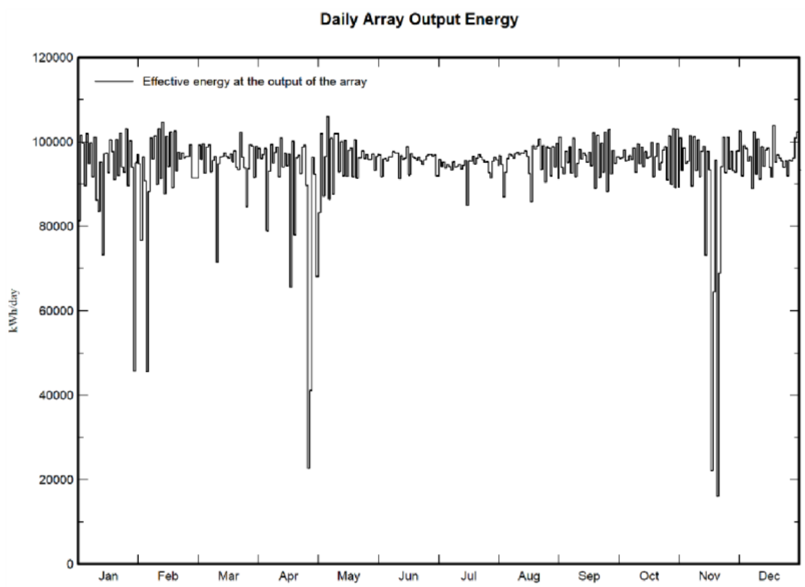


Figure 4-18. Daily Array Output Energy

These details provide a finer knowledge of how many parameters interact to create the behavior of the PV system by delving deeper into the performance intricacies of the system. In conclusion, Figure 4-15 represents the thorough analytical trip performed in this chapter. Its visualization encompasses the various facets of the behavior of the PV system, including energy flow, performance indicators, spatial distribution, and concerns for thermal efficiency. This thorough presentation attests to the methodical methodology and perceptive insights discovered throughout this study.

The term "Estimated Annual Array Power Distribution of Developed PV System" refers to the distribution of power produced over the course of a year by a PV solar array. This distribution may be affected by elements including the climate, the sun's angle, the amount of shade, and the panel's efficiency. The examination of energy generation patterns is prevalent in solar energy systems. The term "Estimated Annual Storage System Efficiency" refers to the effectiveness with which a system of energy storage (such as batteries) stores and releases energy over the course of a year. Temperature, battery deterioration, charge and discharge losses, and others may have an impact on its efficiency. Understanding the total performance of a renewable energy configuration requires calculating and evaluating the efficiency of the storage system.

4.12 System Loss Diagram

No system is perfect, and while applying the prediction algorithms, one has to consider losses to have a realistic prediction when validating at the system level. Figure 4-19 presents the annual loss diagram that is incorporated in the developed ML model for the prediction of power and energy from the given system.

It is necessary to account for losses due to, for example, variations in environmental conditions such as weather patterns and solar irradiation, as well as operational aspects like system effectiveness and maintenance considerations. The system's energy production may be predicted more precisely and realistically by including these considerations in the ML model output. Hence, an in-depth knowledge of these elements must be incorporated into the prediction algorithms in order to make sure that forecasts are as accurate as possible. A yearly loss diagram shown in Figure 4-19 has been put into the ML model prediction of the system's power and energy output. This loss diagram offers a thorough explanation of the possible losses the system may incur over the course of a year. For the purpose of forecasting power and energy from the specified system, the yearly loss diagram has been added to the constructed ML model. The loss diagram takes into account a number of variables that may have an impact on the system's ability to produce energy, including the weather, solar radiation, system efficiency, and maintenance issues. The kind and quality of the components used, the system's design and installation, the local weather patterns and sun irradiation, as well as the system's maintenance and operation all affect how efficient a renewable energy system is.

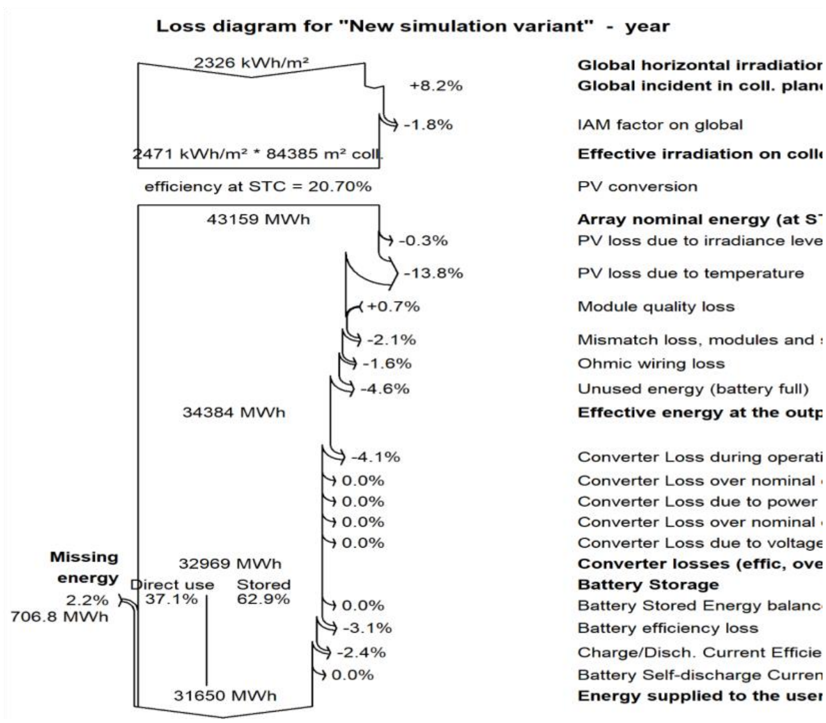


Figure 4-19. Detailed Annual Loss Diagram

4.13 Site Instrumentation and IoT-based Sensor Node for Data

A monitoring box was designed and developed and connected to the PV site system for remote monitoring for dual verification in addition to the Inverter App [96] used for precisely measuring the site variables for accurate prediction as presented in Figure 4-20.

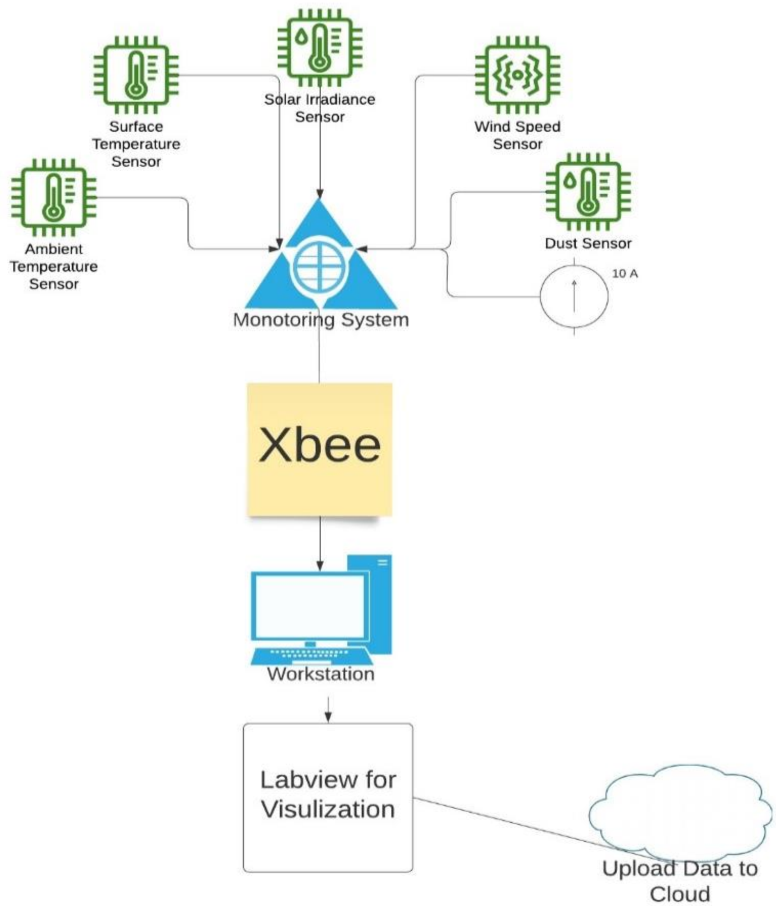


Figure 4-20. Data Acquisition System

The monitoring system has a number of transducers that are installed in relevant parts of the system to measure PV voltage, current, panel or environmental temperature, irradiance, wind speed and humidity. The monitoring box, by measuring these parameters can offer a thorough picture of the PV system's performance, which may be used to detect problems and increase system effectiveness. The positioning of the transducers is critical since the monitoring box must be able to produce accurate and reliable data. Throughout the system, transducers are positioned at key points, such as the input and output of the inverter, on certain solar panels, and other essential places. In order to provide reports on system performance, the data gathered by the monitoring box is sent to a remote monitoring system for analysis. Technicians may utilize this data to pinpoint problems and modify the system as necessary, enhancing overall performance and efficiency. The monitoring box is a crucial part of the PV site system because it enables thorough monitoring and analysis of system performance, both of which are vital for increasing efficiency and guaranteeing that the system is operating at its best.

All the results presented were produced using the following PC and software configuration:

Table 4-2 PC and Software Configuration

CPU	Intel Core i7-11600H 2.4GHz
RAM	16 GB
SOFTWARE	Google Colab

4.14 Purpose of Using Arduino

4.14.1 Prototyping

Arduino boards are an essential tool for electronics professionals who want to prototype quickly. They allow developers to test and fine-tune ideas before moving on with full-scale implementation. These boards provide an easy-to-use platform that makes it quicker to assemble and test electronic devices, and because of their adaptability and simplicity of usage, developers may quickly validate and iterate concepts before moving on to more sophisticated and thorough phases of development by experimenting with different configurations, sensor integrations, and code implementations [115]. The creativity and design process in electronics is greatly accelerated by this quick prototyping capacity.

4.14.2 Automation

Arduino boards are essential to automation because they allow a wide range of actuators, such as lights, motors, sensors, and more, to be controlled in response to real-time inputs or preset conditions [116]. Their capabilities cover a wide range of automation applications, from industrial installations controlling machinery and sensors to home automation systems controlling lighting and temperature. These boards use preprogrammed logic to analyse input signals or particular circumstances, which causes linked devices or systems to respond appropriately. Because of its programmability and ability to interface with a broad range of components, Arduino boards are an essential tool for developing automated solutions for a variety of applications and daily situations.

4.14.3 Data Collection

By virtue of its capacity to communicate with a wide range of sensors, Arduino makes it possible to collect data on a wide range of characteristics, including temperature, humidity, motion, and more [117]. This gathered data turns into a useful resource, providing information for analysis, judgement calls, and subsequent actions. Users are able to access and utilize recorded or real-time data by Arduino's capacity to connect and collect data from a variety of sensors, whether in research settings, IoT applications, or environmental monitoring [118]. Enlightened decisions, process optimization, and the development of more intelligent, responsive systems are all made possible by this data-driven methodology.

4.15 Sampling Requirements and Signal Conditioning

4.15.1 Sampling Requirements

Analog-to-digital converter (ADC) of Arduino is built-in and typically has a 10-bit resolution, which allows analogue signals to be converted into digital values [119]. External ADCs can be integrated to improve resolution for applications that require a higher level of precision [127]. Critical to signal acquisition, the sampling rate changes according to application requirements [130]. Most importantly, greater sampling rates are required to accurately capture and represent the signal in its dynamic state due to the quick changes in signals [131]. Finding the ideal sample frequency requires striking a balance as too high rates could result in unnecessary data volume, while too low rates run the risk of missing important signal variations [128]. In order to guarantee accurate data capture and optimize processing and storage resources in accordance with application requirements, sampling requirements must be calibrated.

4.15.2 Signal Conditioning

Signal conditioning is essential for maximizing the communication between the Arduino and external sensors since it protects the microcontroller from possible electrical risks and guarantees accurate and consistent readings. In order to accurately convert analogue data to digital data, Arduino needs to know how to manipulate the input signal [120].In signal conditioning, compatibility with various voltage levels is crucial. Since many sensors have different operating voltage ranges, attaching them directly to the pins of the Arduino may result in overvoltage or under voltage conditions that could harm the microcontroller. By helping to align these voltage levels, usually with the use of voltage dividers or level shifters, signal conditioning ensures that the Arduino gets signals that are within its permitted voltage range, safeguarding its integrity.

Another crucial element is noise suppression, where sensor signals can be distorted by outside influences like electromagnetic interference or background noise, which can lead to readings that are not correct. Filtering is one of the signal conditioning techniques that helps reduce or eliminate this interference [115]. By reducing undesirable frequencies, low-pass, high-pass, or band-pass filters improve measurement accuracy by lowering noise. Amplification or attenuation of the signal amplitude can also be a part of signal conditioning. This procedure makes sure that the dynamic range of the Arduino's ADC and the analogue signal range match. Amplification increases sensitivity and resolution by bringing weak signals into the ideal range of the ADC. On the other hand, attenuation keeps the ADC from being saturated by extremely high input voltages, preventing distortion and preserving accuracy. When working in high-voltage or possibly noisy environments, isolation is essential. To protect the Arduino from noise or voltage spikes that could damage the microcontroller, it is necessary to isolate it from external circuits using parts like optocouplers or galvanic isolation [122]. Additionally, when interacting with high-power or potentially dangerous systems, this isolation improves safety. To put it simply, signal conditioning serves as a link between various sensors and the Arduino, guaranteeing accuracy, reliability, and safety. Signal conditioning maximizes the quality and dependability of data gathered by the Arduino, improving the system's overall performance by addressing problems with voltage levels, noise, amplitude, and isolation.

4.15.3 Expansion and Justification

Accurate and trustworthy data collection requires signal conditioning, which makes sure that the Arduino correctly records these values within its ADC range in solar energy projects utilizing sensors to monitor voltage or current generated by photovoltaic panels, reducing electrical harm and improving precision [123]. Furthermore, some sensors may need to be conditioned in order to match signal levels or remove noise. For example, amplification may be required for temperature sensors with tiny voltage outputs in order to achieve accurate measurement. Precise sensor readings in the context of solar energy systems allow for more accurate prediction models, which enhances the forecasting of energy generation overall. Consequently, signal conditioning must be included in the sensor interface process in order to provide accurate and dependable data gathering, which is necessary for machine learning-based forecasting models to work.

4.16 Conclusions

The chapter demonstrated the seamless integration of ML-based forecasting approaches into the complex process of PV system design and analysis by turning theoretical claims into concrete actions. It essentially filled the gap between the theoretical level and actual application and demonstrated how to successfully incorporate ML-based forecasting approaches into a practical PV system design, efficiently converting research ideas into useful results. The chapter highlighted the potential of ML-based forecasting in forming practical solutions for using solar energy resources by outlining a thorough route from concept to execution.

5 Discussion of Technical Findings

Prior to delving into the discussion of the results, a comprehensive overview of the machine learning methods employed for prediction in this study will be provided. Each method brings distinct characteristics to the predictive analysis, allowing us to explore and interpret the behaviour patterns of the system effectively.

5.1 Machine Learning Methods

5.1.1 K-Nearest Neighbours (KNN)

K-Nearest Neighbours is a classification algorithm that operates based on the concept of similarity. It predicts the class of a data point by considering the classes of its "K" nearest neighbours in the training dataset. KNN is particularly suitable for capturing local patterns and is robust against noisy data. It does not assume underlying data distributions and can handle both binary and multiclass classification tasks. In this study, KNN was utilized to classify behaviour types based on the proximity of data points in the feature space.

5.1.2 Support Vector Machine (SVM)

A support Vector Machine is a powerful classification algorithm that identifies optimal decision boundaries between classes. By mapping the data into a higher-dimensional space, SVM aims to find the hyperplane that best separates different classes while maximizing the margin between them. SVM is effective for both linear and non-linear classification tasks and is particularly useful when dealing with complex data relationships. In our analysis, SVM was employed to classify behaviour types based on their feature representations.

5.1.3 Artificial Neural Networks (ANN)

Artificial Neural Networks are computational models inspired by the structure of the human brain. ANNs consist of interconnected nodes, or "neurons," organized into layers. Information flows through these layers, with each neuron processing inputs and passing its output to the next layer. ANN, a subset of deep learning, involves architectures with multiple hidden layers, enabling the model to capture intricate patterns in the data. ANNs excel in tasks that require complex feature extraction and representation learning. In our study, we harnessed ANN's capabilities to predict behaviour patterns based on the provided features.

The systems under consideration are explained in this chapter. Both the environmental elements used as inputs to the regression model and the data gathered and used to train and evaluate the models. Results from modelling with multiple linear regression and ANNs are shown and compared [97]. The computer and software combinations shown previously in Table 2 were used to create all of the findings that are displayed.

By introducing these machine learning methods and their respective characteristics, we establish a foundation for comprehending the subsequent discussion of results. The following discourse will delve into the outcomes yielded by each method, shedding light on their individual strengths and contributions in unravelling the system's behaviour complexities.

5.2 Data Description

In order to evaluate Qatar's PV power output during the almost two-year period between November 2014 and October 2016, data were gathered by Khandakar et al. [98]. The measured values of six environmental factors and the output power of the PV system were gathered and stored by a monitoring system. The regression model was then used to forecast the power output using these characteristics, which were afterwards monitored on a regular basis using calibrated sensors. The following environmental factors were chosen because they had the most impact on the electricity production of the PV panels:

- **Irradiance**: The solar irradiance incident on the PV panel's surface is measured in kW/m^2. The sun's brightness changes throughout the day, peaking at roughly 1 $kW/m^{2,}$ and is very location-sensitive. It is the most crucial element impacting the output power of PV panels. A SP110 solar irradiance sensor is used to calculate the incident irradiance in W/m^2.
- **PV surface temperature**: The surface of the PV panel is heated by solar radiation, which affects the panel's performance. To track this variable, in Celsius, a PT 100 temperature sensor was fastened behind (below) the PV panel. The sensors are located on the back of the PV module because this limits heat transmission prevents surface cooling, and produces the most accurate readings of the solar module's actual temperature while also protecting the sensors.
- **Air temperature**: The performance of electrical devices and components is greatly influenced by the ambient temperature. The ambient air temperature was measured in Celsius using an LM35DT analogue temperature sensor.
- **Relative humidity**: The lifespan and efficiency of PV panels are impacted by the amount of water vapor in the air. It lessens the amount of actual sunlight that hits the panel surface and is thought to be a factor that might affect the PV output power. The relative humidity of the ambient air around the panels is measured in % using a calibrated HSM-20G sensor.
- **Dust accumulation**: By preventing sunlight from shining directly on the solar panel surface, dust on the PV panel surface tends to reduce the output power of the panel. A Sharp photoelectric dust sensor, model number GP2Y1010AU0F, measured dust density on the face of the panel in milligrams/m^3.
- **Wind speed**: The wind speed has an indirect impact on PV panel performance since it can change some of the parameters mentioned above, such as dust, humidity, and ambient temperature. An

anemometer (also known as a vane anemometer, thermal anemometer, thermal anemometer with velocity/temperature profiling, or cup anemometer) measures wind speed in kilometers per hour.

The regression model produces the simulated PV output power, and voltage and current transducers are used to measure the PV panel's output voltage and current to record the power output. Afterwards, P-V curves are shown on LabView for further data processing.

Table 5-1 Range of On-site Environmental Parameters Between November 2014 and October 2016 Used for Predicting Power

Environmental Parameter	**Maximum**	**Minimum**	**Unit**
Temperature	61.00	14.63	degree Celsius
Relative Humidity	90.76	27.82	%
PV Surface Temperature	74.40	9.30	degree Celsius
Solar Irradiance	1033.5	38.01	w/m2
Dust Accumulation	1.1142	0.0553	mg/m^3 (or mg/m^2)
Wind Speed	34.24	0.5893	km/h
Power	114.2	0.0368	w

5.3 Data Acquisition

A data-accessing system using calibrated sensors and an inexpensive Arduino microcontroller housed in a monitoring box on a roof was a practical and affordable option. Data was sent from the controller to a workstation using wireless Xbee adapters. The LabView interface was used for additional data processing. Also, data was uploaded onto the cloud using the open Internet of Things (IoT) data platform ThingTalk to improve data accessibility. The block diagram shown previously in Figure 4-20 provides a summary of the data collection system. Assuring the box's stability and safety, shielding it from elements like rain and wind, and having access to a dependable power supply are all special concerns. To ensure the quality and dependability of the data acquired by the system, these aspects would need to be taken into consideration. Data transfer using wireless Xbee adapters made it possible to perform remote monitoring without having to physically reach the monitoring box. The block diagram illustrated in Figure 5-1 provides a comprehensive overview of the system's functional flow, encompassing the stages of training, validation, and operational use. This structured process ensures a well-rounded and systematic approach to harnessing the capabilities of the system.

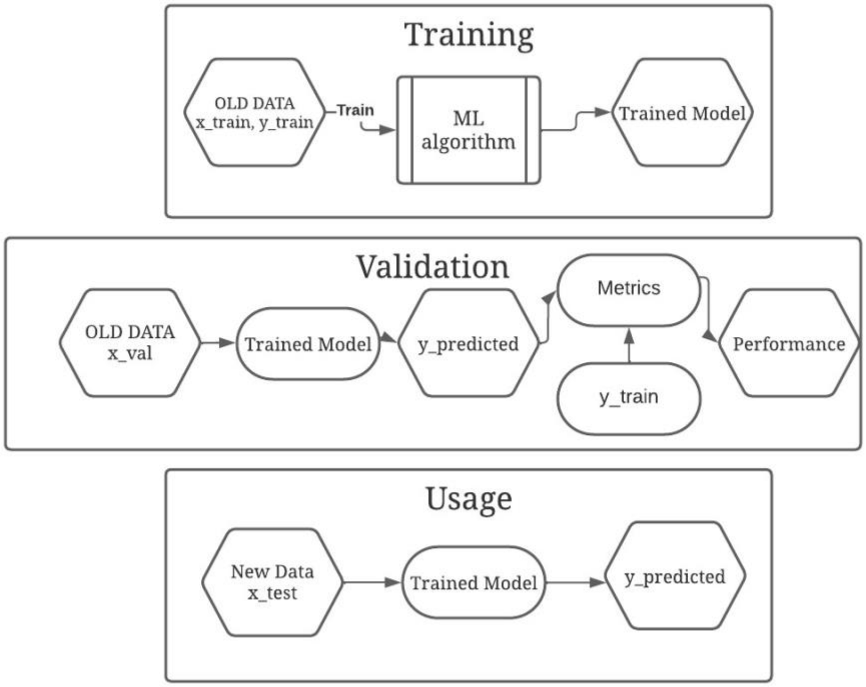


Figure 5-1. Machine Learning Pipeline

Table 5-2 Characteristics of the PV Panels Used

Module	**Poly-crystalline silicon**
Maximum Power	80 W
Area	0.6426 m^2
Open Circuit Voltage	21 V
Short Circuit Current	5.24 A

Uncertainty in how the environmental parameters impact the electrical system can make it difficult to design and operate the electrical grid. Some of the elements that cause solar power production to fluctuate include solar irradiance, ambient temperature, cloud cover, and surface temperature of the panels. The purpose of this study is to identify the most important factors affecting PV energy production, to identify relationships among the various parameters, and to produce more precise forecasts of energy output from PV panels based on these features. The data used is local data supplied by the Qatar Meteorological Centre and other government sources and have been pre-processed by the competent authorities before being obtained and exported to CSV datasets that contain the history of a PV module.

The third and the forth chapters various machine learning methods were described and applied in real life cases consecutively. The ideas have formed the basis upon which the entire concepts of solar power output forecasting through various regression methods are manifested and applied in the field. The concepts have also helped in this chapter in looking at irradiance, PV surface temperature, air temperature, relative humidity, dust accumulation and wind speed.

5.4 Dataset

Table 5-3 Dataset Supplied by Qatar University, Qatar

	Temperature	Relative humidity	PV temperature	Irradiance	Dust	Wind speed	Power
0	22.57	62.39	70.89	714.92	0.24	6.25	84.44
1	20.84	73.81	69.63	660.80	0.20	5.45	81.69
2	26.62	57.83	70.07	383.07	0.71	3.33	41.93
3	20.21	81.82	62.66	616.22	0.45	3.45	70.66
4	24.81	71.60	70.31	669.89	0.67	4.16	74.62
490	35.31	31.95	56.54	592.08	0.86	10.36	35.44
491	33.85	35.14	60.62	560.20	0.85	6.61	33.84
492	31.77	52.80	58.58	508.80	0.84	6.75	31.21
493	34.19	43.25	60.29	545.19	0.86	5.65	32.29
494	34.98	40.26	60.57	672.64	0.86	9.31	40.91

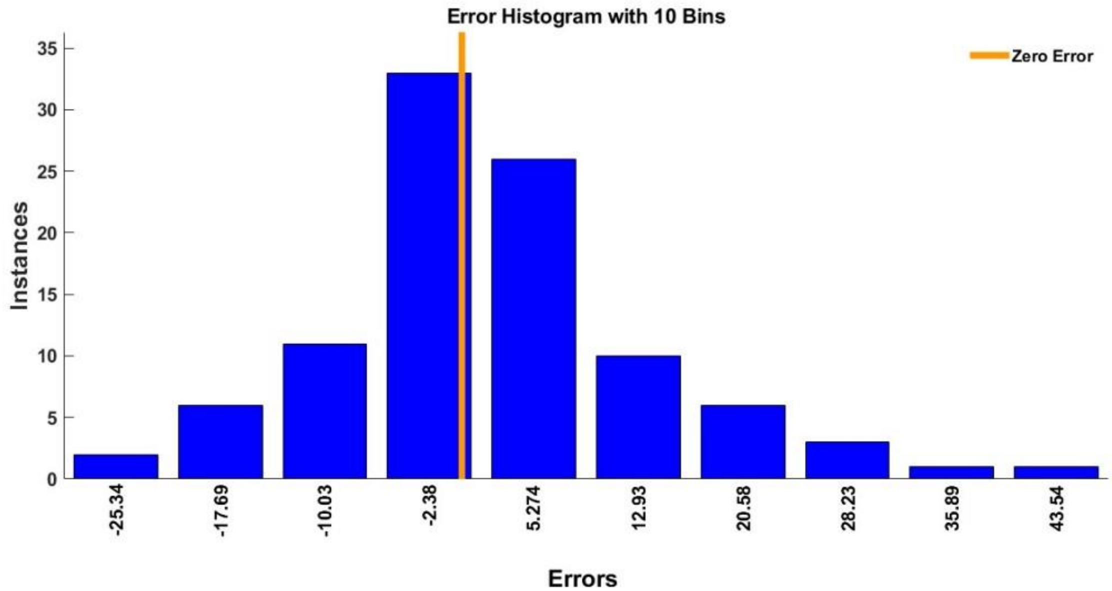
One of the most popular open-source, functional programming languages, Python, is utilized extensively for data analysis [99]. Developers may use Google Colab, a free cloud service, to execute Python scripts and perform data analysis. A variety of libraries, such as Pandas, that are available in Python for data analysis could be utilized. Pandas is a well-known Python data analysis toolkit that offers a simple and adaptable data structure for data analysis. Data from a variety of sources, including CSV files, Excel, SQL, and many more formats, may be loaded and analyzed using Pandas.

To perform data analysis, developers first need to load the data into Panda's data frames. Pandas provide various functions to load data from different sources, such as read_csv() for loading CSV files, read_excel() for loading Excel files, and read_sql() for loading data from SQL databases. Once the data is loaded, Pandas provides various functions to clean and analyse the data, such as dropna() for removing missing values, groupby() for grouping data, and many other functions. With the help of tools like Google Colab and libraries like Pandas, data analysis has become easier and more accessible than ever before. Developers can use these tools to quickly analyse data and extract meaningful insights, which can then be used for making informed decisions in various domains, including finance, healthcare, marketing, and many others [100], [101], [102], [103].

5.5 Results of Multiple Linear Regression Analysis

This section describes research that looks at the effectiveness of a model that uses multiple linear regression to forecast the output power of solar panels. Two metrics, R-squared (R^2) and root-mean-square error (RMSE), are used to assess the model's accuracy. The findings demonstrate that, with the majority of occurrences concentrated around the zero-error line, the projected output power is close to the actual output power. The study also shows that irradiance is the primary factor determining the energy produced and that the irradiance is positively correlated with the amount of energy produced by the PV panels. Overall, this study offers helpful information on the variables affecting PV panel performance and the precision of the multiple linear regression model that was used to estimate its output power.

Figure 5-2 shows an error histogram of 10 bins for the model multiple linear regression where the number of instances in each bin is shown on the vertical (ordinate) axis. The figure shows the zero-error line, and it can be seen that the instances are centered around it.



.

Figure 5-2. Error Histogram of the Multiple Linear Predictions

The predicted output power obtained using multiple linear regression, and the actual genuine power are contrasted in Figure 5-3. The anticipated result has a Test average RMSE of 1.57, while the training average RMSE is 2.04. The training average R^2 is 1.0.

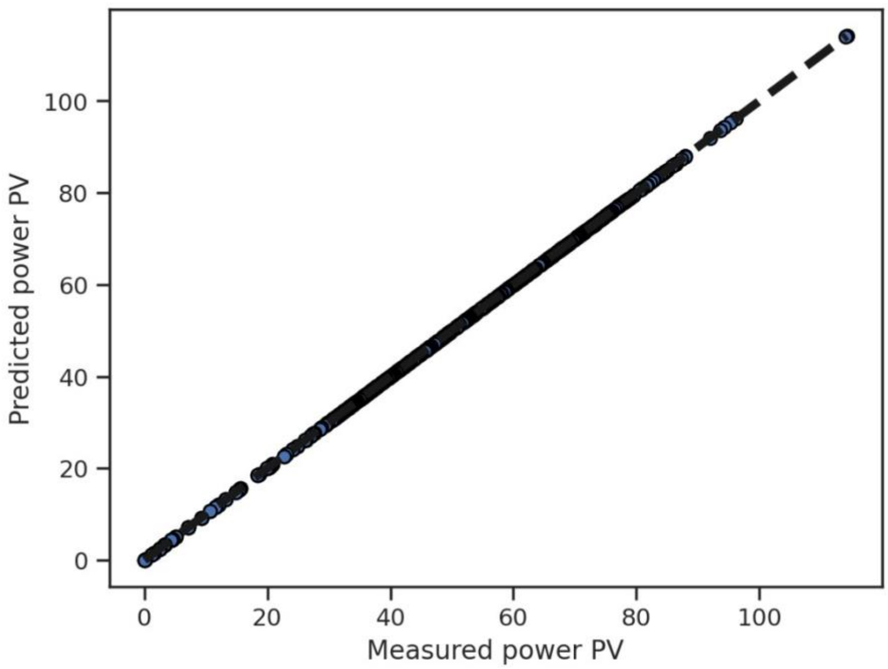


Figure 5-3. Multiple Linear Regression Predictions Vs Measured Values

Multiple linear regression is a statistical technique used to predict the value of a response variable based on multiple predictor variables. In this case, the model was used to predict the output power using a set of given variables. The results show that the model was able to produce accurate predictions, as indicated by the low test average RMSE value of 1.57. The lower the RMSE value, the more accurate the model is in predicting the output power.

The training average RMSE value of 2.04 suggests that the model is less accurate in predicting the output power during training. This may be due to overfitting, which is when a model becomes too complex and fits the training data too closely, resulting in reduced accuracy when applied to new data. However, the training average R^2 value of 1.0 indicates that the model fully explains the variability in the data, suggesting that overfitting may not be a concern in this case.

Overall, the multiple linear regression model appears to be a reliable and robust approach for predicting the output power based on the given variables. The results suggest that the model has the potential to be used for practical applications in the field of solar power prediction. However, further evaluation and validation may be necessary to determine the generalizability and reliability of the model across different datasets and contexts.

Irradiance and the energy generated by the PV panels are directly and positively correlated, as shown in Figure 5-4. It can be seen that irradiance is the most influential factor in the energy produced. This makes sense because the amount of sunlight that falls on the panels is the primary source of energy that drives the electricity generation process; the relationship between energy and irradiance is direct, meaning that the higher the irradiance, the greater the energy produced. Other factors, such as temperature, shading, and panel orientation, can also affect the energy output of PV panels, but irradiance is the most critical factor. Figure 5-5 emphasizes the importance of irradiance as the primary driver of energy generation in PV systems and suggests that maximizing the amount of sunlight that falls on the panels is crucial for optimizing their energy output.

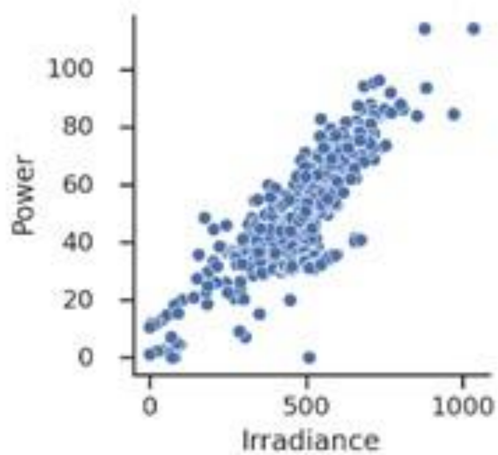


Figure 5-4. PV Power Output Watts Vs Irradiance W/m²

The link between wind speed and energy generation is weakly inverse, that the stronger the wind speed, the lower the solar energy generated). Wind speed in Qatar is typically higher in winter and autumn seasons, and the air is colder than in summer. From Figure 5-5, it appears that the wind speed was below 20 m/s for the overwhelming majority of the time, and in this range, the R^2 value indicated that under these conditions, any significant variation in power output depended on other conditions. On the few occasions when wind speed was more than 25 m/s, typically in winter conditions with a lower air temperature, the power output was relatively low, suggesting that with the natural increase of wind speed, energy output decreases slightly, hence the weak inverse relation. Note that when the speed was less than 20 m/s, the irradiance ranged between 350 and 800 W/m^2.

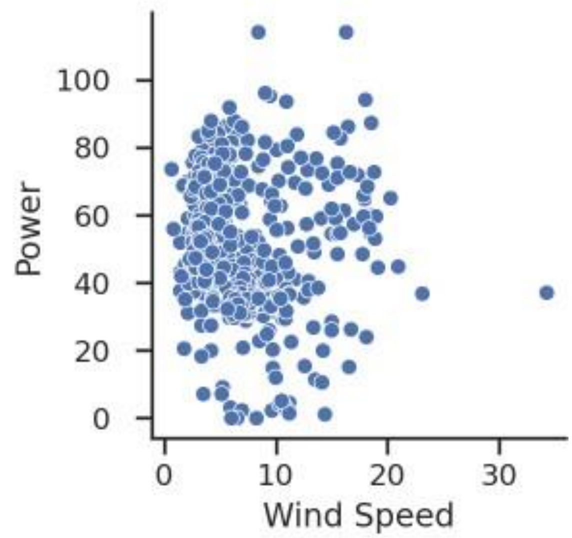


Figure 5-5. The relationship of wind speed m/s to power Watts

The best fit straight line for the plot of electrical power output vs. level of dust, see Figure 5-6, shows a small negative link between dust and solar power. When dust is less than 0.7 mg/m^3, power may reach up to 80 W; however, when dust increases more than 75 mg/m^3, power will tend to be closer to 50 W. Dust on the face of a PV panel would be expected to reduce irradiation level, but the small value of R^2 could be explained by the range of values used to represent the presence of dust.

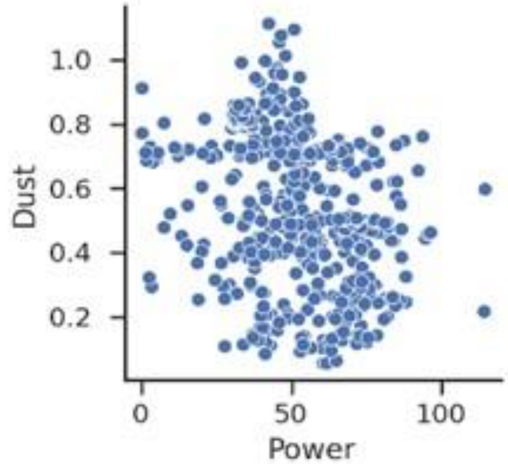


Figure 5-6. Relationship between Electrical Power Output Watts and Presence of Dust mg/m^3

Figure 5-7 is a plot of solar output power as a function of relative humidity. However, in general, relative humidity is not considered a critical factor in PV electrical power generation.

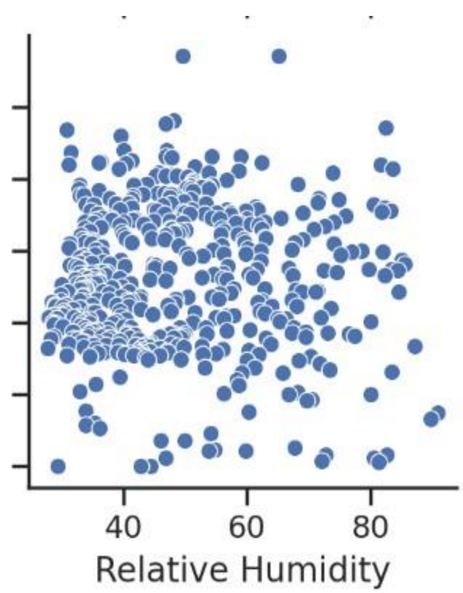


Figure 5-7. Relationship between Electrical Power Output Watts and Relative Humidity RH%

Nevertheless, if the value of relative humidity is high enough, there could be the formation of water droplets on the surface of the panels, and that might lower the quantity of sunlight that reaches the panel, hence reducing energy production. In summary, while relative humidity may have some impact on the energy output of PV panels, it does not have a significant effect unless dew formation takes place, i.e., the surrounding air drops to a temperature below its dew point, and this is not a frequent occurrence in Qatar once the sun has risen. However, any dew formation overnight can be reduced, or even eliminated, by routinely maintaining and cleaning the panels. The primary drivers of energy generation in PV systems are the intensity of sunlight and the efficiency of the panels.

Figure 5-8 is a plot of output power as a function of ambient temperature. As the temperature increases, the efficiency of the semiconductor material decreases, which can reduce the amount of energy that the panel can generate. The temperature coefficient of the panel represents the rate at which the electrical output decreases as the temperature rises.

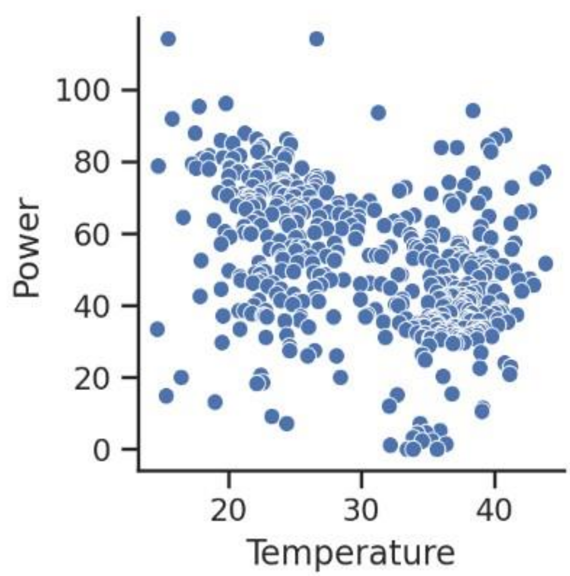


Figure 5-8. Relationship between Electrical Power Output Watts and Ambient Temperature ^oC

Regarding the negative correlation between ambient temperature and power generated, this could be important, particularly in very hot environments. However, the reduction in energy output due to an increase in ambient temperature is not likely to be significant under normal operating conditions. In cooler environments, such as during winter winds, the reduced temperature can actually increase the efficiency of the panel, resulting in higher energy output. In summary, ambient temperature is an important factor in PV electrical power generation. A best-fit straight line can be used to illustrate the relationship between temperature and electrical power, with the R^2 value and correlation coefficient providing a measure of the goodness of fit and strength of correlation, respectively. While there may be a small negative correlation between temperature and power generated in some situations, the effect is generally not significant under normal operating conditions.

Figure 5-9 is a plot of output power as a function of the temperature of the solar panel. There is a small negative link between panel temperature and electrical power. However, the small value of R^2 could be explained by the range of values used to represent the panel temperature. Possibly, even the highest panel temperature considered was not sufficient to have a significantly detrimental effect on electrical output.

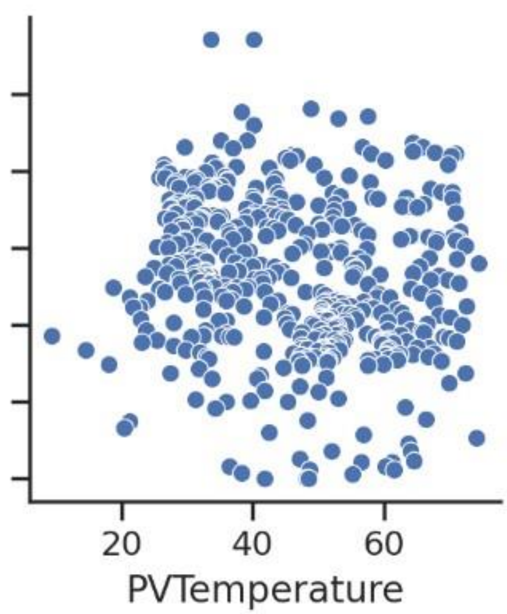


Figure 5-9. Relationship between Electrical Power Output and PVTemperature

The distribution of R^2 and correlation coefficients between features and their effect on energy is small except for irradiance and ambient temperature, but the factor most directly affecting energy is irradiance.

There is a small negative link between the temperature of the solar panel and the electrical power generated. This is due to the temperature coefficient, which represents the decrease in efficiency of the semiconductor material as the temperature increases. However, it is noted that the small value of R^2, which represents the goodness of fit of the line, could be due to the limited range of values used to represent panel temperature in the analysis. It is possible that even the highest panel temperature considered was not sufficient to have a significant detrimental effect on electrical output.

The values of R^2 and correlation coefficients demonstrate that the effects of the given parameters on solar energy are small, except for irradiance. This confirms that irradiance is the factor most directly affecting energy generated by the PV panel. Other factors, such as shading, panel orientation, and wind, may have some impact on the energy output of PV panels, but their effect is relatively small compared to the influence of irradiance, though temperature does have an effect.

In Figure 5-10, the relationships between the features and their influence on each other are shown. Save for the relation between irradiance and power, the values of R^2 indicate that the different parameters have little or no direct effect on each other. For training, validation, and test data sets, the model has a small and unbiased discrepancy between the actual, observed and predicted values. We examine the performance of an ML model and quantify the performance of the models using RMSE.

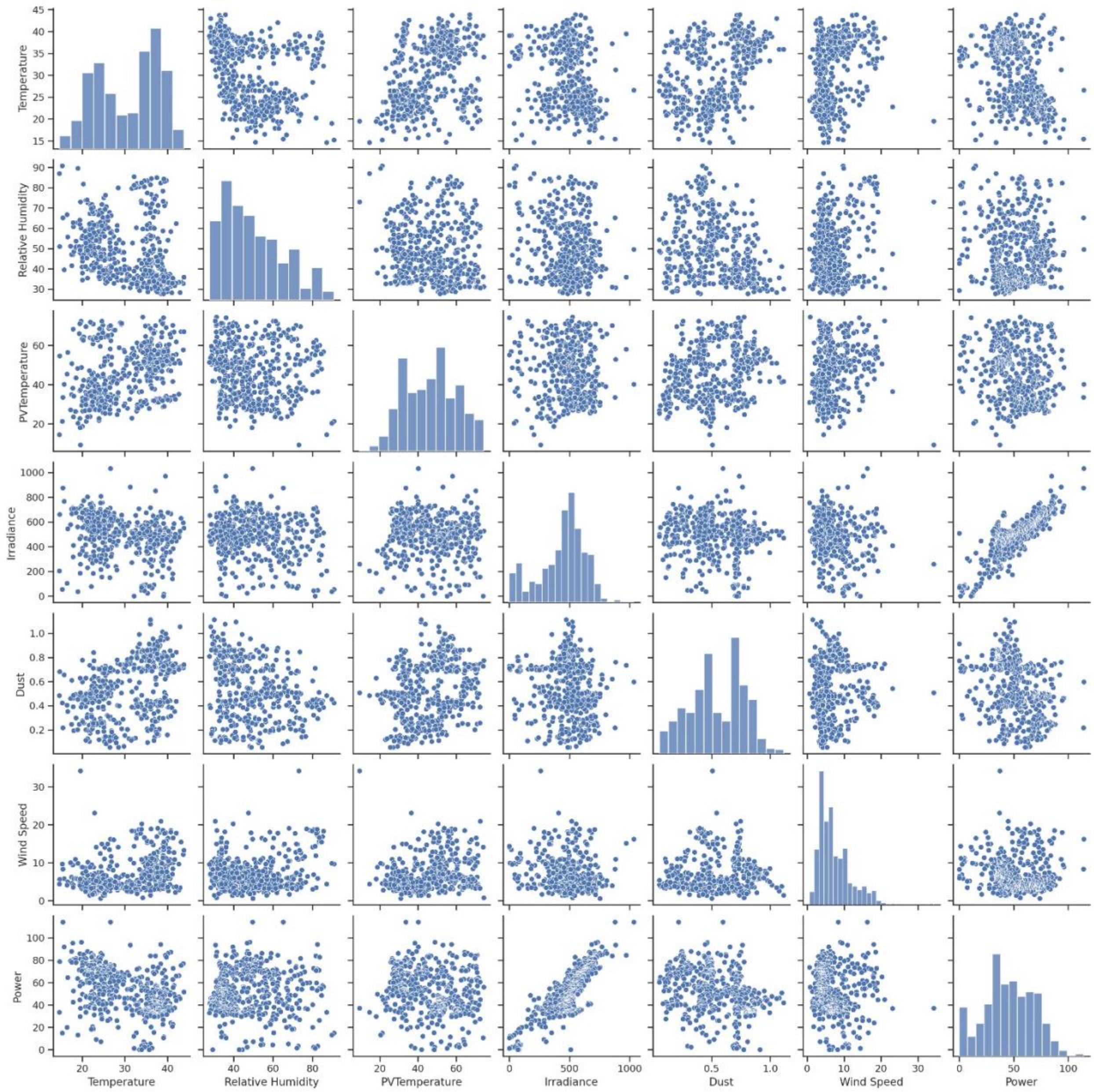


Figure 5-10. Predicted Results

The results of the study have significant ramifications for solar energy system design and operation. The performance and efficiency of solar energy systems may be optimized by maximizing the available irradiance by system designers and operators who are aware of the direct correlation between irradiance and energy generation. This may be accomplished by taking steps such as choosing the best sites for solar panels, monitoring the path of the sun to position panels for optimum irradiance, and routinely cleaning and maintaining solar panels to keep them clear of impediments. The results of the study can also be used to construct forecasting and prediction models for solar energy production that are more precise. It should be able to enhance the precision and dependability of forecasts by including irradiance data in these models, which would eventually aid in improving the effectiveness and performance of solar energy systems.

The study's validation of the causal link between irradiance and energy production offers insightful information about how solar energy systems operate. Researchers and industry experts may use this knowledge to design more effective and efficient solar energy systems, ultimately aiding in the transition to a future with more renewable and sustainable energy sources.

Table 5 lists the ambient temperature, relative humidity, PV temperature, irradiance, dust, and wind speed as the six environmental parameters that most affect the output power generated by solar energy. The top row of the table lists the total number of observations in each case. Each variable's mean value is shown in the second row. The third row displays each variable's standard deviation, which is a measure of how the data varies around the mean. The fourth row of the table displays the minimum value for each variable. The next three rows display the 25th, 50th (median), and 75th percentile values for each variable. The final row shows the maximum value of each variable. The information is presented for a better understanding of the distribution and range of the different variables, as well as to identify any potential outliers or unusual observations. This information is essential in developing accurate forecasting and prediction models for solar energy generation.

Table 5-4 Statistical Analysis of the Six Environmental Parameters Affecting PV Output Power

	Ambient Temperature (ºC)	Relative Humidity (%)	PV Temperature (ºC)	Irradiance (W/m²)	Dust (mg/m³)	Wind Speed (m/s)	Power (W)
No. of Observations	495.00	495.00	495.00	495.00	495.00	495.00	495.00
MEAN	30.49	48.29	46.25	492.6	0.5433	6.958	51.35
STD	7.30	14.63	13.47	159.46	0.2347	4.31	19.72
MIN	14.63	27.73	9.30	0.000	0.0553	0.5893	0.0069
25%	23.98	35.70	34.05	430.4	0.3885	3.84	37.04
50%	31.65	45.21	47.26	504.9	0.5096	5.80	50.73
75%	37.12	57.34	56.40	592.6	0.7316	8.87	66.87
MAX	43.84	90.76	74.39	1033.5	1.111	34.25	114.2

The six environmental factors (ambient temperature, relative humidity, PV temperature, irradiance, dust, and wind speed) given in Table 5 are all included in each of the models used below. This makes sure that the effects of each variable are appropriately taken into consideration and that the final model is as precise as feasible. The resultant model would not accurately depict the complicated interactions between the many environmental elements and their effects on solar energy output if any of these variables were omitted. Therefore, when creating predictive models for solar energy generation, it is essential to include all pertinent variables.

5.6 Results of the K-nearest Neighbours Model

The number of neighbors, K, and the distance metric are the two hyper parameters for the KNN model that need to be modified and chosen, as previously described:

- K, the number of neighbors to be considered has a great effect on the results of the regression. Choosing a low K value means that the prediction depends only on the nearest point from the training set, which is very risky and leads to high values of variance in the output, which emulates the case of overfitting in parametric ML models. However, choosing a high K value results in flat-biased predictions, which emulate the case of under fitting. Values of K between 1 and 100 were tested to find the most suitable value.
- The distance metric is the mathematical representation of the distance between two points, which is used to compare the closeness of the training input to the test input. Three main distance equations are used for the KNN model:
  - The Manhattan distance. Using this metric, the distance between two n-dimensional points x and y that correspond to two inputs of n features is defined as D

$$D(x, y) = \sum_{1}^{n} |x_i - y_i| \tag{5-1}$$

  - The Euclidean distance. The conventional definition of distance is

$$D(x, y) = \sqrt{\sum_{1}^{n} (x_i - y_i)^2} \tag{5-2}$$

- The Minkowski distance. A generalized form of the previous two definitions. The Minkowski distance of order p is defined as

$$D(x, y) = \sqrt[p]{\sum_{1}^{n} |x_i - y_i|^p} \tag{5-3}$$

It is worth noting that the Minkowski distance with order 1 is the Manhattan distance, and with order 2 is the Euclidean distance.

The tested metrics were the Manhattan, Euclidean and Minkowski distances with orders from 3 to 10. In addition to the former two hyper parameters, another factor which can be selected is the degree of effect each of the K neighbors has on the prediction of the output. If the output of all K neighbors has the same effect, it is called a uniform KNN. However, if the effect is based on the distance between the training point and the test point, then a weighted KNN is used. Both uniform and weighted models were tested. The weights of each of the K training inputs are defined as the inverse of the distance between this input and the test input

$$w_j = \frac{1}{D(x_{j.xtest})}; 1 \leq j \leq k \tag{5-4}$$

- and the final output is defined as

$$\hat{Y} = \frac{\sum_{J=1}^{K} W_J Y_J}{\sum_{J=1}^{K} W_J} \tag{5-5}$$

- Where Y_j is the training output of each of the K neighbors.

In this test, 2000 models were examined, including both uniform and weighted models, with K ranging from 1 to 100 and the degree of the distance metric ranging from 1 to 10. A cross-validation approach was used to test each model five times with five-fold, and the mean results were collected for each model for each test. Results of mean square error (MSE) and (R^2). Due to the large number of evaluated models, only the best K value for each combination of distance metric and model type is displayed.

Upon evaluating the outcomes of the study, it was discerned that among the various models tested, the weighted model exhibited the highest level of effectiveness. Specifically, when utilizing a value of K equal to 9 and employing the Manhattan distance as the measurement metric, this particular configuration yielded the most favourable results. In essence, this model's predictive performance outshined others, suggesting that its parameter settings, particularly the chosen value of K and the distance metric, were particularly well-suited for capturing the underlying patterns within the dataset and making accurate predictions. The resulting MSE is 89.6, with an average coefficient of determination of 0.898 over five cross-validations. Because the performance of the third fold was most similar to the mean of the five folds, it was selected for validation and to display the prediction performance of the best model.

5.7 Results of the Support Vector Regression Model

There are several hyper parameters for the SVR model, which, when working with linear regression and sklearn, are:

- The kernel function: this function is applied to the input data as a pre-processing stage to make the support vector algorithm nonlinear. By adding this pre-processing stage, linear regression on the processed data would result in nonlinear regression on the original data.

This is termed the *kernel trick.* This kernel function replaces the dot product in the original mathematical support vector model. Three kernel functions were tested:

- The linear kernel function: this is the normal dot product

$$k(x_i, x_j) = x_i, x_j \tag{5-6}$$

- The Gaussian kernel function:

$$k(x_i, x_j) = e^{\left\| \frac{x_i, x_j}{2\sigma} \right\|} \tag{5-7}$$

- The Gaussian RBF kernel function:

$$k(x_i, x_j) = e^{-\lambda \| x_i, x_j \|^2} \tag{5-8}$$

- The Polynomial kernel function:

$$k(x_{i,} x_j) = (1 + x_i, x_j)^d \tag{5-9}$$

- Where d is the degree of the polynomial. For d = 2, the kernel is the quadratic kernel function, and for d = 3, it is the cubic kernel function.

The linear, Gaussian, Gaussian RBF, quadratic and cubic kernel functions were tested in order to achieve the best regression performance.

- The solver: the solver determines the algorithm that would be used to train and tune the support vector regression model. Three solvers were tested to ensure the best performance:
    - The Iterative Single Data Algorithm (ISDA) solver. o The Sequential Minimal Optimization (SMO) solver. o The L1 soft-margin minimization by quadratic programming (L1QP) solver.

Using the previous two parameters of the model, fifteen combinations were tested and compared to choose the best SVR performance.

Each model was tested using a k-fold cross-validation with the five-fold method. This entailed creating five randomly selected train-test subgroups that were divided 80-20. The results of each model's training and testing for each of these five combinations were then summed to determine the model's ultimate performance.

The mean square error and the coefficient of determination were employed as performance indicators to compare the model. The outcomes demonstrated the superior performance of the quadratic kernel function models. Also, the SMO solver method performs fairly similarly with an MSE of 88.4 and an R^2 value of 1.0, while the L1QP analyzer provided the best result with a squared kernel function, having an MSE of 88.33 and an R^2 value of 1.0. These findings represent the average of five sets, each with a separate training test set, as was previously described. The five visualization folds with the best results were used to visualize the SVR findings. We can observe that the SVR predictions accurately forecast the real outputs, as seen by the outcome metrics.

The error histogram, Figure 5-11, displays how well the model worked. The frequency of occurrence for each range of error values is depicted via ten bins. This demonstrates that the error values are roughly centered around the zero error.

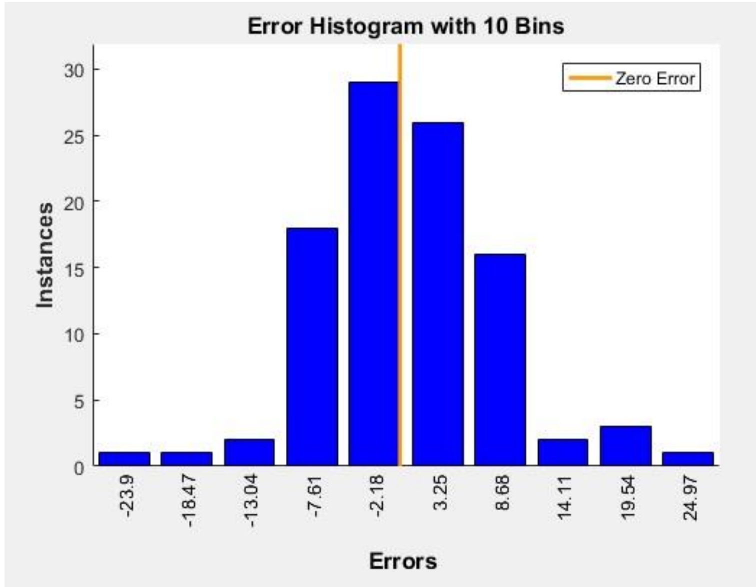


Figure 5-11. Error Histogram for SVR Model with Sequential Minimal Optimization

5.8 Results of Artificial Neural Network Model

The ANN algorithm was trained and tested using the raw data captured by the sensors. A 5-fold cross-validation method was used to improve the ANN's hyperparameters. This approach separated the complete data set into five folds, with the model trained on the first four folds and evaluated and validated on the fifth fold.

Sensitivity analysis was carried out for hyperparameter selection by gradually increasing the hidden layers of the network from 1 to 100. The ideal number of hidden layers was reached when the MSE was at its lowest.

As an example of a prediction, consider case 490; the test index real energy generation was 35.45 W, predicted energy generation was 48.02 W, the training R^2 rating was 1.0, and the RMSE = 2.1. Random data distribution of the data set was utilized for training, verifying, and testing the model after choosing the best hyperparameters to run the ANN. For verification, 15% of the data set was used to verify the model, 15% was used to test it, and 70% was used to train it. Figure 5-13 depicts the performance development over each training period. Up to 60 epochs into the training phase, rapid learning improvement was seen. During epoch 499, the best training performance of 4.761 was attained.

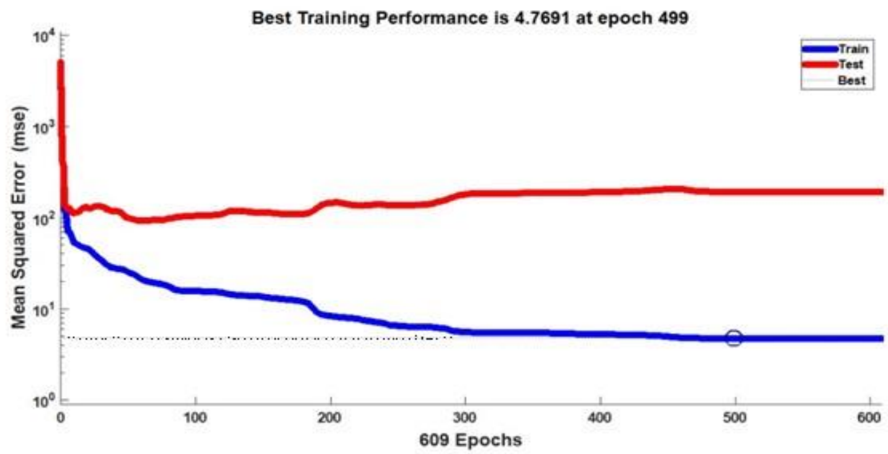


Figure 5-12. Mean Squared Error vs Epochs

Figure 5-13 shows the corresponding error histogram with ten bins. The error of the ANN predictions varied from -25.34 to 43.54. The figure shows the zero-error line, and it is clear the maximum number of instances occurs on this line.

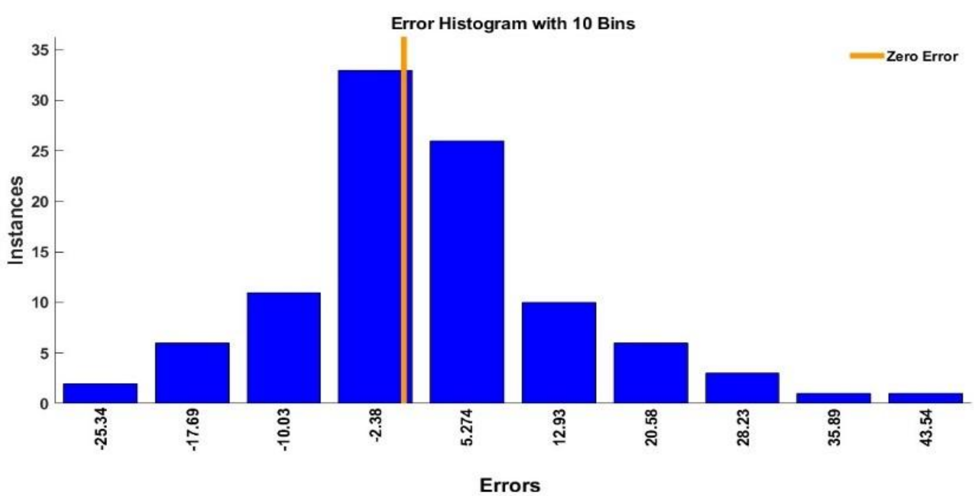


Figure 5-13. ANN Output Error Histogram

With ANN, R^2 of 1.0 was seen in the anticipated values when the prediction was compared to the target. The anticipated output values are reasonably close to the desired values. Also, it can be seen that the anticipated value deviates from the desired value for lower and greater power outputs. The created model is not very sensitive to these parameters: MSE = 98, Train RMSE = 2.04, Testing. RMSE = 1.57 and R^2 = 1.0 are the predicted findings.

5.9 Comparative Analysis of Three Learning Algorithms for Predictive Modelling

The accuracy and effectiveness of the predictions produced in the field of predictive modelling are greatly influenced by the choice of an appropriate learning algorithm. Artificial Neural Networks (ANN), K-Nearest Neighbors (KNN), and Support Vector Machines (SVM) have emerged as strong candidates among the variety of alternatives available, each providing particular advantages and prowess. This research examines a comparative analysis of these three algorithms—ANN, KNN, and SVM—with the goal of elucidating their individual complexities, assessing their effectiveness as predictive models, and identifying the situations in which they work best. For educated model selection and the best prediction results, it is essential to have a detailed grasp of these algorithms' properties and their implications as the field of data-driven decision-making continues to grow. This exploration reveals the distinguishing characteristics that set these algorithms unique by meticulously analyzing their underlying mechanisms, training procedures, and real-world applications. It also equips us with the knowledge we need to make wise decisions when it comes to predictive modelling projects. Table 5-5 provides performance comparison results of machine learning algorithms used in this study.

Table 5-5 Comparison of Machine Learning Algorithms Performance

Algorithm	Accuracy	R^2
ANN	98%	1.0
KNN	89.6%	0.898
SVM	88.4%	1.0

The results demonstrate the accuracy and coefficient of determination (R^2) of three different learning algorithms: Artificial Neural Networks (ANN), K-nearest neighbors (KNN), and Support Vector Machine (SVM). The outcomes show that the ANN algorithm attained a remarkable accuracy rate of 98%, along with a perfect R^2 score of 1.0. This demonstrates the ANN's outstanding predictive power and capacity to faithfully incorporate data variance, leading to a strong model fit. Comparatively, the KNN method displayed an impressive R^2 value of 0.898 and an accuracy of 89.6%. The KNN method maintains a strong predictive power, as evidenced by its significant R^2 score, albeit somewhat falling short of the ANN in accuracy. This demonstrates the algorithm's capacity to explain the variability of the data and provide precise predictions. Similar to this, the SVM method performed admirably, with an R^2 score of 1.0 and an accuracy rate of 88.4%. A strong R^2 value for the SVM confirms its usefulness in accurately modelling the data, and its acceptable accuracy further confirms its dependability in making predictions.

These findings demonstrate the effectiveness of the examined algorithms in predictive modelling tasks. Both the KNN and SVM algorithms demonstrate great predictive skills, further highlighting their usefulness for a variety of applications, even if the ANN emerges as the top performer in terms of accuracy and R^2. The detailed insights gained from this comparison study enable a greater understanding of each algorithm's strengths, assisting in the selection of an algorithm that is appropriate for a given set of modelling criteria and goals.

6 Conclusions and Future Work

6.1 Conclusions

The exploration of machine learning (ML) algorithms for solar energy generation forecasting has demonstrated significant potential, highlighting a significant improvement in accuracy over traditional approaches. The model's strong prediction ability for solar energy outputs was confirmed by constant successes in the coefficient of determination (R2) exceeding 0.88 during the course of the investigation. This notable metric demonstrated the efficacy and dependability of the ML-based forecasts, highlighting their potential to completely transform the forecasting of solar energy.

However, even with these successes, assessment at high power output limits revealed discrepancies between the predicted and measured values. These dramatic variations highlighted some of the model's accuracy limitations, especially when predicting very high or low power outputs. These cases showed subtle difficulties the model had in correctly predicting results in harsh environments, requiring more investigation and improvement.

The model's ability to predict solar energy generation was validated by the R2 values, which were regularly above 0.88. These values were an effective quantitative indicator. However, the observed limits at extremes indicated that continuous improvements and adjustments are required to support the model's correctness globally and guarantee its dependability in a wider range of situations.

6.2 Link to Objective

For the objective to improve the precision of solar energy forecasting, the obtained R^2 value of more than 0.88 shows a significant improvement in accuracy against conventional techniques. Furthermore, with respect to the creation of trustworthy predictive models for solar energy generation, the machine learning model performs exceptionally well in new data sets following extensive training, making it a dependable forecasting instrument. In addition, the model performs well across most ranges, but deviations at extreme power outputs point to areas that require improvement and further model development. These findings are consistent with the objective of addressing limitations and improving accuracy across the board.

6.3 Quantitative Evidence

Strong predictive strength is indicated by the consistently high R^2 values above 0.88 across several test sets, which validates the model's accuracy in predicting solar energy generation.

Notwithstanding the generally strong performance, deviations at extreme power outputs, where the expected values differ from the actual values, highlight areas that still require development.

6.4 Practical implications

The study's findings have a big impact on the renewable energy industry as Decision-making procedures for maximizing the performance of solar energy systems can be guided by the proven efficacy of machine learning algorithms in accurately predicting solar energy generation. Furthermore, even though the model exhibits great accuracy, recognizing its limitations at extreme power outputs highlights the necessity of continued research and development to further enhance the model's capabilities.

6.5 Potential Future Work

In conclusion, our research endeavours have paved the way for more precise and reliable forecasts of solar energy generation using machine learning algorithms. As we look towards the future, it becomes evident that there are several realistic avenues for further refinement and enhancement of our predictive capabilities. One crucial aspect of future research involves expanding the scope of our dataset. By incorporating more granular data collected at intervals such as per day, per hour, and even per minute, we can significantly enhance the accuracy of our predictions. This meticulous approach allows us to account for short-term variations and dynamic changes in solar energy generation, making our forecasts even more trustworthy.

Furthermore, the application of more advanced machine learning models, such as deep learning algorithms, holds substantial promise. These models have the potential to further elevate the precision and reliability of our forecasts. Through rigorous experimentation and practical examples, we can explore the capabilities of these sophisticated algorithms and assess their real-world performance. Recognizing the regional disparities in solar energy generation, future research can take a tailored approach. Creating location-specific models that account for the unique characteristics of different geographical areas and climates can substantially increase predictability. These models can be fine-tuned to capture the intricacies of solar energy production in specific regions, offering more accurate forecasts tailored to local conditions. To bolster the existing dataset, researchers may consider integrating supplementary data sources. Technologies such as satellite photography and remote sensing can provide valuable insights into factors like cloud cover and environmental conditions, contributing to more comprehensive forecasting models.

Lastly, the exploration of hybrid models presents an exciting avenue for future study. These hybrid models, which amalgamate machine learning strategies with traditional forecasting techniques like statistical models and time-series analysis, have the potential to provide a holistic approach to solar energy prediction. By leveraging the strengths of multiple forecasting techniques, these models can further enhance precision and reliability. In summary, the horizon of machine learning-based solar energy generation prediction is ripe with opportunities for future research and development. By continuously refining and advancing these techniques, researchers and industry experts can play a pivotal role in improving the performance and efficiency of solar energy systems. This ongoing pursuit holds the promise of a future characterized by increased reliance on renewable and sustainable energy sources, ultimately contributing to a more environmentally conscious and energy-efficient world.

7 List of References

[1] Khare, V., Nama, S. and Baredar, P. (2016) “Solar–wind hybrid renewable energy system: A review,” *Renewable and Sustainable Energy Reviews*, 58, pp. 23–33. doi: 10.1016/j.rser.2015.12.223.

[2] Ahmed, R. *et al.* (2020) “A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization,” *Renewable and Sustainable Energy Reviews*, 124(109792), p. 109792. doi: 10.1016/j.rser.2020.109792.

[3] Fan, J. *et al.* (2019) “Evaluation and development of empirical models for estimating daily and monthly mean daily diffuse horizontal solar radiation for different climatic regions of China,” *Renewable and Sustainable Energy Reviews*, 105, pp. 168–186. doi: 10.1016/j.rser.2019.01.040.

[4] Khandakar, A. *et al.* (2019) “Machine learning based photovoltaics (PV) power prediction using different environmental parameters of Qatar,” *Energies*, 12(14), p. 2782. doi: 10.3390/en12142782.

[5] Elrahmani, A. *et al.* (2021) “Status of renewable energy in the GCC region and future opportunities,” *Current Opinion in Chemical Engineering*, 31(100664), p. 100664. doi: 10.1016/j.coche.2020.100664.

[6] World Bank Group, ESMAP, and SOLARGIS, “GLOBAL SOLAR ATLAS.” https://globalsolaratlas.info/map.

[7] Inman, R. H., Pedro, H. T. C. and Coimbra, C. F. M. (2013) “Solar forecasting methods for renewable energy integration,” *Progress in energy and combustion science*, 39(6), pp. 535–576. doi: 10.1016/j.pecs.2013.06.002.

[8] Lari, A. J. *et al.* (2021) “Power generation voting prediction model of floating photovoltaic system,” in *2021 International Conference on Engineering and Emerging Technologies (ICEET)*. IEEE.

[9] Lari, A. J. and Egwebe, A. (2021) “Switched-mode power supplies comparison: PI, cascade PI and poiscast controller,” in *2021 International Conference on Electrical, Computer and Energy Technologies (ICECET)*. IEEE.

[10] Varotsos, C. A., Efstathiou, M. N. and Christodoulakis, J. (2019) “Abrupt changes in global tropospheric temperature,” *Atmospheric Research*, 217, pp. 114–119. doi: 10.1016/j.atmosres.2018.11.001.

[11] Nwaigwe, K. N., Mutabilwa, P. and Dintwa, E. (2019) “An overview of solar power (PV systems) integration into electricity grids,” *Materials science for energy technologies*, 2(3), pp. 629–633. doi: 10.1016/j.mset.2019.07.002.

[12] *Iea.org*. Available at: https://www.iea.org/data-and-statistics/charts/solar-pv-powergeneration-in-the-sustainable-development-scenario-2000-2030. (Accessed: August 28, 2023).

[13] Leggett, J. A., Logan, J. and Mackey, A. (2019) “China’s greenhouse gas emissions and mitigation policies,” *Congressional Research Service*.

[14] *Iea.org*. Available at: https://www.iea.org/reports/global-energy-review-2021. (Accessed: August 28, 2023).

[15] Impram, S., Varbak Nese, S. and Oral, B. (2020) "Challenges of renewable energy penetration on power system flexibility: A survey," *Energy strategy reviews*, 31(100539), p. 100539. doi: 10.1016/j.esr.2020.100539.

[16] *Renewablegreenenergypower.com*. Available at: https://www.renewablegreenenergypower.com/solar-energy/solar-energy-facts-concentratedsolar-power-csp-vs-photovoltaic-pv-panels. (Accessed: August 28, 2023).

[17] "Grid integration of large-capacity Renewable Energy sources and use of large-capacity Electrical Energy Storage" (2012) *International Electrotechnical Commission*.

[18] Hassan, A. *et al.* (2020) "Thermal management and uniform temperature regulation of photovoltaic modules using hybrid phase change materials-nanofluids system," *Renewable energy*, 145, pp. 282–293. doi: 10.1016/j.renene.2019.05.130.

[19] Ahmad, L. *et al.* (2020) "Recent advances and applications of solar photovoltaics and thermal technologies," *Energy (Oxford, England)*, 207(118254), p. 118254. doi: 10.1016/j.energy.2020.118254.

[20] Mohamed, H. A. *et al.* (2016) "Design, control and performance analysis of DC-DC boost converter for stand-alone PV system," in *2016 Eighteenth International Middle East Power Systems Conference (MEPCON)*. IEEE.

[21] Sahu, P., Verma, D. and Nema, S. (2016) "Physical design and modelling of boost converter for maximum power point tracking in solar PV systems," in *2016 International Conference on Electrical Power and Energy Systems (ICEPES)*. IEEE.

[22] *Marks, J. (2019). Abutted IRLED infrared scene projector design and their characterization. University of Delaware.* in *2019 Eighteenth International Middle East Power Systems Conference (MEPCON)*. IEEE.

[23] Abbassi, R. *et al.* (2019) "An efficient salp swarm-inspired algorithm for parameters identification of photovoltaic cell models," *Energy conversion and management*, 179, pp. 362– 372. doi: 10.1016/j.enconman.2018.10.069.

[24] Belcher, B. *et al.* (2017) "The effects of major solar integration on a 21-Bus system: Technology review and PSAT simulations," in *SoutheastCon 2017*. IEEE.

[25] Mellit, A. *et al.* (2020) "Advanced methods for photovoltaic output power forecasting: A review," *Applied Sciences (Basel, Switzerland)*, 10(2), p. 487. doi: 10.3390/app10020487.

[26] Babatunde, A. A. and Abbasoglu, S. (2019) "Predictive analysis of photovoltaic plants specific yield with the implementation of multiple linear regression tool," *Environmental progress & sustainable energy*, 38(4), p. 13098. doi: 10.1002/ep.13098.

[27] Shi, J. *et al.* (2012) "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," *IEEE Transactions on Industry Applications*, 48(3), pp. 1064–1069. doi: 10.1109/tia.2012.2190816.

[28] Wang, J. *et al.* (2018) "A short-term photovoltaic power prediction model based on the gradient boost decision tree," *Applied Sciences (Basel, Switzerland)*, 8(5), p. 689. doi: 10.3390/app8050689.

[29] Fentis, A. *et al.* (2017) "Short-term solar power forecasting using Support Vector Regression and feed-forward NN," in *2017 15th IEEE International New Circuits and Systems Conference (NEWCAS)*. IEEE.

[30] Lee, D. and Kim, K. (2019) "Recurrent neural network-based hourly prediction of photovoltaic power output using meteorological information," *Energies*, 12(2), p. 215. doi: 10.3390/en12020215.

[31] Abdel-Nasser, M. and Mahmoud, K. (2019) "Accurate photovoltaic power forecasting models using deep LSTM-RNN," *Neural computing & applications*, 31(7), pp. 2727–2740. doi: 10.1007/s00521-017-3225-z.

[32] Kayri, M., Kayri, I. and Gencoglu, M. T. (2017) "The performance comparison of Multiple Linear Regression, Random Forest and Artificial Neural Network by using photovoltaic and atmospheric data," in *2017 14th International Conference on Engineering of Modern Electric Systems (EMES)*. IEEE.

[33] Markovics, D. and Mayer, M. J. (2022) "Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction," *Renewable and Sustainable Energy Reviews*, 161(112364), p. 112364. doi: 10.1016/j.rser.2022.112364.

[34] Mastromauro, R. A., Liserre, M. and Dell'Aquila, A. (2012) "Control issues in single-stage photovoltaic systems: MPPT, current and voltage control," *IEEE Transactions on industrial informatics*, 8(2), pp. 241–254. doi: 10.1109/tii.2012.2186973.

[35] Aslam, Sheraz *et al.* (2021) "A survey on deep learning methods for power load and renewable energy forecasting in smart microgrids," *Renewable and Sustainable Energy Reviews*, 144(110992), p. 110992. doi: 10.1016/j.rser.2021.110992.

[36] Lin, G., Lin, A. and Gu, D. (2022) "Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient," *Information sciences*, 608, pp. 517–531. doi: 10.1016/j.ins.2022.06.090.

[37] Elsaraiti, M. and Merabet, A. (2022) "Solar power forecasting using deep learning techniques," *IEEE Access: practical innovations, open solutions*, 10, pp. 31692–31698. doi: 10.1109/access.2022.3160484.

[38] Muhammad, A. *et al.* (2019) "Deep learning application in power system with a case study on solar irradiation forecasting," in *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*. IEEE.

[39] Ghimire, S. *et al.* (2022) "Stacked LSTM sequence-to-sequence autoencoder with feature selection for daily solar radiation prediction: A review and new modeling results," *Energies*, 15(3), p. 1061. doi: 10.3390/en15031061.

[40] Wang, Y. *et al.* (2021) "A review of wind speed and wind power forecasting with deep neural networks," *Applied Energy*, 304(117766), p. 117766. doi: 10.1016/j.apenergy.2021.117766.

[41] Lari, A. J. *et al.* (2021a) "Photovoltaic System Ensemble Prediction System," in *2021 International Conference on Electrical, Computer and Energy Technologies (ICECET)*. IEEE.

[42] Nespoli, A. *et al.* (2022) "Machine Learning techniques for solar irradiation nowcasting: Cloud type classification forecast through satellite data and imagery," *Applied Energy*, 305(117834), p. 117834. doi: 10.1016/j.apenergy.2021.117834.

[43] Yang, D. and Perez, R. (2019) "Can we gauge forecasts using satellite-derived solar irradiance?," *Journal of Renewable and Sustainable Energy*, 11(2), p. 023704. doi: 10.1063/1.5087588.

[44] Antonanzas-Torres, F. *et al.* (2019) "Clear sky solar irradiance models: A review of seventy models," *Renewable and Sustainable Energy Reviews*, 107, pp. 374–387. doi: 10.1016/j.rser.2019.02.032.

[45] Nespoli, A. *et al.* (2022) "Machine Learning techniques for solar irradiation nowcasting: Cloud type classification forecast through satellite data and imagery," *Applied Energy*, 305(117834), p. 117834. doi: 10.1016/j.apenergy.2021.117834.

[46] Das, U. K. *et al.* (2018) "Forecasting of photovoltaic power generation and model optimization: A review," *Renewable and Sustainable Energy Reviews*, 81, pp. 912–928. doi: 10.1016/j.rser.2017.08.017.

[47] Zazoum, B. (2022) "Solar photovoltaic power prediction using different machine learning methods," *Energy reports*, 8, pp. 19–25. doi: 10.1016/j.egyr.2021.11.183.

[48] Nam, K., Hwangbo, S. and Yoo, C. (2020) "A deep learning-based forecasting model for renewable energy scenarios to guide sustainable energy policy: A case study of Korea," *Renewable and Sustainable Energy Reviews*, 122(109725), p. 109725. doi: 10.1016/j.rser.2020.109725.

[49] Sharifzadeh, M., Sikinioti-Lock, A. and Shah, N. (2019) "Machine-learning methods for integrated renewable power generation: A comparative study of artificial neural networks, support vector regression, and Gaussian Process Regression," *Renewable and Sustainable Energy Reviews*, 108, pp. 513–538. doi: 10.1016/j.rser.2019.03.040.

[50] Abuella, M. and Chowdhury, B. (2015) "Solar power forecasting using artificial neural networks," in *2015 North American Power Symposium (NAPS)*. IEEE.

[51] Dairi, A. *et al.* (2020) "Short-term forecasting of photovoltaic solar power production using variational auto-encoder driven deep learning approach," *Applied sciences (Basel, Switzerland)*, 10(23), p. 8400. doi: 10.3390/app10238400.

[52] Alomari, M. H. and Adeeb, J. (2019) "PVPF tool: An automated web application for real-time photovoltaic power forecasting," *International Journal of Electrical and Computer Engineering*, 9(1), pp. 34–41.

[53] Wang, X. *et al.* (2021) "A comprehensive review on the application of nanofluid in heat pipe based on the machine learning: Theory, application and prediction," *Renewable and Sustainable Energy Reviews*, 150(111434), p. 111434. doi: 10.1016/j.rser.2021.111434.

[54] Singh, U. *et al.* (2021) "A machine learning-based gradient boosting regression approach for wind power production forecasting: A step towards smart grid environments," *Energies*, 14(16), p. 5196. doi: 10.3390/en14165196.

[55] Mishra, M. *et al.* (2020) "Deep learning and wavelet transform integrated approach for short-term solar PV power prediction," *Measurement: journal of the International Measurement Confederation*, 166(108250), p. 108250. doi: 10.1016/j.measurement.2020.108250.

[56] Alomari, M. H., Adeeb, J. and Younis, O. (2018) "Solar photovoltaic power forecasting in Jordan using Artificial Neural Networks," *International Journal of Electrical and Computer Engineering (IJECE)*, 8(1), p. 497. doi: 10.11591/ijece.v8i1.pp497-504.

[57] Lari, A. J. *et al.* (2021c) "Reliable photovoltaics output power prediction in Qatar," in *2021 14th International Conference on Developments in eSystems Engineering (DeSE)*. IEEE.

[58] Ahmed, R. *et al.* (2020) "A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization," *Renewable and Sustainable Energy Reviews*, 124(109792), p. 109792. doi: 10.1016/j.rser.2020.109792.

[59] Zhou, Y. *et al.* (2020) "Prediction of photovoltaic power output based on similar day analysis, genetic algorithm and extreme learning machine," *Energy (Oxford, England)*, 204(117894), p. 117894. doi: 10.1016/j.energy.2020.117894.

[60] Khandakar, A. *et al.* (2019) "Machine learning based photovoltaics (PV) power prediction using different environmental parameters of Qatar," *Energies*, 12(14), p. 2782. doi: 10.3390/en12142782.

[61] Mishra, M. *et al.* (2020) "Deep learning and wavelet transform integrated approach for short term solar PV power prediction," *Measurement: journal of the International Measurement Confederation*, 166(108250), p. 108250. doi: 10.1016/j.measurement.2020.108250.

[62] Musbah, H., Aly, H. H. and Little, T. A. (2021) "Energy management of hybrid energy system sources based on machine learning classification algorithms," *Electric power systems research*, 199(107436), p. 107436. doi: 10.1016/j.epsr.2021.107436.

[63] Chahboun, S. and Maaroufi, M. (2022) "Performance comparison of K-nearest neighbor, random forest, and multiple linear regression to predict photovoltaic panels' power output," in *Advances on Smart and Soft Computing*. Singapore: Springer Singapore, pp. 301–311.

[64] Lin, G., Lin, A. and Gu, D. (2022) "Using support vector regression and K-nearest neighbors for short-term traffic flow prediction based on maximal information coefficient," *Information sciences*, 608, pp. 517–531. doi: 10.1016/j.ins.2022.06.090.

[65] Pan, M. *et al.* (2020) "Photovoltaic power forecasting based on a support vector machine with improved ant colony optimization," *Journal of Cleaner Production*, 277(123948), p. 123948. doi: 10.1016/j.jclepro.2020.123948.

[66] Ghimire, S., Deo, R. C., Casillas-Pérez, D., *et al.* (2022) "Improved Complete Ensemble Empirical Mode Decomposition with Adaptive Noise Deep Residual model for short-term multistep solar radiation prediction," *Renewable energy*, 190, pp. 408–424. doi: 10.1016/j.renene.2022.03.120.

[67] Korkmaz, D. (2021) “SolarNet: A hybrid reliable model based on convolutional neural network and variational mode decomposition for hourly photovoltaic power forecasting,” *Applied Energy*, 300(117410), p. 117410. doi: 10.1016/j.apenergy.2021.117410.

[68] Lu, Y., Wang, S. and Hu, Z. (2021) “Solar photovoltaic power prediction using convolutional neural network,” *Energy Procedia*, 185, pp. 3003–3008.

[69] Wang, L. *et al.* (2023) “Accurate solar PV power prediction interval method based on frequency-domain decomposition and LSTM model,” *Energy (Oxford, England)*, 262(125592), p. 125592. doi: 10.1016/j.energy.2022.125592.

[70] du Plessis, A. A., Strauss, J. M. and Rix, A. J. (2021) “Short-term solar power forecasting: Investigating the ability of deep learning models to capture low-level utility-scale Photovoltaic system behavior,” *Applied Energy*, 285(116395), p. 116395. doi: 10.1016/j.apenergy.2020.116395.

[71] Gundu, V. and Simon, S. P. (2021) “Short-term solar power and temperature forecast using recurrent neural networks,” *Neural processing letters*, 53(6), pp. 4407–4418. doi: 10.1007/s11063021-10606-7.

[72] Qamber, I. S. and Al-Hamad, M. Y. (2020) “Photovoltaic design for smart cities and demand forecasting using a truncated conjugate gradient algorithm. Optimization, Learning, and Control for Interdependent Complex Networks,” pp. 277–295.

[73] Lipu, M. H. *et al.* (2021) “Data-driven hybrid approaches for renewable power prediction toward grid decarbonization: Applications, issues and suggestions,” *Journal of Cleaner Production*, 328. Sahin, E. S., Bayram, I. S. and Koc, M. (2019) “Demand side management opportunities,

framework, and implications for sustainable development in resource-rich countries: Case study Qatar," *Journal of cleaner production*, 241(118332), p. 118332. doi: 10.1016/j.jclepro.2019.118332.

[74] Grätzel, M. (2005) "Solar energy conversion by dye-sensitized photovoltaic cells," *Inorganic chemistry*, 44(20), pp. 6841–6851. doi: 10.1021/ic0508371.

[75] Singh, R. (2020) "Approximate rooftop solar PV potential of Indian cities for high-level renewable power scenario planning," *Sustainable energy technologies and assessments*, 42(100850), p. 100850. doi: 10.1016/j.seta.2020.100850.

[76] Johari, F. *et al.* (2020) "Urban building energy modeling: State of the art and future prospects," *Renewable and Sustainable Energy Reviews*, 128(109902), p. 109902. doi: 10.1016/j.rser.2020.109902.

[77] Sadeghian, H. and Wang, Z. (2020) "A novel impact-assessment framework for distributed PV installations in low-voltage secondary networks," *Renewable energy*, 147, pp. 2179–2194. doi: 10.1016/j.renene.2019.09.117.

[78] Khan, M. *et al.* (2018) "A novel off-grid optimal hybrid energy system for rural electrification of Tanzania using a closed loop cooled solar system," *Energies*, 11(4), p. 905. doi: 10.3390/en11040905.

[79] Raina, G., Mathur, S. and Sinha, S. (2022) "Behavior of bifacial and monofacial photovoltaic modules under partial shading scenarios," *International Journal of Energy Research*, 46(9), pp. 12837–12853. doi: 10.1002/er.8057.

[80] Senthil, R. and Yuvaraj, S. (2019) "A comprehensive review on bioinspired solar photovoltaic cells," *International Journal of Energy Research*, 43(3), pp. 1068–1081. doi: 10.1002/er.4255.

[81] Ali, H., Muhammad *et al.* (2017) "Effect of dust deposition on the performance of photovoltaic modules in Taxila, Pakistan," *Thermal Science*, 21(2), pp. 915–923. doi: 10.2298/tsci140515046a.

[82] Kurtz, B., Mejia, F. and Kleissl, J. (2017) "A virtual sky imager testbed for solar energy forecasting," *Solar energy (Phoenix, Ariz.)*, 158, pp. 753–759. doi: 10.1016/j.solener.2017.10.036.

[83] Fan, J. *et al.* (2019) "Evaluation and development of empirical models for estimating daily and monthly mean daily diffuse horizontal solar radiation for different climatic regions of China," *Renewable and Sustainable Energy Reviews*, 105, pp. 168–186. doi: 10.1016/j.rser.2019.01.040.

[84] Hafez, A. Z., Yousef, A. M. and Harag, N. M. (2018) "Solar tracking systems: Technologies and trackers drive types-A review," *Renewable and Sustainable Energy Reviews*, 91, pp. 754–782.

[85] *PVsyst – logiciel photovoltaïque Pvsyst.com*. Available at: https://www.pvsyst.com/ (Accessed: August 28, 2023).

[86] Mehedi, I. M. *et al.* (2021) "Critical evaluation and review of partial shading mitigation methods for grid-connected PV system using hardware solutions: The module-level and array level approaches," *Renewable and Sustainable Energy Reviews*, 146(111138), p. 111138. doi: 10.1016/j.rser.2021.111138.

[87] Liu, Y.-H., Chen, J.-H. and Huang, J.-W. (2015) "A review of maximum power point tracking techniques for use in partially shaded conditions," *Renewable and Sustainable Energy Reviews*, 41, pp. 436–453. doi: 10.1016/j.rser.2014.08.038.

[88] Castellano, N. N. *et al.* (2015) "Optimal displacement of photovoltaic array's rows using a novel shading model," *Applied Energy*, 144, pp. 1–9. doi: 10.1016/j.apenergy.2015.01.060.

[89] Kosorić, V. *et al.* (2018) "General model of Photovoltaic (PV) integration into existing public high-rise residential buildings in Singapore-Challenges and benefits," *Renewable and Sustainable Energy Reviews*, 91, pp. 70–89.

[90] Abdelshafy, A. M., Hassan, H. and Jurasz, J. (2018) "Optimal design of a grid-connected desalination plant powered by renewable energy resources using a hybrid PSO-GWO approach. Energy conversion and management," 173, pp. 331–347.

[91] Riza, A. and Gilani, D. F. (2014) "Standalone photovoltaic system sizing using peak sun hour method and evaluation by TRNSYS simulation," *International Journal of Renewable Energy Research*, 4(1), pp. 109–114.

[92] Akikur, R. K. *et al.* (2013) "Comparative study of stand-alone and hybrid solar energy systems suitable for off-grid rural electrification: A review," *Renewable and Sustainable Energy Reviews*, 27, pp. 738–752. doi: 10.1016/j.rser.2013.06.043.

[93] El-Sattar, H. A. *et al.* (2021) "Optimal design of stand-alone hybrid PV/wind/biomass/battery energy storage system in Abu-Monqar, Egypt," *Journal of energy storage*, 44(103336), p. 103336. doi: 10.1016/j.est.2021.103336.

[94] *Solar panel manufacturers, half-cell solar panels, bifacial PV system Powerbluesun.com*. Available at: https://www.powerbluesun.com/modules_c1?gclid=CjwKCAjwitShBhA6EiwAq3RqA6gz9nBY 3iOcqe0ZZpniCeQYMCtOMZl8K-uWnng2jMpmVrMpD_EMTxoCgj4QAvD_BwE (Accessed: August 28, 2023).

[95] *SolarEdge inverter SetApp App Store*. Available at: https://apps.apple.com/us/app/solaredgeinverter-setapp/id1381441516 (Accessed: August 28, 2023).

[96] *PVEducation Pveducation.org*. Available at: https://www.pveducation.org/ (Accessed: August 28, 2023).

[97] Khandakar, A. *et al.* (2019) "Machine learning based photovoltaics (PV) power prediction using different environmental parameters of Qatar," *Energies*, 12(14), p. 2782. doi: 10.3390/en12142782.

[98] *Opensource.com*. Available at: https://opensource.com/resources/python,%20https://www.python.org/ (Accessed: August 28, 2023).

[99] Molin, S. (2019) *Hands-On Data Analysis with Pandas: Efficiently perform data collection, wrangling, analysis, and visualization using Python*. Packt Publishing Ltd.

[100] *Google colaboratory Google.com*. Available at: https://colab.research.google.com/notebooks/intro.ipynb (Accessed: August 28, 2023).

[101]*The Python language references Python documentation*. Available at: https://docs.python.org/3/reference/index.html (Accessed: August 28, 2023).

[102]Hao, J. and Ho, T. K. (2019) "Machine learning made easy: A review of Scikit-learn package in Python programming language," *Journal of Educational and Behavioral Statistics: a quarterly publication sponsored by the American Educational Research Association and the American Statistical Association*, 44(3), pp. 348–361. doi: 10.3102/1076998619832248.

[103]Xu, S., Xue, Y. and Chang, L. (2021) "Review of power system support functions for inverter-based distributed energy resources-standards, control algorithms, and trends," *IEEE Open Journal of Power Electronics*, (2), pp. 88–105.

[104]Kabir, M. N. *et al.* (2014) "Coordinated control of grid-connected photovoltaic reactive power and battery energy storage systems to improve the voltage profile of a residential distribution feeder," *IEEE Transactions on industrial informatics*, 10(2), pp. 967–977. doi: 10.1109/tii.2014.2299336.

[105]Johnson, D. O. and Hassan, K. A. (2016) "Issues of power quality in electrical systems," *International Journal of Energy and Power Engineering*, 5(4), pp. 148–154.

[106]Mansour, O. N. *et al.* (2019) "A geographical information system-based multi-criteria method for the evaluation of solar farms locations: A case study in Souss-Massa area, southern Morocco," *Energy (Oxford, England)*, 182, pp. 900–919. doi: 10.1016/j.energy.2019.06.063.

[107]Bolwig, S. *et al.* (2020) "Climate-friendly but socially rejected energy-transition pathways: The integration of techno-economic and socio-technical approaches in the Nordic-Baltic

region," *Energy research & social science*, 67(101559), p. 101559. doi: 10.1016/j.erss.2020.101559.

[108]Iweh, C. D. *et al.* (2021) "Distributed generation and renewable energy integration into the grid: Prerequisites, push factors, practical options, issues and merits," *Energies*, 14(17), p. 5375. doi: 10.3390/en14175375.

[109]Tröndle, T. *et al.* (2020) "Trade-offs between geographic scale, cost, and infrastructure requirements for fully renewable electricity in Europe," *Joule*. doi: 10.1016/j.joule.2020.07.018.

[110]Das, S. R. *et al.* (2021) "A comprehensive survey on different control strategies and applications of active power filters for power quality improvement," *Energies*, 14(15), p. 4589. doi: 10.3390/en14154589.

[111]El-Hady B Kashyout, A. *et al.* (2021) "Hybrid renewable energy/hybrid desalination potentials for remote areas: selected cases studied in Egypt," *RSC advances*, 11(22), pp. 13201– 13219. doi: 10.1039/d1ra00989c.

[112]Chugh, S. *et al.* (2019) "Machine learning approach for computing optical properties of a photonic crystal fiber," *Optics Express*, 27(25), pp. 36414–36425. doi: 10.1364/OE.27.036414.

[113] Zhengxuan Liu, Ying Sun, Chaojie Xing. (2022, November). *Artificial intelligence powered large-scale renewable integrations in multi-energy systems for carbon neutrality transition: Challenges and future perspectives*. https://www.sciencedirect.com/science/article/pii/S2666546822000428

[114] Geasa, M.M., 2021. Development of an Arduino-based universal testing apparatus. Archives of Agriculture Sciences Journal, 4(3), pp.121-130.

[115] Kondaveeti, H.K., Kumaravelu, N.K., Vanambathina, S.D., Mathe, S.E. and Vappangi, S., 2021. A systematic literature review on prototyping with Arduino: Applications, challenges, advantages, and limitations. Computer Science Review, 40, p.100364.

[116] Kuria, K.P., Robinson, O.O. and Gabriel, M.M., 2020. Monitoring temperature and humidity using Arduino Nano and Module-DHT11 sensor with real-time DS3231 data logger and LCD display. Health Hyg, 6(7), p.8.

[117] Trevathan, J., Schmidtke, S., Read, W., Sharp, T. and Sattar, A., 2021. An IoT general-purpose sensor board for enabling remote aquatic environmental monitoring. Internet of Things, 16, p.100429.

[118] Hu, Q., Tang, X. and Tang, W., 2020. A smart chair sitting posture recognition system using flex sensors and FPGA implemented artificial neural network. IEEE Sensors Journal, 20(14), pp.8007-8016.

[119] Soler-Llorens, J.L., Galiana-Merino, J.J., Nassim-Benabdeloued, B.Y., Rosa-Cintas, S., Ortiz Zamora, J. and Giner-Caturla, J.J., 2019. Design and implementation of an Arduino-based plug-and-play acquisition system for seismic noise measurements. Electronics, 8(9), p.1035.

[120] Jang, J.H., Cho, M., Kim, J., Kim, D. and Kim, Y., 2023. Paralyzing Drones via EMI Signal Injection on Sensory Communication Channels. In NDSS.

[121] Sarker, B., Radin, S.Y., Zahedee, S.H. and Shanto, T.T.I., 2021. Arduino-based power system protection (Doctoral dissertation, Brac University).

[122] SHOMPA, M.S. and HOQUE, A., 2022. Design And Implementation Of An IoT-Based Solar Power Monitoring And Data Logger System.

[123] IRENA (2023). Renewable Power Generation Costs in 2022.

[124] SolarPower Europe (2023). Global Market Outlook For Solar Power 2023 - 2027.

[125] Reichelstein, S., & Sahoo, A. (2023). Levelized cost estimates of solar photovoltaic electricity in the United Kingdom until 2035. Applied Energy, 339, 121903.

[126] IRENA (2023). Cost of Power Generation from Concentrated Solar Power (CSP) Technology.

[127] Chen, D., Cui, X., Zhang, Q., Li, D., Cheng, W., Fei, C. and Yang, Y., 2022. A survey on analog-to-digital converter integrated circuits for miniaturized high resolution ultrasonic imaging system. Micromachines, 13(1), p.114.

[128] Tredennick, A.T., Hooker, G., Ellner, S.P. and Adler, P.B., 2021. A practical guide to selecting models for exploration, inference, and prediction in ecology. *Ecology*, *102*(6), p.e03336.

[129] Dangi, N. (2020, November 5). *Solar energy*. Share & Discover Presentations | SlideShare. https://www.slideshare.net/NishaDangi99/solar-energy-239101949

[130] Svatos, J., Fischer, J. and Holub, J., 2023. High-speed equivalent-time sampling virtual instrument based on microcontroller ADC. Measurement, 220, p.113392.

[131] Gozum, J., DESIGN OF A LOW-COST, MICROCONTROLLER-BASED PHASOR MEASUREMENT UNIT.

8 Appendices

8.1 Equations

The PV cell equivalent circuit has the following generic representation of the PV cell current [6]:

$$I = N_P I_{PH} - N_P I_S \left(exp\left(\frac{V_{cell} + I_{cell} R_S}{V_{th}} \right) - 1 \right) - \left(\frac{V_{cell} + I_{cell} R_S}{R_P} \right) \quad (1)$$

$$I_{PH} = G\left(I_{SC} + K_I (T_C + T_{REF}) \right) \quad (2)$$

$$I_{RS} = \frac{I_{SC}}{exp\left(\frac{qV_{OC}}{KAT_C N_S} \right) - 1} \quad (3)$$

$$I_S = I_{RS} \left(\frac{T_C}{T_{REF}} \right)^3 \left(exp\left(\frac{qE_G \left(\frac{1}{T_{REF}} - \frac{1}{T_C} \right)}{KA} \right) \right) \quad (4)$$

linear regression model using the following formula:

$$Y_i = \alpha_0 + \alpha_1 X_1 \quad (5)$$

Several independent factors using the multivariant linear regression model using the following formula:

$$Y_i = \alpha_0 + \alpha_1 X_1 + \cdots + \alpha_p X_p \quad (6)$$

The formula below describes the input x and output y of each neuron in an ANN:

$$y = F(wx + b) \quad (7)$$

Anticipated value $\widehat{y_i}$ as:

$$e_i = y_i - \widehat{y_i} \quad (8)$$

Predicted value is to the true:

$$e_{p,i} = \frac{e_i}{y_i} \text{ x } 100\% \quad (9)$$

Mean Absolute Error (MAE):

$$MAE = \frac{1}{n} \sum_{1}^{n} |e_i| \quad (10)$$

Mean Absolute Percentage Error (MAPE):

$$MAPE = \frac{1}{n} \sum_{1}^{n} |e_{p,i}| \quad (11)$$

Mean Squared Error (MSE):

$$MSE = \frac{1}{n}\sum_{1}^{n} e_i^2 \tag{12}$$

Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{n}\sum_{1}^{n} e_i^2} \tag{13}$$

The coefficient of determination is also widely used. It has the following formula:

$$R^2 = 1 - \frac{\sum(y_i - \widehat{y_i})^2}{\sum(y_i - \bar{y})^2} \tag{14}$$

- The Manhattan distance. Using this metric, the distance between two n-ddimensional points x and y that correspond to two inputs of n features is defined as $D(x, y) = \sum_1^n |x_i - y_i|$.
- The Euclidean distance. The conventional definition of distance as $D(x, y) = \sqrt{\sum_1^n (x_i - y_i)^2}$
- The Minkowski distance. A generalized form of the previous two definitions. The Minkowski distance of order p is defined as $D(x, y) = \sqrt[p]{\sum_1^n |x_i - y_i|^p}$.

Final output is defined as $\hat{Y} = \frac{\sum_{j=1}^{k} w_j Y_j}{\sum_{j=1}^{k} w_j}$ where Y_j is the training output of each of the K neighbours.

- The linear kernel function: this is the normal dot product $k(x_i, x_j) = x_i . x_j$
- The Gaussian kernel function: $k(x_i, x_j) = e^{-\frac{\left\|x_i - x_j\right\|^2}{2\sigma}}$
- The Gaussian RBF kernel function: $k(x_i, x_j) = e^{-\lambda\left\|x_i - x_j\right\|^2}$
- The Polynomial kernel function: $k(x_i, x_j) = (1 + x_i . x_j)^d$ where d is the degree of the polynomial. For $d = 2$, the kernel is called the quadratic kernel function and for $d = 3$ it is called the cubic kernel function.

8.2 Appendix 2: Python Code

```
import pandas as pd from google.colab import files uploaded = files.upload() df = pd.read_csv("powers.csv")
df df.dtypes
%matplotlib inline
%config InlineBackend.figure_formats = {'png', 'retina'}


# Plot the pairplot to discover correlation between power generation and other variables.  import seaborn as
sns sns.set(style="ticks")
 sns.pairplot(df) df.describe() from sklearn.preprocessing import
LabelEncoder le = LabelEncoder() df['condition'] =
le.fit_transform(df.Temperature) df
power_train = df.drop(['Temperature', 'Relative
Humidity','PVTemperature','Irradiance','Dust','Wind Speed','Power'], axis=1) power_target = df.Power
from sklearn.model_selection import LeaveOneOut from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score from sklearn.metrics import accuracy_score
 def rmse(actual, predicted):     from sklearn.metrics import
mean_squared_error     from math import sqrt


    return sqrt(mean_squared_error(actual, predicted))
 loo = LeaveOneOut()
loo.get_n_splits(power_train)


import numpy as np train_r2_scores =
np.array([]) test_r2_scores = np.array([])
train_rmse_scores = np.array([])
test_rmse_scores = np.array([])
```

```
predicted_powers = np.array([])

actual_powers = np.array([])


# Train Linear Regression model # It is small data, so  for train_index,

test_index in loo.split(power_train):     print("Test

index:{}".format(test_index))     #print("TRAIN:", train_index, "TEST:",

test_index)     regr = LinearRegression()


    X_train, X_test = power_train.iloc[train_index], power_train.iloc[test_index]     y_train, y_test =

power_target.iloc[train_index], power_target.iloc[test_index]     regr.fit(X_train,y_train)


    #print(X_test, y_test)     y_train_pred =

regr.predict(X_train)     y_test_pred =

regr.predict(X_test)     #print(y_test.values,

y_test_pred)

        train_r2_score = regr.score(X_train, y_train)     train_r2_scores =

np.append(train_r2_scores, train_r2_score)     test_r2_score = r2_score(y_test.values,

y_test_pred)     test_r2_scores = np.append(test_r2_scores, test_r2_score)

        train_rmse_score = rmse(y_train, y_train_pred)     train_rmse_scores =

np.append(train_rmse_scores, train_rmse_score)     test_rmse_score = rmse(y_test.values,

y_test_pred)

    test_rmse_scores = np.append(test_rmse_scores, test_rmse_score)


    actual_powers = np.append(actual_powers, y_test.values[0])     predicted_powers =

np.append(predicted_powers, y_test_pred[0])     print("Actual energy generation: {}\tPredicted energy

generation: {}".format(y_test.values[0], y_test_pred[0]))
```

```
    print("Train R^2 score: {}\tTest R^2 score:{}".format(train_r2_score, test_r2_score))
print("Train RMSE: {}\tTest RMSE:{}\n".format(train_rmse_score, test_rmse_score))


pd.DataFrame.std(df.Power)
print("Train average RMSE: {}\tTest average
RMSE:{}".format(np.average(train_rmse_scores), np.average(test_rmse_scores))) print("Train average R2:
{}\tTest average
R2:{}".format(np.average(train_r2_scores), np.average(test_r2_scores))) import matplotlib.pyplot as plt
# Plotting LOO predictions
# http://scikit-learn.org/stable/auto_examples/plot_cv_predict.html y = actual_powers fig,
ax = plt.subplots() ax.scatter(y, predicted_powers, edgecolors=(0, 0, 0)) ax.plot([y.min(),
y.max()], [y.min(), y.max()], 'k--', lw=4) ax.set_xlabel('Measured') ax.set_ylabel('Predicted')
plt.show() from sklearn.metrics import accuracy_score
# Create model with whole data regr = LinearRegression() regr.fit(power_train, power_target)
power_pred = regr.predict(power_train) print("RMSE: {}\tR2 score:
{}".format(rmse(power_target.values, power_pred), r2_score(power_target.values, power_pred)))
 print('Coefficients: \n', regr.coef_)
print(power_train.columns) print('Intercepts: \n',
regr.intercept_ ) from sklearn.metrics import r2_score def
performance_metric(y_true, y_predict):
    """ Calculates and returns the performance score between          true and predicted
values based on the metric chosen. """
    # TODO: Calculate the performance score between 'y_true' and 'y_predict'     score =
r2_score(power_pred,actual_powers)
    # Return the score     print ("Model has a coefficient of determination, R^2, of
```

{:.3f}.".format(score))   return score

Calculate the performance of this model

performance_metric(actual_powers, power_pred)