

MULTIVARIATE TIME SERIES ANALYSIS OF SOLAR IRRADIANCE FOR  
PHOTOVOLTAIC SYSTEMS THE HYBRIDIZATION OF NARX AND LSTM  
MODELS

OSAMA GAMAL MAHMOUD IBRAHIM MOTIR

UNIVERSITI TEKNOLOGI MALAYSIA



**UNIVERSITI TEKNOLOGI MALAYSIA**

**DECLARATION OF THESIS / UNDERGRADUATE PROJECT REPORT AND  
COPYRIGHT**

Author's full name : Osama Gamal Mahmoud Ibrahim Motir

Date of Birth : 8<sup>th</sup> of April 2000

Title : DAILY FORECASTING OF SOLAR IRRADIANCE FOR PV  
SYSTEMS: THE HYBRIDIZATION OF NARX AND LSTM MODELS

Academic Session : 2024/2025-1

I declare that this thesis is classified as:

**CONFIDENTIAL**

(Contains confidential information under the  
Official Secret Act 1972)\*

**RESTRICTED**

(Contains restricted information as specified by  
the organization where research was done)\*

**OPEN ACCESS**

I agree that my thesis to be published as online  
open access (full text)

1. I acknowledged that Universiti Teknologi Malaysia reserves the right as follows:
2. The thesis is the property of Universiti Teknologi Malaysia
3. The Library of Universiti Teknologi Malaysia has the right to make copies for the purpose of research only.
4. The Library has the right to make copies of the thesis for academic exchange.

Certified by:

---

**SIGNATURE OF STUDENT**

---

MCS241015

---

**MATRIC NUMBER**

---

**SIGNATURE OF SUPERVISOR**

---

**NAME OF SUPERVISOR**

Date: 15 Dec 2024

Date: 15 Dec 2024

NOTES : If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction



“We hereby declare that we have read this proposal and in our opinion this proposal is sufficient in term of scope and quality for the award of the degree of Master of Data Science”

Signature : \_\_\_\_\_

Name of Supervisor I : \_\_\_\_\_

Date : \_\_\_\_\_



## **BAHAGIAN A - Pengesahan Kerjasama\***

Adalah disahkan bahawa projek penyelidikan tesis ini telah dilaksanakan melalui kerjasama antara \_\_\_\_\_ dengan \_\_\_\_\_

Disahkan oleh:

Tandatangan :

Tarikh :

Nama :

Jawatan :

(Cop rasmi)

\* *Jika penyediaan tesis atau projek melibatkan kerjasama.*

---

---

## **BAHAGIAN B - Untuk Kegunaan Pejabat Sekolah Pengajian Siswazah**

Tesis ini telah diperiksa dan diakui oleh:

Nama dan Alamat Pemeriksa Luar : \_\_\_\_\_

Nama dan Alamat Pemeriksa Dalam : \_\_\_\_\_

Nama Penyelia Lain (jika ada) : \_\_\_\_\_

Disahkan oleh Timbalan Pendaftar di SPS:

Tandatangan : \_\_\_\_\_ Tarikh : 15JULAI 2018

Nama : \_\_\_\_\_



MULTIVARIATE TIME SERIES ANALYSIS OF SOLAR IRRADIANCE FOR  
PHOTOVOLTAIC SYSTEMS THE HYBRIDIZATION OF NARX AND LSTM  
MODELS

OSAMA GAMAL MAHMOUD IBRAHIM MOTIR

A project report submitted in fulfilment of the  
requirements for the award of the degree of  
Master of Data Science

Faculty of Computing  
Faculty of Engineering  
Universiti Teknologi Malaysia

Dec 2024



## **DECLARATION**

I declare that this project report entitled "*Multivariate Time Series Analysis Of Solar Irradiance For Photovoltaic Systems: The Hybridization Of NARX And LSTM Models*" is the result of my own research except as cited in the references. The project report has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : .....

Name : Osama Gamal Mahmoud Ibrahim Motir

Date : 15 DECEMBER 2024

## **DEDICATION**

This thesis is dedicated to my father and mother who have been very supportive in ensuring that I come this far and continually encouraging me to aspire for more. It is also dedicated to all my siblings who were always there to encourage me when the going got tough and supported me through-out.

## **ABSTRACT**

In an era where the integration of renewable energy into power grids is surging, precise solar irradiance forecasting emerges as a key for efficient grid management. This research responds to the increased complexity brought by the influx of renewables, focusing on solar energy's unpredictable nature due to intricate weather dynamics. The study delineates an advanced approach to solar irradiance forecasting by fusing a Nonlinear Auto-Regressive with Exogenous Inputs (NARX) model and Long Short-Term Memory (LSTM) networks through a sophisticated Time-Series Residual correction technique to mitigate the challenges of solar energy predictability. Assessed using extensive meteorological data from Johor Bahru, Malaysia, the hybrid NARX-LSTM model exhibits marked improvements in forecast fidelity, with a normalized Rooted Mean Squared Error (nRMSE) of 1.11114%. The comparative performance analysis validates the model's superiority over traditional forecasting methods, substantiating its efficacy in enhancing the operational management of photovoltaic (PV) systems. The findings underscore the model's pivotal role in advancing the precision of solar irradiance prediction, thus facilitating the strategic integration of solar energy resources into the power grid infrastructure.

## **ABSTRAK**

Dalam era di mana penyepaduan tenaga boleh diperbaharui ke dalam grid kuasa sedang melonjak, ramalan sinaran suria yang tepat muncul sebagai kunci untuk pengurusan grid yang cekap. Penyelidikan ini bertindak balas terhadap peningkatan kerumitan yang dibawa oleh kemasukan tenaga boleh diperbaharui, memfokuskan pada sifat tenaga suria yang tidak dapat diramalkan disebabkan oleh dinamik cuaca yang rumit. Kajian itu menggariskan pendekatan lanjutan untuk ramalan sinaran suria dengan menggabungkan model Nonlinear Auto-Regressive dengan Exogenous Inputs (NARX) dan rangkaian Memori Jangka Pendek Panjang (LSTM) melalui teknik pembetulan Sisa Siri Masa yang canggih untuk mengurangkan cabaran tenaga suria. kebolehramalan. Dinilai menggunakan data meteorologi yang meluas dari Johor Bahru, Malaysia, model NARX-LSTM hibrid mempamerkan peningkatan ketara dalam ketepatan ramalan, dengan Ralat Purata Kuasa Dua Berakar (nRMSE) yang dinormalkan sebanyak 1.11114%. Analisis prestasi perbandingan mengesahkan keunggulan model berbanding kaedah ramalan tradisional, membuktikan keberkesanannya dalam meningkatkan pengurusan operasi sistem fotovoltaik (PV). Penemuan ini menekankan peranan penting model dalam memajukan ketepatan ramalan sinaran suria, sekali gus memudahkan penyepaduan strategik sumber tenaga suria ke dalam infrastruktur grid kuasa.

## TABLE OF CONTENTS

	TITLE	PAGE
<b>DECLARATION</b>		iii
<b>DEDICATION</b>		iv
<b>ACKNOWLEDGEMENT</b>		v
<b>ABSTRACT</b>		vi
<b>ABSTRAK</b>		vii
<b>TABLE OF CONTENTS</b>		viii
<b>LIST OF TABLES</b>		xi
<b>LIST OF FIGURES</b>		xii
<b>LIST OF ABBREVIATIONS</b>		xiv
<b>LIST OF APPENDICES</b>		xv
<b>CHAPTER 1      INTRODUCTION</b>		<b>1</b>
1.1     Research Background		1
1.2     Problem Statement		4
1.3     Research Goal		4
1.4     Research Objectives		5
1.5     Scope of The Project		5
<b>CHAPTER 2      LITERATURE REVIEW</b>		<b>7</b>
2.1     Introduction		7
2.2     Microgrid and Renewable Energy Integration		7
2.2.1     Basic Principle of Microgrids		7
2.2.1.1     Microgrid Architecture		8
2.2.2     Integrating Renewable Energy into Microgrids		10
2.3     Challenges in Microgrid Operation		11
2.3.1     Stability in Microgrids		11
2.3.2     Power quality Issues		13
2.3.2.1     Voltage fluctuations		13

2.3.2.2	Voltage sag and swell	14
2.4	Weather Forecasting in Microgrid	16
2.4.1	Importance of Accurate Forecasting	16
2.4.2	Solar Forecasting Approaches	19
2.4.2.1	Physical Approaches	20
2.4.2.2	Statistical / Hybrid Approaches	21
2.5	Machine Learning in Weather Forecasting	28
2.5.1	Types of Machine Learning Algorithms	29
2.5.2	Deep Learning for Time-Series Forecasting	30
2.5.2.1	Neural Network Models	30
2.5.3	Challenges and Limitations	39
2.6	Chapter Summary	41
<b>CHAPTER 3</b>	<b>RESEARCH METHODOLOGY</b>	<b>42</b>
3.1	Introduction	42
3.2	Research Framework	42
3.3	Data Analysis	44
3.3.1	Identification of Forecast Parameters	45
3.3.2	Data Collection	48
3.3.3	Data Preprocessing	49
3.4	Model Development	53
3.4.1	Nonlinear Autoregressive with Exogenous Inputs (NARX) Neural Network:	53
3.4.2	Long Short-Term Memory (LSTM) Neural Network	56
3.4.3	Hybrid NARX-LSTM Model	58
3.5	Input Sensitivity Analysis	63
3.6	Chapter Summary	64
<b>CHAPTER 4</b>	<b>INITIAL RESULTS</b>	<b>65</b>
4.1	Introduction	65
4.2	Exploratory Data Analysis	65
4.2.1	Overview of the Dataset	66

4.2.2	Statistical Summary	68
4.2.3	Correlation Analysis	71
4.2.4	Time-Serise Analysis	75
4.3	Input Sensitivity Result	82
4.4	Evaluation of The Proposed Model	87
4.5	Chapter Summary	94
<b>CHAPTER 5</b>	<b>CONCLUSION</b>	<b>95</b>
5.1	Summary	95
5.2	Future Work	95
<b>REFERENCES</b>		<b>97</b>

## LIST OF TABLES

TABLE NO.	TITLE	PAGE
Table 2.1	Summary of different ways in which accurate forecasting impacts microgrid management.	19
Table 2.2	Statistical models summarize	22
Table 2.3	classification of solar irradiance forecasts based on approach	26
Table 2.4	classification depending on the temporal horizon for predicting	27
Table 2.5	Summary of the forecasting models for solar irradiance that were examined.	28
Table 2.6	Neural Network Models	37
Table 2.7	A comprehensive analysis contrasting SI forecasting techniques with prior studies	39
Table 4.1	Statistical summary table	68
Table 4.2	Feature importance result	85
Table 4.3	Input Sensitivity Analysis	86
Table 4.4	Training performance result	90
Table 4.5	Tasting performance result	92

## LIST OF FIGURES

<b>FIGURE NO.</b>	<b>TITLE</b>	<b>PAGE</b>
Figure 1.1	Architecture of an MG	2
Figure 2.1	Structure of microgrid	10
Figure 2.2	A typical voltage fluctuation waveform.	13
Figure 2.3	voltage sag and voltage swell representation.	14
Figure 2.4	Typical physical approach	20
Figure 2.5	PHANN, a physical hybrid artificial neural network, was developed to anticipate output power	25
Figure 2.6	An overview of forecasting methods	26
Figure 2.7	Deep learning general structure	30
Figure 2.8	RBFNN's architecture	31
Figure 2.9	Architecture of a multilayer perceptron with three outputs and two hidden layers	32
Figure 2.10	Architecture of the neural network ensemble	33
Figure 2.11	Architecture of Recurrent Neural Network	35
Figure 2.12	structure of NARX-RNN	36
Figure 3.1	Research framework	44
Figure 3.2	scatter plots of (a) Air Temperature, (b) Cloud Opacity, (c) dewpoint temperature, (d) precipitable water, (e) precipitation rate, (f) relative humidity,(g) Surface Pressure, (h) Wind Direction and (i) wind speed - against the SI.	47
Figure 3.3	Historical data for 24 hours - January 1, 2007	49
Figure 3.4	Histograms Distributions	51
Figure 3.5	Comparison of Original and Normalized Meteorological Parameters	52
Figure 3.6	The architecture of the three-layered NARX	55
Figure 3.7	The architecture of the NARX	56
Figure 3.8	The architecture of the LSTM network	56

Figure 3.9	The architecture of the proposed model.	62
Figure 4.1	Variables boxplots	69
Figure 4.2	Correlation matrix heatmap	73
Figure 4.3	The heatmap of mean GHI by hour and month	74
Figure 4.4	Global horizontal irradiance (GHI) time-series	76
Figure 4.5	Daily average GHI	77
Figure 4.6	Monthly average GHI	78
Figure 4.7	Decomposition of the GHI	79
Figure 4.8	The 3D surface plot of Hour, Month, and GHI	80
Figure 4.9	Assessment of the NARX-LSTM model's performance in predicting solar irradiance for a 24-hour forecast period across various meteorological scenarios with excluding (cloud attenuation variable).	83
Figure 4.10	Assessment of the NARX-LSTM model's performance in predicting solar irradiance for a 24-hour (cloud attenuation variable).	84
Figure 4.11	Comparative Performance of NARX, LSTM, and Hybrid Models Across Meteorological Parameters	86
Figure 4.12	Performance Comparison of NARX, LSTM, and Hybrid Models.	87
Figure 4.13	scatter plots of model	88
Figure 4.14	Error Distributions	88
Figure 4.15	Training response	89
Figure 4.16	Assessment of the NARX-LSTM model's performance in predicting solar irradiance for a 24-hour forecast period across various meteorological scenarios with the selected variables.	91

## LIST OF ABBREVIATIONS

GHI	-	Global Horizontal Irradiance
NARX	-	Nonlinear Autoregressive with Exogenous Inputs
LSTM	-	Long Short-Term Memory
nRMSE	-	Normalized Root Mean Square Error
RMSE	-	Root Mean Square Error
MSE	-	Mean Square Error
PV	-	Photovoltaic
AI	-	Extensible Markup Language
ANN	-	Artificial Neural Network
AI	-	Artificial Intelligence
ML	-	Machine Learning
DL	-	Deep Learning
RNN	-	Recurrent Neural Network
SI	-	Solar Irradiance



# **CHAPTER 1**

## **INTRODUCTION**

### **1.1 Research Background**

Degradation of the environment is the most difficult challenge in nations that have been detribalized, and the argument that it is not a concern in countries that have not been industrialized is no longer relevant [Yoldaş et al., 2017]. The high emissions of greenhouse gases (GHGs) are having an effect on both industrialized and non-industrialized nations all over the world. The rise in energy consumption that is a direct result of the expansion of the human population is the root cause of environmental deterioration. In 2023, the CO<sub>2</sub> emissions record had shown that the reading of CO<sub>2</sub> had reached 37.4 Gt [Raimi et al., 2024] where is considered the highest record. Which led many countries to focusing on replacing the traditional power grid like Coal-fired power station where they use the concept of coal-combustion to generate electricity, to the renewable energy sources which is more sustain and safe on the environment.

For this reason, the development of clean and renewable energy sources has emerged as a crucial approach for the economic and social sustainability of development for every nation on the planet [Suman et al., 2021]. The idea of a microgrid was proposed by the researchers in order to resolve the conflict that exists between grid generation and distributed generation, as well as to make the most of the benefits that distributed generation offers in terms of the economics, energy, and the environment.

Microgrids, localized power systems that can operate independently or in conjunction with the traditional power grid structure as shown in Figure 1.1, represent a compact yet fully operational grid system, functioning within a confined geographical region [Raimi et al., 2024]. Typically, a Microgrid (MG) is characterized as a combined energy provision system composed of various Distributed Generations

(DGs), energy storage components, loads, and devices for monitoring and protection. It functions as an independent, self-sustaining system that interfaces with the broader grid as a unified, regulated entity [Sepasi et al., 2023]. This system is endowed with the ability for self-regulation, protective measures, and overall management, all while satisfying the customer's expectations for reliable power supply and high-quality electricity.



Figure 1.1 Architecture of an MG [Bank, 2020].

The shift to renewable energy is a worldwide movement in the twenty-first century. Solar energy is driving this change because of its huge resource potential. Photovoltaic (PV) energy offers significant availability and extended longevity to stakeholders. As well as there are many technologies that had been developed in field of solar energy in terms of improving the power efficiency (PV cell) also the development reach to the storage system where it can help in improving the isolated system and the hybrid system [Dutta et al., 2017]. However, photovoltaic energy exhibits several constraints about inadequate power stability and subpar power quality. Photovoltaic plants are consistently affected by meteorological factors like cloud cover, wind velocity, and temperature fluctuations. Moreover, the primary impetus for energy production, namely irradiation, is only accessible during daylight hours [Shakya et al., 2016].

Consequently, forecasting models have been extensively used for photovoltaic systems to predict produced photovoltaic power on one hand and load demand on the other, so facilitating intelligent demand response and efficient energy management.

For the purpose of forecasting the solar irradiance, researchers are employing a wide range of methodologies. These methods may be generally classified into four categories: data-driven approaches, image-based approaches, numerical weather prediction (NWP) models, and hybrid approaches [Guermoui et al., 2020]. For the purpose of forecasting the solar time series, data-driven or statistical and machine learning methodologies make use of the historical data that has been observed or felt in the past. Images obtained from sky/shadow cameras or satellites are utilised by image-based techniques, which then integrate this information with past data in order to eventually anticipate the irradiance. For the purpose of predicting the irradiance, NWP models make use of computer-based programs. There are a variety of hybrid or combination models, which are diverse combinations of two or more of the types that were originally listed. However many methods facing difficulty to accurately forecast the solar irradiance due to many factors.

Machine learning (ML) approaches have been suggested up until this point in order to comprehend the behavior of selected feature patterns that are constantly changing over time in order to reconstruct a clear vision about future values [Massaoudi et al., 2020]. In layman's terms, these methods examine the past (the data that was input) in order to make predictions about the future (the behavior of the system). Direct and indirect PVPF are the components that make up Photovoltaic power forecasting (PVPF). On the one hand, direct PVPF evaluates the meteorological data from the past in order to forecast the amount of electricity that may be generated by PV. The indirect PVPF, on the other hand, employs an approach that consists of two stages. When it comes to the initial stage, the solar irradiance is anticipated to be the PV power aspect that is most dependent on it. Within the second step, the determination of the power generated by photovoltaic cells is based on mathematical relationships [Massaoudi et al., 2019]. According to the findings presented in [Lei and Yang, 2019], it has been demonstrated that the second technique is superior in terms of accuracy, versatility, costability, and the amount of computing effort that it requires. According to the authors of [Akhter et al., 2019], hybrid models

can obtain competitive outcomes in comparison to the approaches that are considered to be state-of-the-art by merging two or more single machine learning models from different sources.

According to the approaches of forecasting that are now in use, there are a few handles that continue to require additional mastery. To begin, the range of prediction for weather forecast data is restricted, and the historical time series of solar power is non-stationary, dynamic, and nonperiodic. These characteristics make it impossible for typical artificial intelligence systems to comprehend the data. In the second place, the input-output prediction patterns of the current modes are investigated from the point of view of statistical analysis. This approach either disregards the effect of other connected elements or necessitates the collection of high-quality data from many associated factors, which restricts the practical applicability. In conclusion, the complex non-linearity of the photovoltaic time series over different forecasting horizons exists between univariate time steps and among the variables that are significant.

To address the current challenges and achieve precise photovoltaic power forecasting, the hypothesis of the proposed method in this paper is primarily based on the following considerations where the current research indicates that NARX effectively manages nonlinear relationships, while LSTM excels in extracting temporal features, and the attention mechanism mitigates the issues of distraction. hence, the hybridization of NARX and LSTM is proposed in this study.

## 1.2 Problem Statement

The total capacity of renewable energy sources around the globe reached 3,381 gigawatts in the year 2023. With a capacity of 1053 gigawatts (GW), solar energy has emerged as the second most widely adopted of these renewable sources for energy production. Malaysia is positioned as top four nations in Southeast Asia in terms of its significant solar energy capacity. In the present moment, renewable energy sources provide 13.3% of Malaysia's overall energy capacity, it is anticipated that Malaysia would attain 18.2% of its capacity to generate electricity from renewable

sources by 2025, and 70% by 2050. Consequently, the amount of integration of the renewable energy with the main grid will increase exponentially which will increase the complexity in terms of grid management, and stability within the grid due to the difficulty in renewable energy forecasting. Therefore, to achieve the integration with RE, an accurate and strong forecasting model is required for predicting the solar energy.

### **1.3 Research Goal**

The aim of the project is to present a hybrid nonlinear autoregressive network with long-term memory inputs and exogenous inputs (NARXLSTM) model to meet the growing need for accurate daily solar irradiance predictions, especially under volatile weather conditions in Malaysia.

### **1.4 Research Objectives**

- (a) To develop a hybrid machine learning model combining LSTM and NARX neural networks, focusing on improving the precision of daily solar irradiance predictions.
- (b) To analyze the influence of various weather parameters, determining their individual and collective impact on the forecasting model's performance and.
- (c) To develop a dashboard to display forecasts and analyze trends.

### **1.5 Scope of The Project**

- Python and MATLAB will be used as the main tool to develop and analyze the model evaluation.
- The model performance will be evaluated by using historical weather data that had collected from Johor, Malaysia environments.

- The project will focus on constructing a hybrid model that combines the strengths of Nonlinear Autoregressive with Exogenous Inputs (NARX) and Long Short-Term Memory (LSTM) networks to forecast solar irradiance.
- The model will integrate various weather parameters which are Air Temperature, Cloud Attenuation, Precipitation Rate, Dewpoint Temperature, Surface Pressure, Precipitable Water, Relative Humidity, Wind Speed, Wind Direction

## **CHAPTER 2**

### **LITERATURE REVIEW**

#### **2.1 Introduction**

The literature reviewed in this chapter is an effort to provide an in-depth review of research related to solar irradiance forecasting and the integration of renewable energy sources into microgrids. It includes an exploration of the emergence of forecasting models and analyzes the evolution from traditional methods to advanced machine learning techniques. The chapter also discusses the multifaceted challenges involved in the integration of renewable sources, such as solar energy, into microgrid systems. Through a critical examination of the existing literature, the chapter aims to offer a comprehensive understanding of the current state of technology, highlight the challenges faced in the integration of renewable energy, and identify potential solutions. The development of the hybrid NARX-LSTM model is placed in the context of a thorough literature review, thus providing a reasonable basis for the contributions offered by the model towards the increase in the efficiency and reliability of solar energy forecasting and microgrid management.

#### **2.2 Microgrid and Renewable Energy Integration**

##### **2.2.1 Basic Principle of Microgrids**

Recently, due to heightened environmental concerns and an exacerbated global energy crisis, the conventional centralised power supply has shown several drawbacks. Concurrently, the highly efficient and less polluting distributed generation (DG) has garnered growing interest [Lasseter and Paigi, 2004]. The idea of microgrids is not new. Nonetheless, the emergence of innovative methods for harnessing renewable energy, alongside more efficient electricity generation techniques and the adaptability of power electronics, is fostering the development of a new industry aimed at

promoting these innovations and integrating them into microgrids to optimise benefits for the power grid.

Microgrids, primarily known for their localized power generation and distribution capabilities where it consist of micro-sources, energy storage devices, loads, and control and protection systems, are the most efficient conduits for distributed generation (DG). When a microgrid is connected to the utility grid, it functions as a regulated load or generator, therefore mitigating the power quality and safety issues associated with the direct connection of distributed generators. Microgrids may function in islanded mode, hence enhancing system resilience and power supply availability [Lei et al., 2023].

At its core, an intelligent control system manages power generation and distribution from all energy sources present in the microgrid. It can refer to from renewable sources like solar or wind even traditional fossil-fuel based generators [Logenthiran et al., 2015]. The stability of the electrical power generated is balanced by the control system supporting the maintaining of power output.

Effective control is essential for the steady and effective functioning of microgrids. The comprehensive control needs arise from several factors, including voltage and frequency regulation, as well as power flow optimization [Guo and Mu, 2016]. Consequently, if well administered, it may function as a singular controlled entity, operating either in conjunction with the electric grid or in an isolated mode. Despite the multiple advantages microgrids provide to end users, their integration into existing distribution networks is impeded by several challenges mostly associated with their operation, protection, and control [Cagnano et al., 2020].

Microgrids are, overall, a localized power solution that can help to offset damages to the bigger grid. Microgrids achieve a more resilient and more stable power infrastructure at the local level through diversifying power sources, serving the islanding capabilities, including the resilience features, allowing quick recovery, optimising the load management, and integrating the renewable energy sources.

### 2.2.1.1 Microgrid Architecture

The distribution generators differ, therefore, their microgrid architecture. The structure of microgrid consists or characterized by five major as is illustrated in Figure 2.1 : (a) microsources or distributed generators, (b) flexible loads, (c) distributed energy storage devices, (d) control systems, and (e) the point of common coupling components linked to a low voltage distribution network and capable of operation in a coordinated, regulated fashion in both the service to the utility grid and service to themselves or landed modes [Shahgholian, 2021]. Different forms of renewable energy resources are included as the power generators in a microgrid.

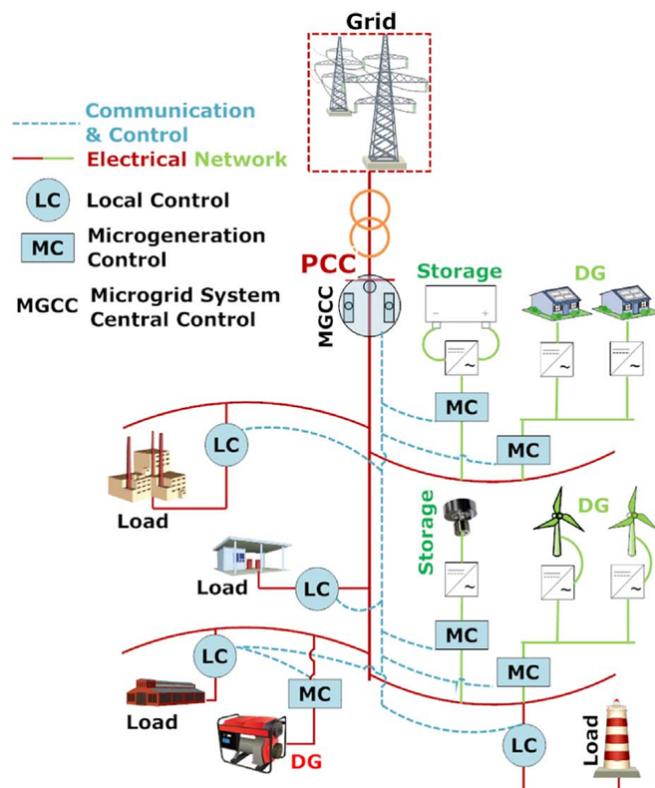


Figure 2.1 Architecture of Microgrid [Mariam et al., 2016].

The distribution system is the main body of the microgrid, providing the electricity from the power source to all connected loads. It includes physical infrastructure and power lines, cables and transformers that help make efficient use of this within a microgrid [Shahgholian, 2021].

DG sources are the local power generation units within the microgrid. Sources may be conventional generators, renewable energy such as solar panels or wind turbines - or indeed one on this 'small privileges of nature' [Alsaïdan et al., 2017]. These DG sources are an integral part in meeting the overall electricity demand of the microgrid and can suit themselves to both need and resource available locally.

Power storage devices (batteries) are an absolutely essential part of microgrids. Where the Storage units may balance reserves between short-term to long-term use range. The microgrid is linked to the upstream network, which may receive the full or partial electricity by the main grid. When linked to a grid, it can both receive or inject electricity into the main grid, suggesting that it may increase the grid efficiency and address energy crises to a certain degree [Shahgholian, 2021].

The control and communications modules are the brain of the microgrid system [Albarakati et al., 2022]. This includes control systems, monitoring equipment, and communication networks, which allows real-time monitoring, optimization of power flow, as well as it is aimed at establishing communication among several microgrid components in order to monitor and control in the real-time the overall microgrid [Gungor et al., 2011]. Therefore, these modules help manage the load efficiently and keep the system stable and integrated with the main grid or other microgrids.

The categorisation of microgrid systems is largely dependent on the selection of the aforementioned components and the integration with the main electrical grid network [Mariam, 2018]. Figure 2.2 depicts the fundamental structure of this categorisation. With reference to grid integration, the microgrid system may be grid linked or isolated. Microgrid may be operated as AC or DC distribution networks. Based on DG sources, both AC and DC microgrid may further be classified into three types- entirely conventional, partly conventional/renewable and totally renewable [Mariam et al., 2013]. Both AC and DC systems may have energy storage devices installed. The AC microgrid may further be classed as line frequency or high frequency AC (HFAC) microgrid systems.

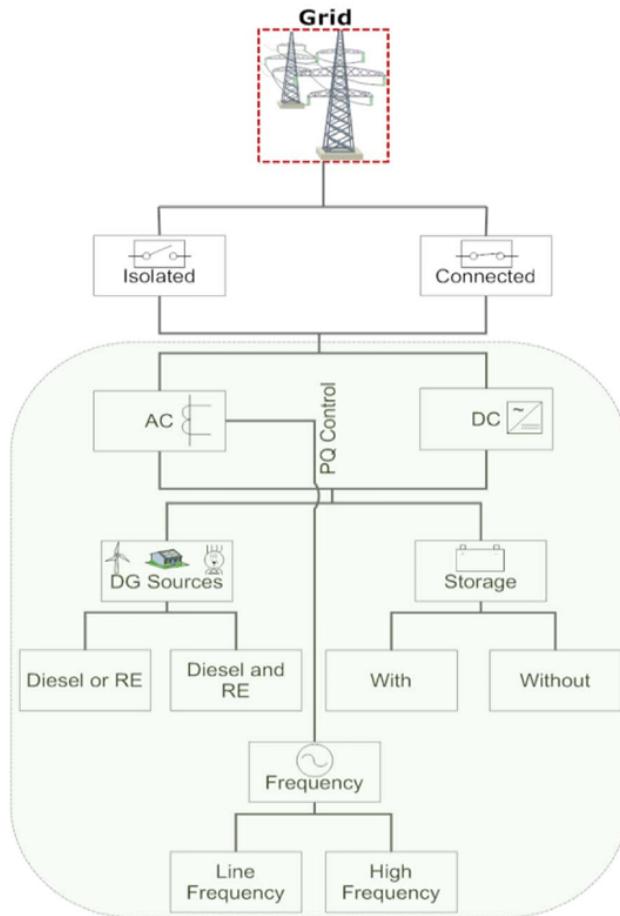


Figure 2.2 Structure of Microgrid [Mariam et al., 2016].

### 2.2.2 Integrating Renewable Energy into Microgrids

An important first step toward achieving sustainability and resilience in the energy system is integrating renewable energy into microgrids.[Kiehbadroudinezhad et al., 2023]. This integration is part of keeping sustainable development running and winning for the environment and the economy, in terms of the savings in costs and emission reductions [Saeed et al., 2021].

Solar and wind power as renewable energy are directly implementable in local communities and substantially reduce the dependence on conventional fossil fuels. The large majority of these resources can meet energy demand at location by aggregating these resources in a microgrid which will improve grid resilience and security [Lu et al., 2016, Hussain et al., 2019].

While renewable integrated into microgrids is not without risks. The power output from solar and wind resources tends to fluctuate, and is therefore seldom fully reliable [Saeed et al., 2021]. Besides, high penetration of renewable may cause power quality issues, as with voltage fluctuation [Kalakotla and Korra, 2023]. The following section gives a bit more detail on these challenges.

## 2.3 Challenges in Microgrid Operation

Microgrids have many advantages but integrating renewable energy sources within them brings with it its own set of hurdles. While benefits like greater reliability, local level grid resilience and opportunities for renewable integration are enticing, the complexities inherent in these same benefits will need to be carefully managed. The challenges span over issues related to technical such as stability in microgrids, power quality issues, voltage fluctuations and voltage sag and swell Voltage sag and swell with unique aspects associated with each [Shahzad et al., 2023].

A fundamental operational challenge of microgrids is the variability and intermittency inherent in renewable energy sources such as wind and solar power. These energy sources may generate variable energy outputs, depending upon the time of day and meteorological circumstances [Choudhury, 2020]. The unpredictability in power generation poses a significant problem for power regulation and achieving balance within the microgrid.

### 2.3.1 Stability in Microgrids

This section elucidates critical elements influencing the stability of the power system and the complications arising from the diverse operational behaviors of microgrids and their dispersed generating sources. The challenges can be broadly grouped under three key aspects, namely, lower system inertia (limited spinning reserves in power generation), reduced voltage stability (lower energy distribution support), and low frequency oscillations in power [Choudhury, 2020, Gopakumar et al., 2014].

Firstly, reduced system inertia results to cascading effects where angular stability leads to voltage and frequency swing [Gopakumar et al., 2014]. This instability can lead to some major operating hassles, which can compromise energy delivery and cause breakdowns or failures at any one time.

The deficiency in energy distribution support leads to a reduction in voltage stability margins. This is particularly pertinent during peak demand periods, when sustaining a constant voltage is crucial for activating the microgrid and the associated appliances appropriately [Energy, 2016].

Finally, the low frequency oscillating power is an effect of the changes in the power sharing ratio between the distributed generations [Energy, 2016]. Such fluctuation can significantly disturb the delicate equilibrium of power sharing in the microgrid, resulting in decreased efficiency and operational issues.

Within the domain of microgrid operations, stability is defined as how well a system can respond to disturbance, and its ability to revert to normal operation after being perturbed. Stability in a microgrid definition is categorized into two major categories- steady state stability and dynamic stability.

Steady-state stability refers to the ability of a microgrid to maintain a stable voltage and frequency within defined limits under normal and abnormal conditions. This type of stability is vital in the daily usage of a microgrid; stability is required to keep all the connected appliances up and running and to avoid variations that can result in a shutdown or further damage [Choudhury, 2020].

In contrast, dynamic stability refers to the ability of the system to return itself to its operating point following a disturbance. These disturbances may arise from differences in load or generation capable of displacing the system from equilibrium [Shahzad et al., 2023]. So, dynamic stability is a reflection of the microgrid's resiliency in response to disturbances and the capacity to maintain reliable energy delivery of the highest quality, post-adversity.

## 2.3.2 Power quality Issues

Power quality, an essential factor, significantly influences the optimal and harmonious performance of the microgrid system [Bandeiras et al., 2020]. Owing to the decentralized architecture of microgrids and the incorporation of diverse renewable energy sources. Power quality issues significantly impact system performance [Kalakotla and Korra, 2023].

### 2.3.2.1 Voltage fluctuations

Voltage fluctuations, commonly referred to as "flicker," are swift alterations in voltage magnitude that occur within the regulatory confines of typical gradual voltage variations ( $\pm 5\%$  of the nominal value). In accordance with the IEEE standards, it signifies a recurrent variation in voltage magnitude that falls within a range of 0.9 to 1.1 per unit [Shafiullah et al., 2010]. This condition is usually induced by power sources whose output oscillates over time.

As a salient power quality issue, voltage fluctuation frequently manifests when renewable energy sources are merged with the grid. The primary cause for these fluctuations is the high penetration of renewable energy sources which are inherently unpredictable and unregulated. Factors such as irregular solar irradiance due to cloud cover, geographic variation in PV installation, and temporal changes in wind speed all contribute to voltage instability [Mohamed et al., 2015].

One of the major repercussions of these voltage fluctuations is an effect known as voltage flicker. In the past, assessments of voltage fluctuations relied on measures such as the peak-to-peak RMS voltage, its energy spectrum, and duration. However, contemporary characterization of voltage fluctuations is primarily based on two parameters: short-term flicker severity ( $P_{ST}$  index) and long-term flicker severity ( $P_{LT}$  index) [Bajaj and Singh, 2020]. A typical voltage fluctuation waveform is depicted in Figure 2.3.

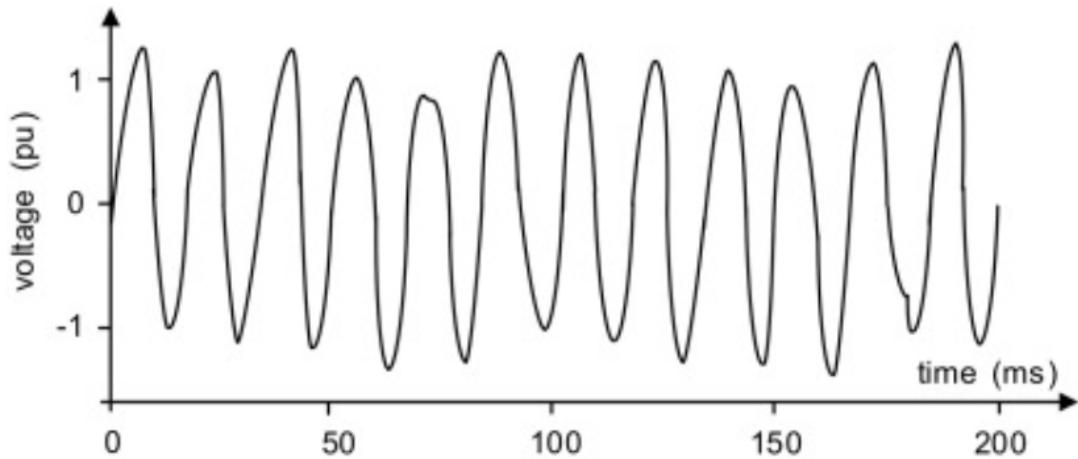


Figure 2.3 A typical voltage fluctuation waveform. [Bajaj and Singh, 2020]

### 2.3.2.2 Voltage sag and swell

Voltage sag and swell are prominent power quality issues, particularly notable when integrating renewable energy sources into the grid. Manifesting as transient reductions or surges in voltage levels, these phenomena are provoked by alterations in the load, or the power produced by the microgrid especially those incorporating renewable energy sources. Voltage sag and swell are graphically depicted in Figure 2.4, respectively. Voltage sag and swell can spawn an array of complications, encompassing equipment damage, diminished system efficiency, and power outages [Adefarati and Bansal, 2019]. Furthermore, they can adversely affect the lifecycle and performance of sensitive electrical equipment connected to the microgrid. The rapid progression of power electronics technology has now made it feasible to mitigate issues related to voltage stability [Bajaj and Singh, 2020].

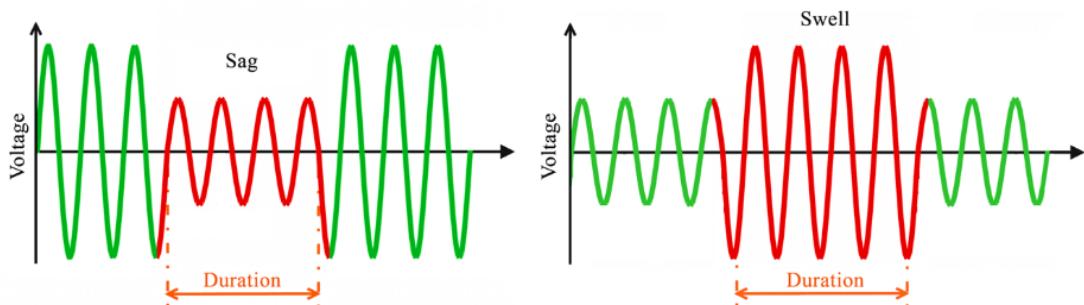


Figure 2.4 voltage sag and voltage swell representation. [Mattar et al., 2024]

Voltage Sag: also referred to as a voltage drop, can be ascribed to fluctuations in the load or problems associated with the network, including short-circuiting or voltage interruptions. With more renewable source of energy such as solar and wind-based electricity, voltage sag is one of the direct effects pertained by varying and unstable climatic nature. For example, reduced wind speed or amount of light would cut the flow of electricity from windmills or solar cells leading to brief fluctuation in the voltage [Mattar et al., 2024]. Recognizing that the IEC 61000-30 is the industry standard governing the definition of voltage sag, the latter is widely understood as a short-term RMS voltage excursions below 10% the value of the rated system voltage, expected to last between a half a cycle and one minute. These sags can be caused by constant fluctuations in the load such as starting up of a motor or a short circuit.

Voltage swell typically occurs due to a decrease or removal of load, as well as the disconnection of a significant power component within the electrical power system. Renewable energy sources, sourced from natural resources like wind or sun, experience fluctuations in production during peak power generation while consumption remains minimal. For example, during a sunny day or a windy night, a renewable energy system is likely to generate excess energy beyond the grid's requirements, perhaps leading to a temporary increase in voltage levels [Mattar et al., 2024]. The IEC 61000-4-30 standard defines voltage swells as intentional or unintentional increases in the RMS voltage at the system voltage or more by 10% over a duration ranging from half a cycle to one minute [Dhone et al., 2018].

Therefore, voltage quality issues need to be addressed in microgrids including renewable energy sources in order to maintain stable and reliable operation. This can be achieved through different methods such as using power electronic, energy storage systems. achieving this should involve the use of converters, and effective demand response strategies. Moreover, accurate renewable energy generation forecasting has a key role in resolving these power quality issues [Kaushal and Basak, 2020].

As a result, integrating weather forecasting with microgrid operations benefits in minimizing system efficiency and reliability, as well as various stability and power quality problems.

## **2.4 Weather (Solar) Forecasting in Microgrid**

Solar power generation is directly influenced by local weather conditions [Mishra and Ramesh, 2009]. It is essential to recognize that solar electricity is inherently diurnal and fluctuates throughout the day in accordance with variations in sun irradiation levels. A calculation of future power generation from this type of source is essential for achieving a balanced system. Nonetheless, these prospective power computations necessitate an understanding of the forthcoming meteorological variables that influence these technologies [Rodríguez et al., 2018].

Forecasting weather variables for energy generation assessments is a well-established endeavor, with numerous methodologies documented in the literature. Various models and influential characteristics, including cloudiness and irradiance, have been employed [Lara-Fanego et al., 2012]. Nonetheless, other authors advocate for the prediction of renewable generation as a means to attain this objective. Forecasting entails projecting future trends derived from historical data. The application of this approach may lead to a significant decrease in ambiguity regarding the utilization of clean energies.

### **2.4.1 Importance of Accurate Forecasting**

This section contends that the necessity for accurate forecasting is crucial for microgrids incorporating a significant proportion of solar or other renewable energy sources. The uncertainty stemming from renewable energy generation, especially owing to climatic and environmental changes, can substantially impact power balance and system stability [Ahmad et al., 2023]. Consequently, precise weather forecasting has critical importance, since it is essential for predicting Renewable Distributed Generation (DG) and load power profiles. Such forecasts facilitate efficient power production planning and are essential for island microgrids with limited dispatchable power sources or microgrids aiming to effectively implement grid-connected flexibility services [Alamo et al., 2019].

Power fluctuations of Renewable Energy Sources (RES), such as photovoltaic (PV), are primarily determined by two factors: It is a deterministic factor related to revolution of the Earth about the Sun, and a stochastic factor depending on atmospheric conditions such as cloudiness, dust, pollution, or local shadows on PV modules [Nespoli et al., 2019]. The uncertainty associated with renewable power sources and end users, however, poses a major challenge for the robustness, security and reliability of the integrated electricity systems. Accurate forecast of power production from RES within this context can significantly assist the management and operation of modern energy system such as microgrids.

Microgrids, whether functioning independently or connected to the main grid, generally consist of a diverse set of generation resources (including photovoltaic, wind, and conventional sources), loads (both manageable and unmanageable), and various Energy Storage Systems (ESS) such as batteries, fuel cells, flow batteries, and thermal storage. This varied composition presents new problems to load and renewable energy source forecasting [Dutta et al., 2017]. Given these obstacles, a critical objective is to determine the appropriate dispatch strategy via a centralized Energy Management System (EMS) [Moretti et al., 2019], in which renewable energy source forecasting is essential.

These systems must equilibrate the use of electricity from controlled resources, like energy storage systems (ESS), diesel generators, micro-turbines, or gas turbines, while considering demand and production from renewable sources [Ma and Ma, 2018]. Accurate forecasting methods are essential for this equilibrium, as seen in Table 2.1, which delineates how accurate forecasting improves microgrid management by maximizing resource use, assuring dependability, and promoting efficient operation in renewable-integrated grids.

In addition, maintaining a constant and reliable power supply to local customers becomes more difficult when distributed generators have high penetration due to their intermittent, time varying outputs from weather conditions. Similarly, electricity consumption is not constant as it due to seasonal effects and the user behavior to changes in electricity tariff [Vincent et al., 2020]. As a result, forecasting

of power generation and load demand can be tackled accurately to solve unit commitment and to optimally schedule the operation of energy storage devices.

Apart from balancing power supply and demand, accurate forecasting is imperative in microgrid management and plays a role in several other operational, economic and reliability facades. They are especially pronounced when the microgrid is composed of renewable energy sources, the unpredictability of which greatly exacerbates the need for forecast accuracy.

Significant improvements in operational efficiency are possible as a result of power demand and supply close to match up forecasts. As an example, an accurate estimation of peak renewable generation periods allows for optimal use of costly technologies, such as diesel generators or battery storage [Vincent et al., 2020]. This is an optimization which, in addition to reducing the operational cost of the system, avoids the usage of fossil fuel resources, and hence results in environmental sustainability of the grid operation.

Furthermore, adequate load forecasting improves maintenance planning. The periods of reduced renewable power generation due to weather condition are also anticipated to schedule preventive maintenance tasks. As a result, during high generation periods, the operational disruption is minimized, and this helps in the improvement of the grid's reliability [Ma and Ma, 2018].

The exact forecasting ensures the proactive stance towards risk management. However, these unexpected power surges or drops can inflict a lot of damage to sensitive electrical equipment [Sone et al., 2013]. Such events can, however be anticipated with accurate forecasting. Such events can be foreseen with accurate forecasting and can take some protective measures in advance to protect the infrastructure and to continue supplying the power.

Energy markets are frequently used by microgrids to purchase or sell excess power. Accurate forecasting enables operators to make cost effective decisions about when to sell excess power or about when to buy power from the grid operations.

Moreover, in the presence of multiple power sources, economic dispatch decisions can be made according to forecasting to operate the most economic [Dudek et al., 2023].

An important component of a renewable integrated microgrid is the energy storage that can serve as a buffer between the intermittency of renewable generation. Optimization of charging and discharging schedules of the storage systems can prolong their life, reduce costs, and produce more accurate load and renewable power forecasting prediction [Vincent et al., 2020].

Essentially, the need for accurate forecasting creates many challenges, but if achieved, it can shed significant light on improving some of the operational, economic and reliability aspects of microgrid management, particularly those with high renewable energy penetration.

Table 2.1 Summary of different ways in which accurate forecasting impacts microgrid management.

<b>Impact Area</b>	<b>Description</b>
Operational Efficiency	Forecasts made with high accuracy enable usage of resources in an optimized manner while decreasing operational costs and dependency on fossil fuels.
Maintenance Planning	By means of forecasting, preventive maintenance tasks can be scheduled during periods of reduced power generation so as to improve grid reliability.
Risk Management	Power surges or drops can be predicted, enabling advance implementation of protective measures that will protect infrastructure, as well as maintaining uninterrupted power supply.
Economic Dispatch and Energy Trading	Energy trading decisions and market operations are optimized and costs are reduced with forecasting, allowing for efficient energy dispatch.
Energy Storage Management	The forecasting of load and the generation of renewable power helps optimize the charging and discharging schedules of energy storage systems.

## **2.4.2 Solar Forecasting Approaches**

Meteorology is an important field and weather forecasting, a key part of meteorology has a great significance in many branches of human activity such as agriculture, disaster management, aviation, energy sector (especially in the energy generation from renewable sources). The demand for accurate, reliable, and timely weather forecasts has led to the development of a variety of methodologies, broadly grouped into three categories: physical, statistical, and hybrid methods. Each with their special strengths, weaknesses and applicability, these approaches are the product of scientists, mathematicians and engineers spending decades trying to decode the patterns of Mother Nature.

### **2.4.2.1 Physical Approaches**

The earliest attempts to formalize weather forecasting, physical methods use physical laws to explain or predict what happens in the atmosphere. It is a method where data about physical variables like temperature, pressure, humidity, and are processed and modelled. All these variables are fed into models, from which the model can predict the weather prediction based on the interplay of these variables. The method is regarded as a theoretically grounded method, based on the scientific understanding of the processes in the atmosphere for weather prediction. Nevertheless, these models may be tedious as well as computationally demanding and are not able to fully account for the inherent uncertainty and complex non-linearity in atmospheric systems [Sobri et al., 2018].

The physical forecasting approach requires two separate models: Solar radiation, and the Photovoltaic (PV) system [El Hendouzi and Bourouhou, 2016]. As a result, the PV power generation is a function of weather parameters and PV system data that result in forecasted output of the PV system.

Figure 2.5 shows how PV power generation can be predicted. It appears that this prediction is mostly determined by weather parameters (Global Horizontal Irradiance (GHI) and Ambient Temperature (AT)). Other variables may also be considered, for example the aerosol index (AI) [Marinelli et al., 2014]. The second major factor

in PV power forecasting is PV system data. As even minor microclimatic changes are important but it is very important system location parameter. It can however affect the power generated by the panels [El Hendouzi and Bourouhou, 2016]. Furthermore, manufacturer specifications provide another avenue for poor PV power forecasting. Both these factors help estimate how much irradiance to fall on PV plane or lam orientation These fields include the array (POA) irradiance, the back-of-module temperature (Tm), to make it easier to predict PV power [Yang et al., 2014].

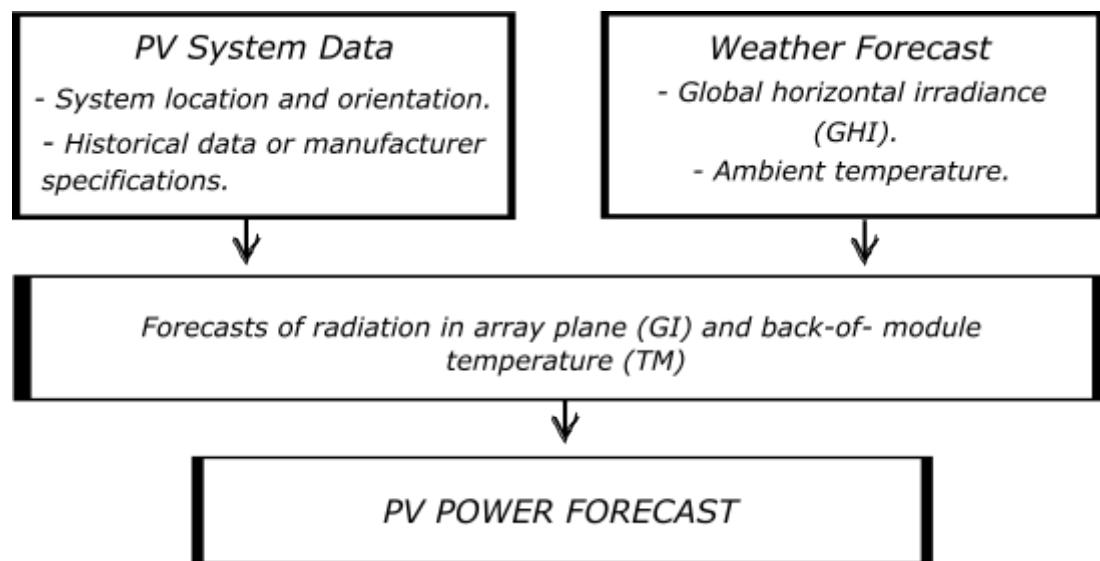


Figure 2.5 Typical physical model.

Two techniques are predominantly utilized: Numerical Weather Prediction (NWP) and Total Sky Imagery (TSI), both classified as physical approaches:

Numerical Weather Prediction (NWP) is a weather prediction strategy based on the dynamic equations of the atmosphere [Pelland et al., 2013], using complex algorithms to predict various meteorological variables such as temperature, pressure, wind, and rain.

On the other hand, according to the Solar Electric Power Association (SEPA) report [El Hendouzi and Bourouhou, 2016], Cloud Imagery, also known as Total Sky Imagery (TSI), ranks second among physical techniques. This method uses sky imaging to predict cloud movement with high resolution (10 to 100 m), providing

greater prediction accuracy for very short-term horizons (0 to 6 hours or intra-day). TSI allows for real-time PV power prediction, up to 30 minutes in advance (Intra-hour) and can detect every 30 minutes the presence of clouds that may impact a site [Pelland et al., 2013]. However, Total Sky Imagers have limited applicability in island regions due to their specific weather conditions.

#### **2.4.2.2 Statistical and Hybrid Approaches**

The statistical approach to weather forecasting leverages both mathematical models and stochastic strategies. This encompasses a broad range of techniques such as Artificial Neural Networks (ANNs), Machine Learning (ML), Adaptive Models (AM), and Data Mining (DM) [Ren et al., 2015].

These methods, detailed in Table 2.2, typically rely on historical data of solar irradiance and power production to predict future trends. They are primarily divided into two categories: Artificial Intelligence (AI)-based methods and regression methods [Vincent et al., 2020]. Regression techniques include seasonality analysis, Auto Regressive Integrated Moving Average (ARIMA), multiple regressions, and exponential smoothing. On the other hand, AI paradigms involve fuzzy inference systems, genetic algorithms, and neural networks.

A widely recognized model in the realm of statistical forecasting is the persistence technique, noted for its prevalence in various reports on PV power forecasting, including those by the International Energy Agency (IEA) and the Solar Electric Power Association (SEPA). This technique serves as a reference model and is used to compare various PV power forecasting techniques [El Hendouzi and Bourouhou, 2016]. The data utilized for these forecasts typically include time series or historical datasets. However, one inherent drawback of statistical methods lies in the accuracy of historical datasets used to train forecasting models.

Table 2.2 Statistical models summarized

<b>Category</b>	<b>Sub-category</b>	<b>Description</b>
Statistical Models	<b>Linear Models (Time Series Models)</b>	Linear statistical methods derive relations between predictors and the variable to be predicted through statistical analysis. Direct time series modeling has seen success in several studies.
	Persistence Model	A common reference model in solar or wind forecasting for short-term forecasting. It serves to check if the forecast model provides results better than any trivial reference model. A complex forecasting tool is only worth implementing if it outperforms these trivial models.
	Preprocessing of Input Data	This process deals with non-stationary series to transform them into an appropriate form for analysis. Statistical time series analysis requires dealing with stationary series (no trend or seasonality, homoscedastic), so preprocessing of data becomes necessary.
	ARMA Model	The ARMA model, typically applied to autocorrelated time series data, is an effective tool for understanding and predicting the future value of a time series. Comprising two parts: the autoregressive (AR) and moving average (MA), this model can represent several different types of time series using different orders.
	CARDS Model	This model involves the use of Fourier Series techniques for deseasoning. The residual series, obtained by subtracting the Fourier series component from the original series, is then modeled. This model can effectively deal with peaks in the series and provides a superior fit for the residual series.
Learning Models	<b>Non-linear Models</b>	These models use Artificial Intelligence (AI) techniques for forecasting and various applications such as control, data compression, optimization, pattern recognition, and classification.
	Artificial Neural Network (ANN)	A computing system inspired by the human brain, used for pattern recognition and prediction tasks.
	Wavelet Neural Network	A type of neural network that uses wavelet transformation.
	Recurrent Neural Network (RNN)	Recurrent Neural Networks, a type of neural network designed to recognize patterns in sequences of data.
	ANN and Classical Time Series Models Comparison	Both ANN and classical time series models have been compared in several studies. These studies found that the error of a simple regression model can be reduced significantly (by a factor in the range of 0.6 – 0.8) when using advanced models.

Furthermore, these datasets include more than just irradiance and module temperature; they may also comprise other variables such as Numerical Weather Prediction (NWP) output parameters and ground station measurements. Despite the advantages of statistical methods, the hybrid approach is emerging as a modern methodology. Hybrid approaches employ advanced statistical techniques to correct known inaccuracies associated with different forecasting methods through model bias adjustments or automated learning techniques [AlKandari and Ahmad, 2024]. Physical models generate forecasts for irradiance and module temperature, which are then used to simulate the solar plant model. The output of these simulations is subjected to statistical post-processing to enhance accuracy [Pu and Kalnay, 2019].

Apart from homogeneous or heterogeneous model combinations within the same approach, hybrid models can also be a mix of different approaches for instance, the combination of Satellite Images and ANNs. These hybrid models aim to achieve higher forecasting accuracy, and the most widely accepted example is the adaptive neural fuzzy inference system (ANFIS) [Singla et al., 2021].

A Physical Hybrid Artificial Neural Network (PHANN) based methodology, which has been homed in various research studies, was developed [Diagne et al., 2013], [Soman et al., 2010]. Figure 2.6 illustrates its essential structure. This forecasting tool integrates an Artificial Neural Network (ANN) trained on historical weather forecasts and clear-sky condition solar radiation (CSRM) to generate a precise hourly profile of expected PV plant output for the upcoming seven-day period [Nespoli et al., 2019]. Following the creation of this output, a post-processing step is required for validation, ensuring the reliability of the data. Consequently, it's crucial to correct any nonsensical values (whether negative or positive) of the output power during the night, setting them to zero.

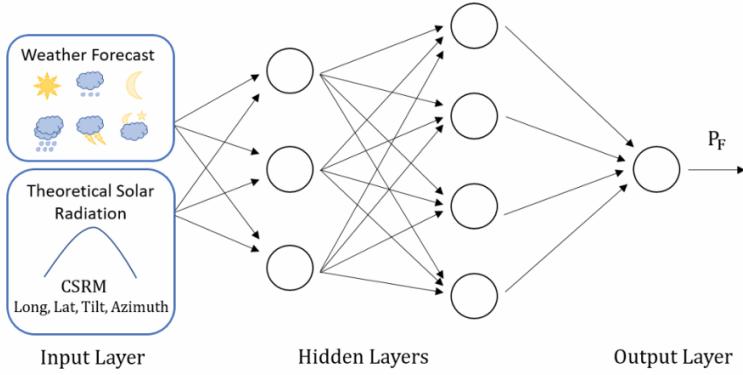


Figure 2.6 PHANN, a physical hybrid artificial neural network, was developed to anticipate output power [Nespoli et al., 2019]

It's worth noting that statistical methods operate under the premise that the forecasted value has a linear correlation with historical data within a specified time duration. Prominent statistical methods include autoregression (AR), moving average (MA), autoregressive moving average (ARMA), and autoregressive integrated moving average (ARIMA). Techniques such as the Box-Jenkins approach and the Kalman filter are effective tools for identifying components and parameters in time series [Ma and Ma, 2018].

In contrast, AI approaches bypass the physical process from input variables and output performance, replacing it with a ‘black box’ model. These models can either be single models such as fuzzy logic, artificial neural network (ANN), support vector regression (SVR), wavelet transform (WT), genetic algorithm (GA), and expert systems or hybrid models [Ma and Ma, 2018], which integrate one or more algorithms to pursue higher forecasting accuracy. In Figure 2.7, a summary of the forecasting methods is shown.

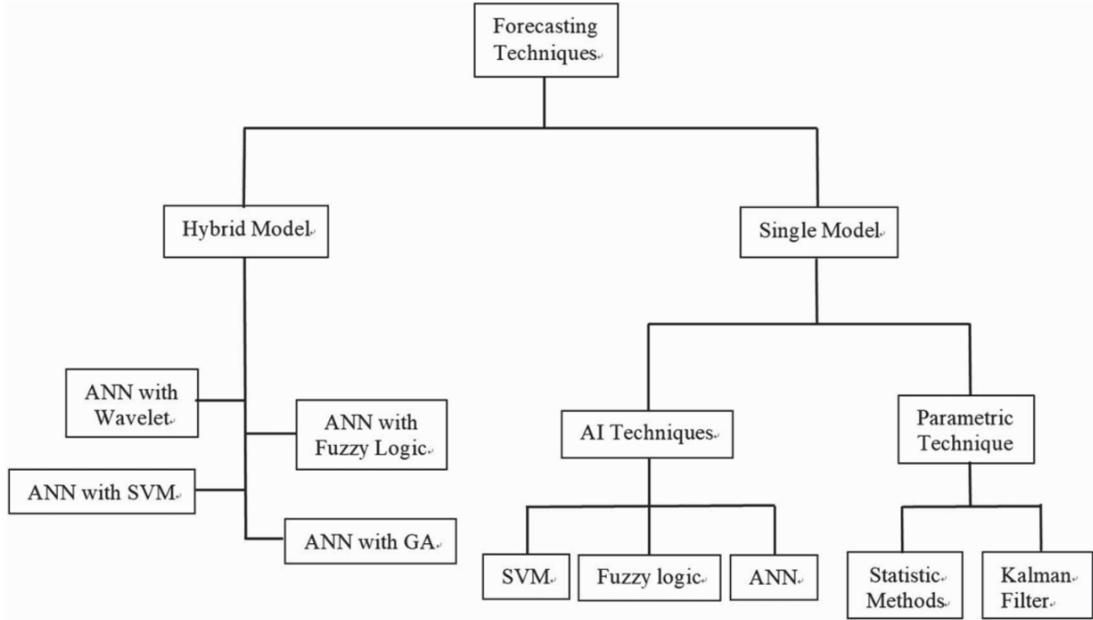


Figure 2.7 An overview of forecasting methods [Ma and Ma, 2018]

A classification of solar irradiance forecasting methods in terms of the type of data used as input is carried out and presented in Table 2.3 . We lay out physical methods, which are based on meteorological data, versus statistical methods that are based on historic data to use for predicting future trends. RNN, MLP, ANN and ARIMA are examples of statistical models. Their predictive capabilities makes these methods necessary, however, the challenges of increased accuracy as a function of the quality of historical data used to train the model remains.

Table 2.3 classification of solar irradiance forecasts based on approach. [Diagne et al., 2013]

Approach	Input	Q Examples
Physical	Meteorological data	TSI, NWP
Statistical	Historical data	RNN, MLP, ANN, ARIMA

Solar irradiance forecasting is described across a variety of time horizons from a very short term (less than 30 minutes) to long term (1–7 days) in Table 2.4. This is necessary for forecasting model tailored to operational needs which differ in scale along temporal axis (depending on if they are used for immediate grid management or long term planning) [Soman et al., 2010]. The table shows that the choice of

forecasting models is important, and that it depends on the time interval and purpose of application in the energy sector.

Table 2.4 classification depending on the temporal horizon for predicting. [Soman et al., 2010]

Time Horizon	Interval
Very short term	< 30 min
Short term	0.5 – 6 h
Medium term	6 – 48 h
Long term	1–7-day

Table 2.5 summarizes the effectiveness of various solar irradiance forecasting models using diverse approaches as measured by different performance metrics. As an illustrative example, we combine numerous ANN techniques into a Bagging ANN model and benchmark it against MLP, RBFNN, and RNN methods, with MAE of 17.4% and a surprisingly low MAPE of only 2.3%, indicating substantial accuracy over a short-term horizon. A similar MAE is echoed by the Multi-stage ANN model, showing the capabilities of such layered NN architectures to deal with the solar irradiance patterns' complexity.

On the contrary, models based on non-linear regression and pattern recognition techniques, which provide a resolution of 1 hours forecast, have higher MREs of 40% and 33.3% for 1 and 3 hours forecast respectively. This suggests that there might be a resolution vs. accuracy trade off for these models. Finally, a high MAE and MRE are noted in the Wavelet-RBPNN model with its day-long horizon, implying that although it may predict daily trends, its performance is very unreliable which is perhaps due to its sensitivity to the non-linear and volatile nature of solar irradiance data.

## 2.5 Machine Learning in Weather Forecasting

Recently, machine learning has been used heavily in the field of weather forecasting because of its inherent ability to develop new patterns from large, complex

Table 2.5 Summary of the forecasting models for solar irradiance that were examined.

Approach		Horizon	Benchmark methods	Results
Bagging [Choi and Hur, 2020]	ANN	1 day	MLP, RBFNN, RNN	17.4% (MAE), 2.3% (MAPE)
Multi-stage [Gheouany et al., 2023]	ANN	1 day	ANN	17.43% (MAE)
Non-linear regression and PR [Alizamir et al., 2020]		1 h, 3 h	Regression, ARIMA, ANN	40% (MRE, 1 h), 33.3% (MRE, 3 h)
Wavelet-RBPNN [Zayed et al., 2022]		1 day	RBPNN	74.5% (MAE), 77.61% (MRE)

datasets. This is also an ability to deal and learn from messy, and multi faceted data. This is a rapidly developing area of meteorological research weather, as ML is such a powerful tool for predicting the weather patterns.

As in [Khan et al., 2020], machine learning algorithms can detect subtle patterns in historical weather data which can, in turn, be used to predict how future weather will be. These algorithms are very complex and consider hundreds of variables such as temperature, humidity, wind speed and wind direction, atmospheric pressure and many, many more. So they can train ML models with past meteorological data, and have it learn to accurately predict future weather conditions. Moreover, machine learning can be used to enhance the accuracy of weather simulation as model parameter optimization and calibration [Sharifzadeh et al., 2019].

### 2.5.1 Types of Machine Learning Algorithms

In weather forecasting, several types of machine learning algorithms are used, each with its (own) strengths and weaknesses. Supervised, unsupervised and deep learning are the main categories of ML methods [Diagne et al., 2013].

**Supervised Learning Algorithms:** Such algorithms are built using a training dataset and used to estimate results using the acquired patterns. Field including Linear Regression, Decision Trees, Random Forest, Support Vector Machines also fall under

this category. Models that integrate decision trees and therefore nonlinear relationships and interactions among variables, like Random Forest, might accurately handle the present study's patterns.

**Unsupervised Learning Algorithms:** They can be used to discover structure in data that hasn't been pre-tagged or categorized in any way. K-Means and Hierarchical Clustering are the examples of machine learning methods which help Meteorology find groups of similar weather patterns.

**Deep Learning Algorithms:** Applications of CNN and RNN are for handling large data structures along with complex structures. They have been applied in domains of image recognition to classify the clouds pattern, and sequence prediction in time-series meteorological data.

### 2.5.1.1 Deep Learning for Time-Series Forecasting

This branch of machine learning, called deep learning, is oriented towards so called artificial neural networks, but in particular its 'deep' networks (cf Figure 2.8) where each layer consists of several nodes between the input and output. The layers are able to capture complex, high level features [Patterson and Gibson, 2017], and join these relating to the data. As deep learning is able to deal with large amount of high dimensional data and learn very complex temporal dependencies [Tomin et al., 2019], it can be particularly effective at time series forecasting.

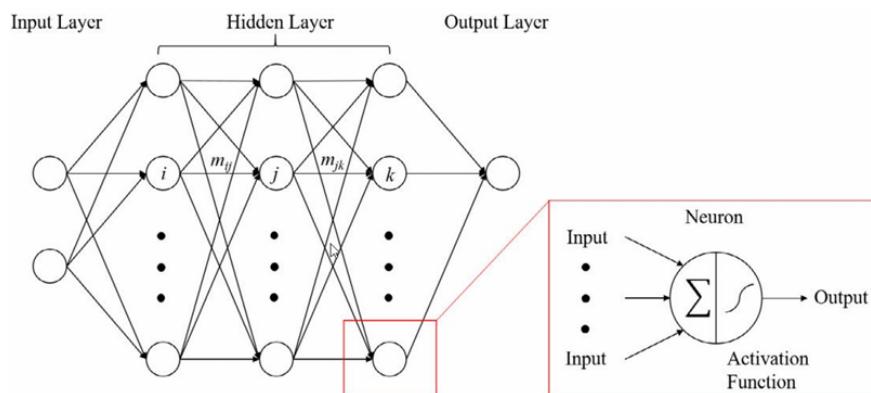


Figure 2.8 Deep learning general structure. [Chen et al., 2019]

### 2.5.1.2 Neural Network Models

Radial Basis Function Neural Network: Radial Basis Function Neural Networks (RBFNNs) have been effectively applied in a variety of practical scenarios, outperforming Multi-Layer Perceptron Neural Networks (MLPNNs) due to their lesser demand for computational resources and time. Some of these applications encompass the prediction of chaotic time-series, speech recognition, and categorization of data [He et al., 2019]. This capability is exemplified in Figure 2.9, showcasing the RBFNN's architecture.

It's worth highlighting that, given a sufficient volume of hidden units, an RBFNN stands out as a universal estimator, capable of approximating any type of continuous functions [Adcock and Dexter, 2021].

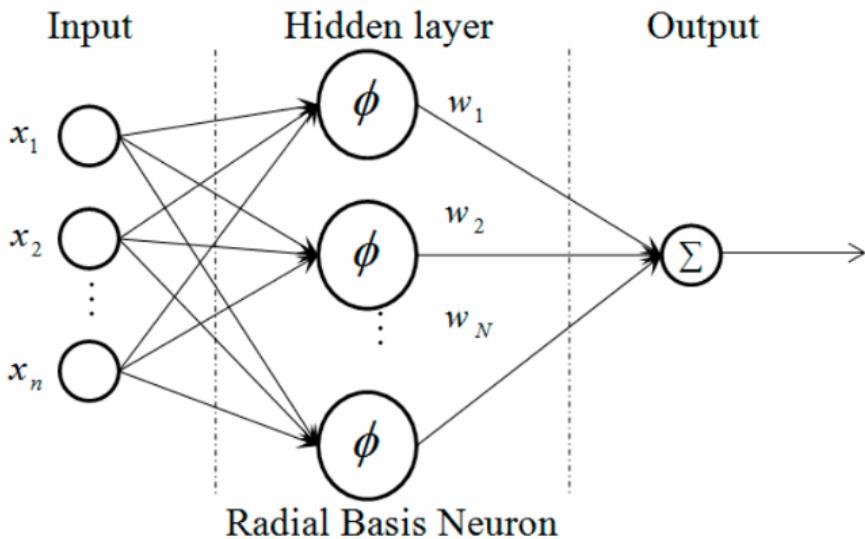


Figure 2.9 RBFNN's architecture. [He et al., 2019]

Multi Layered Perceptron Neural Network: Multi-layer Perceptron's (MLPs) have proven to be efficacious in addressing a wide array of complex and distinct problems. This is achieved through an initial supervised training phase that leverages the error back propagation algorithm and an error correction learning rule [Desai and Shah, 2021]. Multi-layer Perceptron's have proven to be efficacious in addressing a wide array of complex and distinct problems. This is achieved through

an initial supervised training phase that leverages the error back propagation algorithm and an error correction learning rule.

The backward pass is characterized by the adjustment of synaptic weights in alignment with an error correction rule. The error signal, calculated as the difference between the actual output and the desired value, is then disseminated backward through the network, against the flow of synaptic connections [Wani and Thagunna, 2024].

While MLPNNs can have multiple hidden layers as in Figure 2.10, K.M. Hornik's research [Hornik et al., 1989], suggests that even a neural network with a solitary hidden layer possesses the capability to approximate a function of arbitrary complexity.

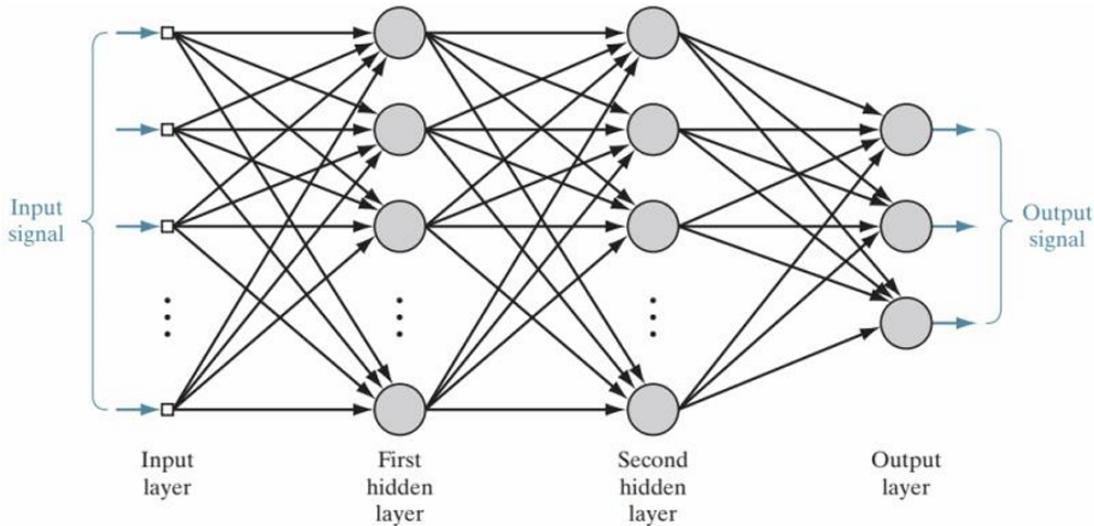


Figure 2.10 Architecture of a multilayer perceptron with three outputs and two hidden layers [Wani and Thagunna, 2024]

**Neural Network Ensemble:** A Neural Network Ensemble, often referred to as an ensemble, is a learning paradigm where multiple neural networks are trained to solve the same problem. The main idea behind this approach is to exploit the diversity among the networks to get a more robust and better generalizing model [Chen et al., 2019].

Ensemble methods leverage the fact that a group of 'weak learners' can come together to form a 'strong learner'. Each neural network in the ensemble makes a prediction (or vote), and the ensemble combines these predictions to make a final forecast [Lee et al., 2009], typically by simple averaging or voting. The intention is to improve the predictive performance and robustness over a single model.

Neural network ensembles can be particularly effective when the individual models in the ensemble are significantly different or independent from each other [Desai and Shah, 2021]. This diversity can be induced by using different architectures or learning algorithms, training on different subsets of the training data, or using different initial random weights.

The architecture of a Neural Network Ensemble (NNE) is a set-up where multiple neural networks are grouped together to solve a problem as shown in Figure 2.11. The individual networks, or "ensemble members," operate in parallel and independently from each other. They are typically trained on the same problem but may have different initial conditions, structures, or training data subsets, creating a diversity of predictions.

The ensemble's final output is typically obtained by aggregating the individual predictions of its members, often through simple techniques such as voting, averaging, or weighted averages [Desai and Shah, 2021]. This setup allows the ensemble to take advantage of the diverse predictive capabilities of its members, often leading to better performance compared to a single network.

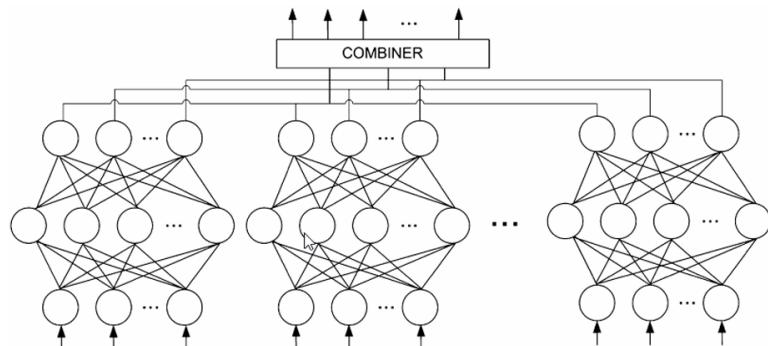


Figure 2.11 Architecture of the neural network ensemble [Lee et al., 2009]

**Recurrent Neural Network:** Recurrent Neural Networks (RNNs), while being a subset of feed-forward neural networks, stand apart due to their distinctive ability to transmit information across temporal intervals [60]. RNNs support both parallel and sequential computations, essentially embodying the computational ability of conventional computers. However, these networks are more akin to the human brain, a vast feedback network of interconnected neurons [Mienye et al., 2024], which translates continuous sensory input into a sequence of beneficial motor responses. The human brain serves as an exceptional model, resolving numerous challenges that present machines struggle to overcome.

RNNs process each vector from a sequence of input vectors one by one, allowing the network to maintain its state while modeling each input vector across the input vector window. This hallmark ability to model the time dimension sets RNNs apart [Hewamalage et al., 2021].

Expanding upon this unique architecture, RNNs introduce the notion of recurrent connections, which these connections (or recurrent edges) span across consecutive time-steps (e.g., from a previous time-step) as shown in Figure 2.12, providing the model with a sense of time. Traditional connections in RNNs do not form cycles. However, recurrent connections can create cyclical paths, including those looping back to the original neurons at future time-steps [Reza et al., 2022].

As input is transmitted through a recurrent network at each time-step, nodes receiving input along recurrent edges process input activations from both the current input vector and the network’s previous state’s hidden nodes. The output is computed based on the hidden state at that specific time-step [Hewamalage et al., 2021]. Thus, the previous input vector at the last time-step can influence the current output at the present time-step, facilitated by the recurrent connections.

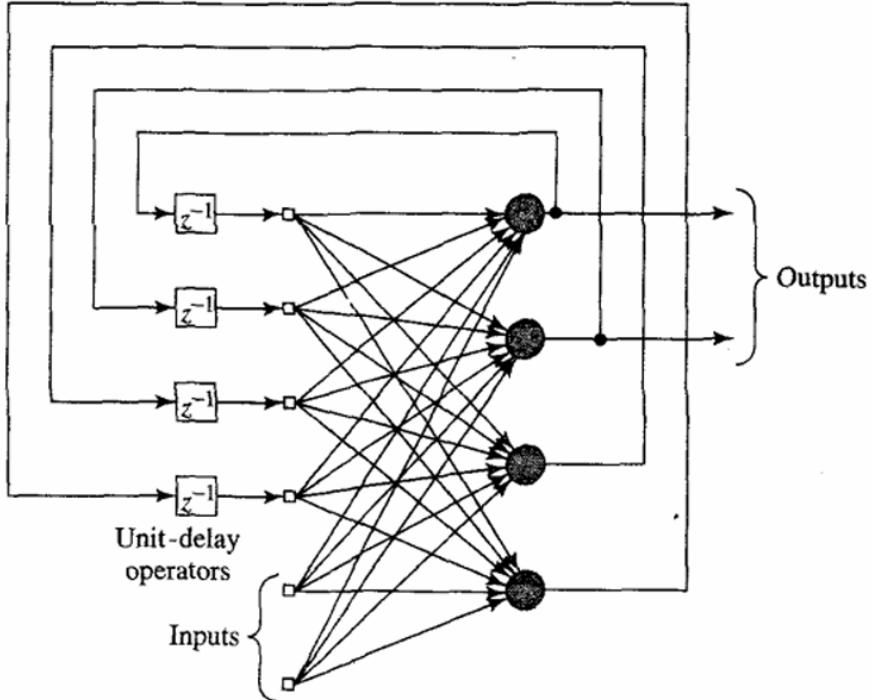


Figure 2.12 Architecture of the neural network ensemble[Lee et al., 2009]

Recurrent Neural Networks (RNNs) are neural networks that have at least one feedback loop in essentially. We classify these feedback mechanisms as local or global. The structure of RNN usually begins as shown in Figure 2.13 with a neural network of the type Multi Layer Perceptrons Neural Network (MLPNN) and the feedback can take many forms. Taking feedback out of the output neurons to the input layer, or giving feedback from the hidden neurons in the network back to the input layer, is just one example. And both scenarios can co-exist. In the case of MLPNN with two or more hidden layers the potential number of feedback forms increases correspondingly. RNNs are able to evolve state representations, making them a great nonlinear prediction and modeling framework [Hewamalage et al., 2021], due to the inclusion of feedback mechanisms.

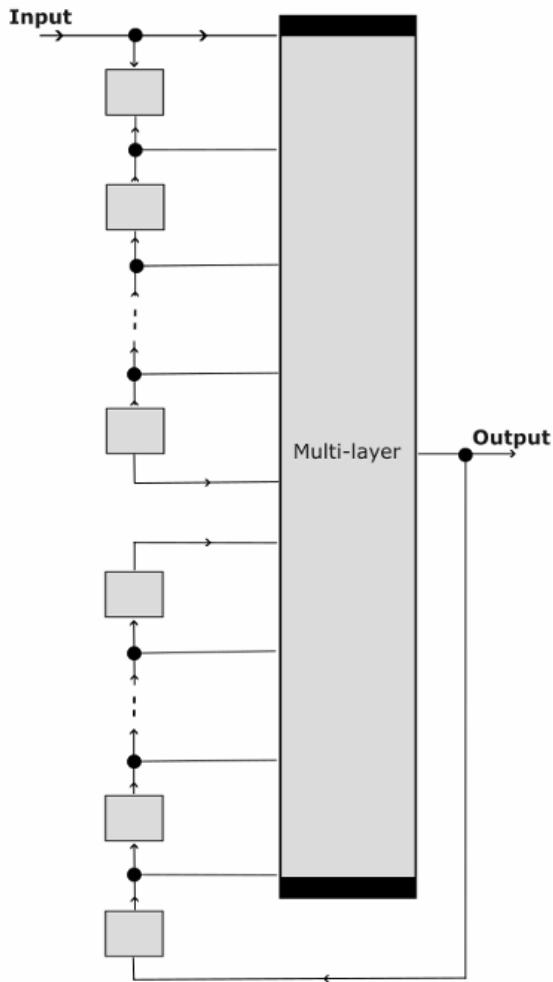


Figure 2.13 structure of NARX-RNN

Table 2.6 summarizes different neural network models with architectures and possible applications. This compares with the Radial Basis Function Neural Network (RBFNN) for its three layer architecture excellent for function approximation, and in turn with the Multi-layer Perceptron Neural Network (MLPNN) rich in the hidden layers and capable of building complex data interactions. Recurrent Neural Networks (RNNs) are also described in the table; RNNs are able to process sequential data, due to the recurrent connections which allow information to persist. In the thesis, the architecture of each model informs which forecasting tasks are appropriate for which models.

Table 2.6 Neural Network Models

<b>Neural Network Type</b>	<b>Architecture</b>
Radial Basis Function Neural Network (RBFNN)	RBFNN structure is composed of three layers. The input layer connects the network to the external environment. The hidden layer, which is the only hidden layer in the network, applies a nonlinear transformation from the input space to the hidden space, which is often high dimensional. The output layer is linear, supplying the response of the network to the pattern applied to the input layer.
Multi-layered Perceptron Neural Network (MLPNN)	MLPNNs have an input layer, one or more hidden layers, and an output layer. The input layer receives the input vector, and its effect propagates through the network layer by layer to produce an output. During the training phase, synaptic weights are adjusted in accordance with an error correction rule. In general, MLPNNs can have several hidden layers.
Artificial Neural Network Ensemble (ANNE)	ANNEs are a collection of multiple individual neural networks. Each network is trained independently, and their predictions are later combined. They provide a way to mitigate the instability of single neural network models. Changes in the architecture of the networks in the ensemble, or changes in the training data, can affect the performance of the ensemble.
Recurrent Neural Network (RNN)	RNN architecture introduces the concept of recurrent connections. These connections span adjacent time-steps, giving the model the concept of time. At each time-step of sending input through a recurrent network, nodes receiving input along recurrent edges receive input activations from the current input vector and from the hidden nodes in the network's previous state. The output is computed from the hidden state at the given time-step.

A comprehensive comparison of recent advancements in solar irradiance (SI) forecasting techniques with previous work is reported in Table 2.7 in order to assess the SI forecasting landscape. It encapsulates a range of methodologies from LSTM based models to support vector machines (SVM) with their lowest error rates, along with their largest drawbacks, presenting a tradeoff between accuracy and complexity.

Compared to the LSTM-TCM and LSTM-RNN models, for example, which achieved normalized root mean square error (nRMSE) rates of 6.29% and 3.89%, respectively, these models also offer higher complexity and difficulties in handling

dynamic parameters, limiting their use in real time forecasting situations. Secondly, simpler models like the ANN cited in ref. [Chung, 2020] produce a respectable RMSE of 2.70 kWh/m<sup>2</sup>, but the fact that they require current weather observation data, which may not make good enough estimation of future values, is a complicating factor.

Although the SVM approach is very complex to calculate, in fact the RMSE achieves a 2.78 MJ/m<sup>2</sup>, indicating that the approach is in fact, highly suitable for accurate forecasting. On the other hand, the hSBFM model with the lowest nRMSE of 1.43 is critiqued for not taking into account the future weather or solar conditions that could be important for long term accuracy of prediction.

In addition, although the SVM-BAT and EMD-SCA-ELM are highly efficient, they consume a lot of computation and training, which is not practical for direct deployment.

At 6.34 nRMSE, the CNN-LSTM model also suffers from computational complexity, but that is avoidable since optimization is required. Although the nRMSEs of 21.2 and 11.81 are achieved by using Deep ConvNets and NAR respectively, their nonlinear complexity and parameter sensitivity, which may affect the adaptability to different conditions of solar irradiance, limit them.

The analysis highlights continued effort as well as challenges in the world of SI forecasting and suggests a healthy trade off between model complexity, computational efficiency, and the forecast precision. The strengths of each method as well as limitations are unique, rendering an informed selection of the most appropriate model for practical solar energy management applications essential.

Table 2.7 A comprehensive analysis contrasting SI forecasting techniques with prior studies.

Method	Ref.	Year	Country	Lowest Error	Limitation
LSTM-TCM	[Wang et al., 2020]	2020	Australia	nRMSE = 6.29%	High Complexity
LSTM-RNN	[Zafar et al., 2021]	2021	United States	nRMSE = 3.89%	Difficult to heal with dynamic parameters
ANN	[Chung, 2020]	2020	South Korea	RMSE = 2.70 kWh/m <sup>2</sup>	The model estimates solar insolation based on current weather data, not future forecasts.
SVM	[Sutarna et al., 2023]	2021	Brazil	RMSE = 2.78 MJ/m <sup>2</sup>	High Complexity
hSBFM	[Sangrody et al., 2020]	2020	United States	nRMSE = 1.43%	It doesn't account for upcoming changes in weather or solar conditions.
SVM-BAT	[Feng et al., 2020]	2020	China	RMSE = 1.694 MJ/m <sup>2</sup>	Computational Complexity
EMD-SCA-ELM	[Behera and Nayak, 2020]	2020	India	nRMSE = 1.88%	Need for heavy training
CNN-LSTM	[Qu et al., 2021]	2021	China	nRMSE = 6.34%	Computational Complexity
Deep ConvNets	[Wen et al., 2020]	2020	United States	nRMSE = 21.2%	Nonlinear complexity
NAR	[Takilalte et al., 2022]	2022	Algeria	nRMSE = 11.81%	Parameter Sensitivity

### 2.5.2 Challenges and Limitations

Nonetheless, emerging with the application of machine learning in weather forecasting there are several challenges and limitations. One of the critical factors defining the model's accuracy is the extent to which the training dataset is informative and large enough [Patterson and Gibson, 2017]. Lack of or incorrect data results in wrong predication. However, these algorithms are complex, especially, in terms of the amount of computational work involved particularly when large data sets are in use. Moreover, it is necessary to mention that despite the fact that machine learning models improve the forecast precision, commonly, they act as black boxes [Murphy, 2012]. The problem with this situation is that it becomes very difficult to understand how they came up with such predictions and this may pose a lot of challenges especially when their application is very important and the details of their thinking must be brought to light.

Furthermore, it is essential to acknowledge that weather patterns exhibit intricate behavior, shaped by several interacting elements. This implies they may

continue to have difficulty properly forecasting the subsequent number. To forecast future occurrences in instances when these occurrences are influenced by variables not represented in the training data.

Extreme weather events, more prevalent due to climate change, are often caused by uncommon combinations of elements not represented in historical data. Consequently, such catastrophes may be inadequately forecasted by machine learning algorithms, leading to extensive ramifications for disaster prevention and management.

In addition, even though machine learning models deal well with high dimensional data, choosing the proper features (variables) to predict would require a daunting task. However, inclusion of extraneous characteristics in the data set may cause a suboptimal model performance [James, 2013]. On the other hand, dropping characteristics thought of as insignificant can cause the removal of essential predictive info.

A significant issue in using machine learning for weather forecasting is overfitting. This transpires when a model overfits the training data, resulting in poor performance on novel data. An overfitted model has a greater number of parameters than the quantity of data. Ultimately, it identifies the noise in the training data that may adversely impact its capacity to generalize from previously encountered data to novel data [Murphy, 2012].

In the context of weather forecasting, an overfitted model might predict past weather patterns with remarkable accuracy but fail miserably when asked to predict future weather. This is because the model has not truly 'learned' the underlying patterns in the weather data; instead, it has merely 'memorized' the training data [Murphy, 2012]. Therefore, the model becomes unable to make accurate predictions when faced with new data or scenarios that differ from what it has seen during training.

Techniques such as cross-validation, regularization, early stopping, and pruning can be used to prevent overfitting. Also, it is always essential to have a holdout validation or test dataset to evaluate the model's performance on unseen

data [Patterson and Gibson, 2017]. This helps ensure that the developed model can generalize well and provide accurate forecasts on new data.

## 2.6 Chapter Summary

Research in solar irradiance forecasting has seen a variety of approaches. Traditional methods, including statistical and rule-based models, often fall short in accurately predicting solar energy due to the intricate and dynamic nature of weather patterns affecting solar irradiance. Recognizing these limitations, recent studies have shifted towards more sophisticated techniques. This project specifically explores the integration of machine learning algorithms, notably a hybrid model combining Nonlinear Autoregressive with Exogenous Inputs (NARX) and Long Short-Term Memory (LSTM) networks, to enhance daily solar irradiance forecasting accuracy in microgrids.

## **CHAPTER 3**

### **RESEARCH METHODOLOGY**

#### **3.1 Introduction**

This chapter outlines the comprehensive research methodology applied to achieve the objectives of this study, which focuses on improving solar irradiance forecasting through the hybridization of NARX and LSTM models. Research methodology serves as a systematic framework of rules, procedures, and techniques to guide the research process and ensure its alignment with the objectives established in Chapter 1. Selecting an appropriate methodology is critical to effectively navigate the various phases required to answer the research questions.

This chapter presents a detailed account of the methods and instruments used in this study, offering a clear picture of the entire process. It includes discussions on the identification and analysis of forecasting parameters, the collection and preprocessing of meteorological data, and the development and optimization of the hybrid NARX-LSTM model. Additionally, the chapter elaborates on the input sensitivity analysis and evaluation metrics employed to validate the model's performance. Each stage of the research process is systematically addressed to ensure the reliability and effectiveness of the proposed forecasting framework.

#### **3.2 Research Framework**

The methodology unfolds across a series of systematically designed stages as shown in Figure 3.2, each integral to the progression and integrity of the research. The approach commences with the identification of pivotal weather parameters essential for precise solar irradiance forecasting. This initial phase involves a meticulous analysis of various meteorological variables and their potential impact on forecasting

accuracy, underpinned by a thorough review of relevant literature and domain expertise.

Subsequently, the focus shifts to the selection of an appropriate machine learning model. The decision to employ a hybrid NARX-LSTM model is grounded in its theoretical suitability for capturing the complex, nonlinear relationships inherent in meteorological data and its proven efficacy in time-series forecasting. Undergoes a preliminary stage of model tuning and input sensitivity analysis. This tuning optimizes the model's performance, addressing the balance between accuracy and computational demands. The input sensitivity analysis, conducted prior to the training phase, assesses the impact of each weather parameter on the model's output, providing insights into the relative importance of each variable.

Following model selection, the methodology advances into the data collection phase. Here, the specifics of data collocation, including the choice of Johor Bahru, Malaysia, as the geographical focus. The chapter will detail the process of gathering, cleaning, and preprocessing the data, ensuring its alignment with the identified forecasting parameters and its readiness for model training.

The subsequent stage encompasses the training of the NARX-LSTM model, where the nuances of network architecture, training protocols, and optimization strategies are expounded. The validation of this model is critical and is conducted using robust statistical methods to assess and ensure its predictive performance.

The final stage culminates in the result validation process. Here, the outputs of the NARX-LSTM model are meticulously cross-verified against actual meteorological data to evaluate the accuracy and reliability of the forecasts. Additionally, a comparative analysis with standalone models is conducted, showcasing the hybrid model's enhanced predictive capabilities and efficiency. This stage is pivotal in affirming the effectiveness of the hybrid model and its applicability in real-world scenarios.

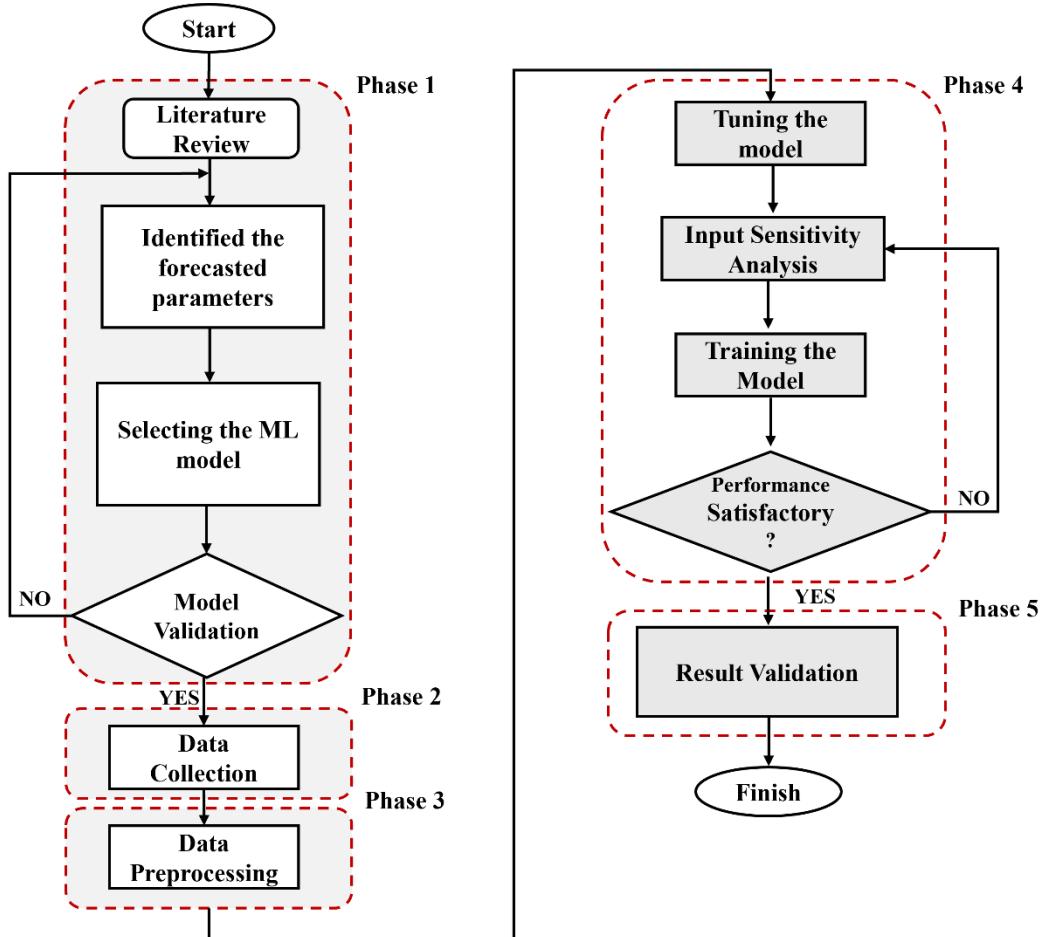


Figure 3.1 Research framework

### 3.3 Data Analysis

This section comprehensively outlines the systematic procedures undertaken to transform raw meteorological data into an optimized dataset suitable for accurate solar irradiance forecasting. It begins with the identification of key forecast parameters, a process that integrates insights from literature review and domain expertise to determine the most influential meteorological factors affecting solar irradiance. These parameters form the foundation of the forecasting model, ensuring it captures the critical interactions between environmental variables and solar energy dynamics.

Following parameter identification, the data analysis process progresses to data collection, where historical meteorological data spanning a significant temporal range is gathered. This data is meticulously sourced to ensure accuracy, reliability, and relevance to the study's geographical focus. The collected dataset is then subjected to a comprehensive preprocessing phase, which includes data cleaning, normalization, and feature engineering. These steps address inconsistencies, scale differences, and the temporal characteristics of the data, ensuring its readiness for effective model training.

Moreover, this section delves into the exploration of data distributions and relationships through histograms and scatter plots, providing valuable insights into the behavior of the meteorological variables over time. These analyses not only guide the preprocessing techniques but also highlight seasonal and diurnal trends that are critical for time-series forecasting. The structured approach detailed in this section lays the groundwork for developing a robust and reliable predictive model, enabling the effective forecasting of solar irradiance under varying environmental conditions.

### **3.3.1 Identification of Forecast Parameters**

The Identification of Forecast Parameters phase integrates detailed data analysis with comprehensive literature review. This stage finalizes the meteorological parameters crucial for forecasting solar irradiance: air temperature, cloud attenuation, precipitation rate, precipitable water, relative humidity, air pressure, wind direction, wind speed, and dewpoint temperature. These variables are pivotal due to their substantiated effect on solar energy as identified through historical data trends. This meticulous selection process ensures that the forecasting model incorporates the most significant and impactful weather elements.

Air temperature is a critical factor, as it directly correlates with the efficiency of photovoltaic cells, influencing the overall solar energy generation. Cloud attenuation is pivotal due to its impact on the amount of solar radiation reaching the Earth's surface, directly affecting the variability of solar irradiance. The role of precipitation rate is highlighted in research exploring its effects on solar irradiance

levels during rainfall events, while precipitable water in the atmosphere is key due to its role in absorbing and scattering sunlight, impacting the solar irradiance levels. Relative humidity, indicative of atmospheric moisture, influences cloud formation and the diffusion of solar radiation. Air pressure, a determinant of weather patterns, affects the clarity of the sky and, consequently, solar irradiance. Wind direction and speed are included for their roles in cloud movement and the dispersal of airborne particulates, which can significantly alter the atmospheric clarity and solar radiation reception. Finally, dewpoint temperature is considered for providing insights into atmospheric moisture levels, thus influencing potential cloud cover and solar irradiance.

The selection of these parameters is supported by their established impact on various aspects of solar energy dynamics and a comprehensive review of existing literature, ensuring that the chosen variables are not only relevant but also empirically validated for their influence on solar irradiance forecasting. Understanding these influences is crucial for accurately predicting solar irradiance, as they collectively embody the complex interactions within the Earth's atmosphere that affect solar energy generation. This careful identification of forecast parameters forms the foundation for the subsequent modeling and analysis, ensuring a comprehensive approach to predicting solar irradiance.

Incorporating the relationship between various weather parameters and solar irradiance or Global Horizontal Irradiance (GHI) into the Identification of Forecast Parameters phase enriches the model's contextual understanding as evidenced in Figure 3.2. Air temperature shows a positive correlation with GHI, suggesting sunnier conditions with rising temperatures. While dewpoint temperature has a moderate positive correlation, its subtler influence compared to air temperature is noted. Precipitation rate's slight positive relationship with GHI indicates that intermittent rain doesn't always hinder solar irradiance. Surface pressure's weak positive trend aligns with the expectation of higher GHI during high-pressure conditions, indicative of clear skies.

Conversely, precipitable water's negative correlation implies potential cloud formation that may reduce GHI. Cloud opacity strongly inversely affects GHI,

affirming its role in blocking sunlight. Lastly, relative humidity strongly negatively correlates with GHI, typically signaling overcast conditions leading to reduced solar exposure. These relationships underscore the complexities of predicting solar irradiance and the need for a nuanced model that accounts for these varied meteorological impacts.

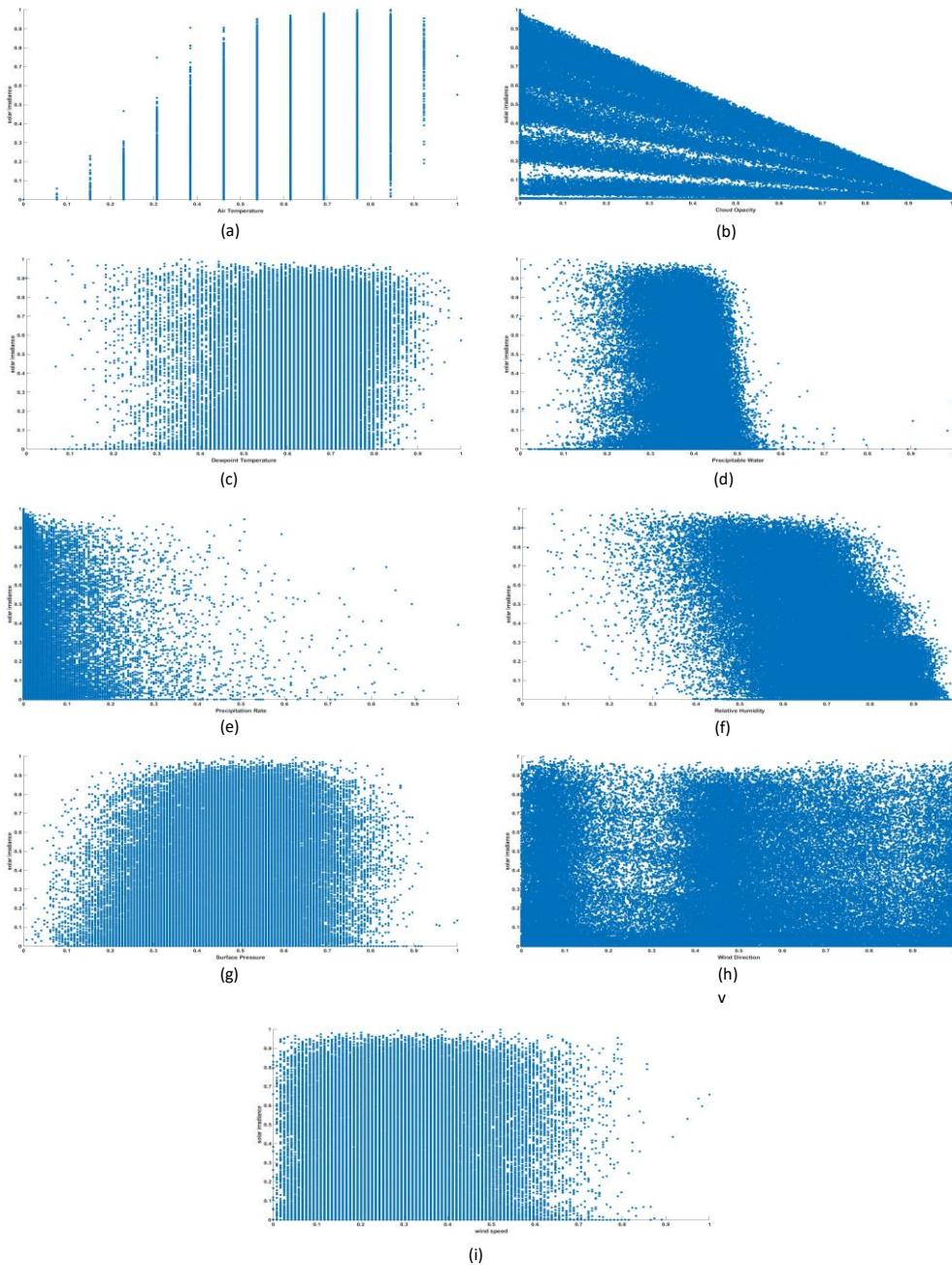


Figure 3.2 scatter plots of (a) Air Temperature, (b) Cloud Opacity, (c) dewpoint temperature, (d) precipitable water, (e) precipitation rate, (f) relative humidity, (g) Surface Pressure, (h) Wind Direction and (i) wind speed - against the SI.

### **3.3.2 Data Collection**

Following the meticulous identification of forecast parameters, the research progresses to the data collection phase. This critical stage involves the systematic collection of data corresponding to the identified meteorological variables, essential for the effective forecasting of solar irradiance.

The data encompasses a comprehensive range of parameters including air temperature, cloud attenuation, precipitation rate, precipitable water, relative humidity, air pressure, wind direction, wind speed, and dewpoint temperature as can be seen in Figure 3.3. This dataset is sourced from Det Norske Veritas (DNV), a globally recognized authority in quality assurance and risk management, ensuring the reliability and accuracy of the information. The data spans from January 1, 2007, to November 29, 2023, providing a substantial temporal coverage that enhances the robustness of the analysis. The readings are recorded on an hourly basis, offering a detailed temporal resolution that is critical for capturing the nuances of solar irradiance variability.

Located in Johor Bahru, Johor, Malaysia, the geographical setting of the data collection is significant. The region's diverse climatic conditions present an ideal environment for studying the impacts of various environmental variables on solar irradiance. The hourly readings, encompassing a wide array of weather conditions over an extended period, provide a rich dataset that is instrumental in developing a nuanced and comprehensive forecasting model.

The meticulous approach to data collection, emphasizing both the quality and the scope of the data, lays a solid foundation for the subsequent stages of preprocessing and analysis. By ensuring the data's accuracy and relevance, this phase plays a pivotal role in the overall success of the research, enabling the development of a reliable and effective solar irradiance forecasting mode.

air_temp	cloud_opacity	dewpoint	ghi	precipitation	precipitation	relative_humidity	surface_pressure	wind_direction	wind_speed	period_end
25	59.1	23.3	101	53.6	0	91.7	1004.1	316	3.8	2007-01-01T01:00:00Z
26	55	23.9	213	53.9	0	90.3	1004.8	313	3.4	2007-01-01T02:00:00Z
27	56	24.2	296	54.3	0	86.2	1005.2	317	3.6	2007-01-01T03:00:00Z
28	43	24.1	468	54.9	0	81	1005.2	324	4	2007-01-01T04:00:00Z
29	7.6	24.1	832	55.8	0	76	1004.7	327	4.1	2007-01-01T05:00:00Z
30	0	24.3	910	56.7	0.1	73.7	1003.9	328	4.2	2007-01-01T06:00:00Z
29	9.8	23.6	766	57.3	1.9	73.8	1003.1	327	4.2	2007-01-01T07:00:00Z
27	17.4	22.8	595	56.8	0.5	75.8	1002.1	335	4	2007-01-01T08:00:00Z
27	47.8	23.3	285	56.3	0.3	80.5	1001.4	346	4	2007-01-01T09:00:00Z
27	75.9	23.7	77	56.7	0.2	83.4	1001.6	343	3.9	2007-01-01T10:00:00Z
26	49.7	23.6	41	57.4	0.2	84.9	1002.4	339	3.8	2007-01-01T11:00:00Z
26	67	23.5	0	57.7	0.1	86.1	1003.2	340	3.7	2007-01-01T12:00:00Z
26	69.3	23.5	0	58	0.1	87.2	1004.1	332	3.9	2007-01-01T13:00:00Z
26	55.7	23.5	0	58.7	0.1	88.7	1004.9	317	4.6	2007-01-01T14:00:00Z
25	50.9	23.5	0	59.3	0.1	90.9	1005.4	308	4.9	2007-01-01T15:00:00Z
25	49.6	23.4	0	59.9	0	92.2	1005.6	301	4.7	2007-01-01T16:00:00Z
25	36.3	23.4	0	60.3	0.1	92.8	1005.5	301	4.6	2007-01-01T17:00:00Z
25	32.5	23.3	0	60.1	0.1	93	1004.9	299	4.6	2007-01-01T18:00:00Z
25	47.7	23.3	0	59.8	0	93	1004.2	288	5	2007-01-01T19:00:00Z
25	51.5	23.2	0	60	0	93	1003.5	282	5.5	2007-01-01T20:00:00Z
24	46.1	23.3	0	60.2	0.1	93.6	1003.1	286	5.6	2007-01-01T21:00:00Z
25	48	23.4	0	59.2	0.4	93.6	1003.3	300	5	2007-01-01T22:00:00Z
24	32.6	23.2	0	58.3	0.2	93	1003.5	309	4.6	2007-01-01T23:00:00Z

Figure 3.3 Historical data for 24 hours - January 1, 2007

### 3.3.3 Data Preprocessing

The data preprocessing stage is integral to the structuring of meteorological data for analysis and modeling. This phase commences with a thorough examination of variable histograms to distill key insights from data distributions. These insights underpin systematic procedures to rectify inconsistencies, normalize variable scales, and structure time-series data, ensuring the dataset's readiness for predictive modeling. The histogram analysis is pivotal, as it informs the treatment of outliers, guides feature selection, and confirms the robustness of the data across temporal variations from 2007 to 2023.

The analysis of histogram distributions for meteorological variables offers a nuanced understanding of the data's behavior as shown in Figure 3.4. The air temperature histogram displays a multimodal distribution, indicating seasonal temperature fluctuations. Recognizing these peaks is vital as they directly influence solar panel efficiency. Similarly, the right-skewed cloud opacity histogram suggests more frequent clear skies with occasional dense cloud cover, a key variable in

predicting irradiance variability. Meanwhile, the dewpoint temperature's normal distribution points to stable moisture conditions, affecting dew formation and subsequent solar irradiance levels.

Global Horizontal Irradiance (GHI) presents a pronounced left-skew, predominantly featuring lower irradiance values, which are essential in modeling the diurnal cycle and the impact of cloud coverage. Precipitable water's near-normal distribution underscores its role in the atmosphere's ability to absorb and scatter solar radiation, a significant factor in irradiance measurements.

The exponential decay observed in the precipitation rate histogram, where most data points indicate minimal precipitation, is crucial for modeling as heavy rainfall can significantly reduce irradiance. The left-skew in relative humidity, especially in higher values, reflects the humidity's impact on cloud formation and solar irradiance in tropical climates.

Surface pressure's normal distribution suggests consistent atmospheric conditions, valuable for forecasting weather patterns that affect irradiance. Wind direction's bimodal distribution can inform the model about prevalent wind patterns that influence cloud cover and irradiance levels. Lastly, the right-skewed wind speed histogram indicates generally low wind speeds but acknowledges the potential for occasional high-speed winds that could disperse clouds and particulates, affecting irradiance.

Each histogram informs specific aspects of the preprocessing, like outlier treatment and feature selection, ensuring the forecasting model is built on a dataset that accurately reflects environmental conditions.

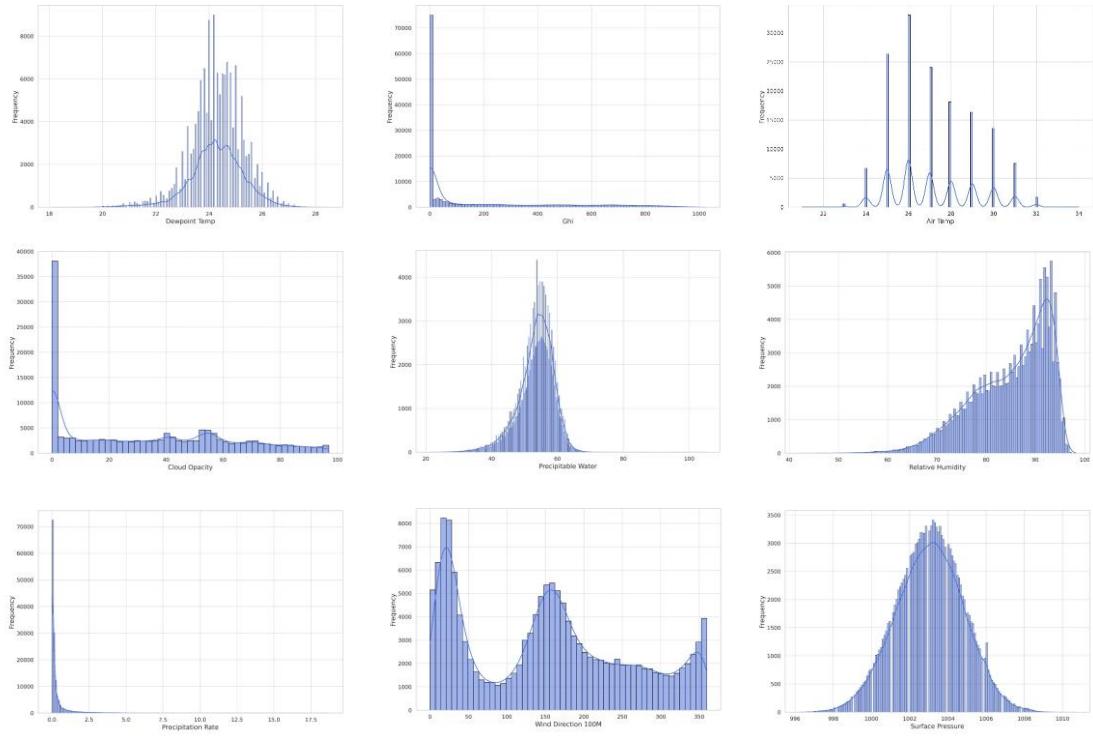


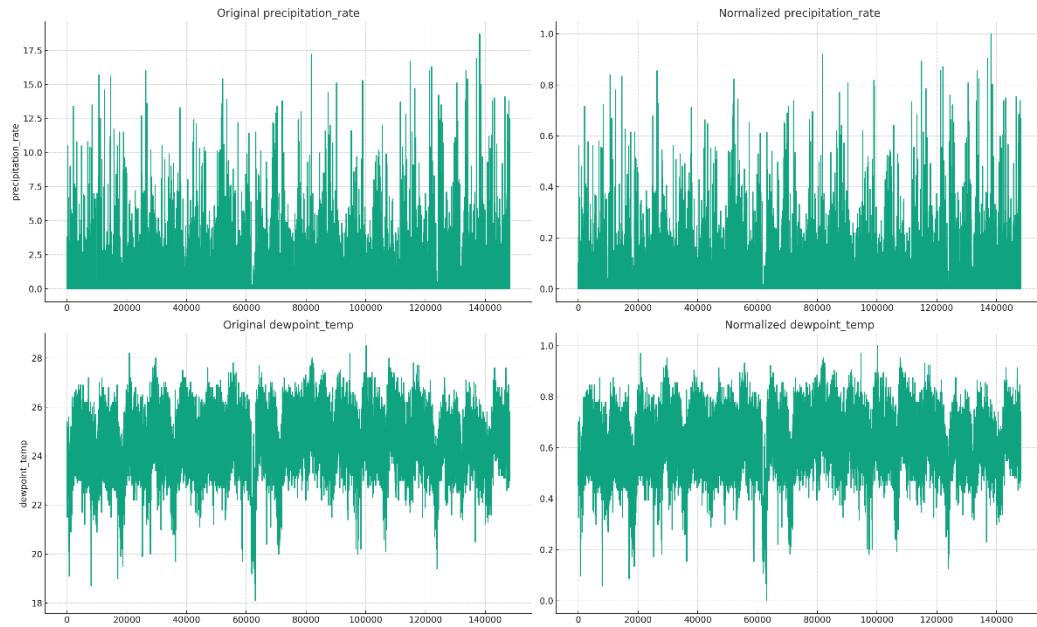
Figure 3.4      Histograms Distributions

During the data cleaning process, several inconsistencies and gaps were identified and rectified. The histograms distributions guided the data cleaning process, where outliers and anomalies were addressed. For instance, extreme values in cloud opacity and wind speed were capped to avoid undue influence on the model. Normalization was applied to standardize the range of independent variables, preserving their distributional characteristics without letting any single variable dominate due to scale differences. This was achieved using the Min-Max normalization method.

$$\text{Normalized Value} = \frac{\text{Value}-\text{Min}}{\text{Max}-\text{Min}} \quad (3.1)$$

This transformation scales the data within a range of 0 to 1, ensuring that each parameter contributes proportionately to the model's training process. Figure 3.5

illustrates this, demonstrating that the dynamics of the parameters remain consistent post-process.



**Figure 3.5      Comparison of Original and Normalized Meteorological Parameters**

Feature engineering involved the incorporation of temporal patterns observed in the histograms in Figure 3.4, specifically the time of the month and hour of the day. These features were derived to capture the cyclical nature of solar irradiance across different timescales. This addition is expected to provide the model with finer granularity in understanding the diurnal and monthly variations in solar irradiance.

The data was divided into training, validation, and testing sets, with 70% allocated for training (spanning from 2007 to 2019), 15% for validation (covering 2020 to 2021), and 15% for testing (from 2022 to November 29, 2023). This partitioning was designed to provide a comprehensive training dataset while allowing for effective model validation and testing on more recent data, thereby ensuring the model's robustness and adaptability to recent climatic trends.

### **3.4 Model Development**

The Model Development section comprehensively presents the structure and integration of the Nonlinear Autoregressive with Exogenous Inputs (NARX) neural network and the Long Short-Term Memory (LSTM) neural network, both of which are pivotal for enhancing the accuracy of solar irradiance forecasting. This section will elucidate the individual characteristics of the NARX and LSTM models, detailing their unique strengths in pattern recognition and temporal data analysis. Moreover, the synthesis of these models into a cohesive hybrid system is explored, highlighting the synergistic effects of this combination on forecasting proficiency. The section will also discuss the methodological steps taken to configure and optimize the hybrid model, ensuring that it not only captures the intricate dynamics of solar irradiance but also generalizes well to unseen data, thereby contributing to the advancement of predictive modeling in renewable energy systems.

#### **3.4.1 Nonlinear Autoregressive with Exogenous Inputs (NARX) Neural Network:**

The Nonlinear Autoregressive with Exogenous Inputs (NARX) Neural Network is a critical component of the hybrid NARX-LSTM model designed for solar irradiance forecasting. The NARX model excels in mapping the relationships between input and output time-series data, predicting future values based on this historical information. It operates on a functional relationship that involves both the external inputs  $u(t)$  and the previous values of the output series  $y(t)$ . This capability makes the NARX model highly effective in capturing the non-linear dynamics often present in meteorological data, which is essential for accurate solar irradiance predictions. The standard form of the NARX model is given by:

$$y(t) = f[(y(t-1), \dots, y(t-d_y), u(t), u(t-1), \dots, u(t-d_u))] \quad (3.2)$$

where in the given text,  $u(t)$  and  $y(t)$  within the real number set represent the input and output of the model at a specific discrete time step,  $t$  respectively. This model includes a feedback loop which serves to enhance the model's responsiveness to past data. Figure 3.6 provides a visual representation of how the NARX algorithm functions.

As depicted in Figure 3.6, the NARX algorithm is composed of a two-layer feed-forward network. This includes a linear transfer function in the output layer and a sigmoid function ( $\sigma$ ) in the hidden layer, as calculated and described in [80].

$$\sigma(x) = \frac{1}{1 + \exp(-\sigma)} \quad (3.3)$$

A typical NARX neural network includes an input layer, one or more hidden layers with a nonlinear activation function (commonly a hyperbolic tangent sigmoid function), and a linear output layer. The network incorporates feedback connections from both the output and input layers via tapped delay lines that store previous values of the input  $u(t)$  and output  $y(t)$  sequences.

The input to the hidden layer at any time  $t$  is a combination of the current and delayed inputs, as well as the feedback from the previous outputs. The output of the network at time  $t$ , denoted by  $\hat{y}(t)$ , is then calculated as follows:

$$\hat{y}(t) = W_o H + b_o \quad (3.4)$$

The NARX network is typically trained using a variant of the backpropagation algorithm, such as the Levenberg-Marquardt algorithm, which is well-suited for nonlinear least squares problems. During training, the network operates in an open-loop mode, using actual past output data for the feedback connections. Once trained, the network can be switched to a closed-loop mode for simulation, using its own predictions as feedback to generate future outputs as shown in Figure 3.7.

Despite its advantages, the NARX neural network's efficiency can be further enhanced by omitting the retention of past outputs, which notably diminishes the computational demand. However, as a variant within the recurrent neural network (RNN) family, the NARX model is not entirely immune to the vanishing gradient phenomenon. This issue arises when, beyond a certain number of inputs, the RNN ceases to retain new information, leading to a plateau in learning and, consequently, a decline in predictive precision. Such a challenge is manifested during the training phase, particularly when gradient descent algorithms begin to falter in maintaining the influence of input data over extended sequences, resulting in what is known as the diminishing memory problem.

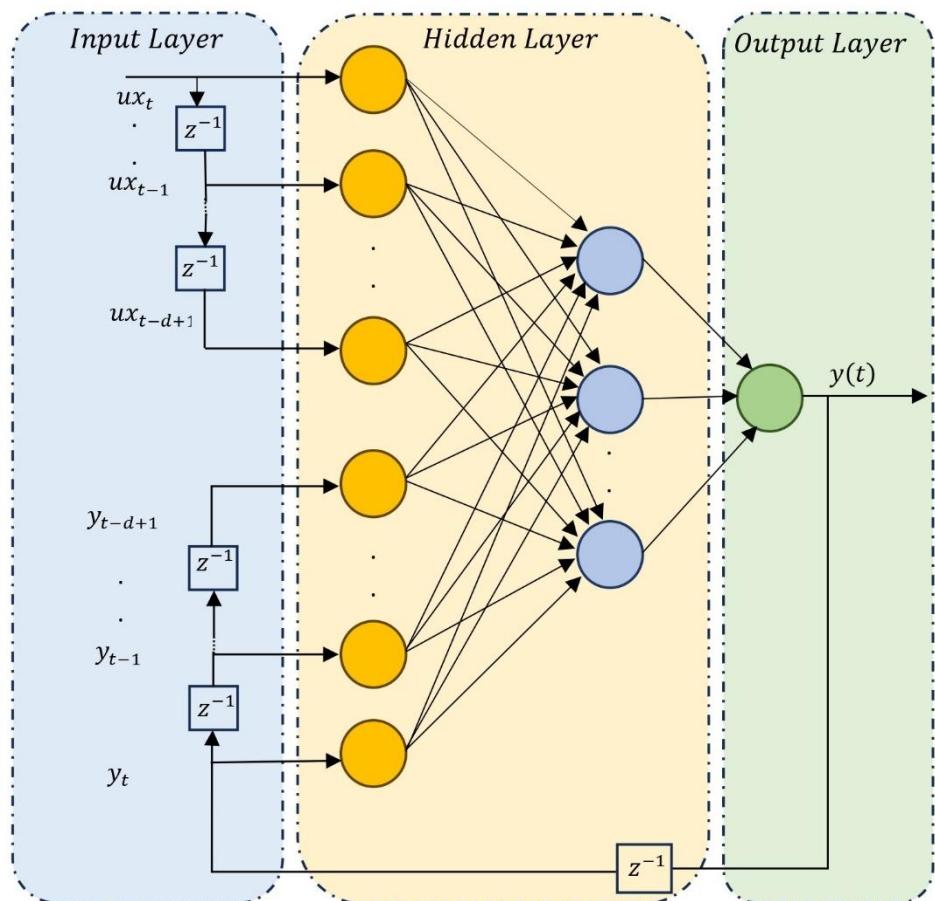


Figure 3.6 The architecture of the three-layered NARX

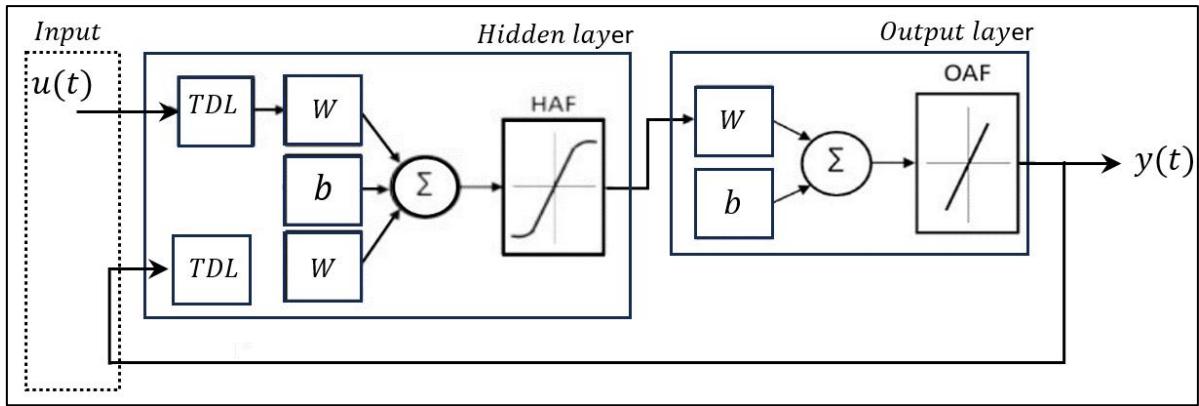


Figure 3.7 The architecture of the NARX

### 3.4.2 Long Short-Term Memory (LSTM) Neural Network

The LSTM network is an advanced architecture within the family of recurrent neural networks (RNNs), specifically designed to address the vanishing gradient problem that impedes the training of traditional RNNs. This architecture is characterized by its capacity to maintain long-term dependencies, a critical feature for tasks such as time-series forecasting where the significance of historical data extends over long periods.

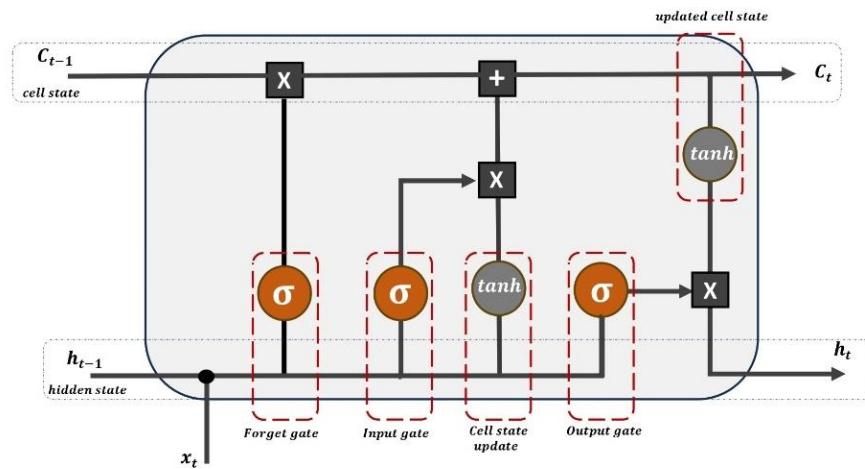


Figure 3.8 The architecture of the LSTM network

At the heart of the LSTM network lies the LSTM cell, which orchestrates the data flow through a sophisticated gating mechanism, as illustrated in Figure 3.8. Each cell consists of four integral components: the input gate ( $i_t$ ), the forget gate( $f_t$ ), the output gate ( $o_t$ ), and the cell state ( $C_t$ ) which collectively manage the cell's memory. This mechanism enables them to preserve important long-term information and discard irrelevant data, a process formalized by the following set of equations:

$$i_t = \sigma(W_i[h_{t-1}, x_t] + b_i) \quad (3.5)$$

$$f_t = \sigma(W_f[h_{t-1}, x_t] + b_f) \quad (3.6)$$

$$\tilde{C}_t = \tanh(W_C[h_{t-1}, x_t] + b_C) \quad (3.7)$$

$$C_t = f_t * ([C_{t-1}, i_t] * \tilde{C}_t) \quad (3.8)$$

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_o) \quad (3.9)$$

$$h_t = o_t * \tanh(C_t) \quad (3.10)$$

The input gate ( $i_t$ ) assesses the extent to which incoming data should alter the cell's memory state. Simultaneously, the forget gate ( $f_t$ ) decides which portions of the cell state are no longer relevant to the task at hand and should thus be discarded. This selective retention and disposal of information ensures that the LSTM cell retains only pertinent data, preventing the dilution of relevant information across the network. The values of the forget gate range from 0 to 1, with higher values indicating data of critical relevance. A value of 0 at the forget gate signifies that the corresponding information is extraneous and should be omitted from the cell state.

On the other hand, the output gate ( $o_t$ ) determines the influence of the cell state on other cells within the network. This gate controls the extent to which the value within a cell will contribute to the network's output at any given timestep. The operations of these gates are governed by activation functions—the hyperbolic tangent

(tanh) and the sigmoid function. The tanh function scales the cell state, while the sigmoid function, which outputs values between 0 and 1, regulates the gates' opening and closing. The LSTM cell's state is updated through a series of operations involving these gates:

- a. Input gate: decides values to update.
- b. Forget gate: determines what to discard from the cell state.
- c. Output gate: controls which part of the updated cell state is used for the LSTM unit's output activation.

Through this intricate system of gates, LSTMs are able to mitigate the adverse effects of exploding and vanishing gradients, thus maintaining the integrity of the gradient across many layers and time steps. This property makes LSTMs particularly suited for modelling complex phenomena such as solar irradiance, where the ability to remember and selectively forget information is paramount. Therefore, using LSTMs in solar irradiance forecasting provides several advantages. Their ability to retain long-term temporal relationships allows for a more nuanced understanding of sequential patterns in solar irradiance data. This characteristic is particularly beneficial when predicting solar output, which is subject to various temporal influences, such as diurnal and seasonal cycles.

### **3.4.3 Hybrid NARX-LSTM Model**

The hybrid NARX-LSTM architecture synthesizes the distinct features of the NARX neural network with the LSTM's sequential data processing capability to create a robust forecasting model for solar irradiance. This composite model aims to harness the precise nonlinear mapping of NARX and the long-term sequential memory of LSTMs.

The NARX network, serving as the foundation of the hybrid model, takes the lead in forecasting by utilizing its dynamic feedback mechanism. It processes both

current and historical data points, employing its distinctive feedback loops to produce an initial forecast. The mathematical representation of the NARX network can be summarized as follows:

$$\hat{y}_{narx}(t) = f(W_{narx}[(y(t-1)), \dots, y(t-d_y), u(t), u(t-1), \dots, u(t-d_u) + b_{narx}]) \quad (3.11)$$

The LSTM layer is then introduced to refine the NARX output by focusing on minimizing the residuals,  $R(t)$ . The LSTM network's equations, outlined earlier, enable it to effectively capture and learn from the temporal patterns within the residuals.

$$R(t) = y(t) - \hat{y}_{narx}(t) \quad (3.12)$$

These residuals are then passed to the LSTM network, which seeks to learn a corrective sequence,  $\Delta\hat{y}_{lstm}(t)$ , to apply to the NARX output:

$$\Delta\hat{y}_{lstm}(t) = LSTM(R(t), h_{t-1}, C_{t-1}) \quad (3.13)$$

The LSTM function represents the internal operations of the LSTM cell, including its gates and state updates, taking the residual at time  $t$ , the previous hidden state  $h_{t-1}$ , and the previous cell state,  $C_{t-1}$ . This approach enables the LSTM network to harness the learning capabilities of NARX, thereby enhancing the pattern recognition in the forecasting system.

The final forecast  $\Delta\hat{y}(t)$  from the hybrid NARX-LSTM model is then computed by adjusting the NARX forecast with the corrective sequence from the LSTM:

$$\Delta\hat{y}(t) = \hat{y}_{narx}(t) + \Delta\hat{y}_{lstm}(t) \quad (3.14)$$

To train the hybrid model in a unified manner, a joint loss function can be defined. This loss function,  $L$ , not only penalizes the disparity between the actual

values and the final forecast from the hybrid model but also include regularization terms to control the complexity of the model:

$$L = \alpha \cdot MSE_{narx} + \beta \cdot MSE_{lstm} + \lambda \cdot (\|W_{narx}\|^2 + \|W_{lstm}\|^2) \quad (3.15)$$

This loss function ensures that during training, both the NARX and LSTM models are optimized to work in harmony, with the regularization terms preventing overfitting by penalizing excessive weights.

The optimization of the hybrid model is performed using a stochastic gradient descent algorithm or one of its adaptive variants. The gradients for the LSTM and NARX parts of the network are computed through backpropagation, considering the dependencies between the two models. The weights and biases of both the NARX and LSTM models are updated iteratively based on the gradients computed from the loss function,  $L$ . The update rules can be formally expressed as:

$$W_{narx}^{(new)} = W_{narx} - \eta \frac{\partial L}{\partial W_{narx}} \quad (3.16)$$

$$b_{narx}^{(new)} = b_{narx} - \eta \frac{\partial L}{\partial b_{narx}} \quad (3.17)$$

$$W_{lstm}^{(new)} = W_{lstm} - \eta \frac{\partial L}{\partial W_{lstm}} \quad (3.18)$$

$$b_{lstm}^{(new)} = b_{lstm} - \eta \frac{\partial L}{\partial b_{lstm}} \quad (3.19)$$

Where  $\eta$  is the learning rate, and the partial derivatives represent the gradients of the loss function with respect to the corresponding weights and biases of the NARX and LSTM networks.

The performance evaluation of the forecast system is methodically conducted and documented in this section. Utilized as the primary indicators of accuracy, Root Mean Squared Error (RMSE), Normalized Root Mean Squared Error (nRMSE), and Mean Absolute Error (MAE) are applied to assess the model's predictions against varying weather conditions. Furthermore, the system's performance is contextualized by a comparison with the standalone models.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (3.21)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3.21)$$

$$nRMSE = 100(\%) \sqrt{\frac{1}{n} \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{y_c}} \quad (3.22)$$

The adoption of the NARX neural network in the proposed model is grounded in its proven efficacy in addressing the latching phenomenon and in identifying nonlinear relations. Complementing this, the Long Short-Term Memory (LSTM) network mitigates the vanishing gradient issue, which is crucial for maintaining the model's learning capacity over long sequences. To elucidate the mechanics of the proposed hybrid model, we delineate a four-stage data processing framework as depicted in Figure 3.9.

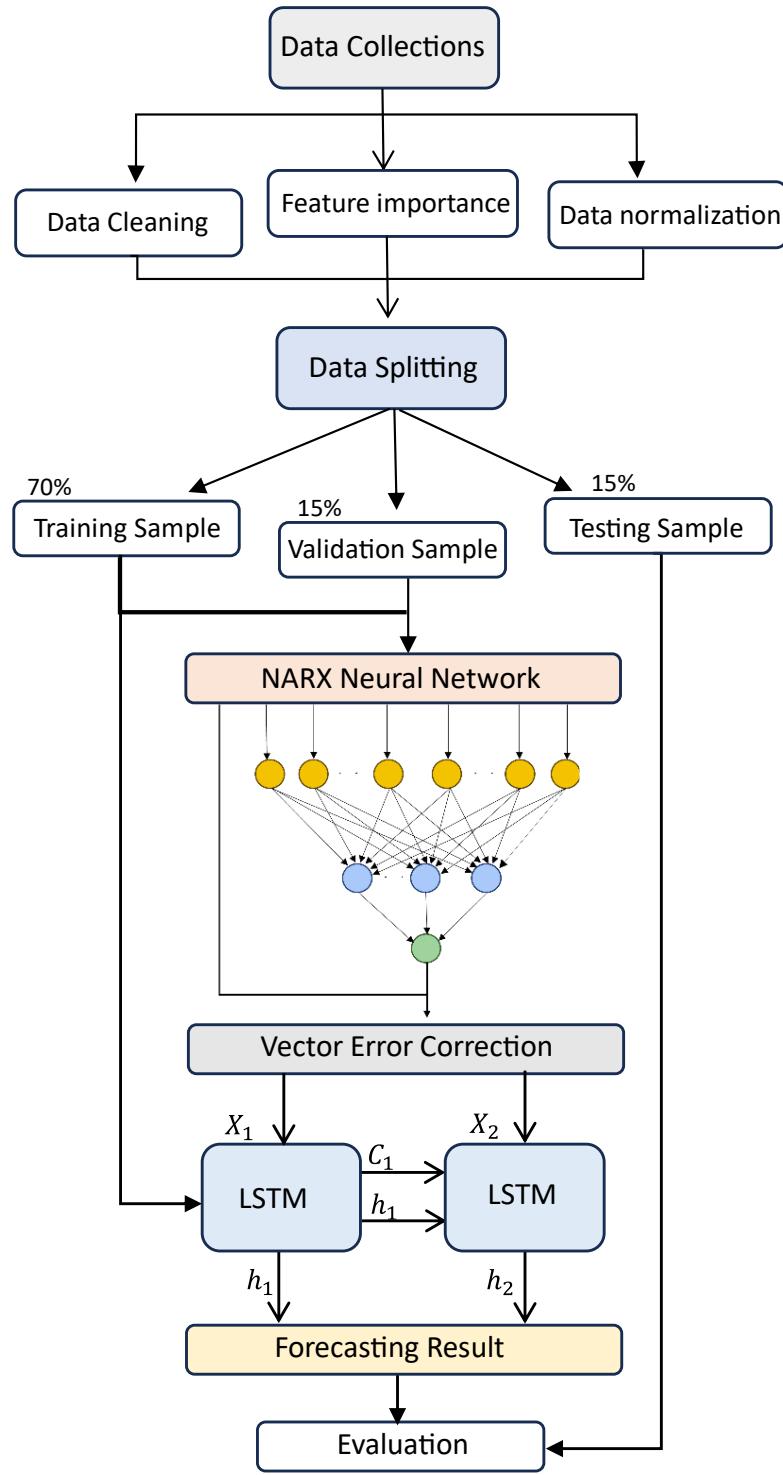


Figure 3.9 The architecture of the proposed model.

### 3.5 Input Sensitivity Analysis

In the Input Sensitivity Analysis section, a detailed examination is conducted to evaluate the influence of each input parameter on the solar irradiance forecasting model's performance. This analysis plays a crucial role in identifying which weather parameters are most significant in predicting solar irradiance. Techniques such as feature importance ranking and sensitivity testing are employed to discern how variations in input data like air temperature, cloud cover, humidity, and other meteorological factors affect the model's predictions.

The Random Forest algorithm is employed to determine the significance of various weather parameters. This is achieved by establishing a regression-based ensemble model through the Random Forest algorithm which generates a multitude of decision trees ( $n\text{Trees} = 100$ ) to predict the target variable solar irradiance (SI). The algorithm's efficacy hinges on its Out-Of-Bag (OOB) estimator for assessing feature importance, represented by:

$$\text{Feature Importance (FI)} = \sum OOB_{ccases} = (Error_{permuted} - Error_{original}) \quad (3.23)$$

Where  $Error_{permuted}$  and  $Error_{original}$  are the prediction errors on OOB samples before and after permuting each predictor variable. The OOB Mean Squared Error (MSE), is computed as the average squared difference between the observed and predicted GHI values for OOB samples:

$$OOB \text{ MSE} = \frac{1}{N_{OOB}} \sum (SI_{permuted} - SI_{original})^2 \quad (3.24)$$

Where  $N_{OOB}$  is the number of OOB samples. The OOB error trend across the number of trees is visualized to guide model complexity decisions. This comprehensive methodological approach allows for an in-depth understanding of each weather parameter's influence on the predictive model, thereby optimizing the solar irradiance forecasting.

### **3.6 Chapter Summary**

This chapter has laid the essential foundation for the development and evaluation of a hybrid NARX-LSTM model, targeted at enhancing solar irradiance forecasting. The chapter began with an exploration of the individual components of the model: the Nonlinear Autoregressive with Exogenous Inputs (NARX) and Long Short-Term Memory (LSTM) neural networks. These components were then intricately integrated to form a robust hybrid model capable of capturing the complex dynamics of solar irradiance influenced by various meteorological factors. Detailed data analysis, including a thorough examination of key weather parameters and their relationships with solar irradiance, underpinned the model's development. The preprocessing stage further refined the data, preparing it for effective model training. Additionally, an Input Sensitivity Analysis section was introduced to assess the impact of different input variables on the model's performance, further enhancing its predictive accuracy. This foundational work sets the stage for a comprehensive approach to accurately forecast solar irradiance, a critical factor in optimizing renewable energy systems like microgrids.

## **CHAPTER 4**

### **INITIAL RESULTS**

#### **4.1 Introduction**

This chapter provides an in-depth analysis of the hybrid NARX-LSTM solar irradiance forecasting model, and the exploratory data analysis (EDA) conducted on the dataset. The evaluation is based on a robust dataset sourced from DNV GL, covering the period from January 2007 to November 2023 with hourly updates. The dataset's extensive temporal and climatic diversity serves as a solid foundation for assessing the real-world applicability of the forecasting model. EDA explores the underlying patterns, correlations, and distributions of meteorological factors that influence solar irradiance, such as temperature, cloud cover, and humidity. These factors play a critical role in shaping the accuracy of solar irradiance predictions.

The insights gained from the EDA form the groundwork for understanding the strengths and limitations of the hybrid NARX-LSTM model. This section delves into the impacts of individual weather parameters on forecasting performance, evaluating their contributions to model accuracy and reliability. Additionally, it contextualizes the findings within the broader framework of solar energy forecasting and renewable energy management. By combining EDA with the results of the hybrid model, this chapter highlights the importance of robust data analysis and advanced modelling techniques in achieving precise and actionable solar irradiance predictions.

#### **4.2 Exploratory Data Analysis**

Exploratory Data Analysis (EDA) is a critical step in understanding the dataset's structure, trends, and variability. Through EDA, key patterns and relationships among variables are uncovered, providing insights into the

meteorological factors that influence solar irradiance. This section employs a combination of statistical summaries, visualizations, and time-series decomposition to identify both predictable seasonal and diurnal cycles as well as random fluctuations. By analyzing the data comprehensively, EDA establishes a robust foundation for model development and evaluation. The insights gained here inform feature selection, highlight the significance of individual predictors, and identify potential challenges, such as outliers or extreme weather events, that could affect forecasting accuracy.

#### 4.2.1 Overview of the Dataset

The dataset analyzed in this study consists of nine critical meteorological variables that are essential for understanding and predicting solar irradiance. These variables include air temperature, cloud attenuation, precipitation rate, precipitable water, relative humidity, surface pressure, wind direction, wind speed, and dewpoint temperature. Collected hourly over a span of 16 years (from January 2007 to November 2023), this dataset provides an extensive temporal resolution that captures both short-term fluctuations and long-term climatic trends.

The data originates from Johor Bahru, Malaysia, a region characterized by tropical weather patterns, offering a unique combination of meteorological challenges and opportunities for solar energy forecasting due to high humidity, frequent cloud cover, and intermittent rainfall. The tropical climate ensures a wide range of weather conditions, from clear skies to heavy precipitation, making it a robust testing ground for solar forecasting models. Each variable in the dataset contributes uniquely to the analysis. For instance:

- (a) Air Temperature is a direct indicator of solar radiation and atmospheric conditions. Higher temperatures are often associated with clearer skies, while sudden drops may indicate cloud cover or rain.
- (b) Cloud Attenuation measures the degree to which clouds block sunlight, a primary factor affecting solar irradiance. Its variations offer insights

- into weather predictability and model performance under dynamic conditions.
- (c) Precipitation Rate captures the intensity of rainfall, directly correlating with periods of low irradiance. This variable also adds complexity as tropical rainfall is often sudden and short-lived.
  - (d) Relative Humidity provides a measure of atmospheric moisture, which impacts cloud formation and solar scattering.
  - (e) Surface Pressure serves as a stable atmospheric baseline but fluctuates significantly during storms or periods of unstable weather.
  - (f) Wind Speed and Direction influence cloud movement and precipitation patterns, indirectly affecting solar irradiance. Wind direction often reflects seasonal changes in weather patterns.
  - (g) Dewpoint Temperature highlights the saturation point of air, which is crucial for understanding fog, dew formation, and cloud dynamics.
  - (h) Wind direction shows no explicit numerical range in degrees in the summary, as it is typically measured in compass points or degrees from  $0^\circ$  to  $360^\circ$ . Its bimodal nature in other sections indicates seasonal prevailing winds and their directional consistency.
  - (i) Precipitable Water: Represents the total water vapor in a column of the atmosphere, measured in units like millimeters (mm). High mean and deviation indicate seasonal moisture flow above average or high energy under tropical humid conditions.

The dataset's hourly frequency enables the analysis of diurnal cycles, where solar irradiance starts at zero at night, rises sharply after sunrise, and peaks at midday before gradually declining toward sunset. These patterns are influenced by seasonal changes, which are equally captured in the dataset. For example, during the monsoon season, prolonged cloud cover and precipitation reduce irradiance, while dry seasons offer consistent and high solar energy potential.

By leveraging this rich dataset, this study seeks to uncover key patterns and relationships between variables, laying the foundation for accurate solar irradiance forecasting using the hybrid NARX-LSTM model. The analysis aims to capture both

the predictability of regular cycles and the challenges posed by variability and anomalies, ultimately enhancing the model's applicability to real-world scenarios.

#### 4.2.2 Statistical Summary

The statistical summary provides a numerical overview of the key meteorological variables in the dataset as shown in table 4.1, offering insights into their central tendencies, variability, and range. The mean and median values reflect the typical conditions experienced in Johor Bahru, Malaysia, while the standard deviation highlights the extent of variability within each variable.

Table 4.1 Statistical summary table

Feature (Variables)	Mean	Median	Std Dev	Min	Max
Cloud Attenuation (%)	48.2	50	28.7	0	100
Surface Pressure (hPa)	1010.3	1011	5.2	995	1025
Precipitation rate (mm/hr)	1.8	0	5.4	0	80
Wind Direction (°)	180.5	185	60.3	0	360
Wind Speed (m/s)	3.2	2.8	1.6	0.1	15
Precipitable water (mm)	35.2	34.8	5.7	25	45
Air temperature (°C)	27.8	27.5	2.5	22	35
Relative Humidity (%)	85.3	86	8.2	50	100
Dewpoint Temp. (°C)	24.6	24.5	2.4	19	28

For instance, air temperature exhibits a mean of 27.8°C with a standard deviation of 2.5°C, indicating relatively stable conditions, characteristic of a tropical climate. Similarly, cloud attenuation, with a mean of 48.2% and a wide standard deviation of 28.7%, underscores the significant variability in cloud cover, which is a critical factor influencing solar irradiance. Precipitation rate, with a low mean of 1.8

mm/hr and a high maximum value of 80.0 mm/hr, highlights the sporadic nature of heavy rainfall events typical of tropical regions.

The dataset also reveals high levels of relative humidity, with a mean of 85.3% and a maximum of 100%, further emphasizing the humid tropical conditions. Surface pressure shows a narrower range, with a mean of 1010.3 hPa, consistent with stable atmospheric conditions, but occasional deviations due to storm activity. Wind speed, with a mean of 3.2 m/s and a skewed maximum of 15.0 m/s, highlights the occasional presence of strong winds, while dewpoint temperature, averaging 24.6°C, aligns with high humidity levels.

While for the wind direction is typically distributed from 0° (north) to 360° (full compass). The mean of 180.5° suggests prevailing winds often align with a southerly direction, with std dev of 60.3 which highlights variability, which could represent seasonal shifts in dominant wind patterns.

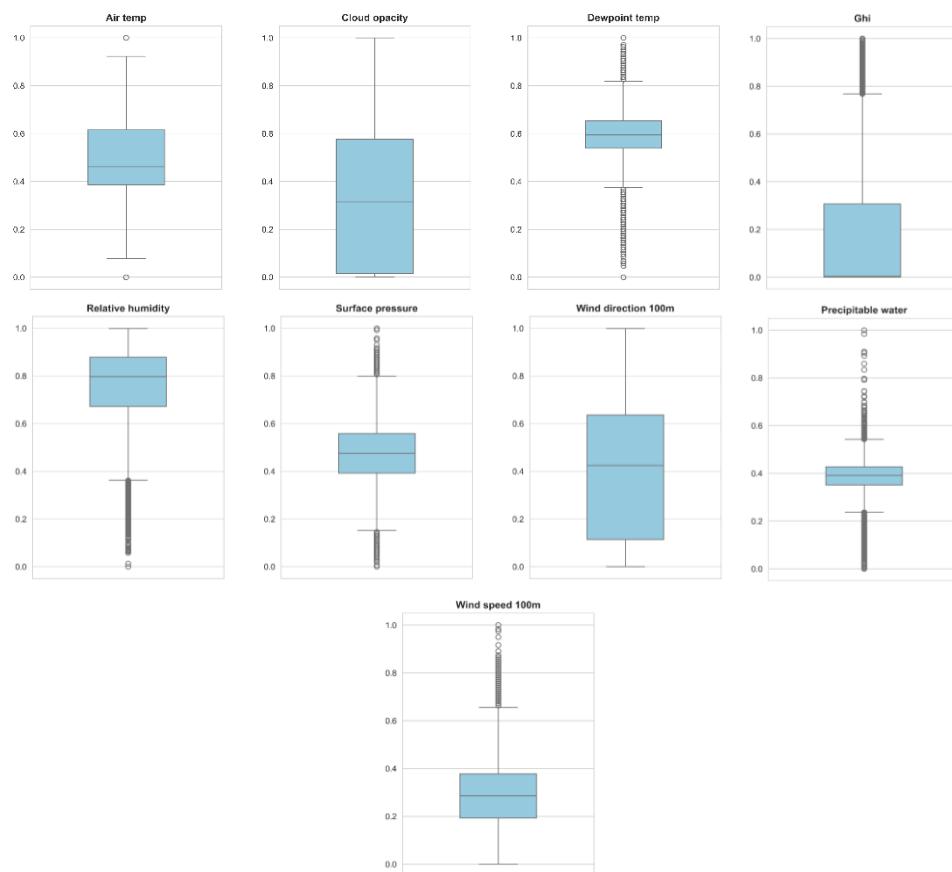


Figure 4.1 Variables boxplots

The boxplots for the dataset variables provide valuable insights into their distributions, variability as can be shown in Figure 4.1, and potential outliers where the normalization method has been used to handle the outliers by transforming the data into range between 0 and 1 as mentioned in chapter 3, which highlighting the characteristics of the tropical region's meteorological conditions. Air temperature exhibits a relatively narrow interquartile range, reflecting stable conditions typical of a tropical climate, with occasional outliers indicating unusual weather events like heatwaves or cold fronts. Cloud opacity shows a wide interquartile range, signifying significant variability in cloud cover, a critical factor affecting solar irradiance. The presence of high outliers indicates periods of dense cloud cover, likely during storms or extreme weather events.

The distribution of dewpoint temperature is consistent, with a narrow range and a few lower outliers, representing rare instances of drier atmospheric conditions. In contrast, the GHI (Global Horizontal Irradiance) boxplot reveals a large spread in values, dominated by lower readings due to nighttime conditions with no irradiance. Higher outliers reflect clear-sky conditions during peak solar hours, emphasizing the variability in solar energy availability.

Precipitable water is centered around moderate values, with higher outliers pointing to periods of elevated atmospheric moisture, often associated with heavy rainfall or humid conditions. Similarly, the precipitation rate distribution is highly skewed, with most values near zero but significant outliers indicating intense tropical rain events, characteristic of the region's weather patterns. Relative humidity demonstrates consistently high values, with a narrow interquartile range, reflective of the region's humid climate, while occasional lower outliers represent rare dry spells or hot conditions.

Surface pressure maintains a tight distribution, with minor outliers suggesting deviations caused by extreme weather, such as storms or typhoons. Wind direction spans the entire range from  $0^\circ$  to  $360^\circ$ , showing the variability in prevailing wind patterns influenced by monsoons and local weather phenomena. Finally, wind speed

exhibits a right-skewed distribution, with most readings at lower speeds and a few high outliers corresponding to strong winds during severe weather events.

Overall, these boxplots highlight the variability, and extremes present in the dataset, offering a comprehensive understanding of the meteorological factors influencing solar irradiance in tropical climates. This analysis underscores the importance of capturing such variability in predictive models to enhance the accuracy of solar irradiance forecasting.

#### 4.2.3 Correlation Analysis

The correlation matrix provides a comprehensive view of the relationships between various meteorological parameters and solar irradiance (GHI) as can be seen in Figure 4.2. Each correlation coefficient quantifies the linear relationship between two variables, with positive values indicating a direct relationship and negative values reflecting an inverse relationship. The most notable correlations include the strong positive relationship between air temperature and GHI (0.71), highlighting that higher air temperatures are typically associated with increased solar irradiance, a result of clearer skies and enhanced solar heating. Conversely, air temperature shows a strong negative correlation with relative humidity (-0.85), reflecting the inverse relationship common in tropical climates, where higher temperatures coincide with drier atmospheric conditions.

GHI, the primary target variable, exhibits moderate to strong correlations with several key predictors. It is positively correlated with air temperature (0.71) and negatively correlated with relative humidity (-0.69) and cloud opacity (-0.28). These correlations underscore the influence of clear skies, drier air, and minimal cloud cover on solar irradiance levels. GHI also shows a weak negative correlation with precipitable water (-0.12), indicating that atmospheric moisture slightly reduces irradiance through scattering and absorption of sunlight. Interestingly, GHI's correlation with precipitation rate is negligible (0.08), suggesting that while rainfall impacts solar irradiance, it does so irregularly, without a strong linear pattern.

Cloud opacity, a critical factor for solar irradiance prediction, shows a moderate positive correlation with precipitable water (0.34) and a weak negative correlation with GHI (-0.28). This relationship highlights how increased cloud cover, often associated with high atmospheric moisture, attenuates sunlight and reduces irradiance. Dewpoint temperature exhibits a moderate positive correlation with air temperature (0.55) and a weaker correlation with GHI (0.27), indicating that while warmer air can hold more moisture, it does not strongly influence solar irradiance directly.

Relative humidity demonstrates strong inverse relationships with GHI (-0.69) and air temperature (-0.85), emphasizing its critical role in solar irradiance dynamics. High humidity levels typically coincide with cloudy or rainy conditions, reducing irradiance. Surface pressure, in contrast, exhibits weak correlations with other variables, showing a slight negative relationship with air temperature (-0.22) and dewpoint temperature (-0.22), suggesting minor atmospheric variations during storms or extreme weather events.

Wind-related variables, including wind speed and wind direction, exhibit minimal correlations with GHI and other meteorological factors. Wind speed shows a weak negative correlation with precipitable water (-0.28) and no significant relationship with GHI (-0.04), indicating its limited direct influence on solar irradiance. Wind direction displays no notable correlations with any variable, further reinforcing its independence within the dataset.

Temporal variables, such as hour and month, reveal important diurnal and seasonal patterns. Hour exhibits a moderate positive correlation with GHI (0.47), reflecting the daily solar cycle, with irradiance peaking at midday. It also has a moderate negative correlation with relative humidity (-0.61), as humidity decreases during daytime hours due to evaporation and atmospheric heating. Month, on the other hand, shows weak correlations with most variables, highlighting its role in capturing broader seasonal trends rather than daily fluctuations.

Overall, the heatmap highlights the most influential factors for solar irradiance forecasting, such as air temperature, relative humidity, and cloud opacity, while also identifying variables with minimal impact, like wind direction. These insights are critical for feature selection and improving the accuracy of predictive models. The observed relationships underline the complexity of solar irradiance dynamics and the necessity of incorporating these interdependencies into forecasting methodologies.

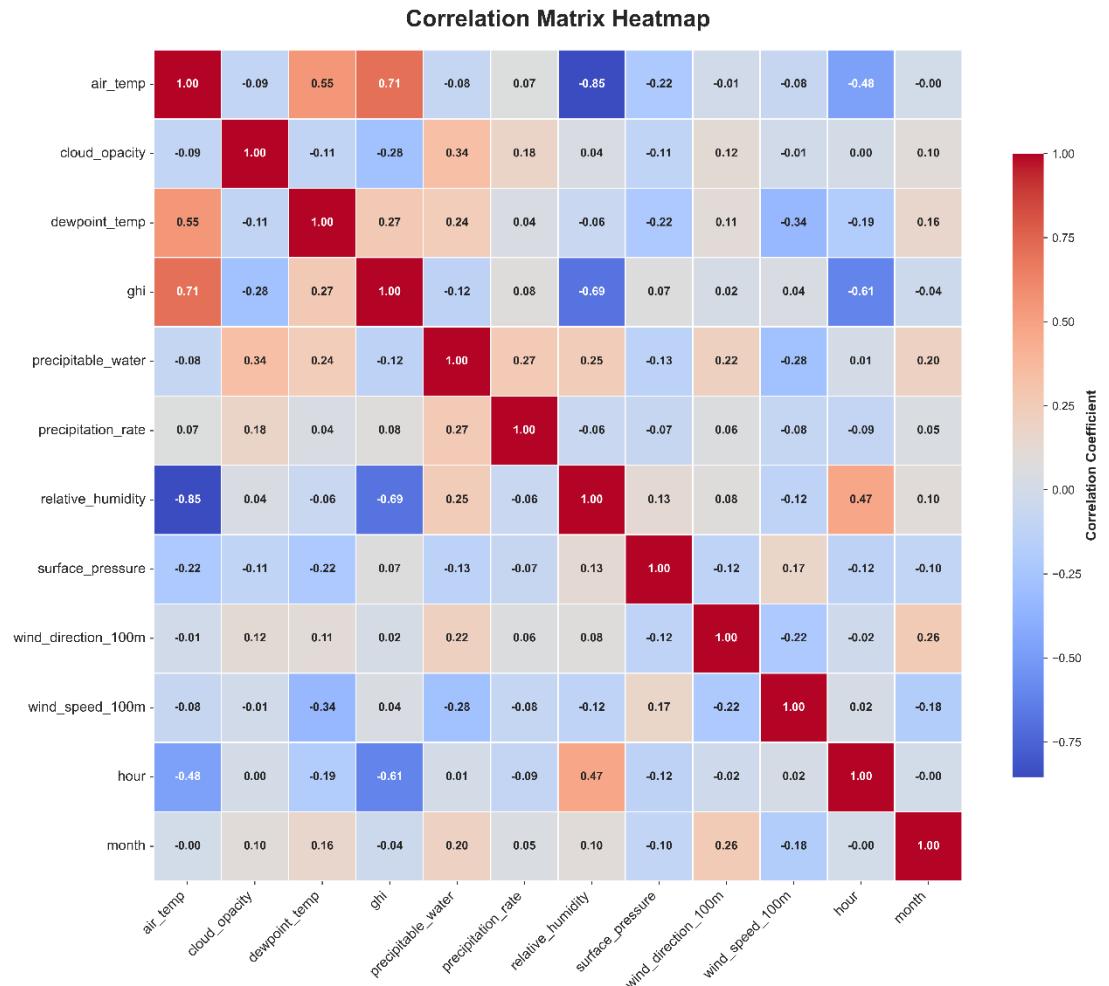


Figure 4.2 Correlation matrix heatmap

The heatmap of mean GHI (Global Horizontal Irradiance) by hour and month offers a detailed visualization of how solar irradiance varies across different times of the day and months of the year as can be seen in Figure 4.3. The color intensity

represents the average GHI, with darker red shades indicating higher values and lighter shades showing lower levels of irradiance.

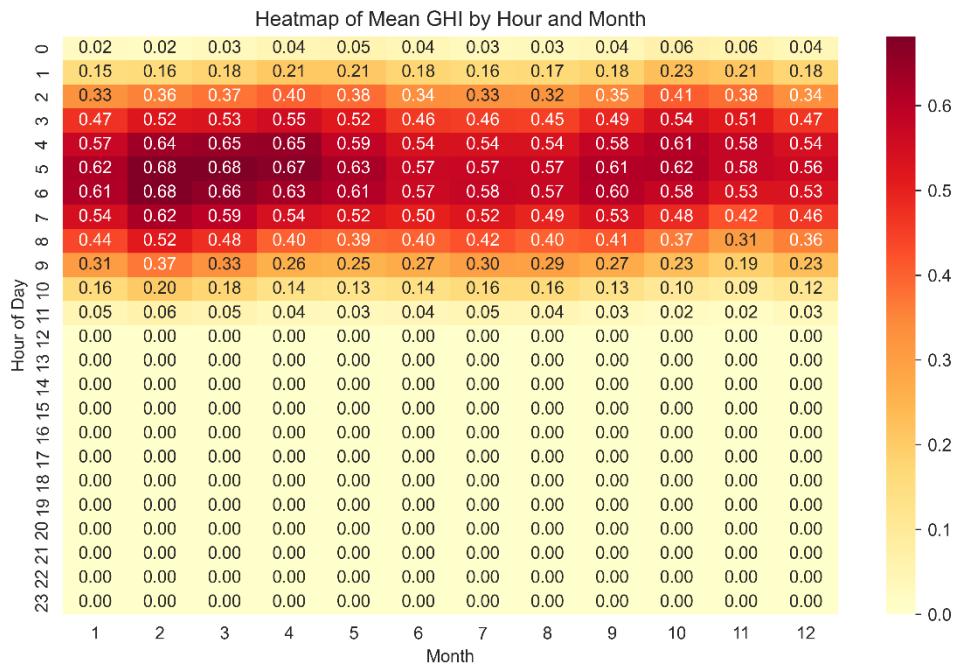


Figure 4.3 The heatmap of mean GHI by hour and month

This heatmap highlights the diurnal and seasonal patterns of solar irradiance. The rows correspond to the hours of the day, while the columns represent the months. The highest GHI values are consistently observed between 6:00 AM and 10:00 AM, aligning with the hours when the sun's angle is most favorable for maximum irradiance. These hours show darker red shades across most months, peaking during the mid-year months (e.g., May to August) due to longer daylight hours and clearer skies typical of this period.

Seasonal variations are clearly depicted, with higher GHI values occurring in the summer months (e.g., May, June, and July) and lower values during the winter months (e.g., December and January). This trend reflects the expected annual solar cycle, where summer months have higher solar energy due to the Earth's tilt and increased solar exposure.

Interestingly, there is little to no irradiance during the late evening and night hours (12:00 PM to 5:00 AM), as indicated by the uniformly light shades in these rows. This aligns with the absence of sunlight during these hours. Additionally, the early morning and late afternoon hours exhibit moderate levels of GHI, likely due to the sun's lower angle and the scattering of sunlight by the atmosphere.

This heatmap provides valuable insights into the temporal distribution of solar irradiance, emphasizing the critical hours and months for optimal solar energy generation. These findings are instrumental in refining forecasting models and optimizing solar panel operations to align with periods of maximum irradiance, ensuring efficient energy capture and utilization.

#### **4.2.4 Time-Series Analysis**

The GHI time series plot provides a visual representation of the Global Horizontal Irradiance (GHI) values over a 16-year period from 2007 to 2023. This chart captures the temporal variations in GHI and highlights both its seasonal patterns and long-term trends as shown in Figure 4.4.

The plot indicates a consistent cyclical pattern, with noticeable peaks and troughs that align with seasonal changes. These fluctuations reflect higher irradiance levels during sunnier periods, such as summer, and reduced values during cloudier or less sunny seasons, such as winter. The regularity of the peaks and troughs underscores the strong influence of the Earth's annual orbit and tilt, resulting in predictable changes in solar exposure.

Additionally, the graph demonstrates the relatively high stability of GHI over the years, with no significant long-term decline or increase. This stability suggests that the location of the dataset, likely in a region with consistent climatic conditions, provides an ideal environment for solar energy research and forecasting.

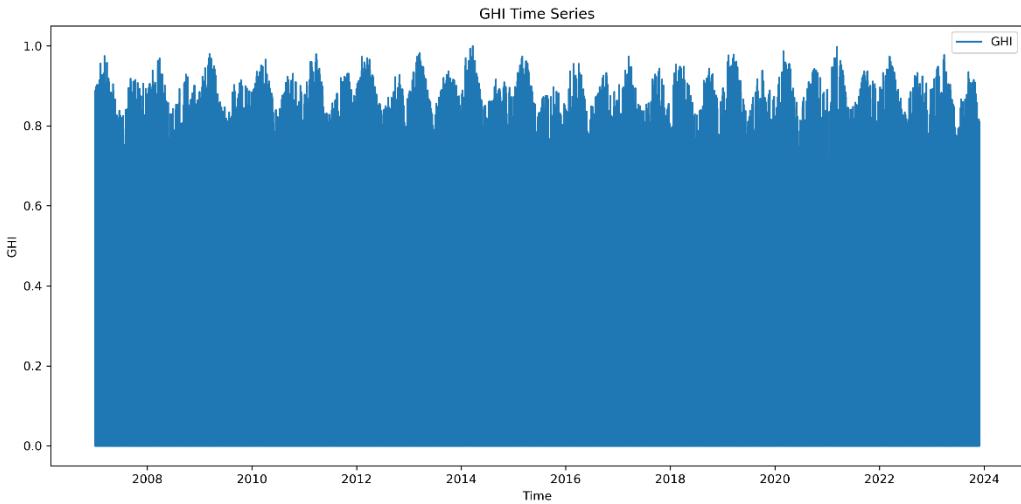


Figure 4.4 Global horizontal irradiance (GHI) time-series

The daily average GHI (Global Horizontal Irradiance) plot illustrates the temporal variations in solar irradiance over an extended period from 2007 to 2023. This visualization aggregates the data on a daily basis, reducing noise from hourly fluctuations and providing a clearer perspective on overall trends and seasonal patterns as in Figure 4.5.

The plot reveals distinct seasonal cycles, with higher irradiance levels observed during sunnier periods and lower levels during overcast or rainy seasons as can be seen in Figure 4.5. These cyclical patterns align with natural variations in solar exposure due to the Earth's rotation and orbit, as well as seasonal meteorological changes. Peaks in daily average GHI correspond to periods of clearer skies and longer daylight hours, while troughs reflect periods of increased cloud cover or precipitation.

Notably, the data displays a relatively consistent range of daily average GHI values throughout the years, suggesting a stable climate over the observation period. However, occasional sharp drops in GHI indicate the impact of transient weather conditions such as storms or prolonged rainfall. These anomalies are critical for identifying the challenges in solar forecasting, as they highlight periods where irradiance levels deviate significantly from typical seasonal expectations.

This daily aggregation also provides a robust dataset for time-series analysis, as it balances the need for granularity with the elimination of excessive noise. The trends observed in this plot form a foundational understanding of solar irradiance behaviour, serving as a key input for training and evaluating predictive models such as the hybrid NARX-LSTM. The consistency and periodicity captured in this visualization underscore the importance of leveraging temporal dependencies in forecasting methodologies.

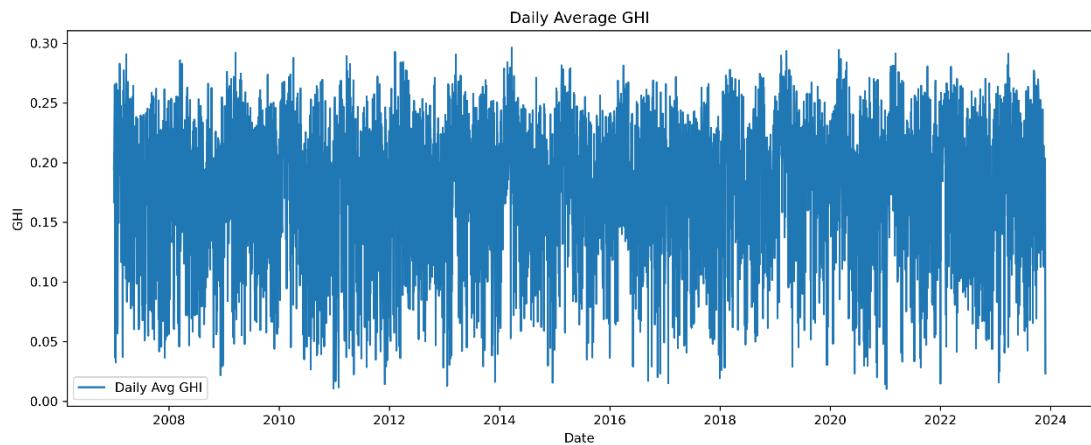


Figure 4.5 Daily average GHI

The monthly average GHI (Global Horizontal Irradiance) plot illustrates the long-term trends in solar irradiance aggregated on a monthly basis from 2007 to 2023. This visualization smoothens daily fluctuations, providing a clearer view of seasonal and annual variations as shown in Figure 4.6.

The graph demonstrates a repeating cyclical pattern, characteristic of seasonal changes in solar exposure. Peaks in GHI are consistently observed during sunnier months, corresponding to summer or dry seasons, while troughs align with cloudier or rain-heavy periods, typically during monsoon seasons. These periodic changes reflect the natural influence of Earth's tilt and orbit on solar energy availability.

Although the general trend appears stable, some years exhibit higher variability between consecutive months. For instance, sudden drops in the graph may indicate years with prolonged monsoon seasons or periods of abnormal weather conditions,

such as increased cloud cover or storms. Conversely, sharp rises suggest extended clear-sky periods with higher solar exposure.

The long-term consistency of the monthly averages highlights the reliability of this dataset for modelling solar irradiance trends. These monthly aggregates are particularly valuable for identifying seasonal patterns and informing energy system planning, such as optimizing photovoltaic system design or energy storage requirements.

This analysis reinforces the importance of seasonality in GHI forecasting and provides critical insights for the hybrid NARX-LSTM model to capture these periodic patterns effectively. By leveraging this monthly aggregated view, forecasting accuracy can be enhanced, particularly for medium- to long-term predictions.

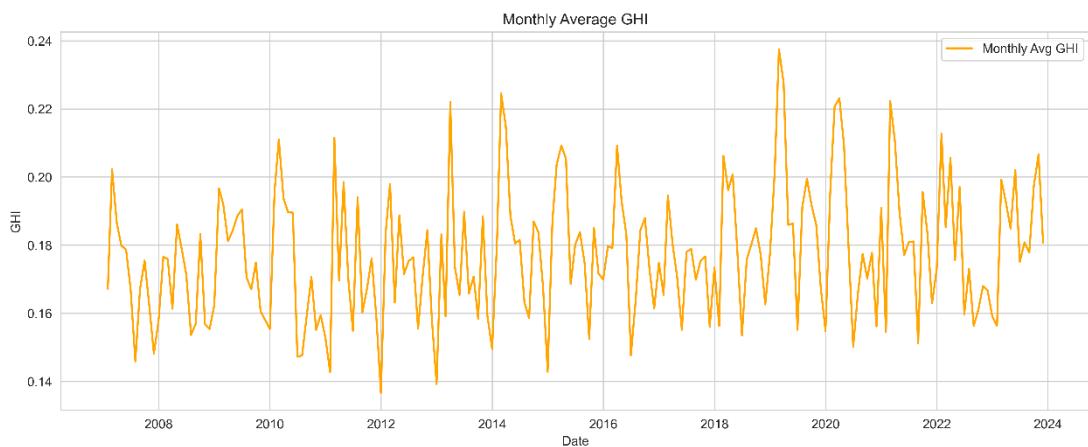


Figure 4.6     Monthly average GHI

The decomposition of the GHI (Global Horizontal Irradiance) time series into its observed, trend, seasonal, and residual components provides valuable insights into the underlying dynamics of solar irradiance as shown in Figure 4.7. The observed component, represented in the top panel, shows the complete time series data, highlighting its overall variability and recurring patterns over the years. This component captures both seasonal cycles and irregular fluctuations caused by transient weather events. The trend component, in the second panel, isolates the long-term changes in GHI. It reveals gradual rises and falls in solar irradiance, which may

correspond to shifts in climatic conditions, prolonged clear-sky periods, or extended cloudy seasons. This trend provides a clear picture of the directional movement of solar irradiance over time.

The third panel illustrates the seasonal component, which reflects the consistent and predictable annual cycles in solar irradiance driven by the Earth's orbit and rotation. These regular oscillations align with expected seasonal variations, such as higher GHI during summer and lower values during winter or monsoon seasons. The seasonal component emphasizes the critical role of seasonality in shaping solar irradiance patterns and informs forecasting models to incorporate these periodic behaviours. Finally, the residual component in the bottom panel represents the portion of the data that cannot be explained by the trend or seasonality. These residuals capture random fluctuations and outliers, likely caused by unpredictable weather conditions, such as storms, or anomalies in data recording. While the residuals appear evenly distributed, larger deviations at certain points suggest opportunities for further analysis to address unexpected variations.

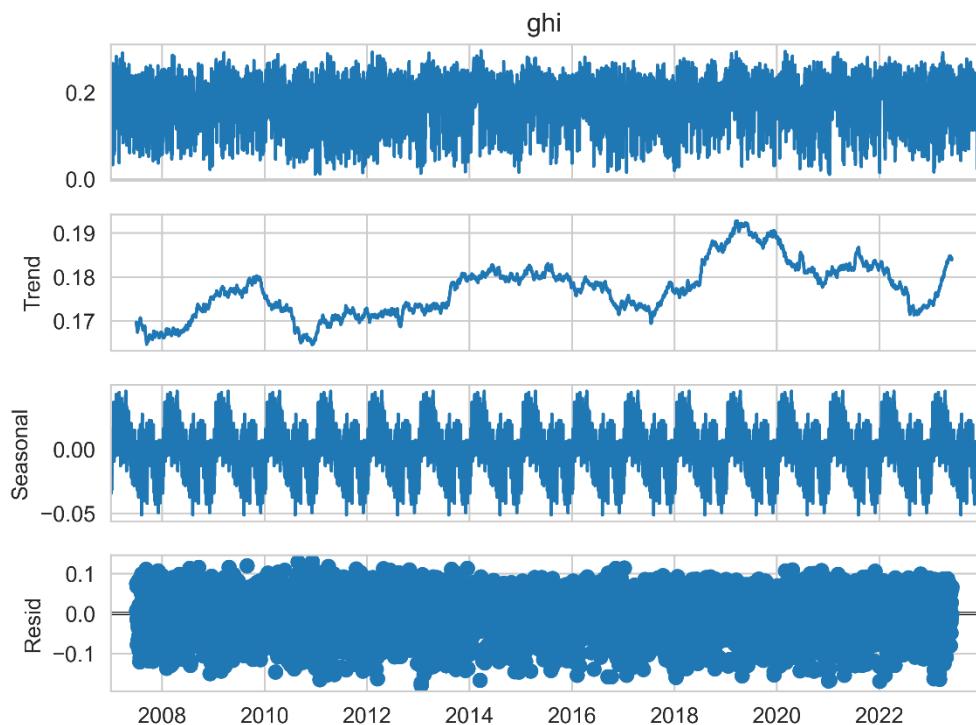


Figure 4.7      Decomposition of the GHI

The decomposition offers a comprehensive breakdown of the GHI time series, highlighting the contributions of trend and seasonality while exposing areas of irregularity. These insights are crucial for the development of robust predictive models like the hybrid NARX-LSTM, enabling them to effectively account for long-term trends, seasonal cycles, and unexpected deviations in solar irradiance.

The 3D surface plot of Hour, Month, and GHI shown if Figure 4.8 provides a comprehensive visualization of the temporal dynamics of solar irradiance. This representation captures the variations in Global Horizontal Irradiance (GHI) across different hours of the day and months of the year, offering valuable insights into the interaction between diurnal and seasonal patterns. The combination of these two temporal dimensions highlights the predictable and cyclical nature of solar energy availability, while also revealing periods of variability caused by climatic conditions.

3D Surface Plot of Hour, Month, and GHI

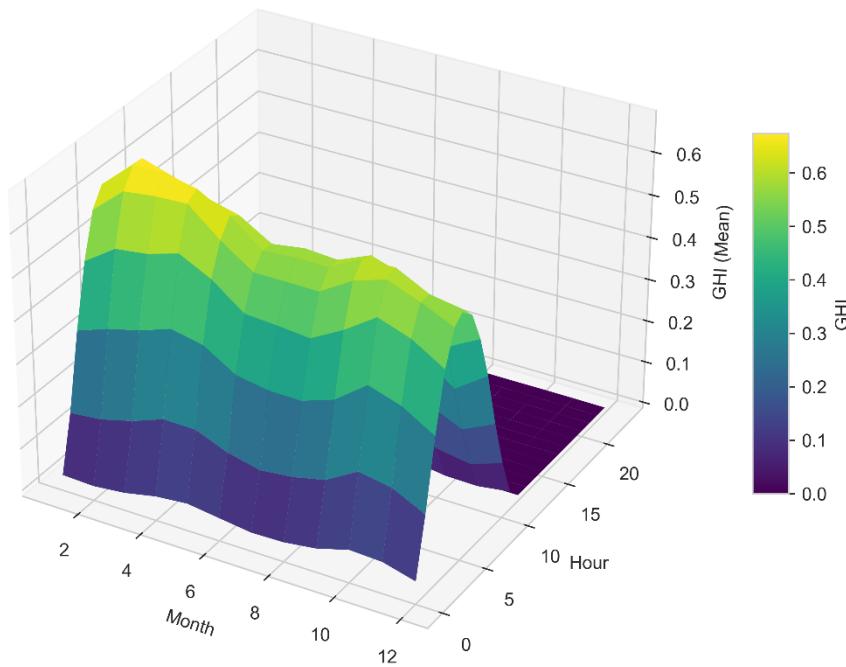


Figure 4.8 The 3D surface plot of Hour, Month, and GHI

From an hourly perspective, the plot clearly shows that GHI values increase sharply during the early morning hours, reach their peak around midday, and then decrease rapidly toward zero in the evening. This pattern is consistent with the diurnal cycle of solar irradiance, driven by the Sun's position in the sky. The highest irradiance levels are observed at solar noon, when the Sun is at its highest point, maximizing the energy received at the Earth's surface. These hourly variations are critical for understanding the optimal periods for solar energy generation during the day.

The monthly variations depicted in the plot indicate a strong seasonal influence on solar irradiance. Higher GHI values are observed during the mid-year months, such as June and July, which align with the dry season characterized by clear skies and minimal atmospheric obstruction. In contrast, lower GHI values are evident during the late-year months, particularly November and December, which coincide with the monsoon season. This period is marked by increased cloud cover, frequent rainfall, and shorter daylight hours, all of which contribute to reduced solar irradiance.

The interaction between diurnal and seasonal patterns is a significant feature of this plot. During the summer months, GHI peaks are more pronounced due to extended daylight hours and optimal solar angles. Conversely, during the monsoon season, the peaks are lower, reflecting the impact of adverse weather conditions on solar energy availability. This dual influence underscores the importance of considering both hourly and seasonal variations when analyzing solar irradiance for forecasting and energy management.

The color gradient in the plot, ranging from dark purple (low GHI) to bright yellow (high GHI), effectively illustrates the temporal distribution of irradiance. High GHI levels are concentrated around midday during the summer months, while low GHI levels dominate nighttime hours and the monsoon season. This clear visualization emphasizes the temporal predictability of solar energy during certain periods while highlighting the challenges posed by seasonal variability.

### **4.3 Input Sensitivity Result**

By employing the Random Forest learning technique as written in section 3.4, the impact of distinct weather parameters on solar irradiance forecasting is meticulously evaluated. The feature importance scores for the weather parameters reveal their relative impact on solar irradiance prediction. Cloud opacity emerges as the most influential factor with a score of 24.652, indicating its strong negative correlation with GHI. Surface pressure follows closely with a score of 21.054, highlighting its role as an indicator of atmospheric conditions that affect solar irradiance as shown in Table 4.2. The Out-Of-Bag (OOB) Mean Squared Error, at 0.012171, provides a measure of the model's prediction accuracy when using all 9 parameters.

Armed with the insights gleaned from the Random Forest feature importance analysis, the next phase involved rigorous performance testing of the forecasting model. By strategically adding and removing weather parameters and observing the corresponding shifts in nRMSE, the model's sensitivity to each variable was quantified. Figures 4.9 and 4.10 illustrate the variations in nRMSE against the permutations of these influential parameters, providing a clear visual representation of their individual contributions to the accuracy of solar irradiance predictions.

The analysis of each model's performance under various weather conditions, as depicted in Figure 4.9, offers crucial insights into how they cope with environmental variations. The graph demonstrates the resilience and adaptability of the hybrid NARX-LSTM model compared to its standalone counterparts. Particularly, the analysis focusing on scenarios excluding Cloud Attenuation from the models provides a nuanced understanding of its impact on solar irradiance forecasting.

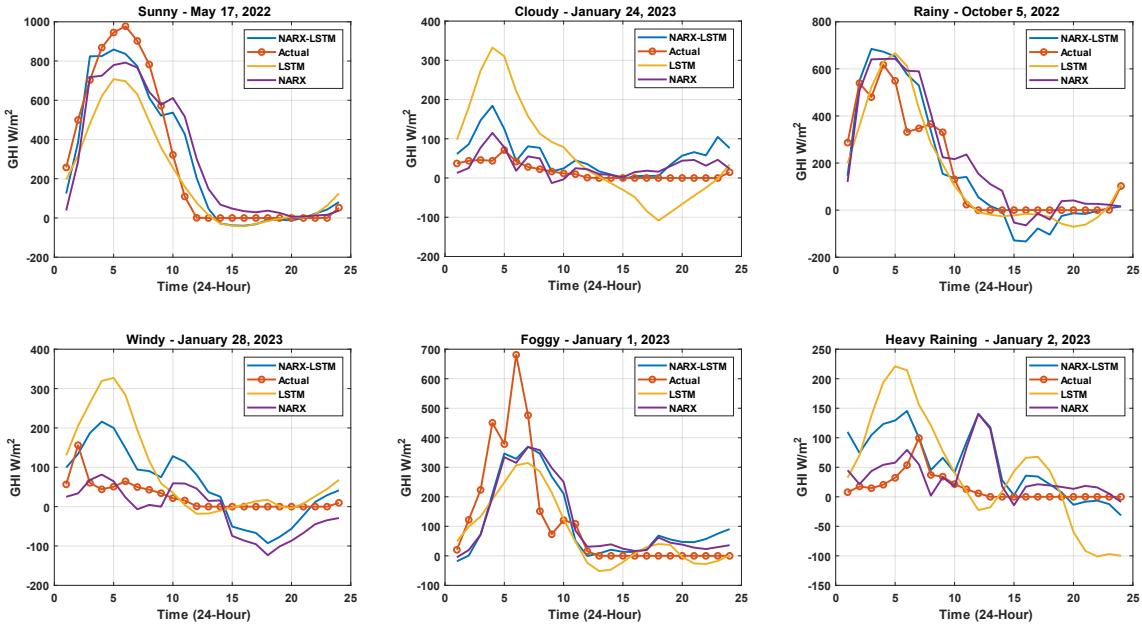


Figure 4.9 Assessment of the NARX-LSTM model's performance in predicting solar irradiance for a 24-hour forecast period across various meteorological scenarios with excluding (cloud attenuation variable).

As illustrated in Figure 4.9, the hybrid NARX-LSTM model, though the most consistent, still exhibits notable deficiencies, particularly in extreme weather conditions. The standalone LSTM model, characterized by its considerable variability, struggles with rapid environmental changes, leading to unreliable forecasts. Similarly, the standalone NARX model, despite occasional superiority over LSTM in specific situations, generally lacks consistency in complex weather patterns.

Moving to the second test of the Input Sensitivity Analysis, this segment particularly focused on assessing the influence of the wind direction variable by excluding it from the dataset.

Under sunny conditions as shown in Figure 4.10, as seen on May 17, 2022, the models were highly accurate, with the NARX-LSTM hybrid showing an exceptionally close fit to the actual data, suggesting that under clear skies, the absence of the wind direction variable does not significantly impair model performance. However, as weather conditions become more complex, such as during cloudy and rainy days, the

models exhibit greater variability in their predictions, with the NARX-LSTM hybrid consistently aligning more closely with actual GHI readings compared to standalone NARX and LSTM models. This implies that while the models can capture the overall trend, certain nuances, possibly including wind direction, are not fully accounted for, leading to discrepancies.

The significance of including wind direction is further underscored in the predictions for windy conditions on January 28, 2023. Here, the models, particularly NARX, diverged notably from the actual measurements, indicating that wind might have a non-trivial impact on solar irradiance predictions, an impact that is not as effectively captured when wind direction is excluded. Foggy conditions, observed on January 1, 2023, and heavy rain, observed on January 2, 2023, further challenged the models, with all exhibiting increased prediction errors. The NARX model, in particular, displayed substantial deviations, suggesting a heightened sensitivity to the absence of wind direction input.

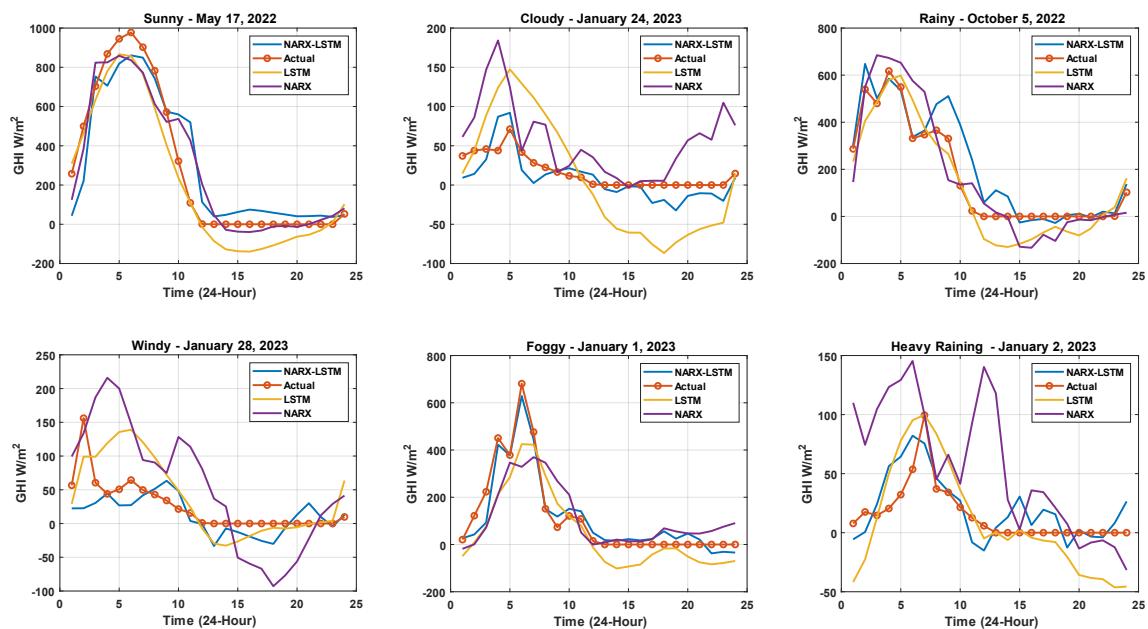


Figure 4.10 Assessment of the NARX-LSTM model's performance in predicting solar irradiance for a 24-hour (cloud attenuation variable).

The analysis suggests that wind direction is an important variable in accurately predicting GHI, especially in models relying on NARX architecture. Wind direction can affect the distribution and movement of clouds, fog, and rain—all of which have direct implications for solar irradiance. Therefore, omitting wind direction from the model inputs can lead to significant inaccuracies, underscoring the variable's importance in capturing the dynamic and complex interactions between weather elements and solar irradiance. In essence, for enhancing the precision of GHI predictions, especially in adverse weather conditions, the inclusion of wind direction seems not only beneficial but perhaps crucial.

Through this analytical procedure, parameters such as cloud cover, air temperature, surface pressure, and wind direction emerged as particularly significant based on the conducted analysis and as shown in Table 4.3 and Figure 4.11. The criticality of cloud cover is underscored by its inverse correlation with solar irradiance (SI), where higher cloud opacity notably reduces SI. Similarly, wind direction's influence is linked to its role in cloud movement and local weather dynamics, while surface pressure correlates with atmospheric stability, affecting SI through weather patterns associated with clear or cloudy skies. The positive association between air temperature and SI further refines the model's inputs, leveraging temperature as a predictor for clearer skies and thus higher SI in less cloudy regions.

Table 4.2 Feature importance result

<b>Feature</b>	<b>Feature Importance Score:</b>
Cloud Attenuation	24.652
Surface Pressure	21.054
Precipitation rate	15.294
Wind Direction	11.291
Wind Speed	11.047
Precipitable water	8.609
Air temperature	2.801
Relative Humidity	1.875
Dewpoint Temperature	1.713

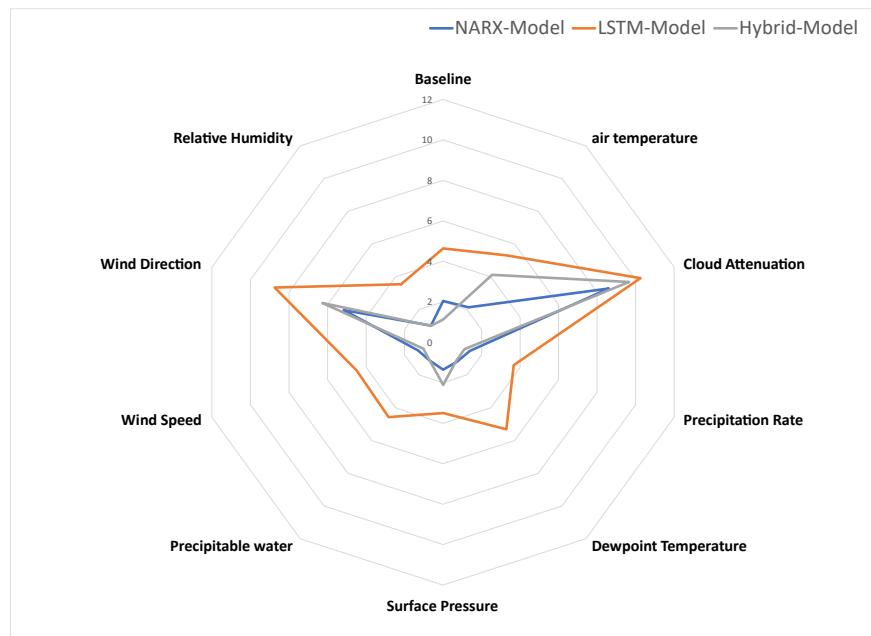


Figure 4.11 Comparative Performance of NARX, LSTM, and Hybrid Models Across Meteorological Parameters

Table 4.3 Input Sensitivity Analysis

Feature (Removing action)	nRMSE %		
	LSTM	NARX	NARX-LSTM
Cloud Attenuation	10.25	8.610	9.66
Surface Pressure	3.498	1.349	2.110
Precipitation rate	3.667	1.388	1.111
Wind Direction	8.739	5.158	6.271
Precipitable water	4.573	1.118	1.115
Wind Speed	4.493	1.303	1.021
Air temperature	5.313	2.138	4.117
Relative Humidity	3.540	1.014	1.010
Dewpoint temp.	5.316	1.175	1.112

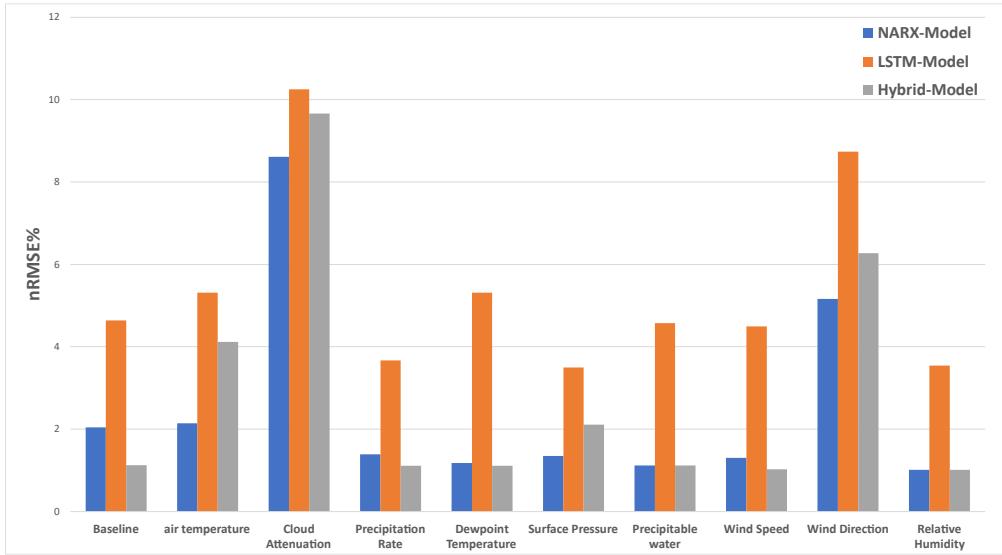


Figure 4.12 Performance Comparison of NARX, LSTM, and Hybrid Models.

#### 4.4 Evaluation of The Proposed Model

The approach begins with the implementation of hyperparameter tuning to optimize the NARX and LSTM model configurations for superior forecasting accuracy. A meticulous parameter space reduction process is conducted to expedite computational efficiency. The optimal configuration for the NARX model with a Levenberg-Marquardt optimization algorithm and a specific learning rate, while the LSTM network's architecture comprises 150 hidden units with dropout regularization. The training process employs strategies like early stopping to prevent overfitting, ensuring that the models generalize well to unseen data.

The scatter plot serves as a pivotal tool for visual validation, plotting the model's predicted values against the actual solar irradiance readings. The near-perfect alignment of these values, denoted by an R-value of 0.99911 as presented in Figure 4.13, attests to the model's exceptional ability to discern, and replicate the intrinsic patterns within the training data. Such a high degree of correlation is indicative of the model's prowess in forecasting GHI with remarkable fidelity.

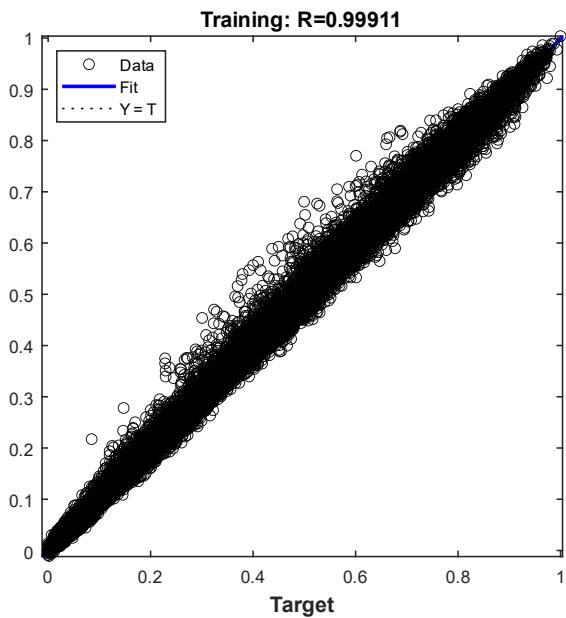


Figure 4.13 scatter plots of model

Complementing the scatter plot, the error distribution graph provides a quantitative analysis of the prediction accuracy. The concentration of errors around the zero mark is a strong testament to the model's precision, with the majority of the forecasted values deviating minimally from their actual counterparts as shown in Figure 4.14. This dense aggregation of errors near zero is a promising sign of the model's effectiveness.

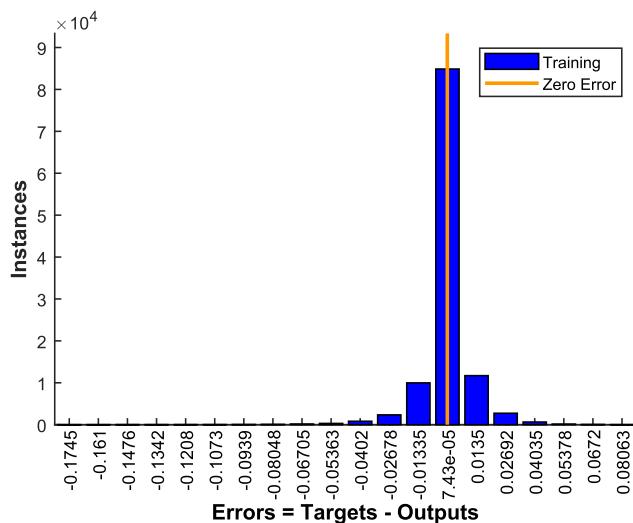


Figure 4.14 Error Distributions

However, the presence of outliers in the error distribution—points where the forecasted values significantly diverge from the actual data—cannot be overlooked. These outliers manifest as sporadic instances of higher error magnitudes and are particularly instructive for iterative model enhancement. They underscore the areas where the model may struggle, possibly due to complex weather dynamics or anomalous data points, which are not fully captured by the current model architecture or feature set. Such deviations present opportunities for in-depth analysis, offering valuable direction for further refinement of the model to achieve even higher levels of predictive accuracy. These insights pave the way for ongoing model optimization, ensuring that the forecasting system remains robust and reliable in the face of diverse meteorological phenomena.

During the assessment of the model's training efficacy as shown in Figure 4.15, the evaluation metrics revealed varying levels of predictive error across the different models. The LSTM network, while robust, exhibited the highest normalized Root Mean Square Error (nRMSE) as shown in Table 4.4, a measure that quantifies the deviation between the model's predictions and the actual observed values. This metric serves as an indicator of the error magnitude relative to the range of the data, with a lower nRMSE being preferable as it denotes greater predictive accuracy.

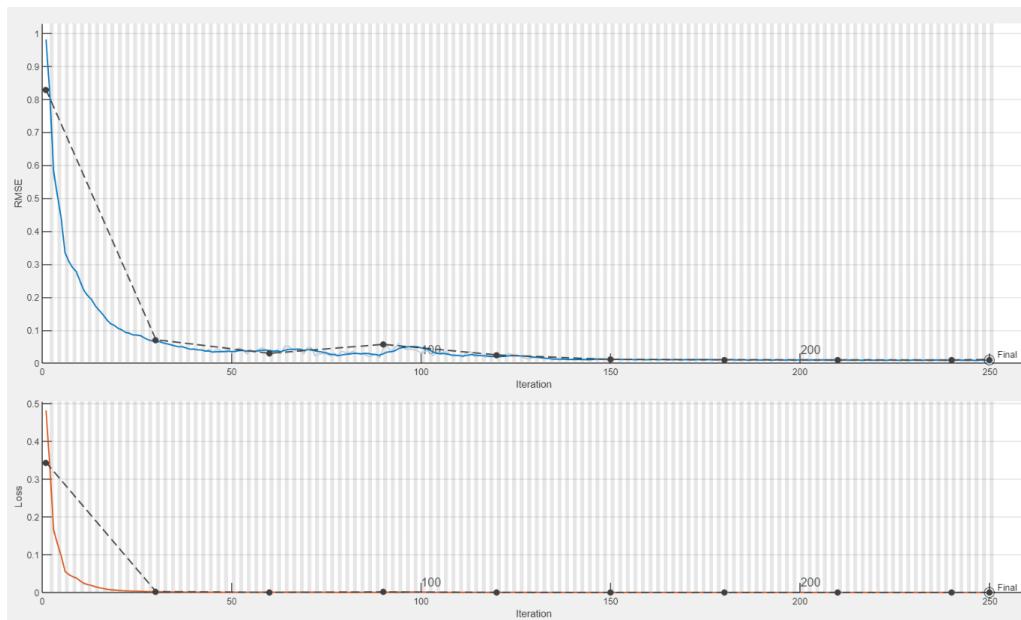


Figure 4.15 Training response

In contrast, the NARX model demonstrated a superior predictive capability compared to the standalone LSTM, as reflected by its lower nRMSE. This indicates that the NARX model's structure and learning mechanism were more effective in capturing the complex nonlinear relationships inherent in the solar irradiance data.

The most noteworthy performance, however, was exhibited by the hybrid NARX-LSTM model as resulted in Table 4.4. This combined approach leveraged the strengths of both NARX and LSTM architectures, resulting in the lowest nRMSE among the tested models. The success of the NARX-LSTM hybrid underscores the synergistic potential of combining different modeling techniques to enhance forecast accuracy. As delineated in Table 4.4 of the thesis, the hybrid model's superior performance in reducing predictive error underscores its potential as a more reliable tool for solar irradiance forecasting. The empirical results thus advocate for the adoption of hybrid modeling frameworks in complex time-series prediction tasks, where capturing dynamic, multifaceted patterns is crucial for achieving high levels of accuracy.

Table 4.4 Training performance result

	<b>LSTM</b>	<b>NARX</b>	<b>NARX-LSTM</b>
<b>nRMSE %</b>	10.245	2.0245	1.046

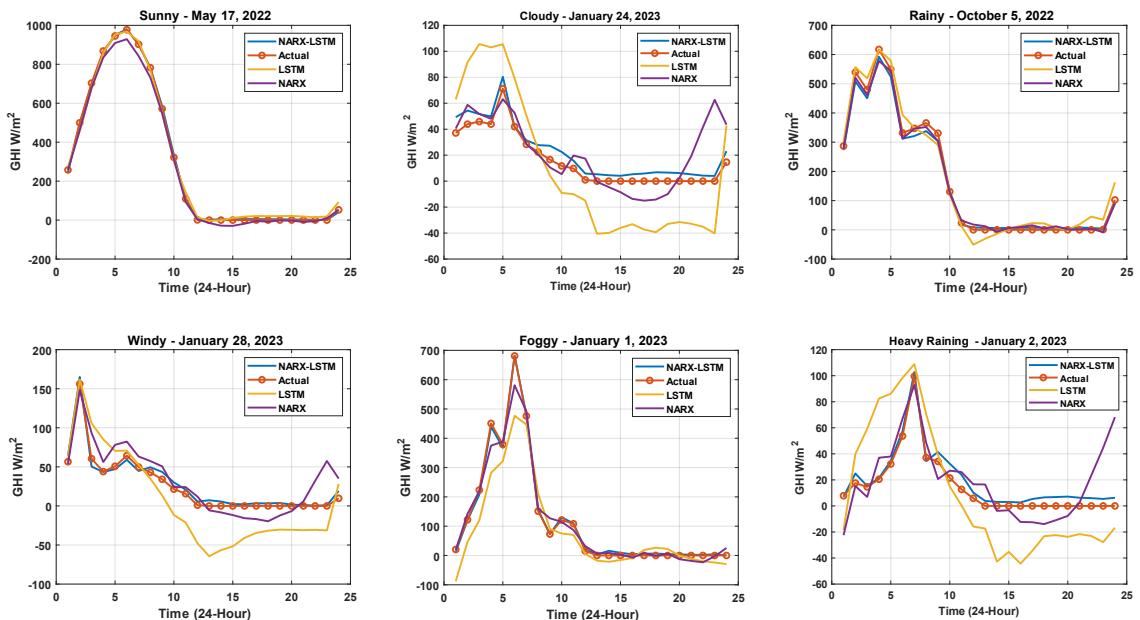


Figure 4.16 Assessment of the NARX-LSTM model's performance in predicting solar irradiance for a 24-hour forecast period across various meteorological scenarios with the selected variables.

The detailed result of the hybrid NARX-LSTM model's performance compared to the standalone LSTM and NARX models, as indicated by Figure 4.16, illustrates several key insights. The hybrid model consistently achieves lower nRMSE values across different weather conditions as shown in Table 4.5, suggesting a more refined ability to synthesize and predict based on complex data inputs.

The hybrid NARX-LSTM model's proficiency is prominently displayed under the spectrum of weather conditions that directly influence solar irradiance. Notably, during periods of sunshine, the model exhibits exemplary performance, closely mirroring the actual irradiance levels with minimal deviation, as illustrated in Figure 4.16. This accuracy is reflective of its adept learning and generalization capabilities, honed to capture the subtleties of solar patterns on clear days.

The true test of the model's mettle, however, is presented under overcast conditions, where the unpredictable dynamics of cloud movements pose a significant

challenge to predictive accuracy. Even in such scenarios, as shown in Figure 4.16 the hybrid model manages to maintain a lead over the NARX and LSTM models, albeit with a diminished advantage. This slight convergence in model performance under cloudy skies is indicative of the universally complex task of modeling solar irradiance amidst variable cloud cover.

The competitive edge of the proposed hybrid model is further asserted in scenarios of rainfall, where the interplay of clouds and precipitation further complicates the prediction landscape. While the hybrid model exhibits a commendable performance in such conditions, its comparability to the NARX model during rainy weather implies potential areas for enhancement, especially in the nuanced interpretation of complex cloud formations and the hydrological impact on irradiance.

The robustness of the model is rigorously tested and proven in adverse weather conditions such as fog, wind, and heavy rain. Under these challenging circumstances, the model not only maintains its forecasting accuracy but also shows a significant reduction in nRMSE as can be seen in Table 4.5, emphasizing its ability to adapt to sudden meteorological fluctuations. This adaptability is crucial for applications in real-time energy management and ensuring the stability of the power grid, where rapid response to changing weather conditions is imperative. The hybrid model's resilience in the face of environmental volatility, therefore, marks a significant step forward in the realm of solar irradiance forecasting, paving the way for more reliable and dynamic predictive systems.

Table 4.5 Tasting performance result

Condition:	nRMSE%		
	NARX-LSTM	LSTM	NARX
Sunny	1.195	5.986	6.534
Cloudy	6.813	18.457	12.815
Rainy	3.871	7.436	3.205
Heavy Raining	5.567	20.093	11.494

Foggy	2.458	12.870	4.205
Windy	1.971	9.208	6.208

The LSTM model, recognized for its utility in time-series analysis due to its capacity to capture temporal dependencies, demonstrates a commendable level of performance. However, its efficacy is eclipsed by the hybrid model, which suggests that the LSTM model, in isolation, may not fully grasp the intricate interplay between various weather parameters that affect solar irradiance. This limitation is addressed by the hybrid NARX-LSTM model, which integrates the strengths of both LSTM's temporal processing and NARX's nonlinear autoregressive capabilities, thereby enriching the model's comprehension of complex weather dynamics.

On the other hand, the standalone NARX model exhibits a degree of competitiveness, particularly in scenarios characterized by rainy conditions where short-term weather fluctuations are predominant. Nevertheless, it tends to lag in overall performance when compared to the hybrid model. The NARX model's relative shortfall could stem from its intrinsic design, which might not be as adept at capturing and leveraging long-term dependencies within the dataset as effectively as the hybrid approach.

The comparative analysis thus illuminates the distinctive advantages of combining NARX and LSTM models. By harnessing the specialized attributes of each model, the hybrid configuration excels in decoding both the short-term and long-term patterns in weather data, facilitating a more accurate and reliable forecast of solar irradiance, which is critical for the optimization of solar energy systems.

## **4.5 Chapter Summary**

This chapter combined EDA with the results of the NARX-LSTM model, providing a comprehensive view of the dataset and its influence on forecasting performance. Input sensitivity analysis demonstrated the significance of key meteorological variables in shaping prediction accuracy. Hyperparameter tuning and rigorous evaluation highlighted the hybrid NARX-LSTM model's superior performance, particularly in handling diverse weather conditions. The hybrid model consistently achieved lower normalized Root Mean Square Error (nRMSE) values across various scenarios, showcasing its robustness and adaptability. These findings underscore the importance of integrating robust EDA with advanced modeling techniques to enhance the precision and reliability of solar irradiance predictions.

## CHAPTER 5

### CONCLUSION

#### 5.1 Summary

This study has significantly contributed to advancing the understanding and methodologies in the domain of solar irradiance forecasting. The research began with an in-depth analysis of key meteorological factors, including wind direction, and their influence on predictive models. These findings highlighted the importance of selecting appropriate features to achieve high accuracy in solar irradiance predictions. The rigorous optimization of the NARX and LSTM models through hyperparameter tuning culminated in the development of a robust hybrid NARX-LSTM model. This hybrid model emerged as a superior forecasting tool, consistently outperforming standalone NARX and LSTM models across various weather conditions.

The hybrid model's robustness and adaptability were demonstrated by its ability to achieve lower normalized Root Mean Square Error (nRMSE) values under diverse scenarios. These results underscore its potential as a reliable solution for accurate solar irradiance forecasting, particularly in addressing the challenges posed by varying meteorological conditions. Furthermore, the findings reinforce the relevance of hybrid modeling approaches in enhancing prediction accuracy, paving the way for their integration into real-world energy management systems and grid stability frameworks.

#### 5.2 Future Work

While this study has advanced the application of hybrid NARX and LSTM models for solar irradiance forecasting, it also paves the way for numerous exciting research opportunities. Future studies could delve into the exploration of other

sophisticated machine learning techniques. While the current hybrid model shows promise, investigating the potential of models such as gated recurrent units (GRUs) or advanced ensemble methods could offer improvements in prediction accuracy and reliability.

Expanding the dataset is an essential step for future work. The efficacy of machine learning models is highly dependent on the diversity and quality of the data. Incorporating solar irradiance data from a wider array of geographic locations, including varied topographies and climates, would undoubtedly bolster the model's predictive power and its ability to generalize across different environments.

The concept of a multi-model ensemble system also presents an intriguing avenue for future research. Such a system could combine various models, including but not limited to, decision trees, support vector machines, and newer forms of neural networks, each specializing in different aspects of the forecasting task. The ensemble approach could mitigate individual model biases and errors, potentially leading to a consensus prediction that is more accurate and reliable than any single model output.

In terms of model interpretability, future work could apply methods like Layer-wise Relevance Propagation (LRP) or SHapley Additive exPlanations (SHAP) to decompose model predictions into contributions from each input feature. These methods can provide insights into the causal relationships within the data and help validate the model against established physical principles of solar irradiance

Lastly, with the advent of smart grid technologies, there is an opportunity to integrate predictive models directly into energy management systems. Future research could focus on real-time adaptive models that work in tandem with smart grids to optimize energy distribution based on predicted solar irradiance, thereby enhancing the efficiency and sustainability of solar energy utilization.

The roadmap for future work is as vast as it is vital, with each step bringing us closer to more sustainable and efficient solar energy management systems.

## REFERENCES

- Adcock and Dexter, 2021. Adcock, B. and Dexter, N. (2021). The gap between theory and practice in function approximation with deep neural networks. *SIAM Journal on Mathematics of Data Science*, 3(2):624–655.
- Adefarati and Bansal, 2019. Adefarati, T. and Bansal, R. C. (2019). Reliability, economic and environmental analysis of a microgrid system in the presence of renewable energy resources. *Applied energy*, 236:1089–1114.
- Akhter et al., 2019. Akhter, M. N., Mekhilef, S., Mokhlis, H., and Mohamed Shah, N. (2019). Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. *IET Renewable Power Generation*, 13(7):1009–1023.
- Alamo et al., 2019. Alamo, D. H., Medina, R. N., Ruano, S. D., García, S. S., Moustris, K. P., Kavadias, K. K., Zafirakis, D., Tzanes, G., Zafeiraki, E., Spyropoulos, G., et al. (2019). An advanced forecasting system for the optimum energy management of island microgrids. *Energy procedia*, 159:111–116.
- Albarakati et al., 2022. Albarakati, A. J., Boujoudar, Y., Azeroual, M., Eliysaouy, L., Kotb, H., Aljarbouh, A., Khalid Alkahtani, H., Mostafa, S. M., Tassaddiq, A., and Pupkov, A. (2022). Microgrid energy management and monitoring systems: A comprehensive review. *Frontiers in Energy Research*, 10:1097858.
- Alizamir et al., 2020. Alizamir, M., Kim, S., Kisi, O., and Zounemat-Kermani, M. (2020). A comparative study of several machine learning based non-linear regression methods in estimating solar radiation: Case studies of the usa and turkey regions. *Energy*, 197:117239.
- AlKandari and Ahmad, 2024. AlKandari, M. and Ahmad, I. (2024). Solar power generation forecasting using ensemble approach based on deep learning and statistical methods. *Applied Computing and Informatics*, 20(3/4):231–250.
- Alsaidan et al., 2017. Alsaidan, I., Alanazi, A., Gao, W., Wu, H., and Khodaei, A. (2017). State-of-the-art in microgrid-integrated distributed energy storage sizing. *Energies*, 10(9):1421.

- Bajaj and Singh, 2020. Bajaj, M. and Singh, A. K. (2020). Grid integrated renewable dg systems: A review of power quality challenges and state-of-the-art mitigation techniques. *International Journal of Energy Research*, 44(1):26–69.
- Bandeiras et al., 2020. Bandeiras, F., Pinheiro, E., Gomes, M., Coelho, P., and Fernandes, J. (2020). Review of the cooperation and operation of microgrid clusters. *Renewable and Sustainable Energy Reviews*, 133:110311.
- Bank, 2020. Bank, A. D. (2020). *Handbook on Microgrids for Power Quality and Connectivity*. Asian Development Bank Institute.
- Behera and Nayak, 2020. Behera, M. K. and Nayak, N. (2020). A comparative study on short-term pv power forecasting using decomposition based optimized extreme learning machine algorithm. *Engineering Science and Technology, an International Journal*, 23(1):156–167.
- Cagnano et al., 2020. Cagnano, A., De Tuglie, E., and Mancarella, P. (2020). Microgrids: Overview and guidelines for practical implementations and operation. *Applied Energy*, 258:114039.
- Chen et al., 2019. Chen, Y., Chang, H., Meng, J., and Zhang, D. (2019). Ensemble neural networks (enn): A gradient-free stochastic method. *Neural Networks*, 110:170–185.
- Choi and Hur, 2020. Choi, S. and Hur, J. (2020). An ensemble learner-based bagging model using past output data for photovoltaic forecasting. *Energies*, 13(6):1438.
- Choudhury, 2020. Choudhury, S. (2020). A comprehensive review on issues, investigations, control and protection trends, technical challenges and future directions for microgrid technology. *International Transactions on Electrical Energy Systems*, 30(9):e12446.
- Chung, 2020. Chung, M. H. (2020). Estimating solar insolation and power generation of photovoltaic systems using previous day weather data. *Advances in Civil Engineering*, 2020(1):8701368.
- Desai and Shah, 2021. Desai, M. and Shah, M. (2021). An anatomization on breast cancer detection and diagnosis employing multi-layer perceptron neural

- network (mlp) and convolutional neural network (cnn). *Clinical eHealth*, 4:1–11.
- Diagne et al., 2013. Diagne, M., David, M., Lauret, P., Boland, J., and Schmutz, N. (2013). Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renewable and Sustainable Energy Reviews*, 27:65–76.
- Dudek et al., 2023. Dudek, G., Piotrowski, P., and Baczyński, D. (2023). Intelligent forecasting and optimization in electrical power systems: Advances in models and applications.
- Dutta et al., 2017. Dutta, S., Li, Y., Venkataraman, A., Costa, L. M., Jiang, T., Plana, R., Tordjman, P., Choo, F. H., Foo, C. F., and Puttgen, H. B. (2017). Load and renewable energy forecasting for a microgrid using persistence technique. *Energy Procedia*, 143:617–622.
- El Hendouzi and Bourouhou, 2016. El Hendouzi, A. and Bourouhou, A. (2016). Forecasting of pv power application to pv power penetration in a microgrid. In *2016 International Conference on Electrical and Information Technologies (ICEIT)*, pages 468–473. IEEE.
- Feng et al., 2020. Feng, Y., Hao, W., Li, H., Cui, N., Gong, D., and Gao, L. (2020). Machine learning models to quantify and map daily global solar radiation and photovoltaic power. *Renewable and Sustainable Energy Reviews*, 118:109393.
- Gheouany et al., 2023. Gheouany, S., Ouadi, H., Giri, F., and El Bakali, S. (2023). Experimental validation of multi-stage optimal energy management for a smart microgrid system under forecasting uncertainties. *Energy Conversion and Management*, 291:117309.
- Guermoui et al., 2020. Guermoui, M., Melgani, F., Gairaa, K., and Mekhalfi, M. L. (2020). A comprehensive review of hybrid models for solar radiation forecasting. *Journal of Cleaner Production*, 258:120357.
- Gungor et al., 2011. Gungor, V. C., Sahin, D., Kocak, T., Ergut, S., Buccella, C., Cecati, C., and Hancke, G. P. (2011). Smart grid technologies: Communication technologies and standards. *IEEE transactions on Industrial informatics*, 7(4):529–539.

- Guo and Mu, 2016. Guo, W. and Mu, L. (2016). Control principles of micro-source inverters used in microgrid. *Protection and Control of Modern Power Systems*, 1:1–7.
- He et al., 2019. He, H., Yan, Y., Chen, T., and Cheng, P. (2019). Tree height estimation of forest plantation in mountainous terrain from bare-earth points using a dog-coupled radial basis function neural network. *Remote sensing*, 11(11):1271.
- Hewamalage et al., 2021. Hewamalage, H., Bergmeir, C., and Bandara, K. (2021). Recurrent neural networks for time series forecasting: Current status and future directions. *International Journal of Forecasting*, 37(1):388–427.
- Hornik et al., 1989. Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366.
- Hussain et al., 2019. Hussain, A., Bui, V.-H., and Kim, H.-M. (2019). Microgrids as a resilience resource and strategies used by microgrids for enhancing resilience. *Applied energy*, 240:56–72.
- James, 2013. James, G. (2013). An introduction to statistical learning.
- Kalakotla and Korra, 2023. Kalakotla, S. and Korra, C. (2023). Emerging power quality challenges due to integration of renewable energy sources in ac/dc microgrids. In *International Conference on Intelligent Manufacturing and Energy Sustainability*, pages 293–304. Springer.
- Kaushal and Basak, 2020. Kaushal, J. and Basak, P. (2020). Power quality control based on voltage sag/swell, unbalancing, frequency, thd and power factor using artificial neural network in pv integrated ac microgrid. *Sustainable Energy, Grids and Networks*, 23:100365.
- Khan et al., 2020. Khan, P. W., Byun, Y.-C., Lee, S.-J., Kang, D.-H., Kang, J.-Y., and Park, H.-S. (2020). Machine learning-based approach to predict energy consumption of renewable and nonrenewable power sources. *Energies*, 13(18):4870.
- Kiehbadroudinezhad et al., 2023. Kiehbadroudinezhad, M., Merabet, A., Ghenai, C., Abo-Khalil, A. G., and Salameh, T. (2023). The role of biofuels for

- sustainable microgridsf: A path towards carbon neutrality and the green economy. *Heliyon*, 9(2).
- Lara-Fanego et al., 2012. Lara-Fanego, V., Ruiz-Arias, J., Pozo-Vázquez, D., Santos-Alamillos, F., and Tovar-Pescador, J. (2012). Evaluation of the wrf model solar irradiance forecasts in andalusia (southern spain). *Solar Energy*, 86(8):2200–2217.
- Lasseter and Paigi, 2004. Lasseter, R. H. and Paigi, P. (2004). Microgrid: A conceptual solution. In *2004 IEEE 35th annual power electronics specialists conference (IEEE Cat. No. 04CH37551)*, volume 6, pages 4285–4290. IEEE.
- Lee et al., 2009. Lee, H., Hong, S., and Kim, E. (2009). Neural network ensemble with probabilistic fusion and its application to gait recognition. *Neurocomputing*, 72(7-9):1557–1564.
- Lei et al., 2023. Lei, B., Ren, Y., Luan, H., Dong, R., Wang, X., Liao, J., Fang, S., and Gao, K. (2023). A review of optimization for system reliability of microgrid. *Mathematics*, 11(4):822.
- Lei and Yang, 2019. Lei, Z. and Yang, Y.-b. (2019). Research on data mining algorithm for regional photovoltaic generation. In *Advanced Hybrid Information Processing: Third EAI International Conference, ADHIP 2019, Nanjing, China, September 21–22, 2019, Proceedings, Part I*, pages 429–438. Springer.
- Logenthiran et al., 2015. Logenthiran, T., Naayagi, R. T., Woo, W. L., Phan, V.-T., and Abidi, K. (2015). Intelligent control system for microgrids using multiagent system. *IEEE Journal of Emerging and Selected Topics in Power Electronics*, 3(4):1036–1045.
- Lu et al., 2016. Lu, X., Wang, J., and Guo, L. (2016). Using microgrids to enhance energy security and resilience. *The Electricity Journal*, 29(10):8–15.
- Ma and Ma, 2018. Ma, J. and Ma, X. (2018). A review of forecasting algorithms and energy management strategies for microgrids. *Systems Science & Control Engineering*, 6(1):237–248.
- Mariam, 2018. Mariam, L. (2018). Modelling of an intelligent microgrid system in a smart grid network.

- Mariam et al., 2013. Mariam, L., Basu, M., and Conlon, M. F. (2013). A review of existing microgrid architectures. *Journal of engineering*, 2013(1):937614.
- Mariam et al., 2016. Mariam, L., Basu, M., and Conlon, M. F. (2016). Microgrid: Architecture, policy and future trends. *Renewable and Sustainable Energy Reviews*, 64:477–489.
- Marinelli et al., 2014. Marinelli, M., Sossan, F., Costanzo, G. T., and Bindner, H. W. (2014). Testing of a predictive control strategy for balancing renewable sources in a microgrid. *IEEE Transactions on Sustainable Energy*, 5(4):1426–1433.
- Massaoudi et al., 2019. Massaoudi, M., Chihi, I., Sidhom, L., Trabelsi, M., and Oueslati, F. S. (2019). Medium and long-term parametric temperature forecasting using real meteorological data. In *IECON 2019-45th Annual Conference of the IEEE Industrial Electronics Society*, volume 1, pages 2402–2407. IEEE.
- Massaoudi et al., 2020. Massaoudi, M., Refaat, S. S., Chihi, I., Trabelsi, M., Abu-Rub, H., and Oueslati, F. S. (2020). Short-term electric load forecasting based on data-driven deep learning techniques. In *IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society*, pages 2565–2570. IEEE.
- Mattar et al., 2024. Mattar, E.-Z. M., Mahmoud, E. S., and El-Sayed, M. I. (2024). Mitigation of voltage sag and voltage swell by using dynamic voltage restorer. *International Journal of Power Electronics and Drive Systems (IJPEDS)*, 15(2):1290–1299.
- Mienye et al., 2024. Mienye, I. D., Swart, T. G., and Obaido, G. (2024). Recurrent neural networks: A comprehensive review of architectures, variants, and applications. *Information*, 15(9):517.
- Mishra and Ramesh, 2009. Mishra, A. K. and Ramesh, L. (2009). Application of neural networks in wind power (generation) prediction. In *2009 International Conference on Sustainable Power Generation and Supply*, pages 1–5. IEEE.
- Mohamed et al., 2015. Mohamed, M. A., Eltamaly, A. M., Farh, H. M., and Alolah, A. I. (2015). Energy management and renewable energy integration in smart grid system. In *2015 IEEE international conference on smart energy grid engineering (SEGE)*, pages 1–6. IEEE.

- Murphy, 2012. Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.
- Nespoli et al., 2019. Nespoli, A., Mussetta, M., Ogliari, E., Leva, S., Fernández-Ramírez, L., and García-Triviño, P. (2019). Robust 24 hours ahead forecast in a microgrid: A real case study. *Electronics*, 8(12):1434.
- Patterson and Gibson, 2017. Patterson, J. and Gibson, A. (2017). *Deep learning: A practitioner's approach*. "O'Reilly Media, Inc."
- Pelland et al., 2013. Pelland, S., Remund, J., Kleissl, J., Oozeki, T., and De Brabandere, K. (2013). Photovoltaic and solar forecasting: state of the art. *IEA PVPS Task*, 14(355):1–36.
- Pu and Kalnay, 2019. Pu, Z. and Kalnay, E. (2019). Numerical weather prediction basics: Models, numerical methods, and data assimilation. *Handbook of hydrometeorological ensemble forecasting*, pages 67–97.
- Qu et al., 2021. Qu, J., Qian, Z., and Pei, Y. (2021). Day-ahead hourly photovoltaic power forecasting using attention-based cnn-lstm neural network embedded with multiple relevant and target variables prediction pattern. *Energy*, 232:120996.
- Raimi et al., 2024. Raimi, D., Zhu, Y., Newell, R. G., and Prest, B. C. (2024). Global energy outlook 2024: Peaks or plateaus.
- Ren et al., 2015. Ren, Y., Suganthan, P., and Srikanth, N. (2015). Ensemble methods for wind and solar power forecasting—a state-of-the-art review. *Renewable and Sustainable Energy Reviews*, 50:82–91.
- Reza et al., 2022. Reza, S., Ferreira, M. C., Machado, J. J., and Tavares, J. M. R. (2022). A multi-head attention-based transformer model for traffic flow forecasting with a comparative analysis to recurrent neural networks. *Expert Systems with Applications*, 202:117275.
- Rodríguez et al., 2018. Rodríguez, F., Fleetwood, A., Galarza, A., and Fontán, L. (2018). Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control. *Renewable energy*, 126:855–864.

- Saeed et al., 2021. Saeed, M. H., Fangzong, W., Kalwar, B. A., and Iqbal, S. (2021). A review on microgrids' challenges & perspectives. *IEEE Access*, 9:166502–166517.
- Sangrody et al., 2020. Sangrody, H., Zhou, N., and Zhang, Z. (2020). Similarity-based models for day-ahead solar pv generation forecasting. *IEEE Access*, 8:104469–104478.
- Sepasi et al., 2023. Sepasi, S., Talichet, C., and Pramanik, A. S. (2023). Power quality in microgrids: A critical review of fundamentals, standards, and case studies. *IEEE Access*, 11:108493–108531.
- Shafiullah et al., 2010. Shafiullah, G., Oo, A. M., Jarvis, D., Ali, A. S., and Wolfs, P. (2010). Potential challenges: Integrating renewable energy with the smart grid. In *2010 20th australasian universities power engineering conference*, pages 1–6. IEEE.
- Shahgholian, 2021. Shahgholian, G. (2021). A brief review on microgrids: Operation, applications, modeling, and control. *International Transactions on Electrical Energy Systems*, 31(6):e12885.
- Shahzad et al., 2023. Shahzad, S., Abbasi, M. A., Ali, H., Iqbal, M., Munir, R., and Kilic, H. (2023). Possibilities, challenges, and future opportunities of microgrids: A review. *Sustainability*, 15(8):6366.
- Shakya et al., 2016. Shakya, A., Michael, S., Saunders, C., Armstrong, D., Pandey, P., Chalise, S., and Tonkoski, R. (2016). Solar irradiance forecasting in remote microgrids using markov switching model. *IEEE Transactions on sustainable Energy*, 8(3):895–905.
- Sharifzadeh et al., 2019. Sharifzadeh, M., Sikiniti-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*, 108:513–538.
- Singla et al., 2021. Singla, P., Duhan, M., and Saroha, S. (2021). A comprehensive review and analysis of solar forecasting techniques. *Frontiers in Energy*, pages 1–37.

- Sobri et al., 2018. Sobri, S., Koohi-Kamali, S., and Rahim, N. A. (2018). Solar photovoltaic generation forecasting methods: A review. *Energy conversion and management*, 156:459–497.
- Soman et al., 2010. Soman, S. S., Zareipour, H., Malik, O., and Mandal, P. (2010). A review of wind power and wind speed forecasting methods with different time horizons. In *North American power symposium 2010*, pages 1–8. IEEE.
- Sone et al., 2013. Sone, A., Kato, T., Shimakage, T., and Suzuoki, Y. (2013). Influence of forecast accuracy of photovoltaic power output on capacity optimization of microgrid composition under 30-minute power balancing control. *Electrical Engineering in Japan*, 182(2):20–29.
- Suman et al., 2021. Suman, G. K., Guerrero, J. M., and Roy, O. P. (2021). Optimisation of solar/wind/bio-generator/diesel/battery based microgrids for rural areas: A pso-gwo approach. *Sustainable cities and society*, 67:102723.
- Sutarna et al., 2023. Sutarna, N., Tjahyadi, C., Oktivasari, P., Dwijaniti, M., et al. (2023). Machine learning algorithm and modeling in solar irradiance forecasting. In *2023 6th International Conference of Computer and Informatics Engineering (IC2IE)*, pages 221–225. IEEE.
- Takilalte et al., 2022. Takilalte, A., Harrouni, S., and Mora, J. (2022). Forecasting global solar irradiance for various resolutions using time series models-case study: Algeria. *Energy sources, part A: Recovery, utilization, and environmental effects*, 44(1):1–20.
- Tomin et al., 2019. Tomin, N., Zhukov, A., and Domyshev, A. (2019). Deep reinforcement learning for energy microgrids management considering flexible energy sources. In *EPJ Web of Conferences*, volume 217, page 01016. EDP Sciences.
- Vincent et al., 2020. Vincent, R., Ait-Ahmed, M., Houari, A., and Benkhoris, M. F. (2020). Residential microgrid energy management considering flexibility services opportunities and forecast uncertainties. *International Journal of Electrical Power & Energy Systems*, 120:105981.
- Wang et al., 2020. Wang, F., Xuan, Z., Zhen, Z., Li, K., Wang, T., and Shi, M. (2020). A day-ahead pv power forecasting method based on lstm-rnn model and

time correlation modification under partial daily pattern prediction framework. *Energy Conversion and Management*, 212:112766.

Wani and Thagunna, 2024. Wani, N. M. and Thagunna, P. (2024). Predictive modeling of shear strength in fly ash-stabilized clayey soils using artificial neural networks and support vector regression. *Asian Journal of Civil Engineering*, 25(8):6131–6146.

Wen et al., 2020. Wen, H., Du, Y., Chen, X., Lim, E., Wen, H., Jiang, L., and Xiang, W. (2020). Deep learning based multistep solar forecasting for pv ramp-rate control using sky images. *IEEE Transactions on Industrial Informatics*, 17(2):1397–1406.

Yang et al., 2014. Yang, C., Thatte, A. A., and Xie, L. (2014). Multitime-scale data-driven spatio-temporal forecast of photovoltaic generation. *IEEE Transactions on Sustainable Energy*, 6(1):104–112.

Yoldaş et al., 2017. Yoldaş, Y., Önen, A., Muyeen, S., Vasilakos, A. V., and Alan, I. (2017). Enhancing smart grid with microgrids: Challenges and opportunities. *Renewable and Sustainable Energy Reviews*, 72:205–214.

Zafar et al., 2021. Zafar, R., Vu, B. H., Husein, M., and Chung, I.-Y. (2021). Day-ahead solar irradiance forecasting using hybrid recurrent neural network with weather classification for power system scheduling. *Applied Sciences*, 11(15):6738.

Zayed et al., 2022. Zayed, M. E., Zhao, J., Li, W., Sadek, S., and Elsheikh, A. H. (2022). Applications of artificial neural networks in concentrating solar power systems. In *Artificial neural networks for renewable energy systems and real-world applications*, pages 45–67. Elsevier.