Development of an Accurate Solar Photovoltaic Performance Analyser for Performance Ratio Prediction in PV plants: Supported with Experimental Investigation and Ground based Validation

Thesis submitted in
partial fulfilment for the
Award of Degree

Doctor of Philosophy

by

ALMAS

RAJIV GANDHI INSTITUTE OF PETROLEUM TECHNOLOGY
BENGALURU CAMPUS, HOSKOTE, BENGALURU, KARNATAKA,
INDIA-562165

20EB0003 ***2025***

CERTIFICATE

It is certified that the work contained in the thesis titled ***" Development of an Accurate Solar Photovoltaic Performance Analyser for Performance Ratio Prediction in PV plants: Supported with Experimental Investigation and Ground based Validation"*** by "Almas" has been carried out under my/our supervision and that this work has not been submitted elsewhere for a degree.

It is further certified that the student has fulfilled all the requirements of Comprehensive, Candidacy and SOTA.

Supervisor	**Co-Supervisor**
Dr. Sivasankari Sundaram	**Dr. Umakant Dhar Dwivedi**
Assistant Professor	**Associate Professor**
(RGIPT, Bengaluru Campus)	(RGIPT, Jais Campus)

DECLARATION BY THE CANDIDATE

I, ″Almas″, certify that the work embodied in this thesis is my own bona fide work and carried out by me under the supervision of ″Dr. Sivasankari Sundaram and Dr. Umakant Dhar Dwivedi″ from "Aug-2020" to “July-2025”, at the, Energy Institute Bengaluru, A Centre of Rajiv Gandhi Institute of Petroleum Technology, Jais now rechristened to RGIPT Bengaluru Campus. The matter embodied in this thesis has not been submitted for the award of any other degree. I declare that I have faithfully acknowledged and given credits to the research workers wherever their works have been cited in my work in this thesis. I further declare that I have not wilfully copied any other's work, paragraphs, text, data, results, *etc*., reported in journals, books, magazines, reports dissertations, theses, *etc*., or available at websites and have not included them in this thesis and have not cited as my own work.

Date:

Place: Bengaluru (*Almas)*

CERTIFICATE BY THE SUPERVISOR(S)

It is certified that the above statement made by the student is correct to the best of my/our knowledge.

Supervisor	Co-Supervisor
Dr. Sivasankari Sundaram	**Dr. Umakant Dhar Dwivedi**
Assistant Professor	**Associate Professor**
(RGIPT, Bengaluru Campus)	(RGIPT, Jais Campus)

Signature & Seal of Head of Department

CERTIFICATE

CERTIFIED that the work contained in the thesis titled "**Development of an Accurate Solar Photovoltaic Performance Analyser for Performance Ratio Prediction in PV plants: Supported with Experimental Investigation and Ground based Validation**" by Mrs. Almas has been carried out under my/our supervision. It is also certified that he/she fulfilled the mandatory requirement of TWO quality publications arose out of his/her thesis work.

It is further certified that the two publications (copies enclosed) of the aforesaid Mrs. Almas have been published in the Journals indexed by

(a) SCI

(b) SCI Extended

(c) SCOPUS

Dr. Sivasankari Sundaram	**Dr. Umakant Dhar Dwivedi**	**Dr. Sivasankari Sundaram**
(Supervisor)	(Co-Supervisor)	(Convener, DPGC)

COPYRIGHT TRANSFER CERTIFICATE

Title of the Thesis: Development of an Accurate Solar Photovoltaic Performance Analyser for Performance Ratio Prediction in PV plants: Supported with Experimental Investigation and Ground based Validation

Name of the Student: Almas

Copyright Transfer

The undersigned hereby assigns to the Rajiv Gandhi Institute of Petroleum Technology Bengaluru Campus all rights under copyright that may exist in and for the above thesis submitted for the award of the "Doctor of Philosophy**".**

Date: Signature of the Student

Place: Bengaluru (*Almas***)**

Note: However, the author may reproduce or authorize others to reproduce material extracted verbatim from the thesis or derivative of the thesis for author's personal use provided that the source and the Institute's copyright notice are indicated.

RAJIV GANDHI INSTITUTE OF PETROLEUM TECHNOLOGY

(BENGALURU CAMPUS)

* * *

FOR SUBMISSION OF PH.D. THESIS

PERSONAL DETAILS

Name of Student: Almas **Roll No.:** 20EB0003

Term of Registration: 16-Aug-2020

Category of Registration: UR [GEN]

Fulfilled required Minimum Residence period: 4 Semesters / **5** Semesters / **6** Semesters (Pl.tick)

Supervisor: Dr. Sivasankari Sundaram

Co-Supervisor: Dr. Umakant Dhar Dwivedi

Department: Rajiv Gandhi Institute of Petroleum Technology Bengaluru Campus

Title of Thesis: Development of an Accurate Solar Photovoltaic Performance Analyser for Performance Ratio prediction in PV plants: Supported with Experimental Investigation and Ground based Validation

Acknowledgment

I would like to express my sincere appreciation to **Rajiv Gandhi Institute of Petroleum Technology**, **Bengaluru campus** for providing me with the opportunity to fulfil the thesis requirement for the award of my **Doctor of Philosophy** degree. The teaching assistantship, research infrastructure, and knowledgeable faculty at the institution have been instrumental in facilitating my research work.

I am privileged to extend my deepest gratitude to my guide, **Dr. Sivasankari Sundaram**, for her exceptional guidance and unwavering support throughout my research journey. His relentless pursuit of excellence, tireless assistance, and constructive criticism have made my research both engaging and rewarding.

The unwavering support of the **Hon'ble Director of RGIPT** and the **Incharge of RGIPT, Bengaluru campus** was a constant source of motivation for me to complete my research work with exceptional dedication.

Being part of the **Solar Research Lab** at **RGIPT, Bengaluru campus** has been an exceptional privilege, and I would like to express my appreciation to my lab members, **Mr Abhijeet Rathore, Mr Raj Chawla** and **Mr Sudhanshu Kumar** for their assistance and fostering a friendly research environment.

I extend my heartfelt gratitude to my family for their unwavering support throughout my Ph.D. journey. Finally, I humbly acknowledge the grace of the Almighty.

Almas

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AC	Alternating Current	DC	Direct Current
AE-LSTM	AutoEncoder Long Short-Term Memory	DEL_T	Temperature Difference (°C)
AI	Artificial Intelligence	DL	Deep Learning
ANOVA	Analysis of Variance	EDA	Exploratory Data Analysis
ARIMA	Autoregressive Integrated Moving Average	CSV	Comma Separated Values
CF	Capacity Factor	CUF	Capacity Utilization Factor
CI	Confidence Interval	EL	Electroluminescence
CNN	Convolutional Neural Network	FF	Fill Factor
CR	Compression Ratio	FFBP	Feed Forward Back Propagation
GAP	Global Average Pooling	GHI	Global Horizontal Irradiance
GHGs	Greenhouse Gas	GUI	Graphical User Interface
GRU	Gated Recurrent Unit	IMD	Indian Meteorological Department
IoT	Internet of Things	IR	Infrared

JNNSM	Jawaharlal Nehru National Solar Mission	LCOE	Levelized Cost of Energy
LSTM	Long Short-Term Memory	GHI	Global Horizontal Irradiance
MAE	Mean Absolute Error	GRU	Gated Recurrent Unit
MAPE	Mean Absolute Percentage Error	RNN	Recurrent Neural Network
MBE	Mean Bias Error	RT	Recording Time
ML	Machine Learning	R_d	Degradation Rate
MLP	Multi-Layer Perceptron	ReLU	Rectified Linear Unit
MNRE	Ministry of New and Renewable Energy	SAM	System Advisor Model
MSE	Mean Squared Error	SCADA	Supervisory Control and Data Acquisition
NPV	Net Present Value	SDM	Single Diode Model
NREL	National Renewable Energy Laboratory	Tkinter	Python GUI Library
PLR	Performance Loss Rate	V-I	Voltage-Current
POA	Plane of Array	Wp	Watt peak
PPA	Power Purchase Agreement	Ya	Array Yield
PR	Performance Ratio	Yf	Final Yield

PRLSTM	Performance Ratio Long Short-Term Memory	R_{se}	Series Resistance
PV	Photovoltaic	P_m	Maximum output power
RMSE	Root Mean Square Error	SE	Standard Error
SLDC	Single Line Diagram	SD	Standard Deviation
STC	Standard Test Conditions	CI	Confidence Interval
ENV	Environmental Variables	BiLSTM	Bidirectional Long Short-Term M
Yr	Reference Yield	CNN	Convolutional Neural Network
kWp	kilowatt peak	FF	Fill Factor
FFBP	Feed Forward Back Propagation	GAP	Global Average Pooling
IEC	International Electrotechnical Commission	GRU	Gated Recurrent Unit
RTD	Resistance Temperature Detector		

List of Symbols

I	Irradiance	Pm	Measured Power Output
Isc	Short Circuit Current	PR	Performance Ratio
Imp	Current at Maximum Power	P_actual	Actual power output
L	Number of samples	P_rated	Rated maximum power
P	Pooling window size	T	Temperature / Time step
S(t)	Cell state at time t	Ta	Ambient Temperature
K_T	Total Degradation Rate (%/year)	Tm	Module Temperature
R	Discount Rate	Tpv_{max} / Tpv_{min}	Maximum / Minimum Module Temperature Limits
T	Time / current time instance	Ts	Apparent sun temperature (typically 6000 K)

t+1, t+2	Future forecasted time instances	V	Voltage
tanh	Hyperbolic tangent function	V_{OC}	Open Circuit Voltage
A	Temperature coefficient (%/°C)	V_{mp}	Voltage at Maximum Power
η_ex	Exergy Efficiency	k_B	Boltzmann Constant
η	Efficiency	k_H	Rate Constant for Hydrolysis-Based Degradation
σ	Sigmoid activation function	°C	Degrees Celsius
H_{n-1}	Hidden state from previous time step	R_s	Series Resistance
Y_{H1}	Output of neuron in hidden layer 1	C_n	LSTM cell state at time n
Y_{H2}	Output of neuron in hidden layer 2		

PREFACE

Experimental investigations integrated with the development of true assessment tool play a crucial role in understanding the long-term performance behavior of photovoltaic (PV) systems under realistic conditions. Such investigations form the backbone of performance benchmarking, system reliability analysis, and emulate the true remining useful life for solar installations. The motivation for this work stems from the growing need for reliable, real-time monitoring of PV systems to optimize energy production and extend system lifetime.

This research focuses on Development of Solar PV Performance analyser for quantifying performance ratio with advancing prediction accuracy. The framework integrates field-based degradation modelling, exergy efficiency evaluation, and deep learning-based approaches for estimation of Performance Ratio (PR) across varying operational conditions. While existing models tend to use simplified assumptions or laboratory-based data, they fail to account for site-specific failures, environmental variations, and long-term performance degradation, which often lead to underperformance and financial losses in real-world PV systems. This research fills a critical gap by offering a data-driven, site-specific approach to degradation and performance prediction, moving beyond static models and allowing for better management and operation of PV plants.

The primary objective of this research is to develop an accurate and economical assessment approach with its transformation into a user-accessible tool for predicting true operational performance ratio of PV plants. The first step in development of performance assessment tool is to model degradation rate (R_d)

which is based on failure modes. This focuses on the analysis of a rooftop grid-connected PV plant with a rating of 125 kWp located in erstwhile transit campus of the Institute, at Chikkajala, Bengaluru, Karnataka. The PV modules used in the system are polycrystalline Solar panels manufactured by Vikram Solar, which have been operational for 5–6 years. The panels are oriented to face south with a tilt angle of 13° to optimize solar energy absorption. The installation site receives an average of 2365 hours of annual sunshine, making it suitable for a high-efficiency PV setup. However, as with all PV systems, performance degradation due to failure modes and operational stresses is inevitable.

The Performance Ratio (PR) of the 125 kWp system was observed to fluctuate seasonally, with 45.5% PR during summer and 54% PR during winter. These variations in PR are influenced by several factors, primarily module temperature and irradiance fluctuations, which are inherent to the local climate. Additionally, failure modes such as delamination, microcracks, junction box oxidation, and bubbles were observed through on-field quantitative approaches impacting the operational performance. The on-field approaches include visual inspections, thermal imaging, and I–V curve tracing. The experimentally evaluated annual failure mode-based degradation rate (R_d) for Bengaluru based PV system is 7.7%, reflecting the combined effect of these failure modes. This rate is higher than the typical degradation rates (0.5%–1% annually) reported for new PV systems, underscoring the need for more accurate, field-based degradation models. The high R_d also indicates its significant impact on the performance ratio and longevity of PV systems.

To forecast day-ahead degradation rate based on failure modes, a hyper-parameter tuned and optimised two models based on Long Short-Term Memory (LSTM) approach was formulated. Parameters such as module temperature difference (DEL_T), Irradiance (I), Wind speed (W), Relative humidity (H) and Series resistance (R_s) were incorporated in model 1. While Module 2, involves Temperature (Tm), Ambient Temperature (Ta), Irradiance (I), Wind speed (W), Relative humidity (H) and Series resistance (R_s).

The model was trained employing the on-field based experimental investigation performed during Solar Noon over an annual duration and its evaluated metrics. The developed model is also validated for a non-trained location for PV plant at Telangana. The developed R_d model 1 outperformed model 2, with Root Mean Square Error (RMSE) values of 0.00198 and Mean Absolute Percentage Error (MAPE) of 0.0578 demonstrating its superior ability to capture the non-linear behaviour of degradation over time. This approach enables estimation of the failure mode-based degradation rate degradation rate without relying on costly accelerated testing. The proposed Model 1 incorporates key physical indicators that directly correlate with the presence and progression of specific failure modes. Unlike conventional methods that assume a constant degradation factor, often leading to under- or overestimation of performance metric, this model captures the actual time-series trend of on-field degradation.

The second step in performance tool development, involves modelling thermal exergy loss which attributes its impact on PR. In addition to the Bengaluru plant, a 191.9 kWp realistic grid-connected PV system in Khopoli, Maharashtra, was extensively analyzed to evaluate the impact of thermal exergy loss and system

performance under different environmental conditions. This system is equipped with 370 Wp polycrystalline silicon panels, configured into three sets of strings: two sets of 70.3 kWp each and one set of 29.6 kWp. These strings are connected to a 170 kW (AC) inverter. Following identification of the base system, identification of significant attributes affecting exergy efficiency based thermal exergy loss is carried out. A combined Pareto analysis and correlation heatmap revealed that irradiance, module temperature, wind speed, and factor temperature correction ratio 'A' significantly impact daily exergy efficiency. Notably, strong negative correlations were observed with irradiance (–0.94), module temperature (–1.00), and wind speed (–0.91), highlighting their role in increasing thermal losses. Factor "A" also showed a moderate negative correlation (–0.70), indicating its influence on thermal exergy loss and system efficiency.

A multi-layer hybrid model and five Dual layer hybrid Models is proposed to predict exergy efficiency. The multi-layer hybrid model, which incorporated Convolutional Neural Networks (CNN), Gated Recurrent Units (GRU), attention mechanisms, and autoencoders, achieved RMSE values between 0.5498 and 0.6210, and MAPE ranged from 1.88% to 1.93%. This hybrid approach demonstrated strong accuracy, justifying the effect of exergy on PR, which is often overlooked in traditional performance metrics analyser and proved critical for a true and accurate system performance.

Finally, the short-term, minute-ahead predictive tool for estimating the Performance Ratio (PR) of a solar PV plant is developed using pre-estimated inputs such as degradation rates, thermal exergy loss, and irradiance. The predictive model is built on a Long Short-Term Memory (LSTM) architecture comprising four layers, each

with 256 units, followed by a dense output layer with a single neuron. The model is trained using the Mean Absolute Error (MAE) as the loss function and optimized with the Adam optimizer. A total of 1000 training epochs are conducted with a batch size of 32. The objective of the training process is to minimize the deviation between the predicted and actual PR values of the PV system, ensuring high accuracy in real-time performance forecasting.

The developed PR-LSTM analyser is compared with existing assessment approaches, including PVsyst, PRcorrected model and Feedforward Neural Network (FFNN), and was found to yield the lowest MAPE of 1.83% during training. To comprehensively evaluate the robustness and prediction accuracy of the proposed estimator, validation was conducted for non-trained locations namely Bengaluru, India, and Köprübaşı, Turkey. On comparison with the existing PR model approaches, the proposed PR-LSTM model achieved a Mean Absolute Percentage Error (MAPE) as low as 5.81%, demonstrating strong predictive performance for Bengaluru location.

The second validation study was conducted using data from a 30 kWp rooftop PV installation at Köprübaşı, in Turkey. In this case, the PR-LSTM model outperformed all alternative prediction methods, achieving the lowest Root Mean Square Error (RMSE) of 0.0137, and MAPE of 1.48%.

Further to enhance usability, a Graphical User Interface (GUI)-based Performance Ratio (PR) analyser has been developed for short-term PR prediction. Built using Tkinter, a Python library for GUI development, the interface offers an intuitive platform for users to interact with the underlying PR prediction model. Key features of the tool include: (i) a frame for visualizing predicted PR against actual PR values

over a long-term horizon (ii) a frame for plotting short-term PR trends, and (iii) a short-term visualization of thermal exergy loss over time frame.

The advantage of the developed interface lies in its reliance on readily accessible user inputs such as latitude and longitude of location, rated capacity of PV system, wind speed, ambient temperature and irradiance for early estimation of Performance Ratio (PR). Although users do not explicitly input advanced parameters like thermal exergy loss and R_d, these factors are internally accounted for within the underlying PR-LSTM model, which powers the GUI. By integrating this model into an interactive interface, the proposed GUI enables simplified yet intelligent computation, making it a practical tool for managing and optimizing photovoltaic systems and supporting informed decision-making. This research contributes to the growing body of PV Industry for accurate system performance ratio prediction by offering insights on true performance loss of the system.

Key words: Performance Ratio Analyser, Failure mode-based power degradation rate (R_d), Thermal exergy loss, Exergy efficiency, Multi-layer hybrid deep-learning model, dual-layer hybrid model, MAPE and ground-based validation.

Chapter 1: Introduction

Overview:

As the global push toward sustainable energy intensifies, solar photovoltaic (PV) technology has emerged as a cornerstone of renewable energy systems due to its modularity, scalability, and eco-friendly profile with reduced global warming potentials. However, ensuring the long-term reliability and performance of PV installations, especially in diverse and dynamic environments such as India remains a critical challenge. This chapter outlines the evolution of Solar PV adoption in India, the country's energy landscape, and the limitations of conventional performance models that fail to incorporate true on-field system attributes. It highlights the growing relevance of advanced data-driven approaches like deep learning for predictive analysis and the need for integrated tools to monitor system health. The chapter also defines the research motivation, objectives, and contributions, establishing the stage for a comprehensive investigation into development of performance predictive tool. The developed performance metric tool depends on field-based degradation modelling, exergy efficiency analysis, and intelligent forecasting using LSTM and hybrid models. Finally, the thesis structure is presented to guide the reader through each phase of this interdisciplinary research.

Outline of Chapter ***Page***

1.1 Background of the Study

The growing global demand for sustainable energy has accelerated the deployment of renewable energy technologies, among which solar photovoltaic (PV) systems play a prominent role. Solar PV is increasingly recognized for its modularity, minimal environmental impact, and potential to meet rising electricity demands, especially in regions with abundant solar radiation. However, the long-term viability, economic feasibility and scalability of PV systems depend not only on installation but also on their true operational performance attributes such as degradation behaviour which affect system-level reliability under real-world conditions.

Renewable energy systems especially Solar PV systems are becoming a fit and forget model due to its scalability. The enormous growth in PV sector lies economically viable only when the accurate knowledge on the performance ratio is known. This lies as a unique comparative indicator among large-scale PV plants.

Despite technological advancements and real-time monitoring predictive analytics of PV systems remain underutilized in many regions. Performance indicators such as the Performance Ratio (PR), failure mode degradation rate (R_d), and exergy efficiency are essential for evaluating plant effectiveness over time. However, these parameters are often influenced by complex environmental and operational factors such as irradiance, temperature, wind speed, humidity, and failure modes, which are seldom considered together in performance ratio models.

Major traditional modelling aspects as per IEC standards depend on daily/ monthly/ annually monitored energy generation-based input for computation of performance ratio. Also, conventional analytical techniques fall short in capturing the nonlinear, time-dependent interactions that impact PR. This has led to growing interest in machine learning and deep learning approaches, particularly Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), Convolution neural network (CNN) and Transformer-based models which are intuitively capable of learning complex physical interactions for performance ratio forecasting.

1.2 Energy Situation in India and Solar PV Development

As seen in table 1.1, India is one of the fastest-growing economies in the world, and its energy demand is increasing rapidly due to industrialization, urbanization, and population growth. According to the Ministry of Power, Government of India (2025), India's total installed power generation capacity reached approximately 475 GW by

March 2025, with fossil fuels accounting for around 52% and non-fossil fuel sources comprising the remaining 48%. Renewable energy capacity, including solar, wind, hydro, and biomass, stands at about 228 GW, reflecting significant growth in clean energy deployment. The electricity generation mix, however, remains dominated by coal gains support from renewables (excluding large hydro) with solar power alone making up approximately 10% of generation (Ministry of Power, 2025). Despite this rapid growth in renewables, challenges persist. For instance, in the fiscal year ending March 2024, hydropower output declined by 16.3%, the steepest fall in nearly four decades, mainly due to erratic rainfall patterns, which temporarily reduced the share of renewables in the power mix to 11.7% (Reuters, 2024) as seen in figure 1.1. Furthermore, coal dependency remains high due to ongoing demand and concerns over the intermittent nature of renewables. The Government of India has set ambitious renewable energy targets, aiming to reach 500 GW of non-fossil fuel capacity by 2030, with solar energy constituting a significant portion of this goal. Plans to invest up to $385 billion have been announced to expand renewable capacity and associated infrastructure (Reuters, 2024). The Jawaharlal Nehru National Solar Mission (JNNSM), launched in 2010, laid the groundwork for this expansion, helping India install over 70 GW of solar PV capacity by 2024 and paving the way towards achieving 280 GW of solar capacity by 2030.

Despite this growth, several challenges continue to impact solar PV deployment, including land acquisition difficulties, grid integration complexities, high upfront

investment costs, economic feasibility and performance variability driven by environmental factors such as dust accumulation, high ambient temperatures, and humidity. One of the solutions to these challenges involve accurate modelling of Solar PV performance, degradation, and reliability under Indian environmental conditions is critical to achieving these national renewable energy goals. This research contributes to that vision by developing models that incorporate field data, failure mode-based degradation insights, and data-driven forecasting to improve operational performance and decision-making in Solar PV deployment.

Table 1.1. Installed Power Generation Capacity by Source in India (March 2025)

Energy Source	**Installed Capacity (GW)**	**Share of Total (%)**
Coal	205.0	43%
Lignite	6.0	1.2%
Natural Gas	25.0	5.3%
Diesel	0.5	0.1%
Nuclear	7.0	1.5%
Large Hydro	47.0	10%
Solar	72.0	15.1%
Wind	45.0	9.5%
Biomass + Waste	10.0	2.1%
Small Hydro	5.0	1.0%
Total	475.0	100%

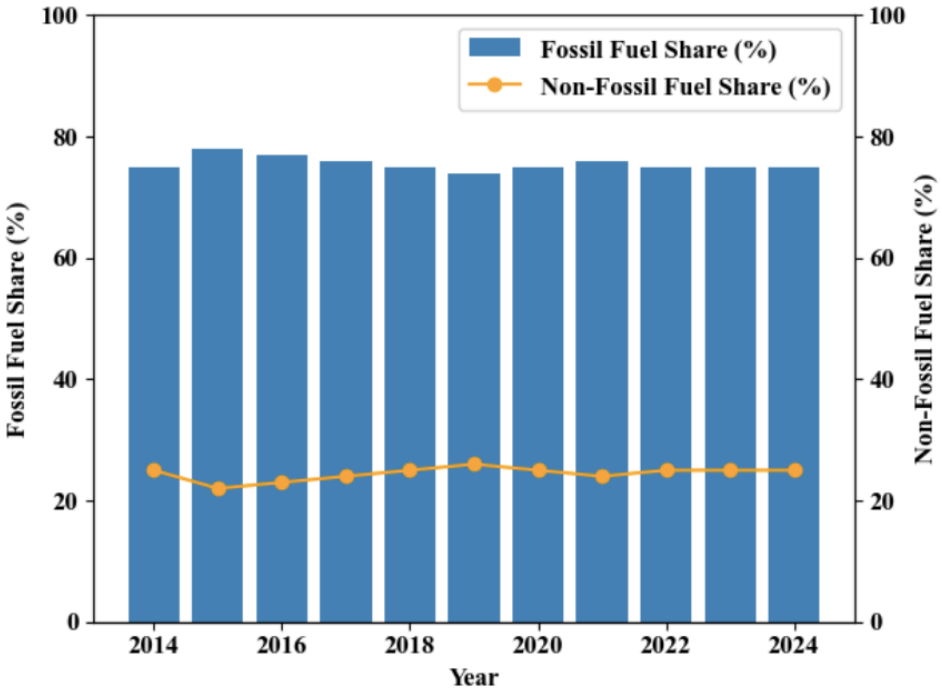


Fig 1.1. India Electricity Generation for a duration from 2014–2024

1.3 Problem Statement

Conventional performance assessments of PV systems typically use assumed degradation rates and simplified efficiency models, failing to reflect the realities of field-induced degradation modes and dynamic environmental interactions. The lack of failure mode-based degradation analysis and real-time integration of predictive models limits the accuracy of performance forecasts. In addition, current forecasting frameworks rarely incorporate thermodynamic (exergy-based) losses or offer deployable user interfaces for real-time monitoring. This results in suboptimal asset management, inaccurate financial projections, and delayed fault detection.

1.4 Key Objective

To develop a comprehensive, real-time performance assessment framework using deep learning and GUI-based visualization tools for grid-connected PV systems. This integrates failure mode-based degradation modelling, exergy based input and environmental factors affecting true on-field performance impacting economic feasibility and remaining useful life.

1.4.1 Specific Objectives

- To assess and quantify the failure mode-based degradation rate (R_d) using on-field experimental investigation such as thermal imaging and I–V curve tracing.
- To model exergy efficiency using a hybrid deep learning approach incorporating environmental and operational parameters.
- To predict /forecasts minute-based performance ratio (PR) for an annual duration, using deep-learning approaches.
- To validate the developed PR based predictor employing ground-based measurements and comparison with benchmark models.
- To develop a GUI-based PR analyser /application that enables real-time monitoring and visualization of PR, R_d and thermal exergy losses.

1.5 Contribution and Significance of the Study

This study contributes a novel framework that blends failure diagnostics, thermodynamic modelling, and deep learning for real-time PV performance evaluation.

It introduces a robust R_d estimation method validated through two distinct geographic sites (Bengaluru, and Telangana), and applies advanced neural network models for accurate PR forecasting. The dual and multi-layer hybrid models for exergy efficiency are among the first to consider field-derived input parameters in depth. The GUI tool enhances the practical applicability of this research, enabling stakeholders to visualize and act upon real-time system insights. This work not only advances the scientific understanding of degradation and performance modeling but also supports the broader goal of sustainable, data-driven solar PV deployment.

1.6 Thesis Structure

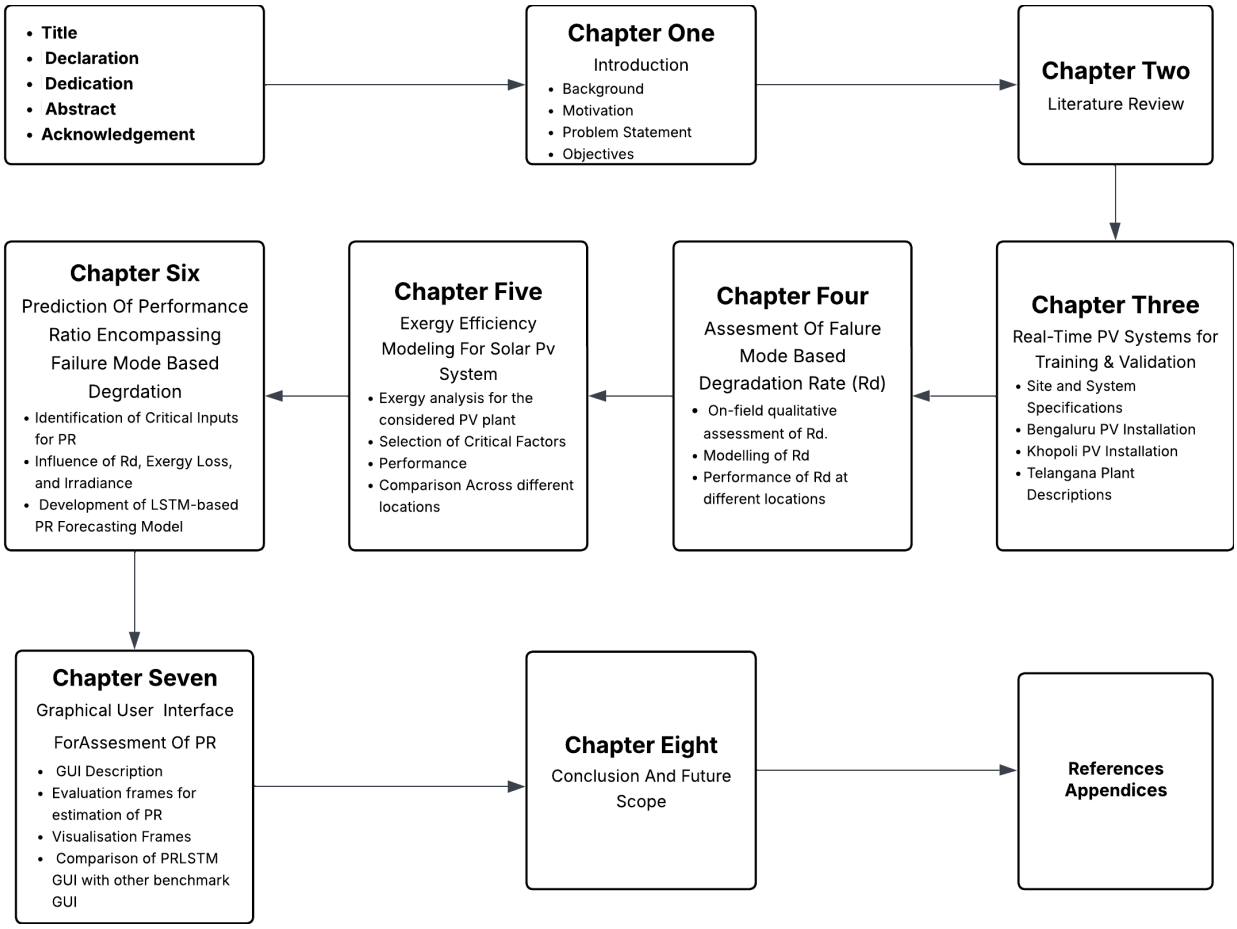


Fig 1.2. Thesis Structure

The thesis as seen in figure 1.2 is structured as follows:

- Chapter 1 provides an overview of the study, including background, motivation, problem statement, objectives, and contributions.
- Chapter 2 presents a critical review of literature on PR estimation, failure mode-based degradation modeling, exergy efficiency-based analytics, and GUI tools.

- Chapter 3 outlines the PV systems selected for training and validation, including detailed site specifications and long-term performance metrics.
- Chapter 4 focuses on the experimental analysis and modeling of failure mode-based degradation rate (R_d).
- Chapter 5 details the exergy efficiency modeling using dual- and multi-layer hybrid models.
- Chapter 6 describes the LSTM-based PR forecasting framework and its validation against alternative models.
- Chapter 7 introduces the GUI platform for real-time assessment of R_d, PR, and exergy losses.
- Chapter 8 concludes the study and discusses potential future research directions, including IoT integration and financial modeling.

The primary motivation for this research stems from the pressing need to enhance the reliability and predictability of PR of Solar PV system under real environmental conditions. Existing models often overlook failure mode-driven degradation or treat it as a static parameter, thereby misrepresenting system behavior over time. Furthermore, while LSTM and related deep learning models show promise, their application in forecasting PR with explicit integration of exergy loss and degradation-based parameters lies unreported. Finally, stakeholders such as operators, policy makers, and investors lack accessible tools for real-time, data-driven decision-making. This thesis aims to fill these gaps by combining theoretical modelling, machine learning, and GUI-based application development.

Chapter 2: Literature Review

Overview:

This chapter reviews key literature aligning the primary objective, focusing on degradation rate (R_d) estimation, exergy efficiency prediction, performance ratio (PR) computation and GUI's for Solar PV system. So, the literature review is also classified into four different sections as mentioned. The review highlights the reported estimation technique, its methodology, application and current limitations. It concludes by identifying critical research gaps which outlines the specific objectives of this study.

Outline of Chapter ***Page***

2.1 Estimation of Failure Mode-Based Degradation Rate in PV Systems

The long-term performance and reliability of photovoltaic (PV) systems are significantly influenced by degradation phenomena, which manifest as a gradual decline in power output over time. This degradation rate (R_d), typically expressed in percentage per year (%/year), serves as a critical metric for analysing the on-field performance, remaining useful life and economic viability. Higher the value of R_d, higher is the levelized cost of energy generated (LCOE) significantly impacting the plant owners and Grid system operator's. (Pagnini *et al.*, 2024). While nominal R_d values for crystalline silicon modules are often reported between 0.5% and 1% annually(Singh, Sharma and Yadav, 2022) field studies have revealed that actual degradation can be substantially higher due to various failure modes and environmental stressors. These failure modes include, but are not limited to, delamination, corrosion, snail trails, hot-spots, and back sheet cracking, which are exacerbated by factors such as high irradiance, humidity, and thermal cycling (Waqar Akram *et al*.,2022 & Abdallah *et al*., 2024).

Traditional methods for estimating R_d often rely on linear assumptions and performance ratio (PR) metrics, which may not adequately capture the nonlinear and site-specific nature of degradation. Therefore, a comprehensive understanding of R_d necessitates reported approaches that integrate accelerated testing, field-based analysis, and modelling techniques.

2.1.1 Accelerated Testing for R_d

Accelerated testing methodologies are designed to simulate the long-term aging of PV modules within a condensed timeframe by subjecting them to intensified environmental stressors as reported in table 2.1.. Standards such as IEC 61215 prescribe a series of tests, including thermal cycling, damp heat exposure, and ultraviolet (UV) irradiation, to evaluate the durability and performance stability of PV modules (IEC 61215, 2021). These tests aim to replicate the mechanical and chemical stresses that modules would encounter over years of field operation (Aghaei *et al.*, 2022).

For instance, thermal cycling tests expose modules to temperature fluctuations between -40°C and 85°C to assess the resilience of solder joints and interconnects. Damp heat tests, involving exposure to 85°C and 85% relative humidity for extended periods, evaluate the potential for moisture ingress and corrosion. UV exposure tests simulate the effects of prolonged sunlight exposure on encapsulant materials and back sheet integrity (Azam *et al.*, 2023).

While accelerated tests are invaluable for initial quality assurance and comparative analysis, their ability to predict realistic degradation accurately is limited. Studies have shown discrepancies between accelerated test results and field performance, primarily due to the complex interplay of environmental factors in actual operating conditions (Jathar *et al.*, 2023) as reported in table 2. 1. Consequently, accelerated testing should be complemented with field-based analyses to obtain a holistic understanding of module degradation.

Table.2.1. Reported Accelerated Test approaches

S.No	Accelerated Approach Reported	System Considered	Location	Limitation(s)	Evaluated Metric(s)	Reference
1	Combined Light-Soaking + Damp Heat (IEC 61215) + Thermal Cycling	1 kWp	Seoul, South Korea	Limited real field variation assumptions	Power degradation %, Latent heat (Q_{mod}), PR, Lifetime prediction	(Nam *et al.*, 2025)
2	Accelerated aging full-spectrum light	2.5 kWp	Beijing, China	Results are based on lab-scale setups; real-world effects like dust, salinity, mechanical load, or inverter degradation not included.	Power degradation %, AF factor, Arrhenius & humidity-light accelerated aging models, real vs simulated loss rates, correction coefficient	(Liu *et al.*, 2025)
3	Combined IEC stress compared with field data	10 kWp	Nigeria	Performance assessed under modelled conditions.	GHI, GTI, DNI, PR, CF and output power (MWh)	(Ebhota and Tabakov, 2023)
4	Thermal cycling + humidity chamber (IEC 61215)	75 kWp	Cairo, Egypt	Lab conditions may not fully mimic real field stresses	PR and energy yield	(Atia *et al.*, 2023)
5	Combined Damp Heat (DH) + Temperature Cycling (TC) per IEC 61215	36 kWp	Mesa, Arizona	Accelerated lab tests need microclimate correlation	Pmax, I_{sc}, V_{oc}, FF, EL imaging, IR thermography, R_{sh}, R_s	(Mahmood *et al.*, 2023)

6	NREL OTF field aging study	7.2 GW	Colorado, USA	Difficulty in isolating permanent soiling from degradation.	Performance Loss Rate (PLR), Degradation Rate, Soiling Rate, Clipping Loss, Performance Ratio (PR)	(Deline *et al.*, 2022)
7	Combined IEC stress compared with field data	6 kWp	Golden, USA	limited to selected back sheet materials	(Pmax, V_{oc}, I_{sc}), corrosion, delamination,	(Owen-Bellini *et al.*, 2021)
8	UV+Temp+Humidity exposure in NIST SPHERE device + sequential mechanical strain (fragmentation test)	1.5 kWp	Clevand, USA	Only PA-based backsheets studied; field exposure limited to 4–6 years	Yellowing index,tensile strength, PCA analysis of degradation vs environment	(Lyu *et al.*, 2020)
9	Accelerated Aging using full-spectrum irradiation	36 kWp	Nigde, Turkey	Limited generalizability due to climatic and topographic specificity	Energy Yield, Performance Ratio (PR), Capacity Utilization Factor (CUF), Mean Daily Energy, System Efficiency	(Kazem *et al.*, 2020)

2.1.2 Field-Based Analysis

Field-based analysis involves qualitative identification of failure modes that occur in PV modules under actual operating conditions to evaluate degradation mechanisms and rates. This approach encompasses a range of diagnostic techniques, including visual inspections, infrared (IR) thermography, electroluminescence (EL) imaging, and current-voltage (I–V) curve tracing (Høiaas *et al.*, 2022)

Visual inspections can identify macroscopic defects such as discoloration, delamination, and snail trails. IR thermography detects thermal anomalies indicative of hot-spots or defective cells, while EL imaging reveals microcracks and inactive regions within the cells (Müller *et al*., 2019). I–V curve analysis provides quantitative data on electrical performance, enabling the detection of increased series resistance or decreased fill factors associated with degradation ((Kumar and Maheshwari, 2022),reported in table 2.2.

Field studies have reported degradation rates exceeding 2% per year in certain climates, significantly higher than laboratory predictions (Aboagye *et al.*, 2022) These elevated rates are often attributed to environmental stressors such as high humidity, temperature fluctuations, and soiling, which are not fully replicated in accelerated tests. Moreover, the heterogeneity of degradation across different modules and installations underscores the importance of site-specific assessments (Abdulla, Sleptchenko and Nayfeh, 2024). Though Field-based analysis exist, these methods do not directly quantify power degradation rate.

Table 2.2. Reported on-field approaches applicable to PV system

S.No.	Type of On-Field Approach Reported	System Considered	Location	Limitation	Evaluated Metric	Reference
1	Field evaluation of a thermoelectric cooling-integrated PV module	3.kWp	Ahmedabad, India	Limited scalability, prototype-level data	Output power, module temperature reduction, energy gain (%)	(Rao, Sahoo and Yanine, 2025)
2	Field-based I–V curve tracing and IR thermography for mismatch and fault detection	3.2 kWp	São Paulo, Brazil	Requires stable irradiance during I–V measurement; IR thermal thresholds can vary	PR, mismatch factor, normalized power loss, temperature distribution, IR + I–V correlation	(Özkalay *et al.*, 2024)
3	Drone-based IR orthogonal imaging and deep learning for defect detection	75 MW	South Africa	Classification accuracy varies with overlap thresholds	Accuracy of defect classification, feature distribution mapping, thermal delta detection	(Masita, Hasan and Shongwe, 2022)
4	Outdoor experimental assessment of anti-soiling coatings using visual inspection and PR monitoring	1 kWp	Biskra, Algeria	Small sample size; coating durability depends on climate and cleaning cycles	PR comparison, visual soiling inspection, transmittance degradation, monthly energy yield	(Rahaman, Urmee and Parlevliet, 2022)
5	Field-based monitoring with IoT-integrated low-cost sensors	3.2 kWp	Bari, Italy	Environmental noise, inverter loss not isolated; Metric: PR, R_d, CUF	System efficiency, irradiance, temperature readings	(Massaro and Starace, 2022)

6	Tracking and Fixed Tilt Comparative Study	1 kWp	Cyprus (Mediterranean climate)	Site-specific degradation rates; degradation mechanisms vary per technology	Degradation rate (%/year), energy yield, temperature coefficient	(Gönül *et al.*, 2022)
7	Performance Testing of Dust and Cleaning Intervals	5 kWp	Qatar	Dust deposition impact varies by panel type; requires regular cleaning and specific climate data	PR, energy yield, dust loss factor, bifacial gain	(Salamah *et al.*, 2022)
8	Field Degradation Assessment via I-V Tracing	22.5 kWp	Tamil Nadu, India	Limited long-term data	PR, energy yield, water cooling impact, temperature drop on module surface	(Bansal, Jaiswal and Singh, 2022)
9	Field and weather-based hybrid degradation model validation	35 kWp	Chennai, India	Region-specific model accuracy	Annual degradation rate, PR, CUF, irradiance-adjusted yield, actual output	(Kandeal *et al.*, 2021)
10	Seasonal Tilt Angle Optimization in Field Conditions	5 kWp PV Array	UAE	Highly dependent on albedo surface stability and ground reflectivity	Seasonal yield, annual PR, CUF	(Mansour *et al.*, 2021)
11	Combined EL, IR thermography, and VI for degradation analysis	3 kWp	Tunisia	Sensor calibration and system cost	Detection of snail trails, delamination, inactive areas, thermal hotspots, PR drop	(Di Tommaso *et al.*, 2022)

2.1.3 Modelling Approaches for R_d

To enhance the accuracy of R_d estimations, researchers have recently reported analytical modelling techniques that incorporate environmental conditions and failure diagnostics. Machine learning (ML) and deep learning (DL) models have demonstrated superior performance in capturing the nonlinear and temporal aspects of PV degradation (Aslam *et al*., 2020; Malhotra *et al.*, 2016).

The models, as summarized in Table 2.3, leverage datasets comprising empirical system metrics such as irradiance, temperature, humidity, output power, and imaging results (e.g., EL, IR).For example, Aslam et al. (2020) employed LSTM networks using time-series inputs like irradiance, temperature, and historical performance ratio (PR) to forecast long-term PV degradation trajectories. Additionally, hybrid AE-LSTM models were developed to process high-dimensional inputs including voltage-current characteristics, degradation flags from image analysis, and sensor-based fault indicators, enabling feature extraction and sequential forecasting. While these studies demonstrate the capability of LSTM in modeling degradation, their input spaces remain relatively constrained, often excluding advanced thermal or exergy loss indicators. In contrast, our approach introduces a richer feature set including degradation rate (R_d), thermal exergy loss, and irradiance as LSTM inputs, allowing more nuanced modeling of real-time PV degradation under field-specific conditions. This shift enhances prediction specificity and addresses limitations of earlier models that relied primarily on environmental and electrical parameters .

The integration of ML and DL models into PV monitoring systems facilitates real-time degradation tracking and predictive maintenance, thereby enhancing system reliability and reducing operational costs. However, the efficacy of these models is contingent upon the quality and comprehensiveness of the input data, necessitating robust data acquisition and preprocessing protocols.(Al-Dahidi *et al.*, 2024)

As inferred from the reported investigations, employing qualitative techniques namely accelerated and on-field testing, for estimation of R_d is time-consuming and non-economical. Data- driven based approach though has the capability to learn the dynamic behaviour of the performance metric yet depend on nature of input data and time horizon. So there lies an evident gap for contribution involving the assessment of failure mode-based degradation for a country like India where Energy Transition lies pivotal.

Table 2.3.Reported approaches for R_d modelling.

S.No.	Type of Modelling Approach Reported for R_d	System Considered	Location	Limitation	Evaluated Metric	Reference
1	Rule-based + Deep hybrid ensemble	3.2 kWp testbed	India	Dependent on initial rule accuracy	Daily R_d, RMSE	(Rathore *et al.*, 2025)
2	Multi-linear Regression (MLR)	Rooftop mono-Si PV	China	Assumes linearity; poor in seasonal transition	Estimated R_d	(Zhang et al., 2025)
3	Physical-Statistical hybrid model (I–V + envt.)	Field PV array (10 kWp)	Brazil	I–V tracing requires stable irradiance	R_d % per year	(Mesquita et al.,2024)
4	Support Vector Regression (SVR)	Rooftop 3 kWp system	India	Requires careful kernel tuning	Prediction error of R_d	(K. Patel et al., 2023)
5	Recurrent Neural Network (RNN)	4.5 kWp residential PV	Tunisia	Sensitive to missing data and overfitting	R_d trend capture	(Ghodhbani et al., 2022)
6	ARIMA with Kalman Filter	Residential PV	Germany	Assumes linear degradation; not robust to outliers	Daily R_d, model confidence	(Ghosh et al. , 2021)
7	Thermography + Regression	Commercial rooftop (50 kWp)	Spain	Limited to surface-visible anomalies	Thermal R_d, fault severity	(Hernandez et al.,2021)
8	Genetic Algorithm-optimized empirical model	Simulated 10-year PV data	Egypt	Not validated on real-world field data	Optimized R_d fit	(Mostafa et al., 2020)

2.2 Estimation of Exergy Efficiency for PV systems

Exergy analysis offers a thermodynamically rigorous approach for evaluating the actual usable energy within a photovoltaic (PV) system, going beyond traditional energy efficiency metrics. Unlike conventional energy efficiency, which merely considers the ratio of output energy to input solar energy, exergy efficiency incorporates the quality and irreversibility of energy transformations. It thereby reflects how effectively a PV system converts the available solar energy into useful work (Aboelwafa *et al*., 2022).

Exergy analysis is crucial for understanding system-level inefficiencies arising from material degradation, temperature mismatch, and entropy generation during solar conversion. The concept is rooted in the second law of thermodynamics and requires accurate environmental data (solar irradiance, ambient temperature) along with system performance data (voltage, current, module temperature) for its estimation. Researchers have established that exergy analysis not only quantifies energy conversion losses but also correlates strongly with degradation phenomena and operational anomalies in PV arrays (Tiwari *et al*., 2022; Aboelwafa *et al*., 2022).

Recent studies emphasize integrating exergy efficiency with real-time monitoring and machine learning tools to facilitate predictive maintenance and optimization. The availability of high-resolution temporal data and computational tools has enabled hybrid models that fuse exergy formulations with statistical learning for performance forecasting (Paliwal *et al*., 2023; Roy *et al*., 2022).

This section reviews the key reported approaches for exergy efficiency estimation and highlights recent advancements on integration of data-driven approaches for exergy efficiency in hybrid PV systems.

2.2.1 Reported Approaches for Exergy Efficiency Estimation

The exergy efficiency (η_ex) of a PV system is commonly and majorly calculated using the Petela model expressed as in equation (2.1).

$$\eta_{ex} = \frac{V \cdot I}{E_{ex}} = \frac{V \cdot I}{A \cdot G \cdot \left[1 - \frac{4}{3} \cdot \left(\frac{T_a}{T_s}\right) + \frac{1}{3} \cdot \left(\frac{T_a}{T_s}\right)^4\right]} \qquad \text{Eq.(2.1)}$$

where:

V and I are the module's voltage and current; A is the PV module area,

G is the incident solar irradiance,

T_a is the ambient temperature,

T_s is the sun's surface temperature (~6000 K).

This formulation considers the thermodynamic potential of solar radiation, thereby incorporating both energy magnitude and quality.

The most adopted methodology for computation of exergy efficiency is derived from the Petela model as in equation (2.1), where the exergy of solar radiation is considered as the maximum work potential of incident solar energy. This approach accounts for the thermodynamic limitations of solar conversion and offers more realistic performance insights compared to simple energy efficiency

In addition, approaches which depend upon factors like entropy generation due to non-uniform irradiance, thermal resistances, and material imperfections are also reported

(Roy *et al*., 2022). These models are particularly important when assessing degraded or partially shaded modules, where traditional energy metrics may overestimate performance. Aboelwafa *et al.* (2022) provided a detailed exergy framework for rooftop PV systems and demonstrated how η_ex varies significantly across different timescales and operational states, reflecting seasonal performance variations and degradation behaviour.

Other researchers (Paliwal *et al.,* 2023) (Rodríguez *et al.*, 2022) have explored empirical-statistical models that combine Petela-based exergy equations with regression, support vector machines, or ensemble learning algorithms to capture variations in performance over time and across geographies. These methods allow predictive and real-time estimation of exergy efficiency using datasets with high temporal granularity, which is essential for rooftop PV systems with fluctuating loading and shading conditions. Tiwari *et al.* (2022) demonstrated that including thermodynamic parameters like exergy destruction and entropy generation improves model sensitivity in anomaly detection tasks. Their findings showed that thermodynamic loss indicators aligned closely with observed failure modes, such as diode bypass issues and interconnect corrosion, highlighting the operational value of exergy-based diagnostics. Furthermore, emerging tools such as hybrid exergy-PR

models and multi-output regression trees are being explored to integrate exergy monitoring into broader asset management and forecasting systems (Roy *et al*., 2022). As inferred, the exergy based key metric evaluation majorly depend on assumptions, a greater number of model constants and are dependent on attributes which cannot be easily accessed or priorly estimable. Development of a dynamic exergy based approach applicable to on-field conditions for solar PV systems still remain a space for contribution leading to attainment of specific research objectives as stated in table 2.4.

Table 2.4.Reported approaches for Exergy estimation in hybrid PV systems

S.No.	Type of Approaches Reported	System Considered	Location	Limitation	Evaluated Metric	Reference
1	Exergy and energy analysis of bifacial PV module on a cloudy day	300 W	Saudi Arabia (Buraydah, Qassim University)	Study limited to one cloudy day; no year-round or different weather comparison; rear irradiance not measured directly	Exergy efficiency, Energy efficiency, Output exergy,	(Almarshoud *et al.*, 2024)
2	Energy and exergy assessment under harsh UAE climate	2.88 kWp	UAE (Sharjah)	Only 4 days studied in different months; limited cooling techniques explored	Exergy efficiency Electrical exergy, Thermal losses	(Salameh *et al.*, 2023)
3	Comparative energy and exergy study of PV, ST, and PV/T systems	63.5 W	Shenzhen, China	Only small-scale systems; limited thermal application evaluation; annual estimates based on short-term tests	Energy gain, Exergy gain, Energy efficiency, Exergy efficiency (daily and annual); semi-empirical and input-output modeling	(Qingyang *et al.*, 2021)
4	Performance analysis using Taguchi optimization on PV vs PVT	20 W	Kumasi, Ghana (KNUST)	Only summer conditions tested; water heating	Exergy efficiency, Energy efficiency, Output power, Heat gain	(Abdul-Ganiyu *et al.*, 2021)
5	Energy and exergy analysis of PVT air collector	37.5 W	India (Delhi)	short-term data	Exergy efficiency,Energy efficiency, Thermal gain	(Arslan, Aktaş and Can, 2020)

6	Energy and exergy analysis of nanofluid-based PVT system integrated with PCM	40 W	Iran (Mashhad, Ferdowsi University)	Small-scale lab system; limited duration (only select days in Aug–Sept); weather-dependent	Exergy efficiency, Energy efficiency Entropy generation	(Hosseinzadeh, Sardarabadi and Passandideh-Fard, 2018)
7	Exergy analysis of utility-scale PV at Cochin Airport	12 MWp	Cochin International Airport, India	No active cooling used; limited to thermal and electrical exergy—chemical and optical losses not included	Energy efficiency (13.33–16.4%), Exergy efficiency (avg 9.77%), Thermal loss (up to 4.44 MW)	(Sukumaran and Sudhakar, 2018)
8	Exergy analysis comparing PV, Thermal, PV+T, and PVT systems under limited roof area	300 W	Detroit, Denver, Phoenix (USA)	Year-round exergy assumed fully consumed; no real-time demand matching considered	Exergy efficiency (hourly & monthly), Total yearly exergy output (MWh)	(Pathak, Sanders and Pearce, 2014)

2.3 Quantifying Performance Ratio (PR)

The Performance Ratio (PR) is a widely accepted metric and a normalized indicator of operational performance that is independent of the location and climatic conditions. As per IEC standard, it is defined as the ratio of the actual energy output of a PV system to the theoretical energy yield calculated under Standard Test Conditions (STC). Unlike energy yield alone, PR allows for comparison across different geographies and technologies by accounting for operational losses due to temperature, inverter inefficiencies, and soiling.

However, the accuracy and relevance of PR estimation are contingent upon the methodology used, data availability, and the system's real-time operational context. The evolution of PR estimation can be broadly classified into three categories: standard deterministic approach and primitive data-driven based techniques. In this work, PR is primarily evaluated on a monthly and annual basis, in line with IEC 61724 and IEA-PVPS recommendations, where PR is typically interpreted as a long-term performance indicator that smooths short-term meteorological variability.

2.3.1 Standard Approaches to PR Estimation

Standard PR estimation techniques are grounded in methodologies outlined by international standards such as IEC 61724 and IEA PVPS guidelines, which define PR as given in Equations (2.2) and (2.3). Currently reported PR models and their formulations are summarized in table 2.5.

$$PR = \frac{E_{AC}}{P_{STC} \times H_{POA}} \qquad \text{Eq.(2.2)}$$

$$PR\ (corrected) = \frac{E_{AC}}{P_{ref} * \frac{Plane\ of\ array\ irradiance}{Irradiance\ at\ STC} * (1+\alpha_t(T_c-25))} \quad \text{Eq.(2.3)}$$

where E_{AC} is the AC energy output, P_{STC} is the nominal installed capacity at standard test conditions, and H_{POA} is the plane-of-array solar irradiance over the measurement period.

Table 2.5. Currently Existing/ Reported PR models

Form of the PR model	**Added inputs**	**References**
PR= $\frac{kWh}{sunhours*area*effieicncy}$	Area and efficiency of the PV module and AC energy output	System Monitoring and Analysis(Abdul-Ganiyu *et al.*, 2021)
PR= $\frac{E_{actual}}{E_{ideal}}$ *100	AC energy output and rated power output	Solar Pro, Taylor and William(Kim *et al.* 2023)
$PR = \frac{100 * \frac{Net\ production}{Total\ incident\ radiation}}{Rated\ PV\ module\ efficiency}$	AC energy output, rated PV module effieicncy and incident irradiance	NREL (Aste *et al.*, 2013)

Studies have shown that PR values derived solely from IEC-based equations often fail to capture temporal and spatial variances induced by environmental factors such as shading, temperature gradients, and soiling (Vector Renewables, 2023; Hukseflux, 2023). Moreover, aging and degradation effects can cause deviations that are not reflected in static PR metrics (Alimi, Meyer and Olayiwola, 2022).

2.3.2 Limitations of Existing PR Estimation Approaches

Despite these advancements, limitations persist across both standard techniques. Standard methods assume homogeneity in module performance and static loss coefficients, which is rarely the case in field environments where degradation is both spatially and temporally non-uniform (Sethi and Kosmopoulos, 2025). Modules affected by delamination, corrosion, snail trails, or backsheet failures exhibit dynamic and non-linear losses that traditional PR estimation fails to capture . ML-based models, on the other hand, are data-hungry and sensitive to sensor calibration errors, data gaps, and site-specific noise. Moreover, many models function as black boxes, limiting their practical application for stakeholders who require transparency for financial, operational, or regulatory decisions (Salazar-Pena et al., 2024; Hamad *et al*., 2025). A summary of literatures with specific approaches employed is tabulated in table 2.5.

Benchmarking PR across geographies or technologies is also challenging due to variability in module design, system configurations, and data logging protocols. Thus, there is a growing call for standardized hybrid models that integrate physical, statistical, and AI-driven layers to improve robustness and scalability of PR assessment.

Table 2.6.Reported Performance loss estimation approaches

S.No.	Estimation approaches for Performance loss	System Considered	Location	Limitation	Evaluated Metric	Reference
1	AI-based forecasting	10 -50 MW	Global	Model dependency on training	PR, RMSE, MAPE, NSE	Garcia *et al*., 2024
2	Degradation-based real-time model	231 kWp	India	High data volume	PR, RMSE, MAE	Patel *et al.,* 2024
3	Soiling correction model	540 kW	Middle East	Cleaning frequency impact	PR, RMSE, Soiling Correction Factor	Ahmed *et al.,* 2023
4	Statistical enhancement of estimate	275 kW	Brazil	Statistical model assumption	PR, RMSE, MBE	Oliveira *et al.*, 2023
5	NREL degradation impact analysis	50 MW	USA	Data aging effects	PR, MAPE, Degradation Rate	Johnson *et al*., 2023

2.4 GUI for Solar PV Systems

Graphical User Interfaces (GUIs) have become an essential component of modern solar PV monitoring and control systems, enabling intuitive visualization, fault detection, and decision support based on real-time system data. As PV installations grow in scale and complexity, operators require accessible tools to interpret environmental, electrical, and thermodynamic parameters at both module and system levels. A GUI acts as a bridge between raw sensor data and actionable insights, enhancing transparency, performance tracking, and system maintainability (Tariq *et al.,* 2020).

In conventional PV systems, GUI-based dashboards are often integrated with SCADA (Supervisory Control and Data Acquisition) platforms or cloud-based analytics software. These systems typically display real-time voltage, current, irradiance, temperature, and power output trends. More advanced GUIs also include failure alerts, thermal maps, PR trend lines, and exergy efficiency plots, allowing asset managers to identify underperforming modules, bypass diode failures, or inverter issues (Sharma *et al.,* 2021).

Recent advances in GUI design leverage open-source platforms such as Python's Dash, LabVIEW, or MATLAB App Designer to create responsive, customizable dashboards. These tools allow seamless integration with I–V curve tracers, pyranometers, and data loggers, enabling real-time data acquisition and processing.

Furthermore, AI-enhanced GUI platforms are being developed to support predictive analytics and dynamic anomaly detection. By incorporating neural network-based performance output into the interface, users can proactively manage PV health, schedule maintenance, or trigger alerts for critical thresholds. This supports better

decision-making, especially for distributed rooftop PV installations where manual inspection is infeasible (Garcia *et al*., 2024; Salazar-Pena *et al.,* 2024).

On inference a need for an adequate tool which transforms a predictive model, increasing user interaction lies necessary. This tool shall serve useful for all the stakeholders of PV Industry.

2.5 Research Gap

1.Inadequate on-field R_d Estimation:

- Existing PV degradation studies primarily rely on accelerated indoor tests conducted under controlled environmental conditions (e.g., elevated temperature, humidity, or UV exposure), which do not reflect real-world outdoor failure modes, environmental variability, and non-linear degradation behaviour.
- Limited research exists on field-based degradation analysis that accounts for actual climatic conditions, operational losses, and site-specific failure mechanisms. Predictive assessment or scalable forecasting of the degradation rate (R_d) under real-world conditions remains largely underexplored.
- Most R_d models assume a linear degradation trend, which does not align with the non-linear and variable nature of degradation observed in real-world field conditions.

2.Underdeveloped Exergy Efficiency Integration:

- Exergy-based evaluation remains largely theoretical and is seldom linked to real-time PV performance monitoring.

- Scarce research on hybrid exergy-AI models incorporating environmental and operational variables for estimation of exergy efficiency

3.Limitations in PR Estimation Models:

- Existing model approaches for PR do not account for real-world failure modes and degradation mechanisms.
- They rely on static assumptions rather than dynamic field data.
- Reported methods lack precision in predicting /forecasting long-term system behavior.
- These models often lack integration with real-time monitoring, limiting their adaptability to changing system conditions.
- Existing approaches lack exergy-based evaluation, focusing only on electrical output without considering thermodynamic losses.
- Often rely on averaged irradiance values rather than incorporating high-resolution time-series data, limiting accuracy under varying irradiance conditions.

4. Lack of Unified, Intelligent GUI Platforms:

- Most PV monitoring GUIs are hardware-locked (e.g., SCADA) or provide only basic visualization.
- Absence of modular, real-time GUIs combining R_d, exergy analytics, and fault alerts into a unified operator-friendly dashboard.

- AI-based performance prediction, and GUI visualization tailored for rooftop PV systems in varied Indian climates remains unreported to the best of knowledge of the researcher.

5. Missing End-to-End Assessment Framework:

- Existing system holistically integrates degradation analysis, thermodynamic modeling, AI-based performance prediction, and GUI visualization tailored for rooftop PV systems in varied Indian climates remains unreported to the best of knowledge of the researcher.

2.6 Chapter Summary

This chapter presented a detailed review of key concepts and methodologies relevant to the performance analysis of photovoltaic (PV) systems, with a particular focus on degradation modelling, exergy efficiency, and performance ratio (PR) estimation.

- In Section 2.1 examined the techniques used for estimating the degradation rate (R_d), including accelerated stress testing, field-based analysis, and modeling approaches. While nominal R_d values are often cited between 0.5% and 1% annually, field studies have shown that environmental stressors and failure modes can lead to significantly higher degradation. The review identified a gap in failure mode-specific R_d estimation models validated with field diagnostics.
- Section 2.2 explored exergy efficiency analysis, a thermodynamically rigorous method that accounts for both the quantity and quality of energy in PV systems. Reported models varied in complexity, with most lacking integration into real-

time monitoring systems. Hybrid models combining exergy principles with AI approaches have shown potential but remain underutilized.

- Section 2.3 focused on quantifying the performance ratio (PR), highlighting the limitations of conventional IEC-based methods and the emergence of machine learning-based forecasting models. While primitive ANN have shown improved accuracy, issues of generalizability, interpretability, and integration with thermodynamic insights persist.
- Section 2.4 reviewed the evolution of Graphical User Interfaces (GUIs) for PV monitoring. Existing systems are often either too simplistic or hardware-bound, lacking dynamic visualization, predictive analytics, or real-time integration of R_d, PR, and exergy data. Emerging GUI platforms are beginning to incorporate machine learning and diagnostics but remain fragmented.
- Section 2.5 outlined the key research gaps across the domains of degradation modeling, PR forecasting, exergy assessment, and user-interface integration. The chapter emphasized the need for a unified, real-time framework capable of combining diagnostics, AI-driven forecasting, thermodynamic modeling, and user-centric visualization tools tailored for rooftop PV systems in diverse environmental conditions.
- Section 2.6 defined the core objectives of the study, which include estimating failure-mode-based degradation, modeling exergy efficiency, predicting PR using deep learning, and integrating these functions into a GUI-based monitoring platform. These objectives aim to bridge the identified gaps and enhance real-time performance evaluation of rooftop PV systems.

Chapter 3: Real-Time Grid Connected

PV Systems for Training & Validation

Overview:

This chapter presents a comprehensive description of three selected rooftop solar photovoltaic (PV) installations situated in distinct climatic zones across India. Each plant varies in capacity, commissioning age, and operational context, allowing for a diverse representation of real-world performance conditions. The chapter begins by explaining the rationale for site selection, emphasizing geographical spread, solar resource variability, and data accessibility. It then describes the technical specifications of each system, including module types, inverter configurations, array structures, and grid connectivity setups. Seasonal and annual performance metrics such as array yield, final yield, reference yield and capacity utilization factor (CUF) are used to characterize each site's operational behaviour. These insights form a critical foundation for the performance ratio modelling, degradation rate prediction, and exergy analysis undertaken in

Outline of Chapter ***Page***

3.1 Site and System Specifications

The selection of test sites plays a critical role in evaluating the robustness and generalizability of solar photovoltaic (PV) performance ratio models. For this study, three PV installations located in Bengaluru (Karnataka), Khopoli (Maharashtra), and Maheshwaram (Telangana) were chosen based on their climatic diversity, system scale, and monitoring capabilities. These sites represent varied geographies, environmental stressors, and operational contexts, which are crucial for validating the proposed degradation rate modelling, exergy efficiency assessment, and deep-learning based performance ratio (PR) prediction. The following description include justification for site-selection.

3.1.1. Climatic Variability:

India's geographic diversity results in a wide range of climate zones—from humid subtropical (Bengaluru), to semi-arid (Telangana), to coastal-influenced temperate (Khopoli) as per Koppen classification. According to the Indian Meteorological Department (IMD, 2023), such variability impacts solar resource availability, temperature profiles, and dust accumulation patterns factors that significantly influence degradation rates and system performance

Specifically, the three locations—Bengaluru (Karnataka), Khopoli (Maharashtra), and Maheshwaram (Telangana)—represent distinct and contrasting climatic regimes as per Köppen and Indian Meteorological Department classifications:

- Bengaluru, Karnataka represents a *moderate, humid subtropical climate* with relatively stable temperatures and long-term operational aging effects.
- Khopoli, Maharashtra represents a *high-humidity, high-rainfall coastal-influenced climate*, enabling assessment of moisture-related degradation and monsoonal impacts.

- Maheshwaram (Telangana) represents a *hot semi-arid climate* characterized by high irradiance, dust exposure, and thermal stress.

These three climates collectively encompass major environmental stressors affecting PV performance in India, including humidity, temperature variation, soiling, and irradiance diversity.

3.1.2 Degradation Pattern Validation:

Different sites are expected to exhibit different dominant degradation modes. For example, Bengaluru's moderate climate is ideal for studying long-term thermal degradation and humidity induced degradation patterns, without extreme environmental stress. Khopoli, located in a humid, high-rainfall zone, presents an opportunity to evaluate moisture-induced failure modes such as delamination and corrosion (Yan.*et.al.,*2025) Telangana, with its dry and dusty environment, helps assess the impact of soiling, thermal cycling, and high irradiance on performance loss and exergy reduction.

3.1.3 Data Availability and Monitoring Infrastructure:

Each site was selected based on the availability of high-resolution monitoring parameters which include irradiance, module temperature, wind speed and DC / AC side energy generation. All three systems are equipped with pyranometers, temperature sensors, and energy loggers, enabling robust data-driven model development and real-time performance tracking.

Each PV installation is equipped with calibrated pyranometers for irradiance measurement, temperature sensors for module temperature monitoring, and inverter-integrated utility-grade energy meters for DC and AC energy measurement, as described in Table 3.1 The data acquisition and performance assessment procedures adhere to **IEC 61724 recommendations**, which specify

appropriate sensor accuracy and calibration requirements for grid-connected PV system performance evaluation.

Module temperature was measured using contact-type temperature sensors affixed to the rear surface of representative PV modules at the center of the backsheet to approximate cell operating temperature. The sensors were installed with thermally conductive adhesive and insulated from ambient air effects in accordance with the recommendations of IEC 61724-1 to ensure accurate and stable temperature measurements.

For the Telangana and Mumbai PV plants, operational data were acquired through commercial Energy Monitoring Systems (EMS), namely **Wattmon and DelREMO** , employed at the respective sites. These platforms integrate calibrated irradiance sensors, temperature sensors, and inverter-level electrical measurements. Few of them include a PYRA-300 solar radiation sensor (pyranometer) for irradiance measurement, ambient temperature sensors integrated through weather-station modules such as WS200 or PT100-based sensors, a WS102 cup anemometer for wind speed measurement, and relative humidity sensors with a measurement range of 0–100 % RH at the Telangana location. However, detailed make and model specifications of relatively other sensors for the Khopoli location were not disclosed to end users. The data provided by these EMS platforms are industry-standard, time-synchronized, and suitable for performance and degradation analysis. Hence, the work focuses on data consistency and validation rather than hardware-level sensor comparison.

Table 3.1 Summarised parameters of employed sensors for Bengaluru Location

Sl. No.	Instrument Name	Model / Type	Measured Parameters	Measurement Range / Accuracy	Calibration / Standard	Sampling Interval	Used in Analysis
1	Solar I–V Curve Tracer	PV200 Curve Tracer	I–V curve, Voc, Isc, Vmp, Imp, Pmax, Irradiance	As per IEC 60904	Factory calibrated; IEC 60904 compliant	Monthly / periodic testing	R_d estimation and PR validation
2	Irradiance Analyzer	PV200 Irradiance Analyzer	Global solar irradiance	±5%	Factory calibrated; traceable to ISO Standards	1–5 min	R_d estimation
3	Module Temperature Sensor	Pt-100 RTD	Module surface temperature	±0.1–0.3 °C	Calibrated RTD (IEC 60751)	1–5 min	R_d estimation and PR model validation
4	Thermal Imaging Camera	FLIR E5-XT	Temperature distribution, hotspots	Sensitivity ≤ 0.05 °C	Factory calibrated	Periodic inspection	Failure diagnosis for R_d

5	Ambient Temperature Sensor	Weather station	Ambient temperature	±0.5 °C	Factory calibrated	1–5 min	R_d estimation and PR model validation
6	Relative Humidity Sensor	Weather station	Relative humidiy	0–100% RH	Factory calibrated	1–5 min	R_d estimation and PR model validation
7	Wind Speed Sensor	Cup anemometer	Wind speed	0–30 m/s	Factory calibrated	1–5 min	R_d estimation, Thermal loss estimation influencing PR during validation
8	Visual Inspection based on-field method	IEC 61215	Physical defects	Qualitative	IEC 61215	Periodic	Failure mode analysis

3.1.4 Representation of Typical Indian Installations:

The chosen locations collectively mirror the conditions under which major PV deployments in India operate. As per MNRE's 2023 solar mission progress report, these regions are representative of over 40% of India's deployed rooftop and small-scale ground-mounted solar capacity. Therefore, findings from these sites are expected to hold relevance for broader applications.

This work utilizes time-series operational data relating to three grid-connected rooftop PV plants located in Bengaluru, Maharashtra (Khopoli) and Maheshwaram (Telangana). The details are summarized below.

(a) Data duration and period (per location):

Bengaluru (125 kWp): Long-term operational data from a plant commissioned approximately six years prior to the study. Performance indicators are evaluated using monthly mean of daily observations over a complete annual cycle ranging from February 2022- January 2023

Maharashtra, Khopoli (191.9 kWp): Continuous time-series data spanning January 2021 to December 2022, covering two complete calendar years and capturing seasonal and monsoonal variations.

Maheshwaram, Telangana (235 kWp): Operational data from a newly commissioned plant (commissioned in March 2022), covering a complete year of post-commissioning performance.

(b)Sampling interval and time of measurement:

For all three sites, electrical and meteorological parameters, including DC power, AC energy, plane-of-array irradiance, and module temperature, are recorded at a fixed 1-5 minute sampling interval, with daily, monthly, and annual energy values

obtained through aggregation of these minute measurements in accordance with IEC 61724 guidelines.

To guarantee accurate calculation of performance indicators like final yield, reference yield, performance ratio, and degradation trends, IEC 61724 suggests routine and time-synchronized data collection of critical electrical and environmental variables. The chosen hourly sampling interval balances data resolution, storage needs, and measurement reliability while meeting the standard's requirements for long-term performance evaluation. The standard recommended methodology for assessing operational performance and energy yield consistency in grid-connected PV systems is followed when consolidating small-scale data to daily and monthly measures.

(c)Data-synchronization:

All measured parameters are time-synchronized, as these data are acquired through integrated plant monitoring systems and inverter-based data loggers. This ensures temporal consistency among irradiance, temperature, and electrical output data used for yield calculation, performance ratio estimation, degradation analysis, and model training.

(d)Instrumentation and accuracy class:

Each PV installation is equipped with calibrated pyranometers for irradiance measurement, temperature sensors for module temperature monitoring, and inverter-integrated utility-grade energy meters for DC and AC energy measurement, as described in Section 3.1.3 of page 40. The data acquisition and performance assessment procedures adhere to IEC 61724 recommendations, which specify

appropriate sensor accuracy and calibration requirements for grid-connected PV system performance evaluation.

3.2 Bengaluru PV Installation Overview – with specifications, single-line diagram

One of the photovoltaic (PV) plant selected for the proposed investigation is a 125 kWp rooftop grid-connected solar installation located at the erstwhile transit campus of the RGIPT Bengaluru Campus, situated at Chikkajala, Bengaluru as illustrated in figure 3.1. The geographic coordinates of the installation are 13.1596° N latitude and 77.6362° E longitude. The site receives an average of 2365 sunshine hours annually, making it a favourable location for long-term PV performance modelling.

The system is composed of polycrystalline silicon PV modules manufactured by Vikram Solar, each with a peak power rating of 325 W_p whose specification is detailed in table 3.1. The modules have been in operation for 5–6 years and are installed with a fixed tilt angle of 13°, facing south, to optimize solar radiation capture based on the site's latitude.

Fig 3.1. Actual Field Image of the PV facility of 125kWp at Bengaluru, Karnataka

The string considered for study comprises 20 modules, out of which 13 were selected for detailed degradation analysis based on visual examination and thermal criteria. A detailed single-line diagram of the 125 kWp PV plant in Bengaluru is shown in figure 3.2.

Table 3.2 Electrical parameters of PV Module of 125 kWp at Bengaluru, India

Parameter	**Rating**
Maximum Power Pmax(W)	325 W
Power Tolerance W	0 ~ +5 W
Tolerance (%)	±3%
Open Circuit Voltage V_{oc}(V)	46.40
Short Circuit Current I_{sc}(A)	9.25
Maximum Power Voltage V_{mpp}(V)	38.80
Maximum Power Current I_{mpp}(A)	8.37
Cell Type (mm)	156 × 156 (Poly-Crystalline Silicon)
Temp. Coeff of V_{oc} (%/K)	-0.33%/K
Temp. Coeff of I_{sc} (%/K)	+0.05%/K
Temp. Coeff. Of P_{max} (%/k)	-0.43%/K

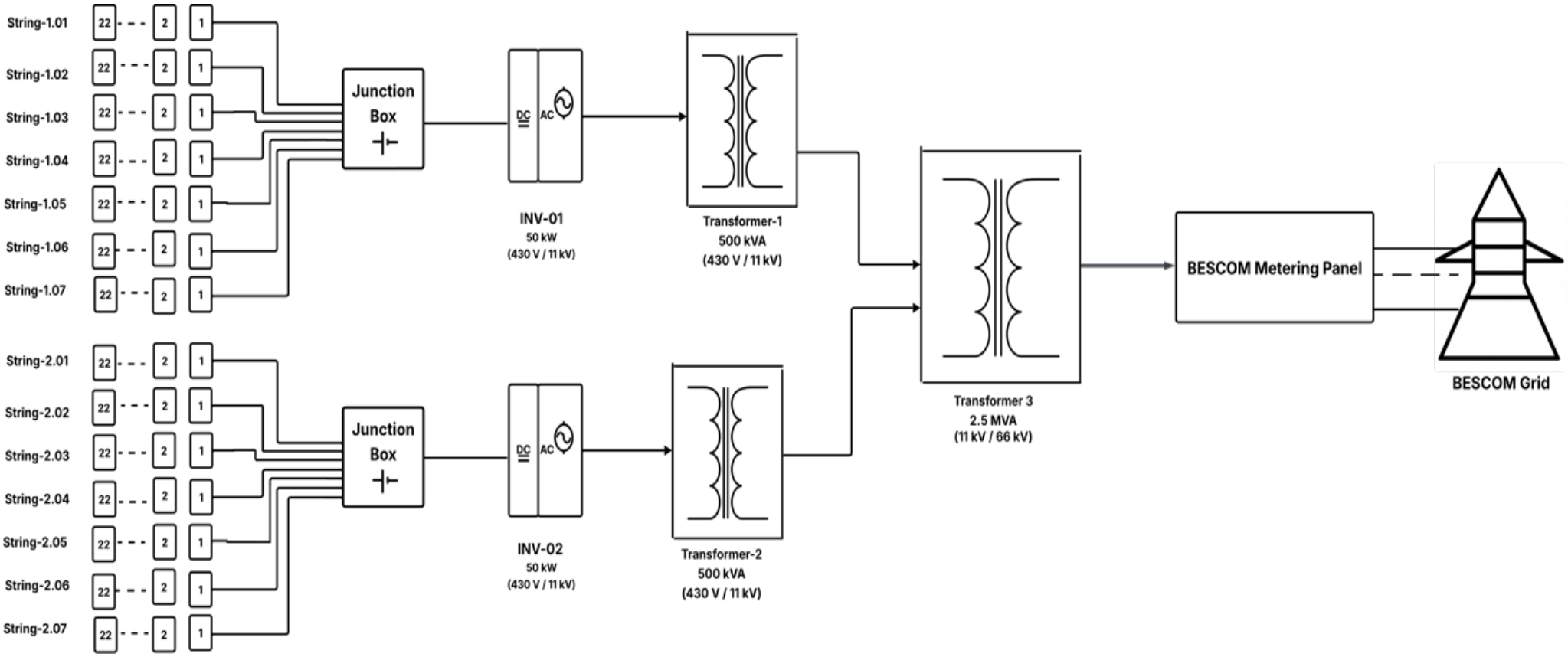


Fig 3.2. Single line diagram considered for investigation at of 125 kWp at Bengaluru, Karnataka

3.2.1.Long-term performance analysis:

To model the long-term operational behaviour of a photovoltaic (PV) system in a specific region, conducting a thorough operational performance investigation is essential. This analysis provides crucial insights into the energy output and efficiency of the system under on-field conditions. The evaluated performance indicators such as reference yield, final yield, array yield, and capacity utilization factor (CUF) serve not only as diagnostic metrics but also as foundational inputs for developing predictive models. These indicators are computed on a monthly mean of daily observation, offering a comprehensive view of the PV system's operational characteristics throughout the year. This plays a vital role in informing stakeholders, including policymakers, developers, and researchers by enabling them to evaluate and compare the actual performance of PV systems across different geographical locations. It also supports evidence-based decision-making for future deployment and optimization of solar power infrastructure. The evaluation methodology adheres to the procedures outlined in the IEC 61724 standard, which emphasizes the importance of real-time monitoring of AC and DC energy outputs. In this study, performance assessments were carried out for PV plants installed at three distinct locations in India: Bengaluru (Karnataka), Khopoli (Maharashtra), and Mallapur (Telangana), capturing regional variations in solar potential and system efficiency.

3.2.2. AC Energy Output (E_{ac}):

The A.C energy generated by the PV system is measured across the inverter end at instants of recording time interval. The data is recorded for every 60 minutes. The

daily and monthly net energy generated by the PV systems are obtained as equation (3.1) and (3.2)

$$P_{AC.d} = \sum_{t=1}^{Rtp} V_{ac} * I_{ac} * R_T \quad \text{Eq. (3.1)}$$

$$P_{ac.m} = \sum_{d=1}^{N} P_{AC,d} \quad \text{Eq. (3.2)}$$

R_T is the recording time interval and R_{tp} is the reporting period and N is the number of operating days of plant in a month.

3.2.3 Array Yield (Y_a) of 125 kWp PV Plant

As per IEC 61724 array yield represents the number of hours per day that the PV array module would require to operate at its rated output power to produce the same daily array energy to the system. It is given by equation (3.3)

$$Y_A = \frac{E_{A.DC}}{E_0} \text{ (kWh/kWp/month)} \quad \text{Eq. (3.3)}$$

where $E_{A.DC}$ is the total AC energy output and E_o is installed array rated output power. It is highest in March (87.3 kWh/kWp/month) and February (80.8 kWh/kWp/month). A sharp decline is seen during May through September, where values drop to 47.9–51.0 kWh/kWp month reflecting seasonal irradiance reduction and potential issues such as mismatch losses as seen in Figure 3.3.

3.2.4. Reference Yield (Yr) of 125 kWp PV Plant

Reference yield represents the maximum potential available at a particular location. It is provided by the ratio of in-plane irradiance H_T to the standard test conditions of irradiance as seen in equation (3.4).It increases from February(182 kWh/kWp) and reaches its peak in March(210.8 kWh/kWp), aligning with peak solar insolation and longer daylight hours as illustrated in Figure 3.3

$$Y_R = \frac{H_T}{G_0} \text{ (h/d) or kWh/kW}_p \quad \text{Eq. (3.4)}$$

3.2.5. Final Yield (Y_f) of 125 kWp PV Plant

The net energy output (E_{AC}) observed daily, monthly or annually of the PV plant per kW of installed peak power output PV array (E_0) is denoted as final yield. In the previous reported literature ,usually this parameter is scaled to Standard Test conditions (STC) i.e. 1000 W/m^2 solar irradiance, 25^o C module temperature and Air mass of 1.5 AM as seen in equation (3.5). This parameter is useful in comparing PV systems at different geographical location.It increases from February (107.7 kWh/kWp) and reaches its peak in March (116.4 kWh/kWp) when both irradiance and system output are favourable as represented in figure 3.3.

$$Y_F = \frac{E_{AC}}{E_O} \text{ (h/d) or kWh/kW}_p\text{/month} \qquad \text{Eq. (3.5)}$$

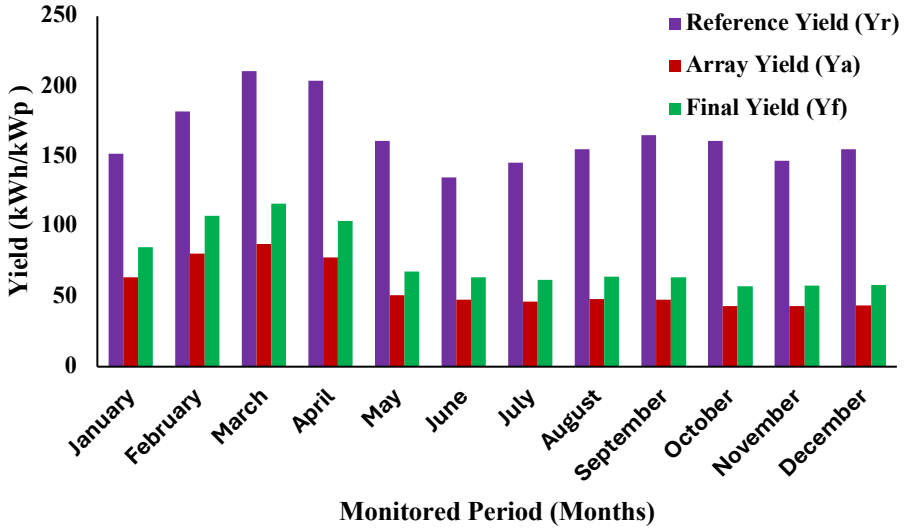


Fig. 3.3 Yields of 125 kWp Rooftop Plant at Bengaluru

3.2.6. Capacity Utilization Factor (CUF) of 125 kWp PV Plant

CUF is location dependent as it depends on the solar radiation received by the modules, total number of sunny days available at the site of installation of the PV

system and the type of modules used as suggested by. (Shravanth Vasisht *et.al.,*2016) It signifies the usage factor of the PV plant to generate AC power output, which is 14% annually of AC output whose long-term annual variation is illustrated in figure 3.4. It is expressed in % and is given by equation (3.6).

$$\text{CUF}(\%) = \frac{Y_F}{24*365} = \frac{E_{AC}}{P_0*24*365} * 100 \qquad \text{Eq.(3.6)}$$

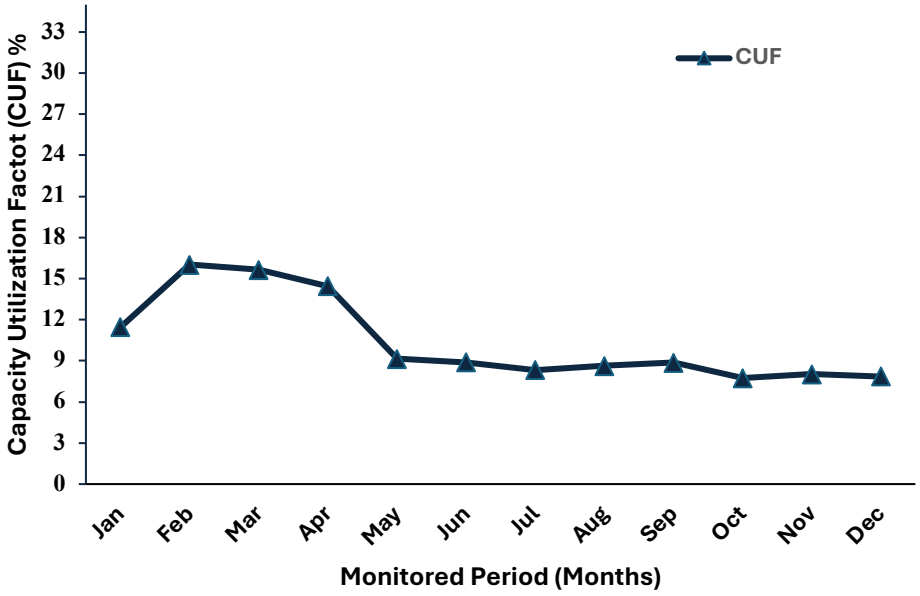


Fig 3 4 .CUF of 125kWp Rooftop Plant at Bengaluru

3.3 Khopoli PV Installation Overview

The Khopoli photovoltaic installation under investigation is a grid-connected solar PV plant with an installed capacity of 191.9 kWp, located at Arkose Industrial Estate, Adoshi Road Dekhu, Sajgaon, Taluka Khalapur, Khopoli, Maharashtra, India. The geographical coordinates of the site are approximately 18° 47' 8.23" N latitude and 73° 20' 45.20" E longitude, placing it in a region with moderate to high solar irradiance. The plant comprises 460 polycrystalline PV modules, each rated

at 370 Wp whose specification is detailed in table 3.2 and connected to the DC side of the system. These modules are arranged into three string of groups of two with 70,300 Wp each and one with 29,600 Wp, resulting in a total DC capacity of 170,200 Wp. The DC power is interfaced to the AC grid via a 170 kW inverter system, which includes three string inverters two rated at 70 kVA and one at 30 kVA configured in a series-parallel layout as seen in figure 3.5 and 3.6.

Fig 3.5. Actual field Image of 191.9 kWp at Khopoli,Maharashtra

The inverter output is routed through an AC distribution board for grid integration, enabling real-time energy export and system monitoring. The dataset considered for performance evaluation spans for a continuous annual period from January 2021 to December 2022, capturing seasonal variability and operational dynamics across two full calendar years.

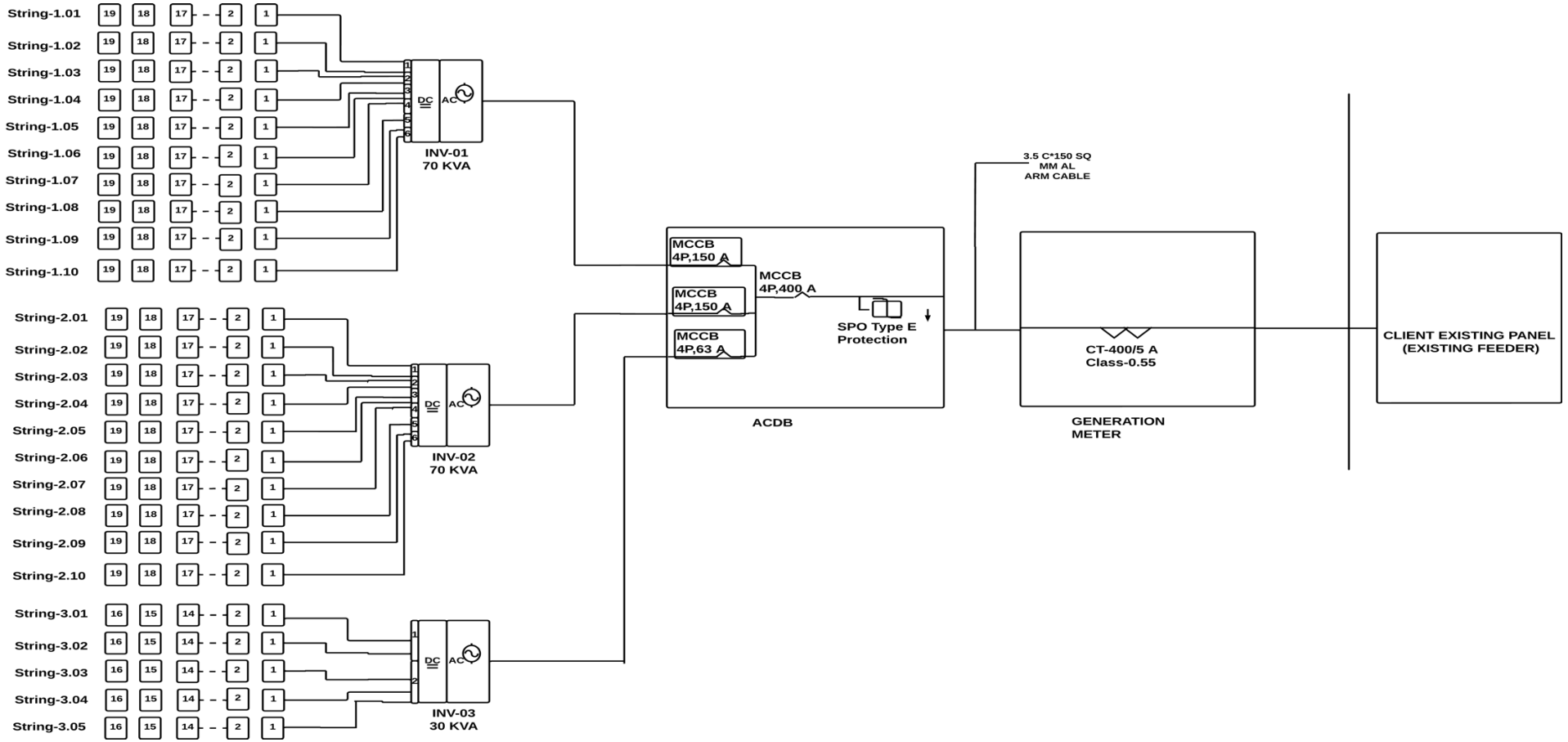


Fig 3.6. Single line diagram considered for investigation of 191.9 kWp at Khopoli,Maharashtra

Table 3.3 Electrical parameters of PV Module at Khopoli site

S.No	Parameters	Details
1	Module Peak Power (P_{max})	370 W_p
8	Maximum Voltage (V_{mpp})	40.00 V
5	Maximum Current (I_{mpp})	8.26 A
4	Open Circuit Voltage (V_{oc})	48.5 V
8	Short Circuit Current (l_{sc})	9.84 A
0	Inverter Make	Delta
7	PV Module Make	Vikram

3.3.1 Performance analysis of Khopoli PV plant

The performance metrics of the Khopoli rooftop PV system was computed through equation (3.1) to (3.6). The Reference Yield is highest during April (306.3 kWh/kWp/month) and remains elevated in March (273.4 kWh/kWp/month) and May (270.0 (kWh/kWp/month), aligning with longer daylight hours and higher solar irradiance of 845 W/m^2. In contrast, it dips during July to September, with the lowest observed in September (149 kWh/kWp/month) due to monsoonal cloud cover minimising the Solar Irradiance to 350 W/m^2. The Final Yield follows a similar pattern reaching its maximum in April (153.15 kWh/kWp/month) and maintaining high levels in May (140.4 kWh/kWp/month) and October (128.2

kWh/kWp/month) as inferred from Figure 3.7. The Array Yield which tracks DC-side generation, is also highest in April (114.86 kWh/kWp/month) and May (105.30 kWh/kWp/month), with reduced values during August to December as seen in figure 3.7. The Capacity Utilization Factor (CUF) peaks in April (21.27%), indicating optimal performance under ideal irradiance and temperature conditions, and remains strong in May (18.87%) and October (17.23%). Conversely, CUF is at its lowest in January (9.23%) and February (10.64%), correlating with both lower solar availability and cooler conditions as seen in figure 3.8. On the whole, the Khopoli plant demonstrates acceptable operation in spring and post-monsoon periods, while showing performance dips in winter and monsoon-transition months, aligning with expected seasonal dynamics.

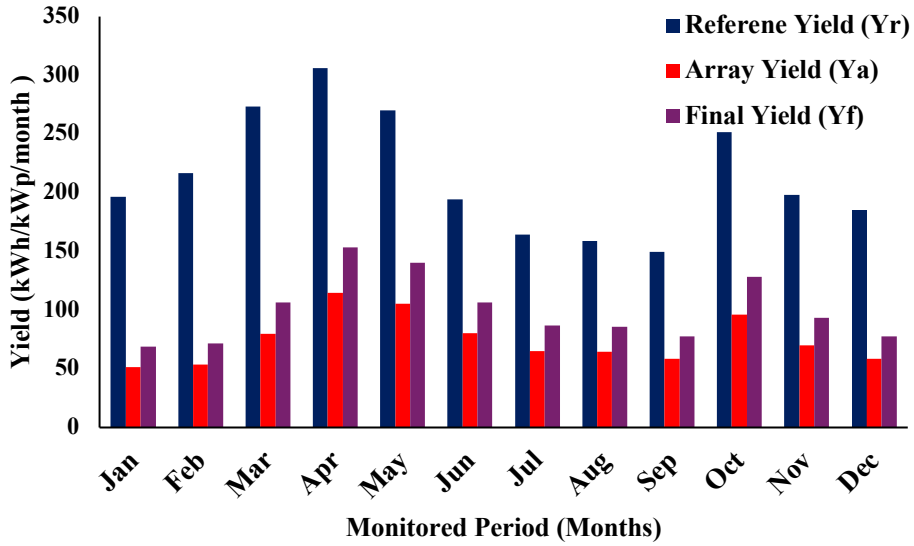


Fig. 3.7. Yields of 191.9 kWp at Khopoli, Maharashtra

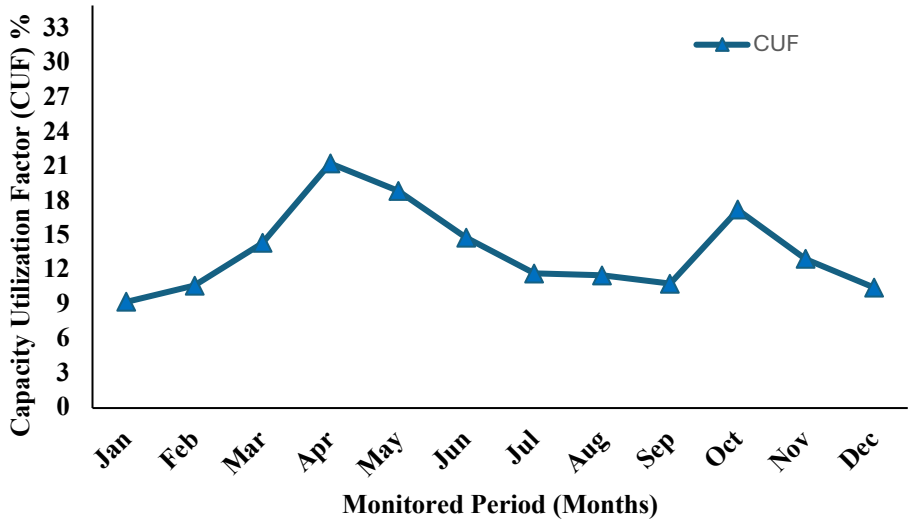


Fig 3.8. CUF of 191.9 kWp at Khopoli, Maharashtra

3.4 Telangana Plant Descriptions Overview – with specifications, single-line diagram and performance analysis of the system

The third site chosen in this research investigation is a 235 kW grid-connected rooftop solar PV system located in Maheshwaram, Telangana (latitude: 17° 13′ 16.40″ N, longitude: 78° 41′ 69.92″ E) and commissioned in March 2022. The system comprises a total DC capacity of 235 kWp whose specification is tabulated in table 3.3. The DC system is structured through multiple PV module strings interfaced via string combiner boxes. the plant architecture includes a combination of 433 panels string inverters that are integrated into the AC distribution board, as seen in figure 3.9 and 3.10.

Fig.3.9. Actual field Image of 235 kWp at Maheshwaram, Telangana

Table 3.4 Electrical parameters of PV Module at Bengaluru.

Parameters	**Rating**
Project Capacity	235 KW AC (294.845 kWp DC)
Module Make and Rating	Axitec Solar - 545 Wp Mono crystalline Model Module Size : 2278 mm × 1133 mm × 40 mm
Inverter Make and Rating	Sungrow - 125KW × 01No, 55 KW -02 No
Maximum rated Power (Pmax)	545 Wp
Maximum Power Voltage-V_{mpp} (V)	41.91
Maximum Power Current-I_{mpp} (A)	13.00
Open Circuit Voltage-V_{OC} (V)	59.76
Short Circuit Current-I_{SC} (A) :	13.85

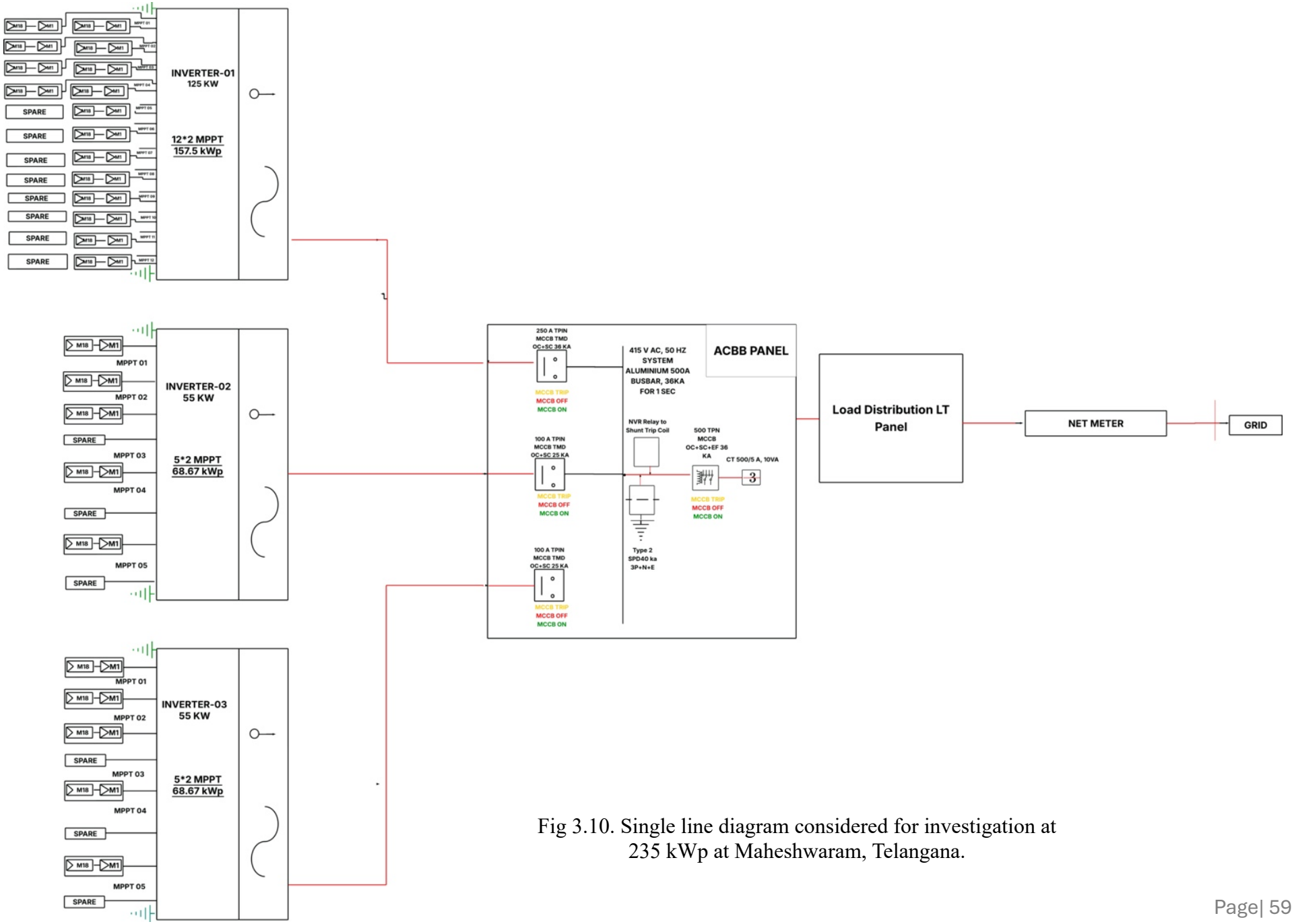


Fig 3.10. Single line diagram considered for investigation at 235 kWp at Maheshwaram, Telangana.

3.4.1 Performance analysis of 235 kWp

The Telangana PV system demonstrates consistently strong performance across all months, as shown by its key energy yield indicators expressed in kWh/kWp. The Reference Yield, which represents the total available solar irradiation, varies from 225.68 kWh/kWp/month in January to a peak of 324.88 kWh/kWp/month in March, driven by high insolation and longer daylight durations during the spring season. The Final Yield, which quantifies the actual AC-side energy delivered per installed kWp/month, follows a similar pattern, with the highest values observed in December (245.72 kWh/kWp/month) and March (224.17 kWh/kWp/month), indicating highly efficient conversion even during winter as seen in figure 3.11. The Array Yield, reflecting the DC-side output of the modules, consistently remains high exceeding 130 kWh/kWp/month for most of the year and peaking at 184.29 kWh/kWp/month in December, suggesting strong module-level performance with low conversion losses as illustrated in figure 3.11. The Capacity Utilization Factor (CUF) also follows a positive trajectory, starting at 20.32% in January and reaching 33.03% in December, showing that the plant is effectively utilizing its installed capacity, particularly toward the year-end. The Performance Ratio (PR) a critical measure of overall system efficiency after accounting for irradiance remains robust across all months, ranging from 67% in January to an impressive 83% in December, far exceeding the industry threshold of 75% for new systems as seen in figure 3.12. Even during the monsoon months of June to August, PR values stay between 58% and 62%, likely due to moderated temperatures and stable inverter performance. Compared to the plants at Bengaluru and Khopoli which is 6 year old, the Telangana system which is newly commissioned exhibits

higher CUF, PR, and yield values throughout the year, reflecting optimized system design, better maintenance, and favorable operating conditions, and thus stands out as the best-performing plant among the three.

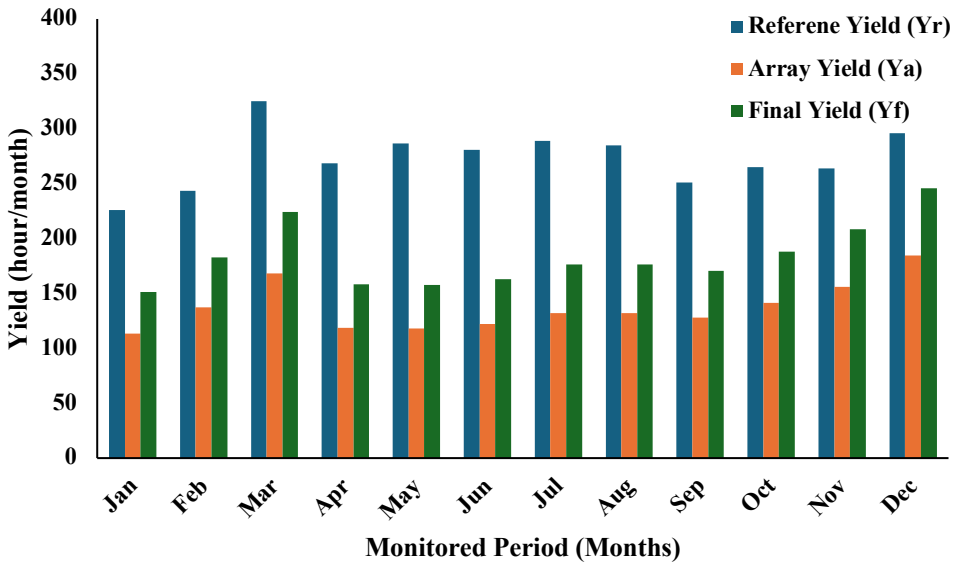


Fig. 3.11 Yields of 235 kWp at Maheshwaram, Telangana.

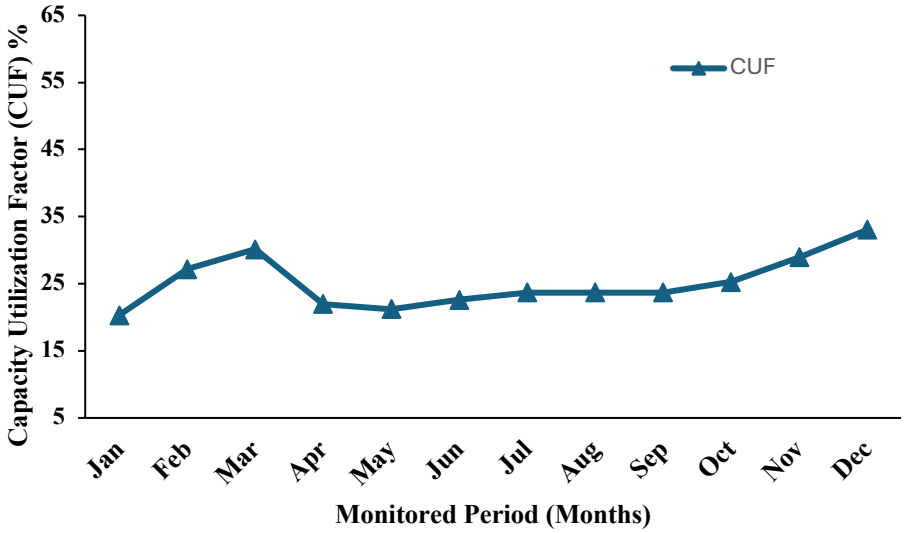


Fig 3.12. CUF of 235 kWp at Maheshwaram, Telangana.

3.5 Summary

This chapter documented the technical configurations and operational profiles of three rooftop PV systems across India. The comparative insights on system layout, yields, and PR provide a foundational basis for the performance modeling and degradation analysis in the subsequent chapters.

- Section 3.1 justifies the choice of three geographically diverse rooftop solar PV plants—located in Bengaluru (Karnataka), Khopoli (Maharashtra), and Maheshwaram (Telangana). The sites were selected based on the availability of reliable performance data, varying climatic conditions, and representativeness for urban, industrial, and semi-arid operational contexts.

- Section 3.2 outlines the configuration and operational profile of the 125 kWp rooftop PV plant at Sri Venkateshwara College of Engineering. It covers panel types, orientation, inverter setup, and yearly energy output, which serves as a reference point for degradation analysis and PR evaluation under semi-urban climatic conditions.

- Section 3.3 details the 191.9 kWp rooftop PV system at Arkose Industrial Estate in Khopoli, Maharashtra. It presents a full system overview including module ratings, string configuration, inverter sizing, and grid integration via multi-string inverter blocks. The plant's seasonal performance behavior is highlighted to support upcoming analysis of thermal losses and PR variability.

- Section 3.4 introduces the newly commissioned 235 kWp PV plant in Maheshwaram, Telangana. It includes plant layout, single-line diagram

interpretation, inverter and combiner box specifications, and a month-wise performance summary. The plant's high CUF and PR values across the year are noted, supporting its role as a validation benchmark in subsequent modeling chapters.

Chapter 4: Assessment of Failure Mode Based (R_d)

Overview:

This chapter presents a comprehensive methodology for evaluating and modelling the rate of power degradation (R_d) in photovoltaic (PV) systems under on-field conditions. Beginning with a detailed description of the experimental setup, it outlines the rationale for estimating R_d by integrating both performance monitoring and failure mode analysis. Qualitative on-field investigation methods, such as visual inspections, infrared thermography, and I-V curve tracing, are employed to capture degradation signatures. Subsequently, the experimental R_d is quantified using the Single Diode Model under STC reference baselines. The chapter further explores the selection of critical influencing parameters based on exploratory data analysis (EDA) and proposes predictive models including model uncertainty quantification to assess R_d. A comparative analysis of model performance across two distinct geographic sites, Telangana and Bengaluru, is also presented to validate the robustness and transferability of the approach.

Outline of Chapter **Page**

4.1 Importance for assessment of R_d (unitless normalized degradation factor)

The national push towards solar power has intensified, especially as India focuses on revolutionizing green hydrogen production. While large-scale deployment of renewable energy plants continues, a critical gap persists in ensuring their sustained performance. Societal acceptance and active engagement in energy usage are pivotal and not merely the commissioning of plants based on their rated peak capacities. In practice, these systems rarely operate at their rated peak due to numerous field-level inefficiencies, resulting in what is increasingly being referred to as the *"installed and ignored"* paradigm.

This disconnect between installed capacity and actual operational output underscores the need to move beyond installation metrics to performance-based evaluations. Field data often reveals significant deviations from expected output, which can compromise the economic viability, levelized cost of electricity (LCOE), remaining useful life (RUL), power scheduling, and overall system reliability of photovoltaic (PV) installations (Wieser.et.al.,2024; Dirk *et al.,* 2020). Among the various on-field factors such as MPPT efficiency, temperature coefficient of modules, soiling losses, and array mismatch the most critical yet under-monitored is module-level degradation. Studies suggest that module degradation alone can contribute to power loss rates of up to 10% per year in worst-case scenarios (Makrides *et al*., 2012).

The failure-mode-based power degradation rate, commonly denoted as R_d, quantifies the decline in power output of a PV system over time, accounting for losses beyond what would be expected from natural irradiance and temperature variability. These losses are typically failure mode-based aspects that occur during operation of the PV plant. R_d serves as a crucial metric for predicting remaining

useful life (RUL), planning maintenance schedules, and accurately estimating future energy yields. When the decline in power exceeds the expected environmental variation, the system is considered to be under degradation or experiencing fault conditions (Gerardo *et.al*.,2023).

Traditional models often simplify degradation as a linear decline typically 0.5–1% per year, but real-world long-term monitoring reveals a more complex, non-linear behaviour influenced by failure modes like thermal cycling, humidity ingress, potential-induced degradation (PID), delamination, and soiling (Sharma *et al*., 2018) as seen in figure 4.1. If this non-linearity is overlooked and inaccurately estimated, it can lead to serious financial and operational risks, including unexpected downtimes and erroneous LCOE projections.

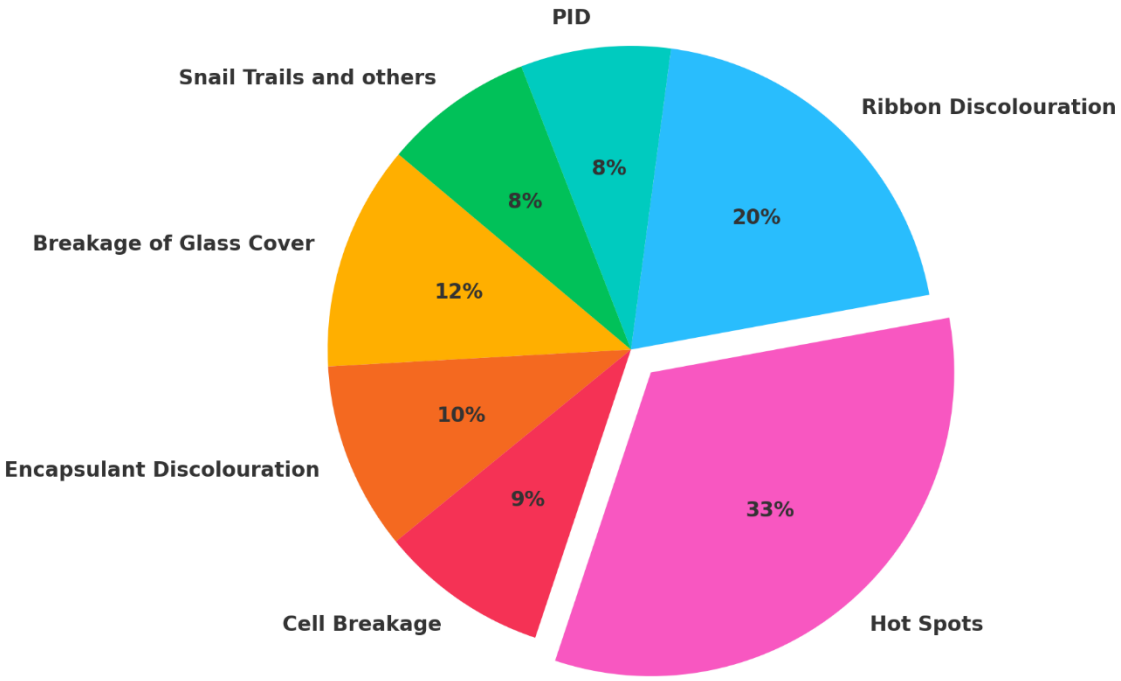


Fig 4.1. Power degradation percentage observed in silicon solar PV modules

The PV system studied in this analysis consists of a grid-connected crystalline silicon-based rooftop installation located in Bengaluru, India, with a total installed

capacity of 125 kW_p. The system includes 20 monocrystalline PV modules rated at 250 W each, interfaced with a maximum power point tracking (MPPT) inverter, and continuously monitored via real-time data acquisition. Key environmental parameters such as solar irradiance, wind speed, and ambient temperature conditions are logged at one-minute interval with the help of irradiance meter, wind speed sensor, temperature and humidity sensor with solar shield and two-channel data logger. This comprehensive data allows for the accurate evaluation of performance metrics and facilitates the estimation of R_d under realistic operating conditions.

Accurate prediction of R_d enables a realistic evaluation of the PV system's health, strengthens investor and policy confidence, supports predictive maintenance, and ensures long-term operational sustainability of renewable assets. In this research R_d is a dimensionless, normalized failure-mode-based power degradation rate derived from relative deviations in power output and does not represent an absolute or scaled power quantity

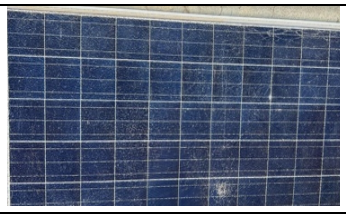
4.2 On-field qualitative investigation for assessment of R_d (unitless normalized failure-mode-based power degradation rate)

This research adopts a multi-pronged on-field approach incorporating visual inspection, thermal imaging, and I–V curve tracing to derive a comprehensive understanding of failure-mode-based power degradation rate and to justify its non-linear behaviour observed in field conditions.

4.2.1 Visual Inspection of the PV system under study

Visual inspection is a fundamental step in identifying early-stage failure modes in PV modules that may not immediately show up in electrical performance metrics but contribute to long-term degradation. In this research, a visual assessment was conducted on a rooftop 125 kW_p grid-connected PV system in Bengaluru. The inspection followed IEC 61215 standards, and 13 out of 20 modules exhibited visible defects as seen in table 4.1.

Table 4.1 Visual inspection Report carried out at investigated location

S.No	Type of Defect	No. of Modules Impacted	Image of Module
1	Decolorization	12	
2	Snail trail	5	
3	Soiling	8	
4	Cell cracks	2	

5	Back sheet delamination	2	
6	Junction box oxidation	10	
7	Glass breakage	0	-
8	Frame defects	0	
9	Bubbles on back sheet	2	

4.2.2 Thermal Imaging based Inspection

One of the primary applications of Forward Looking Infrared (FLIR) cameras in photovoltaic (PV) systems is the detection of cell and module-level anomalies Sharma *et al*. (2018). Thermal imaging helps identify hotspots, thermal irregularities that may result from bypass diode failures, cell cracks, or poor electrical contacts, each of which can significantly reduce energy yield and system efficiency. Given the current objective of evaluating non-linear power degradation rates in PV modules, thermal imaging was conducted in accordance with IEC 60904-12 guidelines, as illustrated in figure 4.2.

Prior to image acquisition, environmental parameters such as ambient temperature, relative humidity, incident solar irradiance, and wind speed were carefully recorded to ensure contextual accuracy as shown in figure 4.3 and 4.4. The FLIR E5-XT thermal camera, used in this investigation, offers a temperature detection range

from -20°C to 400°C and a thermal sensitivity of ≤0.05°C, making it suitable for field diagnostics .

Fig 4.2. On-field usage of FLIR Thermal scanner

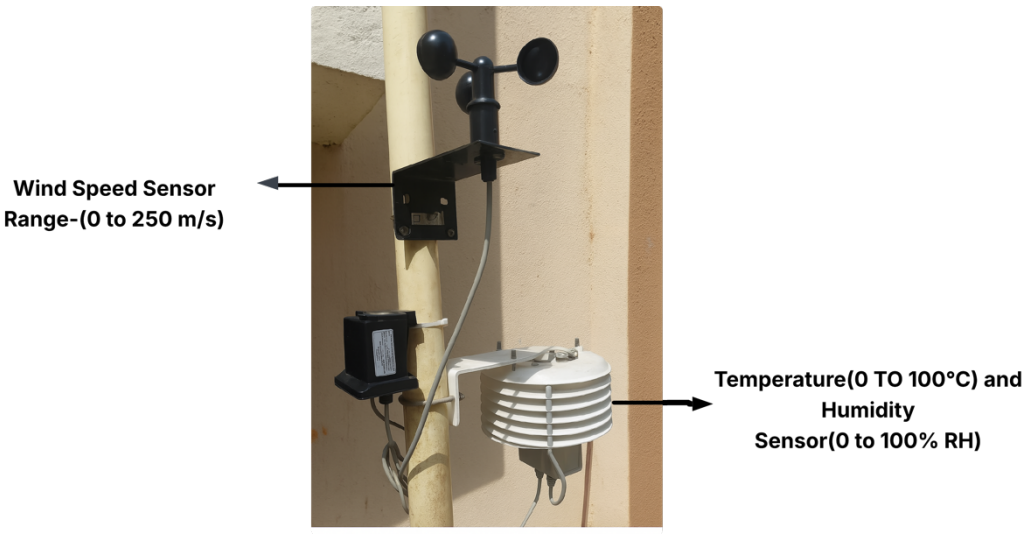


Fig 4.3. On-field installed Wind speed, Temperature and Humidity sensor at Bengaluru

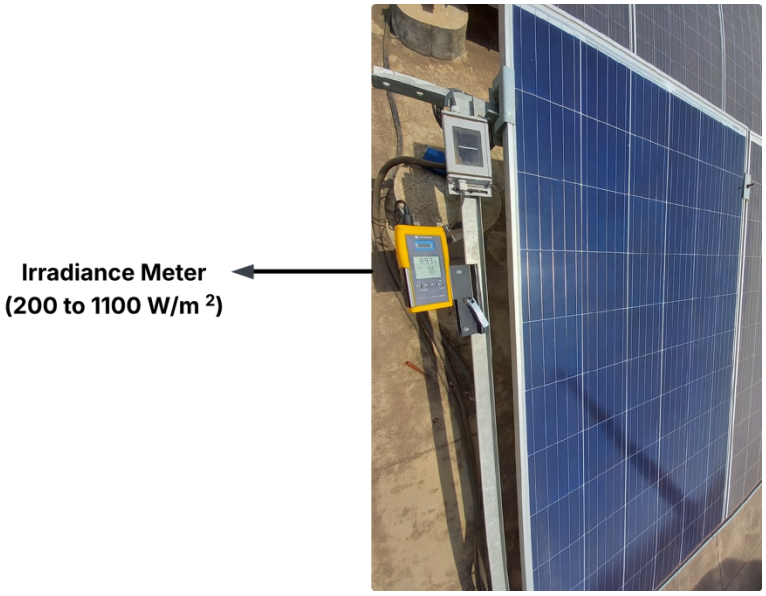


Fig 4.4. On-field Solar irradiance meter integrated to the PV system at Bengaluru

While thermal imaging method effectively detects surface hotspots remaining qualitative, correlating these anomalies to quantifiable power degradation requires additional analytical inputs. To this end, a new parameter termed DEL_T was introduced. DEL_T represents the difference between the maximum and minimum surface temperatures recorded on a module and is defined as in equation 4.1.

$$DEL_T = Tpv_{max} - Tpv_{min} \quad \text{Eq.(4.1)}$$

In this case, a maximum module surface temperature of 58°C and a minimum of 42.4°C were observed, resulting in a DEL_T value of 16°C as seen from the report generated employing FLIR Tools (Nesmachnow and Hernández,2019) in figure 4.5. This high thermal gradient serves as a significant predictive indicator for assessing the impact of localized heating on long-term power degradation R_d.

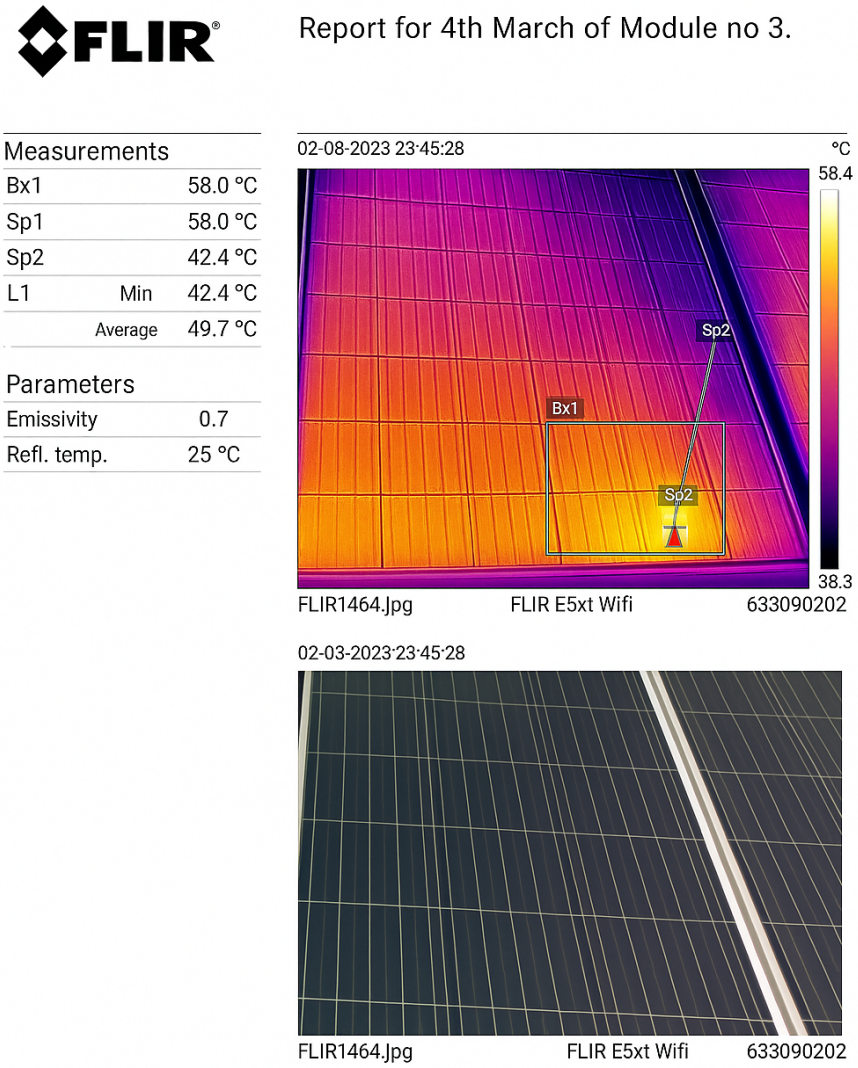


Fig 4.5. FLIR Report for calculation of DEL_T.

A significant thermal gradient exceeding 10°C was observed between the maximum and minimum module operating temperatures DEL_T over the monitored period, as illustrated in figure 4.5, indicating the formation of a hotspot. This observation underscores the critical role of DEL_T in influencing the power degradation behaviour of PV modules.

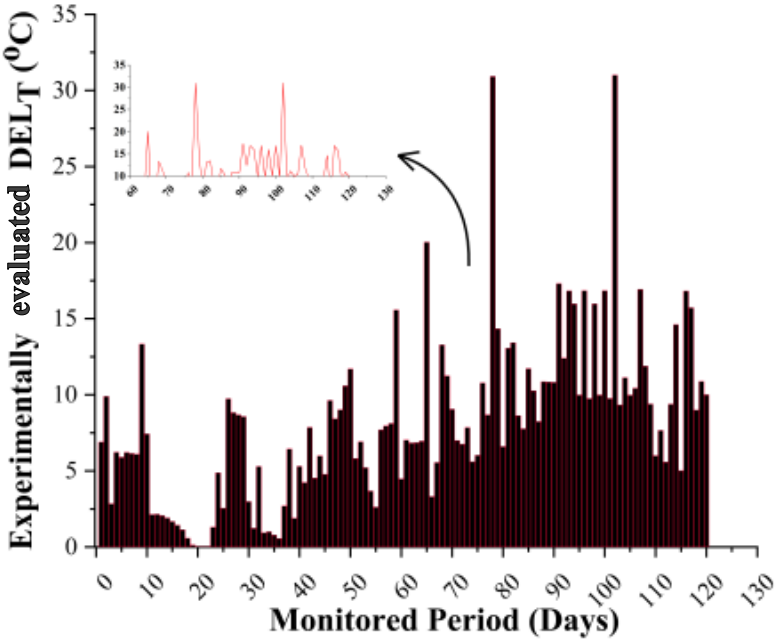


Fig 4.6. Experimentally evaluated DEL_T variation throughout the experimental period

As depicted in figure 4.6, the DEL_T parameter fluctuated throughout the monitoring period, with the maximum value reaching 30°C and the minimum dropping to 1.5°C. The highest thermal gradients were recorded between the 60th and 110th day, a duration that also corresponded with elevated solar irradiance levels ranging from 900 W/m² to 1100 W/m². The intensified irradiance contributed to sharp rises in module operating temperatures, which, when sustained, can result in irreversible thermal stress, leading to permanent degradation of the photovoltaic system. Thus, monitoring DEL_T is essential for anticipating failure onset and quantifying non-linear degradation effects under operational conditions.

4.2.3 On-field curve tracer-based Investigation

Following visual inspection and thermal imaging based on-field experimental investigation ensuring failure modes, a Solar V-I tracer-based examination was conducted at solar noon. The conduction of this investigation at solar noon ensures

standardized irradiance conditions for comparison and accurate assessment of R_d as shown in figure 4.7. PV module temperature was measured using contact-type RTD (PT100) sensors mounted on the rear surface of representative modules, following IEC 61724 recommendations.

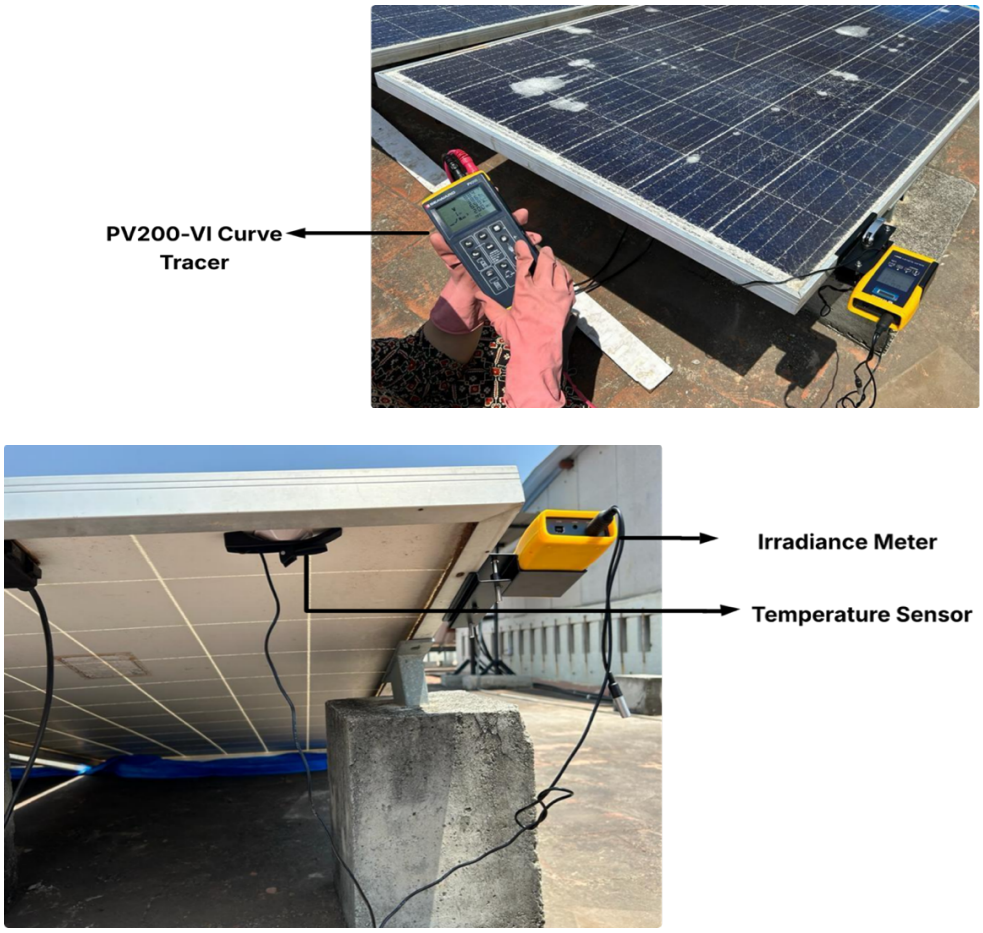


Fig 4.7. VI Curve tracer coupled with irradiance meter and temperature sensor for the considered system

This step is critical as the evaluation of field-level degradation must be benchmarked against Standard Test Conditions (STC), thereby ensuring the validity of performance comparisons. The methodological framework employed for estimating the unitless normalized failure-mode-based power degradation rate (R_d) through I-V field characterization is depicted in figure 4.8.

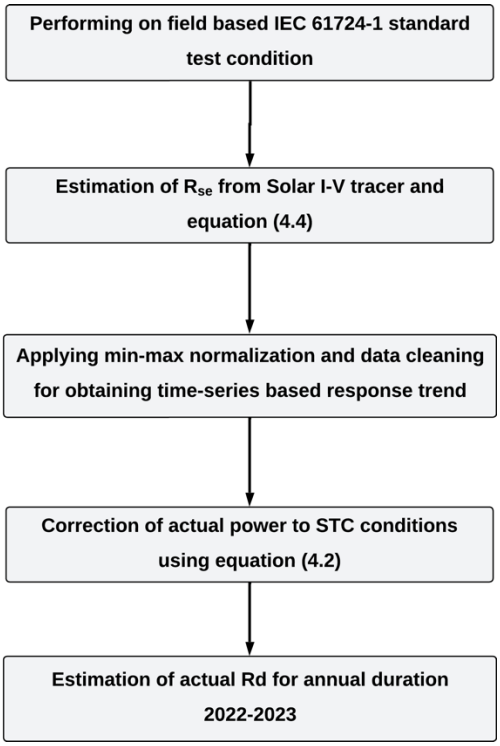


Fig 4.8. Methodology used for V-I Curve tracer-based estimation of R_d

To account for ambient field conditions, the actual power output was corrected using Equation (4.2), enabling normalization of performance data. One of the critical influencing parameters in this analysis is the series resistance (R_{se}), which significantly contributes to performance losses. A higher R_{se} leads to increased internal voltage drops, thereby reducing the module's output voltage and overall energy conversion efficiency. R_{se} represents a physically measurable electrical loss parameter associated with ohmic and contact-related losses in PV modules. An increase in R_{se} leads to a reduction in maximum power output. Previous studies like (Hansen and King,2019) have shown that long-term degradation mechanisms such as contact ageing, interconnect corrosion, solder bond fatigue, and metallization wear contribute to an increase in R_{se}. The parameter R_d used in this work is a dimensionless degradation indicator, defined to quantify the performance impact of failure-induced electrical losses, including those arising from increased R_{se}. Thus, R_d does not represent resistance itself but captures the normalized reduction in

power generation caused by degradation mechanisms, of which series resistance is a dominant contributor.The series resistance was calculated using Equations (4.3) and (4.4), and subsequently incorporated into the estimation of R_d (unitless normalized failure-mode-based power degradation rate) via Equation (4.5).

$$P_{CR} = P_M * \frac{G_{STC}}{G} * \frac{1}{[1+\gamma*(T_{cell}-T_{STC})]} \quad \text{Eq. (4.2)}$$

$$I_E=(1 - \frac{I_{MP}}{I_{SCO}}) * G_0 \quad \text{Eq. (4.3)}$$

$$R_{se}=\frac{V_{OC}(I_E)-V_{MP}}{I_{MP}} \quad \text{Eq. (4.4)}$$

$$R_d=\frac{P_{CR}}{P_{rated}} * 100 \quad \text{Eq. (4.5)}$$

Where, G_0 denotes the irradiance at which the measurement is taken, while $I_{MP}$and V_{MP} represent the current and voltage at the maximum power point, respectively. The short-circuit current under standard test conditions is denoted by I_{SCO} , and R_{se} refers to the series resistance of the PV module, which influences internal losses. The corrected irradiance value is expressed as I_E, and $V_{OC}(I_E)$corresponds to the open-circuit voltage adjusted to the irradiance level and referenced to STC. The corrected power at STC conditions is given by P_{CR} , while $P_M$is the actual measured maximum power output of the PV module and P_{rated} is 325 W.

Furthermore, G_{STC} is the standard irradiance level under STC, typically fixed at 1000 W/m², and G represents the actual irradiance obtained through the V-I curve tracer during field testing. The parameter γ is the power temperature coefficient (expressed as a percentage per degree Celsius), which quantifies the effect of temperature on power output. Finally, T_{cell} is the module surface temperature during testing, and T_{STC} is the standard cell temperature of 25°C under 1000 W/m² irradiance.

The resulting degradation factor, R_d (unitless normalized failure-mode-based power degradation rate) represents the rate of decline in module performance over time. Unlike the traditionally assumed linear decline, the actual R_d values were found to be non-linear, as illustrated in figure 4.9. This trend indicates an average annual degradation rate of 7.7%, strongly suggesting the occurrence of field-induced failure modes such as interconnect degradation, delamination and hot spots are actively influencing operational performance.

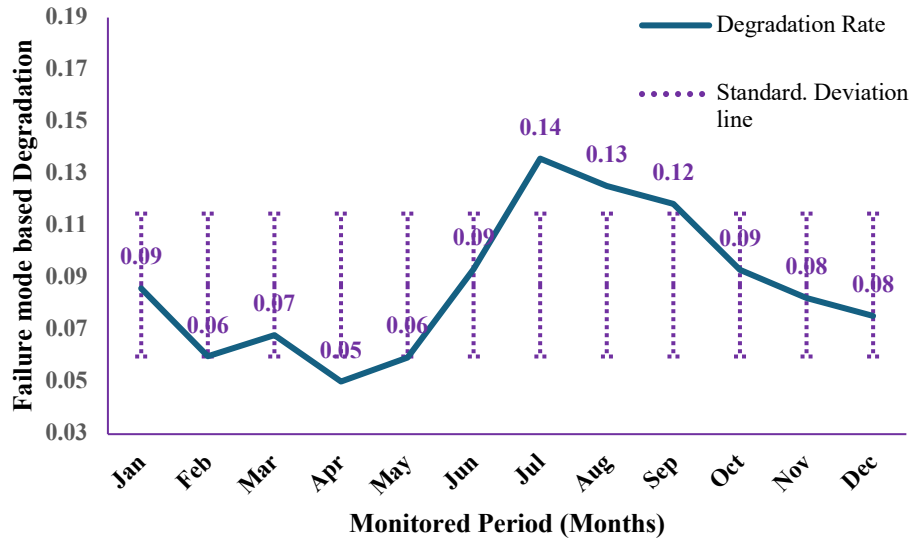


Fig 4.9 R_d (unitless normalized degradation factor) variation throughout the experimental period for the PV system under study

Figure 4.9 captures the statistical spread of R_d values, where the interquartile range (IQR) encompasses 50% of the dataset, analogous to the densest region of a Gaussian distribution. The whiskers of the plot extend outward, capturing outlier and extreme values, and demonstrating how frequency diminishes with increasing deviation from the median, further reinforcing the presence of non-uniform degradation behaviour in real-world operational conditions. The degradation indicator R_d used in this work represents a performance-based degradation factor, reflecting the cumulative impact of all failure modes on energy yield, rather than the intrinsic material degradation of individual failure modes/ modules alone.

Therefore, the reported value should be interpreted as a **realistic, field-**observed degradation rate under on-field conditions**.**

4.2.4 Validation of experimental R_d for the considered PV system (Single Diode Model and STC Baselines)

To validate the presence of degradation and failure modes within the considered PV system, the output derived from the on-field V-I curve tracer was systematically compared with the theoretical output predicted by the benchmarked Single Diode Model (SDM) (Lidaighbi.et.al.,2020) under the actual/ field monitored irradiance and module temperature conditions. The SDM, being a well-established electrical model that characterizes the ideal behaviour of PV modules, provides a benchmark against which on-field deviations can be evaluated.

By applying identical field conditions including irradiance, cell temperature, and series resistance into the single diode model, the expected performance was computed as seen from figure 4.10 to figure 4.16. The PV module parameters considered under rated conditions are expressed in table 4.2. This was then compared with the actual performance data obtained from the V-I tracer to further justify the presence of failure mode.

Table 4.2. Rated values used in the SDM model

Parameters of considered PV panel	**Rated Values**
Rated Power(V_{mp})	325 W
Voltage at maximum power(V_{mp})	38.80 V
Current at maximum power(I_{mp})	8.37 A
Open circuit voltage(V_{oc})	46.40 V
Short circuit current(I_{sc})	9.25 I
Total number of cells in series (N_s)	72
Total number of cells in parallel(N_p)	1

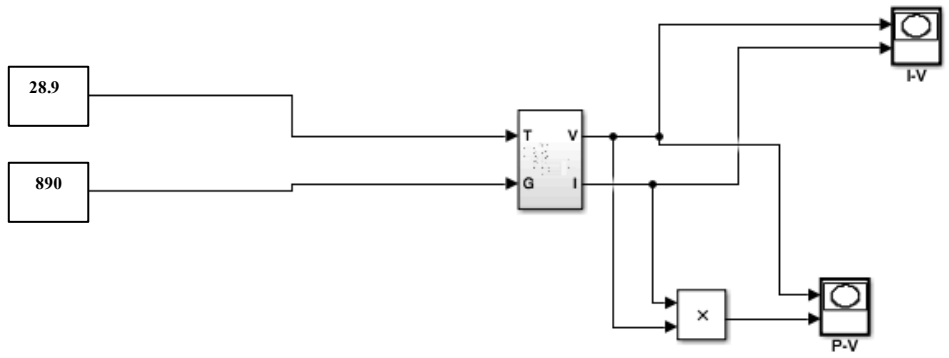


Fig 4.10 Simulink Model for the considered 325 W solar panel

This comparative approach offers a reliable diagnostic framework for pinpointing underperformance due to hidden or evolving failure modes that may not be immediately evident through visual inspection alone. As summarized in table 4.4, the discrepancies observed confirm the onset of system degradation, thereby validating the importance of incorporating physical and model-based diagnostics in the failure-mode-based power degradation rate (R_d) assessment methodology.

Table 4.3 Equations used in Simulink model (Yao.et.al.,2023)

S.No	Model components	Computation method
1	Photo current	$I_{ph} = [I_{sc} + K_i(T - 298)] * G/1000$
2	Reverse Saturation Current	$I_{rs} = \frac{I_{sc}}{\exp\left(\frac{qV_{oc}}{N_s kT}\right) - 1}$
3	Saturation Current	$I_0 = I_{rs}\frac{T}{T_n}^3 \; exp\left[\frac{nE_{g0}}{nk}\left(\frac{1}{T_n} - \frac{1}{T}\right)\right]$
4	Current Through Shunt Resistor	$I_{shunt} = \frac{(V + IR_{series})}{R_{shunt}}$
5	Output Current	$I_{op} = I_{photo} - I_0 exp\left[\frac{q(V + I_{rs}}{nkN_s T} - 1\right] - I_{sh}$

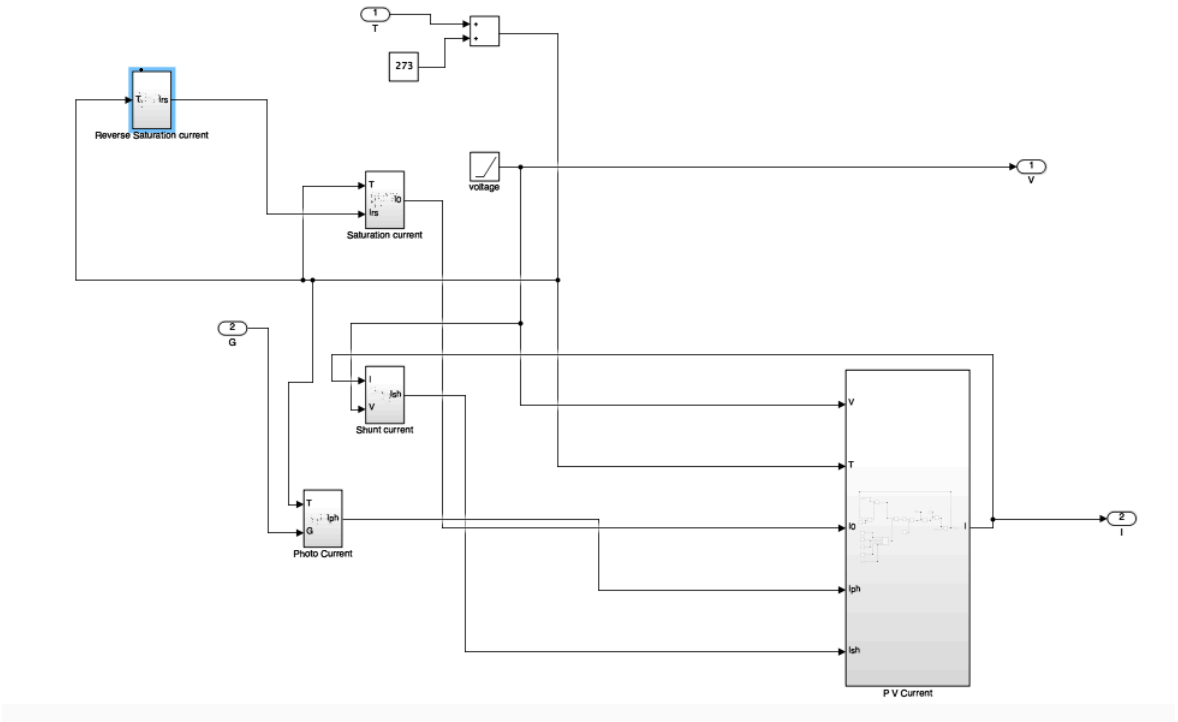


Fig 4.11 Subsystem of Modelled Solar PV Panel

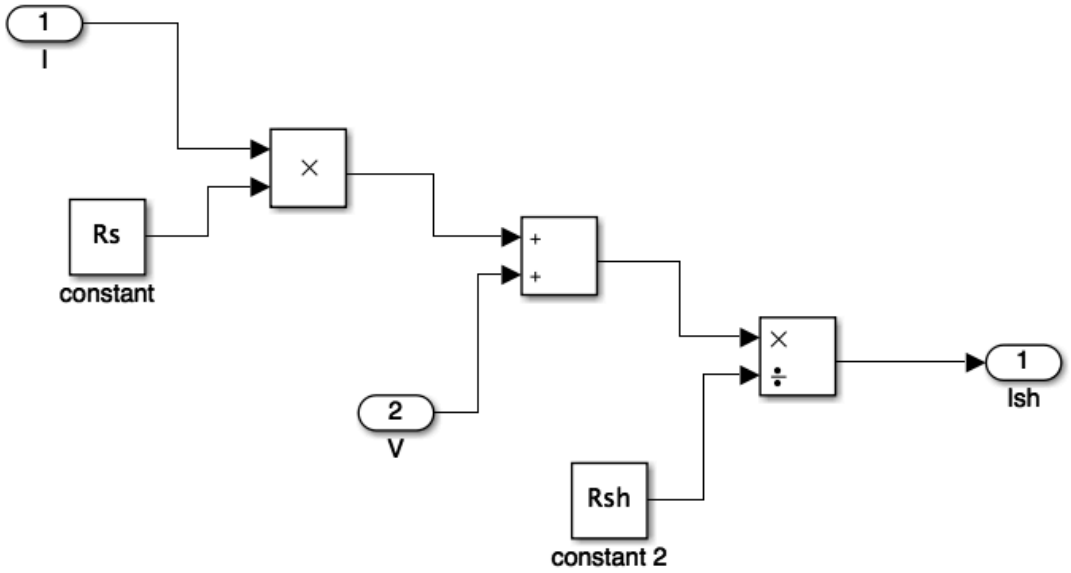


Fig 4.12 Computation of Shunt current (Ish) as represented in table 4.3

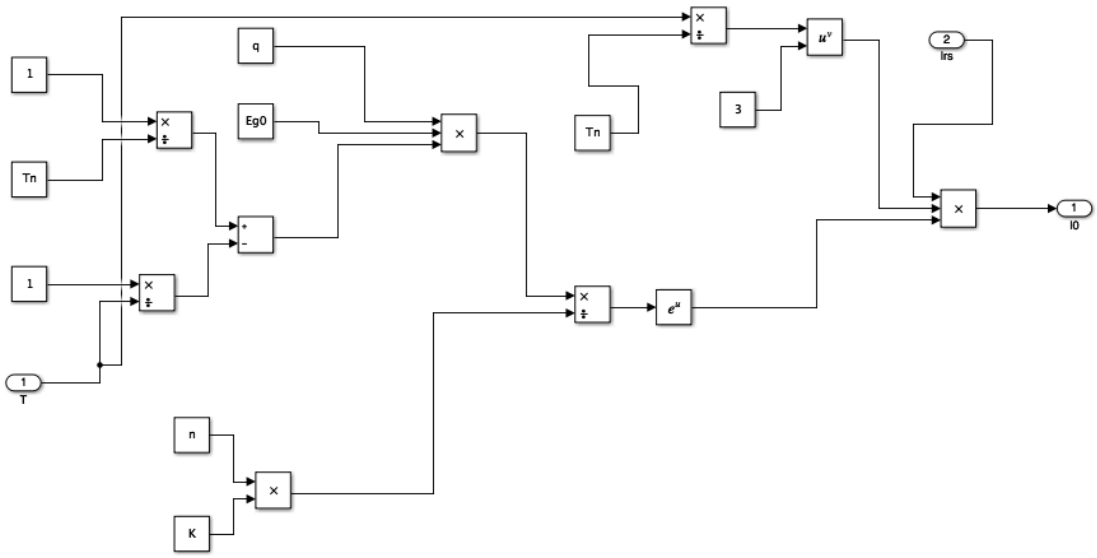


Fig 4.13 Computation of Saturation Current (I_o) as represented in Table 4.3

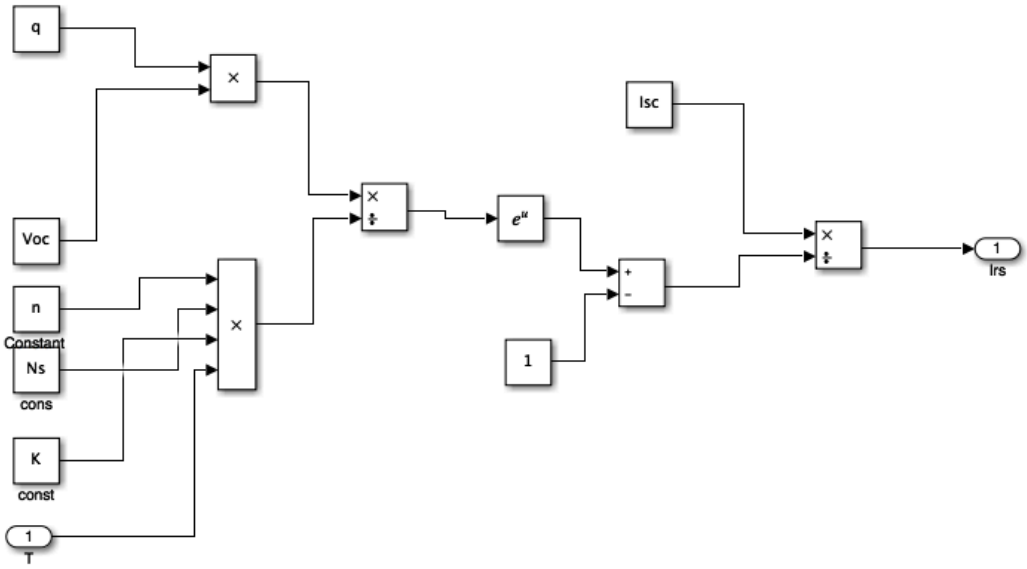


Fig 4.14 Computation of Reverse Saturation Current (I_{rs}) as represented in Table 4.3

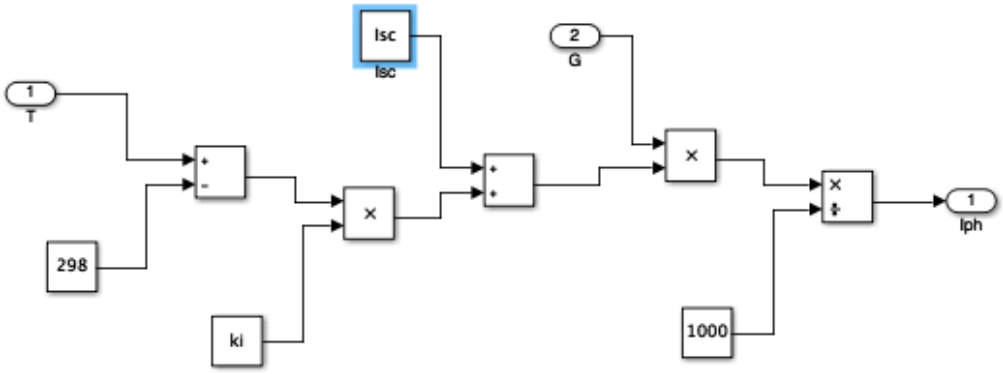


Fig 4.15 Computation of Photo current as represented in table 4.3

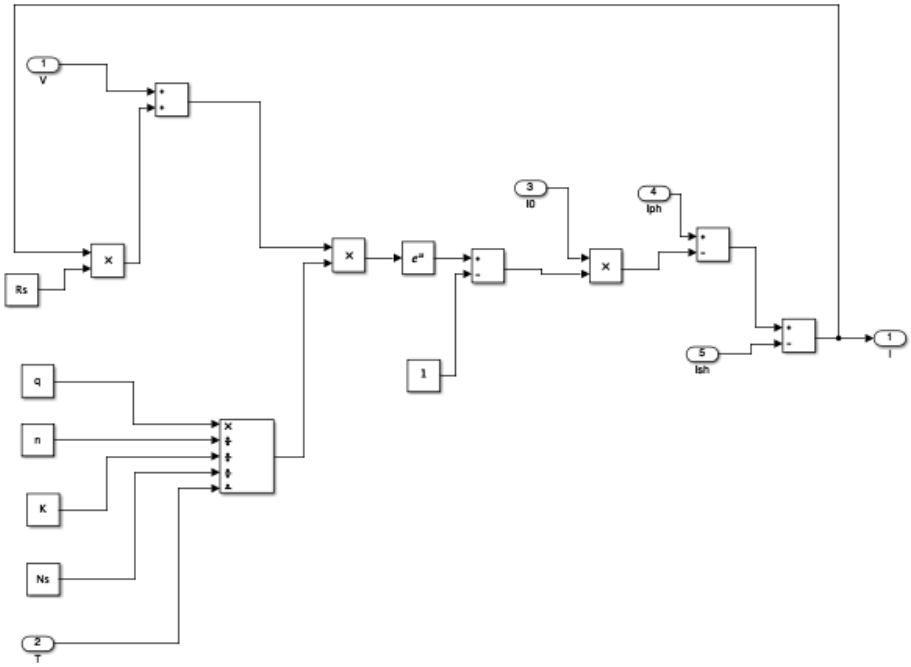


Fig 4.16 Computation of PV Output Current as represented in table 4.3

Table 4.4 Evaluation of actual measured power output against the single-diode model under field conditions to ensure failure mode occurrence

S.No	Type	I (W/m^2)	T_m (^0C)	V_{OC} (V)	I(A)	V_m (V)	I_m(A)	T_{ci} ($/^0C$)	P_m (W)
1	Standard test conditions of a 325 W_p PV module	1000	25	45.7	9.22	37.4	8.7	0.05	325
2	Single diode model	1058	64.38	45	9.2	32	12	0.125	310
3	I-V Curve Tracer (Experimental)	1058	64.38	39.2	8.056	31.19	7.7	0.125	245.63

4.3 Selection of critical factors for development of R_d model

Identifying statistically significant input variables is essential for enhancing the prediction accuracy of the proposed power degradation (R_d) estimation. In this study, the inputs considered include Irradiance (I), Wind Speed (W), Relative Humidity (H), Series Resistance (R_s), Module Temperature (T_m), Ambient Temperature (T_a), and the DEL_T parameter, which reflects the thermal gradient observed on-field.

The statistical significance of these input variables was assessed using Pearson's correlation coefficient, which quantifies the linear relationship between two

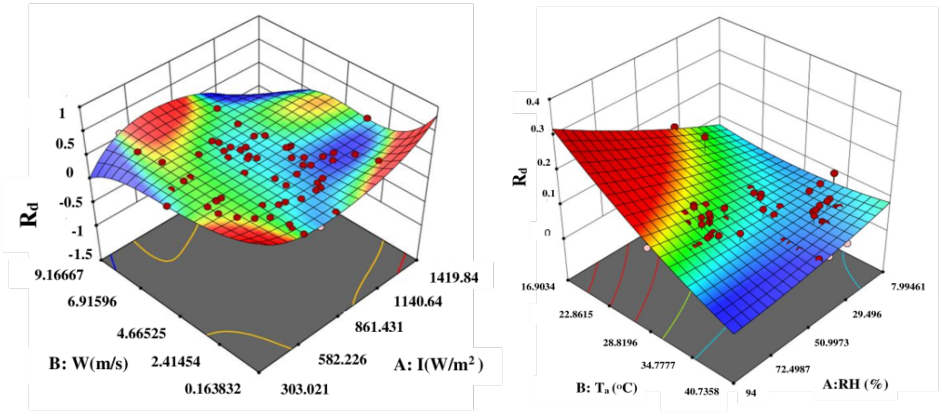
variables. In this context, a p-value less than 0.05 indicates a statistically significant correlation, meaning that the input variable has a meaningful influence on the output parameter, R_d. As detailed in table 4.5, the computed p-values confirm that each of the selected input variables contributes significantly to the output prediction.

Table 4.5. Model Input parameters p-value.

Model Parameters	**p -value**
I	0.00617
W	0.00958
H	0.0087
R_{se}	0.00254
DEL_T	0.00046
T_m	1.0737×10^{-10}
T_a	9.2437×10^{-9}

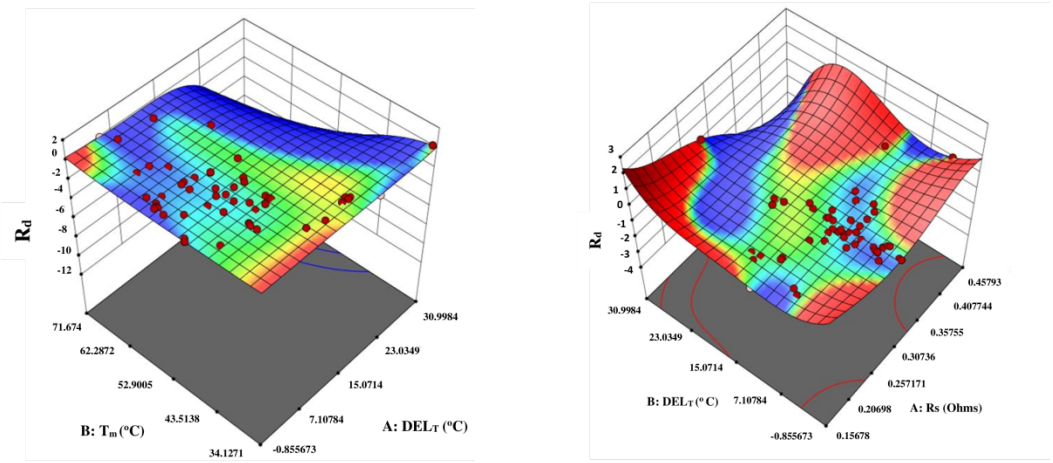
Furthermore, the influence of variations in independent input parameters on the failure-mode-based power degradation rate (R_d) is illustrated in figures 4.17 and 4.18. In particular, figure 4.17 presents a two-dimensional contour plot where each contour line denotes a constant R_d value. The spacing and distribution of these lines represent the sensitivity of R_d to changes in the input parameters. The contour map depicting the relationship between irradiance and wind speed shows a broad and well-defined gradient, indicating their substantial impact on the prediction of power degradation in the solar PV module. This widespread variation across the contour space justifies the selection of irradiance and wind speed as critical input features for the present study. Figure 4.18 presents a surface plot illustrating the combined impact of series resistance (R_s) and temperature difference (DEL_T) on the failure-

mode-based power degradation rate (R_d). The relatively flat surface region indicates that areas with higher R_s and DEL_T values are associated with elevated failure-mode-based power degradation rate. Specifically, degradation tends to be more significant when R_s lies in the range of 0.25 to 0.3 ohm, and DEL_T ranges between 15.06°C and 30.99°C. Regions located at the minimum side of the surface contour represent data conditions with lower R_s and DEL_T values, corresponding to minimal R_d. This suggests that controlling R_s and thermal gradients is critical for minimizing degradation. These two parameters are treated as consequence-based indicators of failure modes, and their strong influence on R_d highlights their importance in evaluating a PV module's long-term performance. Therefore, inclusion of as R_s and DEL_T as input features significantly contributes to improving the predictive accuracy of the proposed model.



(a) (b)

Fig 4.17.(a) 3-D Surface graph of R_d with respect to Irradiance and wind speed (W) (b) 3-D Surface graph of R_d w.r.t Ambient Temperature (T_a) and Relative Humidity (RH)



(a) (b)

Fig 4.18 (a) 3-D Surface plot of R_d w.r.t Module Temperature and DEL_T. (b) Surface plot of R_d w.r.t Series Resistance R_s and DEL_T

To assess the effect of increasing the number of input parameters on the prediction accuracy of the model, Bland-Altman plots were generated for three cases: using 2 inputs, 5 inputs, and 6 inputs. These plots display the mean of the actual and predicted failure-mode-based power degradation rate (R_d) on the x-axis and the difference (actual − predicted) on the y-axis, along with indicating the mean bias and ±1.96 standard deviation (SD) limits, which define the 95% confidence interval.

As evident from the figure 4.19 (2 inputs), the differences are relatively more scattered and lie closer to the upper and lower SD limits, indicating a higher degree of estimation uncertainty. The spread of the data points across the vertical axis implies less agreement between actual and predicted value of R_d in consideration of only two input factors.

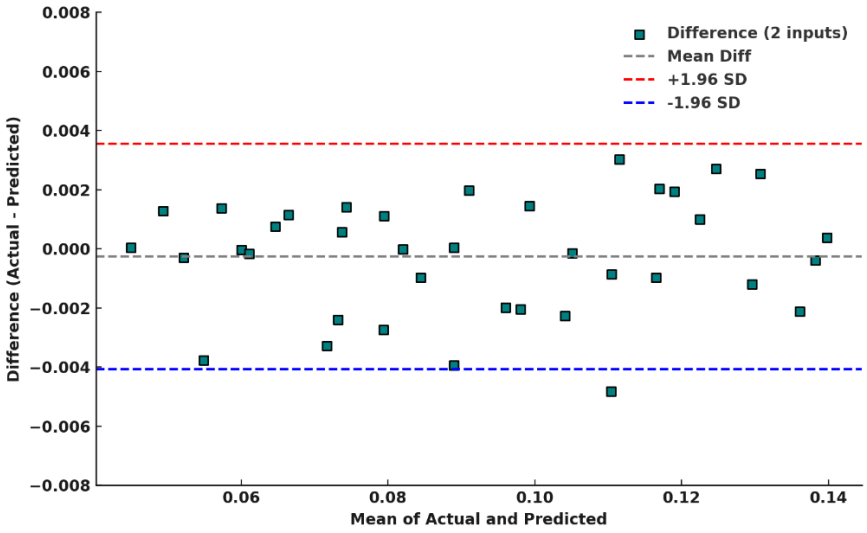


Fig 4.19. Bias and Limits of Agreement for Predicted vs Actual R_d with Two inputs

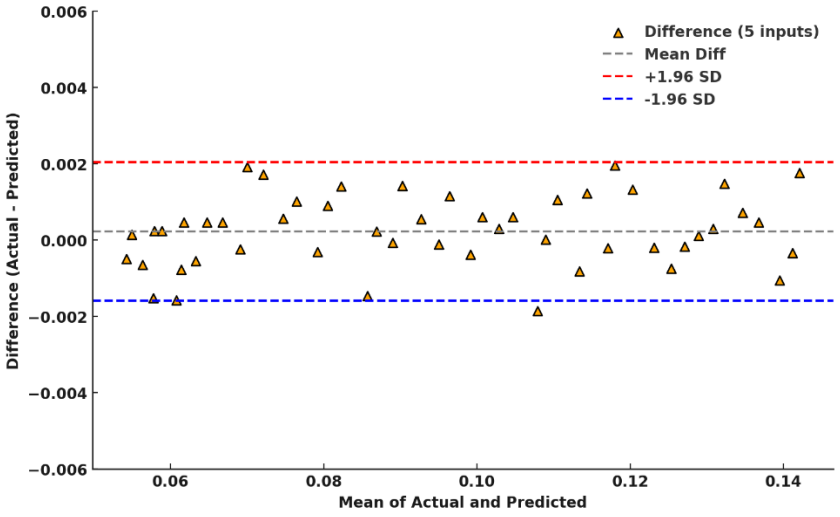


Fig 4.20. Bias and Limits of Agreement for Predicted vs Actual R_d with Five inputs

When the number of inputs is increased to 5, figure 4.20 shows a noticeable reduction in the spread of differences existing between the mean of actual and predicted values. A higher concentration of data points lies closer to the zero-difference line (mean bias), and fewer points approach the SD boundaries. This tighter clustering reflects improved prediction consistency and reduced bias on increase in the number of input attributes from 2 to 5.

Figure 4.21illustrates further refinement in the predicted estimate of R_d on considering 6 input parameters. Most data points are even more concentrated near the zero-difference line, and the vertical spread is narrower than in the previous cases. Importantly, none of the observations lie outside the ±1.96 SD bounds, highlighting higher model reliability and tighter agreement on inclusion of six attributes towards modelling failure-mode-based power degradation rate (R_d).

This progression across the plots validate that the inclusion of additional meaningful and significant input features improves the model's ability to estimate R_d accurately. The narrowing of the confidence interval and the concentration of residuals near zero support the conclusion that incorporating five or six input variables enhances model robustness and minimizes error variance, ultimately leading to improved predictive performance.

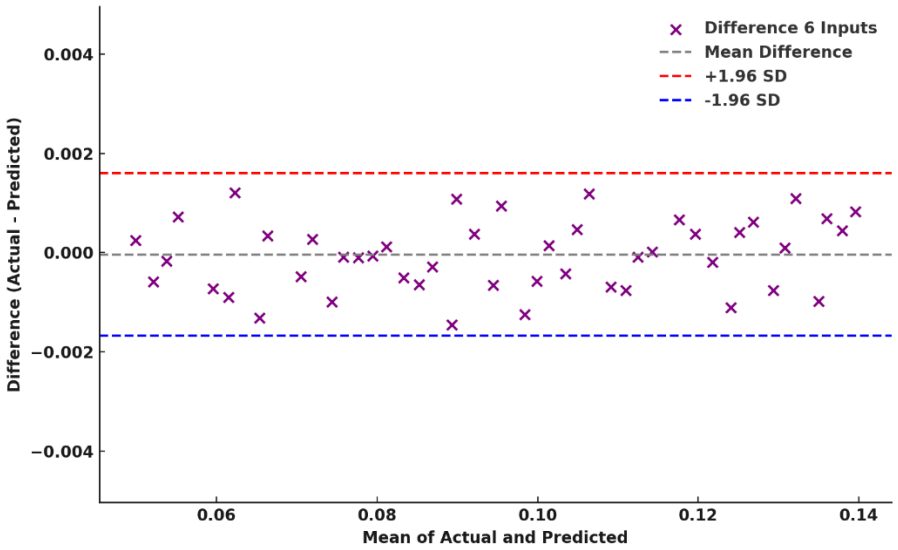


Fig 4.21. Bias and Limits of Agreement for Predicted vs Actual R_d with Six inputs

4.3.1. Data Normalization and Data cleaning

The raw experimental input data initially exhibited several outliers, which were carefully identified and replaced using a weighted average of the nearest significant observations to maintain data integrity while minimizing distortion. Data points with solar irradiance values below 200 W/m² were also excluded, as such low irradiance conditions do not contribute meaningfully to photovoltaic output under standard test conditions and typically correspond to non-failure operational states. In addition to irradiance, thresholds were defined for other input parameters based on domain knowledge, exploratory data analysis (EDA) methods such as box plots and scatter plots, and reference to industry standards including IEC and NREL guidelines. Specifically, data was filtered to exclude wind speed values below 0.5 m/s or above 12 m/s, relative humidity values below 15% or above 95%, module temperature readings below 15°C or above 85°C, and ambient temperatures outside the range of 10°C to 50°C. Similarly, corresponding operational electrical parameters were filtered to remove outliers, with open-circuit voltage (V_{oc}) restricted to the range of 10 V to 50 V, maximum power current (I_{mp}) to 1 A to 9 A, and maximum power voltage (V_{mp}) to 15 V to 45 V. These cut-off points were selected based on observed data distributions and the typical operating ranges of silicon-based PV modules under field conditions. After cleaning, all input parameters were scaled using the min-max normalization technique, which transforms each value based on its relative position between the observed minimum and maximum for that variable. This method was chosen because it introduces minimal error and improves model convergence while preserving the interpretability of results. The normalized value of R_d was computed using the equation (4.6).

$$Y_t = \frac{Y_{i,t} - Y_{i,t}^{min}}{Y_{i,t}^{max} - Y_{i,t}^{min}} \quad \text{Eq.(4.6)}$$

where $Y_{i,t}$ is the original observation and $Y_{i,t}^{min}$ and $Y_{i,t}^{max}$ are the corresponding variable's minimum and maximum values. This comprehensive filtering and normalization approach ensured that only meaningful, consistent, and relevant data was used for model development and analysis.

4.4 Proposed models for prediction of R_d

Two distinct predictive frameworks referred to as Model 1 and Model 2 have been developed utilizing machine learning methodologies, specifically Long Short-Term Memory (LSTM) and Feed-Forward Back Propagation (FFBP) networks. The architectural differentiation between the models is rooted in the variation of input parameters, which are systematically detailed in table 4.6. These variations facilitate the classification and functional distinction between the two models.

Table 4.6. Parameters for Model 1 and 2.

R_d model attributes derived using LSTM and FFBP approach	**Parameters**
Model 1	I, W, H, R_s and DEL_T
Model 2	I, T_m, T_a, W, H and R_s

4.4.1 R_d model based on LSTM approach

Recurrent Neural Networks (RNNs), particularly the Long Short-Term Memory (LSTM) variant, were developed to overcome the vanishing gradient problem that commonly affects traditional RNN architectures during long sequence learning [39]. In this context, the present study leverages an LSTM-based modelling approach specifically designed to predict power degradation due to failure modes in photovoltaic systems.

LSTM networks, a specialized form of RNNs are characterized by their internal memory structure referred to as the LSTM cell which enables them to capture long-term dependencies within sequential data. This cell comprises three principal components or gates: the input gate, forget gate, and output gate. These gates collaboratively regulate the flow of information into, within, and out of the memory cell, thereby ensuring that relevant features are retained and irrelevant ones are discarded over time.

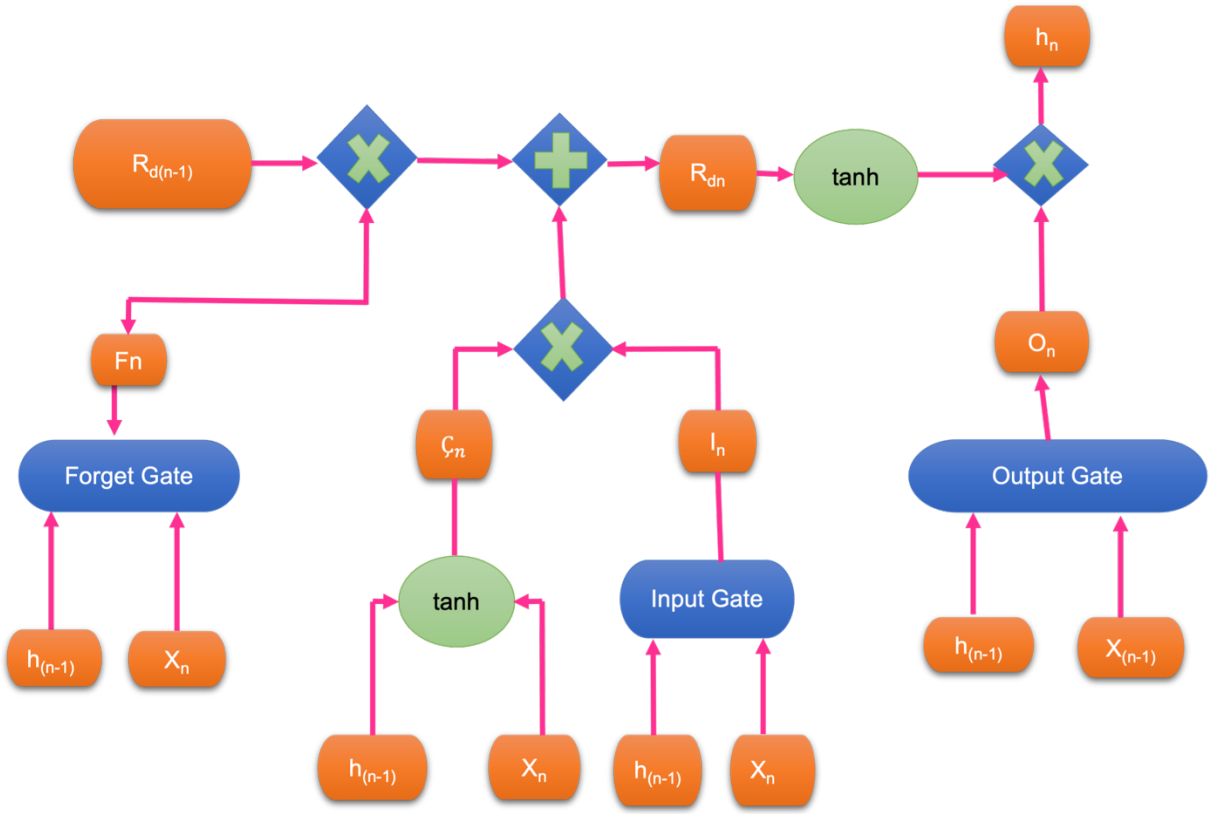


Fig 4.22. Proposed Long Short-Term Memory(model) architecture

In the proposed model, at any given time step n, the network receives a multivariate input X_n, which consists of environmental and system-specific parameters such as wind speed, solar irradiance, series resistance, humidity, ambient temperature, and

module temperature. Based on this input, along with the previous hidden state h_{n-1}, the LSTM cell updates its internal memory to produce the current hidden output h_n, the cell state $Ç_n$, and the predicted failure-mode-based power degradation rate R_{dn}.This prediction is informed by both the current input and the prior degradation state R_{dn} as seen in figure 4.22.

The functional behaviour of the gates input gate I_n, forget gate F_n, and output gate O_n is mathematically defined through a series of equations (4.7) to (4.9), which govern the transformation and propagation of information across the LSTM architecture.

$$I_n = \sigma(W_1^I.x_n + W_h^I.h_{n-1} + b_I) \quad \text{Eq.(4.7)}$$

$$F_n = \sigma(W_1^F.x_n + W_h^F.h_{n-1} + b_F) \quad \text{Eq.(4.8)}$$

$$O_n = \sigma(W_1^O.x_n + W_h^O.h_{n-1} + b_O) \quad \text{Eq.(4.9)}$$

$$Ç_n = \tanh(W_1^c.x_n + W_h^c.h_{n-1} + b_c) \quad \text{Eq.(410)}$$

In the LSTM cell architecture, the input gate I_n, forget gate F_n and output gate $O_n$regulate the information flow, while $Ç_n$ represents the cell input activation. The weights associated with these components include W_1^I , W_1^F , W_1^O & W_1^c which connect the current input Xn to the respective gates and the cell input. Additionally, the weight matrices W_h^I , W_h^F W_h^O & W_h^c are responsible for linking the hidden state from the previous time step h_{n-1} to the gates and cell input.

Each gate also includes a bias term: b_I , b_F , b_O & b_c, which are added to the gate calculations to enhance learning flexibility. The non-linear activation functions used include the sigmoid function $\frac{1}{1+\exp(-x)}$ and the hyperbolic tangent function $\frac{\exp(x)-\exp(-x)}{\exp(x)+\exp(-x)}$ enabling the model to handle both linear and non-linear temporal dependencies.

The cell state update is governed by equation (4.10), where the degradation output at time step n, denoted as R_{dn} as given in equation (4.11) which represents the proposed model of each cell for estimation of daily R_d. 256 cells along with a batch size of 128, considered for better performance of the models.

$$R_{dn} = I_n . Ç_n + F_n . Rd_{n-1} \quad \text{Eq.(4.11)}$$

$$h_n = O_n . \tanh(C_t) \quad \text{Eq.(4.12)}$$

$$\tilde{x}_{n+1} = W_2 . h_t + b \quad \text{Eq.(4.13)}$$

The hidden state output h_n is derived using equation (4.12). Finally, the cell output $\tilde{x}_{n+1}$is determined through equation (4.13). W_2 represents the weight matrix connecting the hidden layer to the output layer, and b is the output layer bias. These operations collectively enable the proposed model derived employing LSTM approach to predict the daily failure-mode-based power degradation rate based on current and historical environmental parameters and failure modes.

4.4.2 R_d model based on FFBP approach

In addition to LSTM derived R_d models as developed and detailed above, a Feed Forward Back Propagation (FFBP) neural network architecture was employed to predict the daily failure-mode-based power degradation rate (R_d) of solar PV modules. Two FFBP based models employing the input attributes as mentioned in table 4.6 (for model-1 and model-2) are correspondingly formulated. The network architecture leverages a multilayer perceptron (MLP) structure wherein the input layer feeds data forward through two hidden layers, and the weights are updated using the back-propagation algorithm as described by (Woldegiyorgis.et.al,2023)..

In this configuration, a sigmoid activation function (tansig) is used in the hidden layers, and a linear transfer function (purelin) is applied at the output layer. The hidden layer comprises 10 neurons, each receiving weighted input from five selected input parameters. The mathematical representation of the layer-wise computations is provided below from equation (4.19) to equation (4.25).

$$Y_{H1} = \sum_{n=1}^{5}\left((W_{i1H1}X_1 + W_{i2H1}X_2 \pm -- + W_{i5H1}X_5) + b\right) \quad \text{Eq. (419)}$$

$$Y_{H2} = \sum_{n=1}^{5}\left((W_{i1H2}X_1 + W_{i2H2}X_2 -- + W_{i5H5}X_5) + b\right) \quad \text{Eq. (4.20)}$$

Similarly

$$Y_{H10} = \sum_{n=1}^{5}\left((W_{i1H10}X_1 + W_{i2H10}X_2 -- + W_{i5H10}X_5)b\right) \quad \text{Eq. (4.21)}$$

Y_{H1} represents the output of neuron 1 of hidden layer 1; similarly, Y_{H10} represents the output of neuron 10 of hidden layer 1; The formulated network has 10 number of hidden neurons. X_1 to X_5 represents the input's considered as per Model-1; W_{i1H1} is the value of weight connecting input 1 to first neuron in hidden layer 1. Similarly, W_{i1Hk} is the value of weight connecting input 1 to k^{th} neuron in hidden layer 1; where k is the number of neurons; b and b_2 represents the bias connected between input and hidden layer1 and hidden layer 2 and output layer respectively.

$$Y_{H2-1} = \sum_{k=1}^{10}(W_{h11-h21}Y_{H1} + W_{h12-h21}Y_{H2} -- + W_{h110-h21}Y_{H10}) + b1 \quad \text{Eq. (4.22)}$$

$$Y_{H2-2} = \sum_{k=1}^{10}(W_{h11-h22}Y_{H1} + W_{h12-h22}Y_{H2} -- + W_{h110-h22}Y_{H10}) + b1 \quad \text{Eq. (4.23)}$$

Similarly

$$Y_{H2-k} = \sum_{k=1}^{10}(W_{h11-h2k}Y_{H1} + W_{h12-h2k}Y_{H2} -- + W_{h110-h2k}Y_{H10}) + b1 \quad \text{Eq.(4.24)}$$

$$R_d = \sum_{k=1}^{10}(Y_{H2-1}W_{OH2-1} + Y_{H2-2}W_{OH2-2} + - - + Y_{H2-K}W_{OH2-10} + b_2) \quad \text{Eq.(4.25)}$$

Where Y_{H2-1} represents the output of the first neuron in the second hidden layer; $W_{h11-h2k}$ represents the weight connecting the neuron 1 of hidden layer 1 to k^{th} neuron in hidden layer 2. W_{OH2-1} represents the weight connecting the output layer and the first neuron of the second hidden layer. Equation (4.25) based R_d model can be employed for prediction of daily R_d involving inputs at current time instant.

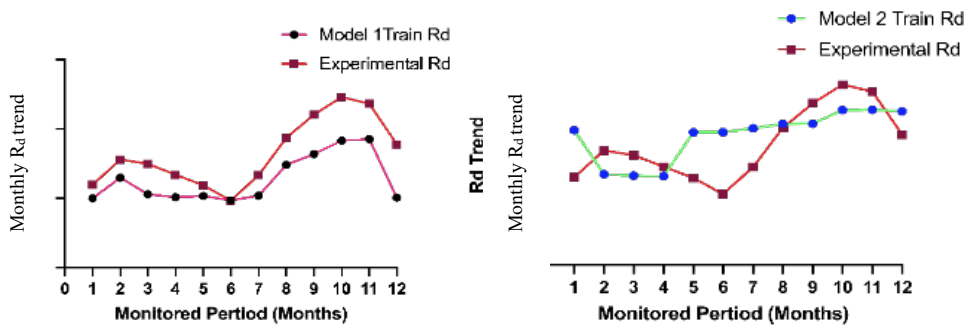
4.5 Results and Discussion

4.5.1 Predictive assessment of Proposed LSTM based R_d Model

80% of the experimental R_d values were allocated for training R_d model-1 and model-2, while the remaining 20% were reserved for validation. Model -1 and Model-2 of LSTM based with proposed architecture as in figure 4.22 and expressed as in equation (4.11) was provided with appropriate field attributes of the day to predict the daily values of R_d. The daily predicted values are averaged monthly, and a monthly Mean of daily observation of R_d is found to closely align with the training set as seen in figure 4.23 (a) and 4.23 (b), indicating good generalization. The Mean Absolute Percentage Error (MAPE) for Model 1 and Model 2 was found to be 0.03283 and 0.02832, respectively. The inclusion of the DEL_T parameter in Model 1 enhanced the model's accuracy, yielding a 5.7%

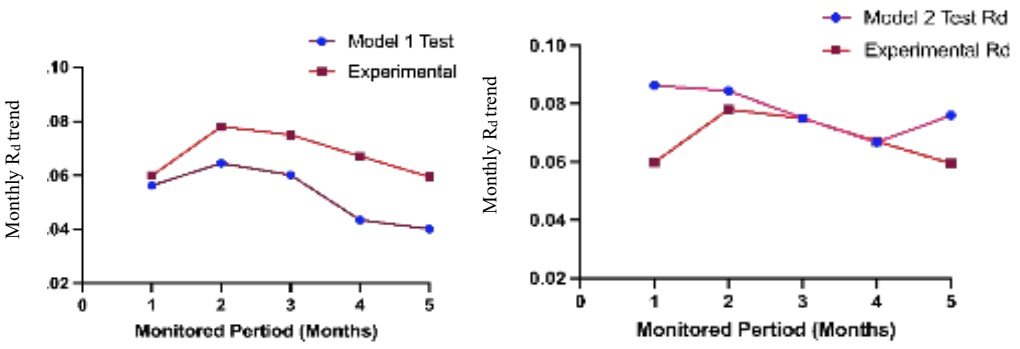
reduction in RMSE, thus demonstrating its significance in estimating the failure-mode-based power degradation rate (R_d).

As depicted in figure 4.24 (a) and 4.24 (b) during validation/testing, Model 2 demonstrates a closer fit to the actual test values, whereas Model 1 effectively captures the overall trend. This is reflected through the MAPE value of 0.023 for model -2 in comparison with proposed model-1 during testing /validation. The training RMSE for Model 1 is 0.000612, while that for Model 2 is 0.0261835, indicating higher accuracy in Model 1. The testing RMSE for Model 1 is 0.001984, while that for Model 2 is 0.002062, indicating higher accuracy in model 1. This performance difference is further supported by the average of RMSE based training and testing errors shown in figure 4.25 where Model 1 and 2 compete.



(a) (b)

Fig 4.23. (a) and (b) Comparison of monthly mean of daily observations predicted and actual monthly R_d during training for model 1 and model 2 using LSTM based approach



(a) (b)

Fig 4.24 (a) and (b) Comparison of monthly mean of daily observations predicted and actual R_d during testing for model 1 and model 2 using LSTM based approach

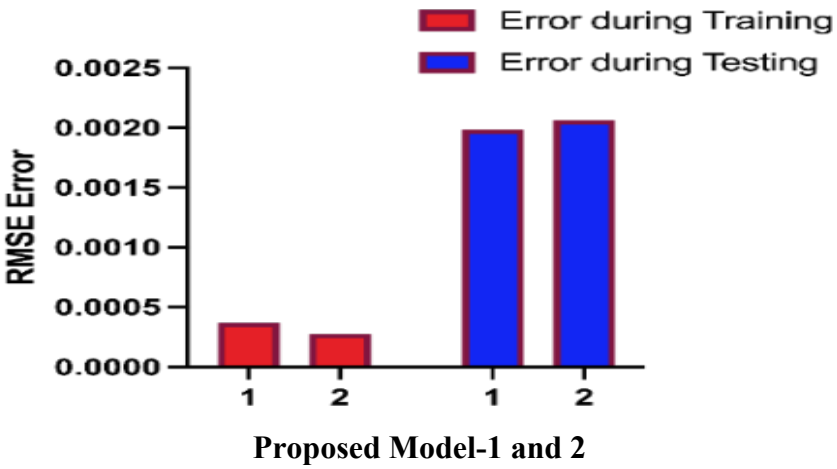


Fig 4.25. RMSE for proposed models 1 and 2 of LSTM based approach

As inferred from figure 4.25, both the proposed LSTM based R_d models behave competitively during training and testing. Nevertheless, the behaviour of the model during testing or validation lies essential for its deployment. The error generated during testing for the proposed model -1 is least than model-2 which shall be attributed to the inclusion of DEL_T which implicitly accounts for hot-spot-induced failure modes, making it more reflective of real-world PV module degradation scenarios.

The proposed sequential LSTM layered network namely Model-1 and Model-1 is scaled to test its ability to forecast day ahead failure-mode-based power degradation

rate (R_d), with change in time-horizon of the considered input attributes. Input of previous day time stamps of attributes as per table 4.6 are provided to Model-1 to forecasts R_d for time 't'. For instance, the forecast of R_d at 372nd day (non-rained period) as illustrated in Figure 4.28, is estimated employing the attributes governing Model-1 and Model -2 of 371st time stamps. As seen in figure 4.26, the proposed models are scalable for forecasting R_d, indicating their ability to predict beyond the training dataset for Bengaluru. This capability highlights the practical expansive utility of the proposed approach in forecasting failure-mode-based power degradation rate without the need for resource-intensive accelerated aging tests or extended real-time monitoring setups. Thus, the model offers a cost-effective and scalable solution for predictive maintenance of solar PV modules in field conditions.

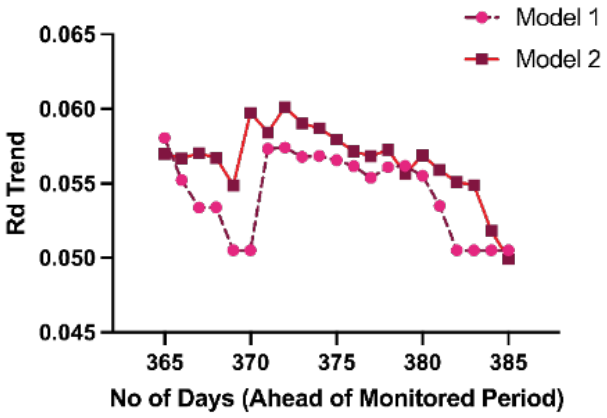


Fig.4.26. Forecasted day -ahead degradation rate for non-trained duration with model 1 and model 2 (LSTM)

When evaluating the reliability and statistical significance of the input coefficients in a predictive model, t-statistics (t-stat) and standard error (SE) are key indicators. The standard error quantifies the variability or uncertainty of the estimated coefficient, while the t-statistic assesses the meaningful contribution of attributes to the model (i.e whether the coefficient is significantly different from zero). Standard

Error (SE) of coefficient $\hat{\beta}_j$ is computed using equation (4.26) and t-statistic for coefficient $\hat{\beta}_j$ is estimated using equation 18 (Kaplan.et.al.,2025)

$$SE(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 \cdot (X^T X)^{-1}_{jj}} \quad \text{Eq. (4.26)}$$

where $\hat{\sigma}^2$ is the estimated variance of the error term, and $(X^T X)^{-1}_{jj}$ is the j^{th} diagonal element of the inverse of the design matrix.

$$t_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad \text{Eq. (4.27)}$$

In this study, Model 1 shows a standard error of 0.010, whereas Model 2 demonstrates an improved estimate with a lower standard error of 0.003, as visualized in figure 4.27 (a) and 4.27 (b), respectively. A lower standard error in Model 2 indicates increased precision in the coefficient estimates, implying a more stable and reliable model. Additionally, the presence of negative t-statistics in figure 4.24 (b) points to possible non-linearity or inverse influence of certain input variables in Model 2, suggesting complex interactions between predictors and the response variable. Standard Error (SE) of coefficient $\hat{\beta}_j$ is given in equation (4.28) and t-statistic for coefficient $\hat{\beta}_j$ is given in equation (4.29) (Kaplan.et.al.,2025)

$$SE(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2 \cdot (X^T X)^{-1}_{jj}} \quad \text{Eq. (4.28)}$$

where $\hat{\sigma}^2$ is the estimated variance of the error term, and $(X^T X)^{-1}_{jj}$ is the j^{th} diagonal element of the inverse of the design matrix.

$$t_j = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad \text{Eq. (4.29)}$$

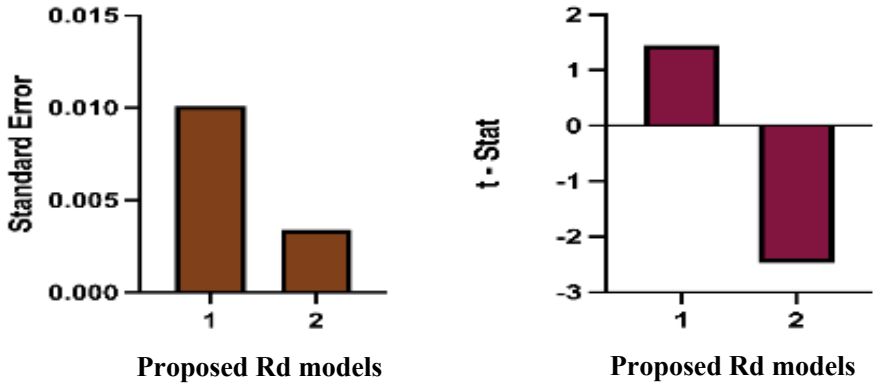


Fig.4.27 (a) Standard Error and (b) t-stat for Proposed LSTM based Model-1 & 2

4.5.2. Model Uncertainty Assessment for proposed R_d

To ensure the robustness and generalizability of the proposed R_d model uncertainty was estimated using bootstrap resampling considering model-1 and evaluated through the Root Mean Squared Error (RMSE) distribution given in equation (4.30). Bootstrap sampling is a well-established resampling technique increasingly used in machine learning and environmental modelling to quantify uncertainty in predictions without assuming underlying data distributions (Hyndman & Athanasopoulos, 2018; Vandal *et al*., 2017).

In this study, 1000 bootstrap iterations were performed on the original actual and the predicted daily R_d value. For each iteration, a random sample was drawn with replacement, and RMSE was calculated to generate a distribution of performance scores. From this distribution, a 95% confidence interval for RMSE was computed, providing an estimate of the model's variability and prediction stability under sampling variation as seen in figure 4.28.

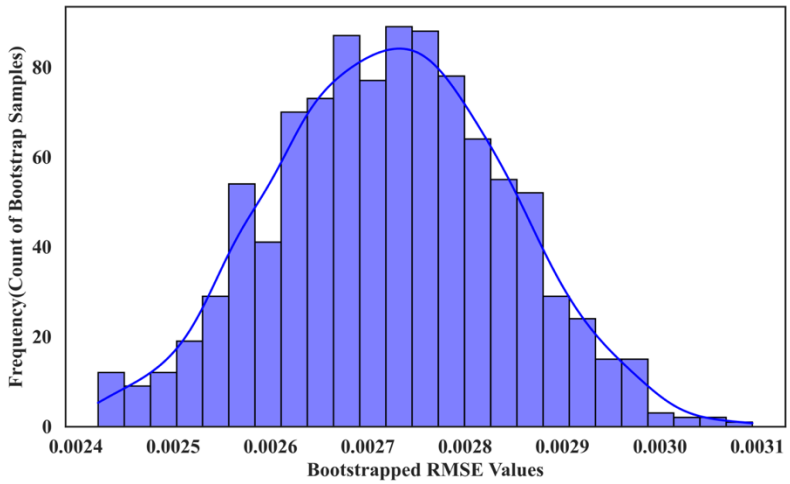


Fig 4.28. Histogram of Bootstrap-Derived RMSE Values for Model 1 for Uncertainty Quantification

$$\text{RMSE}_b = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (y_i - \hat{y}_i)^2} \qquad \text{Eq. (4.30)}$$

where:

- y_i is the actual R_d value for sample i=119,
- $\hat{y}_i$ is the predicted R_d value for sample i,
- n is the number of samples in each resampled set,
- and b denotes the b_{th} bootstrap iteration.

From b=1000 iterations, a distribution of RMSE values is obtained. The mean RMSE gives the central tendency of prediction error, while the confidence interval (CI) is computed as in equation (4.31) (Chen.et.al.,2025)

$$\text{CI}_{95\%} = [\text{RMSE}_{2.5\%}, \text{RMSE}_{97.5\%}] \qquad \text{Eq.(4.31)}$$

This confidence interval given in equation (4.31) quantifies the uncertainty in the model's generalization capability.

The overall model uncertainty $\sigma_{\text{uncertainity}}$ given in equation (4.32) can also be described as the standard deviation of the RMSEs across bootstrap samples:

$$\sigma_{\text{uncertainty}} = \sqrt{\frac{1}{B-1}\sum_{b=1}^{B}\left(\text{RMSE}_b - \overline{\text{RMSE}}\right)^2} \quad \text{Eq.(4.32)}$$

The narrow spread and low standard deviation of the bootstrap RMSE values as seen in figure 4.23 suggested that the R_d model maintained consistent performance across variations in the input data, reinforcing its reliability and robustness in real-world deployment scenarios. Bootstrap-based uncertainty estimation is particularly suited for small- to medium-sized datasets, such as in photovoltaic and environmental studies, and is considered a best practice for performance validation in recent literature (Zhao *et al*., 2019; Chakraborty *et al*., 2021).

4.5.3 Predictive assessment of Proposed FFBP based R_d Model

FFBP model as described in section 4.4.2 with model expressed as in equation (4.25) is employed for daily prediction of R_d with considered attributes in accordance with Model-1 and Model-2. A comparative performance of the predicted R_d value against the actual experimental R_d is illustrated in figure 4.29. As evident from the plot, both FFBP Model-1 and FFBP Model-2 closely follow the experimental R_d trend, with FFBP Model-2 demonstrating slightly improved alignment.

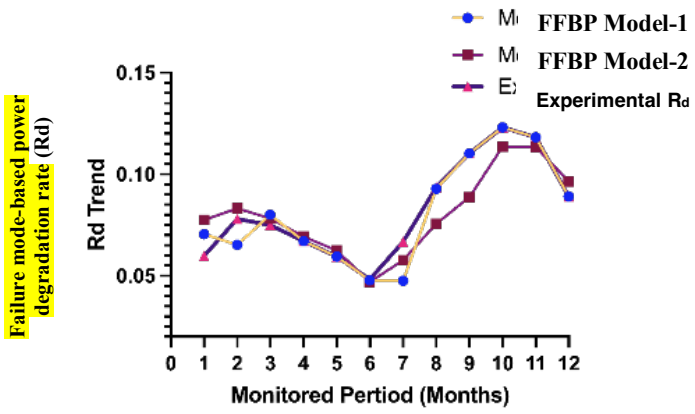


Fig 4.29. Predicted and Actual monthly mean of daily observations of R_d as obtained from FFBP model -1 and 2

The predictive accuracy of the models is quantified using the Root Mean Square Error (RMSE), which was found to be 0.001833 for model 1and 0.001305 for model 2.

4.5.4 Performance Comparison of the proposed R_d models for Bengaluru

Though all the proposed models of R_d behave appropriately and lie applicable for a predictor, an error performance radar chart as in figure 4.30 provides a comparative performance evaluation among the four different degradation prediction models, two based on Long Short-Term Memory (LSTM) networks and two based on Feedforward Backpropagation (FFBP) neural networks—specifically applied to the Bengaluru plant location. The chart presents three key error metrics: Mean Bias Error (MBE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE), all normalized for visualization.

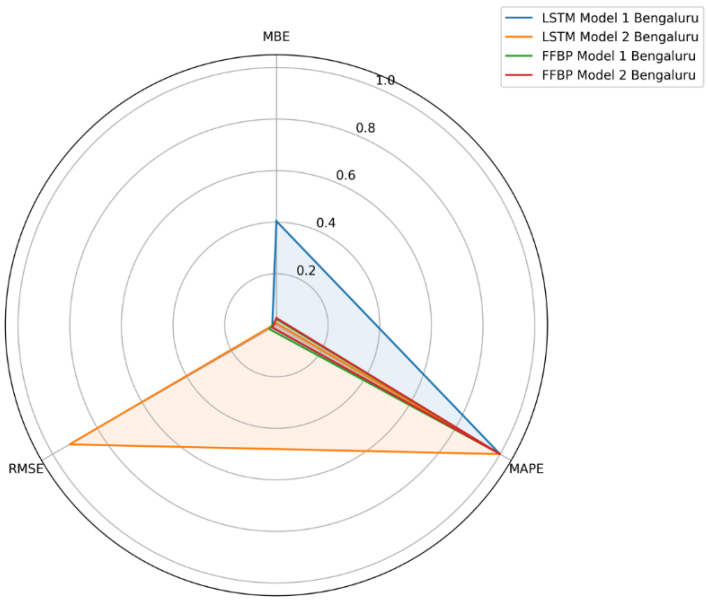


Fig.4.30. Error Performance Radar for the proposed R_d models for Bengaluru

From the plot, it is evident that the FFBP Model-1 demonstrates the most compact and centralized shape, indicating superior performance with the lowest values of

MBE, RMSE, and MAPE among all models. LSTM Model-1 also shows considerable improvement over LSTM Model-2, particularly with a drastic reduction in RMSE and MBE of -0.0132 and 0.0006, suggesting that an increase in the number of input parameters (e.g., inclusion of DEL_T and Rs) enhances model accuracy. In contrast, LSTM Model-2 exhibits the largest spread, especially in MBE and MAPE, highlighting greater prediction errors and bias.

The radar plot effectively illustrates how Model-1 for both LSTM and FFBP leads to a significant enhancement in prediction accuracy. It supports the conclusion that FFBP Model-1 is the most reliable for the Bengaluru site, followed by LSTM Model-1, making a strong case for both architecture selection and critical input prioritization in power degradation modelling.

4.5.5 Validation of the proposed R_d models for Telangana

To validate the proposed methodology for predicting the failure-mode-based power degradation rate (R_d), a 235 kWp on-grid rooftop Solar PV system located in Maheshwaram, Telangana (17°13′16.40″N, 78°41′69.92″E) was selected. The plant, commissioned in March 2022, has an observed annual degradation rate of approximately 8.5%, inferred from monitored performance indicators.

The predictive models Model-1 and Model-2, developed using the Long Short-Term Memory (LSTM) architecture, were applied to this validation site. On estimation, both models closely follow the actual degradation trend, with Model-2 demonstrating superior accuracy, as indicated by its Mean Absolute Percentage Error (MAPE) of 0.0277, compared to 0.05758 for Model-1.

Furthermore, the Root Mean Square Error (RMSE) values for Model-1 and Model -2 during testing phases were low amounting to 0.0019 and 0.0020 respectively and lie within acceptable limits indicating close conformity to ideal performance (i.e.,

minimal deviation). This reinforces the generalizability and reliability of the proposed models for field-level estimation of power loss due to degradation-induced failure mechanisms.

The Error Performance Radar chart in figure 4.31 complements the validation analysis of the proposed R_d models (Model-1 and Model-2) for predicting daily failure-mode-based power degradation rate (R_d) at the Telangana site. As evident from the chart, LSTM Model-1 exhibits lower values across all three-error metrics MBE, RMSE, and MAPE compared to Model-2, particularly in terms of MAPE. This observation is supported by the quantitative results, where Model-1 achieved a MAPE of 0.0277, significantly lower than the 0.05758 reported for Model-2. The compact area enclosed by Model-1 on the radar plot further suggests better generalization and minimal deviation from actual experimental values.

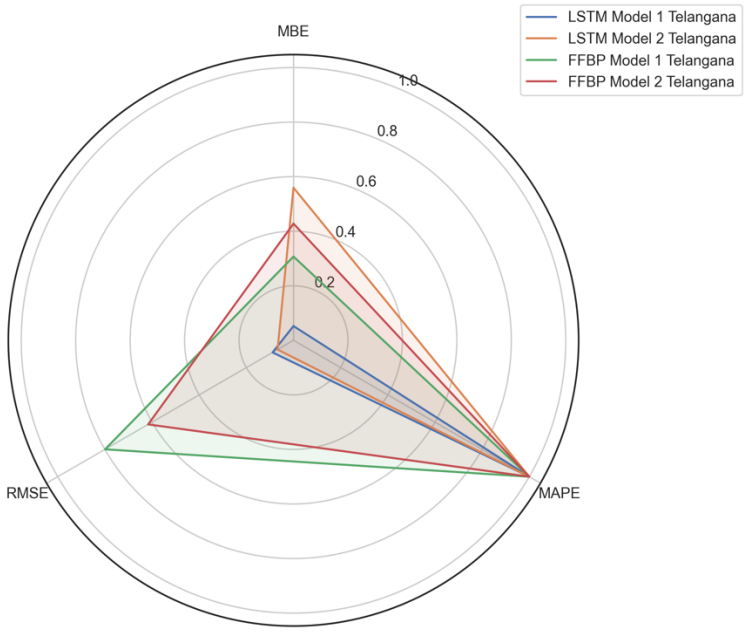


Fig 4.31 Error Performance Radar for the proposed R models for Telangana

In addition to the LSTM models, the radar chart also compares Feedforward Backpropagation (FFBP) Model-1 and Model-2. FFBP Model-1, while more stable than FFBP Model-2 in terms of MAPE, still shows relatively higher RMSE and MBE values, indicating limited learning capacity in capturing complex time-series patterns.

In conclusion, among all evaluated architectures, LSTM Model-1 demonstrates superior accuracy and robustness, owing to its inherent capability to model sequential temporal dependencies, making it a reliable estimator for R_d in PV system diagnostics.

4.6 Summary

This chapter presented a comprehensive approach to estimating and modelling the failure-mode-based power degradation rate (R_d) encompassing failure modes in photovoltaic systems through both experimental techniques and predictive modelling. Beginning with on-field diagnostics and data processing, it led to the development and evaluation of machine learning models for R_d estimation. The key takeaways from this chapter are summarized below:

- A grid-connected rooftop PV plant at Bengaluru, Karnataka was selected for the experimental study, with a focus on long-term degradation behaviour.
- On-field diagnostic techniques such as visual inspection, thermal imaging, and I-V curve tracing were employed to assess failure modes contributing to R_d.
- The Single Diode Model (SDM) under STC conditions was used to validate the experimental R_d for various degradation scenarios.
- Exploratory data analysis (EDA) and domain knowledge were utilized to identify critical input parameters affecting R_d, leading to data cleaning, normalization, and outlier handling.
- Two types of machine learning architectures—LSTM and Feed Forward Back Propagation (FFBP)—were developed for R_d prediction using different input configurations.
- Model uncertainty was quantified using bootstrap-based RMSE confidence intervals to evaluate robustness.

- Bland-Altman and Radar variations were used to compare the accuracy of models across different numbers of input parameters, justifying that higher input dimensionality (e.g., six inputs) improved prediction precision.
- Validation with unseen field data from a different geographic location (Telangana) confirmed the transferability of the proposed models.
- Model 1 incorporating DEL_T and Rs, consistently outperformed Model 2 across various metrics such as RMSE and MAPE.

Having successfully established a predictive framework for power degradation using experimental and machine learning techniques, the next step involves analyzing other attributes which affect the performance ratio of the PV plant. This includes thermodynamic losses in PV systems. Chapter 5 extends this understanding by incorporating exergy efficiency modeling, providing a deeper perspective into the usable exergy loss mechanisms in solar PV modules across diverse locations.

Chapter 5 : Exergy Efficiency Modelling for Solar PV System

Overview:

This chapter presents a comprehensive framework for analyzing and predicting the exergy efficiency of photovoltaic (PV) systems using hybrid deep learning approaches. It begins by establishing the relevance of thermodynamic based exergy analysis in capturing the true performance of PV systems under real-world conditions. The methodology includes preprocessing of environmental and operational data through cleaning, interpolation, and normalization to ensure consistency and accuracy. Various dual-layer and multi-layer hybrid models combining architectures like CNN, GRU, Transformer, and BiLSTM are developed and trained to estimate exergy efficiency. Model performance is assessed across three diverse climatic regions Khopoli, Telangana, and Bengaluru using statistical error metrics such as RMSE, MBE, and MAPE. The findings highlight the models' ability to generalize across sites while effectively capturing complex nonlinear relationships influencing PV performance. The proposed model approach for exergy efficiency depends on easily accessible or pre-estimable inputs for prediction.

Outline of Chapter ***Page***

5.1. Exergy Fundamentals, Relevance and Objective

Solar energy, being abundant, renewable, and globally accessible, holds immense potential for meeting future energy demands while reducing reliance on fossil fuels. While the performance of distributed photovoltaic (PV) systems is commonly evaluated using metrics such as energy yield, capacity utilization factor, and performance ratio as outlined in IEC 61724 standards (Fuster.*et.al*,2022), these indicators are primarily based on electrical output alone. They do not provide insights into the internal energy conversion mechanisms within the PV module. In practice, the photovoltaic process inherently results in thermal losses, leading to a simultaneous reduction in power output. Therefore, the actual on-field efficiency of a solar PV system is more accurately reflected through exergy analysis, which accounts for energy quality and losses during power conversion (Sun.*et.al*,2023). Although energy efficiency considers only the quantity of energy output, it does not account for irreversibility or the quality of energy lost as heat, which is particularly relevant in photovoltaic systems subjected to varying thermal and environmental conditions (Elavarasan.*et.al*,2022). In contrast, exergy efficiency incorporates the second law of thermodynamics, allowing to assess not just the quantity of energy output, but also quantifies the actually available amount for useful work. This makes exergy efficiency a more comprehensive and physically meaningful performance indicator for real-world PV system analysis.

Several studies have also employed simulation tools focused on electrical performance metrics (Panicker.*et.al.,*2023) (Gao.*et.al*.,2024). However, upon reviewing the investigations enlisted in Chapter 2, two key limitations emerge. Firstly, most existing studies rely on thermodynamic models involving assumptions for calculating thermal

exergy, exergy input, and exergy output parameters that are not directly measurable and typically depend on empirical estimations of heat losses and heat transfer coefficients. These coefficients, in turn, require real-time monitoring of output power, limiting the practicality of predictive deployment.

Secondly, there is a notable lack of research focused on utility-scale, grid-connected solar PV systems for evaluating exergy efficiency. Comprehensive exergy assessments of megawatt-scale PV installations using actual field data and their influence on power degradation and system performance remain underexplored.

To address these gaps, this study advocates for the application of machine learning (ML) and artificial intelligence (AI) techniques to estimate exergy efficiency. These data-driven approach for modelling exergy efficiency offer several advantages like enabling prior prediction, reduce dependency on complex thermodynamic assumptions, minimize manual intervention, and utilize easily measurable input parameters. Such a digital-domain approach not only enhances estimation accuracy but also supports system optimization and operational decision-making, thereby serving as a valuable tool for stakeholders in the solar PV industry.

The primary objective of this study is to develop a robust, digitally deployable model for predicting the exergy efficiency of large-scale photovoltaic (PV) power plants. The model leverages machine learning techniques and is designed to depend on readily and economically available input parameters, making it practically feasible and anticipative in nature. Daily exergy efficiency prediction is formulated based on historical operational variables for a two-year period from a 191.9 kWp grid-connected PV system. Multiple architectures, including five distinct dual-layer and multi-layer hybrid models, are

explored for model development. The best-performing model is identified and proposed for accurate short-term daily prediction of exergy efficiency in solar PV systems.

5.1.1 Exergy Analysis for the considered PV plant under study

Exergy analysis quantifies the maximum useful work obtainable from a system as it interacts with its environment. In PV systems, exergy input is defined by the solar radiation, and exergy output is the electrical energy generated. The relevance of exergy analysis lies in its ability to highlight areas of irreversibility and guide improvements in system design and operation. This method is particularly crucial for assessing degradation effects under varying climatic and operational conditions. The overall thermodynamic energy balance of the solar PV can be expressed as in equation (5.1) (Sudhakar and Srivastav.,2014) and for the steady-state flow process over a time-period.

$$Ex_{in} = Ex_{out} + Ex_{loss} + irreversibility \quad \text{Eq.(5.1)}$$

The process of energy conversion in Solar PV results in electrical output and thermal heat loss. The part of energy generated and utilised by the load is the electrical energy which is considered for computation of energy efficiency. The irreversibility or thermal energy loss dissipated to the ambient, is considered for computation of exergy efficiency and is modelled in equations (5.2) to (5.11). This destruction in energy conversion is called exergy loss. As inferred from the traditional thermodynamic based approach for computation of exergy efficiency, the convective heat losses are modelled empirically and found dependent on model constants.

$$Ex_{loss} = Ex_{in} - Ex_{out} \quad \text{Eq.(5.2)}$$

$$Ex_{in} = A_{PV} * G * \left[1 - \frac{4}{3}\left(\frac{T_a}{T_s}\right) + \frac{1}{3}\left(\frac{T_a}{T_s}\right)^4\right] \quad \text{Eq.(5.3)}$$

The assumptions for radiative Exergy Modelling includes the following aspects:-

- **Spectral assumption:**

The incoming solar radiation is treated as blackbody radiation emitted from the sun. Spectral variations are not explicitly resolved; instead, measured broadband global irradiance is used. This assumption is consistent with standard PV exergy formulations and enables practical implementation using field-measured data.

- **Reference sun temperature:**

The apparent sun temperature (Ts) is assumed to be 6000 K, which is a commonly accepted reference value in solar exergy studies and is adopted in Eqs. (5.2)–(5.4). This assumption allows consistent estimation of the maximum theoretical work potential of incident solar radiation.

- **Radiative Exergy Input: Assumptions and Uncertainty**

In the present work, the radiative exergy input is estimated using the Petela-type formulation as adopted by Aziz *et al.* (2022), which is widely employed in solar photovoltaic exergy analysis. The formulation assumes that the incident solar radiation originates from a blackbody source at an apparent sun temperature and is expressed in terms of measurable broadband global irradiance.

- **Uncertainty associated with radiative exergy estimation:**

The primary sources of uncertainty in radiative exergy estimation arise from (i) the blackbody radiation assumption, (ii) the fixed sun temperature assumption, and (iii) measurement uncertainty in irradiance and ambient temperature. These uncertainties systematically affect all data points and therefore do not influence the relative temporal

trends, seasonal variations, or comparative performance analysis conducted in this study. Since the same formulation and assumptions are uniformly applied across all datasets, their impact on degradation analysis, exergy efficiency trends, and AI/ML model training remains consistent.

- **Relevance to AI/ML framework:**

The thermodynamic exergy model is used to generate the target variable for supervised learning. The proposed AI/ML framework does not re-derive radiative exergy but learns the system behaviour from measured operational inputs. Consequently, minor uncertainties associated with radiative exergy assumptions do not compromise the predictive validity of the developed models.

$$Ex_{th} = (h_{con} + h_{rad}) * A_{PV}(T_a - T_{PV}) * \left(1 - \frac{T_a}{T_m}\right) \quad \text{Eq.(5.4)}$$

$$Q = (h_{con} + h_{rad}) * A_{PV} * (T_{\mathrm{m}} - T_{\mathrm{a}}) \quad \text{Eq.(5.5)}$$

$$h_{\mathrm{con}} = 2.8 + 3V_w \quad \text{Eq.(5.6)}$$

$$h_{\mathrm{rad}} = \varepsilon\sigma\left(T_{\mathrm{sky}} + T_{\mathrm{m}}\right)\left(T_{\mathrm{sky}}^2 + T_{\mathrm{m}}^2\right) \quad \text{Eq.(5.7)}$$

$$T_{\mathrm{sky}} = T_a - 6 \quad \text{Eq.(5.8)}$$

$$Ex_{ele} = V_{OC} * I_{sc} * FF \quad \text{Eq.(5.9)}$$

$$Ex_{out} = Ex_{ele} - Ex_{ther} \quad \text{Eq.(5.10)}$$

$$Exe_{\eta} = \frac{Exe_{out}}{Exe_{in}} \quad \text{Eq.(5.11)}$$

Where,

Ex_{in} is Exergy input – maximum work potential of incident solar radiation; A_{PV} is Area of the PV module (m²); G is Global solar irradiance (W/m²); T_a is Ambient temperature (K); T_s is Apparent sun temperature, typically assumed as 6000 K; T_m is module

temperature (K); h_{con} is convective heat transfer coefficient (W/m²·K); h_{rad} is radiative heat transfer coefficient (W/m²·K); ε is emissivity of the PV surface (0.85); σ is Stefan–Boltzmann constant (5.67×10^{-8} W/m²·K⁴); T_{sky} is Effective sky temperature (K); V_{OC} is Open circuit voltage (V); I_{sc} is Short circuit current (A); FF is Fill factor (0.85); Ex_{out} is Total exergy output; Exe_{η} is the Exergy efficiency of the solar PV system

In this research, the fill factor (FF) is assumed constant at 0.85, corresponding to the rated module value under standard test conditions. Although FF varies with temperature and irradiance, its use as a constant acts as a uniform scaling factor on the electrical exergy term.As seen in Eq(5.9) V_{OC} and I_{SC} are directly and significantly influenced by operating temperature and irradiance. Thus, the primary environmental dependencies are inherently captured through the measured variations in V_{OC} and I_{SC} .So, investigation concerning exergy analysis typically assume fill factor (FF) as a constant at 0.85, corresponding to the rated module fill factor under Standard Test Conditions (STC). This assumption is also aligned with PV performance and exergy studies for instance as in (Sudhakar, K., & Srivastava, T. 2013)

The convective heat-transfer coefficient used in this study is estimated using the empirical correlation $h_{\mathrm{con}} = 2.8 + 3V_w$, which represents combined natural and forced convection for outdoor flat-plate surfaces. This formulation has been widely adopted in photovoltaic thermal and exergy analyses and is suitable for PV modules exposed to ambient wind conditions. The radiative heat-transfer coefficient is computed using the Stefan–Boltzmann relation for surface–sky radiative exchange. These correlations and constants are adopted from established PV thermal modeling literature (Skoplaki &

Palyvos,2009)(Dubey et al.,2013).Although empirical in nature, their consistent application across all datasets ensures that relative trends and comparative analyses remain reliable.

5.1.2 Assessment of exergy efficiency for 191.9 kWp PV plant through thermodynamic based approach

The thermodynamic assessment of exergy efficiency as expressed in equations (5.2) to (5.11) are employed for computing the daily variation of exergy efficiency of the considered 191.9 kWp PV plant. The computation is performed for a duration from January 2021 to December 2022 whose variation is as shown in figure 5.1. The variation of the target or the response lies essential for development of dual-layer and multi-layer hybrid models.

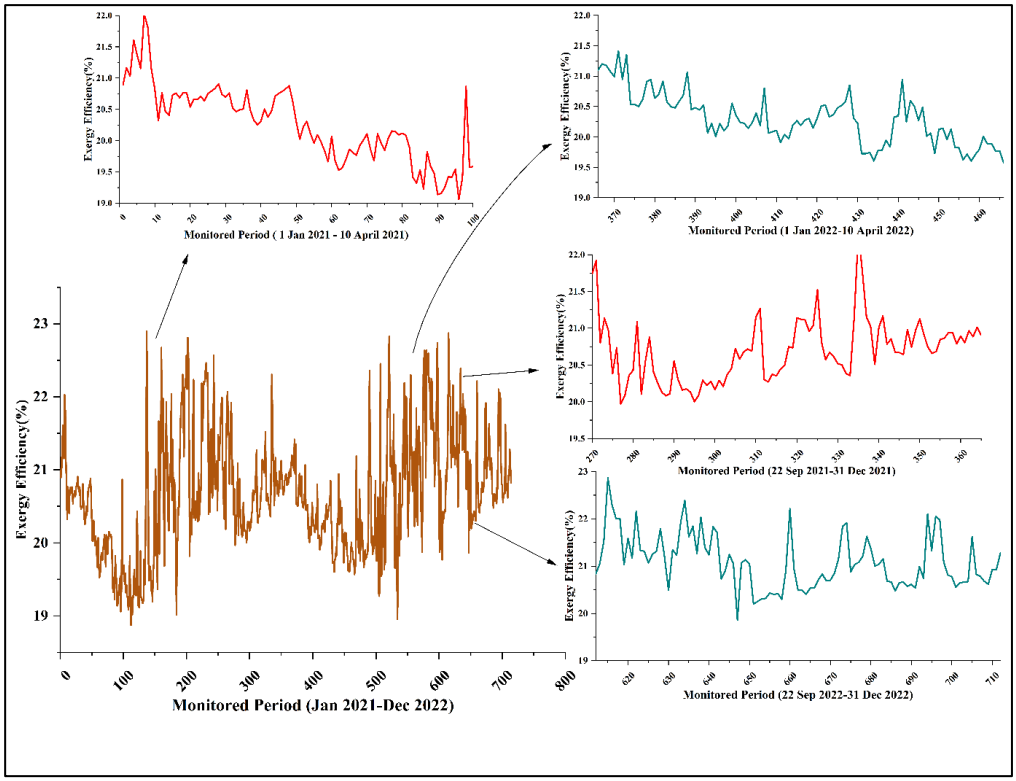


Fig. 5.1 Daily Exergy Efficiency variation from January 2021 to December 2022 for 191.9 kWp PV plant

As depicted in figure 5.1, the actual daily exergy efficiency of the system exhibits a seasonal trend, with discernible variation between the two monitored years. Year 1 (January -December 2021) the initial portion of the time series exhibits comparatively higher exergy efficiency value of 20.35%, whereas Year 2 (January -December 2022) displays greater variability with intermittent peaks observed around the mid-year period of 2022. Elevated exergy efficiency correlates with enhanced energy output and a reduction in the cost per unit of generated energy, with annual average exergy efficiency of 21.32% thereby positioning it as a critical indicator in PV performance assessment.

A notable period of high exergy efficiency occurs between 1st January and 10th January 2021, where the system achieves values up to 22.3%, primarily due to lower module temperatures averaging 44°C and moderately high wind speeds of approximately 4.7 m/s, both of which reduce thermal losses. Conversely, between 1st April and 10th April 2021, exergy efficiency dips due to elevated module temperatures averaging 64.9°C and lower wind speeds around 3.2 m/s, conditions which elevate irreversibility in the system. Throughout the entire assessment window, the exergy efficiency varies between a minimum of 18.87% and a maximum of 22.89%, highlighting the thermodynamic impact of environmental and operational parameters on PV performance.

5.1.3 Quasi-Steady-State Assumption and Effect of Temporal Variations

The exergy and performance analyses presented in this study are formulated under a quasi-steady-state assumption, which is commonly adopted in photovoltaic performance and exergy studies based on field measurements. Although solar irradiance, ambient temperature, and module temperature vary continuously under real operating conditions,

the analysis is performed using time-averaged measured data at hourly resolution and further aggregated to daily and monthly scales.

Over these short sampling intervals, the dynamic response of the PV system is assumed to be sufficiently slow, allowing steady-state thermodynamic relations to remain valid. Temporal variations in irradiance and temperature are inherently reflected in the time-series data and directly influence the calculated exergy input, thermal losses, and performance indicators at each time step.

When results are aggregated over daily or monthly periods, short-term transients such as passing clouds or rapid wind-induced cooling effects are naturally smoothed, enabling meaningful assessment of seasonal trends, degradation behaviour, and long-term performance. While instantaneous transient effects are not explicitly modeled, the consistent application of the same analytical framework across all datasets ensures that relative comparisons and trend analyses remain robust.

5.2. Selection of Critical Factors for modelling exergy efficiency

Accurate identification of influential parameters and understanding their impact on the modelled response are essential steps in the development of robust predictive model. This step becomes particularly critical when dealing with thermodynamic parameters such as exergy efficiency, which is sensitive to dynamic climatic and operational conditions. A comprehensive examination was conducted using Pareto analysis and a correlation heatmap as seen in figure 5.2 and figure 5.3 to evaluate the statistical relationship and sensitivity of the identified and proposed input parameters namely irradiance, wind speed, module temperature, and temperature correction factor "A" on the daily predicted exergy

efficiency of the rooftop solar PV system. The proposed inputs possessing recognizable impact on the outcome i.e. exergy efficiency, are daily variation of irradiance (kWh/m^2), module temperature (°C), wind speed (m/s) and temperature correction ratio represented as 'A'. 'A' represents the second order power of ratio of the ambient temperature to the sun's temperature as represented in equation (5.12).

Where $A = \left(\frac{T_a}{T_S}\right)^2$ Eq.(5.12)

Among these inputs, module temperature in the present work, is not measured for every individual PV module across the plant, which would be impractical for large-scale systems as described in chapter 3, section 3.1.3. Instead, temperature measurements are obtained from representative modules or array-level sensors, consistent with standard PV monitoring practices and recommendations in IEC 61724 and related PV performance literature (Madeti and Singh,2017).

The selected modules are located in centrally positioned, unshaded strings, ensuring that the recorded temperature captures the dominant thermal behaviour influencing electrical performance. This approach is widely adopted in field-scale PV performance and degradation studies, where representative measurements are used to characterize array-level behaviour rather than assuming identical temperatures across all modules. The work does not assume that all modules operate at the same temperature. Instead, it assumes that representative module or array-level temperature measurements as cited above under section (a) capture the aggregate thermal trends governing plant-level power output. Spatial temperature variations within the plant are implicitly reflected in the electrical

output (DC power and energy yield), which is measured at the inverter and plant level varying with time-horizon under consideration

The correlation heatmap as in figure 5.2 reveals a strong inverse relationship between exergy efficiency and irradiance, with a Pearson correlation coefficient of -0.94. This suggests that as irradiance increases, exergy efficiency declines significantly. The underlying reason for this trend is attributed to the rise in module temperature that accompanies higher irradiance. In the present work, module temperature is not measured for every individual PV module across the plant, which would be impractical for large-scale systems. Instead, temperature measurements are obtained from representative modules or array-level sensors, consistent with standard PV monitoring practices and recommendations in IEC 61724 and related PV performance literature (Madeti and Singh,2017).

The selected modules are located in centrally positioned, unshaded strings, ensuring that the recorded temperature captures the dominant thermal behaviour influencing electrical performance. This approach is widely adopted in field-scale PV performance and degradation studies, where representative measurements are used to characterize array-level behaviour rather than assuming identical temperatures across all modules.

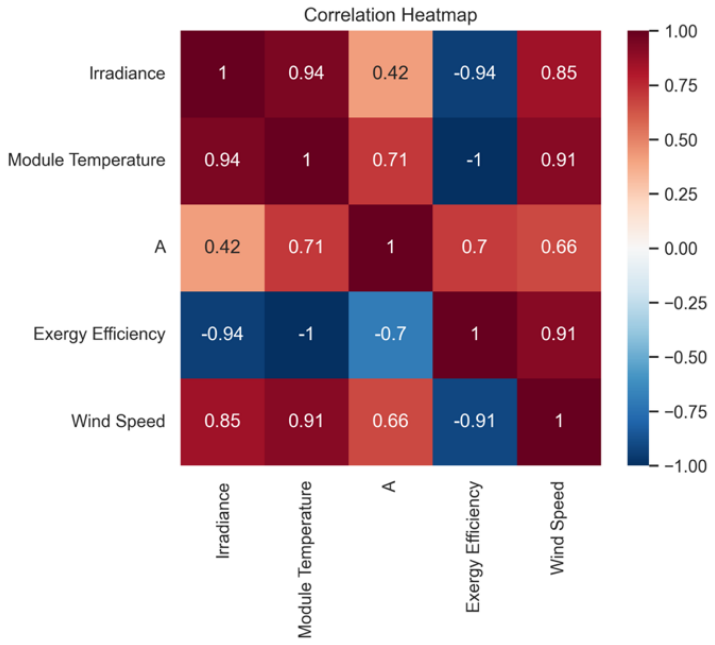


Fig 5.2. Correlation heatmap validating the choice of model input parameters towards model development

Elevated module temperatures enhance thermal losses through radiative and convective mechanisms, thereby reducing the amount of useful energy that can be converted into electrical output. This phenomenon is further validated by the perfect negative correlation of -1.00 observed between module temperature and exergy efficiency, underscoring that temperature elevation is a primary contributor to exergy loss.

The correlation analysis reveals a very strong negative correlation between module temperature and exergy efficiency. This behavior is primarily physics-driven, as module temperature explicitly appears in the analytical formulation of exergy efficiency. An increase in module temperature leads to higher thermal exergy destruction and temperature-induced voltage drop, thereby reducing the electrical exergy output and overall exergy efficiency.

Although the dependence is strong, it is noted that in practical field conditions the relationship is influenced by measurement uncertainty, transient effects, and the interaction of other variables such as irradiance and wind speed. Therefore, the observed high correlation should be interpreted as a model- and physics-constrained dependency rather than an independent empirical linear relationship

Similarly, factor "A", a temperature correction ratio included in the model to account for thermal behaviour under variable conditions, shows a moderate negative correlation of -0.70 with exergy efficiency. Though its impact on exergy efficiency is not as influential as irradiance or module temperature, its variation introduces measurable effects on the energy conversion characteristics of the system. Temperature co-efficient also serves as a parameter representing the spectral and temperature differential between the ambient and module, helping to approximate convective loss contributing to exergy thermal loss. Furthermore, wind speed was observed to have a strong negative correlation of -0.91 with exergy efficiency. Although wind is generally considered beneficial for heat dissipation, in this case, higher wind speeds might disturb the local thermal equilibrium, causing fluctuating module temperatures that negatively influence stable energy conversion and impact exergy efficiency.

The Pareto chart in figure 5.3 quantitatively ranks the significance of these variables. Module temperature emerges as the most critical factor, followed by irradiance, wind speed, and the correction factor "A". The statistical evaluation as computed from ANOVA as in table 5.1 confirms that the temperature correction factor of 0.7, introduced in this study plays a significant role and cannot be omitted from the model. Its presence enhances

the accuracy of exergy efficiency estimation by compensating for temperature-induced nonlinearities in the system's thermal response.

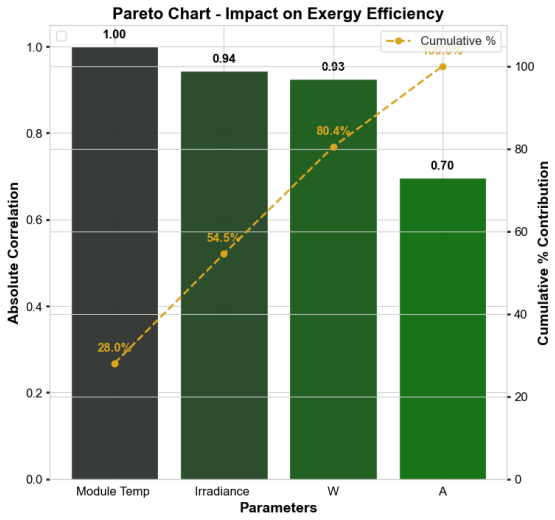


Fig 5.3. Pareto chart for analysing the significant impact of the considered attributes over exergy efficiency model

In summary, the combination of correlation and Pareto analysis not only provides a data-driven justification for the selection of inputs but also highlights the thermodynamic interplay between considered variables and system performance.

Table 5.1. Analysis of Variance (ANOVA) for selected Inputs

Variable	**Sum of Squares**	**degrees of freedom (df)**	**F-value**	**P-value**
Irradiance	5.09×10^{-08}	1	17.96	0.0017
A	3.82×10^{-08}	1	13.48	0.0003
Module Temperature	1.79×10^{-08}	1	6.32	0.0123
Wind speed [1]	5.35×10^{-08}	1	7.31	0.0071

Such insights are invaluable for the development of predictive models with greater generalizability and reliability across diverse climatic zones.

The significant impact of considered attributes for modelling exergy efficiency is also justified employing ANOVA analysis. The ANOVA results indicate that irradiance ($p = 0.0017$), variable A ($p = 0.0003$), module temperature ($p = 0.0123$), and wind speed ($p = 0.0071$) all have statistically significant effects on exergy efficiency ($p < 0.05$). These results confirm that all four variables make a substantial contribution to the variation observed in system performance. Notably, variable A shows a strong statistical relationship, suggesting its importance in the model. These findings reinforce the need to include irradiance, temperature, and wind-related factors when predicting exergy efficiency and demonstrate the robustness of the model in capturing key environmental drivers.

5.2.1 Data Cleaning and Normalization

This two-step approach, imputation followed by min-max normalization ensures that the dataset is both complete and scaled appropriately for downstream machine learning algorithms. By addressing gaps and outliers in the data, the continuity and integrity of time-dependent relationships are preserved. Simultaneously, normalization places all input variables on a uniform scale, which helps prevent bias during training and improves model convergence. These preprocessing steps are crucial for sensitive applications like exergy efficiency prediction, where even small numerical inconsistencies can lead to significant errors in learning on-field behaviour of physical systems (i.e) Solar PV. The cleaned and normalized dataset is then ready for further refinement through noise handling and temporal alignment as discussed in the following sections.

a. Imputation (for time-series continuity)

Real-world solar datasets often suffer from missing data due to sensor failures, communication dropouts, or shading effects.

- Missing entries were interpolated using time-based linear interpolation, maintaining the temporal continuity of the data. If a value x_t at time t is missing and values at surrounding timestamps t_1<t<$t_2$are available, the missing value is estimated as in equation (5.13) (Demirhan and Renwick,2018).

  $$x_t = x_{t_1+} \frac{(t-t_1)}{(t_2-t_1)} . (x_{t_2} - x_{t_1}) \quad \text{Eq.(5.13)}$$

- To eliminate high-frequency noise (e.g., cloud flicker in irradiance), a rolling median filter with window size w (24 ×60 minute) considering daily prediction is applied. For each timestamp t the filtered value is given in equation (5.14).

  $$\tilde{x_t} = median(x_{t-w/2}, \dots, x_t, \dots., x_{t+w/2}) \quad \text{Eq.(5.14)}$$

b. Min-Max Normalization

For each input variable (e.g., wind speed, temperature correction factor, irradiance and module temperature) its corresponding normalized value is calculated as in equation (5.15) (Gao.*et.al*,2025)

$$x_{norm} = \frac{x-x_{min}}{x_{max}-x_{min}} \quad \text{Eq.(5.15)}$$

Where:

- x is the original value
- x_{min}, $x_{max}$are the minimum and maximum values in the dataset for that feature
- $x_{norm}$is the normalized output
-

5.2.2 Scaling and Temporal Alignment

The training dataset included observations with different sampling frequencies (e.g., module temperature monitored and recorded for every 1 minute, wind speed monitored and recorded for every 10 min) where temporal alignment was essential.

- All data streams encompassing the attributes were resampled to a common 10-minute interval using time-based averaging as in equation 5.16.

$$\overline{x_d} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{Eq.(5.16)}$$

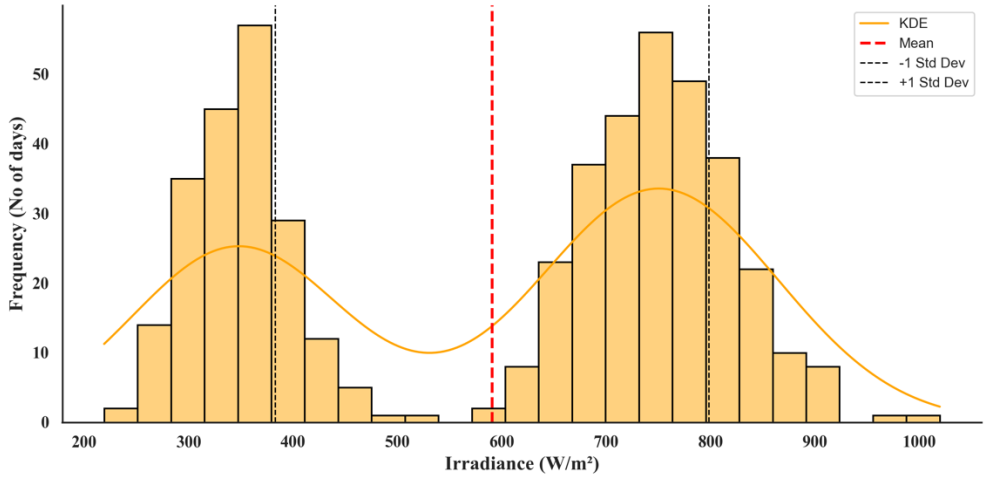
where , $x_i$are values of input for day d , x_d is daily-averaged value of variable x and n is number of measurements per day

- All inputs were then scaled using standardization (z-score normalization) prior to development of predictor, especially important for exergy efficiency estimation which is sensitive to small thermodynamic fluctuations. (Arora.*et.al*.2020),

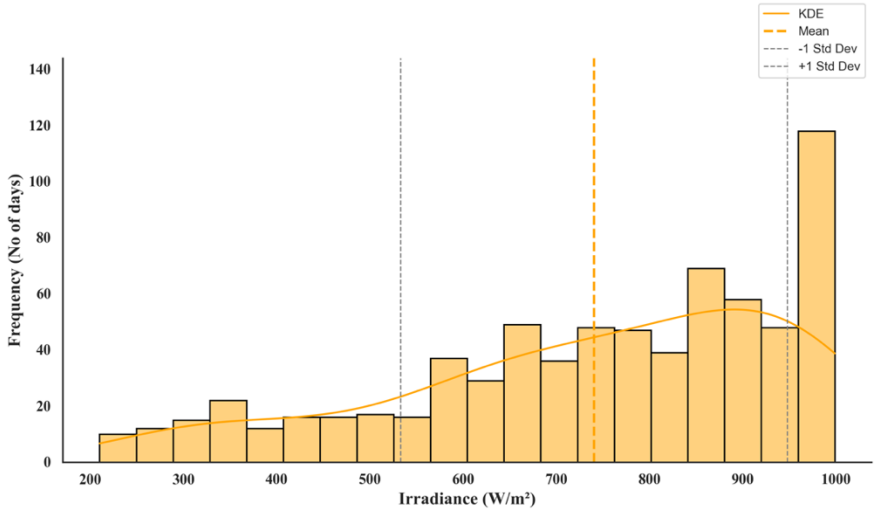
$$z_d = \frac{x_d - \mu_x}{\sigma_x} \qquad \text{Eq.(5.17)}$$

x_d is daily value of the variable,μ_x is the mean of the variable across all days, σ_x standard deviation of the variable and z_d is the normalized (standardized) value as seen in equation 5.17. Following the data cleaning, imputation, normalization, and alignment steps described in Sections 5.2.1 to 5.2.3, a set of diagnostic plots was generated to validate the effectiveness of preprocessing. These visualizations as illustrated in figures 5.4 and 5.5 indicate the spread and distribution of temporal behaviour of key input parameters such as irradiance, module temperature, wind speed and A implying confirmation of the transformation and readiness of the dataset for further analysis. By comparing the raw and cleaned datasets, the day wise frequency distribution as indicated in figures 5.4 and

5.5 demonstrate the manner in which the outliers, missing values, and irregularities have been addressed. This ensures that the trained data is suitable for time-series modelling. The uncleaned data as in figure 5.4 (a) exhibits a multimodal distribution with scattered peaks and visible skewness, suggesting potential sensor noise, abrupt weather fluctuations, or irregular measurement intervals. Additionally, the kernel density curve displays erratic undulations, which could interfere with time-series model learning. In contrast after data preprocessing, the cleaned histogram as in figure 5.4 (b) appears more uniformly distributed with reduced variance and a well-defined central tendency. Outliers below 300 W/m² and above 1000 W/m² were removed, leading to a smoother probability density and a shift in the mean closer to the operational irradiance range .The original raw time-series dataset contained intermittent missing samples and irregular measurements due to sensor communication gaps and short periods of signal loss. These missing values created fragmented representations of clear-sky midday irradiance conditions in the unprocessed dataset as seen in Figure 5.4 (a).



(a)

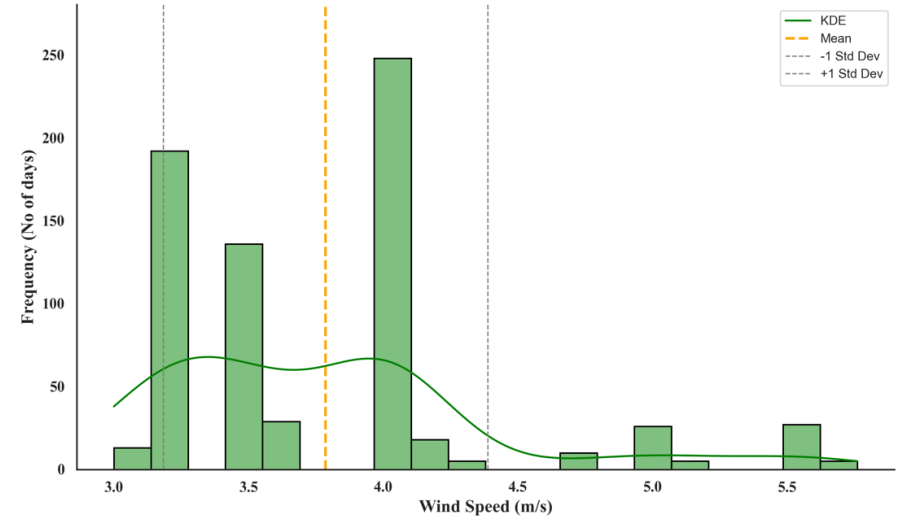


(b)

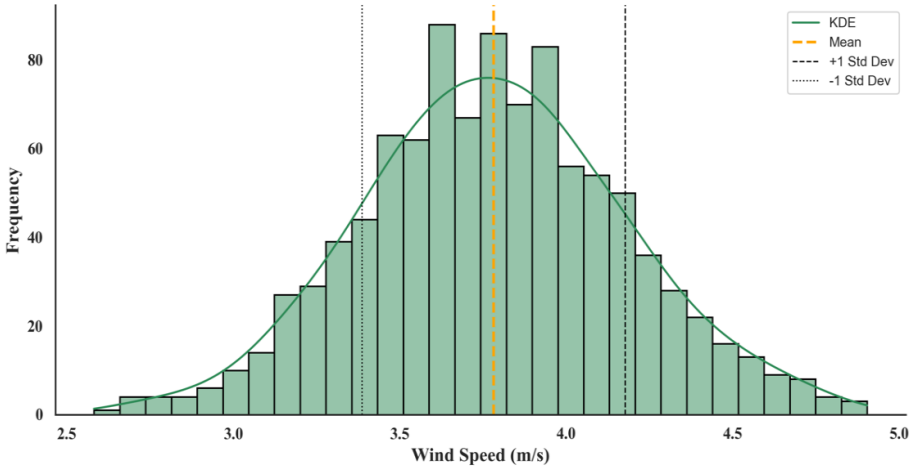
Fig.5.4 (a) Uncleaned and (b) cleaned distribution of Input irradiance / attribute at over the monitored annual period (Jan 2021– Dec 2022)

During preprocessing, missing samples were reconstructed using time-based linear interpolation to maintain continuity in the irradiance time series. In addition, a rolling median filter was applied to remove high-frequency noise caused by transient atmospheric disturbances and sensor fluctuations. These preprocessing steps as

mentioned above reconstructed physically consistent clear-sky periods that were previously fragmented in the raw dataset.



(a)



(b)

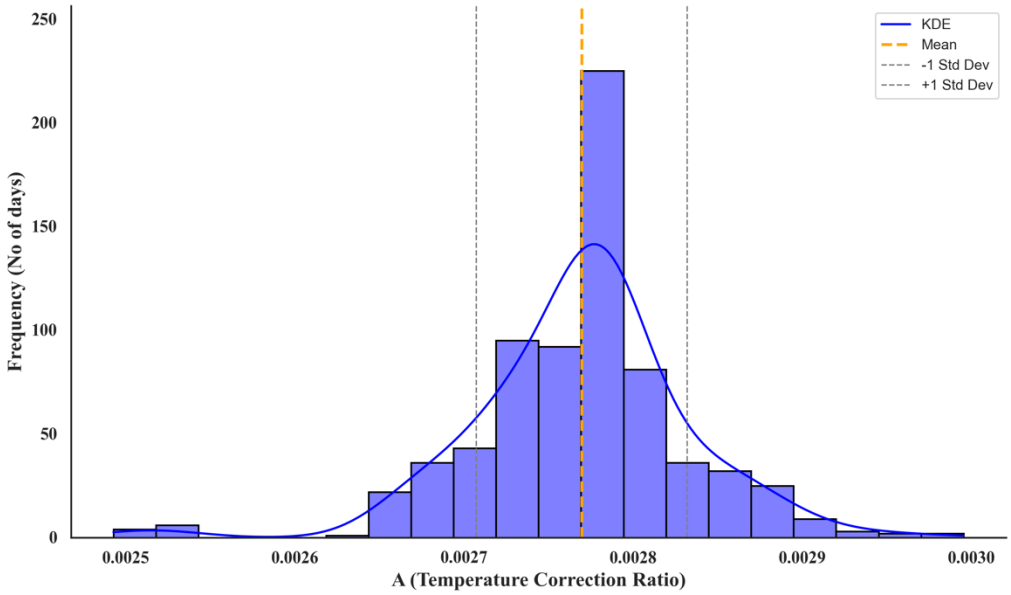
Fig 5.5. (a) Uncleaned and (b) cleaned distribution of Input wind-speed attribute over the monitored annual period (Jan 2021– Dec 2022)

Figure 5.5 (a) representing the uncleaned data exhibits a skewed and irregular distribution with visible peaks and inconsistencies, suggesting the presence of sensor noise, inconsistent recording intervals, or environmental interference. These distortions can mislead model learning, especially for time-series forecasting tasks.

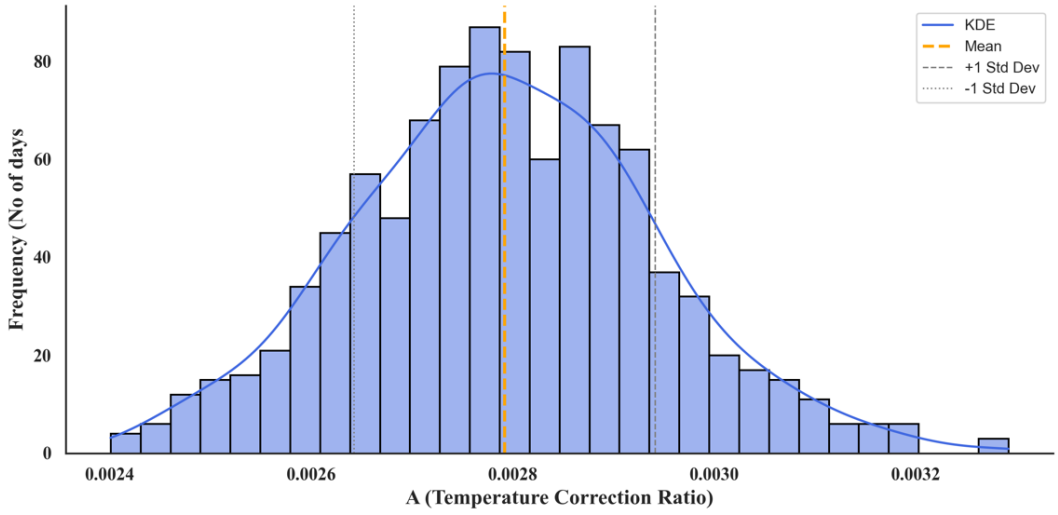
After cleaning, the histogram as demonstrated in figure 5.5 (b) possess a more symmetrical and Gaussian-like distribution centred around the mean (~3.8 m/s), with tighter bounds within ±1 standard deviation. This transformation not only improves the interpretability of the wind data but also enhances model stability by ensuring uniformity in statistical properties.

The uncleaned histogram of the temperature correction ratio (A) as seen in figure 5.6 (a) exhibits moderate dispersion, with a visibly skewed tail toward lower values. The spread outside the ±1 standard deviation range indicates potential noise or anomalous entries that could adversely impact model accuracy and exergy efficiency estimation.

Post-cleaning, the histogram as in figure 5.6 (b) becomes more symmetrical and exhibits a sharp peak near the mean, closely resembling a normal distribution. The narrower spread and increased concentration within ±1 standard deviation boundaries suggest that data smoothing and outlier removal have effectively improved distribution quality. In the raw module temperature data as seen in figure 5.7 (a) , noise is evident through discontinuities and dual peaks reflecting possible erroneous readings or influence of shading and rapid atmospheric changes. Upon cleaning as seen in figure 5.7 (b), the temperature distribution becomes unimodal and smoother. The cleaned data better approximates the expected diurnal temperature behaviour of PV modules, which is essential for degradation analysis and exergy-based energy flow modelling.



(a)

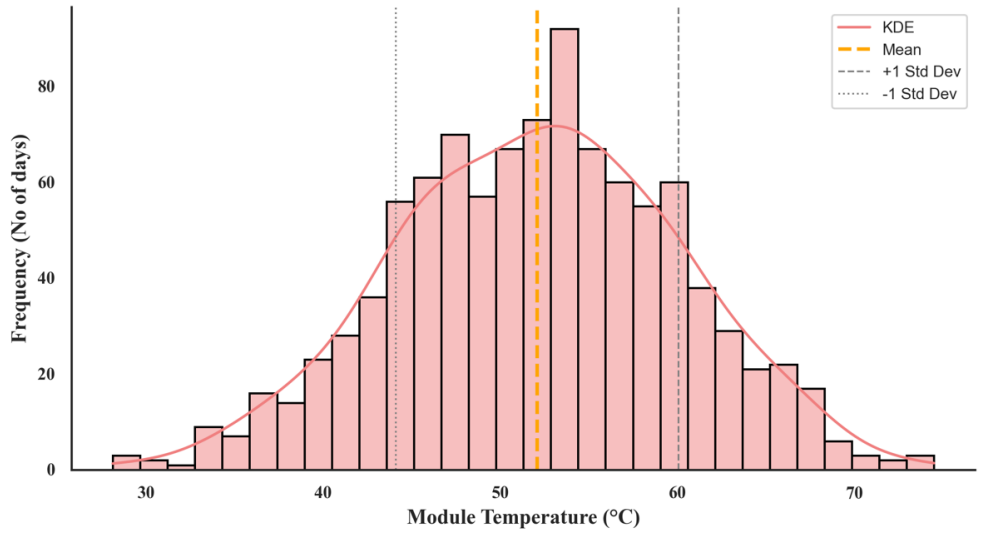


(b)

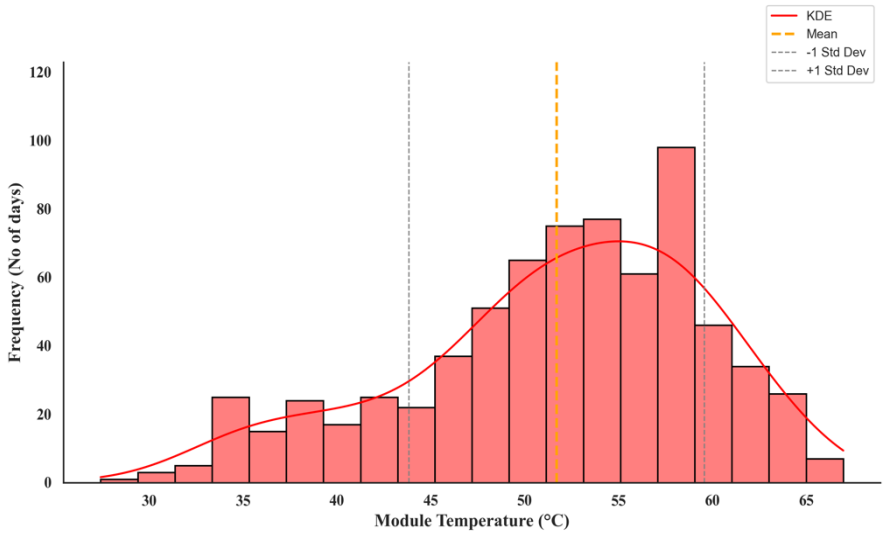
Fig.5.6 (a) Uncleaned and (b) cleaned variation of Factor "A' at the Khopoli PV installation site over the monitored period (Jan 2021– Dec 2022)

5.3 Proposed dual-layer hybrid models for prediction of exergy efficiency

To enable prediction of exergy efficiency and enhance its accuracy under varying environmental and operational conditions, five distinct dual-layer hybrid models are developed.



(a)



(b)

Fig 5.7 (a) Uncleaned and (b) cleaned variation of Module Temperature at the Khopoli PV installation site over the monitored period (Jan 2021– Dec 2022)

The dual-layer hybrid models are co-integrated with a Bidirectional Long Short-Term Memory (BiLSTM) architecture. These hybrid models incorporate BiLSTM alongside CNN, GRU, Attention, Autoencoder, and Transformer mechanisms, as illustrated in figure 5.8 .

A Bidirectional LSTM (BiLSTM) processes the input sequence in both forward and backward temporal directions. This configuration enables the model to capture both past and future context dependencies within the data, which is crucial for accurate time-series prediction. The combined hidden state of the BiLSTM is represented mathematically as inn equation (5.18).

$$H_t^{BiLSTM} = H_t^{\rightarrow} \oplus H_t^{\leftarrow} \quad \text{Eq.(5.18)}$$

Where, $H_t^{\rightarrow}$ represents the hidden state from the forward LSTM (processing from past to future), $H_t^{\leftarrow}$ denotes the hidden state from the backward LSTM (processing from future to past) and $\oplus$ indicates concatenation of the two vectors.

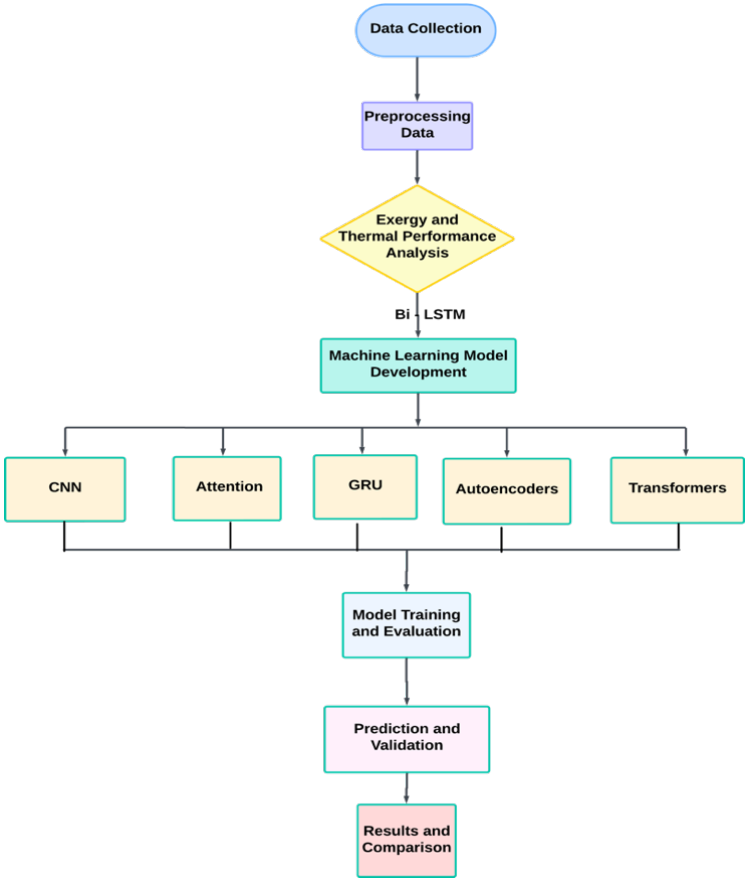


Fig.5.8. Block Diagram for the proposed dual layer hybrid models for prediction of daily exergy efficiency

Each hybrid configuration integrated to a BiLSTM is designed to capture complementary temporal, spatial, and contextual features that affect PV exergy performance. Model tuning was carried out using the Bayesian Optimization strategy in Keras Tuner, which supports parallel execution, early stopping, and visualization of convergence. This setup enabled efficient training across a wide hyperparameter space. The following subsections elaborate on the specific structure, motivation, and contribution of each dual-layer hybrid model.

5.3.1 Proposed Dual Layer Model Configurations

a) CNN-BiLSTM dual-layer hybrid Model

The CNN-BiLSTM hybrid architecture integrates the spatial feature extraction capability of Convolutional Neural Networks (CNN) with the sequential modeling strength of Bidirectional LSTM (BiLSTM) (Rai.*et.al.,*2021). The model utilizes 64 convolutional filters with a kernel size of 3, followed by 128 BiLSTM units. A dropout rate of 0.2 is applied to mitigate overfitting, and a pooling layer with a window size of 2 is used for dimensionality reduction. The ReLU activation function is adopted to introduce non-linearity, and the model is optimized using a learning rate of 0.001.

The overall transformation of the CNN-BiLSTM model is mathematically represented in equation (5.19).

$$C_B = f_{Dense}\left(f_B\left(f_{dr}\left(f_p\left(f_{co}(X)\right)\right)\right)\right) \quad \text{Eq.(5.19)}$$

Here, f represents the transformations at each layer including Convolutional (co), Pooling (p), Dropout (dr), BiLSTM (B), and Dense layers.

The individual layer operations are defined as in equations (5.20) to (5.22).

- Convolutional Layer:

$$f_{CO}(X)=\sigma(W_{conv} * X + b_{conv}) \quad \text{Eq.(5.20)}$$

where W_{conv} and b_{conv} are the convolutional kernel weights and biases, and σ denotes the ReLU activation function.

- Pooling Layer:

$$f_p\left(f_{co}(X)\right)=max_{j\,\epsilon\,k}\left(f_p\left(f_{co}(X)\right)\right) \quad \text{Eq.(5.21)}$$

where max pooling is applied over each region k.

- Dropout Layer:

$f_{dr}\left(f_p\left(f_{co}(X)\right)\right)=f_p\left(f_{co}(X)\right)\odot \mathrm{M}$ Eq.(5.22)

where M is a binary mask drawn from a Bernoulli distribution with dropout probability p=0.2, and ⊙ denotes element-wise multiplication.

b) GRU-BiLSTM dual- layer hybrid model

The GRU-BiLSTM hybrid model synergistically combines the temporal efficiency of Gated Recurrent Units (GRUs) with the bidirectional sequence-learning capability of BiLSTM (Michael.*et.al*.,2024). This model is structured with 64 GRU units and 128 BiLSTM units, configured with a learning rate of 0.0005, a dropout rate of 0.2, and a batch size of 64. A lookback window of 3-time steps is employed to capture temporal dependencies in input data sequences. The functional representation of the GRU-BiLSTM model is described by equation (5.23).

$$G_B = f_{GRU+BiLSTM}(X_{t-timestep}) \quad \text{Eq.(5.23)}$$

The GRU mechanism at time step t is expressed in equation.5.24 involving computations as expressed in equation (5.25) and equation (5.27).

$$f_{GRU}(X_t) = \{1 - z_t) \odot f_{GRU}(X_{t-1}) + z_t \odot f^{-}_{GRU}(X_t) \quad \text{Eq.(5.24)}$$

where,

candidate Activation $f^{-}_{GRU}(X_t)$=tanh $(W_G X_t + U_G\ (r_t \odot\ h_{t-1}) + b_h)$ Eq.(5.25)

Update Gate z_t=σ $(W_z X_t + U_z f_{GRU}(X_{t-1}) + b_z)$ Eq.(5.26)

Reset Gate r_t = σ $(W_r X_t + U_r f_{GRU}(X_{t-1}) + b_r)$ Eq.(5.27)

X_t denotes the input feature vector at time step t, and σ represents the sigmoid activation function. The symbol ⊙ indicates element-wise multiplication. However, its capability to capture spatial correlations within input features remains limited compared to models that integrate convolutional or attention mechanisms.

c) Attention-BiLSTM dual layer hybrid model

The Attention-BiLSTM hybrid model enhances sequential learning by integrating an attention mechanism that dynamically assigns varying importance to different time steps in the input sequence. This design allows the model to selectively focus on temporally significant features during prediction, improving its ability to capture contextual dependencies . (Hu.*et.el.,*2025)

The architecture comprises 64 encoder units and 128 BiLSTM-based decoder units, using ReLU as the activation function. A lookback window of 2 is employed, along with a learning rate of 0.0002, to fine-tune the learning process.

The attention-based formulation of the model is mathematically represented as in equation (5.28).

$A_B = \sum_{t=1}^{T} a_t h_t$ Eq.(5.28)

Where, h_t denotes the hidden state output from the BiLSTM decoder at time step t; $a_t$is the attention weight corresponding to each ; T is the total number of time steps.

This weighted aggregation of hidden states enables the model to emphasize time steps that contribute most significantly to the prediction task.

d) Autoencoder-BiLSTM dual layer hybrid model

The Autoencoder-BiLSTM hybrid model is designed to learn a compressed representation of the input sequence X through an encoder-decoder architecture) (Demirhan and Renwick.,2018) Initially, the encoder employs 64 BiLSTM units to transform the input into a lower-dimensional latent representation. This is followed by a decoder with 128 BiLSTM units, which attempts to reconstruct the original input from the encoded features.The reconstruction process is mathematically expressed in equation (5.29) and (5.30)

$$AU_B = Dense(H_f) \quad \text{Eq.(5.29)}$$

Where,$H_F = BiLSTM_{predictor}X$ Eq.(5.30)

Where:

- AU_B is the final reconstructed output; $H_F$is the encoded feature representation obtained from the BiLSTM-based encoder.

The model uses tanh as the activation function, a batch size of 16, and mean squared error (MSE) as the reconstruction loss to measure the discrepancy between the original and reconstructed inputs. This structure allows the model to learn underlying patterns in the time-series data, making it suitable for applications involving anomaly detection or dimensionality reduction before prediction.

e) Transformer-BiLSTM dual layer hybrid model

The Transformer-BiLSTM hybrid model integrates the sequence learning strength of BiLSTM with the contextual awareness of Transformer attention mechanisms (Yin.*et.al*.,2024). The model first processes the input feature vector X through 128 BiLSTM units, producing a contextual representation H_f. This output is then passed

through a Transformer block comprising 8 attention heads, which enhances the model's ability to capture long-range dependencies within the sequence.as seen in equation (5.31) to (5.34).

$Tr_B = W_O y + b_o$ Eq.(5.31)

$y = Dense(H_T)$ Eq.(5.32)

$H_T = Multihead\ Attention\ (H_f)$ Eq.(5.33)

$H_f = BiLSTM(X)$ Eq.(5.34)

The model uses a batch size of 32, an attention dropout rate of 0.1, and a learning rate of 0.0001. This architecture leverages the strengths of both sequential memory and global attention, making it well-suited for tasks requiring nuanced temporal and contextual understanding in PV performance data.

The architecture, training parameters, and mathematical formulations of all the proposed dual layer hybrid model variants for daily prediction of exergy efficiency are summarized in table 5.2.

5.4 Proposed multi-layer hybrid model for prediction of exergy efficiency

The proposed multi-layer hybrid model is designed to operate on digital platforms for the accurate prediction of daily exergy efficiency in photovoltaic systems. The step-by-step formulation methodology is elaborated in figure 5.9. This model strategically integrates multiple neural network architectures to enhance predictive performance and generalization. Initially, the significant input features are passed through an Autoencoder, which learns a compressed representation of the data by eliminating redundancy and capturing essential patterns. The encoded representation is then propagated into a

Bidirectional LSTM (BiLSTM) network, which captures both past and future dependencies in the temporal data sequence.

Table 5.2. Summary of optimized parameter configurations for the proposed dual-layer architectures

Dual layer hybrid Models	**Parameters**	**Optimal Values**	**Mathematical Equation**
CNN +BiLSTM	Number of Filters	64	$C_B = f_{Dense}(f_B\left(f_{dr}\left(f_p(f_{co}(X))\right)\right))$ Where f is transformation applied to respective layer of Dense,BiLSTM,Dropout,pooling, convulutional
	Kernel Size	3	
	LSTM Units	128	
	Learning Rate	0.001	
	Dropout	0.2	
	Batch Size	32	
	Lookback Window	6	
	Pooling Layer Size	2	
	Activation Function	ReLU	
GRU +BiLSTM	GRU Units	64	$G_B = f_{GRU+BiLSTM}(X_{t-time_{step}})$ Where X_t is feature vector at timestep t
	LSTM Units	128	
	Learning Rate	0.0005	
	Dropout	0.2	
	Batch Size	64	
	Lookback Window	6	
Attention+ BiLSTM	Encoder Units	64	$A_B = \sum_{t=1}^{T} a_t h_t$ Where h_t is the hidden state output from the BiLSTM layer at time step t. a_t is the attention weight assigned to each time step t
	Decoder Units	128	
	Attention Activation	ReLU	
	Lookback Window	6	

	Learning Rate	0.0002	T is total number of timestep in input
	Batch Size	50	
Autoencoder +BiLSTM	Encoder Units	64	$AU_B = Dense(H_f)$ Where $H_F = BiLSTM_{predictor}X$ X is reconstructed Input sequence
	Decoder Units	128	
	Activation Function	Tanh	
	Reconstruction Loss	Mean Squared Error	
	Batch Size	16	
Transformer+ BiLSTM	LSTM Units	128	Tr_B=$W_O y$+b_o Where $W_O$is Weights of the output layer b_o is Bias term y is predicted exergy efficiency
	Transformer Heads	8	
	Attention Dropout	0.1	
	Learning Rate	0.0001	
	Batch Size	32	

To further refine the predictions, the output from the BiLSTM is processed through a Transformer-based attention mechanism, enabling the model to identify and weigh complex interdependencies between the input features and thermal exergy efficiency. This attention-enhanced architecture strengthens the model's ability to handle dynamic and nonlinear behaviours inherent in solar energy systems.

Subsequently, a Global Average Pooling (GAP) layer is employed to reduce the risk of overfitting by minimizing the number of trainable parameters. Finally, the output is passed through a Dense layer, which yields the predicted exergy efficiency. This proposed multi-layered structure harnesses the combined strengths of Autoencoders (feature extraction), BiLSTM (temporal modelling), and Transformers (contextual attention),

making it a robust solution for high-fidelity exergy efficiency forecasting. To ensure reproducibility of the LSTM and hybrid deep learning models, fixed random seeds were employed during model initialization and training. This practice controls stochastic variations arising from weight initialization and mini-batch selection, thereby enabling consistent and fair comparison across different model architectures. The use of controlled random seeds ensures that observed performance differences are attributable to model structure and learning capability rather than random initialization effects.

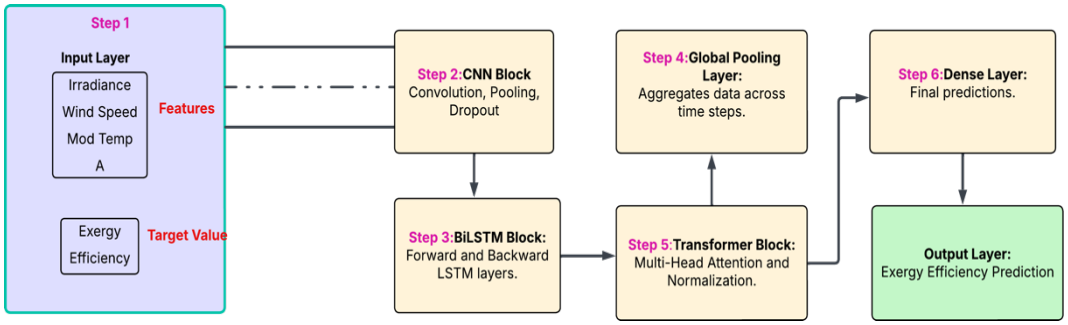


Fig.5.9. Proposed Multilayer Hybrid Model for prediction of exergy efficiency

5.4.1 Configuration of Model Layers in Multilayer Model Architecture

Step 1 Structured Dataset Creation

In the initial step, a dataset is constructed by structuring the relevant features irradiance, module temperature, wind speed, and factor A along with the target variable, exergy efficiency, over preceding time steps. This results in a time-series formatted data frame represented by equation (5.35).

$$X \in L^{T*F} \qquad \text{Eq.(5.35)}$$

Where:

- **L** is the number of samples
- **T** is the number of time steps (in this study, **T = 6**)
- **F** is the number of features (4 in this case)

This configuration allows the model to capture short-term temporal dependencies essential for accurate forecasting.

Step 2 Convolutional Layer & Feature Extraction

To extract spatial and local temporal patterns from the input data, a 1D Convolutional Layer is employed. This layer applies filters over the time-series sequence, transforming the input data frame (X) into feature maps as seen in equation (5.36).

$$h_1[i,j]=f_{conv}(\sum_{k=1}^{3} W_{conv}[j,k].X[i,k]+b_{conv}[j]) \quad \text{Eq.(5.36)}$$

Where:

- $W_{conv} \in L^{F*K}$ is the learnable weight matrix for the convolution kernel, where K is the kernel size=3.
- $b_{conv} \in L^{K}$ is the learnable bias term for the convolutional layer as shown in Appendix 1.
- $f_{conv}(.)$ is the activation function applied after convolution (ReLU).
- $h_1 \in L^{T'*K}$ is the feature maps generated after convolution, where T′ is the reduced temporal size depending on padding and stride.
- i is the time step index.

A Max-Pooling layer is applied next to retain dominant features and reduce temporal dimensions as seen in equation (5.37).

$$h_2[i.j] = f_{pool}(h_1[i{:}\, i + P, j]) \quad \text{Eq.(5.37)}$$

Where, P is the pooling window size (P=2 for a MaxPooling of size 2),$f_{pool}(.)$ is the pooling operation for max pooling., $h_2 \in L^{T''*K}$,$T''=\left[T'/_P\right]$.To mitigate overfitting, a dropout layer with a probability p = 0.3 is used.

Step 3 Bidirectional LSTM Integration

Standard LSTM captures only past temporal dependencies. To overcome this limitation, a Bidirectional LSTM (BiLSTM) is used, which processes sequences both forward and backward. The BiLSTM is modelled as represented in equations (5.38) to (5.40).

$$h_3[i,j] = \begin{pmatrix} h_2[i,j] & with(1-p)\ probability \\ 0 & with\ p\ probability \end{pmatrix} \quad \text{Eq.(5.38)}$$

$$h_t^{bwd} = LSTM_{bwd}(h_{t+1}^{bwd}, x_t; W_{bwd}, b_{bwd}) \quad \text{Eq.(5.39)}$$

$$h_t^4 = \left[h_t^{for}, h_t^{bwd}\right] \quad \text{Eq.(5.40)}$$

Where, x_t is input at time step t.

Step 4 Transformer-Based Attention Layer

To further enhance the BiLSTM output, a Transformer Attention mechanism is applied. This allows the model to learn dynamic attention weights over different time steps as expressed in equation (5.41).

$$Attention(\boldsymbol{Q}, \boldsymbol{K}, \boldsymbol{V}) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad \text{Eq.(5.41)}$$

Where $Q\ is\ query$ representing current time step; $K\ is\ Key$ representing all time steps to compare the query with; $V\ is\ value$ against each key; $d_k\ is$ used for scaling to stabilize gradients. This mechanism allows the model to prioritize the most influential observations dynamically.

Step 5 Global Average Pooling

To reduce feature dimensionality and retain the most critical sequential information, Global Average Pooling (GAP) is used as seen in equation (5.42)

$$h_7 = \frac{1}{T}\sum_{t=1}^{T} h_6(t) \quad \text{Eq.(5.42)}$$

This operation aggregates the Transformer output h_6 over all time steps T = 6, generating a compact and robust feature vector h_7 for the final prediction layer.

Step 6 Dense Output Layer and Final Prediction

The final prediction of daily exergy efficiency is computed by applying a dense (fully connected) layer on the pooled output as expressed in equation (5.43).

$$Predicted\ Exergy\ Efficency\ (\hat{y}_{t+1}) = W.\frac{1}{6}\sum_{i=1}^{6} h_7^i + b \quad \text{Eq.(5.43)}$$

Where, h_7 is the output from Global Average Pooling layer ; W is weight matrix of dense /output layer ranging from -0.7 to 0.9; b is Bias vector of dense/ output layer ranging from -0.02 to 0.04

A complete pipeline of the proposed multi-layer hybrid model architecture as against the dual layer architecture for predicting daily exergy efficiency is represented in figure 5.10. While Steps 1 to 6 detailed the formulation of the core predictive model integrating convolutional, BiLSTM, attention, and dense layers, a unified operation of distinct dual-layer hybrid models namely CNN-BiLSTM, GRU-BiLSTM, Attention-BiLSTM, Autoencoder-BiLSTM, and Transformer-BiLSTM with Multi-layer hybrid model are evaluated within the same framework. These sub-models function in parallel with the main predictive pathway, providing comparative insights and allowing modular extensibility for performance benchmarking.

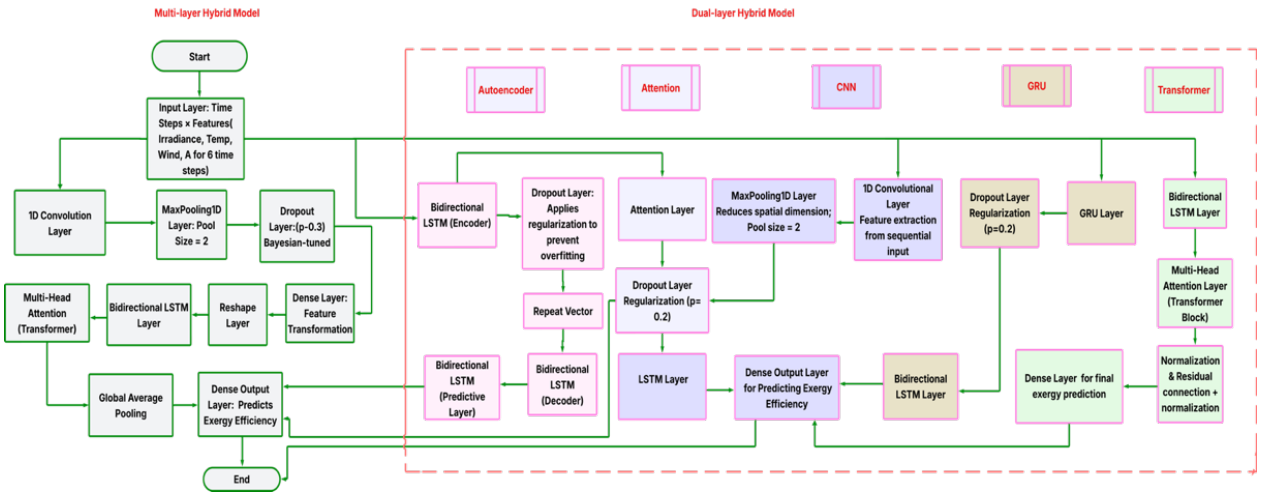


Fig.5.10. Architectural Flow of Multi-Layer and Dual-Layer Hybrid Models for Exergy Efficiency Prediction

5.4.2 Model Architecture Design and Hyperparameter Justification

1. Look-back window justification

The look-back window was selected based on the short-term temporal dependency of PV system performance, where irradiance, module temperature, wind speed, and degradation effects exhibit strong correlations over limited time horizons. Multiple window lengths were evaluated, and shorter windows (2–6 time steps) provided optimal prediction accuracy while avoiding overfitting and excessive model complexity.

2. Number of units (LSTM / BiLSTM / Encoder–Decoder)

The number of hidden units in LSTM, BiLSTM, encoder, and decoder layers was selected considering the dataset size, temporal resolution, and non-linear behavior of PV degradation and exergy efficiency. Values between 64 and 128 units were found sufficient to capture system dynamics without causing training instability or over-parameterization.

3. CNN filter and kernel size

For CNN-based hybrid models, a kernel size of 3 was adopted to capture local temporal variations in irradiance and temperature, which is consistent with standard time-series feature extraction practices. The number of filters was limited to 64 to balance feature richness and computational efficiency. A kernel size of 3 was chosen to capture short-term local temporal variations in PV operating parameters such as irradiance and exergy loss, which typically exhibit strong correlations over adjacent time steps. Using 64 filters provides sufficient feature diversity to model the nonlinear behavior of the PV system while maintaining computational efficiency and avoiding overfitting. This configuration offered the best trade-off between prediction accuracy and model complexity based on empirical evaluation.

4. Learning rate & batch size

Learning rates in the range of 0.0001–0.001 were selected after empirical evaluation to ensure stable convergence without gradient explosion or vanishing. Batch sizes were chosen based on model depth and memory constraints to ensure smooth training convergence. Different learning rate values were used to ensure stable and efficient convergence across models with varying depths and architectures. Lower learning rates were adopted for deeper or hybrid models to prevent unstable updates, while relatively higher rates were suitable for simpler architectures. The selected values were based on empirical tuning to minimize validation error and avoid gradient instability.

5. Dropout and regularization

Dropout regularization with a rate of 0.2 was employed across all hybrid models to mitigate overfitting, especially given the supervised nature of the learning task and the presence of seasonal clustering in the data.

6. Final selection criterion

The final hyperparameter configurations reported in Table 5.2 correspond to the models that achieved the lowest validation RMSE and MAPE while maintaining stable convergence across multiple training runs. The selected parameters represent a trade-off between predictive accuracy, generalization capability, and computational efficiency.

5.5 Training, Validation, and Robustness Framework

To ensure reliable performance evaluation and reproducibility of the proposed multi-layer hybrid deep learning model, a structured training–validation framework was adopted. The framework addresses prevention of temporal leakage, reproducibility, hyperparameter selection, and fair comparison across model architectures.

5.5.1 Prevention of temporal leakage

In the developed PRLSTM framework, temporal leakage was prevented by adopting a strict chronological train–test split. Specifically, 80% of the earliest time-series observations of Rd, Exergy loss, Irradiance, and PR were used for training, while the remaining 20% of subsequent unseen data were reserved for testing. The split was implemented using sequential indexing (df.iloc[0:train_days]), thereby preserving the natural time order of PV system operation.

Although batch-level shuffling (shuffle=True) was applied during training to improve gradient convergence, this shuffling was restricted exclusively within the training dataset and did not mix future test observations into the learning phase. Therefore, the PRLSTM model was trained only on past information and evaluated on truly unseen future data, ensuring realistic forecasting conditions.

5.5.2.Random-seed control and reproducibility

During PRLSTM model development, reproducibility was ensured by fixing the random seeds associated with NumPy, TensorFlow, and Python's random module prior to model initialization. This controlled stochastic variations arising from random weight initialization of the stacked LSTM layers, dropout regularization (rate = 0.6), and mini-batch selection during training (batch size = 2).

By maintaining fixed seeds, repeated executions of the model on the same dataset produced consistent training–validation loss convergence trends and stable RMSE values, thereby enabling fair comparison between LSTM and other deep learning architectures implemented in this work.

5.5.3.Hyperparameter selection

The hyperparameters of the developed PRLSTM model were selected through an iterative empirical tuning process aimed at minimizing validation loss while preventing overfitting. The architecture consists of four stacked LSTM layers, each with 256 hidden neurons, followed by dropout regularization layers. A dropout rate of 0.6 was applied after each LSTM layer to mitigate overfitting by reducing neuron co-adaptation during training.

The model was trained using the Adam optimizer with a batch size of 2 over 500 epochs. The final hyperparameter configuration was determined based on validation loss convergence behavior and generalization performance on unseen test data. Different combinations of hidden units, dropout rates, and epoch values were experimentally evaluated, and the configuration reported in Table 5.2 (Page 129) provided the most stable convergence and lowest RMSE without significant overfitting.

5.5.4.Hybrid model comparison

For fair architectural comparison, an identical training and evaluation protocol was applied to all implemented deep learning models, including LSTM, GRU, CNN, and FFNN. A chronological 80–20 train–test split was adopted to prevent temporal leakage, ensuring that only past observations were used for training and future data were reserved for evaluation.

Fixed preprocessing steps, including normalization and consistent input window size (n_input = 2, n_output = 1), were uniformly applied across all models. Furthermore, the same optimizer (Adam), batch size (2), number of epochs (500), and evaluation metrics (RMSE, MAE, and MAPE) were maintained for consistency.

All deep learning models were implemented in Python using TensorFlow/Keras and trained on a MacBook CPU-based system without GPU acceleration. Bayesian hyperparameter optimization with 20 trials and early stopping (patience = 20, maximum epochs = 300) was employed for all models.

- Autoencoder–BiLSTM: -60–70 minutes
- CNN–BiLSTM: -60–70 minutes
- GRU–BiLSTM: -65–75 minutes
- BiLSTM + Attention: -50-80 minutes
- BiLSTM + Transformer: ~55–85 minutes
- Hybrid CNN–BiLSTM–Transformer: ~120–150 minutes

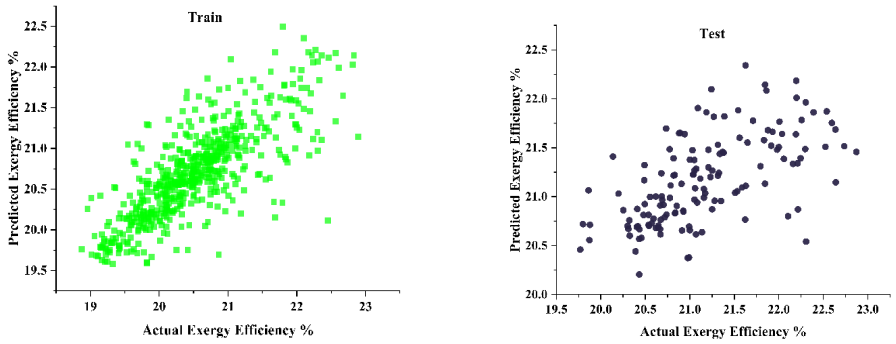
5.6. Results and Discussions for Predictive Assessment of Exergy efficiency

The proposed dual-layer hybrid and multi-layer hybrid models are deployed to generate daily predictions of thermal exergy efficiency for a 191.9 kWp grid-connected photovoltaic (PV) plant situated in Mumbai, India. These predictions are evaluated comprehensively across the training, testing, and validation phases to ensure robust model performance under realistic operating conditions. The hybrid architectures incorporate deep learning constructs such as CNN-BiLSTM, GRU-BiLSTM, Attention-BiLSTM, Autoencoder-BiLSTM, and Transformer-BiLSTM, each designed to capture both spatial and temporal dependencies in the multivariate time-series data.

To quantify and compare the predictive capability of each model, established error metrics Mean Bias Error (MBE), Mean Absolute Percentage Error (MAPE), and Root Mean Square Error (RMSE) are computed during both model training and independent testing phases. These metrics provide insight into systematic deviation (MBE),

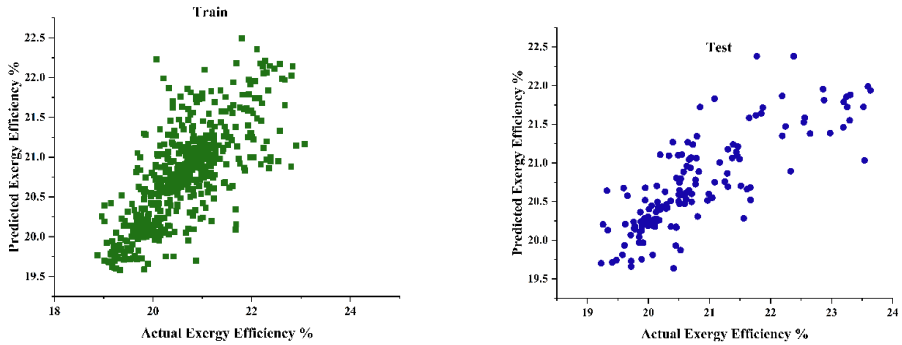
percentage-based accuracy (MAPE), and overall prediction dispersion (RMSE), respectively. Figures 5.11 to 5.16 present the comparative results between the predicted and actual values of exergy efficiency for each of the developed hybrid models.

The training-phase based variation of predicted and actual outcome evaluate the model capability to learn the temporal dynamics and relationships between meteorological and operational features.



(a) (b)

Fig 5.11. Variation of predicted and actual Exergy Efficiency during (a) training and (b) testing employing multi-layer hybrid model.



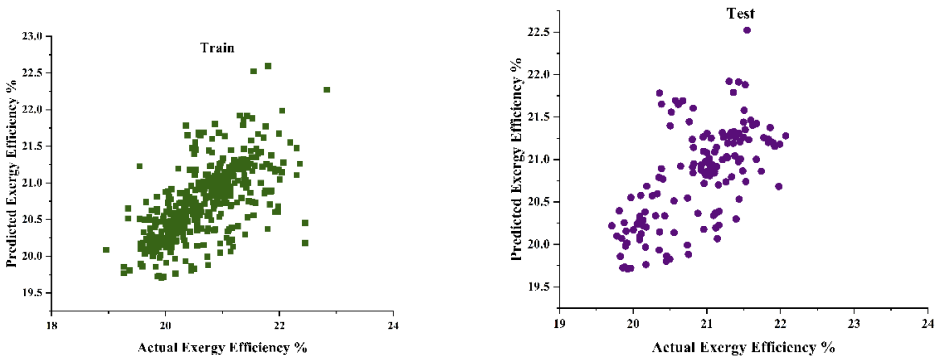
(a) (b)

Fig.5.12. Variation of predicted and actual Exergy Efficiency during (a) training and (b) testing employing dual layer hybrid CNN + BiLSTM model

The testing-phase results demonstrate the generalization ability of each model when exposed to unseen data. Collectively, these analyses allow for performance benchmarking across models, helping identify the most reliable architecture for real-world implementation of PV system performance forecasting.

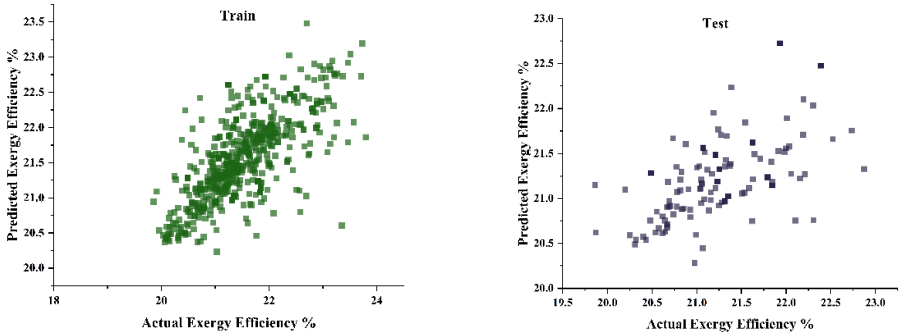
During the training phase as in figure 5.11 (a), the proposed multi-layer hybrid model exhibits strong predictive performance, as evidenced by the tight clustering among the predicted actual exergy efficiency values. The close alignment of the data points along the ideal diagonal line indicates a high degree of correlation, suggesting that the model has effectively learned the complex nonlinear interactions between the input features such as irradiance, module temperature, wind speed, and factor A and the corresponding exergy efficiency. This concentrated distribution reflects minimal prediction error and affirms the model's capacity to generalize within the training dataset. However, during the testing phase as in figure 5.11 (b), the scatter of predicted values displays a mild divergence from the actual values, implying a modest decline in predictive accuracy when exposed to unseen data.

Figures 5.12 to 5.16 further support these observations for the proposed dual-layer hybrid models, including CNN-BiLSTM, GRU-BiLSTM, Attention-BiLSTM, Autoencoder-BiLSTM, and Transformer-BiLSTM respectively.



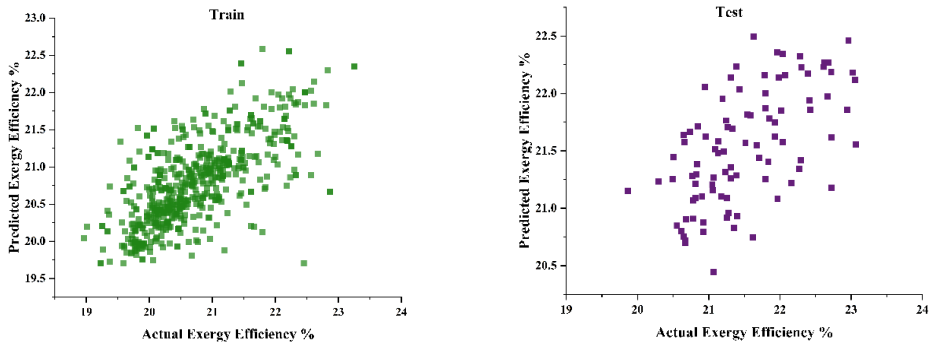
(a) (b)

Fig.5.13. Variation of predicted and actual exergy efficiency during (a) training and (b) testing employing dual layer hybrid GRU +BiLSTM model



(a) (b)

Fig.5.14. Variation of predicted and actual Exergy Efficiency during (a) training and (b) testing employing dual layer hybrid Attention + BiLSTM model



(a) (b)

Fig.5.15. Variation of predicted and actual Exergy Efficiency during (a) training and (b) testing employing dual layer hybrid Autoencoder model

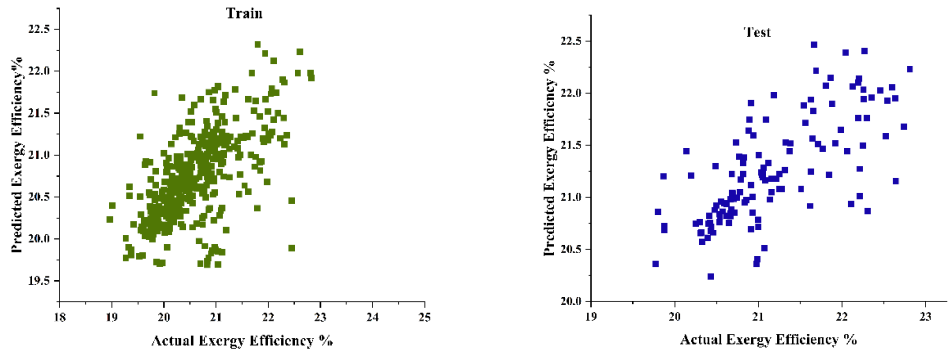


Fig.5.16.Variation of predicted and actual Exergy Efficiency during (a) training and (b) testing employing dual layer hybrid Transformer +BiLSTM model

In each case, the training plots exhibit high fidelity with reduced residuals, while the corresponding testing plots reveal slight dispersion, indicative of generalization behavior. The relatively lower error metrics (MBE, MAPE, RMSE) achieved by these models in both training and testing phases validate their effectiveness in capturing temporal-spatial dependencies and exergy degradation trends under realistic PV operating conditions.

Table 5.3. Training and Testing Accuracy for the developed dual layer and multi-layer hybrid models at Khopoli.

S.No	Model Name	Training accuracy	Testing accuracy
1	Proposed Multi-layer hybrid	0.987	0.976
2	CNN + BiLSTM	0.971	0.975
3	GRU+ BiLSTM	0.980	0.977
4	Attention+ BiLSTM	0.981	0.966
5	Autoencoder+ BiLSTM	0.979	0.976
6	Transformer+ BiLSTM	0.977	0.980

To evaluate the effectiveness of the proposed multi-layer hybrid model, a comparative analysis of its training and testing accuracies is presented in table 5.3. The multi-layer hybrid model achieves the highest training accuracy of 0.987, demonstrating superior capability in learning intricate patterns and relationships from the training dataset. Furthermore, its testing accuracy of 0.976 is closely comparable to that of the Transformer-BiLSTM (0.976) and Autoencoder-BiLSTM (0.980) dual-layer hybrid models, indicating strong predictive performance on unseen data. A key observation is that the minimal gap of 0.011 between training and testing accuracies, suggests that the multi-layer hybrid architecture exhibits robust generalization with negligible signs of overfitting. Across all the proposed hybrid models, accuracies lie within a narrow and consistent range of 0.971 to 0.987, reflecting the reliability of each model under realistic operational conditions.

5.6.1. Validation of Model Across Telangana and Bengaluru locations

To evaluate the prediction accuracy of different hybrid models, two realistic operations of Solar PV plants with a capacity of 294.8 kWp situated in 17.2212° N, 78.6944° E at Mallapur, Hyderabad, Telangana, India and a 125 kWp photovoltaic facility located at

13.1596° N, 77.6362° E are considered. The 125 kWp was situated within the transit campus of the Energy Institute in Bengaluru. It has a humid subtropical environment with frequent overcast conditions, lower solar availability, and higher humidity. Its corresponding input ranges for irradiance is 350–900 W/m² and wind speed ranging between 0.8–3.9 m/s. In contrast, Telangana is a semi-arid climate zone, where high irradiance and sharp temperature swings are common. During the observation period, irradiance varied between 500–1050 W/m² and wind speed from 1.2–5.3 m/s.

Mumbai, a coastal location, experiences high relative humidity typically in the range of 70–85%, along with higher air density due to moist air. Elevated humidity increases atmospheric scattering and absorption, reducing effective plane-of-array irradiance, while salt-laden air accelerates soiling and surface degradation. Ambient temperatures often exceed 30–35 °C, leading to higher module operating temperatures and increased thermal exergy destruction. Moderate wind speeds (2–3 m/s) provide limited convective cooling, contributing to exergy efficiency of 21.32 %.

Telangana, representing a semi-arid inland climate, is characterized by lower humidity levels (30–50%) and high solar irradiance. However, ambient temperatures frequently exceed 35–40 °C, resulting in elevated module temperatures and pronounced temperature-induced voltage losses. Wind speeds are typically moderate (1.2–4.3 m/s), offering insufficient cooling to fully offset thermal losses, which adversely impacts both electrical and exergy performance despite high irradiance availability. Hence the exergy efficiency of the plant in Telegana is 21.2%.

Bengaluru benefits from a comparatively moderate climate, with ambient temperatures generally in the range of 20–30 °C and average wind speeds (0.8–3.9 m/s). Enhanced convective cooling lowers module operating temperature, reducing thermal losses and voltage degradation. Although relative humidity is moderate (50–65%) and irradiance levels are slightly lower than in Telangana, the favorable thermal environment results in improved exergy efficiency of 22.5 %.

Three featured statistical error metrics were considered for predictive comparison namely the Root Mean Square Error (RMSE), Mean Bias Error (MBE), and Mean Absolute Percentage Error (MAPE) which offers a comprehensive view of model performance among the considered locations.

Considering the behaviour of the proposed models for Telangana, Transformer+BiLSTM model achieved the lowest RMSE of 0.5122, indicating the most accurate predictions in terms of magnitude, closely followed by GRU+BiLSTM with an RMSE of 0.5209 and Autoencoder+BiLSTM rendering 0.5253 as tabulated in table 5.4. While GRU+BiLSTM delivered strong RMSE and the lowest MAPE of 1.81%, its very high negative MBE (−0.4518) highlighted a substantial underprediction bias, making it less reliable despite its precision. In contrast, the Transformer+BiLSTM model maintained a competitive MAPE of 1.86% with a more moderate MBE (−0.0546), balancing accuracy and bias effectively. Other models such as CNN+BiLSTM and the Multilayer hybrid showed slightly higher RMSE and MAPE values but offered more stable bias behavior (MBE ≈ +0.012), suggesting they may provide more consistent predictions under real-world conditions. Overall, while GRU+BiLSTM leads in pure error magnitude with closeness

as represented in figure 5.17, the Transformer+BiLSTM model emerges as the more robust and reliable option for Telangana due to its favourable trade-off between accuracy and directional bias.

Table 5.4. Error Metrics delivered by the proposed models for Telangana Location

S.NO	Model	RMSE	MBE	MAPE
1	Multilayer hybrid model	0.5560	0.0117	0.0192
2	CNN +BiLSTM	0.5520	0.0121	0.0196
3	GRU +BiLSTM	0.5209	-0.4518	0.0181
4	Attention +BiLSTM	0.5264	-0.0043	0.0189
5	Autoencoder +BiLSTM	0.5253	0.0050	0.0198
6	Transformer +BiLSTM	0.5122	-0.0546	0.0186

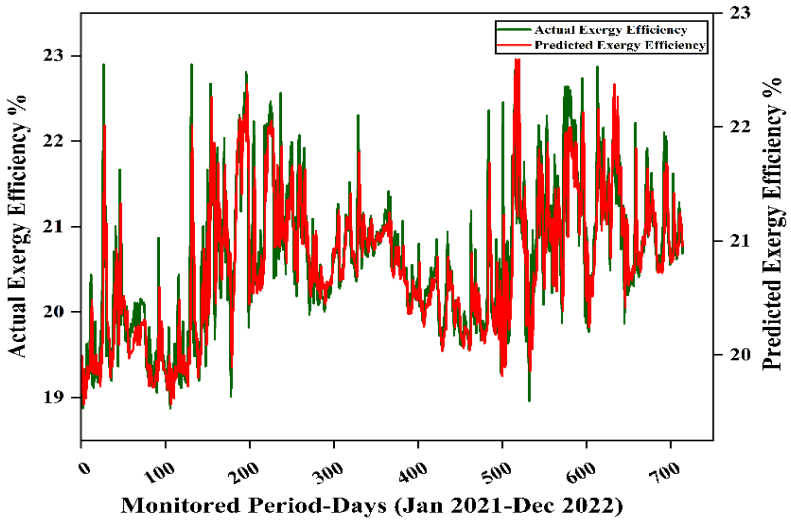


Fig.5.17. Comparison of Actual and Predicted Exergy efficiency for the location of Telangana as delivered by dual layer hybrid GRU+BiLSTM model

For the Bengaluru location, model performance was again evaluated using RMSE, MBE, and MAPE metrics to understand accuracy, bias, and reliability. Among all models, the

Multilayer hybrid model showed the most balanced performance with a low RMSE of 0.62103, the lowest MAPE of 1.88%, and a minimal bias (MBE = −0.0123), indicating accurate and stable predictions as in table 5.5 . In contrast, although the GRU+BiLSTM model also had a low RMSE of 0.6229 and competitive MAPE amounting to 1.89%, it exhibited a severe underprediction bias with MBE of −0.5502, making its outputs less reliable for real-world application. The Transformer+BiLSTM model, while maintaining a good MAPE (1.91%), showed moderate bias with MBE of −0.0816. Models like CNN+BiLSTM, Autoencoder+BiLSTM, and Attention+BiLSTM maintained consistent performance with slightly higher RMSE and MAPE but showed minimal bias, suggesting robustness across scenarios. Overall, the Multilayer hybrid model stands out as the most effective for the Bengaluru dataset, offering a strong trade-off between accuracy and generalization. Figure 5.18 depicts the variation between predicted and actual exergy efficiency for the Bengaluru location, obtained using the multi-layer hybrid model.

Table 5.5. Error Metrics delivered by the proposed models for Bengaluru Location

S.NO	**Model**	**RMSE**	**MBE**	**MAPE**
1	Multilayer hybrid model	0.62103	-0.0123	0.0188
2	CNN +BiLSTM	0.6512	-0.0129	0.0191
3	GRU +BiLSTM	0.6229	-0.5502	0.0189
4	Attention+BiLSTM	0.6462	-0.0113	0.0196
5	Autoencoder +BiLSTM	0.6924	-0.0069	0.0192
6	Transformer +BiLSTM	0.6621	-0.0816	0.0191

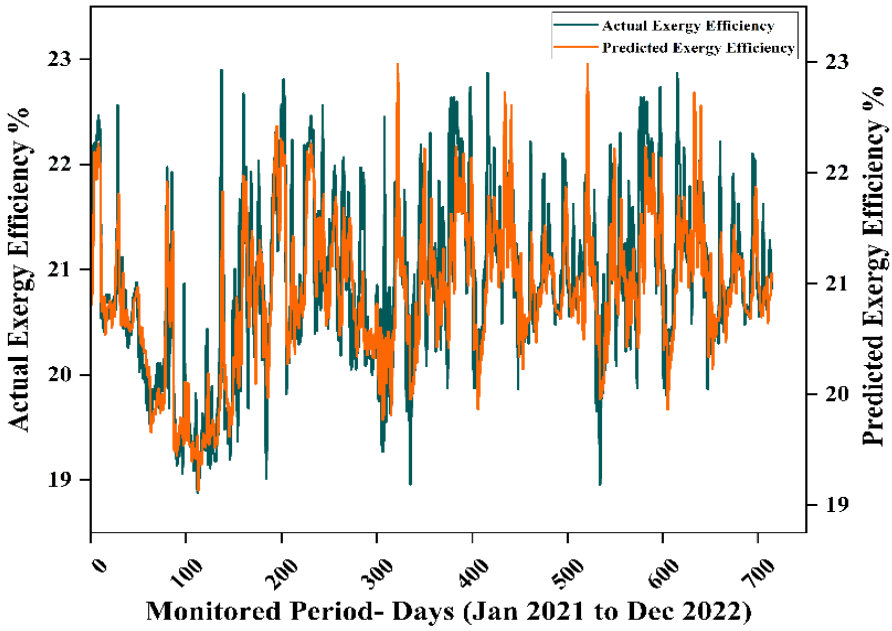


Fig.5.18. Comparison between actual and predicted exergy efficiency for the Bengaluru, location, as estimated by the proposed multi-layer hybrid model.

Overall, the multilayer hybrid model emerges as the most effective choice for the Bengaluru site, offering a well-rounded balance between predictive accuracy and stability. This is visually illustrated in figure 5.19, which compares the predicted and actual exergy efficiency values for this location.

5.6.2. Radar plot-based Model Performance Comparison across different locations

A radar plot which compares the error behaviour for all the developed models with respect to training and validated location is developed and illustrated in figure 5.19 (a) –(c). This include Multilayer+BiLSTM, CNN+BiLSTM, GRU+BiLSTM, Attention+BiLSTM, Autoencoder+BiLSTM, and Transformer+BiLSTM—across three distinct geographic locations which are Khopoli, Bengaluru, and Telangana. The first radar plot as in figure

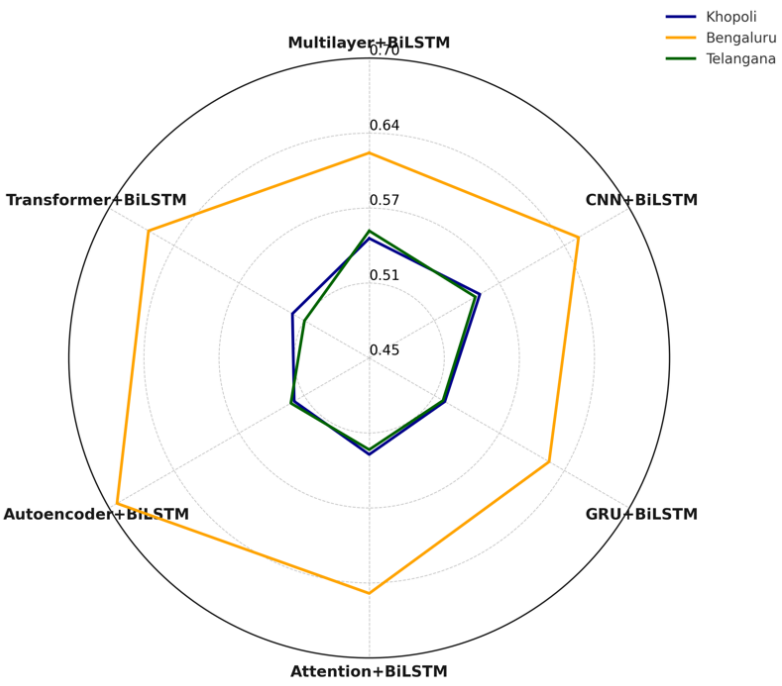
5.19 (a) shows the maximum variation of RMSE (Root Mean Square Error), where Bengaluru region exhibits the highest RMSE values across all models occupying relatively larger radar area in comparison to other locations. However, it lies appropriate in range of prediction from actual values. On the other hand, Telangana and Khopoli maintain lower and closely grouped RMSEs, reflecting higher accuracy, as the respective area occupied under the considered radar is relatively small as seen from figure 5.19 (a).

The radar chart in figure 5.19 (b) displays the variation in MAPE across all models and locations. Again, Bengaluru shows the widest spread, indicating relatively higher percentage error between predicted and actual values. This suggests that while the models are still effective, there is slightly less consistency in capturing the magnitude of change for the Bengaluru dataset.

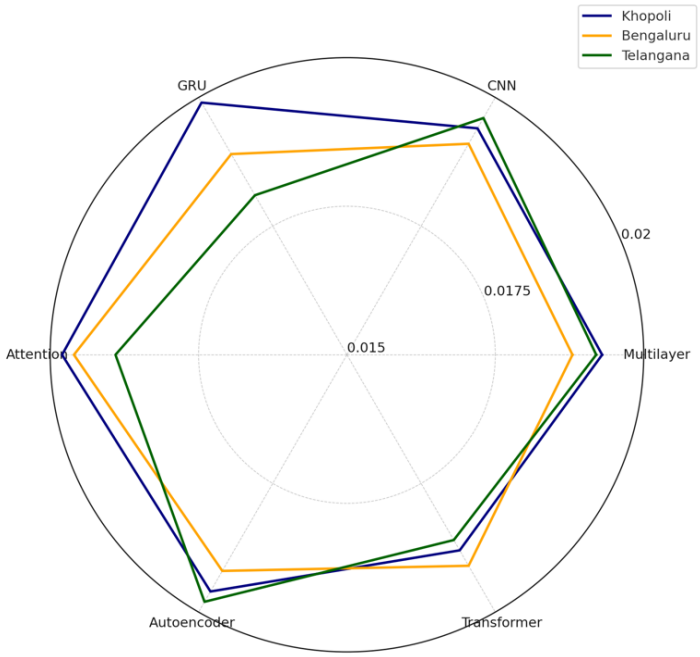
In contrast, Telangana and Khopoli demonstrate more compact and tightly grouped areas, reflecting lower MAPE values and thus greater accuracy in proportionate error. The smaller the radar area, the more reliable the model is in estimating exergy efficiency with minimal deviation from actual observed values. This reinforces the robustness of developed models for prediction for these two sites.

Figure 5.19 (c) illustrates the bias tendencies of each model using MBE.A larger area indicates stronger directional bias (either under or over prediction). Among the three regions, Telangana shows the most irregular and expanded area, suggesting that some models consistently underpredict or overpredict exergy efficiency in this location likely due to more extreme or variable climatic conditions. Bengaluru and Khopoli, on the other hand, exhibit more symmetric and smaller regions, implying that the models have more

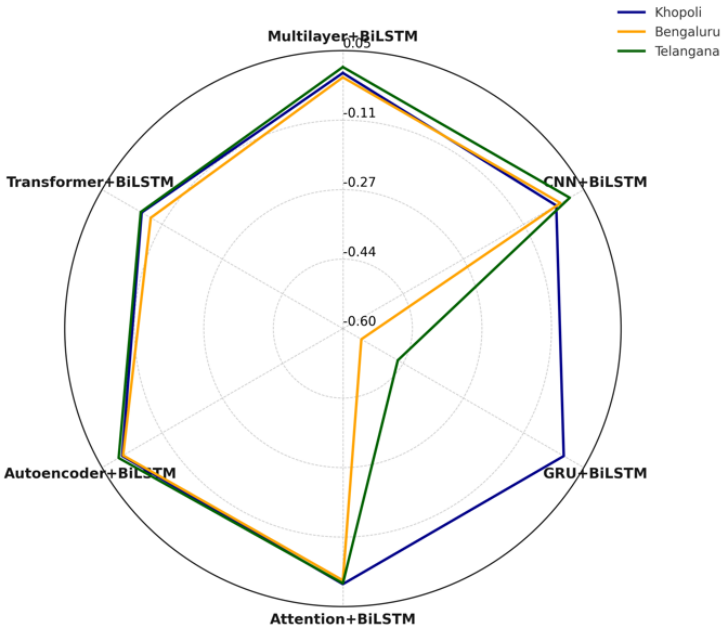
neutral bias behaviour, with minimal systematic overestimation or underestimation. This consistency in MBE suggests better generalization and reliability across the models when applied to these two locations.The radar plot-based analysis reinforces the robustness for spatially diverse solar PV datasets. The differences in the behaviour of the proposed models for all the locations is analysed and appropriate architecture selection among the proposed models for a particular location can be evidently made.



(a)



(b)



(c)

Fig.5.19 Performance evaluation of the proposed Multilayer Hybrid model and Dual layer hybrid models on the metric (a) RMSE and (b) MAPE and (c) MBE.

5.6.3 Methodological Limitations and Scope

Despite the encouraging predictive performance of the proposed AI-driven framework for exergy efficiency and performance ratio estimation, certain methodological limitations are inherently associated with the adopted modelling strategy, data sources, and analytical assumptions. These limitations are explicitly acknowledged to clearly define the scope, applicability, and boundaries of the proposed approach.

1. Dependency on supervised labels

The proposed AI/ML models are developed using a supervised learning paradigm, where analytically computed exergy efficiency and performance ratio values serve as target labels. Consequently, the predictive accuracy of the models is intrinsically dependent on the assumptions and uncertainties associated with the underlying thermodynamic formulations and measured input parameters. This limitation arises because exergy quantities are not directly measurable in field-scale photovoltaic systems and must be estimated analytically. While this dependency ensures thermodynamic consistency, it also implies that the model performance is bounded by the accuracy and reliability of the generated labels.

2. Climatic and site-specific bias

Although the models are trained and validated using datasets from multiple geographical locations, the primary training data correspond to photovoltaic systems operating under specific climatic and environmental conditions. As a result, the learned relationships may exhibit reduced accuracy when applied to regions characterized by substantially different

irradiance spectra, temperature regimes, or atmospheric conditions. While site-wise independent training helps capture local operational characteristics and limits overfitting, complete climatic generalization cannot be guaranteed without broader multi-climate training datasets or adaptive learning strategies.

3. Sensor accuracy and measurement uncertainty

The reliability of the proposed framework depends on the accuracy of measured electrical and meteorological parameters, including irradiance, ambient temperature, module temperature, and power output. Measurement uncertainty associated with these sensors propagates into the computation of performance ratio, exergy efficiency, and degradation indicators, thereby influencing both analytical calculations and AI-based predictions. Although IEC-compliant instrumentation and standard data pre-processing procedures are employed, residual uncertainty remains unavoidable in long-term field measurements.

To mitigate the impact of these limitations, consistent thermodynamic formulations, identical data pre-processing steps, chronological train–test separation, and repeated training runs were adopted throughout the study. These measures enhance robustness and reduce sensitivity to stochastic variability and measurement noise.

Overall, the identified limitations do not undermine the validity of the proposed framework but instead define its current operational envelope. They also highlight clear directions for future enhancements, including uncertainty-aware modelling, physics-informed learning, and broader climatic generalization through multi-site data integration.

5.7. Summary

The present study builds a robust data-driven modelling framework to evaluate and predict exergy efficiency in PV systems under real-world conditions. Using a combination of thermodynamic principles and deep learning techniques, each section contributes a specific aspect to the overall methodology from foundational analysis to model design and performance comparison. The following is a section-wise summary of the chapter:

- Section 5.1 introduces the thermodynamic basis for exergy analysis in photovoltaic systems and explains its significance over traditional energy-based metrics. Using real-time data from a 191.9 kWp PV plant in Mumbai, actual exergy efficiency is computed to quantify system performance under real field conditions.
- In section 5.2 key input variables such as irradiance, module temperature, wind speed, and current are identified. The section details data cleaning to remove outliers, interpolation of missing values, smoothing of noisy data, and normalization techniques to bring variables onto a common scale. These steps ensure high-quality, consistent input data for model training.
- Section 5.3 section presents hybrid architectures combining CNN, GRU, Autoencoder, Transformer with BiLSTM. Each model's configuration, mathematical formulation, and activation strategies are discussed to capture spatial-temporal dependencies in the data.
- In section 5.4 a novel multi-layer model is developed using Autoencoder, BiLSTM, and Transformer layers in a sequential pipeline. The architecture is designed to extract deep temporal patterns and contextual relevance, using

attention mechanisms and global average pooling for improved prediction accuracy. The superior performance of the Transformer + BiLSTM model arises from its ability to jointly capture short-term sequential dependencies and long-range temporal variability present in irradiance and temperature data.

- Section 5.5 presents a systematic training, validation, and robustness framework developed to ensure reliable performance evaluation, reproducibility, and fair comparison of the proposed multi-layer hybrid deep learning models. Chronological data splitting was adopted to prevent temporal leakage and preserve physical causality in time-series photovoltaic data.
- In section 5.6 the predictive performance of the proposed models is evaluated using RMSE, MBE, and MAPE across three sites: Khopoli, Telangana, and Bengaluru. ECDF plots and radar charts are used to visualize and compare the accuracy and generalization ability of each model under different climatic conditions.

Chapter 6: Prediction of Performance Ratio Encompassing Failure Mode Based Degradation

Overview:

This chapter presents a comprehensive framework for forecasting Performance Ratio (PR) of photovoltaic (PV) systems by incorporating true on-field loss parameters. It begins with a detailed performance analysis of a 191.9 kWp PV system and identifies key thermoelectric indicators namely, failure mode-based degradation rate (R_d), thermal exergy loss derived from exergy analysis, and global irradiance as critical inputs for accurate PR prediction. A Long Short-Term Memory (LSTM) based deep learning model is developed to capture temporal dependencies among these inputs. The chapter compares the predictive performance of the proposed PRLSTM model against other sequential models and benchmark tools like PVsyst and PRcorrected. Finally, extensive validation across two geographically distinct locations demonstrates the model's robustness and adaptability, reinforcing its utility as a reliable tool for short-term PR estimation under varying climatic conditions.

Outline of Chapter ***Page***

6.1. Importance of prediction of Performance Ratio (PR)

Accurate prediction and early estimation of the Performance Ratio (PR) in grid-connected photovoltaic (PV) systems is vital for ensuring energy reliability, cost-effectiveness, and regulatory adherence for both plant operators and grid managers. Traditional methods of estimating PR (Kim.et.al.,2023, Aste.et.al,2013,Sobri et al., 2018) primarily rely on mathematical formulations that necessitate the availability of multiple real-time dependent variables especially DC and AC energy output. However, such system-derived inputs are often difficult to obtain or forecast in advance. To address this limitation, the present study proposes a deep learning-based predictive framework that leverages pre-estimable parameters reflecting the electrical and thermal behaviour of PV systems. Specifically, it integrates novel thermo-electric indicators, including the failure mode-based power degradation rate (R_d) and thermal exergy loss, to offer a more practical and accurate tool for assessing on-field PR, which acts as a unique comparative health indicator among PV plants.

The proposed approach for estimating the annual performance ratio (PR) is based on the real-time operation of a 191.9 kWp grid-connected solar PV plant located at` 18°47'8.23" N latitude and 73°20'45.20" E longitude from January to December 2021.The plant is installed at Arkose Industrial Estate, Adoshi Road Dekhu, Sajgaon, Tal-Khalapur, Khopoli, Maharashtra, India. It comprises 460 PV modules, each rated at 370 Wp, as detailed in Chapter 3. The ground-based observations monitored at the site primarily involve solar irradiance and module temperature which are critical in understanding the energy generation potential of the PV system. As shown in figure 6.1, annual fluctuations in solar irradiance significantly impact the monitored plant's energy output at the AC end. For instance, in January,

an average irradiance of 791.25 W/m²/day corresponds to an energy yield of 12,634.41 kWh/day. This output rises in March with an irradiance of 808.75 W/m²/day, generating 16,244.73 kWh/day. The peak irradiance is observed in May at 1088.75 W/m²/day, resulting in the highest energy generation of 21,286.11 kWh/day. Comparatively, October registers a slightly lower irradiance of 1013.75 W/m²/day, leading to a reduced energy output of 17,961.3 kWh/day. These trends highlight the direct influence of photo-induced flux on energy production. However, it is important to note that other environmental and operational factors such as module temperature, meteorological conditions, and system efficiency also play a significant role in determining the actual energy output.

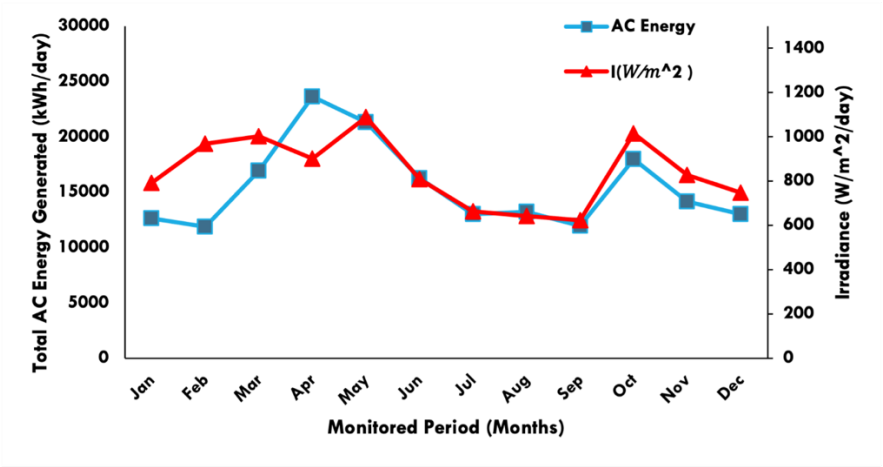


Fig.6.1. Monthly average daily variation of Solar Irradiance and AC energy produced by 191.9 kW_p PV plant under consideration

The monthly average temperature of the PV modules at the considered site ranges from a minimum of 43.76 °C to a maximum of 54.2 °C, with an annual average of 51.4 °C. Elevated module temperatures are known to contribute to several degradation mechanisms such as thermal stress, material fatigue, hot spots, and the development of micro-cracks (Madurai Elavarasan et al., 2022). Figure 6.2 illustrates the variation in recorded module temperature (T_m) throughout the year.

Notably, April records the lowest average T_m at 41.9 °C/day, during which the AC energy output peaks at 23,626.25 kWh/day. This inverse relationship indicates that higher module temperatures are generally associated with occurrence of failure modes resulting in reduced energy generation.

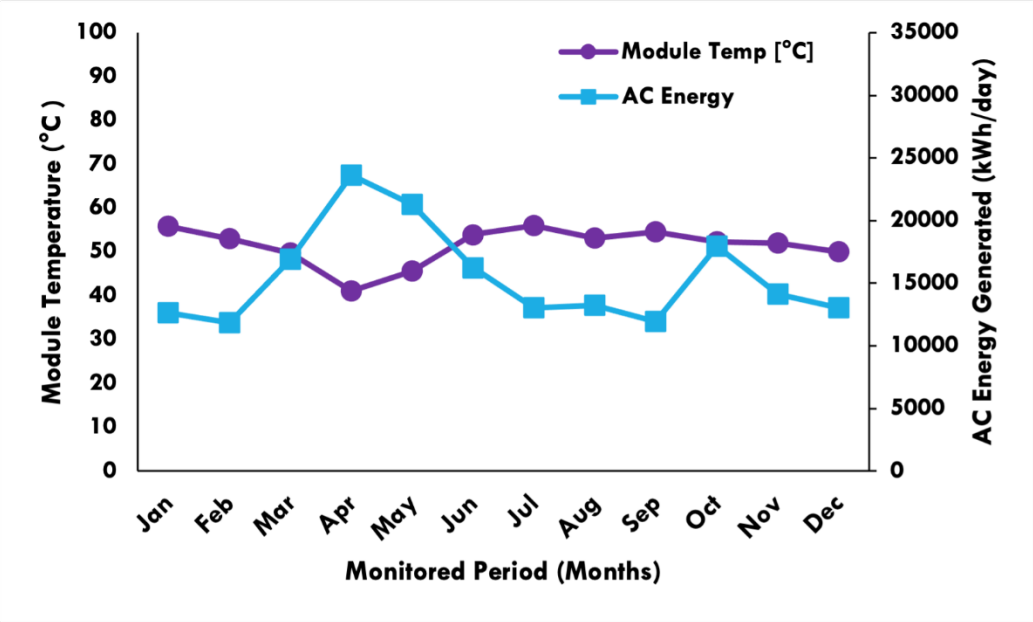


Fig.6.2. Monthly average daily module temperature and AC energy generation at PV plant site Khopoli, India

The evaluated performance indicators such as CUF, Y_a, Y_f, and PR are 13.6%, 74.82 kWh/kWp/month, 99.76 kWh/kWp/month and 46.9% respectively as described in Chapter 3. A comparative summary of performance metrics of reported distributed PV systems and the investigated are summarized across different locations in table 6.1. While the table primarily reflects the variation in PR across locations, this variation provides an initial basis for exploring the impact of environmental, operational factors and failures on PR. A detailed investigation into the critical inputs influencing PR, including degradation effects, exergy loss, and irradiance, is presented in subsequent sections of this chapter. These factors collectively impact the technical reliability and economic viability of PV systems.

This insight forms the basis for identifying critical input parameters essential for developing a robust model capable of accurately estimating PR.

Table 6.1. Comparison of PR for the PV plant under study with different reported systems at different Indian locations

S.No	Location	Rated Capacity	Monitored Duration	Y_F	Y_A	CUF %	PR%	Reference
1	Khatkar-Kalan, Karnataka,	190 kW	Annual	2.23	-	-	60.8	(Sharma & Chandel, 2013)
2	Gujarat, India (PV1)	500 kW	Annual	2.79	-	-	75.3	(Tripathi et al., 2014)
3	Gujarat, India (PV2)	500 kW	Annual	5.14	-	-	70.8	(Tripathi et al., 2014)
4	Rajkot, India	50.5 kW	Annual	4.49	4.71		70	(Dobaria et al., 2016)
5	Telangana, India	10 MW	Annual	4.44	-	17.68	85.2	(B. S. Kumar & Sudhakar, 2015)
6	Tamil Nadu, India	5 MW	Annual	4.81	5.46	-	89.15	(Sundaram & Babu, 2015)
7	Karnataka, India	3 MW	Annual	3.73	-	-	70	(Padmavathi & Daniel, 2013)
8	Tamil Nadu, India	1 MWp	Annual	4.26	-	-	-	(Sundaram, 2022)
9	Maharashtra, India	191.9 kWp	Annual	3.57	8.01	9.2	43.09	Present Study
10	Haryana, India	186 KWp	Annual	4.28	4.57	17.8	82.79	(Arora et al.,2022)

6.2 Identification of Critical Inputs impacting PR

The development of the PR analyzer and user-interface-based tool begins with the identification and selection of key input parameters that influence the Performance Ratio (PR). In the proposed approach, PR is primarily affected by thermal exergy loss dependent on exergy efficiency, plane-of-array irradiance, and the power degradation rate arising from failure modes. A comprehensive analysis, both physical and statistical relationship between these selected attributes and the response variable, PR, is presented to ensure accurate modelling and prediction.

6.2.1 Influence of Exergy Loss, Irradiance and R_d

a) Thermal Exergy loss and its impact on PR:

Exergy represents the quality of the useful power output delivered by a system, accounting for internal irreversibilities and inefficiencies. It provides a more meaningful measure of how effectively the available energy potential is utilized. Unlike conventional energy analysis, exergy analysis considers both the quantity of energy and its usability or quality.

For a solar PV system operating under steady-state conditions over a given time period, the overall exergy balance is expressed by Eq. (5.1), as presented in Chapter 5. Exergy loss was evaluated at each measurement interval using steady-state exergy relations applied to measured, time-varying operational data. Although the governing equations are steady-state in form, their application to sequential field measurements enables representation of the PV system under continuously changing environmental conditions.

This quasi-steady-state approach is widely adopted in field-based PV exergy analyses and allows transient operating behaviour to be captured through time-resolved data rather than explicit transient modeling.

In solar PV systems, the energy conversion process produces electrical output along with unavoidable thermal losses. The electrical energy delivered to the load is used to evaluate energy efficiency. In contrast, thermal losses and internal irreversibilities represent the portion of energy degraded during the conversion process and form the basis for exergy efficiency evaluation.

These irreversibilities, collectively referred to as exergy loss, are quantified using Eqs. (5.2)–(5.11) in Chapter 5. These equations capture the degradation of energy quality occurring during the PV energy conversion process.

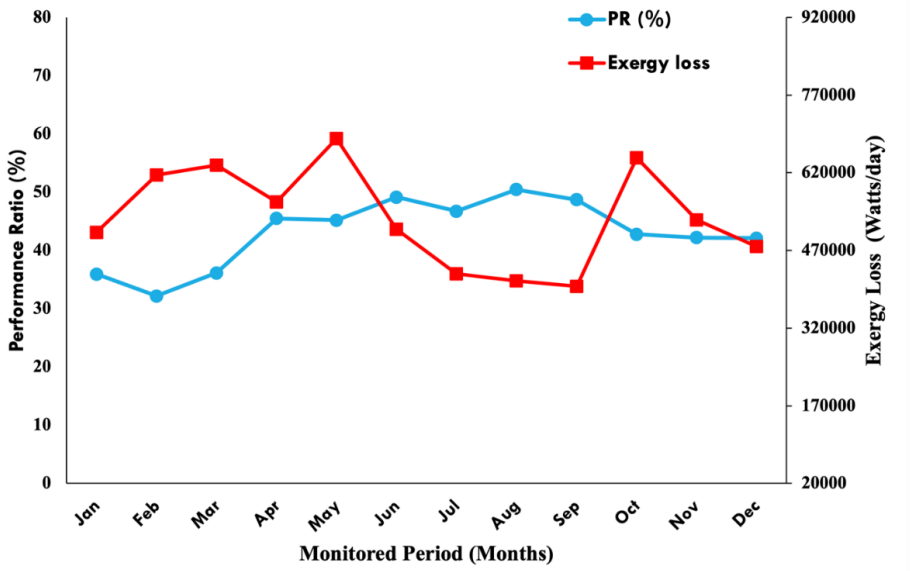


Fig.6.3. Monthly variation of mean daily thermal exergy loss and performance ratio over the analysis period for the 191.9 kWp PV system.

The daily thermal exergy loss values were averaged to obtain the monthly mean of daily thermal exergy loss for the PV system, as shown in Fig. 6.3. The figure indicates an inverse relationship between thermal exergy loss and the performance ratio (PR). An increase in thermal exergy loss corresponds to a reduction in PR, whereas a decrease in thermal exergy loss results in an improvement in PR.

For instance, between January and February, the thermal exergy loss increases from 505,267 W/day to 615,495 W/day. During the same period, the PR decreases from 35.94% to 32.19%. This clearly demonstrates the detrimental effect of thermal losses on the overall performance of the PV system.

b) Global Solar Irradiance and its Impact on PR

The dependence of the performance ratio (PR) on global solar irradiance for the studied PV plant is analyzed using monthly average daily data, as shown in figure. 6.4. The solar irradiance at the site varies from a minimum of 4.9 kWh/m²/day to a maximum of 10.2 kWh/m²/day over the monitoring period.

As illustrated in figure. 6.4, PR shows a decreasing trend during months with higher solar irradiance. This occurs because increased incident irradiance raises the reference yield (Yr) of the PV plant. Since PR is defined as the ratio of actual yield to reference yield, an increase in Yr without a proportional rise in actual power output results in a lower PR.

Consequently, the highest monthly mean PR values are observed during June and July, when solar irradiance is relatively lower, ranging from 662 W/m²/day to 808 W/m²/day. This behavior highlights the inverse relationship between irradiance intensity and PR for the considered PV system.

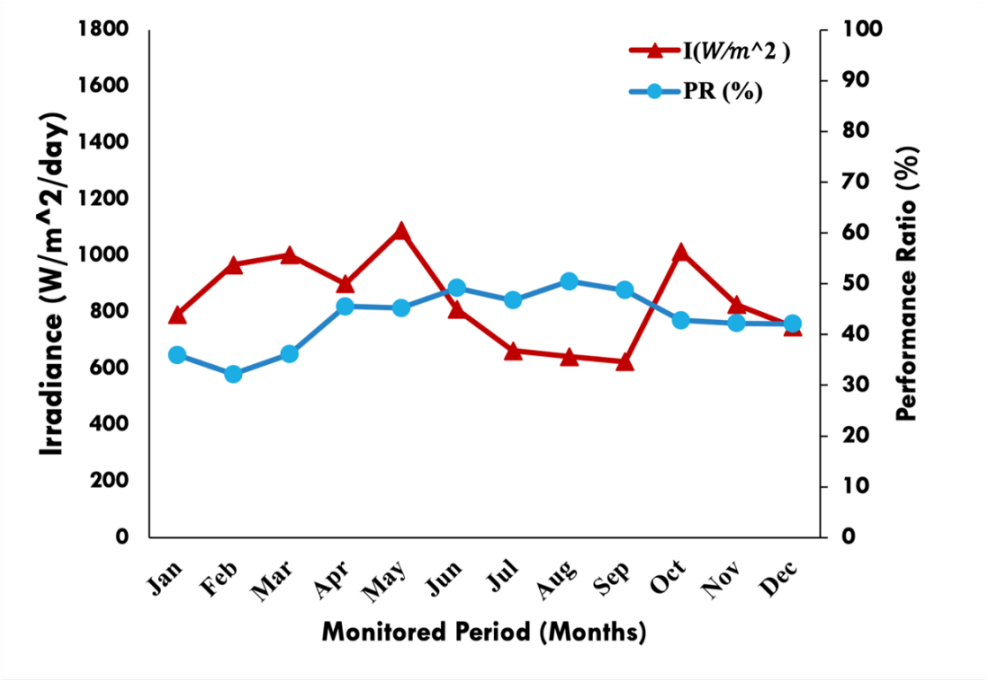


Fig 6.4. Monthly mean of daily variation of global solar irradiance and performance ratio (PR) for the 191.9 kWp PV system under study.

c) Impact of R_d on PR:

The power degradation rate (R_d), arising from various failure modes, plays a significant role in the operational performance of photovoltaic (PV) systems. Since the 191.9 kWp plant under investigation has been operational since 2019, early-stage degradation mechanisms have a pronounced impact on performance. These

mechanisms include visual degradation, hot spots, micro-cracks, and encapsulant delamination, all of which contribute to power loss over time.

The effect of degradation is reflected in the observed variation of the performance ratio (PR), which ranges from a minimum of 32.19% to a maximum of 50.46%, with an annual average value of 43.09%. In addition, the influence of module temperature on PR is clearly evident, as shown in Fig. 6.5. Lower module temperatures correspond to higher PR values, indicating improved system performance.

This correlation suggests that increased thermal stress may initiate hot-spot-related failures, thereby accelerating performance degradation. To accurately capture this relationship, a sequentially tuned and optimized LSTM model was developed for R_d prediction. The architecture and topology of the model are detailed in Section 4.4 of Chapter 4.

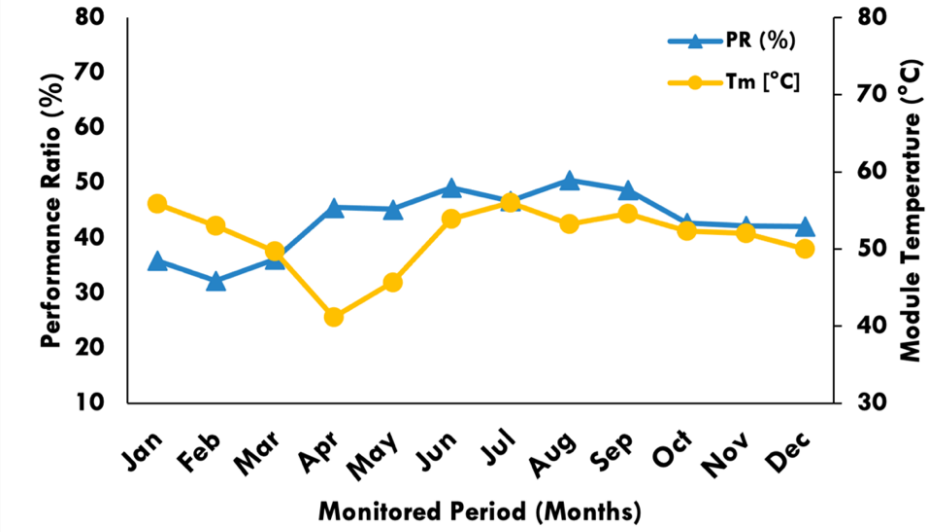


Fig.6.5. The trend of variation of monthly mean of daily Performance Ratio (PR%) on Module Temperature (T_m) for PV plant under study

The R_d - based LSTM estimator primarily uses module temperature and relative humidity as input features to predict the power degradation rate (R_d). The proposed LSTM model was validated against the established linear degradation rate

formulation reported by Hajjaj et al. (2020), as governed by equations (6.1) and (6.2).

For the 191.9 kWp PV system, the estimated degradation rate varies over the monitoring period, as illustrated in Fig. 6.6. The minimum R_d value of 9.8% was observed in January 2021, while the maximum value of 11.6% occurred in May 2021. The annual average degradation rate was estimated to be 10.12% per year. The relatively high failure mode–based degradation rate is attributed to the plant's operational age and its comparatively lower observed performance ratio (PR). Although the system is relatively new, with an operational age of six years, early-stage degradation mechanisms have already begun to influence its performance

$$.Linear\ Degradation\ Model = \frac{(P_{act}-P_{cr})}{P_{cr}} \times \frac{1}{operational\ age\ of\ the\ plant} \quad \text{Eq.(6.1)}$$

where, $P_{cr} = P_{max} * \frac{G_{STC}}{G_i} * \frac{1}{[1+\alpha*(T_m-T_{STC})]}$ Eq.(6.2)

P_{cr} denotes the power output corrected to Standard Test Conditions (STC), while P_{max} represents the rated maximum power of the photovoltaic (PV) module. G_{STC} is the reference irradiance under STC (1000 W/m²), and G_i corresponds to the actual irradiance measured under field conditions. The coefficient α captures the temperature sensitivity of the power output (in %/°C), with T_{STC} set at 25°C.

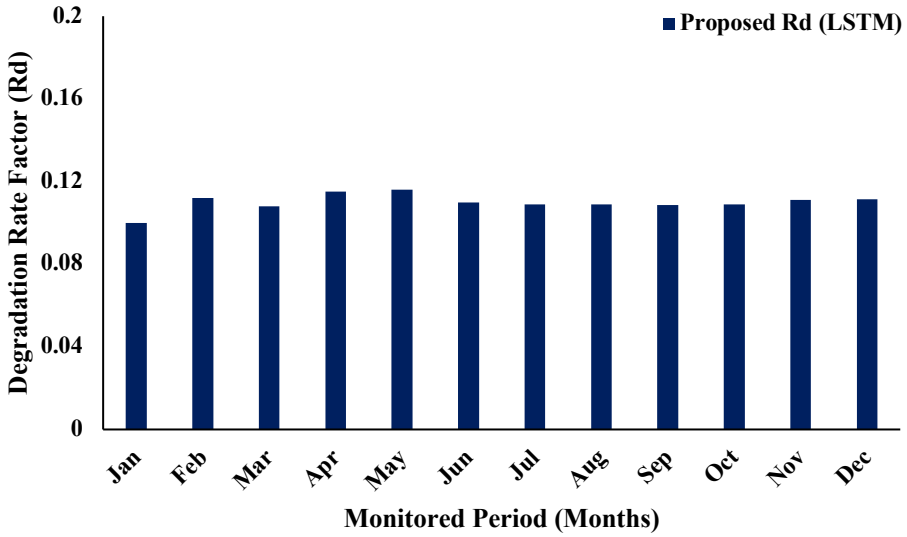


Fig.6.6 Evaluation of monthly mean of daily R_d through LSTM based model for the considered PV plant under study

The physical relationship between the power degradation rate (R_d) and the performance ratio (PR) is illustrated in Fig. 6.7, which presents the monthly average daily variation of both parameters over the analysis period. An inverse relationship between R_d and PR is clearly observed.

During February, PR decreases sharply to 32.19%, while R_d increases from 9.6% to 10.9% per day. Similarly, from June to September, PR shows a gradual improvement as R_d decreases over time. In May, a higher R_d is again associated with a reduced PR of 45.1%, further confirming this trend.

These results indicate a strong and non-negligible inverse dependence of PR on R_d, underscoring the significant role of degradation in determining the operational performance of the PV system.

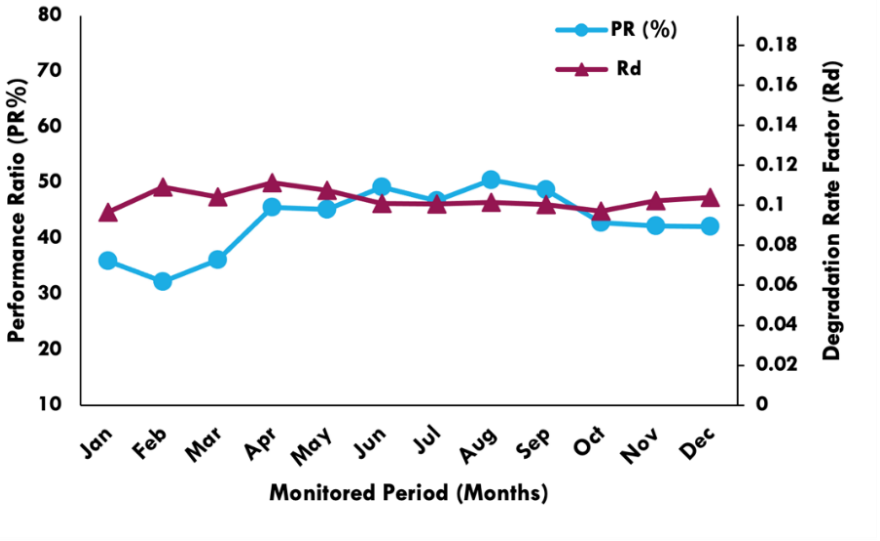


Fig.6.7 Variation of monthly mean of daily R_d and Performance ratio of 191. 9 kWp PV throughout the Investigated Period

6.2.2 ANOVA and Ablation Analysis

Beyond the physical interpretation and operational relevance of these parameters in influencing the Performance Ratio (PR), the statistical justification for including the proposed variables such as thermal exergy loss, degradation rate (R_d) and Solar Irradiance is supported using Analysis of Variance (ANOVA), as presented in table 6.2. This emphasizes the significance of the selected inputs in accurately modelling PR variations under real-world PV operating conditions.

Table 6.2. ANOVA analysis of considered Input Parameters over PR

Proposed Input attributes	Sum of Squares	df	F-value	p-value	Model R^2
A- Solar Irradiance (I)	168.56	1	33.37	0.0089	0.9366
B- Degradation rate factor (R_d)	68.09	1	13.48	0.0144	
C- Thermal exergy loss	169.56	1	33.57	0.0022	

The statistical analysis demonstrates that the performance ratio (PR) of the photovoltaic (PV) system is significantly influenced by the selected input variables. As shown in Table 6.2, this conclusion is supported by high F-values and very low p-values (all < 0.05), indicating that the observed effects are statistically significant and not due to random variation.

Global horizontal irradiance exhibits a strong influence on PR, with an F-value of 33.37 and a p-value of 0.0089. This confirms that irradiance variability plays a key role in determining system performance. This result is physically consistent, as irradiance governs photon availability for electricity generation.

The failure mode–based power degradation rate (R_d) also has a significant impact on PR. An F-value of 13.48 and a p-value of 0.014 indicate that R_d accounts for a substantial portion of PR variability. This is further supported by its sum of squares value of 68.09. The result aligns with practical observations that degradation mechanisms such as micro-cracks, delamination, and thermal cycling progressively reduce conversion efficiency.

Thermal exergy loss shows an even stronger statistical influence, with a low p-value of 0.0022. This highlights the thermodynamic importance of irreversibility and heat dissipation in PV performance assessment, particularly under varying climatic conditions.

The combined significance of irradiance, R_d, and thermal exergy loss is confirmed by a high coefficient of determination (R^2) of 0.9366. This indicates that more than 93% of the variation in PR is explained by the proposed model. Overall, the results justify the inclusion of degradation rate and exergy loss in the performance modelling framework. They also emphasize the need to move beyond conventional irradiance-only evaluations toward a more comprehensive, degradation- and

quality-based assessment of PV system performance.

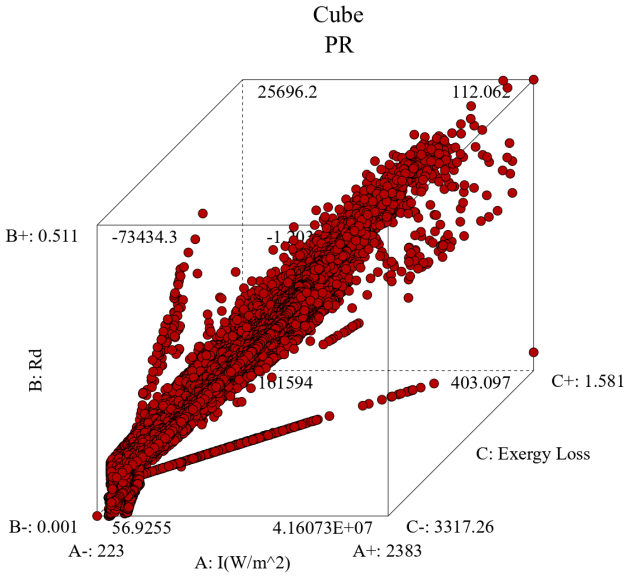


Fig.6.8. Dispersion showing the interaction of Input parameters over PR

The dispersion of data points shown in figure 6.8 indicates a strong interdependence between the input variables solar irradiance, thermal exergy loss, and degradation rate (R_d) and the output variable, performance ratio (PR). Variations in any of these inputs result in noticeable changes in PR, highlighting their critical role in PV system performance.

To further quantify the contribution of each input, an ablation study was performed by systematically removing individual features from the model. The results, presented in figure 6.9, identify R_d as the most influential parameter. Excluding R_d caused a substantial increase in RMSE to 0.18 and a reduction in R^2 from 0.92 to 0.84, indicating a significant loss in prediction accuracy.

Solar irradiance emerged as the second most important factor. Its removal increased RMSE to 0.15 and reduced R^2 to 0.80. Thermal exergy loss, while comparatively less influential, still affected model performance, increasing RMSE to 0.08 and lowering R^2 to 0.70.

Overall, the results confirm that all three inputs are essential for accurate PR estimation, with R_d exerting the strongest influence among them. Feature sensitivity was evaluated using a combination of ANOVA-based statistical analysis and ablation studies, providing consistent and complementary insights into the relative importance of each input parameter.

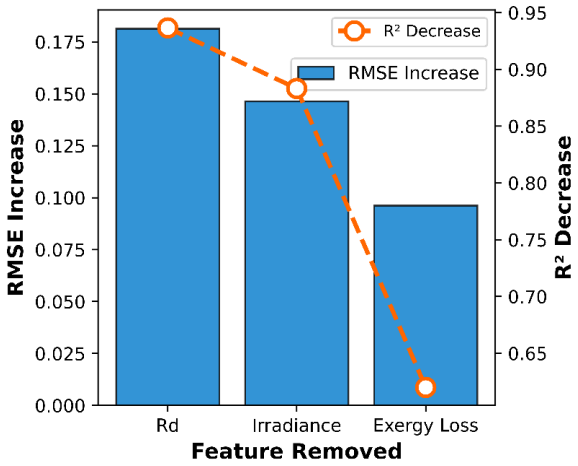


Fig.6.9. Ablation Analysis for the considered input factors for estimation of PR

The development of the PR analyser involves systematic preprocessing of the input features solar irradiance, power degradation rate (R_d), and thermal exergy loss. Sampling, filtering, and normalization were applied at a minute-level resolution to ensure data consistency.

To assess prediction uncertainty and model robustness, bootstrap resampling was employed. This approach yielded a 95% confidence interval for RMSE in the range of 0.0957 to 0.103, indicating stable and reliable model performance with low variability.

As shown in figure 6.10, the RMSE distribution is symmetric and well centred, with most error values lying within the confidence bounds. This consistency across resampled datasets confirms the robustness of the PR-LSTM model and supports its suitability for reliable real-world deployment.

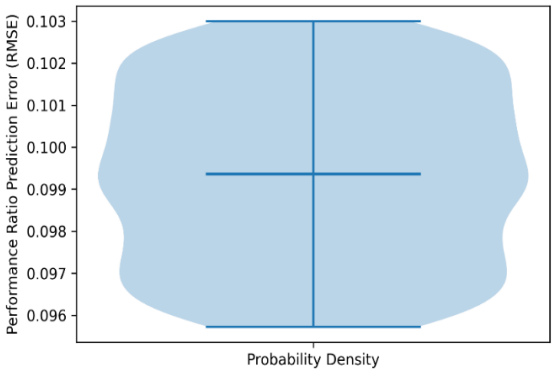
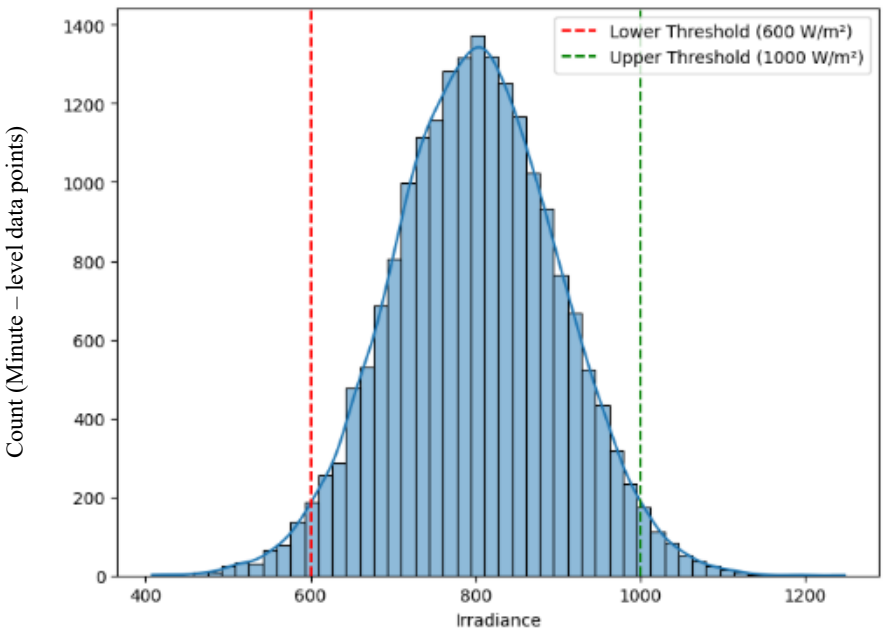


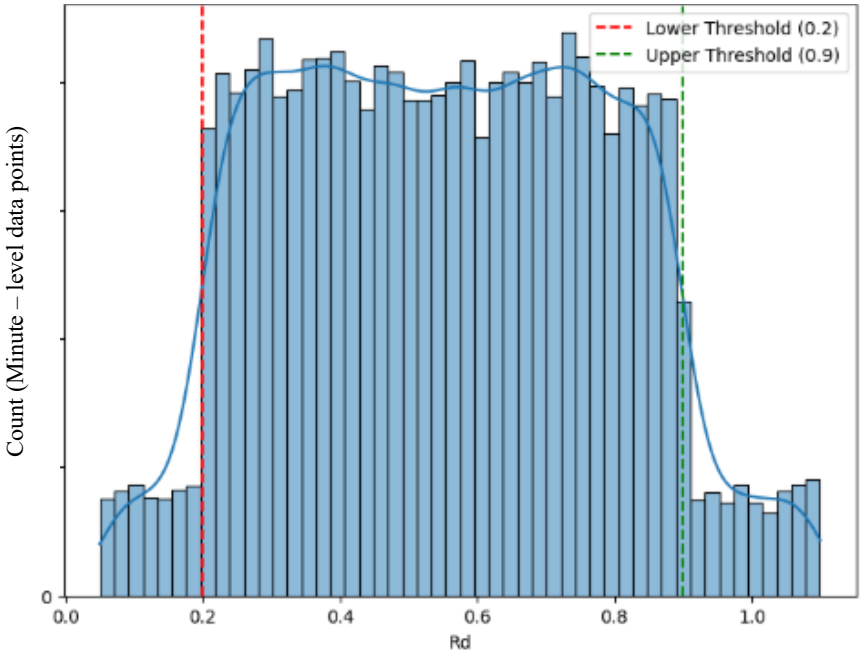
Fig.6.10. Violin plot of Bootstrap resampling for the considered attributes and response

Threshold ranges for irradiance (600–1000 W/m²) and degradation rate (R_d) (0.2–0.9) were established through exploratory data analysis (EDA) to preserve meaningful data variation while filtering out extreme outliers that could skew the model's learning process. Histogram analysis, as illustrated in figure 6.11, revealed that values outside these ranges were either infrequent or potentially erroneous, validating their exclusion. The concentration of data points within these thresholds ensures that essential patterns are maintained, thereby enhancing model generalization and reducing bias. To enable effective time-series modelling, min-max normalization was applied to the performance ratio (PR) and its influencing variables, as expressed by equation (6.3), standardizing their values across a comparable scale.

$$Y_B = \frac{Y_A - Y_A^{min}}{Y_A^{max} - Y_A^{min}} \qquad \text{Eq.(6.3)}$$



(a)



(b)

Fig. 6.11. Histogram Plot of (a) Irradiance and (b) R_d justifying the selection of threshold range of attributes

6.2.3 SHAP-Based Feature Importance and Model Interpretability Analysis

To further enhance the interpretability of the proposed PRLSTM model and quantitatively assess the contribution of each input variable to the predicted Performance Ratio (PR), a SHapley Additive exPlanations (SHAP) analysis was

conducted. Unlike ANOVA, which evaluates statistical significance, and ablation studies, which measure performance degradation upon feature removal, SHAP provides a model-specific, prediction-level attribution mechanism that quantifies the marginal contribution of each feature toward the output.

SHAP values were computed using a KernelExplainer framework applied to the trained LSTM model. Since the PRLSTM operates on sequential inputs, SHAP contributions were aggregated across time steps to obtain global feature importance. The analysis considered the three primary inputs: failure mode-based degradation rate (R_d), thermal exergy loss, and global solar irradiance.

The SHAP bar plot (figure 6.12) reveals that R_d and thermal exergy loss exhibit the highest mean absolute SHAP values, indicating that these parameters contribute most significantly to PR prediction. In contrast, irradiance shows a comparatively smaller independent SHAP magnitude. This result does not imply that irradiance is unimportant; rather, it reflects the intrinsic definition of PR as a ratio of actual yield to reference yield, where irradiance influences both numerator and denominator, thereby reducing its isolated contribution in the model's attribution framework.

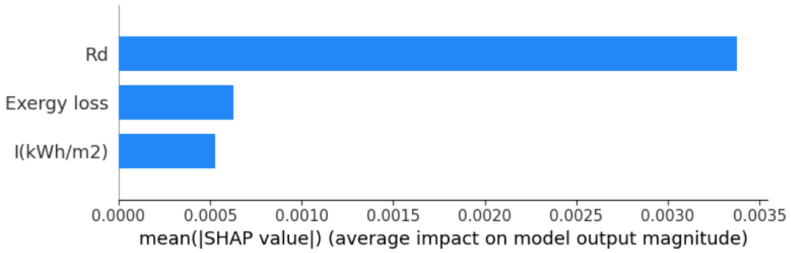


Fig.6.12. Global SHAP feature importance for PR prediction using the proposed PRLSTM model.

The SHAP beeswarm plot (figure 6.13) further illustrates the directional impact of each feature. Higher values of R_d correspond to negative SHAP contributions, indicating a reduction in PR due to degradation effects. Similarly, increased thermal exergy loss produces negative contributions, confirming the thermodynamic

principle that higher irreversibility lowers system performance. These nonlinear relationships demonstrate the capability of the PRLSTM model to capture complex degradation–performance interactions beyond linear statistical assumptions.

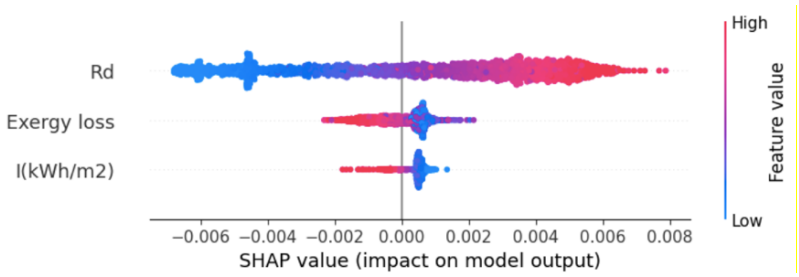


Fig.6.13.Mean absolute SHAP values indicating the global importance of R_d, thermal exergy loss, and irradiance in PR prediction.

When compared with the ablation results presented in Figure 6.9, a strong consistency is observed. The ablation study identified R_d as the most influential feature, followed by irradiance and thermal exergy loss in terms of performance degradation impact. SHAP analysis similarly confirms the dominance of R_d while additionally highlighting the thermodynamic importance of exergy loss at the prediction level. Thus, ANOVA, ablation, and SHAP collectively provide complementary validation of feature importance.

The convergence of physical reasoning, statistical significance testing, ablation sensitivity, and SHAP-based interpretability strengthens the robustness of the proposed modelling framework. This multi-layered validation confirms that PR variability in the investigated PV system is governed predominantly by degradation mechanisms and thermodynamic losses, rather than irradiance magnitude alone.

6.3 Development of LSTM-based PR Prediction Model - PRLSTM

The proposed model architecture utilizes time-series inputs namely, degradation rate (R_d), thermal exergy loss, and solar irradiance fed into a sequential deep learning framework based on Long Short-Term Memory (LSTM) networks.

LSTM is selected over traditional recurrent neural networks (RNNs) due to its ability to mitigate the vanishing gradient problem and effectively capture long-range temporal dependencies.

As illustrated in figure 6.14, the model comprises four stacked LSTM layers, each consisting of 256 memory units, followed by a dense output layer with a single neuron for final prediction with 2 input stamp (t-1) and (t) and one output at (t). The model is trained using the Adam optimizer and Mean Absolute Error (MAE) as the loss function, over 1000 epochs with a batch size of 32.

The objective of the training process is to minimize the discrepancy between the predicted and actual performance ratio (PR) of the 191.9 kWp PV plant, thereby enhancing predictive accuracy under real-world operational conditions.

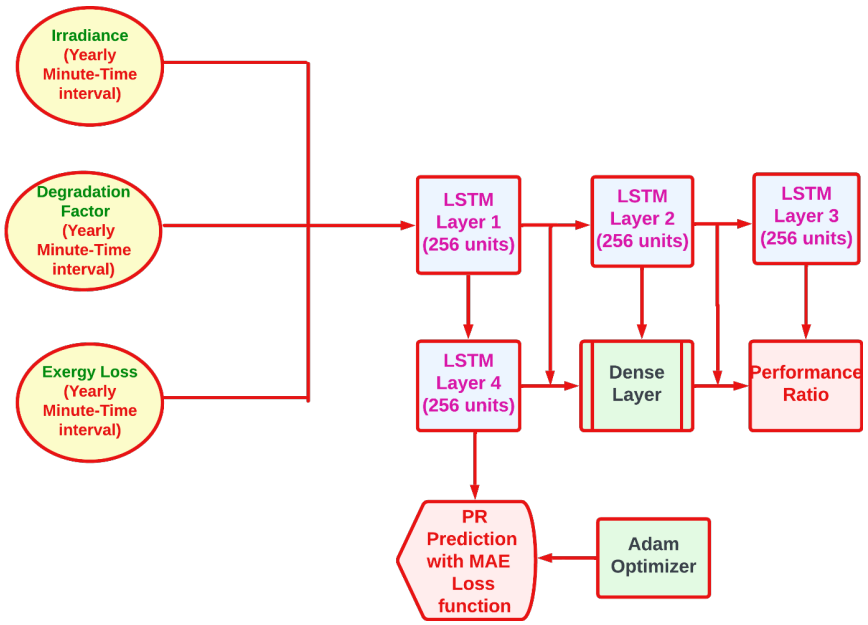


Fig.6.14 Architecture of the proposed LSTM-based model (PR LSTM) for short-term prediction of Performance Ratio (PR)

Each LSTM layer in the proposed model incorporates key components namely the cell state, forget gate, input gate, and output gate as illustrated in figure 6.14. These components interact through operations such as pointwise addition, multiplication,

and activation functions (sigmoid and hyperbolic tangent) to regulate the flow of information.

The cell state is updated by combining the retained information from the previous state $S_{(t-1)}$ scaled by the forget gate $F_{(t)}$, and the new candidate information $\zeta_{(t}$ modulated by the input gate $I_{(t)}$. This update mechanism, governed by equation (6.4), enables the network to preserve long-term dependencies and learn complex temporal patterns effectively.

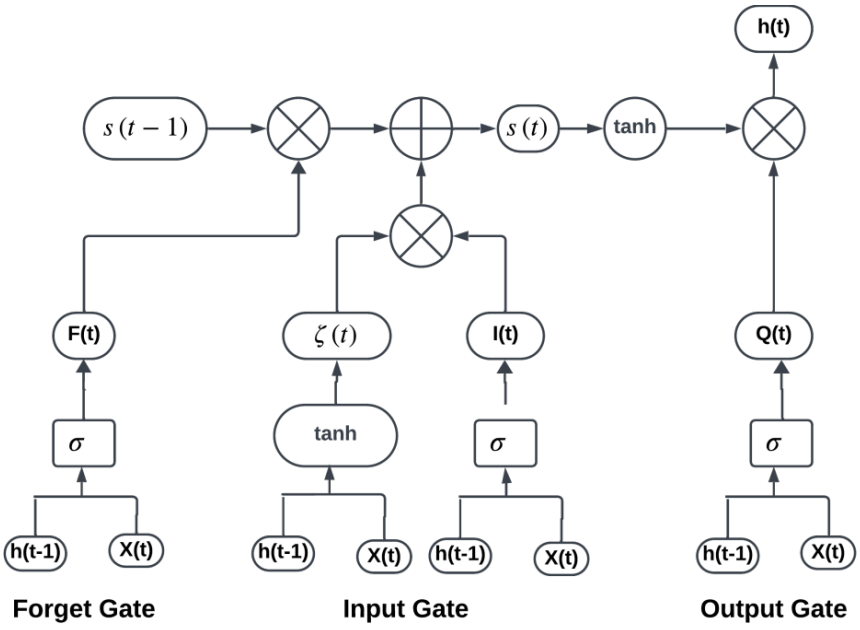


Fig. 6.15. Internal LSTM layer for forecasting (short -term minute-ahead) PR.

The forget gate plays a crucial role in resetting the internal memory of the LSTM cell. As defined in equation (6.5), the forget gate function $F_{(t)}$determines the proportion of the previous memory state to be retained or discarded at time step t as seen in figure 6.15.

$$s_{(t)} = F_{(t)}s_{(t-1)} + I_{(t)}\zeta_{(t)} \quad \text{Eq.(6.4)}$$

$$F_{(t)} = \sigma(b_f + \sum_j U_f x_{(t)} + \sum_j W_f h_{(t-1)}) \quad \text{Eq.(6.5)}$$

Here, b_f represents the forget gate bias vector, $x_{(t)}$ is the current input, $h_{(t-1)}$ is the previous hidden state, and U_f, W_f denote the weight matrices for input and recurrent connections, respectively.

The input gate computes a gating vector using the current input and hidden states to determine the new information to store in the cell state. A sigmoid activation function first normalizes the input and hidden states between 0 and 1, as given in equation (6.6).

Simultaneously, a candidate cell state $\zeta_t$is generated using a hyperbolic tangent function, as shown in equation (6.7)

$$I_{(t)} = \sigma(b_g + \sum_j U_g, x_{(t)} + \sum_j W_g h_{(t-1)}) \quad \text{Eq.(6.6)}$$

$$\zeta_t = \tanh\,(b + \sum_j U\, x_{(t)} + \sum_j W\, h_{(t-1)}) \quad \text{Eq.(6.7)}$$

The output gate then determines the final hidden state $h_{(t)}$, which is passed to the next time step. The output is a product of the cell state $s_{(t)}$ passed through a hyperbolic tangent and the output gate function $Q_{(t)}$ as seen in equation (6.8) and (6.9).

$$h_{(t)} = \tanh\,(s_{(t)})Q_{(t)} \quad \text{Eq.(6.8)}$$

$$Q_{(t)} = \sigma(b + \sum_j U\, x_{(t)} + \sum_j W\, h_{(t-1)}) \quad \text{Eq.(6.9)}$$

By implementing the LSTM operations described in equations (6.4–6.9), the proposed model performs minute-ahead forecasting of the Performance Ratio (PR). It uses current and preceding time-step inputs such as degradation rate, exergy loss, and irradiance to predict the PR at the next time instance, thus enabling short-term, high-resolution forecasting.

To enhance prediction accuracy, the model undergoes rigorous hyperparameter tuning. A dropout regularization of 0.4 is applied to mitigate overfitting by

randomly disabling a fraction of the neurons during training. The network depth is optimized to four LSTM layers to balance complexity with learning capacity for temporal dependencies. Transfer learning was not employed in this work. Instead, the trained models were directly evaluated on external datasets without re-training or cross-site weight reuse.

6.3.1 Comparison of PRLSTM with other sequential model approaches

To validate the effectiveness of the proposed PRLSTM model for short-term performance ratio (PR) prediction in grid-connected solar PV systems, a comparative analysis was conducted against other widely used sequential learning models namely, Gated Recurrent Unit (GRU) and Convolutional Neural Network (CNN). All models were evaluated on a common dataset comprising 1500 samples to ensure a consistent basis for comparison. As presented in table 6.3, the PRLSTM model outperforms both GRU and CNN across all key performance metrics namely the Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and Mean Bias Error (MBE).

The PRLSTM achieves the lowest RMSE of 0.0896 and a MAPE of 2.71%, indicating its superior capability to minimize both absolute and percentage prediction errors. Additionally, its moderate MBE value of 0.4324 suggests a well-balanced bias, avoiding consistent under- or over-predictions. In contrast, the CNN model performs the weakest, recording the highest RMSE (0.1291) and MAPE (3.80%). This is likely due to CNN's architectural emphasis on spatial rather than temporal features, making it less suitable for time-series modelling like PR prediction. The GRU model, while better than CNN, still falls short of LSTM's performance with a higher RMSE (0.0977), MAPE (3.11%), and an even greater

bias (MBE: 0.5879), indicating inconsistency in capturing temporal dependencies over the evaluated horizon.

Table 6.3. Comparative analysis of other sequential models in comparison to PRLSTM for Khopoli location

Sequential Models	RMSE	MAPE	MBE
PRLSTM (proposed)	0.0896	0.0271	0.4324
CNN	0.1291	0.0380	0.1334
GRU	0.0977	0.0311	0.5879

Thus, this comparative evaluation reaffirms that the LSTM-based approach is not only more accurate but also more stable and reliable for capturing the complex time-series dynamics of performance ratio. Its capacity to retain long-term dependencies, handle vanishing gradients, and model temporal variations makes it the most appropriate model among the tested architectures for short-term PR forecasting.

6.4 Results and Discussion for PR assessment

6.4.1 Performance of the PRLSTM during training and testing

Short-term prediction using the PRLSTM model is visualized in Figure 6.14, where the predicted PR closely tracks the actual values during the training phase which comprises of 21020 samples. The model is trained using a two-time-step input window (n_input = 2) comprising significant attributes like degradation rate (R_d), exergy loss, and irradiance (I) to predict the performance ratio (PR) at the current time step (t). Each input sample comprises real-time operational parameters namely degradation rate (R_d), exergy loss, and irradiance (I). These inputs exhibit natural fluctuations due to environmental and system conditions throughout the day, which the model learns to associate with short-term changes in PR. The dynamic behavior

of the predicted and actual PR values shows strong alignment, with low error metrics evaluated in terms of MBE of 0.0231, MPE of 0.0831, and RMSE of 0.0726 as reported in table 6.4. This is further confirmed by the tight clustering of predicted and actual values in figure 6.16, indicating high model precision within the observed operational envelope.

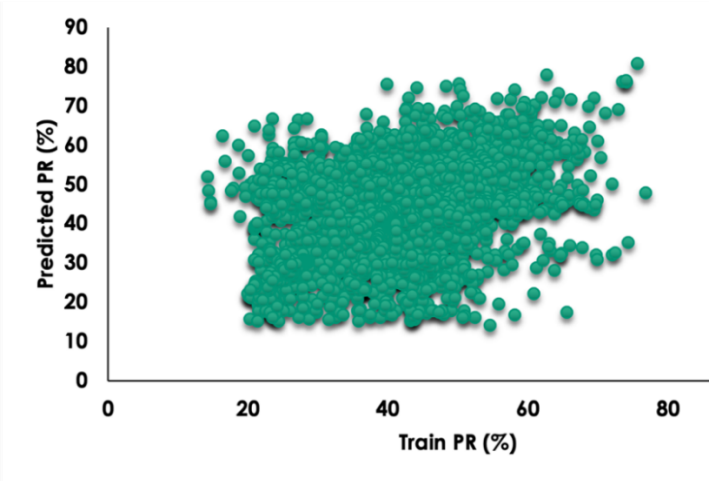


Fig.6.16. Comparison between predicted and actual performance ratio (PR) during training using the proposed PR LSTM model

During the testing phase, the model is evaluated on 4,204 previously unseen samples. The PR LSTM model achieves an MBE of 0.0211, indicating a minor underestimation tendency. Additionally, it maintains a low RMSE of 0.0961 and MAPE of 0.0938, confirming robust generalization capabilities. The predictive behaviour of the proposed PR LSTM model is evaluated using test input sequences consisting of two time steps in alignment with the training input structure to predict the performance ratio at time (t). As shown in figure 6.17, the predicted values form a compact distribution around the diagonal line, reinforcing the model's efficacy as a reliable tool for real-time PR estimation.

The proposed PR LSTM model exhibits consistently strong performance during both the training and testing phases. Its low Root Mean Square Error (RMSE) demonstrates the model's ability to closely track the actual performance ratio (PR) values with minimal deviation.

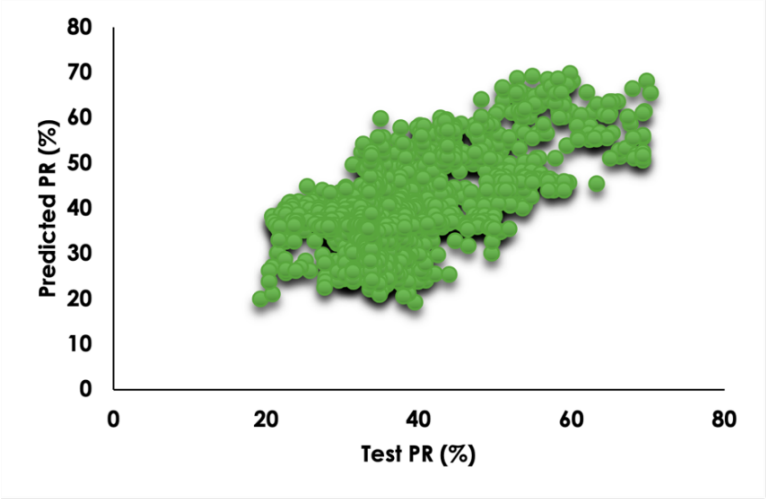


Fig.6.17. Comparison between predicted and actual performance ratio (PR) during testing using the developed PR LSTM model.

Table 6.4. Model performance metrics for PR LSTM during stages of Training and Testing for Khopoli

PRLSTM MODEL	RMSE	MBE	MAPE
Training	0.0726	0.0231	0.0831
Testing	0.0961	0.0211	0.0938

The Mean Bias Error (MBE) and Mean Absolute Percentage Error (MAPE) further reinforce its predictive robustness by indicating low average bias and percentage error across the dataset. These characteristics make the model a reliable choice for modeling dynamic, nonlinear dependencies within solar photovoltaic (PV) systems.

From a computational standpoint, the training duration on a standard MacBook (approximately 8 hours) reflects the model's complexity and the constraints of personal computing resources. However, this limitation is not a barrier in real-world deployment, especially for industries equipped with high-performance computing infrastructure. With access to GPUs like NVIDIA A100 or RTX 4090, or specialized accelerators such as Tensor Processing Units (TPUs), training time can be reduced drastically to minutes through hardware-level parallelization and cloud-based distributed learning platforms (e.g., AWS EC2, Google Cloud TPUs, Google Colab Pro).

Moreover, after the initial training, the PRLSTM model exhibits low inference latency, meaning it can deliver real-time or near-real-time predictions with minimal delay as it is trained and tested with field based inputs. This capability is critical for smart grid operations, automated monitoring systems, and real-time fault detection in solar power plants. In contrast, traditional tools such as PVsyst or analytical correction models like PRcorrected, though simpler and faster to execute, rely on predefined rules or simplified assumptions. These lack the adaptive learning mechanisms of LSTM, making them less responsive to on-site variability such as shading, weather fluctuations, or thermal inefficiencies.

Therefore, while the LSTM approach incurs a higher computational load during model development, its high prediction accuracy, fast inference, and adaptability to temporal patterns make it exceptionally well-suited for modern, scalable, and intelligent solar PV system monitoring and optimization frameworks.

6.4.2. Performance comparison of PRLSTM with other models

To validate the predictive capability of the proposed PRLSTM model, a comparative analysis was conducted against established benchmark methods including PVsyst, PRcorrected, and a reported Feedforward Neural Network (FFNN) based PR prediction model, all evaluated for the same geographic location Khopoli, India. The comparison is based on Mean Absolute Percentage Error (MAPE), a widely accepted metric for evaluating forecasting accuracy in PV performance studies.

The PRcorrected model, which adjusts PR using analytical corrections and loss assumptions, exhibits the poorest predictive accuracy with a MAPE of 22.50 %. This high error percentage highlights the model's inability to capture nonlinear operational losses, dynamic weather variations, and degradation factors that influence real-time PR behavior. This is also visually substantiated in figure 6.16, where a significant deviation is observed between the predicted and actual PR values with MBE of -0.0604, reflecting the model's limited capacity to track intermittent variability in true loss factors.

In contrast, the FFNN-based model and PVsyst show moderate improvements, with MAPE values of 12.43% and 11.85%, respectively. While these models outperform PRcorrected, they still fall short in precisely capturing the complex temporal dependencies and thermodynamic parametric variations that are inherent in solar PV systems. PVsyst, although widely used for system design and simulation, operates on static assumptions and simplified loss models, whereas FFNN, being a shallow network, lacks memory and struggles with sequential forecasting in time series data.

The proposed PR LSTM model, on the other hand, demonstrates superior learning capability and prediction accuracy. It not only achieves the lowest MAPE among all models as seen in table 6.5, but also closely tracks the actual PR trajectory over time, as inferred from figure 6.18. The close alignment between the predicted and actual PR values highlights the PRLSTM model's strength in capturing long-term dependencies, dynamic environmental factors, and degradation behaviours, all of which are critical for accurate short-term forecasting.

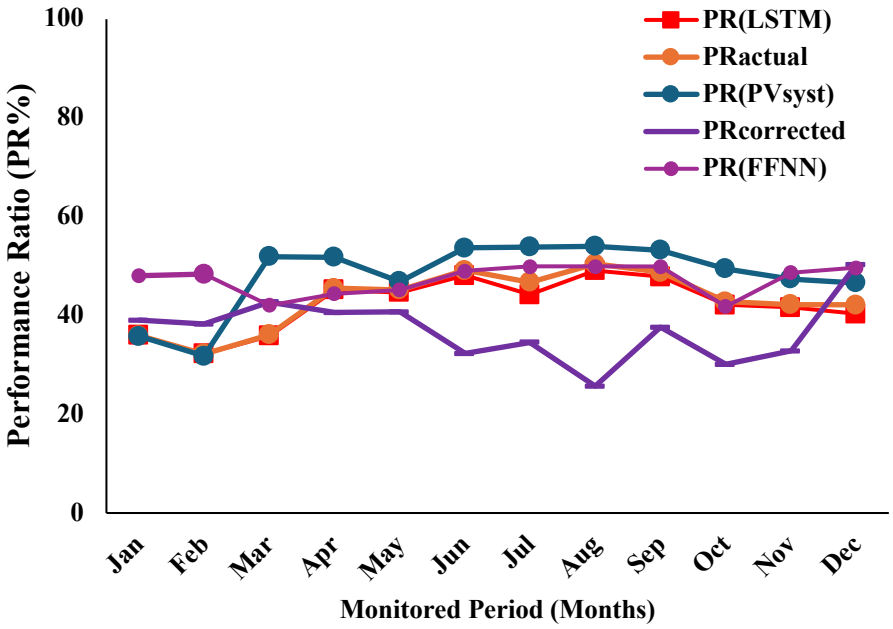


Fig. 6.18. Comparative evaluation of the proposed PR LSTM model against benchmark models for PR estimation at the Khopoli site.

This comparative evaluation reaffirms that the proposed model, owing to its memory-based learning and temporal modelling capacity, offers a substantial improvement over traditional deterministic and shallow learning models, making it a robust and scalable tool for real-time PR estimation in field-deployed PV systems. To enhance interpretability and provide a macro-level understanding of the model's performance trends, the minute-level training data spanning one year was

temporally aggregated to monthly averages for visualization purposes in figures and comparative analyses.

Table 6.5. Comparison of Errors of PRLSTM (proposed) and reported models for Khopoli Location

PR Models	RMSE	MAPE	MBE
PR LSTM (Proposed)	0.0110	0.0183	-0.0082
PVsyst	0.0637	0.1185	0.04944
PRcorrected	0.1161	0.2250	-0.0604
PR(FFNN)	0.0681	0.1243	0.0415

However, the PR LSTM model was developed, trained, and validated using unaggregated minute-level data, ensuring the model captures high-frequency temporal dependencies and subtle fluctuations in system behaviour.

6.4.3 Performance comparison of PRLSTM during validation

To rigorously assess the generalizability and predictive strength of the proposed PR LSTM model, two independent validation exercises were conducted using datasets from geographically distinct and previously unseen locations Bengaluru, India, and Köprübaşı, Turkey (Murat.et.al.,2021). The external validation was designed to assess the robustness and applicability of the proposed PRLSTM framework across distinct climatic regimes beyond the primary training location.

a. The **Koprubasi, (Turkey)** site represents a Mediterranean climate with pronounced seasonal irradiance variation and comparatively lower operating temperatures. This contrast enables evaluation of model

behaviour outside the climatic envelope of the primary study site, thereby strengthening the assessment of model generalizability.

b. **Data source and instrumentation comparability**

The Turkey dataset was obtained from a peer-reviewed published study (Murat *et al*., 2021), while the Indian datasets were acquired from on-site plant monitoring systems. Although sourced independently, the datasets comprise equivalent attributes namely irradiance, ambient and module temperature, and energy output—measured in accordance with IEC 61724–based PV monitoring practices, enabling meaningful comparison at the system-performance level

The first validation was based on on-field data from a 125 kWp rooftop PV system in Bengaluru for the year 2021. During this phase, the PR LSTM model achieved a low MAPE of 5.81%, demonstrating its effectiveness in accurately predicting the performance ratio. This high prediction accuracy is also evident in figure 6.17, where only marginal deviations between predicted and actual PR are observed across major of the observation.

To benchmark its performance, the PRLSTM model was compared with conventional tools including the PVsyst simulation tool (Ravi.et.al.'2020), the corrected PR model (PRcorrected) (Gopi.et.al.,2021) and a Feedforward Neural Network-based PR predictor (PR-FFNN) (Shapsough et.el.,2019). As seen in figure 6.19, the PRcorrected model exhibits substantial deviations from the actual PR, particularly during January–April and July–September, resulting in higher error values (MBE: -0.0853, RMSE: 0.0779, MAPE: 17.59%). Similarly, the PR-FFNN model displays poor alignment with actual PR, especially between April and June, yielding a high MAPE of 12.43%. The PVsyst model also underperforms in comparison, with a MAPE of 13.42%.

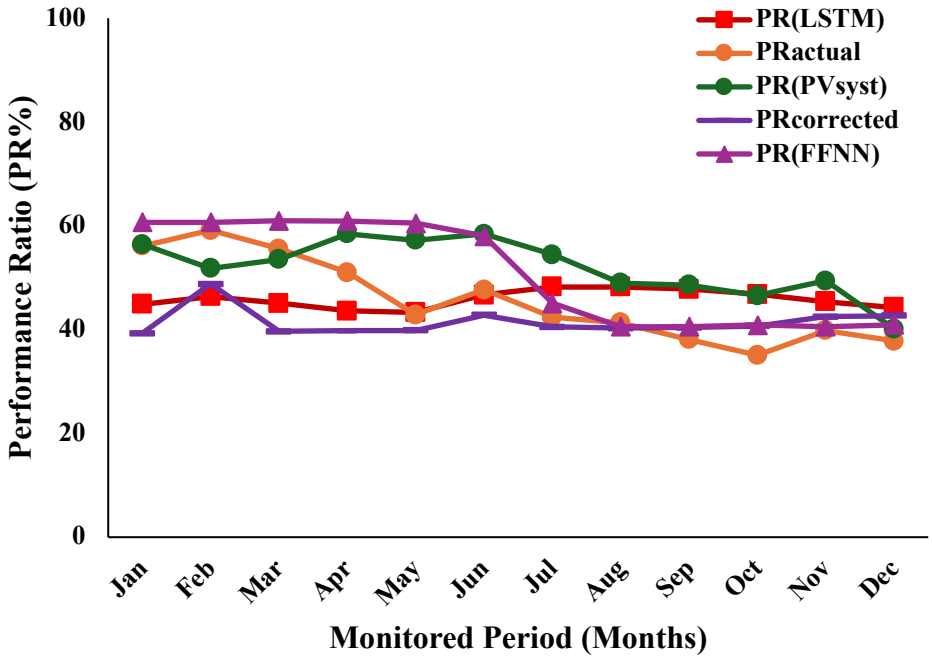


Fig.6.19. Comparison of PR LSTM Model with PVsyst, PR corrected and PR (FFNN) models for Bengaluru, India location.

In contrast, the PR LSTM model demonstrates the highest consistency and lowest bias (MBE: -0.0247), clearly outperforming the other methods in both predictive accuracy and robustness. These findings are further confirmed by the close alignment between predicted and actual values illustrated in table 6.6, validating the PRLSTM model's suitability for reliable, short-term, minute-ahead PR prediction in diverse real-world settings.

Table 6.6. Comparison of Errors of PRLSTM (Proposed) and other benchmark-based models for Bengaluru, India Location.

PR Model	**RMSE**	**MAPE**	**MBE**
PRLSTM (Proposed)	0.0537	0.0581	-0.0247
PVsyst	0.0647	0.1342	-0.0770
PRcorrected	0.0779	0.1754	-0.0853
PR (FFNN)	0.0681	0.1243	0.0451

To further evaluate the generalizability of the proposed PR LSTM model, a second validation study was conducted on a 30 kWp rooftop PV installation located at Koprübaşı Vocational School in Turkey. The system comprises 116 polycrystalline solar modules mounted 15 cm above the roof, with a tilt angle of 12° and an azimuth orientation of 20°.

The actually reported annual performance ratio (PR) for this plant was 83.61%, as depicted in figure 6.20. Despite the geographical and climatic differences between Turkey and the model's training region, the PRLSTM model demonstrated strong predictive capability with a closer agreement existing between the predicted PR and the actual. It achieved the lowest error metrics among all the reported approaches with RMSE of 0.0137, MAPE of 1.48%, and MBE of 0.0815 as shown in table 6.7, closely matching the actual measured PR.

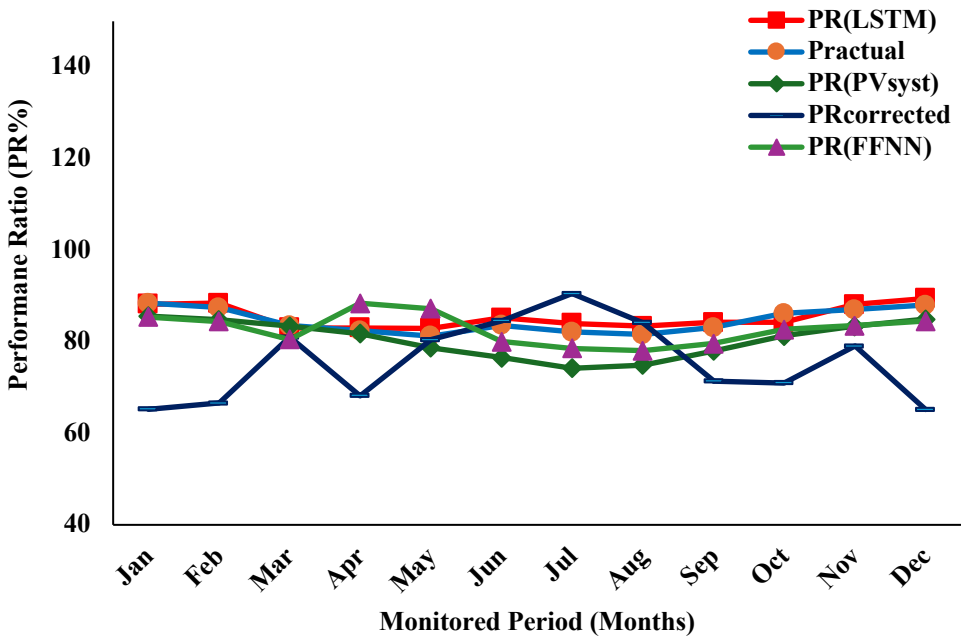


Fig.6.20. Comparison of PR LSTM Model with PVsyst, PR corrected and PR (FFNN) models for Koprübaşı, Turkey location

In contrast, the PRcorrected model produced the highest deviation, with a MAPE of 12.71%. The PR-FFNN and PVsyst methods also showed comparatively higher errors, with MAPEs of 4.81% and 4.65%, respectively as seen in table 6.7. The superior performance of the PR LSTM model can be attributed to its incorporation of critical features such as R_d, thermal exergy loss, and irradiance, as well as effective hyperparameter tuning that enhances learning from temporal patterns. Although the validation is constrained by the limited availability of uniform inputs across all models, the PR LSTM consistently captures the dynamic variation in attributes and maintains an acceptable prediction error, thereby affirming its suitability for short-term PR forecasting in diverse operating conditions.

This comparison demonstrates that photovoltaic system performance is governed not solely by the magnitude of incident irradiance, but by the combined influence of ambient temperature, humidity, air density, and wind-induced convective cooling, which together determine thermal losses and the available electrical exergy.

Table 6.7. Comparison of Errors of PRLSTM (Proposed) and other benchmark-based models for Koprubaşı, Turkey Location.

Model	**RMSE**	**MAPE**	**MBE**
PRLSTM (Proposed)	0.0137	0.0148	0.0815
PVsyst	0.0456	0.0465	-0.1916
PRcorrected	0.1356	0.1271	0.0892
PR (FFNN)	0.0428	0.0481	-0.0912

The observed differences in PV performance across Mumbai, Telangana, and Bengaluru are primarily attributed to site-specific climatic conditions. In Mumbai, high coastal humidity and increased air density enhance atmospheric attenuation and thermal retention, leading to elevated module temperatures and increased

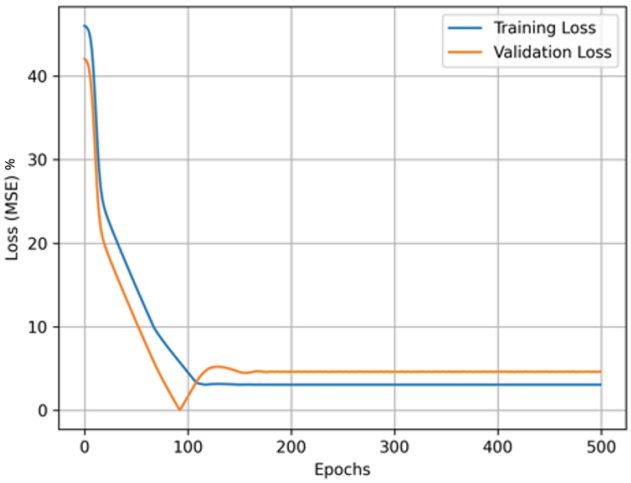
optical and thermal losses. Telangana, characterized by high irradiance and elevated ambient temperatures, experiences intensified thermal exergy destruction and temperature-induced voltage losses, despite favorable solar availability. In contrast, Bengaluru benefits from relatively moderate temperatures and higher wind speeds, which enhance convective cooling, reduce module operating temperature, and consequently improve both electrical performance and exergy efficiency. The proposed framework of PR LSTM is intended for performance prediction under operating conditions similar to those represented in the training data. Its accuracy will be impacted under extreme climatic conditions, sensor anomalies, abrupt system changes, or long-term extrapolation beyond the training horizon.

6.4.4. Training Convergence and Learning Stability Analysis

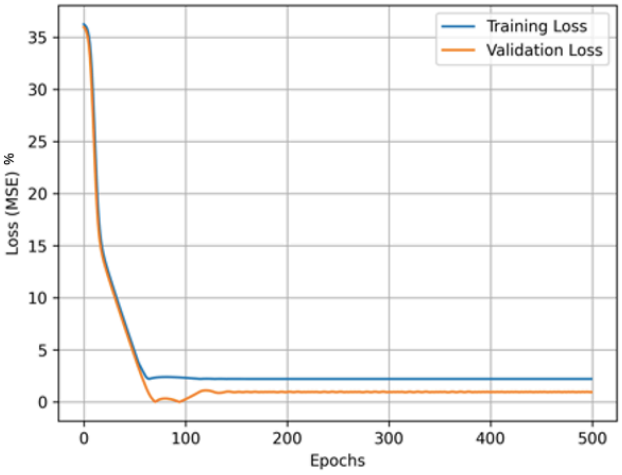
Figure 6.21 illustrates the training and validation loss convergence of the PR–LSTM model for Khopoli, Bengaluru, and Turkey locations. For all sites, the model exhibits rapid loss reduction during the initial epochs followed by stable convergence, with no evidence of divergence between training and validation losses.

Convergence is achieved within approximately 60–80 epochs, indicating effective learning of temporal dependencies in irradiance, temperature, wind speed and A (temperature correction ratio). The consistent convergence behaviour across geographically and climatically diverse locations confirms the robustness and generalization capability of the proposed PR–LSTM model.

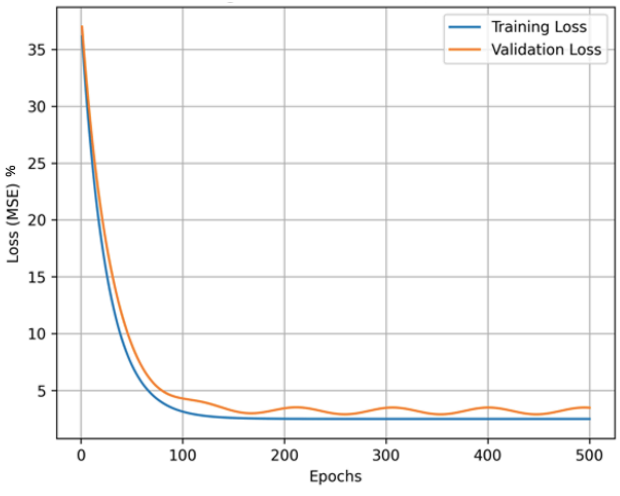
The learning stability of the model is further substantiated through loss–epoch convergence plots, which demonstrate consistent training behavior across all validation locations. Trained models were directly evaluated on external datasets without Re-training.



(a)



(b)



(c)

Figure 6.21 (a) PR–LSTM loss vs epochs – Khopoli (training site)(b): PR–LSTM loss vs epochs – Bengaluru (validation site)(c): PR–LSTM loss vs epochs – Turkey (validation site)

6.4.5. Conditions Affecting Predictive Accuracy of the Proposed Framework

1.Out-of-distribution climatic conditions:

The framework is trained primarily on datasets from specific climatic regimes namely tropical semi-arid and warm temperate (Mediterranean-type), corresponding to the Indian (Bengaluru, Telangana) and Turkish validation sites. When applied to locations with significantly different irradiance spectra, temperature extremes, or atmospheric conditions (e.g., high-latitude or snow-dominated regions), predictive accuracy may degrade due to distribution mismatch. However, since the model inputs (irradiance, ambient/module temperature, temperature difference, and electrical output) are physically meaningful and derived from on-field measurements, the framework can still provide reasonable first-order estimates, provided the operating conditions remain within physically justifiable bounds.

2.Sensor faults or prolonged data gaps:

The proposed approach relies on continuous and reliable measurements of irradiance, temperature, and electrical output. Sensor malfunction, calibration drift, or extended missing data periods can propagate errors into derived performance indicators and, consequently, into AI-based predictions.

3.Abrupt system-level changes:

Sudden changes such as inverter replacement, reconfiguration of array layout, major maintenance interventions, or retrofitting of components introduce non-stationarity that is not captured in historical training data, potentially reducing short-term prediction accuracy.

4.Severe or localized failure events:

Although temperature difference is provided as a physically meaningful input to the R_d model and is consistently used in the PR and exergy prediction frameworks, the proposed approach is primarily designed to capture system-level and gradual degradation behaviour. Localized or abrupt failure events such as cracked modules, hot spots, or connector faults affecting a limited subset of strings may not be fully reflected in aggregated temperature or performance indicators unless they induce a sufficiently strong and persistent thermal or electrical signature in the measured data. Consequently, prediction accuracy may be reduced for such isolated failures unless explicitly represented in the training dataset or monitored through string-level or module-level diagnostics.

6.5. Summary

This chapter proposed and validated a comprehensive deep learning-based framework for predicting the Performance Ratio (PR) of a grid-connected solar PV system by incorporating failure mode-based degradation (R_d), thermal exergy loss dependent on exergy efficiency, and irradiance as critical inputs. The system under study, a 191.9 kWp PV plant, was first analyzed to understand site-specific energy patterns, followed by identification of statistically and physically significant predictors. A custom LSTM architecture was developed to forecast minute-level PR, outperforming benchmark models like PVsyst, PRcorrected, and Feedforward Neural Network (FFNN) in both training and testing phases, with minimal error margins (MAPE as low as 2.71%). Ablation studies further highlighted R_d as the most influential input. Validation across two distinct locations Bengaluru, India, and Koprübaşı, Turkey confirmed the model's generalizability, achieving MAPE values of 5.81% and 1.48% respectively. Overall, the PRLSTM model

demonstrated superior predictive capability, adaptability, and reliability, making it a viable tool for real-time performance monitoring and early fault detection in operational PV plants.

- Section 6.1 provided an overview of the 191.9 kWp PV system under study, located in Khopoli, Maharashtra. Seasonal variations in irradiance and temperature were analysed to highlight their impact on energy output and performance ratio (PR). It was shown that higher module temperatures tend to reduce PR due to increased thermal losses.
- Section 6.2 described key factors influencing PR and were identified, specifically as failure mode-based degradation rate (R_d), thermal exergy loss, and global irradiance.

(a) Visual and statistical analysis confirmed the significant relationship between PR and the selected inputs, with R_d having the strongest correlation.

(b) ANOVA results validated the statistical importance of the inputs, with all p-values < 0.05 and a high R^2 value (0.9366), confirming their strong influence on PR variability.

(c) SHAP-based feature importance analysis further substantiated these findings by quantifying the model-specific contribution of each input to PR prediction. The global SHAP ranking identified R_d as the most influential parameter, followed by thermal exergy loss and irradiance. The beeswarm distribution additionally revealed the directional impact of the features, showing that higher R_d and increased exergy loss contribute negatively to PR. The consistency between statistical testing, ablation analysis, and SHAP

interpretation strengthens the reliability and physical validity of the proposed PRLSTM framework.

- Section 6.3 described a four-layer LSTM model was designed to predict minute-based PR using R_d, exergy loss, and irradiance as time-series inputs using two-minute historical time-series inputs of R_d, exergy loss, and irradiance. The minute-based variations are averaged to estimate daily PR and subsequently averaged to form monthly average daily PR as shown in figure 6.16 to 6.18.

(a) The LSTM network structure, training methodology, and hyperparameter tuning were described. The model used MAE loss and Adam optimizer with dropout regularization for generalization.

(b) Comparison of PRLSTM with Other Sequential Models: PRLSTM was benchmarked against GRU and CNN models which is believed to be adaptive for intermittent applications. PRLSTM achieved the lowest RMSE (0.0896) and MAPE (2.71%), proving its superiority in learning temporal dependencies among the GRU and CNN model approaches.

- Section 6.4 analyzed the model's predictive performance and generalizability.

(a) The model showed strong agreement between predicted and actual PR values with low error metrics. Visualization plots reinforced its consistency and learning stability.

(b) A predictive performance error comparison was carried out on employing reported approaches like with PVsyst, PRcorrected, and FFNN for PR estimation. Compared with PVsyst, PRcorrected, and FFNN, the PRLSTM

model consistently delivered better accuracy across RMSE, MAPE, and MBE.

(c) Validation was carried out using long-term independent datasets from Bengaluru, India and Köprübaşı, Turkey. The PRLSTM model generalized well, achieving MAPE values of 5.81% and 1.48%, significantly outperforming benchmark tools and models.

Chapter 7: Graphical User Interface for Assessment of PR

Overview:

This chapter presents the development, functionality, and comparative evaluation of a Graphical User Interface (GUI)-based tool built upon the proposed PRLSTM model for short-term performance ratio (PR) forecasting in photovoltaic (PV) systems. Designed with user-centric simplicity and real-time forecasting capabilities, the tool integrates three sequential computation frames input preprocessing, derived parameter estimation (thermal exergy loss and R_d), and PR prediction using LSTM. It also incorporates visual analytics and performance evaluation features. Comparative insights with benchmark tools such as PVsyst, SAM, and SolarFarmer are provided to assess the depth, adaptability, and forecasting utility of the proposed GUI. The chapter concludes by highlighting its operational feasibility for real-time industrial use.

Outline of Chapter ***Page***

7.1 Proposed GUI Description

The proposed PR-LSTM model, as described in Chapter 6, is operationalized into a user-friendly graphical interface capable of forecasting performance ratio (PR) over a minute-ahead time horizon. In alignment with its training resolution, the deployed Graphical User Interface (GUI) retains its original capability to generate minute-ahead predictions. Moreover, the model architecture is scalable to longer forecasting horizons such as 5-minute, 15-minute, or hourly intervals by adjusting the input sample size and retraining with appropriately resampled time-series data. ,thereby supporting real-time operational decision-making.

The GUI-based PR analyser is developed using Python's Tkinter library, as illustrated in figure 7.1. To enhance usability, the workflow of PR analyser is segmented into three computational stages.

The first stage allows users to provide directly measurable or easily accessible parameters as inputs. The second stage automatically computes derived parameters namely thermal exergy loss and the degradation rate (R_d) from the fed fundamental inputs. In the final stage, the tool utilizes these processed inputs to predict minute-ahead PR using the pre-trained PRLSTM model as described in chapter 6.

To simplify user interaction, the fundamental input parameters required in Stage 1 of operation, include plane-of-array solar irradiance, geographic coordinates (latitude and longitude), ambient temperature, module temperature, wind speed, PV module dimensions, open-circuit voltage, and short-circuit current as shown in figure 7.2.

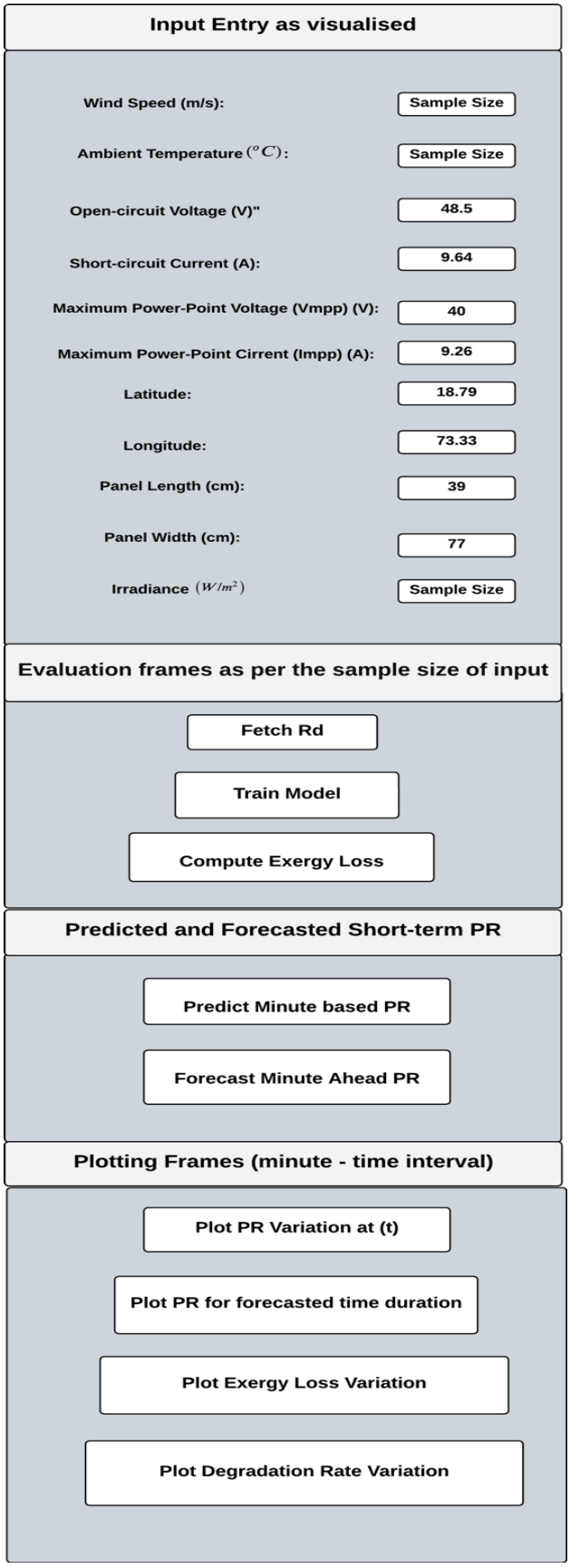


Fig.7.1. PR analyser -Performance Ratio GUI Calculator

Upon receiving these values, the analyser transitions to Stage 2, where it invokes a trained LSTM-based estimator in-line with architecture as discussed in Chapter 4 section 4.4.1 to predict R_d. Concurrently, thermal exergy loss is computed using time-series input data like wind speed and irradiance integrated with thermodynamic models (5.2–5.11), as presented in Chapter 5. These are directly embedded within the Tkinter interface. In Stage 3, the estimated values of R_d, thermal exergy loss, and recorded irradiance are collectively passed into the PRLSTM predictor module for real-time, short-term forecasting of PR under the "Train PR model" operation. This modular design enhances interpretability, minimizes manual computation, and supports efficient on-field PR estimation without prior energy estimation.

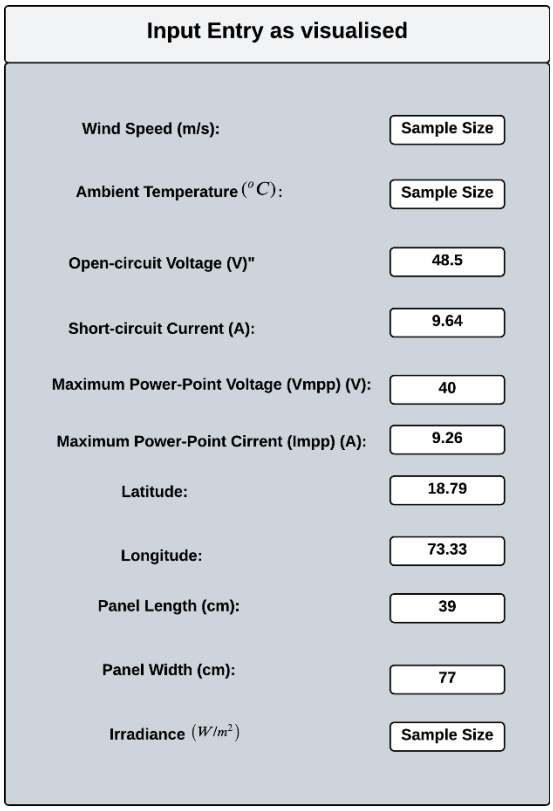
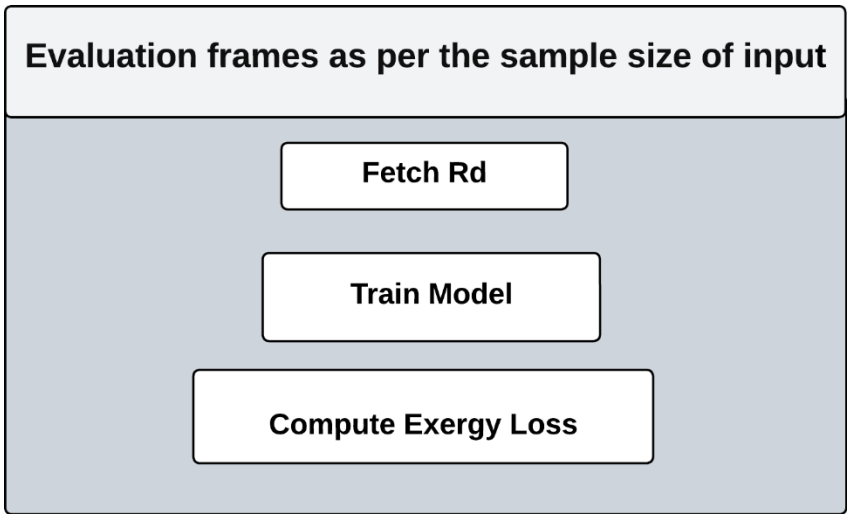


Fig.7.2. GUI Frame showing the Entry of Input parameters for early estimation of PR

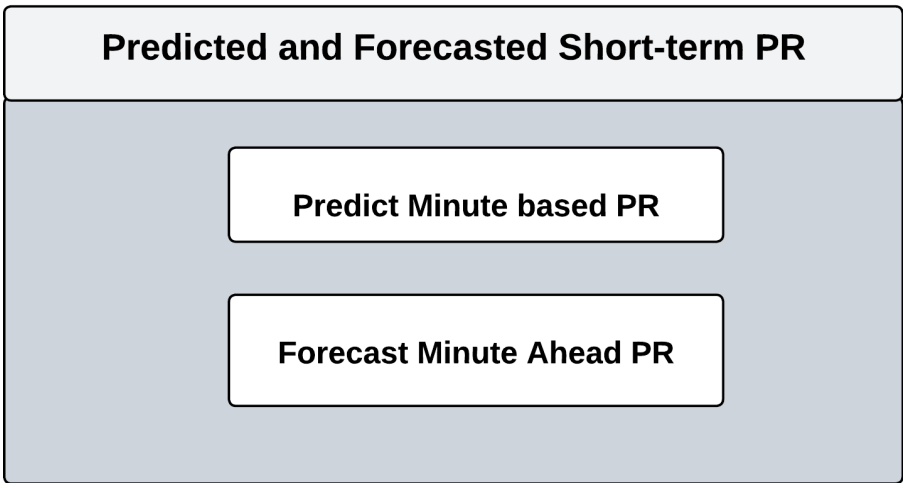
7.2 Evaluation frames for estimation of PR

The Train PR Model interface, illustrated in figure 7.3 (a) and (b), incorporates a built-in sequential workflow designed for the forecasting of performance ratio (PR) using the developed LSTM architecture.

1. The "Train Model" button initiates the training of the PRLSTM model employing the model architecture as seen in figure 6.14 of Chapter 6 and evaluates minute based PR. The Train PRLSTM uses the direct user fed inputs from Stage 1 (fundamental parameters as seen in figure 7.2) and evaluated inputs as in Stage 2 (derived attributes such as R_d and thermal exergy loss).
2. The "Forecasts Minute-Ahead PR" function utilizes the trained LSTM model to forecast the performance ratio on a minute-ahead time horizon, providing near real-time insights into the operational efficiency of the PV plant under investigation. An aggregator based recursive forecasting method is used for forecasting PR as described in sub-section (b) of Section 7.3.



(a)



(b)

Fig.7.3 (a) Frames to calculate Thermal exergy loss, Train Model for PRLSTM at time (t) and fetch R_d and (b) Predict minute based PR and Forecast minute ahead PR

7.3 Visualisation Frame of derived inputs and output

The developed PR analyser tool includes a dedicated data visualization module, which enhances user understanding by graphically representing trends and variations in key input and output parameters. This visualization functionality is facilitated through three distinct plotting frames, as illustrated in figure 7.4. These frames are seamlessly integrated within the GUI and allow users to interactively observe the variation of derived input factors such as degradation rate (R_d) and thermal exergy loss and its impact on the predicted performance ratio (PR) over computed time-horizon.

a. Plot Predicted PR

This visualization frame computes the minute-based predicted value of PR at the current time step t by using a two-time-step input window comprising evaluated parameters degradation rate (R_d), thermal exergy loss, and irradiance. The model performs single-step prediction, estimating PR at time t solely based on inputs up to that point. All, these inputs are fed to the PRLSTM model executed via the Train

PRLSTM operation frame which computes the minute-wise variation of PR for the investigated period. The predicted variation of PR for the considered sample size of 21,419 minute wise observations is shown in figure 7.5 (a). Prediction of PR employing priorly estimable and easily accessible inputs is instrumental in evaluating the reliability of Solar PV system.

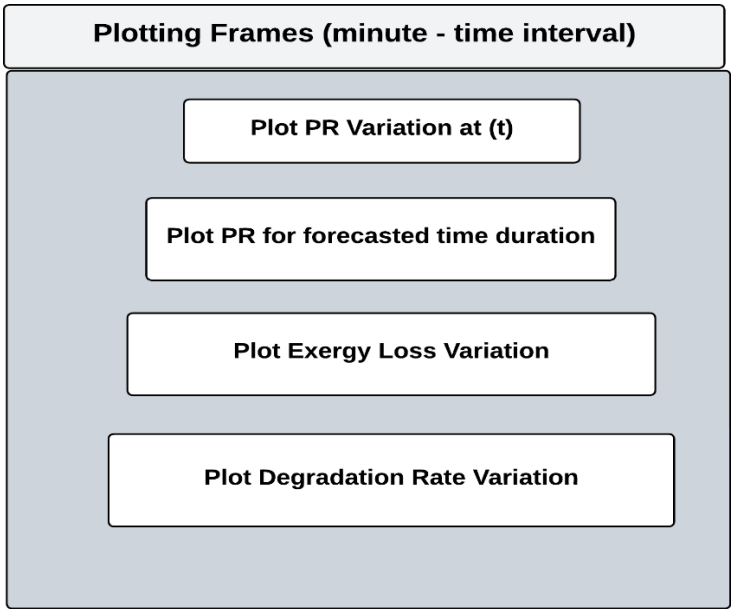


Fig. 7.4. Data Visualisation frames in the developed GUI

For solar analysts, it offers insight on the capability of the proposed model based GUI to capture system dynamics and operational variability. The predicted outcome can serve for technical performance tuning and long-term reliability analysis. So, the proposed PRLSTM predictive GUI depends on easily accessible inputs for prediction of PR as against the reported approaches which mainly depend on the monitored DC/ AC energy generation. The predictive model is also scalable to futuristically forecasts PR making it a priorly estimable indicator as described in sub-section ‘b’.

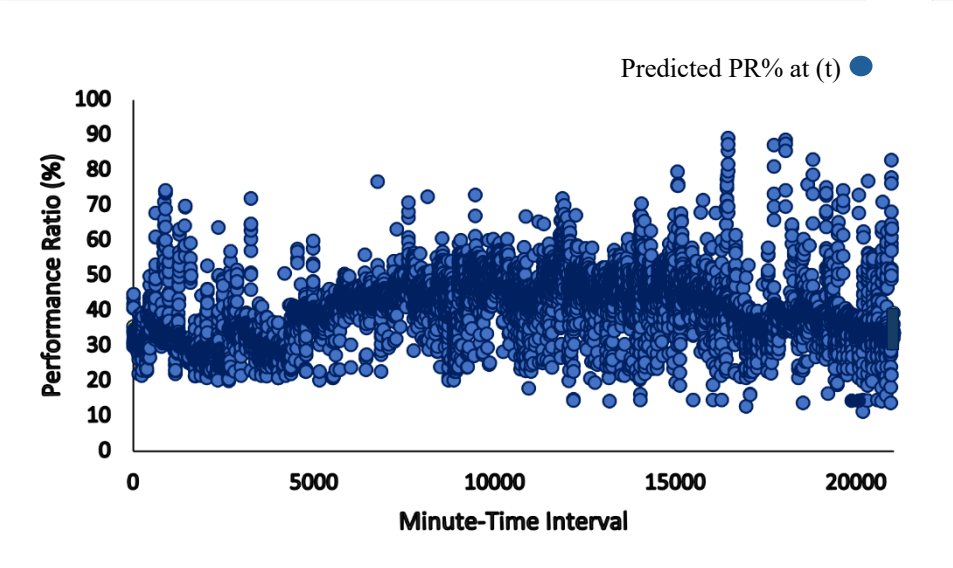
b. Short-Term PR Forecast

For forecasting, in-line with prediction, two-time-step input structure is employed. The forecasted minute ahead PR is then used as an as input to

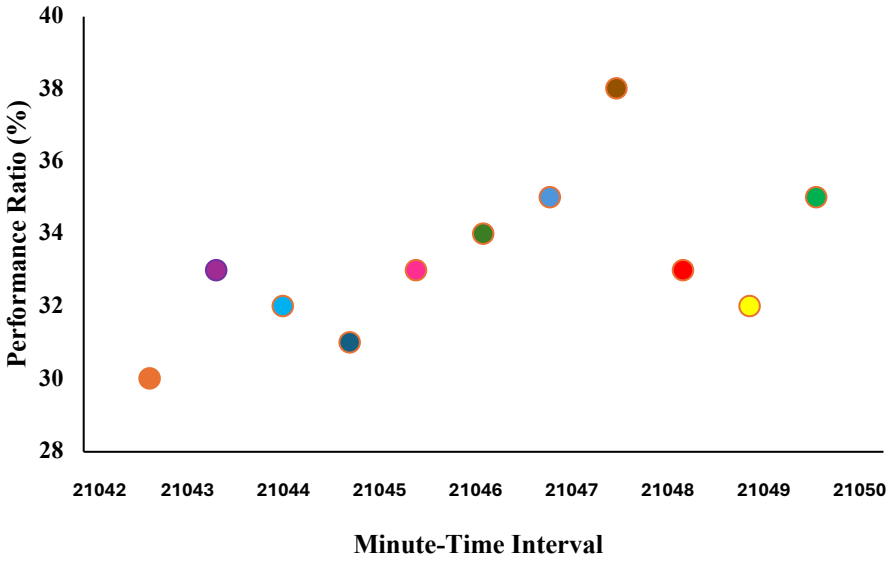
further forecast futuristic minute ahead PR at (t+2), and so on. This recursive forecasting approach allows the model to extend its horizon beyond immediate predictions while maintaining consistency with its learned temporal patterns. figure 7.5 (b) displays the short-term extended variation and minute-ahead forecast of PR value. In comparison to the existing assessments for PR, the proposed approach is capable for prior estimation of PR. This plot enables users to monitor rapid temporal fluctuations in performance, which may result from transient environmental conditions or operational anomalies. For system operators and energy economists, this functionality supports near-real-time decision-making and allows for proactive response to anticipated dips in energy yield. Moreover, forecasting PR at such a fine resolution is vital for short-term revenue projection and operational planning.

c. Minute-Based Thermal Exergy Loss Plot

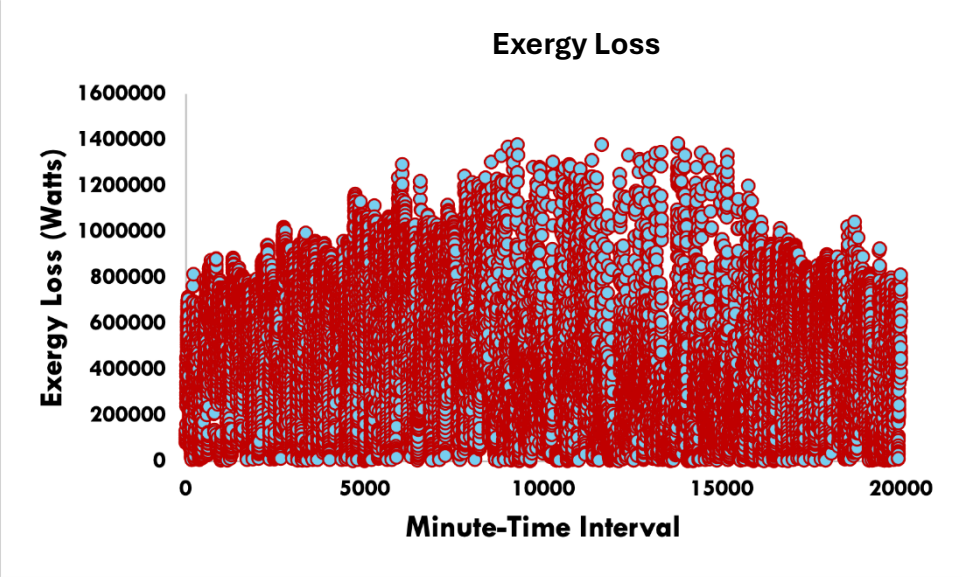
The third visualization frame, shown in figure 7.5 (c), presents a time-series view of thermal exergy loss across minute intervals. This plot is particularly relevant for evaluating the thermodynamic inefficiencies within the PV system. For technical analysts, a performance optimization-based decision making can be adopted for reducing the thermal irreversibilities, while for financial planners, it highlights opportunities to enhance energy conversion efficiency and reduce associated costs. Overall, understanding trends in exergy loss can directly inform system optimization strategies and improve overall profitability of solar energy projects.



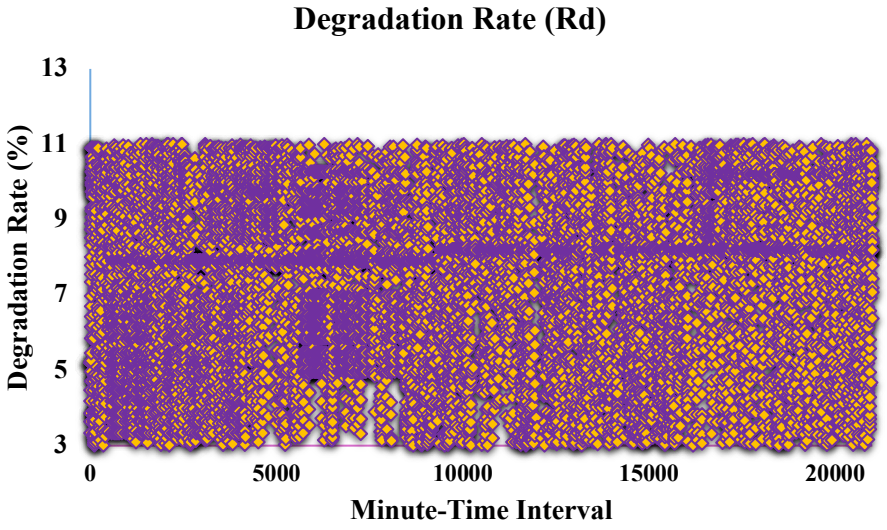
(a) Predicted PR at time(t)



(b) Forecasted PR% (t+1) to (t+11)



(c)



(d)

Fig.7.5. Results of Data Visualisation frames as viewed from developed PR Analyser GUI (a) Predicted PR at (t) (b) Forecasted minute-ahead PR (c)Short-term minute-based variation of thermal exergy loss (d) Short-term minute-based variation of failure mode-based power degradation rate (R_d)

d. Plot of Minute-Based R_d

The fourth visualization frame, illustrated in figure 7.5 (d), presents a time-series analysis of degradation rate (R_d) across minute-level intervals. This plot is essential for justifying the presence of failure modes due to environmental and operational stresses that influence power output. For technical stakeholders, such real-time visibility into R_d variations enable proactive maintenance planning and module

replacement strategies, particularly during periods of accelerated degradation. For asset managers and energy planners, the trends serve as early indicators of efficiency loss and operational anomalies, supporting informed decision-making for yield improvement. Overall, continuous monitoring of R_d not only strengthens system reliability but also enhances the long-term performance and financial viability of the PV installation.

Together, visualization components as discussed above transform the PR analyser into a comprehensive diagnostic and forecasting tool, bridging the gap between mathematical based technical modelling and economic evaluation in PV system analysis.

7.4 Comparison of PRLSTM GUI with other benchmark GUI

To evaluate the usability and forecasting capabilities of the developed PRLSTM GUI, a comparative analysis is conducted against established benchmark tools such as PVsyst (Baqir & Channi, 2022), SAM (System Advisor Model) (Shand et al., 2025), and SolarFarmer (Rajput et al., 2018). These tools are widely recognized for PV system modelling, energy yield estimation, and financial forecasting, and are frequently employed in both academic and industrial settings.

7.4.1.PVsyst

PVsyst is a widely recognized and extensively used simulation tool for the design, performance evaluation, and energy yield analysis of photovoltaic (PV) systems. It offers an interactive graphical interface that allows users to define system configurations, input meteorological data, and simulate system behaviour under

different scenarios. Users can interactively adjust system parameters such as module orientation, inverter sizing, shading conditions, and losses, making it useful for project planning and feasibility analysis.

However, despite its interactivity in terms of system design and energy estimation, PVsyst primarily operates on hourly or daily averaged meteorological inputs and relies on standard test condition-based performance models. It does not incorporate minute-level resolution or dynamic adaptation to real-time operational parameters. Moreover, while it allows the calculation of Performance Ratio (PR) using correction factors, it does not directly model the influence of thermodynamic variables such as thermal exergy loss or predict degradation behavior using time-series learning from failure mode-based degradation rates (R_d).

In contrast, the proposed PRLSTM GUI is designed specifically for short-term, minute-level prediction using real-time or near-real-time operational inputs like R_d, exergy loss, and irradiance. It enables dynamic visualization and forecasting of PR with greater responsiveness to field-level fluctuations. A comparative summary of the key differences between PVsyst and the PRLSTM-based interface is provided in table 7.1.

7.4.2. System Advisor Model (SAM)

Developed by NREL, SAM is designed for techno-economic analysis of renewable energy projects. It includes modules for PV system performance modelling, financial assessment, and policy impact. While SAM allows customization and scenario analysis, it still predominantly operates on hourly or daily timescales and lacks support for advanced machine learning integration. It cannot incorporate degradation modelling from failure diagnostics or thermodynamic inefficiencies like exergy loss in the performance ratio calculation, which is a key feature of the

PRLSTM GUI. A detailed comparison of distinctive features of the existing SAM based GUI and the proposed PRLSTM GUI is tabulated in table 7.2.

7.4.3. SolarFarmer

SolarFarmer offers advanced 3D modelling for shading and terrain analysis. While it is powerful for design-stage simulations, it is less suited for operational forecasting or predictive maintenance. It does not feature machine learning-based predictive capabilities and lacks real-time visual diagnostics based on data-driven temporal inputs like minute-based irradiance, R_d, or thermal losses. A detailed comparison of distinctive features of the existing Solar Famer based GUI and the proposed PRLSTM GUI is tabulated in table 7.3.

7.4.4.PRLSTM GUI (Proposed)

The proposed PRLSTM GUI bridges the gap between machine learning-based forecasting and practical usability. It allows minute-based prediction / forecasting of PR using dynamically computed inputs like thermal exergy loss, R_d, and monitored irradiance, offering superior short-term accuracy. Built using Python's Tkinter framework, it is lightweight, interactive, and user-friendly suitable for field engineers, researchers, and solar economists. Its modular design (input panels, calculators, and visualization frames) enables staged computation and interpretation, making it ideal for operational forecasting rather than just design simulation. While general prediction refers to broader trend estimation over longer periods, minute-ahead forecasting emphasizes high-resolution(frequent, detailed updates), near real-time estimations, making it more scalable for dynamic operational control and rapid response systems. Moreover, the inclusion of ablation

analysis, bootstrap validation, and visual feedback on model performance makes it uniquely suited for both research and deployment in live PV plants.

7.5. Feasibility for real-time applications

The proposed PRLSTM-based GUI has been developed with a clear focus on real-time applicability for industrial PV system monitoring. By leveraging minute-resolution input data and a lightweight yet powerful LSTM architecture, the model ensures rapid computation with high forecasting accuracy. The modular design of the GUI, built using Python's Tkinter library, allows seamless integration with on-site data acquisition systems or cloud-based dashboards. Importantly, the system is scalable not only capable of minute-ahead forecasting but also adaptable for daily performance prediction by adjusting input resolution and retraining the model accordingly, making it suitable for both short-term operational control and longer-term performance planning. Real-time prediction of minute-ahead performance ratio (PR) enables early detection of operational anomalies, performance degradation, or thermal inefficiencies. The average inference time per prediction is under a second on standard computing platforms, making it highly suitable for deployment on low-power edge devices (e.g., Raspberry Pi, Jetson Nano) or integration into SCADA systems.

Moreover, the tool's support for user-driven inputs, automatic computation of derived parameters like R_d and thermal exergy loss, and dynamic visualization features ensures that field operators and energy analysts can derive actionable insights without requiring deep technical expertise.

Table 7.1 Comparison of Inputs for Performance Ratio (PR) Estimation employing proposed PRLSTM and PVsyst

Category	Input Parameter	Used in PVsyst	Used in Proposed PRLSTM Model	Remarks
Irradiance Parameters	Global Horizontal Irradiance (GHI)	YES	YES	Base input for energy production modelling
	Plane of Array (POA)	YES	YES	Used for accurate PR calculation
	Irradiance Incident Angle Modifier	YES	NO	Specific to PVsyst's optical loss modelling
Temperature Inputs	Ambient Temperature	YES	YES	Used to compute cell/module temperature
	Module Temperature	YES	YES	PRLSTM uses real-time or measured module temperature as input for estimation of R_d.
Electrical output Parameters	AC Energy output	YES	NO	Required for PR calculation in PVsyst

Thermodynamic Loss	Thermal Exergy Loss	NO	YES	Unique to proposed model; considers irreversibility in energy conversion
Degradation Factors	R_d (Failure mode-based Degradation Rate)	NO	YES	Fixed value in PVsyst, irrespective of variable field conditions and is not used for computation of PR. However, in PRLSTM its adaptive for dynamic field conditions.
Scalability for Forecasting PR	Historical Time-Series Data based input	NO	YES	PRLSTM uses past irradiance, R_d, and exergy loss for time-series learning
	Minute-level Time Resolution	NO	YES	PVsyst operates hourly; PRLSTM supports minute-ahead prediction

Table 7.2 Comparison of Inputs for Performance Ratio (PR) Estimation employing proposed PRLSTM and SAM

Category	Input Parameter	Used in SAM	Used in Proposed PRLSTM Model	Remarks
Irradiance Parameters	Global Horizontal Irradiance (GHI)	YES	YES	Common base input
	Plane of Array (POA) Irradiance	YES	YES	Used in both for accurate PR modelling
	Diffuse and Beam Irradiance	YES	NO	SAM uses this for PV performance; PRLSTM simplifies input
Temperature Inputs	Ambient Temperature	YES	YES	Used in thermal loss estimation
	Module Temperature	YES	YES	SAM estimates; PRLSTM uses real-time input

Electrical Parameters	AC Energy output	YES	NO	Used for performance tracking
Thermodynamic Factors	Thermal Exergy Loss	NO	YES	Unique to PRLSTM to account for irreversible energy loss
Degradation Factors	R_d (Failure-mode-based Degradation Rate)	NO	YES (optimally tuned LSTM-learned)	PRLSTM incorporates dynamic, failure-mode-specific degradation
Scalability for Forecasting PR	Historical Time-Series Data	YES	YES	SAM uses hourly data; PRLSTM uses minute-wise data
	Minute-level Forecasting Capability	NO	YES	SAM is hourly-focused; PRLSTM allows minute based prediction / forecasting

Table 7.3 Comparison of Inputs for Performance Ratio (PR) Estimation employing proposed PRLSTM and SolarFarmer

Category	Input Parameter	Used in SolarFarmer	Used in Proposed PRLSTM Model	Remarks
Irradiance Parameters	Global Horizontal Irradiance (GHI)	YES	YES	Core input in both
	Plane of Array (POA) Irradiance	YES	YES	Critical for surface-level solar potential
	Albedo / Reflected Irradiance	YES	NO	SolarFarmer includes terrain & reflection effects
	Diffuse and Beam Irradiance	YES	NO	Used in 3D shading and layout optimization as the application is dedicated for evaluating terrain-based feasibility
Temperature Inputs	Ambient Temperature	YES	YES	Required in both for performance loss estimation.
	Module Temperature	YES	YES	PRLSTM uses sensor data; SolarFarmer models it

Electrical Parameters	Inverter Efficiency & MPP tracking Array	YES	NO	Detailed Inverter behavior in SolarFarmer. Prediction of PR withput inverter tracking in PRLSTM
	AC Energy output	YES	NO	For real-time performance mapping
Thermodynamic Loss	Thermal Exergy Loss	NO	YES	Unique to PRLSTM; accounts for energy quality degradation
Degradation Factors	R_d (Failure-mode-specific degradation)	NO	YES (LSTM-learned)	Unique predictive capability of PRLSTM
Scalability for Forecasting PR	Historical Time-Series Data Input	YES	YES	PRLSTM uses real-time minute-wise inputs; SolarFarmer uses TMY or CSV files
	Minute-level Forecasting Capability	NO	YES	SolarFarmer focuses on design simulation, not forecasting

Although the training of the LSTM model initially requires computational resources (e.g., ~8 hours on a standard laptop), once trained, the model operates with low latency and minimal memory usage. With access to higher-end GPUs or cloud platforms like AWS or Google Cloud, even model retraining and large-scale deployment can be expedited.

In contrast to traditional simulation tools like PVsyst or SAM, which rely on predefined configurations and irradiance forecasts, the proposed model dynamically responds to real-time sensor data, making it ideal for adaptive energy forecasting, predictive maintenance, and performance auditing in modern PV plants.

7.6. Summary

- Section 7.1 introduces the design and architecture of the proposed GUI, developed using Python's Tkinter library. It describes the three-step interface flow: (1) user input of basic PV parameters, (2) computation of derived attributes like thermal exergy loss and degradation rate (R_d), and (3) PR minute-based prediction and minute-ahead PR forecasting using the trained LSTM model.
- Section 7.2 describes the internal logic of the developed PRLSTM based PR calculator is detailed. It outlines how user-provided inputs are transformed into model and systematically processed by the LSTM-based framework to generate accurate PR estimations. Each computational stage is modularly structured to facilitate user comprehension and interaction.

- Section 7.3 highlights four visual analytics components embedded within the GUI which includes (i) Predicted PR (ii) Forecasted PR (iii) Plot of minute-based thermal exergy loss plot and (iv) minute-based variation of degradation rate. These features assist users in understanding model reliability, performance trends, and loss analysis.
- Section 7.4 provides a comparative study is conducted between the proposed GUI and established tools like PVsyst, SAM, and SolarFarmer. Input parameter requirements, forecasting granularity, thermal modeling, and user interactivity are compared to illustrate the enhanced predictive capabilities and unique integration of thermal and degradation aspects in the PRLSTM-based GUI.
- Section 7.5 explains the real-time implementation potential of the GUI is examined. It discusses computational speed, inference time, hardware compatibility, and practical applicability in industrial or on-field PV system monitoring setups. The tool is validated as suitable for real-time forecasting with minimal latency.

Chapter 8: Conclusion And Future Scope

Overview:

A consolidated conclusion of the developed degradation rate (R_d) model, exergy efficiency estimators, and the PR prediction framework is presented in this chapter. It highlights how the integration of thermal exergy loss dependent on exergy efficiency, failure-mode-based degradation, and irradiance contributes to improved PV performance assessment. The chapter further discusses the contributions of this work to the Solar PV industry, particularly in enhancing predictive accuracy, enabling early diagnostics, and supporting decision-making through a GUI-based performance analyzer. It also outlines the future scope, including possibilities for integrating IoT for real-time monitoring, embedding financial models such as LCOE and NPV, and deploying the tool via cloud-based platforms for broader scalability. Together, these components position the developed tool as a holistic, data-driven framework for accurate and practical solar PV system evaluation.

Outline of Chapter ***Page***

8.1 Conclusion

The proposed research presented a comprehensive methodology to easily assess and improve the prediction accuracy of Performance Ratio (PR) in solar photovoltaic (PV) systems by incorporating failure mode-based degradation, thermal exergy loss, and advanced machine learning architectures. Through experimental investigation and real-time validation across multiple sites, the study addressed key limitations in existing PR assessment approaches.

8.1.1 Quantification of derived input R_d dependent on PR

In Chapter 4, an on-field experimental procedure was undertaken to evaluate the failure mode-based degradation rate (R_d) in PV modules. Using thermal imaging, visual inspection, and IV curve tracing, the system's physical deterioration was qualitatively identified. Four different R_d models based on two different architectures and input attributes were formulated. The employed architectures involved LSTM and FFBP approaches which were dependent on real-time thermal and environmental parameters. The models are derived or trained for the investigated PV site at Bengaluru, India for an annual duration from February 2022 to January 2023. As observed from figure 4.3 in chapter 4 , during training, FFBP Model-1 demonstrates higher accuracy, with the lowest error values across all three metrics with RMSE of 0.001833, MAPE of 0.059867, and MBE of -0.001452 highlighting its strong predictive capability. LSTM Model-1 also performs well, particularly with a very low RMSE of 0.000612 and MBE of -0.013271, suggesting that the inclusion of parameters like DEL_T and R_s contributes positively to the model's performance. On the other hand, LSTM Model-2 shows comparatively higher RMSE of 0.026183 despite a lower MAPE, indicating greater deviation

from actual values. FFBP Model-2, while having the highest MAPE of 0.075105, maintains relatively low RMSE and MBE, suggesting that it may slightly overestimate or underestimate values but remains consistent in its predictions.

Validation using unseen data for a long -term annual duration of 2022-2023 from Telangana further substantiated the ability of the formulated models to act as predictors with its robustness in prediction.

Figure 4.31 presents a comparative analysis of error metrics MBE, RMSE, and MAPE for the proposed LSTM-based R_d models and Feedforward Backpropagation (FFBP) models at the Telangana PV site. Among the models, LSTM Model-1 clearly outperforms the others with the lowest MAPE of 0.0277, as well as comparatively reduced RMSE and MBE values, reflecting better prediction accuracy and minimal bias. These quantitative results indicate strong alignment with the experimental R_d values and underscore the model's generalization capability.

In contrast, LSTM Model-2, though structurally similar, shows relatively higher MAPE (0.05758), highlighting a reduced ability to capture intricate temporal patterns effectively. The FFBP models exhibit larger errors in comparison to LSTM based R_d models during validation. FFBP Model-1, though slightly more stable than FFBP Model-2 in terms of MAPE, records higher RMSE and MBE, indicating its limited effectiveness in modelling nonlinear temporal dynamics. FFBP Model-2 performs the weakest overall, with the highest error values across all three metrics. Among all, considering the validation behaviour of the model to be predominantly important than training, R_d LSTM Model-1 stands out as the most accurate and robust architecture for estimating failure mode-based degradation rate (R_d) in PV

systems, due to its superior capability to learn and generalize temporal dependencies inherent in field data.

8.1.2 Quantification of derived input of thermal exergy loss based on exergy analysis

In Chapter 5, the research extended into exergy efficiency analysis, offering a thermodynamic perspective to solar performance. The impact of attributes namely irradiance, wind speed, module temperature and proposed temperature correction factor towards exergy efficiency thereby affecting thermal exergy loss is analysed. Data normalization and temporal scaling was adopted before model A rule-based estimation of exergy efficiency was implemented using empirical correlations based on site-specific parameters such as irradiance, module temperature, and wind speed, derived from established thermodynamic principles.These rules are grounded in thermodynamic formulations and statistical correlations observed from the actual field data. By applying these predefined rules to the operational dataset, an estimated baseline of exergy efficiency was obtained, serving as a reference for further model training. Building upon this, two hybrid deep learning architectures a dual-layer and a multi-layer hybrid model were proposed to predict daily exergy efficiency more accurately by learning nonlinear relationships beyond the rule-based estimates.

These models, integrating BiLSTM with CNN, GRU, Autoencoder, Transformer, and Attention mechanisms, were evaluated based on operational metrics from Khopoli, Bengaluru, and Telangana location. The results highlighted region-specific performance variations and demonstrated that the multi-layer hybrid model

showed minimal bias during training, while multi-layer hybrid and GRU with BiLSTM model showed excellent predictive performance during validation.

The error performance metrics evaluated during exergy efficiency model development is tabulated in table 5.3 in chapter 5. It demonstrates that while the Transformer-BiLSTM model achieves the lowest MAPE of 1.88%, indicating minimal percentage-based error, the Multilayer hybrid model offers a more balanced and consistent performance across all three metrics with RMSE of 0.54977, MBE of −0.0021, and MAPE of 1.93%. Its RMSE and MBE are more stable compared to the Transformer model, which, despite its lower MAPE, shows a larger bias of −0.0570, suggesting potential underprediction in certain conditions. This makes the Multilayer hybrid model more reliable and generalizable, especially for scenarios where both absolute error magnitude and bias minimization are critical. Hence, although the Transformer-based architecture excels in minimizing percentage error, the Multilayer hybrid model emerges as a better-performing and well-rounded solution for exergy efficiency estimation at the Khopoli site. Reasonable accuracy, their slightly higher MAPE values suggest marginally less effective modelling of complex temporal dependencies.

The predictive behaviour of exergy efficiency model is analysed based on the evaluated error metrics during validation for the location of Telangana as presented in table 5.4. This also offers insight into the predictive accuracy and reliability of the proposed models for the Telangana location. Among these, the Multilayer hybrid model delivers a robust and balanced performance, with an RMSE of 0.5560, a minimal positive MBE of 0.0117, and a competitive MAPE of 1.92%. While the Transformer-BiLSTM model achieves a slightly lower MAPE of 1.86%, it exhibits a larger negative bias of –0.0546, indicating a tendency to underpredict. Similarly,

the GRU-BiLSTM model offers the lowest MAPE of 1.81% but suffers from a significantly large negative MBE of –0.4518, making it less reliable due to considerable underestimation. In contrast, the Multilayer hybrid model maintains stability across all three metrics, ensuring accuracy, reduced bias, and consistent error margins. This suggests that the model not only captures the complex temporal and nonlinear interactions in the input but also generalizes well without significant deviation. Its low MBE and balanced RMSE-MAPE combination emphasize its effectiveness as a reliable and precise estimator of exergy efficiency for the Telangana site, especially under varying climatic and operational conditions.

The error metrics in table 5.5 highlight the predictive performance of various formulated dual -layer and multi-layer BiLSTM-based models for exergy efficiency estimation at the Bengaluru location. Among the six models evaluated, the Multilayer hybrid model once again demonstrates its balanced and reliable performance, achieving a low RMSE of 0.62103 and the lowest MAPE of 1.88%, along with a modest negative bias of –0.0123. These values suggest that the model maintains both accuracy and consistency across predictions, making it particularly well-suited for dynamic and complex real-world conditions.

While the GRU-BiLSTM model records a slightly better RMSE of 0.6229, its MBE is –0.5502, indicating strong underestimation tendencies. The Transformer-BiLSTM and CNN-BiLSTM models show competitive MAPE values of 1.91% but suffer from either increased RMSE or larger biases. The Autoencoder-BiLSTM model, though comparable in MAPE of 1.92%, exhibits the highest RMSE of 0.6924, reducing its reliability.

In comparison, the Multilayer hybrid model sustains low error margins across all metrics, proving its robustness in handling time-series variability and nonlinear

system dynamics. This validates its efficacy in providing stable and precise exergy efficiency forecasts for the Bengaluru PV installation, reinforcing its utility in diverse climatic regions.

So as against major thermodynamic approach which involve empirical model constants and lie dependent on derived attributes which cannot be easily accessed or priorly estimable, the proposed approach predicts exergy efficiency employing easily accessible or fundamental inputs.

8.1.3 Formulation of Performance ratio PRLSTM quantifier

In Chapter 6, a Performance Ratio (PR) forecasting model was formulated using the key predictors: R_d, thermal exergy loss, and irradiance which act as significant attributes as described in Chapter 4 and Chapter 5. A short-term predictive LSTM based model for estimation of daily performance ratio is attempted. The LSTM-based PR model (PR LSTM) also demonstrated excellent temporal learning ability or scalability for minute-ahead forecasting. The model achieved MAPE of 2.71%, with very low RMSE and MBE, outperforming benchmark models such as CNN and GRU.

The CNN model exhibited the lowest training time (approximately 3–4 hours) due to its simpler structure and absence of recurrent memory units. The GRU model required moderate training time (approximately 5–6 hours), benefiting from a reduced gating mechanism compared to LSTM. The PRLSTM model required a comparatively higher training time (approximately 8 hours), attributed to its enhanced memory cell structure designed to capture long- and short-term temporal dependencies. However, once trained, all models demonstrated low inference latency and minimal memory requirements, making them suitable for real-time deployment..An ablation study revealed R_d as the most critical input, followed by

irradiance and exergy loss. The model was validated on new datasets from Bengaluru and Turkey, demonstrating strong generalization and adaptability across climatic regions, further cementing the model's viability for deployment.

The error comparison tables as in table 6.5 to 6.7 for Khopoli, Bengaluru, and Köprübaşı (Turkey) locations collectively underscore the superior accuracy and generalizability of the proposed PR LSTM model in predicting the performance ratio (PR) of solar photovoltaic systems under varied climatic and geographical conditions.

At Khopoli, as shown in table 6.5, the PR LSTM model outperforms all benchmark models including PVsyst, PRcorrected, and PR(FFNN) across all error metrics. It achieves the lowest RMSE of 0.0110, lowest MAPE of 1.83%, and a minimal bias of –0.0082. In contrast, PRcorrected exhibits a significantly higher MAPE of 22.5% and bias, indicating limited reliability in real-time estimation.

At Bengaluru as shown in table 6.6, the PR LSTM again delivers the lowest MAPE of 5.81% and RMSE of 0.0537 compared to benchmark models, while PRcorrected and PVsyst suffer from relatively high errors and negative bias of –0.0853, suggesting that they may underpredict PR in dynamic operational settings.

At the international test site in Köprübaşı, Turkey as shown in table 6.7 in chapter 6, the PR LSTM model continues to exhibit strong generalization performance, with MAPE as low as 1.48%, RMSE of 0.0137, and acceptable MBE. This is significantly better than PVsyst with MAPE of 4.65% and PRcorrected MAPE of 12.71%, validating the model's applicability beyond Indian datasets.

Across all locations, the proposed PR LSTM consistently achieves lower prediction errors and minimal bias, showcasing its robustness, adaptability, and superior predictive accuracy over traditional and reported benchmark models. This

emphasizes its potential as a reliable forecasting tool for real-time performance monitoring and diagnostics in the solar PV industry

8.1.4 Formulation of a User Interactive Tool

In Chapter 7, a Graphical User Interface (GUI)-based PR Analyser tool was developed using Tkinter. The GUI offers streamlined user interaction encompassing execution of three stages namely input entry, intermediate computation (R_d and exergy loss), and final PR prediction /forecasting. Integrated visualization frames were added for long-term variation of key inputs and predicted response (PR). The operational and predictive performance of GUI was compared against benchmarked commercial tools like PVsyst, SAM, and SolarFarmer, demonstrating its added value in terms of inputs required, real-time PR estimation, failure-aware modelling, and user-friendly interface. The tool is well-suited for both technical diagnostics and performance forecasting in PV systems rendering economic viability.

Overall, the study has successfully demonstrated that integrating physical failure insights with machine learning, and translating them into a real-time GUI, can significantly improve the accuracy, usability, and transparency of PV performance assessment.

8.2 Contribution of the proposed research to Solar PV Industry

This research makes a multifaceted contribution to the solar PV industry by integrating experimental diagnostics, thermodynamic evaluation, and deep learning-based predictive modelling into a cohesive, on-field deployable solution for Performance Ratio (PR) estimation. The proposed approach redefines traditional PR prediction methods by embedding failure-mode-based degradation

analysis (R_d) and thermal exergy loss, which are often overlooked in standard energy-based estimation frameworks.

a) Advanced Diagnostics through Field-Informed Degradation Analysis

Unlike conventional PR estimation methods that rely solely on AC energy yield and irradiance data, this research emphasizes the physical degradation mechanisms affecting system performance. By incorporating on-field diagnostics such as visual inspection, IR thermography, and I–V curve tracing, the study enables prediction/ prior estimation of performance loss rate due to failure modes like hot spots, delamination, and encapsulant failure. This field-derived degradation insight strengthens predictive modelling and improves system transparency.

b) Integration of Thermodynamic Principles into Performance Forecasting

The inclusion of thermal exergy loss into the PR forecasting framework brings a novel thermodynamic perspective to solar performance modelling. Exergy analysis captures the quality of energy conversion, accounting for environmental irreversibility and system inefficiencies beyond simple output-based metrics. This not only enhances the technical robustness of PR predictions but also improves the tool's utility for advanced system planners, thermal auditors, and performance evaluators.

c) Development of AI-Powered Predictive Tools for Real-Time Assessment

Through the development of an hyperparameter tuned LSTM-based PR prediction model, trained on minute-resolution of environmental and electrical data, the study introduces a predictive AI model capable of short-term, minute-based predictions of PV performance. The model is also scalable for minute ahead forecasting by utilizing - input attributes based on previous time stamps. The model is extensively validated using realistic data from geographically diverse installations (e.g.,

Khopoli, Telangana, Bengaluru), showcasing its generalization capability and adaptability to site-specific operating conditions.

d) Transformation of model to Graphical User Interface (GUI)

A major contribution lies in the translation of technical complexity into a user-friendly Python-based GUI tool using Tkinter. The GUI allows plant operators and technicians to input basic parameters, automatically compute derived variables (such as R_d and exergy loss), and generate real-time PR forecasts and visualizations. The modular architecture simplifies usage even for non-experts and bridges the gap between advanced analytics and on-ground implementation.

e) Benchmarking and Comparative Superiority

The research presents a critical comparison of the proposed PR analyser with industry-standard tools like PVsyst, SAM, and SolarFarmer, highlighting the limitations of existing black-box models in accurately representing thermally influenced and degradation-induced performance changes. In contrast, the proposed tool demonstrates lower error margins (RMSE, MAPE, MBE) and improved temporal resolution, validating its superiority in precision forecasting and diagnostics.

f) Contribution to Digitalization and Predictive Maintenance in PV Systems

With the increasing scale of PV deployments, the shift from reactive to predictive maintenance is crucial. This study provides the framework for real-time performance monitoring, enabling timely fault detection and intelligent maintenance scheduling. As a result, the tool can help reduce downtime, improve yield, and lower operating costs, key metrics for enhancing plant profitability and sustainability.

8.3 Future Scope

While the proposed solar PV performance analyser demonstrates robust accuracy, real-time usability, and physical relevance, there remains significant potential to expand its functionality, scalability, and integration with emerging technologies. As solar PV adoption accelerates globally, future work can further align this tool with industrial, grid, and digital infrastructure needs. The following sub-sections outline key directions for enhancing its applicability, reliability, and impact across larger and more diverse PV ecosystems.

8.3.1 Real-Time Monitoring and IoT Integration

The next logical step is integrating the proposed PR analyser tool with Internet of Things (IoT) platforms for real-time data acquisition and decision-making. By connecting environmental sensors (e.g., irradiance, temperature, humidity) and electrical monitoring devices to cloud-based systems, the model can be deployed for live tracking of R_d, exergy loss, and PR. Such integration will enable predictive maintenance, anomaly detection, and alert-based notifications, making it highly suitable for smart solar farms and SCADA-supported PV installations.

8.3.2 Expanded Financial Models and LCOE Tracking

To enhance the practical relevance of the proposed PV performance analyser for project developers, operators, and investors, future work can incorporate **integrated techno-economic assessment capabilities**. This includes embedding financial performance indicators such as Levelized Cost of Electricity (LCOE), net present value (NPV), and dynamic payback period estimation within the existing analytical framework.

By linking degradation rate (R_d), performance ratio (PR), and exergy efficiency with long-term energy yield and revenue streams, the tool can be extended to evaluate **the economic consequences of performance degradation under real operating conditions**. Such integration would allow continuous tracking of financial metrics as system performance evolves over time, rather than relying on static or assumed degradation values.

This enhancement would enable the analyser to function as a **decision-support platform for long-term asset management**, supporting investment planning, operation and maintenance strategies, and performance evaluation under power purchase agreements (PPAs). In particular, it would provide insight into how deviations in field-measured degradation and efficiency influence lifecycle costs and project bankability.

8.3.3 Standardization and Cloud Deployment

The developed PR analyser, currently built as a standalone desktop application, can be upgraded into a web-based platform or mobile application for wider access. Cloud deployment on platforms such as AWS, Azure, or Google Cloud will enable multi-user access, location-wise dashboards, and automated updates. Moreover, standardizing input formats and output reports can facilitate regulatory compliance and interoperability with other digital tools used in the solar industry, such as digital twins or fleet management systems.

Additionally, standardization also invites wider validation across diverse PV plant sites and climates, improving the generalizability and robustness of the tool's performance under varying field conditions.

8.4 Summary

This chapter consolidated the findings from the proposed research, culminating in the development of an accurate, experimentally validated solar photovoltaic (PV) performance analyser tailored for real-time prediction of Performance Ratio (PR). The conclusion emphasized the success of LSTM-based models for predicting power degradation rate (R_d), exergy efficiency, and PR, supported by ground-based observations from multiple locations. A key contribution was the development of a GUI-based tool which integrates developed predictive models, enabling practical application of the model in the field. The chapter further highlighted the significance of this work for the solar PV industry in enhancing prediction accuracy, operational diagnostics, and decision-making. Finally, future directions were identified, including IoT-based real-time monitoring, financial integration via LCOE tracking, and deployment via standardized, cloud-compatible frameworks to scale the solution for industrial-grade PV systems.

List of Journal Publications from Thesis

1. Almas, Sundaram, S. and Dwivedi, U.D. (2024) 'Predictive analysis of power degradation rate in solar PV systems emphasizing hot spots and visual effects-based failure modes', Renewable Energy, 228, p. 120684. https://doi.org/10.1016/j.renene.2024.120684 .
2. Almas and Sundaram, S. (2025) 'A user interactive tool for assessment of performance ratio for commercial solar photovoltaic system: Leveraging exergy and energy based inputs', Energy for Sustainable Development, 87, p. 101734. https://doi.org/10.1016/j.esd.2025.101734.
3. Almas and S. Sundaram (2025) 'Integrating artificial intelligence based hybrid deep learning prediction models to estimate exergy efficiency for realistic solar photovoltaic power plants: validation with ground measurements' Thermal Science and Engineering Progress, vol. 65, p. 103902, Sep. 2025, https://doi.org/10.1016/j.tsep.2025.103902.
4. Almas and S. Sundaram, "Impact of Field Degradation Rates on the Levelized Cost of Energy (LCOE) for a roof-top Solar PV System," *2025 IEEE 53rd Photovoltaic Specialists Conference (PVSC)*, Montreal, QC, Canada, 2025, pp. 0564-0567, doi: 10.1109/PVSC59419.2025.11133123.
5. Almas and S. Sundaram, "Effect of Degradation Rate on Payback Period and Economic Viability of a 125 kWp Rooftop Solar PV System," *2025 IEEE 53rd Photovoltaic Specialists Conference (PVSC)*, Montreal, QC, Canada, 2025, pp. 1257-1259, doi: 10.1109/PVSC59419.2025.11132767.
6. Almas and Sivasankari Sundaram. (2024) 'Degradation Assessment Through Prediction Approaches for Solar PV System in South India', Volume 1. Springer Singapore (Proceedings of the 9th International

Conference on Advances in Energy Research), pp. 477–489. https://link.springer.com/book/10.1007/978-981-97-5415-1?page=3#toc.

7. Sundaram, S. and Almas (2023) ‘An experimental cum computational economical approach for evaluation of performance loss rate or degradation rate for realistic roof-top PV plant in South India’, BIO Web of Conferences, 62, 04001. Available at: https://doi.org/10.1051/bioconf/20236204001.
8. Rathore, A., Almas and Sundaram, S. (2024) ‘Energy, exergy and performance analysis of a 380 kWp roof-top PV plant assisted with data-driven models for energy generation’, Journal of Thermal Engineering, 10(5), pp. 1164−1183. Available at: https://doi.org/10.14744/thermal.0000859.
9. Sundaram, S., Almas, and Darbha, P. (2026). A Comprehensive Review on Digital Initiatives Fostering Improvements in Solar PV Systems. In Digital Twins and ESG (eds S. Mondal, A. Kumar and M. Khan). https://doi.org/10.1002/9781394303243.ch6
10. Sundaram, S. & Almas, Energy Systems Engineering modelling analysis and application-Accepted for publication by WILEY ISTE.

References

A. G. Kaplan and Y. A. Kaplan, 'Developing new models for estimating solar radiation by using Fourier series', Energy, vol. 332, p. 137279, Sep. 2025, doi: 10.1016/j.energy.2025.137279.

A. Gopi, K. Sudhakar, N. Wai Keng, and A. R. Krishnan, 'Comparison of normal and weather corrected performance ratio of photovoltaic solar plants in hot and cold climates', Energy Sustain. Dev., vol. 65, pp. 53–62, Dec. 2021, doi: https://doi.org/10.1016/j.esd.2021.09.005.

A. Murat Ates and H. Singh, 'Rooftop solar Photovoltaic (PV) plant – One year measured performance and simulations', J. King Saud Univ. - Sci., vol. 33, no. 3, p. 101361, May 2021, doi: 10.1016/j.jksus.2021.101361.

Abdallah, A.A. et al. (2024) 'Degradation of photovoltaic module backsheet materials in desert climate', Solar Energy Materials and Solar Cells, 277, p. 113118. Available at: https://doi.org/10.1016/j.solmat.2024.113118.

Abdul-Ganiyu, S. et al. (2021) 'Techno-economic analysis of solar photovoltaic (PV) and solar photovoltaic thermal (PVT) systems using exergy analysis', Sustainable Energy Technologies and Assessments, 47, p. 101520. Available at: https://doi.org/10.1016/j.seta.2021.101520.

Abdulla, H., Sleptchenko, A. and Nayfeh, A. (2024) 'Optimizing Cleaning Schedules for Spatially Distributed Photovoltaic Installations with Site-Specific Variations'. Elsevier BV. Available at: https://doi.org/10.2139/ssrn.5069550.

Aboagye, B. et al. (2022) 'Investigation into the impacts of design, installation, operation and maintenance issues on performance and degradation of installed solar photovoltaic (PV) systems', Energy for Sustainable Development, 66, pp. 165–176. Available at: https://doi.org/10.1016/j.esd.2021.12.003.

Aboelwafa, O., Mostafa, H. M., & Elsayed, A. (2022). Development of exergy-efficiency framework for grid-connected PV systems. Thermal Science and Engineering Progress, 34, 101557.

Aghaei, M. et al. (2022) 'Review of degradation and failure phenomena in photovoltaic modules', Renewable and Sustainable Energy Reviews, 159, p. 112160. Available at: https://doi.org/10.1016/j.rser.2022.112160.

Ahmed, R. and Aziz, T. (2023) 'Quantifying Soiling Impacts on PR and Correction Model Proposals', Journal of Cleaner Production, 396, p. 136142. Available at: https://doi.org/10.1016/j.jclepro.2023.136142.

Al-Dahidi, S. et al. (2024) 'Forecasting Solar Photovoltaic Power Production: A Comprehensive Review and Innovative Data-Driven Modeling Framework', Energies, 17(16), p. 4145. Available at: https://doi.org/10.3390/en17164145.

Alimi, O.A., Meyer, E.L. and Olayiwola, O.I. (2022) 'Solar Photovoltaic Modules' Performance Reliability and Degradation Analysis—A Review', Energies, 15(16), p. 5964. Available at: https://doi.org/10.3390/en15165964.

Almarshoud, A.F. et al. (2024) 'Exergy and Energy Analysis of Bifacial PV Module Performance on a Cloudy Day in Saudi Arabia', Sustainability, 16(17), p. 7428. Available at: https://doi.org/10.3390/su16177428.

AP News. (2024, March). Coal remains dominant in India's electricity generation. Retrieved from https://apnews.com/article/909daa6733be375817d9d37ceb98dc6b

Arora, R., Arora, R., & Sridhara, S. N. (2022). Performance assessment of 186 kWp grid interactive solar photovoltaic plant in Northern India. International Journal of Ambient Energy, 43(1), 128–141. https://doi.org/10.1080/01430750.2019.1630312

Arslan, E., Aktaş, M. and Can, Ö.F. (2020) 'Experimental and numerical investigation of a novel photovoltaic thermal (PV/T) collector with the energy and exergy analysis', Journal of Cleaner Production, 276, p. 123255. Available at: .

Aslam, M., Lee, J.-M., Altaha, M. R., & Lee, S.-J. (2020). AE-LSTM Based Deep Learning Model for Degradation Rate Influenced Energy Estimation of a PV System. Energies, 13(17), 4373.

Aste, N. et al. (2013) 'A simplified model for the estimation of energy production of PV systems', Energy, 59, pp. 503–512. Available at: https://doi.org/10.1016/j.energy.2013.07.004.

A. S. Joshi, I. Dincer, and B. V. Reddy, 'Analysis of energy and exergy efficiencies for hybrid PV/T systems', International Journal of Low-Carbon Technologies, vol. 6, no. 1, pp. 64–69, Mar. 2011, doi: 10.1093/ijlct/ctq045

Atia, D.M. et al. (2023) 'Degradation and energy performance evaluation of mono-crystalline photovoltaic modules in Egypt', Scientific Reports, 13(1). Available at: https://doi.org/10.1038/s41598-023-40168-8.

Azam, M.F. et al. (2023) 'Accelerated UV stress testing and characterization of PV-modules: Reliability analysis using different encapsulants and glass sheets',

Sustainable Energy Technologies and Assessments, 56, p. 103119. Available at: https://doi.org/10.1016/j.seta.2023.103119.

Bai, M., Zhao, X., Long, Z., Liu, J., & Yu, D. (2021). Short-Term Probabilistic PV Forecasting Using LSTM and KDE. arXiv:2107.01343.

Bansal, N., Jaiswal, S.P. and Singh, G. (2022) 'Long term operational performance and experimental on-field degradation measurement of 10 MW PV plant in remote location in India', Energy for Sustainable Development, 67, pp. 135–150. Available at: https://doi.org/10.1016/j.esd.2022.01.007.

Baqir, M., & Channi, H. K. (2022). Analysis and design of solar PV system using Pvsyst software. Materials Today: Proceedings, 48, 1332–1338. https://doi.org/10.1016/j.matpr.2021.09.029

Chakraborty, D., Mallick, J., & Mondal, A. (2021). Uncertainty quantification of machine learning-based streamflow forecasting: A hybrid bootstrap and quantile regression approach. Journal of Hydrology, 598, 126409.

Chen, X., & Zhang, L. (2023). Deep learning-based exergy estimation in large-scale PV systems. Energy and AI, 13, 100213.

C. Scott, M. Ahsan, and A. Albarbar, 'Machine learning for forecasting a photovoltaic (PV) generation system', Energy, vol. 278, p. 127807, Sep. 2023, doi: 10.1016/j.energy.2023.127807.

D. Gao, Y. Zhi, X. Rong, and X. Yang, 'Mismatch analysis of rooftop photovoltaics supply and farmhouse load: Data dimensionality reduction and explicable load pattern mining via hybrid deep learning', Appl. Energy, vol. 377, p. 124520, Jan. 2025, doi: 10.1016/j.apenergy.2024.124520.

Deceglie, M. G., Silverman, T. J., Marion, B., & Kurtz, S. R. (2016). Photovoltaic degradation rates: An analytical review. Progress in Photovoltaics: Research and Applications, 24(7), 1043–1055.

Deline, C. et al. (2022) Reducing Uncertainty of Fielded Photovoltaic Performance (Final Technical Report). Golden, CO (United States): National Renewable Energy Laboratory (NREL). Available at: https://doi.org/10.2172/1880076.

Di Tommaso, A. et al. (2022) 'A multi-stage model based on YOLOv3 for defect detection in PV panels based on IR and visible imaging by unmanned aerial vehicle', Renewable Energy, 193, pp. 941–962. Available at: https://doi.org/10.1016/j.renene.2022.04.046.

Dobaria, B., Pandya, M., & Aware, M. (2016). Analytical assessment of 5.05 kWp grid tied photovoltaic plant performance on the system level in a composite climate of western India. Energy, 111, 47–51.

Dubey, S., Sarvaiya, J. N., & Seshadri, B. (2013). Temperature dependent photovoltaic (PV) efficiency and its effect on PV production in the world – A review. Energy Procedia, 33, 311–321.

E. Fuster-Palop, C. Vargas-Salgado, J. C. Ferri-Revert, and J. Payá, "Performance analysis and modelling of a 50 MW grid-connected photovoltaic plant in Spain after 12 years of operation," Renew. Sustain. Energy Rev., vol. 170, p. 112968, Dec. 2022, doi: 10.1016/j.rser.2022.112968.

Ebhota, W.S. and Tabakov, P.Y. (2023) 'Assessment of solar photovoltaic potential of selected site locations in cities across sub-Saharan Africa', Energy Systems [Preprint]. Available at: https://doi.org/10.1007/s12667-023-00625-9.

G. G. Kim, J. H. Hyun, J. H. Choi, S.-H. Ahn, B. G. Bhang, and H.-K. Ahn, 'Quality Analysis of Photovoltaic System Using Descriptive Statistics of Power Performance Index', IEEE Access, vol. 11, pp. 28427–28438, 2023, doi: 10.1109/ACCESS.2023.3257373..

Garcia, M. et al. (2024) 'AI-Based Forecasting of PR for Grid-Connected PV Systems', Applied Energy, 343, p. 120883. Available at: https://doi.org/10.1016/j.apenergy.2023.120883.

Gonzalez, R. and Torres, J. (2020) 'Performance Ratio Variability in Extreme Climate Conditions', Energy, 212, p. 118654. Available at: https://doi.org/10.1016/j.energy.2020.118654.

Guerra, G., MercadéRuiz, P., Anamiati, G., & Landberg, L. (2023). Longterm PV system modelling and degradation using neural networks. EPJ Photovoltaics, 14, 30.

Gönül, Ö. et al. (2022) 'A comparative techno-economic assessment of manually adjustable tilt mechanisms and automatic solar trackers for behind-the-meter PV applications', Renewable and Sustainable Energy Reviews, 168, p. 112770. Available at: https://doi.org/10.1016/j.rser.2022.112770.

H. Demirhan and Z. Renwick, 'Missing value imputation for short to mid-term horizontal solar irradiance data', Appl. Energy, vol. 225, pp. 998–1012, Sep. 2018, doi: 10.1016/j.apenergy.2018.05.054.

H. Sun, M. Duan, Z. Yang, P. Ding, Y. Wu, and B. Lin, 'Evaluation of the intermittent performance of heating terminals based on exergy analysis: Discriminate the impacts

of heat and electricity input', Appl. Energy, vol. 346, p. 121331, Sep. 2023, doi: 10.1016/j.apenergy.2023.121331.

Hacke, P., Terwilliger, K., Glick, S., et al. (2015). A Practical Guide to PV Module Reliability: Moisture Ingress, Delamination, and Potential-Induced Degradation.

Hacke, P., Terwilliger, K., Glick, S., Smith, R., & Kurtz, S. (2015). Test-to-failure of crystalline silicon modules. IEEE Journal of Photovoltaics, 5(1), 174–182.

Hajjaj, C., Bouaichi, A., Zitouni, H., Alami Merrouni, A., Ghennioui, A., Ikken, B., Benhmida, M., Choukri, M., & Regragui, M. (2020). Degradation and performance analysis of a monocrystalline PV system without EVA encapsulating in semi-arid climate. Heliyon, 6(6), e04079.

Hamad, S. A., Ghalib, M. A., Munshi, A., Alotaibi, M., & Ebied, M. A. (2025). Evaluating ML Models for Predicting PV Performance. Scientific Reports, 15, 10750.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.

Hosseinzadeh, M., Sardarabadi, M. and Passandideh-Fard, M. (2018) 'Energy and exergy analysis of nanofluid based photovoltaic thermal system integrated with phase change material', Energy, 147, pp. 636–647. Available at: .

Hukseflux. (2023). How to Calculate PV Performance Ratio. [Online] https://www.hukseflux.com

Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice (2nd ed.). OTexts. [Open access: https://otexts.com/fpp2/]

Høiaas, I. et al. (2022) 'Inspection and condition monitoring of large-scale photovoltaic power plants: A review of imaging technologies', Renewable and Sustainable Energy Reviews, 161, p. 112353. Available at: https://doi.org/10.1016/j.rser.2022.112353.

I. Arora, J. Gambhir, and T. Kaur, 'Data Normalisation-Based Solar Irradiance Forecasting Using Artificial Neural Networks', Arab. J. Sci. Eng., vol. 46, no. 2, pp. 1333–1343, Feb. 2021, doi: 10.1007/s13369-020-05140-y.

IEA PVPS (2015). Analysis of Long-Term Performance of PV Systems. IEA PVPS Task 13.

IEC 61215:2021. (2021). Terrestrial photovoltaic (PV) modules – Design qualification and type approval. International Electrotechnical Commission.

IEC 61724-1. (2021). Photovoltaic System Performance Monitoring—Part 1: Monitoring. International Electrotechnical Commission.

IEEE Journal of Photovoltaics, 5(5), 1429–1436. https://doi.org/10.1109/JPHOTOV.2015.2444098

Iqbal, M., & Baloch, A. A. (2023). Hybrid hydrogen-solar systems: A thermodynamic perspective. International Journal of Hydrogen Energy, 48(1), 204–214. https://doi.org/10.1016/j.ijhydene.2022.10.057

Jathar, L.D. et al. (2023) 'Comprehensive review of environmental factors influencing the performance of photovoltaic panels: Concern over emissions at various phases throughout the lifecycle', Environmental Pollution, 326, p. 121474. Available at: https://doi.org/10.1016/j.envpol.2023.121474.

Johnson, K. and Morris, L. (2023) 'Degradation Effects on Performance Ratio Using NREL Dataset Analysis', Progress in Photovoltaics: Research and Applications, 31(2), pp. 98–106. Available at: https://doi.org/10.1002/pip.3602.

Jordan, D. C., & Kurtz, S. R. (2013). Photovoltaic degradation rates—an analytical review. Progress in Photovoltaics: Research and Applications, 21(1), 12–29. https://doi.org/10.1002/pip.1182

Joshi, A. S., Tiwari, A., & Shukla, A. (2010). Energy and exergy efficiencies of a hybrid photovoltaic–thermal (PV/T) air collector. Renewable Energy, 35(3), 685–691.

K. Panicker, P. Anand, and A. George, 'Assessment of building energy performance integrated with solar PV: Towards a net zero energy residential campus in India', Energy Build., vol. 281, p. 112736, Feb. 2023, doi: 10.1016/j.enbuild.2022.112736.

K. Yan et al., 'Causes of moisture-induced corrosion around N-TOPCon photovoltaic modules and pre-study of high-reliability packaging solutions', Renew. Energy, vol. 248, p. 123148, Aug. 2025, doi: 10.1016/j.renene.2025.123148.

Kandeal, A.W. et al. (2021) 'Infrared thermography-based condition monitoring of solar photovoltaic systems: A mini review of recent advances', Solar Energy, 223, pp. 33–43. Available at: https://doi.org/10.1016/j.solener.2021.05.032.

Kazem, H.A. et al. (2020) 'Evaluation of aging and performance of grid-connected photovoltaic system northern Oman: Seven years' experimental study', Solar Energy, 207, pp. 1247–1258. Available at: https://doi.org/10.1016/j.solener.2020.07.061.

Kim, G.G. et al. (2023) 'Quality Analysis of Photovoltaic System Using Descriptive Statistics of Power Performance Index', IEEE Access, 11, pp. 28427–28438. Available at: https://doi.org/10.1109/ACCESS.2023.3257373.

Kumar, B. S., & Sudhakar, K. (2015). Performance evaluation of 10 MW grid connected solar photovoltaic power plant in India. Energy Reports.

Kumar, K. A., Sundareswaran, K., & Venkateswaran, P. R. (2014). Performance study on a grid connected 20 kW p solar photovoltaic installation in an industry in Tiruchirappalli (India). Energy for Sustainable Development, 23, 294–304.

Kumar, K. A., Sundareswaran, K., & Venkateswaran, P. R. (2014). Performance study on a grid connected 20 kW p solar photovoltaic installation in an industry in Tiruchirappalli (India). Energy for Sustainable Development, 23, 294–304. https://doi.org/10.1016/j.esd.2014.10.002

Kumar, P., Singh, S., & Sharma, R. (2022). Thermodynamic assessment of PV modules under real environmental conditions. Renewable Energy, 189, 1274–1285. https://doi.org/10.1016/j.renene.2022.03.098

Kumar, S. and Rathi, R. (2021) 'Prediction of PV Performance Ratio Using Artificial Neural Networks Based on Environmental Inputs', Energy Reports, 7, pp. 375–383. Available at: https://doi.org/10.1016/j.egyr.2021.09.084.

Kumar, V. and Maheshwari, P. (2022) 'Advanced analytics on IV curves and electroluminescence images of photovoltaic modules using machine learning algorithms', Progress in Photovoltaics: Research and Applications, 30(8), pp. 880–888. Available at: https://doi.org/10.1002/pip.3469.

Köntges, M., Kunze, I., Kajari-Schröder, S., & Breitenmoser, X. (2014). The Risk of Power Loss in Crystalline Silicon Based Photovoltaic Modules Due to Micro-Cracks. Solar Energy Materials and Solar Cells, 95(4), 1131–1137.

Köntges, M., Kunze, I., Kajari-Schröder, S., et al. (2014). The risk of power loss in crystalline silicon-based photovoltaic modules due to microcracks.

Köntges, M., Kurtz, S., Packard, C., Jahn, U., & Berger, K. A. (2014). Review of Failures of Photovoltaic Modules. IEA PVPS Task 13.

L. Yin and Y. Sun, 'BiLSTM-InceptionV3-Transformer-fully-connected model for short-term wind power forecasting', Energy Convers. Manag., vol. 321, p. 119094, Dec. 2024, doi: 10.1016/j.enconman.2024.119094.

Lee, D. et al. (2022) 'Analyzing Real-World PR Deviations Using the Sandia Model for Utility-Scale PV Plants', IEEE Journal of Photovoltaics, 12(3), pp. 456–463. Available at: https://doi.org/10.1109/JPHOTOV.2022.3145629.

Lee, J., Kim, H., & Park, J. (2022). Impact of partial shading on exergy performance of PV modules. Sustainable Energy Technologies and Assessments, 53, 102666. https://doi.org/10.1016/j.seta.2022.102666

Liu, R. et al. (2025) 'Accelerated Aging Method of Performance Attenuation of Crystalline Silicon Photovoltaic Modules Under Full-Spectrum Conditions', Materials, 18(7), p. 1507. Available at: https://doi.org/10.3390/ma18071507.

Luo, X., & Zhang, D. (2021). An Adaptive Deep Learning Framework for PV Forecasting. arXiv:2109.13442.

Lyu, Y. et al. (2020) 'Impact of environmental variables on the degradation of photovoltaic components and perspectives for the reliability assessment methodology', Solar Energy, 199, pp. 425–436. Available at: https://doi.org/10.1016/j.solener.2020.02.020.

Md. I. Islam, M. S. Bin Jadin, A. Al Mansur, and T. Alharbi, 'Electrical Performance and Degradation Analysis of Field-Aged PV Modules in Tropical Climates: A Comparative Experimental Study', Energy Conversion and Management: X, vol. 24, p. 100719, Oct. 2024, doi: 10.1016/j.ecmx.2024.100719.

M. Shravanth Vasisht, J. Srinivasan, and S. K. Ramasesha, 'Performance of solar photovoltaic installations: Effect of seasonal variations', Sol. Energy, vol. 131, pp. 39–46, Jun. 2016, doi: 10.1016/j.solener.2016.02.013.

M. Yaghoubirad, N. Azizi, A. Ahmadi, Z. Zarei, and S. F. Moosavian, 'Performance assessment of a solar PV module for different climate classifications based on energy, exergy, economic and environmental parameters', Energy Reports, vol. 8, pp. 15712–15728, Nov. 2022, doi: 10.1016/j.egyr.2022.12.070.

Madurai Elavarasan, R., Mudgal, V., Selvamanohar, L., Wang, K., Huang, G., Shafiullah, G. M., Markides, C. N., Reddy, K. S., & Nadarajah, M. (2022). Pathways toward high-efficiency solar photovoltaic thermal management for electrical, thermal and combined generation applications: A critical review. Energy Conversion and Management, 255, 115278. https://doi.org/10.1016/j.enconman.2022.115278

Mahmood, F.I. et al. (2023) 'Potential induced degradation in c-Si glass-glass modules after extended damp heat stress', Solar Energy, 254, pp. 102–111. Available at: https://doi.org/10.1016/j.solener.2023.03.013.

M. Goel, 'Solar rooftop in India: Policies, challenges and outlook', Green Energy & Environment, vol. 1, no. 2, pp. 129–137, Jul. 2016, doi: 10.1016/j.gee.2016.08.003.

Makrides, G., Zinsser, B., Norton, M., & Georghiou, G. E. (2012). Degradation of Photovoltaic Modules: Results of a Survey of 200 PV Systems in Europe. Renewable Energy, 44, 1–8.

Malhotra, P., TV, V., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., & Shroff, G. (2016). Multi-Sensor Prognostics using an Unsupervised Health Index based on LSTM Encoder-Decoder. arXiv preprint arXiv:1608.06154.

Malvoni, M. et al. (2020) 'Assessment of Performance of Large-Scale PV Systems in Tropical Semi-Arid Climates Using IEA PVPS Data', Renewable Energy, 162, pp. 1204–1213. Available at: https://doi.org/10.1016/j.renene.2020.08.012.

Mansour, Ridha Ben et al. (2021) 'Optimizing the Solar PV Tilt Angle to Maximize the Power Output: A Case Study for Saudi Arabia', IEEE Access, 9, pp. 15914–15928. Available at: https://doi.org/10.1109/access.2021.3052933.

Masita, K., Hasan, A. and Shongwe, T. (2022) '75MW AC PV Module Field Anomaly Detection Using Drone-Based IR Orthogonal Images With Res-CNN3 Detector', IEEE Access, 10, pp. 83711–83722. Available at: https://doi.org/10.1109/access.2022.3194547.

Massaro, A. and Starace, G. (2022) 'Advanced and Complex Energy Systems Monitoring and Control: A Review on Available Technologies and Their Application Criteria', Sensors, 22(13), p. 4929. Available at: https://doi.org/10.3390/s22134929.

Ministry of Power, Government of India. (2025). Power Sector at a Glance. Retrieved from

Müller, M., Köntges, M., & Jahn, U. (2019). Analysis of electroluminescence and infrared thermal images of photovoltaic modules. Solar Energy, 183, 173–180.

N. Aste, C. Del Pero, F. Leonforte, and M. Manfren, 'A simplified model for the estimation of energy production of PV systems', Energy, vol. 59, pp. 503–512, Sep. 2013, doi: 10.1016/j.energy.2013.07.004.

N. E. Michael, R. C. Bansal, A. A. A. Ismail, A. Elnady, and S. Hasan, 'A cohesive structure of Bi-directional long-short-term memory (BiLSTM) -GRU for predicting hourly solar radiation', Renew. Energy, vol. 222, p. 119943, Feb. 2024, doi: 10.1016/j.renene.2024.119943.

Nam, W. et al. (2025) 'Predicting Photovoltaic Module Lifespan Based on Combined Stress Tests and Latent Heat Analysis', Energies, 18(2), p. 304. Available at: https://doi.org/10.3390/en18020304.

Owen-Bellini, M. et al. (2021) 'Advancing reliability assessments of photovoltaic modules and materials using combined-accelerated stress testing', Progress in Photovoltaics: Research and Applications, 29(1), pp. 64–82. Available at: https://doi.org/10.1002/pip.3342.

Ozkalay, E. et al. (2024) 'Correlating long-term performance and aging behaviour of building integrated PV modules', Energy and Buildings, 316, p. 114252. Available at: https://doi.org/10.1016/j.enbuild.2024.114252.

Öztürk, A., Doğan, B., & YeşilYurt, M. (2023). Energy, exergy, sustainability, and economic analyses of a grid-connected solar power plant consisting of bifacial PV modules with solar tracking system on a single axis. Science and Technology for Energy Transition (STET), 78.

Padmavathi, K., & Daniel, S. A. (2013). Performance analysis of a 3MWp grid connected solar photovoltaic power plant in India. Energy for Sustainable Development, 17(6), 615–625.

Pagnini, L. et al. (2024) 'Levelized cost of electricity in renewable energy communities: Uncertainty propagation analysis', Applied Energy, 366, p. 123278. Available at: https://doi.org/10.1016/j.apenergy.2024.123278.

Paliwal, K., Mishra, A., & Sharma, R. K. (2023). Real-time exergy analysis and prediction of PV modules using ensemble regression techniques. Solar Energy, 258, 441–454.

Patel, M. et al. (2024) 'Real-Time Monitoring-Based PR Modelling Considering Degradation', Renewable and Sustainable Energy Reviews, 177, p. 113318. Available at: https://doi.org/10.1016/j.rser.2023.113318.

Pathak, M.J.M., Sanders, P.G. and Pearce, J.M. (2014) 'Optimizing limited solar roof access by exergy analysis of solar thermal, photovoltaic, and hybrid photovoltaic thermal systems', Applied Energy, 120, pp. 115–124. Available at: https://doi.org/10.1016/j.apenergy.2014.01.041.

P. P. Angelov, E. A. Soares, R. Jiang, N. I. Arnold, and P. M. Atkinson, 'Explainable artificial intelligence: an analytical review', WIREs Data Min & Knowl, vol. 11, no. 5, p. e1424, Sep. 2021, doi: 10.1002/widm.1424.

Qingyang, J. et al. (2021) 'Energy and exergy analyses of PV, solar thermal and photovoltaic/thermal systems: a comparison study', International Journal of Low-Carbon Technologies, 16(2), pp. 604–611. Available at: https://doi.org/10.1093/ijlct/ctaa092.

R. Kumar, C. S. Rajoria, A. Sharma, and S. Suhag, 'Design and simulation of standalone solar PV system using PVsyst Software: A case study', Mater. Today Proc., vol. 46, pp. 5322–5328, 2021, doi: 10.1016/j.matpr.2020.08.785.

R. Madurai Elavarasan et al., 'Pathways toward high-efficiency solar photovoltaic thermal management for electrical, thermal and combined generation applications: A critical review', Energy Convers. Manag., vol. 255, p. 115278, Mar. 2022, doi: 10.1016/j.enconman.2022.115278.

R. Petela, 'Exergy of undiluted thermal radiation', Solar Energy, vol. 74, no. 6, pp. 469–488, Jun. 2003, doi: 10.1016/S0038-092X(03)00226-3.

Rahaman, S.A., Urmee, T. and Parlevliet, D.A. (2022) 'Investigate the impact of environmental and operating conditions of infrared (IR) imaging on infrared thermography of PV modules to identify defects', Solar Energy, 245, pp. 231–253. Available at: https://doi.org/10.1016/j.solener.2022.09.003.

Rai, A. Shrivastava, and K. C. Jana, 'A CNN-BiLSTM based deep learning model for mid-term solar radiation prediction', Int. Trans. Electr. Energy Syst., vol. 31, no. 9, Sep. 2021, doi: 10.1002/2050-7038.12664.

Rajput, A. S., Ho, J. W., Zhang, Y., Nalluri, S., & Aberle, A. G. (2018). Quantitative estimation of electrical performance parameters of individual solar cells in silicon photovoltaic modules using electroluminescence imaging. Solar Energy, 173, 201–208. https://doi.org/10.1016/j.solener.2018.07.046

Rao, C.K., Sahoo, S.K. and Yanine, F.F. (2025) 'A comprehensive review of smart energy management systems for photovoltaic power generation utilizing the internet of things', Unconventional Resources, 7, p. 100197. Available at: https://doi.org/10.1016/j.uncres.2025.100197.

Rathore, A. et al. (2025) 'Day ahead solar forecast using long short term memory network augmented with Fast Fourier transform-assisted decomposition technique', Renewable Energy, 247, p. 123021. Available at: https://doi.org/10.1016/j.renene.2025.123021.

Reddy, K. S., & Kaushik, S. C. (2021). Exergy-based performance analysis of solar photovoltaic systems. Solar Energy, 225, 741–750. https://doi.org/10.1016/j.solener.2021.08.036

Reuters. (2024, April 1). India hydropower output records steepest fall in nearly four decades. Retrieved from https://www.reuters.com/business/energy/india-hydropower-output-records-steepest-fall-nearly-four-decades-2024-04-01/

Rodríguez, E. et al. (2022) 'Modelling the Exergy of Solar Radiation: A Review', Energies, 15(4), p. 1477. Available at: https://doi.org/10.3390/en15041477.

Roy, A., Ghosh, S., & Chatterjee, A. (2022). Thermodynamic indicators for fault detection in PV arrays using hybrid modeling. Renewable Energy, 182, 906–919.

S. Lidaighbi, M. Elyaqouti, D. Ben Hmamou, D. Saadaoui, K. Assalaou, and E. Arjdal, 'A new hybrid method to estimate the single-diode model parameters of solar photovoltaic panel', Energy Convers. Manag. X, vol. 15, p. 100234, Aug. 2022, doi: 10.1016/j.ecmx.2022.100234.

S. Nesmachnow and L. Hernández Callejo, Eds., Smart Cities: Second Ibero-American Congress, ICSC-CITIES 2019, Soria, Spain, October 7–9, 2019, Revised Selected Papers, vol. 1152. in Communications in Computer and Information Science, vol. 1152. Cham: Springer International Publishing, 2020. doi: 10.1007/978-3-030-38889-8.

S. Shapsough, R. Dhaouadi, and I. Zualkernan, 'Using Linear Regression and Back Propagation Neural Networks to Predict Performance of Soiled PV Modules', Procedia Comput. Sci., vol. 155, pp. 463–470, 2019, doi: 10.1016/j.procs.2019.08.065.

S. Sreenath, K. Sudhakar, and Y. Af, '7E analysis of a conceptual utility-scale land-based

Salamah, T. et al. (2022) 'Effect of dust and methods of cleaning on the performance of solar PV module for different climate regions: Comprehensive review', Science of The Total Environment, 827, p. 154050. Available at: https://doi.org/10.1016/j.scitotenv.2022.154050.

Salameh, T. et al. (2023) 'Energy and exergy assessment for a University of Sharjah's PV grid-connected system based on experimental for harsh terrestrial conditions', Energy Reports, 9, pp. 345–353. Available at: https://doi.org/10.1016/j.egyr.2022.12.117.

Salazar-Pena, N., Tabares, A., & Gonzalez-Mancera, A. (2024). AI-powered dynamic fault detection and GUI-based PR visualization. arXiv preprint arXiv:2409.00052.

Sample, T. et al. (2021). Nonlinear Field Degradation in PV Systems and Its Impact on System Performance Predictions. IEEE Journal of Photovoltaics, 11(4), 1032–1039.

Schwabe, U. and Jansson, P.M. (2009) 'Performance measurement of amorphous and monocrystalline silicon PV modules in Eastern U.S. Energy production versus ambient and module temperature', in 2009 IEEE Intrumentation and Measurement Technology Conference. 2009 IEEE Intrumentation and Measurement Technology Conference

(I2MTC), Singapore. Singapore: IEEE, pp. 1636–1641. Available at: https://doi.org/10.1109/IMTC.2009.5168717.

Sethi, D. and Kosmopoulos, P.G. (2025) 'Rooftop Solar Photovoltaic Potential in Polluted Indian Cities: Atmospheric and Urban Impacts, Climate Trends, Societal Gains, and Economic Opportunities', Remote Sensing, 17(7), p. 1221. Available at: https://doi.org/10.3390/rs17071221.

Shand, J., Glakpe, E., & Ivey, C. (2025). Policy implications of implementing residential PV solar energy systems in developing regions. Energy Policy, 196, 114414. https://doi.org/10.1016/j.enpol.2024.114414

Sharma, A., & Sinha, S. (2019). Thermodynamic performance evaluation of PV modules under partial shading conditions. Energy Reports, 5, 1165–1172. https://doi.org/10.1016/j.egyr.2019.08.069

Sharma, N., Yadav, P., & Mehta, S. (2021). Real-time PV performance monitoring and GUI integration for distributed solar plants. Journal of Cleaner Energy Technologies, 9(5), 1243–1251.

Sharma, V., & Chandel, S. S. (2013). Performance and degradation analysis for long term reliability of solar photovoltaic systems: A review. Renewable and Sustainable Energy Reviews, 27, 753–767.

Sharma, V., Kumar, A., & Tiwari, G. N. (2018). Performance Analysis of Rooftop PV Systems in Composite Climate: An Empirical Case Study. Solar Energy, 173, 39–52.

Simeunović, J., Schubnel, B., Alet, P.-J., & Carrillo, R. E. (2021). Spatio-Temporal Graph Neural Networks for PV Forecasting. arXiv:2107.13875.

Singh, R., & Jain, M. (2024). A hybrid machine learning and thermodynamic model for PV exergy estimation. Renewable Energy, 222, 456–468. https://doi.org/10.1016/j.renene.2023.12.018

S. R. Madeti and S. N. Singh, 'Monitoring system for photovoltaic plants: A review', Renewable and Sustainable Energy Reviews, vol. 67, pp. 1180–1207, Jan. 2017, doi: 10.1016/j.rser.2016.09.088.

Singh, R., Sharma, M. and Yadav, K. (2022) 'Degradation and reliability analysis of photovoltaic modules after operating for 12 years: A case study with comparisons', Renewable Energy, 196, pp. 1170–1186. Available at: https://doi.org/10.1016/j.renene.2022.07.048.

Smith, J. and Allen, R. (2021) 'Performance Ratio Evaluation Methods for Photovoltaic Systems Based on IEC Standards', Solar Energy, 221, pp. 85–92. Available at: https://doi.org/10.1016/j.solener.2021.04.032.

Sobri, S., Koohi-Kamali, S., & Rahim, N. Abd. (2018). Solar photovoltaic generation forecasting methods: A review. Energy Conversion and Management, 156, 459–497. https://doi.org/10.1016/j.enconman.2017.11.019

Solar Energy Materials and Solar Cells, 95(4), 1131–1137. https://doi.org/10.1016/j.solmat.2010.12.018

solar photovoltaic power plant', Energy, vol. 219, p. 119610, Mar. 2021, doi: 10.1016/j.energy.2020.119610.

Sukumaran, S. and Sudhakar, K. (2018) 'Performance analysis of solar powered airport based on energy and exergy analysis', Energy, 149, pp. 1000–1009. Available at: .

Sundaram, S., & Babu, J. S. C. (2015). Performance evaluation and validation of 5MWp grid connected solar photovoltaic plant in South India. Energy Conversion and Management, 100, 429–439. https://doi.org/10.1016/j.enconman.2015.04.069

T. A. Woldegiyorgis et al., 'Harnessing solar power: Predicting photovoltaic potential in fiche, oromia, ethiopia with artificial neural networks', Sci. Afr., vol. 21, p. e01884, Sep. 2023, doi: 10.1016/j.sciaf.2023.e01884.

Tanaka, H. et al. (2022) 'Hybrid PR Modelling: Merging IEC Methods with Machine Learning for Accuracy Improvements', Energy AI, 10, p. 100193. Available at: https://doi.org/10.1016/j.egyai.2022.100193.

Tariq, M., Mahajan, R., & Tiwari, G. N. (2020). Development of SCADA and GUI-based solar PV monitoring system. Solar Energy, 203, 1–10.

Tiwari, S., Yadav, S., & Mahato, R. (2022). Investigation of exergy destruction in photovoltaic arrays under partial shading conditions. Journal of Cleaner Production, 344, 130879.

Tripathi, B., Yadav, P., Rathod, S., & Kumar, M. (2014). Performance analysis and comparison of two silicon material based photovoltaic technologies under actual climatic conditions in Western India. Energy Conversion and Management, 80, 97–102.

T. Rahman et al., 'Investigation of Degradation of Solar Photovoltaics: A Review of Aging Factors, Impacts, and Future Directions toward Sustainable Energy Management', Energies, vol. 16, no. 9, p. 3706, Apr. 2023, doi: 10.3390/en16093706.

U. Schwabe and P. M. Jansson, 'Performance measurement of amorphous and monocrystalline silicon PV modules in Eastern U.S. Energy production versus ambient and module temperature', in 2009 IEEE Intrumentation and Measurement Technology Conference, Singapore. Singapore: IEEE, May 2009, pp. 1636–1641. doi: 10.1109/IMTC.2009.5168717.

U. Singh, S. Singh, S. Gupta, M. A. Alotaibi, and H. Malik, 'Forecasting rooftop photovoltaic solar power using machine learning techniques', *Energy Reports*, vol. 13, pp. 3616–3630, Jun. 2025, doi: 10.1016/j.egyr.2025.03.005.

V. Sharma and S. S. Chandel, 'Performance and degradation analysis for long term reliability of solar photovoltaic systems: A review', Renew. Sustain. Energy Rev., vol. 27, pp. 753–767, Nov. 2013, doi: 10.1016/j.rser.2013.07.046.

Vandal, T., Kodra, E., Ganguly, S. (2017). DeepSD: Generating High Resolution Climate Change Projections through Single Image Super-Resolution. Proceedings of the 23rd ACM SIGKDD. https://doi.org/10.1145/3097983.3098004

Vector Renewables. (2023). Problems with the Calculation of PR. [Online] https://www.vectorenewables.com

Wall, G., & Gong, M. (2020). Exergy analysis of renewable energy systems. Energy, 203, 117867. https://doi.org/10.1016/j.energy.2020.117867

Wang, L. et al. (2021) 'Simulation-Based Estimation of PR in Various PV Configurations', Sustainable Energy Technologies and Assessments, 46, p. 101229. Available at: https://doi.org/10.1016/j.seta.2021.101229.

Wang, Y., & Lu, H. (2021). Exergy prediction using emotional artificial neural networks for PV systems. Applied Energy, 301, 117483. https://doi.org/10.1016/j.apenergy.2021.117483

Waqar Akram, M. et al. (2022) 'Failures of Photovoltaic modules and their Detection: A Review', Applied Energy, 313, p. 118822. Available at: https://doi.org/10.1016/j.apenergy.2022.118822.

X. Chen, T. Karin, and A. Jain, 'Analyzing the impact of design factors on solar module thermomechanical durability using interpretable machine learning techniques', Appl. Energy, vol. 377, p. 124462, Jan. 2025, doi: 10.1016/j.apenergy.2024.124462.

X. Hu, H. Li, and C. Si, ‘Improved composite model using metaheuristic optimization algorithm for short-term power load forecasting’, Electr. Power Syst. Res., vol. 241, p. 111330, Apr. 2025, doi: 10.1016/j.epsr.2024.111330.

Y. Yao, Y. Wang, H. Jia, and Y. Mu, ‘An analytical approach based on coupled multi-physics model for photovoltaic arrays performance simulation’, Electr. Power Syst. Res., vol. 224, p. 109773, Nov. 2023, doi: 10.1016/j.epsr.2023.109773.

Z. Zhang and Z.-Q. J. Xu, ‘Implicit Regularization of Dropout’, IEEE Trans. Pattern Anal. Mach. Intell., vol. 46, no. 6, pp. 4206–4217, Jun. 2024, doi: 10.1109/TPAMI.2024.3357172.

Zhang, Y. and He, J. (2020) ‘Empirical Modelling of Performance Ratio Under Varying Climate Conditions’, Energy Conversion and Management, 213, p. 112835. Available at: https://doi.org/10.1016/j.enconman.2020.112835.

Zhao, Z., Chen, W., Wu, X., Chen, Y., & Hu, J. (2019). Review on uncertainty analysis of machine learning applications in building energy prediction. Energy and Buildings, 202, 109389.

Appendix - A

Python Code for two non- linear failure mode-based R_d models with differences in inputs (LSTM based approach) for Chapter 4

Table A.1.Imput sample for the R_d models

Date	I	W	H	R_s	R_d	DEL_T	Tm	Ta
2022-02-23	739.1166	5.2778	42.0	0.3405	0.0791	6.8971	50.8346	33.1490
2022-02-24	1021.6422	3.6111	32.0	0.3062	0.0527	9.8968	55.6866	31.4308
2022-02-27	1014.2466	1.1111	40.0	0.2914	0.0475	2.8335	63.4070	33.1816
2022-03-02	886.5780	1.6667	46.0	0.2756	0.0538	6.2159	52.5285	32.9670
2022-03-03	942.1162	5.2778	24.0	0.3317	0.0619	5.8904	54.1530	33.3134
...	...	...	...	...	...	...	...	...
2023-12-22	1005.3923	4.1667	37.0	0.2622	0.0641	16.8153	51.3469	28.7530
2023-12-23	994.7615	4.1667	56.0	0.2581	0.0707	15.7153	54.7069	29.5969

```
import pandas as pd
import numpy as np
import pickle
import matplotlib.pyplot as plt
%matplotlib inline
plt.style.use('bmh')
import math
import keras
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense
```

```
from tensorflow.keras.layers import LSTM


from matplotlib import pyplot
import warnings
warnings.simplefilter("ignore", UserWarning)
df = pd.read_excel('/Users/Almas/Desktop/25june.xlsx')
#df = pd.read_excel('25june.xlsx')
df.index=df.Date
df.drop(df.columns[[0]], axis=1, inplace=True)
df
120 rows × 11 columns
df1=df[['DT','Rs','H','W','I','Rd']]
df2=df[['Ta','Tm','I','H','W','Rs','Rd']]

###################MODEL1#########################
def split_data(df,train_size):
    train_days=math.floor(len(df)*train_size/9)
    train_data,test_data=df.iloc[0:train_days*9],df.iloc[train_days*9:len(df)]
    return train_data,test_data
def create_trainable_dataset(dataframe,n_inputs,n_outputs):
    X,Y=list(),list()
    for i in range(len(dataframe)-n_inputs-n_outputs+1):
        X.append(dataframe.iloc[i:(i+n_inputs), :])
        Y.append(dataframe.iloc[i + n_inputs:i + n_inputs + n_outputs, -1])
    return np.array(X), np.array(Y)
n_input=9
n_output=1
train_size=0.8
train_data,test_data=split_data(df1,train_size)
X_train,Y_train=create_trainable_dataset(train_data,n_input,n_output)
X_test,Y_test=create_trainable_dataset(test_data,n_input,n_output)
from tensorflow.keras import Input
```

```
from tensorflow.keras.layers import Dense, LSTM
from tensorflow.keras.models import load_model, Model
# design network with epoch=500, batch size = 128, epoch=2000
model = Sequential()
model.add(LSTM(256, input_shape=(X_train.shape[1],X_train.shape[2]), return_sequences=True))
model.add(LSTM(256, return_sequences=True))
model.add(LSTM(256, return_sequences=True))
model.add(LSTM(256))
model.add(Dense(1))
model.compile(loss='mae', optimizer='adam',metrics=['accuracy'])
# fit network
history = model.fit(X_train, Y_train, epochs=300, batch_size=128, validation_data=(X_test, Y_test), verbose=1, shuffle=True)
# plot histogram
pyplot.plot(history.history['loss'], label='train')
pyplot.plot(history.history['val_loss'], label='test')
pyplot.legend()
pyplot.show()
```

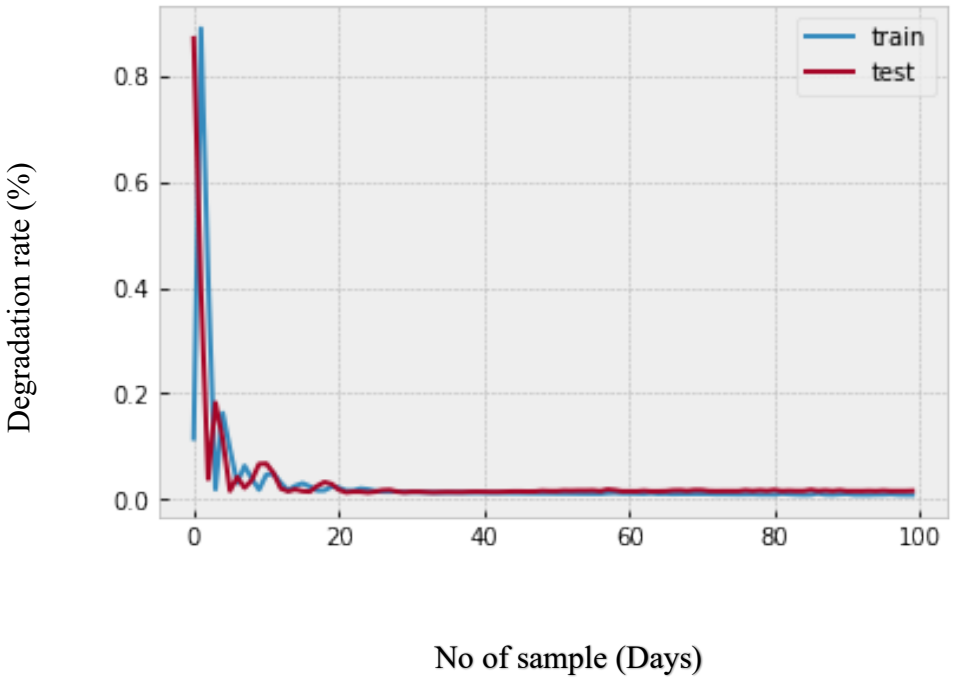


Fig.A.1Training and Testing of model-1.

```
yhat2 = model.predict(X_test)
Epoch 1/100

1/1 [==============================] - 2s 2s/step
train=pd.DataFrame(Y_train)
test=pd.DataFrame(Y_test)
predicted=pd.DataFrame(yhat2)
test=test*(df['Rd'].max()-df['Rd'].min())+df['Rd'].min()
predicted=predicted*(df['Rd'].max()-df['Rd'].min())+df['Rd'].min()
cc=df1*(df['Rd'].max()-df['Rd'].min())+df['Rd'].min()
s=range(90,111)
predicted.index=s
test.index=s


def export_data(predicted):
    df=pd.DataFrame(predicted)
    df.to_excel("errors.xlsx")
    df.to_csv("errors.csv")

x=range(0,len(df1)-14)
plt.plot(x,cc.Rd[0:-14],'r',label='Train')

plt.plot(predicted,'b--',label='Predicted')
plt.plot(test,'g',label='Test')
plt.legend()
<matplotlib.legend.Legend at 0x12eefa230>
################MODEL_2#################
n_input=9
n_output=1
train_size=0.8
```

```
train_data,test_data=split_data(df2,train_size)

X_train,Y_train=create_trainable_dataset(train_data,n_input,n_output)

X_test,Y_test=create_trainable_dataset(test_data,n_input,n_output)

# design network with epoch=2000, batch size = 128, epoch=500

model = Sequential()

model.add(LSTM(256, input_shape=(X_train.shape[1],X_train.shape[2]),
return_sequences=True))

model.add(LSTM(256, return_sequences=True))

model.add(LSTM(256, return_sequences=True))

model.add(LSTM(256))

model.add(Dense(1))

model.compile(loss='mae', optimizer='adam',metrics=['accuracy'])

# fit network

history = model.fit(X_train, Y_train, epochs=500, batch_size=128,
validation_data=(X_test, Y_test), verbose=1, shuffle=True)

# plot history

pyplot.plot(history.history['loss'], label='train')

pyplot.plot(history.history['val_loss'], label='test')

pyplot.legend()

pyplot.show()

loss: 0.0126 - accuracy: 0.0000e+00 - val_loss: 0.0179 - val_accuracy:
0.0000e+00
```

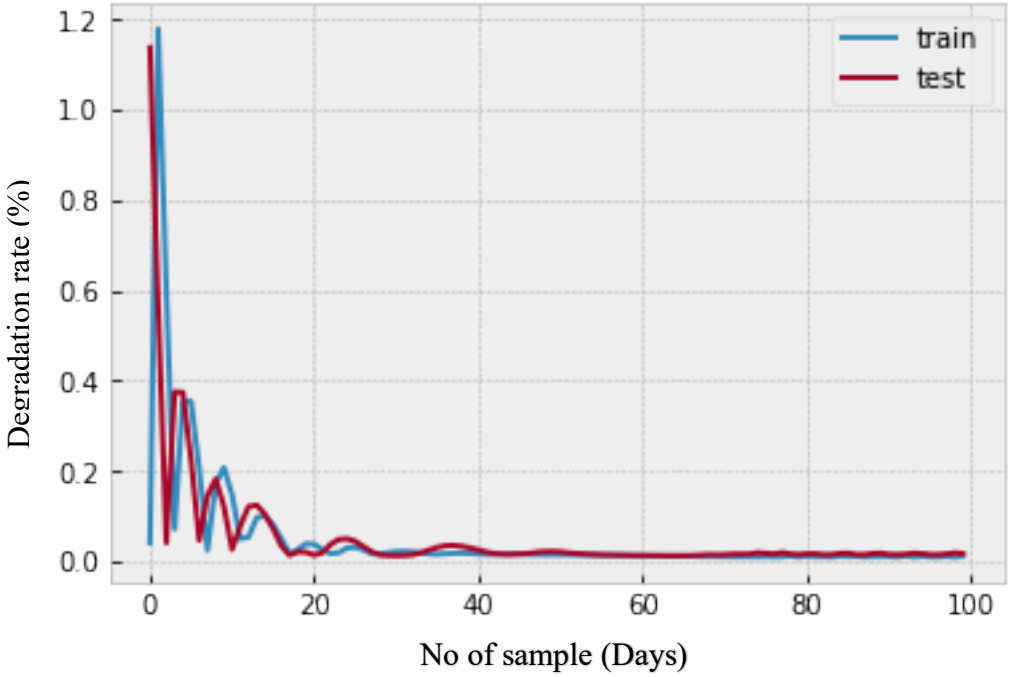


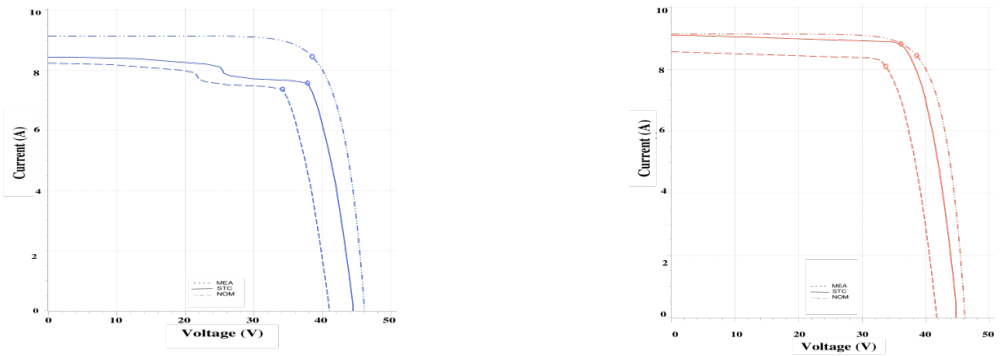
Fig.A.2.Training and testing model-2

```
yhat3 = model.predict(X_test)
1/1 [==============================] - 2s 2s/step
train2=pd.DataFrame(Y_train)
test2=pd.DataFrame(Y_test)
predicted2=pd.DataFrame(yhat3)
test2=test2*(df['Rd'].max()-df['Rd'].min())+df['Rd'].min()
predicted2=predicted2*(df['Rd'].max()-df['Rd'].min())+df['Rd'].min()
cc2=df2*(df['Rd'].max()-df['Rd'].min())+df['Rd'].min()
s=range(99,120)
predicted2.index=s
test2.index=s
x=range(0,len(df1)-14)
plt.plot(x,cc2.Rd[0:-14],'r',label='Train')

plt.plot(predicted2,'b--',label='Predicted')
plt.plot(test2,'g',label='Test')
plt.legend()
```

```
<matplotlib.legend.Legend at 0x1339a17b0>
#final result
x=range(0,len(df1)-14)
plt.plot(x,cc2.Rd[0:-14],'r',label='Train')
plt.plot(predicted,'b--',label='Predicted_model1')
plt.plot(predicted2,'m--',label='Predicted_model2')
plt.plot(test2,'g',label='Test')
plt.legend()
<matplotlib.legend.Legend at 0x134ba3250>
train.to_excel('M1TRAIN.xlsx')
train2.to_excel('M2TRAIN.xlsx')
test2.to_excel('M2TEST.xlsx')
test.to_excel('M1TEST.xlsx')
```

String Report



(a) (b)

Figure A.3. String report for a (a) visually degraded and (b) healthy module.

Figure A.3 compares a healthy PV panel with a visually degraded one. The degraded module shows a 22.4% power loss, generating only 251.7 W, with Vmpp at 34.2 V and Impp at 7.36 A. The distorted V-I curve reflects performance issues affecting output. Key parameters such as global irradiance, ambient and module

temperatures, and tilt angle were monitored using a PV200 irradiance analyser with Pt100 and back-surface vacuum-based temperature sensors**.**

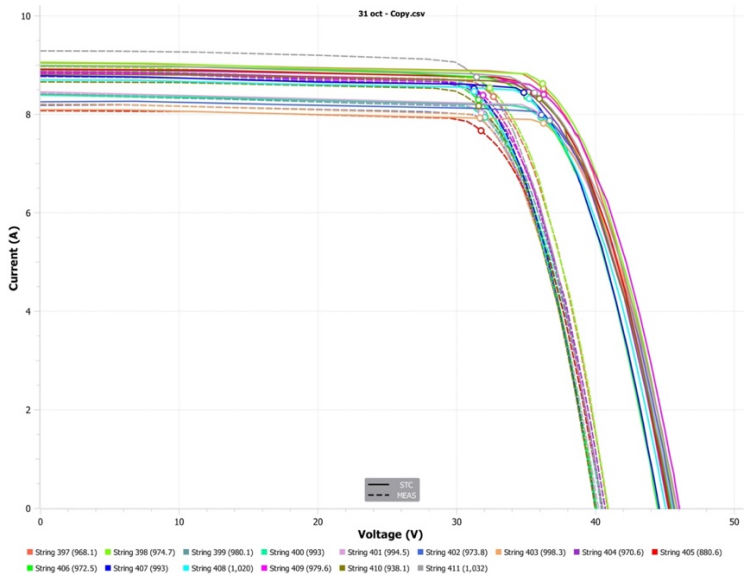


Figure A.4. V-I curve of the selected 13 modules.

The measured curves (MEAS) exhibit reduced knee sharpness and lower fill factors compared to the ideal STC curves, confirming performance degradation. Notably, String 405 (880.6 W) shows significant deviation, while others like String 401 and String 410 also fall short of expected output, validating that all 13 modules are degraded to varying degrees. This degradation is evident through lowered voltage at maximum power and broadened knee regions in the curves as seen in figure A.4.

Appendix – B

Python code for dual layer and multi-layer hybrid based deep learning, artificial intelligence models that predict exergy efficiency in Chapter 5.

Table B.1.Input sample for the Dual layer and Hybrid layer models with Exergy Efficiency(%) as output.

S.No	Date	Irradiance [kWh/m2]	Module Temp [°C]	A	W(m/s)	Exergy Eff %
1	01-Jan-21	5.84	49.82	0.00278996	3.54	20.89
2	02-Jan-21	5.2	46.73	0.00276862	4.36	21.16
3	03-Jan-21	5.85	48.19	0.0027597	4.06	21.03
4	04-Jan-21	4.14	41.7	0.00273504	3.23	21.61
5	05-Jan-21	4.65	44.48	0.0027577	4.76	21.36
6	06-Jan-21	5.42	46.84	0.00275861	5.55	21.15
7	07-Jan-21	3.03	36.98	0.00270994	6.8	22.02
8	08-Jan-21	4.18	39.27	0.00268909	4.66	21.81
9	09-Jan-21	5.14	46.99	0.00277663	5.57	21.14
10	10-Jan-21	5.75	50.62	0.00280955	3.63	20.82
11	11-Jan-21	7.32	56.29	0.00282719	2.93	20.32
12	12-Jan-21	5.87	51.3	0.00281542	4.06	20.76
13	13-Jan-21	7.03	54.68	0.00281359	3.23	20.46
14	14-Jan-21	7.45	55.28	0.00280149	4.76	20.40
15	15-Jan-21	7	51.57	0.00275843	5.55	20.72

16	16-Jan-21	7.02	51.19	0.00275044	3.63	20.75
17	17-Jan-21	6.93	52.07	0.00277135	3.23	20.68
18	18-Jan-21	6.52	51.21	0.00277809	4.76	20.76
19	19-Jan-21	6.46	51.26	0.00278229	5.55	20.76
20	20-Jan-21	7.29	53.76	0.00278247	6.8	20.53
21	21-Jan-21	7.45	52.26	0.00274645	4.66	20.66
22	22-Jan-21	7.09	52.35	0.00276771	5.57	20.66
23	23-Jan-21	6.8	51.8	0.00277353	3.63	20.71
24	24-Jan-21	7.24	52.6	0.00276407	2.93	20.63
25	25-Jan-21	7	51.29	0.00275334	4.06	20.75
26	26-Jan-21	7.11	50.78	0.00273811	3.63	20.79
27	27-Jan-21	7.04	50.31	0.00273341	3.23	20.83
28	28-Jan-21	6.78	49.5	0.00273287	4.76	20.90
29	29-Jan-21	7.14	51.39	0.00274754	5.55	20.74
30	30-Jan-21	6.96	51.92	0.00276698	6.8	20.89

(1) (Transformer+BiLSTM)-Dual Layer hybrid model

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from tensorflow.keras.models import Model
```

```
from tensorflow.keras.layers import Input, Bidirectional, LSTM,
MultiHeadAttention, LayerNormalization, Add, Dense, Dropout
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import EarlyStopping
from keras_tuner import BayesianOptimization
import shutil  # For clearing tuner directories

# Step 1: Load Data from Excel
file_path = '/Users/Almas/Desktop/3 jan.xlsx'  # Adjust this to your file path
data = pd.read_excel(file_path, sheet_name='khopoli', engine='openpyxl')

# Step 2: Select Features and Target
features = ['Irradiance [kWh/m2]', 'A', 'Module Temp [°C]', 'W(m/s)']
target = 'eff'

# Step 3: Normalize Features and Target using MinMaxScaler
scaler = MinMaxScaler(feature_range=(0, 1))
scaled_data = scaler.fit_transform(data[features])

target_scaler = MinMaxScaler(feature_range=(0, 1))
scaled_target = target_scaler.fit_transform(data[[target]])

# Step 4: Create Dataset using a Time Window
def create_dataset(features, target, time_step=1):
    X, y = [], []
    for i in range(len(features) - time_step):
        X.append(features[i:i+time_step])
        y.append(target[i+time_step])
    return np.array(X), np.array(y)

time_step = 6  # Adjustable time window
X, y = create_dataset(scaled_data, scaled_target, time_step)
```

```
# Step 5: Train-Test Split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
shuffle=False)

# Reshape data for LSTM + Transformer input
X_train = X_train.reshape(X_train.shape[0], X_train.shape[1], X_train.shape[2])
X_test = X_test.reshape(X_test.shape[0], X_test.shape[1], X_test.shape[2])

# Step 6: Define the Bayesian Optimization Function for BiLSTM + Transformer
def build_model_bilstm_transformer(hp):
    inputs = Input(shape=(X_train.shape[1], X_train.shape[2]))

    # BiLSTM layer
    lstm_out = Bidirectional(LSTM(
        units=hp.Int('lstm_units', min_value=32, max_value=256, step=32),
        return_sequences=True
    ))(inputs)

    # Transformer block: Multi-Head Attention + Add & Norm
    attention_output = MultiHeadAttention(
        num_heads=hp.Int('num_heads', min_value=2, max_value=8, step=2),
        key_dim=hp.Int('key_dim', min_value=32, max_value=128, step=32)
    )(lstm_out, lstm_out)  # Self-attention on BiLSTM output

    attention_output = Add()([lstm_out, attention_output])  # Residual connection
    attention_output = LayerNormalization()(attention_output)  # Layer
normalization

    # Fully connected layer
    dense_out = Dense(
        units=hp.Int('dense_units', min_value=32, max_value=256, step=32),
        activation='relu'
```

```
    )(attention_output[:, -1, :])  # Use the last time step output


    # Output layer
    outputs = Dense(1)(dense_out)


    model = Model(inputs, outputs)
    model.compile(optimizer=Adam(hp.Choice('learning_rate', [1e-2, 1e-3, 1e-4])),
loss='mean_squared_error')
    return model


# Step 7: Clear Tuner Directory and Run Optimization
shutil.rmtree('model_dir', ignore_errors=True)


# Early stopping
early_stopping = EarlyStopping(monitor='val_loss', patience=20,
restore_best_weights=True)


# Tuning BiLSTM + Transformer Model
tuner_bilstm_transformer = BayesianOptimization(
    build_model_bilstm_transformer,
    objective='val_loss',
    max_trials=20,  # Adjustable
    directory='model_dir',
    project_name='bilstm_transformer_optimization'
)
tuner_bilstm_transformer.search(X_train, y_train, epochs=300,
validation_data=(X_test, y_test), callbacks=[early_stopping], verbose=1)
best_bilstm_transformer_model =
tuner_bilstm_transformer.get_best_models(num_models=1)[0]


# Step 8: Evaluate Model with Test Data
best_bilstm_transformer_model_loss =
best_bilstm_transformer_model.evaluate(X_test, y_test)
```

```
print(f"Best BiLSTM + Transformer Model Test Loss:
{best_bilstm_transformer_model_loss}")


# Step 9: Make Predictions

bilstm_transformer_predictions = best_bilstm_transformer_model.predict(X_test)

bilstm_transformer_predictions =
target_scaler.inverse_transform(bilstm_transformer_predictions)


# Actual inverse-transformed values for comparison

y_test_actual = target_scaler.inverse_transform(y_test)


# Step 10: Calculate Metrics

def calculate_metrics(actual, predicted):

    rmse = np.sqrt(np.mean((actual - predicted) ** 2))

    mape = np.mean(np.abs((actual - predicted) / actual)) * 100

    mbe = np.mean(predicted - actual)

    return rmse, mape, mbe


bilstm_transformer_rmse, bilstm_transformer_mape, bilstm_transformer_mbe =
calculate_metrics(y_test_actual, bilstm_transformer_predictions)

print(f'BiLSTM + Transformer Metrics\nRMSE:
{bilstm_transformer_rmse}\nMAPE: {bilstm_transformer_mape}%\nMBE:
{bilstm_transformer_mbe}')


# Displaying the Predicted and Actual Values

print("\nPredicted vs Actual Values for BiLSTM + Transformer Model:")

result_df = pd.DataFrame({

    'Actual Exergy Loss': y_test_actual.flatten(),

    'Predicted Exergy Loss': bilstm_transformer_predictions.flatten()

})


print(result_df.head())  # Display first few rows of predictions vs actual

Trial 20 Complete [03h 04m 43s]
```

val_loss: 0.062436493784189224

Best val_loss So Far: 0.017258113250136375

BiLSTM + Transformer Metrics

RMSE: 0.5283755334062161

MAPE: 1.878042569441318%

MBE: -0.05097474606289182

Predicted vs Actual Values for BiLSTM + Transformer Model:

	Actual Exergy Loss	Predicted Exergy Loss
0	19.871917	20.688335
1	20.976677	20.360285
2	22.306444	20.871979
3	22.523671	21.589977
4	22.637264	21.953180

(2) (GRU+BiLSTM)-Dual Layer hybrid model

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import GRU, LSTM, Bidirectional, Dense, Dropout, Input
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import EarlyStopping
from keras_tuner import BayesianOptimization
import shutil  # For clearing tuner directories

# Step 1: Load Data from Excel
file_path = '/Users/Almas/Desktop/3 jan.xlsx'
data = pd.read_excel(file_path, sheet_name='khopoli', engine='openpyxl')
```

```
# Step 2: Select Features and Target
features = ['Irradiance [kWh/m2]', 'AA', 'Module Temp [°C]', 'W(m/s)']
target = 'eff'

# Step 3: Normalize Features and Target using MinMaxScaler
scaler = MinMaxScaler(feature_range=(0, 1))
scaled_data = scaler.fit_transform(data[features])

target_scaler = MinMaxScaler(feature_range=(0, 1))
scaled_target = target_scaler.fit_transform(data[[target]])

# Step 4: Create Dataset using a Time Window
def create_dataset(features, target, time_step=1):
    X, y = [], []
    for i in range(len(features) - time_step):
        X.append(features[i:i+time_step])
        y.append(target[i+time_step])
    return np.array(X), np.array(y)

time_step = 6  # Adjustable time window
X, y = create_dataset(scaled_data, scaled_target, time_step)

# Step 5: Train-Test Split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
shuffle=False)

# Reshape data for GRU + BiLSTM input
X_train = X_train.reshape(X_train.shape[0], X_train.shape[1], X_train.shape[2])
X_test = X_test.reshape(X_test.shape[0], X_test.shape[1], X_test.shape[2])

# Step 6: Define the Bayesian Optimization Function for GRU + BiLSTM
def build_model_gru_bilstm(hp):
```

```
    model = Sequential()
    model.add(Input(shape=(X_train.shape[1], X_train.shape[2])))

    # GRU layer
    model.add(GRU(
        units=hp.Int('gru_units', min_value=32, max_value=256, step=32),
        return_sequences=True,
        recurrent_dropout=hp.Choice('gru_recurrent_dropout', [0.0, 0.2, 0.3]),
        activation='tanh'   ))
    model.add(Dropout(hp.Choice('gru_dropout', [0.0, 0.2, 0.3])))

    # BiLSTM layer (Bidirectional LSTM)
    model.add(Bidirectional(LSTM(
        units=hp.Int('lstm_units', min_value=32, max_value=256, step=32),
        return_sequences=False,
        recurrent_dropout=hp.Choice('lstm_recurrent_dropout', [0.0, 0.2, 0.3]) )))
    model.add(Dense(1))  # Output layer
    model.compile(optimizer=Adam(hp.Choice('learning_rate', [1e-2, 1e-3, 1e-4])),
loss='mean_squared_error')
    return model

# Step 7: Clear Tuner Directory and Run Optimization
shutil.rmtree('model_dir', ignore_errors=True)

# Early stopping
early_stopping = EarlyStopping(monitor='val_loss', patience=20,
restore_best_weights=True)

# Tuning GRU + BiLSTM Model
tuner_gru_bilstm = BayesianOptimization(
    build_model_gru_bilstm,
    objective='val_loss',
```

```
    max_trials=20,  # Adjustable
    directory='model_dir',
    project_name='gru_bilstm_optimization'
)
tuner_gru_bilstm.search(X_train, y_train, epochs=300, validation_data=(X_test, y_test), callbacks=[early_stopping], verbose=1)
best_gru_bilstm_model = tuner_gru_bilstm.get_best_models(num_models=1)[0]

# Step 8: Evaluate Model with Test Data
best_gru_bilstm_model_loss = best_gru_bilstm_model.evaluate(X_test, y_test)
print(f"Best GRU + BiLSTM Model Test Loss: {best_gru_bilstm_model_loss}")

# Step 9: Make Predictions
gru_bilstm_predictions = best_gru_bilstm_model.predict(X_test)
gru_bilstm_predictions = target_scaler.inverse_transform(gru_bilstm_predictions)

# Actual inverse-transformed values for comparison
y_test_actual = target_scaler.inverse_transform(y_test)

# Step 10: Calculate Metrics
def calculate_metrics(actual, predicted):
    rmse = np.sqrt(np.mean((actual - predicted) ** 2))
    mape = np.mean(np.abs((actual - predicted) / actual)) * 100
    mbe = np.mean(predicted - actual)
    return rmse, mape, mbe

gru_bilstm_rmse, gru_bilstm_mape, gru_bilstm_mbe = calculate_metrics(y_test_actual, gru_bilstm_predictions)
print(f'GRU + BiLSTM Metrics\nRMSE: {gru_bilstm_rmse}\nMAPE: {gru_bilstm_mape}%\nMBE: {gru_bilstm_mbe}')

# Displaying the Predicted and Actual Values
print("\nPredicted vs Actual Values for GRU + BiLSTM Model:")
```

```
result_df = pd.DataFrame({
    'Actual Exergy eff': y_test_actual.flatten(),
    'Predicted Exergy eff': gru_bilstm_predictions.flatten()
})

print(result_df.head())  # Display first few rows of predictions vs actual
```

```
Trial 20 Complete [04h 01m 05s]
val_loss: 0.01691400073468685
Best val_loss So Far: 0.016774073243141174
Best GRU + BiLSTM Model Test Loss: 0.016774073243141174
GRU + BiLSTM Metrics
RMSE: 0.5209130927750637
MAPE: 1.9981766304735083%
MBE: -0.04018690823484284

Predicted vs Actual Values for GRU + BiLSTM Model:
   Actual Exergy eff  Predicted Exergy eff
0          19.871917             20.599123
1          20.976677             20.304026
2          22.306444             20.715054
3          22.523671             21.723122
4          22.637264             21.934214
```

(3) (CNN+BiLSTM)-Dual Layer hybrid model

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv1D, MaxPooling1D, LSTM, Dense,
Dropout, Input, Flatten
from tensorflow.keras.optimizers import Adam
```

```
from tensorflow.keras.callbacks import EarlyStopping
from keras_tuner import BayesianOptimization
import shutil  # For clearing tuner directories

# Step 1: Load Data from Excel
file_path = '/Users/Almas/Desktop/Book2.xlsx'  # Adjust this to your file path
data = pd.read_excel(file_path, sheet_name='khopoli', engine='openpyxl')

# Step 2: Select Features and Target
features = ['Irradiance [kWh/m2]', 'Ambient Temp [°C]', 'Module Temp [°C]',
'Total Yield [kWh]', 'W(m/s)']
target = 'Exergy loss'

# Step 3: Normalize Features and Target using MinMaxScaler
scaler = MinMaxScaler(feature_range=(0, 1))
scaled_data = scaler.fit_transform(data[features])

target_scaler = MinMaxScaler(feature_range=(0, 1))
scaled_target = target_scaler.fit_transform(data[[target]])

# Step 4: Create Dataset using a Time Window
def create_dataset(features, target, time_step=1):
    X, y = [], []
    for i in range(len(features) - time_step):
        X.append(features[i:i+time_step])
        y.append(target[i+time_step])
    return np.array(X), np.array(y)

time_step = 6  # Adjustable time window
X, y = create_dataset(scaled_data, scaled_target, time_step)

# Step 5: Train-Test Split
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
shuffle=False)


# Reshape data for CNN + LSTM input
X_train = X_train.reshape(X_train.shape[0], X_train.shape[1], X_train.shape[2])
X_test = X_test.reshape(X_test.shape[0], X_test.shape[1], X_test.shape[2])


# Step 6: Define the Bayesian Optimization Function for CNN + LSTM
def build_model_cnn_lstm(hp):
    model = Sequential()
    model.add(Input(shape=(X_train.shape[1], X_train.shape[2])))


    # Convolutional layers
    model.add(Conv1D(
        filters=hp.Int('filters', min_value=32, max_value=256, step=32),
        kernel_size=hp.Choice('kernel_size', [2, 3, 4]),
        activation='relu'
    ))
    model.add(MaxPooling1D(pool_size=2))
    model.add(Dropout(hp.Choice('conv_dropout', [0.0, 0.2, 0.3])))


    # LSTM layer
    model.add(LSTM(
        units=hp.Int('lstm_units', min_value=32, max_value=256, step=32),
        return_sequences=False,
        recurrent_dropout=hp.Choice('lstm_recurrent_dropout', [0.0, 0.2, 0.3])
    ))
    model.add(Dense(1))  # Output layer


    model.compile(optimizer=Adam(hp.Choice('learning_rate', [1e-2, 1e-3, 1e-4])),
loss='mean_squared_error')
    return model
```

```
# Step 7: Clear Tuner Directory and Run Optimization
shutil.rmtree('model_dir', ignore_errors=True)

# Early stopping
early_stopping = EarlyStopping(monitor='val_loss', patience=20,
restore_best_weights=True)

# Tuning CNN + LSTM Model
tuner_cnn_lstm = BayesianOptimization(
    build_model_cnn_lstm,
    objective='val_loss',
    max_trials=20,  # Adjustable
    directory='model_dir',
    project_name='cnn_lstm_optimization'
)
tuner_cnn_lstm.search(X_train, y_train, epochs=300, validation_data=(X_test,
y_test), callbacks=[early_stopping], verbose=1)
best_cnn_lstm_model = tuner_cnn_lstm.get_best_models(num_models=1)[0]

# Step 8: Evaluate Model with Test Data
best_cnn_lstm_model_loss = best_cnn_lstm_model.evaluate(X_test, y_test)
print(f"Best CNN + LSTM Model Test Loss: {best_cnn_lstm_model_loss}")

# Step 9: Make Predictions
cnn_lstm_predictions = best_cnn_lstm_model.predict(X_test)
cnn_lstm_predictions = target_scaler.inverse_transform(cnn_lstm_predictions)

# Actual inverse-transformed values for comparison
y_test_actual = target_scaler.inverse_transform(y_test)

# Step 10: Calculate Metrics
```

```
def calculate_metrics(actual, predicted):
    rmse = np.sqrt(np.mean((actual - predicted) ** 2))
    mape = np.mean(np.abs((actual - predicted) / actual)) * 100
    mbe = np.mean(predicted - actual)
    return rmse, mape, mbe


cnn_lstm_rmse, cnn_lstm_mape, cnn_lstm_mbe =
calculate_metrics(y_test_actual, cnn_lstm_predictions)
print(f'CNN + LSTM Metrics\nRMSE: {cnn_lstm_rmse}\nMAPE:
{cnn_lstm_mape}%\nMBE: {cnn_lstm_mbe}')
rial 20 Complete [03h 00m 47s]
val_loss: 0.022112706676125526
Best val_loss So Far: 0.02206093817949295
Best CNN + LSTM Model Test Loss: 0.02206093817949295
CNN + LSTM Metrics
RMSE: 0.549408596465
MAPE: 2. 43363231308%
MBE: 0.01946032617643
```

(4) (Autoencoder+BiLSTM)-Dual Layer hybrid model

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from tensorflow.keras.models import Model
from tensorflow.keras.layers import LSTM, Dense, Dropout, Input, RepeatVector,
Bidirectional
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import EarlyStopping
from keras_tuner import BayesianOptimization
import shutil  # For clearing tuner directories


# Step 1: Load Data from Excel
```

```
file_path = '/Users/Almas/Desktop/3 jan.xlsx'  # Adjust this to your file path
data = pd.read_excel(file_path, sheet_name='khopoli', engine='openpyxl')

# Step 2: Select Features and Target
features = ['Irradiance [kWh/m2]', 'AA', 'Module Temp [°C]', 'W(m/s)']
target = 'eff'

# Step 3: Normalize Features and Target using MinMaxScaler
scaler = MinMaxScaler(feature_range=(0, 1))
scaled_data = scaler.fit_transform(data[features])

target_scaler = MinMaxScaler(feature_range=(0, 1))
scaled_target = target_scaler.fit_transform(data[[target]])

# Step 4: Create Dataset using a Time Window
def create_dataset(features, target, time_step=1):
    X, y = [], []
    for i in range(len(features) - time_step):
        X.append(features[i:i+time_step])
        y.append(target[i+time_step])
    return np.array(X), np.array(y)

time_step = 6  # Adjustable time window
X, y = create_dataset(scaled_data, scaled_target, time_step)

# Step 5: Train-Test Split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
shuffle=False)

# Reshape data for Autoencoder + BiLSTM input
X_train = X_train.reshape(X_train.shape[0], X_train.shape[1], X_train.shape[2])
X_test = X_test.reshape(X_test.shape[0], X_test.shape[1], X_test.shape[2])
```

```
# Step 6: Define the Bayesian Optimization Function for Autoencoder + BiLSTM
def build_model_autoencoder_bilstm(hp):
    inputs = Input(shape=(X_train.shape[1], X_train.shape[2]))

    # Encoder: BiLSTM Layer
    encoded = Bidirectional(LSTM(
        units=hp.Int('encoder_units', min_value=32, max_value=256, step=32),
        activation='relu',
        return_sequences=False
    ))(inputs)
    encoded = Dropout(hp.Choice('encoder_dropout', [0.0, 0.2, 0.3]))(encoded)

    # Decoder: RepeatVector and BiLSTM Layer
    decoded = RepeatVector(X_train.shape[1])(encoded)
    decoded = Bidirectional(LSTM(
        units=hp.Int('decoder_units', min_value=32, max_value=256, step=32),
        activation='relu',
        return_sequences=True
    ))(decoded)

    # LSTM for prediction
    lstm_out = Bidirectional(LSTM(
        units=hp.Int('lstm_units', min_value=32, max_value=256, step=32),
        return_sequences=False
    ))(decoded)

    # Output layer
    outputs = Dense(1)(lstm_out)

    model = Model(inputs, outputs)
```

```
    model.compile(optimizer=Adam(hp.Choice('learning_rate', [1e-2, 1e-3, 1e-4])),
loss='mean_squared_error')
    return model


# Step 7: Clear Tuner Directory and Run Optimization
shutil.rmtree('model_dir', ignore_errors=True)


# Early stopping
early_stopping = EarlyStopping(monitor='val_loss', patience=20,
restore_best_weights=True)


# Tuning Autoencoder + BiLSTM Model
tuner_autoencoder_bilstm = BayesianOptimization(
    build_model_autoencoder_bilstm,
    objective='val_loss',
    max_trials=20,  # Adjustable
    directory='model_dir',
    project_name='autoencoder_bilstm_optimization'
)
tuner_autoencoder_bilstm.search(X_train, y_train, epochs=300,
validation_data=(X_test, y_test), callbacks=[early_stopping], verbose=1)
best_autoencoder_bilstm_model =
tuner_autoencoder_bilstm.get_best_models(num_models=1)[0]


# Step 8: Evaluate Model with Test Data
best_autoencoder_bilstm_model_loss =
best_autoencoder_bilstm_model.evaluate(X_test, y_test)
print(f"Best Autoencoder + BiLSTM Model Test Loss:
{best_autoencoder_bilstm_model_loss}")


# Step 9: Make Predictions
autoencoder_bilstm_predictions =
best_autoencoder_bilstm_model.predict(X_test)
autoencoder_bilstm_predictions =
target_scaler.inverse_transform(autoencoder_bilstm_predictions)
```

```
# Actual inverse-transformed values for comparison
y_test_actual = target_scaler.inverse_transform(y_test)


# Step 10: Calculate Metrics
def calculate_metrics(actual, predicted):
    rmse = np.sqrt(np.mean((actual - predicted) ** 2))
    mape = np.mean(np.abs((actual - predicted) / actual)) * 100
    mbe = np.mean(predicted - actual)
    return rmse, mape, mbe


autoencoder_bilstm_rmse, autoencoder_bilstm_mape, autoencoder_bilstm_mbe = calculate_metrics(y_test_actual, autoencoder_bilstm_predictions)
print(f'Autoencoder + BiLSTM Metrics\nRMSE: {autoencoder_bilstm_rmse}\nMAPE: {autoencoder_bilstm_mape}%\nMBE: {autoencoder_bilstm_mbe}')


# Displaying the Predicted and Actual Values
print("\nPredicted vs Actual Values for Autoencoder + BiLSTM Model:")
result_df = pd.DataFrame({
    'Actual Exergy eff': y_test_actual.flatten(),
    'Predicted Exergy eff': autoencoder_bilstm_predictions.flatten()
})


print(result_df.head())  # Display first few rows of predictions vs actual
Trial 20 Complete [02h 01m 03s]
val_loss: 0.017170753329992294


Best val_loss So Far: 0.017063194885849953
Total elapsed time: 00h 35m 04s
Best Autoencoder + BiLSTM Model Test Loss: 0.017063194885849953
Autoencoder + BiLSTM Metrics
RMSE: 0.5221831399230905
```

MAPE: 1.9678231423258707%

MBE: -0.004548247212925195

Predicted vs Actual Values for Autoencoder + BiLSTM Model:

	Actual Exergy eff	Predicted Exergy eff
0	19.871917	20.662722
1	20.976677	20.364653
2	22.306444	20.858887
3	22.523671	21.433508
4	22.637264	21.800680

(5) (ATTENTION+BiLSTM)-Dual Layer hybrid model

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from tensorflow.keras.models import Model
from tensorflow.keras.layers import LSTM, Dense, Dropout, Input, Attention, Bidirectional
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import EarlyStopping
from keras_tuner import BayesianOptimization
import shutil  # For clearing tuner directories

# Step 1: Load Data from Excel
file_path = '/Users/Almas/Desktop/Book2.xlsx'  # Adjust this to your file path
data = pd.read_excel(file_path, sheet_name='khopoli', engine='openpyxl')
```

```
# Step 2: Select Features and Target
features = ['Irradiance [kWh/m2]', 'Ambient Temp [°C]', 'Module Temp [°C]', 'Total Yield [kWh]', 'W(m/s)']
target = 'Exergy loss'

# Step 3: Normalize Features and Target using MinMaxScaler
scaler = MinMaxScaler(feature_range=(0, 1))
scaled_data = scaler.fit_transform(data[features])

target_scaler = MinMaxScaler(feature_range=(0, 1))
scaled_target = target_scaler.fit_transform(data[[target]])

# Step 4: Create Dataset using a Time Window
def create_dataset(features, target, time_step=1):
    X, y = [], []
    for i in range(len(features) - time_step):
        X.append(features[i:i+time_step])
        y.append(target[i+time_step])
    return np.array(X), np.array(y)

time_step = 6  # Adjustable time window
X, y = create_dataset(scaled_data, scaled_target, time_step)

# Step 5: Train-Test Split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, shuffle=False)

# Reshape data for BiLSTM + Attention input
X_train = X_train.reshape(X_train.shape[0], X_train.shape[1], X_train.shape[2])
X_test = X_test.reshape(X_test.shape[0], X_test.shape[1], X_test.shape[2])

# Step 6: Define the Bayesian Optimization Function for BiLSTM + Attention
```

```
def build_model_bilstm_attention(hp):
    inputs = Input(shape=(X_train.shape[1], X_train.shape[2]))

    # BiLSTM layer
    bilstm_out = Bidirectional(LSTM(
        units=hp.Int('bilstm_units', min_value=32, max_value=256, step=32),
        return_sequences=True,
        recurrent_dropout=hp.Choice('bilstm_recurrent_dropout', [0.0, 0.2, 0.3])
    ))(inputs)

    # Attention mechanism
    attention_out = Attention()([bilstm_out, bilstm_out])

    # Dropout layer
    attention_out = Dropout(hp.Choice('dropout', [0.0, 0.2, 0.3]))(attention_out)

    # Dense output layer
    outputs = Dense(1)(attention_out[:, -1, :])  # Use the last step in sequence

    model = Model(inputs, outputs)
    model.compile(optimizer=Adam(hp.Choice('learning_rate', [1e-2, 1e-3, 1e-4])),
loss='mean_squared_error')
    return model

# Step 7: Clear Tuner Directory and Run Optimization
shutil.rmtree('model_dir', ignore_errors=True)

# Early stopping
early_stopping = EarlyStopping(monitor='val_loss', patience=20,
restore_best_weights=True)

# Tuning BiLSTM + Attention Model
```

```
tuner_bilstm_attention = BayesianOptimization(
    build_model_bilstm_attention,
    objective='val_loss',
    max_trials=20,  # Adjustable
    directory='model_dir',
    project_name='bilstm_attention_optimization'
)
tuner_bilstm_attention.search(X_train, y_train, epochs=300,
validation_data=(X_test, y_test), callbacks=[early_stopping], verbose=1)
best_bilstm_attention_model =
tuner_bilstm_attention.get_best_models(num_models=1)[0]


# Step 8: Evaluate Model with Test Data
best_bilstm_attention_model_loss =
best_bilstm_attention_model.evaluate(X_test, y_test)
print(f"Best BiLSTM + Attention Model Test Loss:
{best_bilstm_attention_model_loss}")


# Step 9: Make Predictions
bilstm_attention_predictions = best_bilstm_attention_model.predict(X_test)
bilstm_attention_predictions =
target_scaler.inverse_transform(bilstm_attention_predictions)


# Actual inverse-transformed values for comparison
y_test_actual = target_scaler.inverse_transform(y_test)


# Step 10: Calculate Metrics
def calculate_metrics(actual, predicted):
    rmse = np.sqrt(np.mean((actual - predicted) ** 2))
    mape = np.mean(np.abs((actual - predicted) / actual)) * 100
    mbe = np.mean(predicted - actual)
    return rmse, mape, mbe
```

```
bilstm_attention_rmse, bilstm_attention_mape, bilstm_attention_mbe =
calculate_metrics(y_test_actual, bilstm_attention_predictions)

print(f'BiLSTM + Attention Metrics\nRMSE: {bilstm_attention_rmse}\nMAPE:
{bilstm_attention_mape}%\nMBE: {bilstm_attention_mbe}')


# Displaying the Predicted and Actual Values

print("\nPredicted vs Actual Values for BiLSTM + Attention Model:")

result_df = pd.DataFrame({

   'Actual Exergy Loss': y_test_actual.flatten(),

   'Predicted Exergy Loss': bilstm_attention_predictions.flatten()

})


print(result_df.head())  # Display first few rows of predictions vs actual
```

```
Trial 20 Complete [03h 00m 31s]

val_loss: 0.018429750576615334

Best val_loss So Far: 0.017129385843873024

Best BiLSTM + Attention Model Test Loss: 0.017129385843873024

BiLSTM + Attention Metrics

RMSE: 0.5264011569250185

MAPE: 1.9809820736835474%

MBE: -0.0025308423504004375


Predicted vs Actual Values for BiLSTM + Attention Model:

   Actual Exergy Loss  Predicted Exergy Loss

0         19.871917           20.651711

1         20.976677           20.298040

2         22.306444           20.826788

3         22.523671           21.630520

4         22.637264           21.818758
```

6) Multi-layer Hybrid Model

```
import pandas as pd
import numpy as np
from sklearn.preprocessing import MinMaxScaler
from sklearn.model_selection import train_test_split
from tensorflow.keras.models import Model
from tensorflow.keras.layers import (
    Input, Conv1D, MaxPooling1D, Flatten, Dense, Dropout, LSTM, Bidirectional,
    LayerNormalization, MultiHeadAttention, Add, GlobalAveragePooling1D,
Reshape
)
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import EarlyStopping
from keras_tuner import BayesianOptimization
import shutil  # For clearing tuner directories
import tensorflow as tf

# Load Dataset
file_path = '/Users/Almas/Desktop/8 jan.xlsx'  # Adjust this to your file path
data = pd.read_excel(file_path, sheet_name='khopoli', engine='openpyxl')

# Feature Selection
features = ['Irradiance [kWh/m2]', 'Module Temp [°C]', 'W(m/s)','AA']
target = 'eff'

# Data Normalization
scaler = MinMaxScaler(feature_range=(0, 1))
scaled_data = scaler.fit_transform(data[features])
target_scaler = MinMaxScaler(feature_range=(0, 1))
scaled_target = target_scaler.fit_transform(data[[target]])

# Create Time-Windowed Dataset
def create_dataset(features, target, time_step=1):
```

```
    X, y = [], []
    for i in range(len(features) - time_step):
        X.append(features[i:i + time_step])
        y.append(target[i + time_step])
    return np.array(X), np.array(y)


time_step = 6  # Adjustable time window
X, y = create_dataset(scaled_data, scaled_target, time_step)


# Train-Test Split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
shuffle=False)


# Reshape Inputs
X_train = X_train.reshape(X_train.shape[0], X_train.shape[1], X_train.shape[2])
X_test = X_test.reshape(X_test.shape[0], X_test.shape[1], X_test.shape[2])


# Define CNN-Autoencoder + BiLSTM + Transformer with Attention
def build_model(hp):
    inputs = Input(shape=(X_train.shape[1], X_train.shape[2]))


    # CNN-Autoencoder
    x = Conv1D(
        filters=hp.Int('filters', min_value=32, max_value=256, step=32),
        kernel_size=hp.Choice('kernel_size', [2, 3, 4]),
        activation='relu',
    )(inputs)
    x = MaxPooling1D(pool_size=2)(x)
    x = Dropout(hp.Choice('conv_dropout', [0.0, 0.2, 0.3]))(x)


    # Skip Flatten, maintain temporal features
```

```
    x = Dense(hp.Int('dense_units', min_value=32, max_value=256, step=32),
activation='relu')(x)

    # Use Reshape layer to adjust the shape
    x = Reshape((1, -1))(x)  # Reshape to (batch_size, time_steps=1, features)

    # BiLSTM
    x = Bidirectional(LSTM(
        units=hp.Int('lstm_units', min_value=32, max_value=256, step=32),
        return_sequences=True,
        recurrent_dropout=hp.Choice('lstm_recurrent_dropout', [0.0, 0.2, 0.3])
    ))(x)

    # Transformer Block
    attention_output = MultiHeadAttention(
        num_heads=hp.Int('num_heads', min_value=2, max_value=8, step=2),
        key_dim=hp.Int('key_dim', min_value=32, max_value=128, step=16),
    )(x, x)
    attention_output = Add()([attention_output, x])  # Residual connection
    attention_output = LayerNormalization()(attention_output)

    # Global Pooling
    x = GlobalAveragePooling1D()(attention_output)
    x = Dense(1, activation='linear')(x)  # Output Layer

    model = Model(inputs=inputs, outputs=x)
    model.compile(
        optimizer=Adam(hp.Choice('learning_rate', [1e-2, 1e-3, 1e-4])),
        loss='mean_squared_error'
    )
    return model
```

```
# Bayesian Optimization Setup
shutil.rmtree('model_dir', ignore_errors=True)  # Clear any existing tuner directory

tuner = BayesianOptimization(
    build_model,
    objective='val_loss',
    max_trials=20,  # Number of hyperparameter search trials
    directory='model_dir',
    project_name='cnn_bilstm_transformer'
)

# Early Stopping
early_stopping = EarlyStopping(monitor='val_loss', patience=20, restore_best_weights=True)

# Model Tuning
tuner.search(X_train, y_train, epochs=300, validation_data=(X_test, y_test), callbacks=[early_stopping], verbose=1)
best_model = tuner.get_best_models(num_models=1)[0]

# Evaluate Model
test_loss = best_model.evaluate(X_test, y_test)
print(f"Test Loss: {test_loss}")

# Predictions
predictions = best_model.predict(X_test)
predictions = target_scaler.inverse_transform(predictions)

# Actual Values
y_test_actual = target_scaler.inverse_transform(y_test)

# Metrics Calculation
```

```
def calculate_metrics(actual, predicted):
    rmse = np.sqrt(np.mean((actual - predicted) ** 2))
    mape = np.mean(np.abs((actual - predicted) / actual)) * 100
    mbe = np.mean(predicted - actual)
    return rmse, mape, mbe

rmse, mape, mbe = calculate_metrics(y_test_actual, predictions)
print(f"Metrics:\nRMSE: {rmse}\nMAPE: {mape}%\nMBE: {mbe}")

# Predictions vs Actual Values
results_df = pd.DataFrame({
    'Actual Exergy Efficiecy': y_test_actual.flatten(),
    'Predicted Exergy Efficiency': predictions.flatten()
})
print(results_df.head())
import os
import time
import pandas as pd

# Create a directory for saving Excel files
os.makedirs("model_weights", exist_ok=True)

Trial 20 Complete [04h 02m 32s]
val_loss: 0.026814928278326988

Best val_loss So Far: 0.019106239080429077
Test Loss: 0.019106239080429077
Metrics:
RMSE: 0.5559473147419626
MAPE: 0.019381031069111%
MBE: -0.021124523327554006
   Actual Exergy Efficiency  Predicted Exergy Efficiency
```

0	22.028599	21.478889
1	21.815967	21.744724
2	21.146011	21.697275
3	20.827744	21.176590
4	20.321071	20.95099

Appendix-C

Python code for A User Interactive Tool for Assessment of Performance Ratio for Commercial Solar Photovoltaic system used in Chapter 6 and 7.

Table C.1.Input sample for the PR model

S.No	Date	I(kWh/m2)	Exergy loss	Rd
1	01/01/21	0.936	80629	5.2
2	01/01/21	1.016	87710	5.2
3	01/01/21	1.112	95998	5.2
4	01/01/21	1.224	105438	5.2
5	01/01/21	1.328	114397	5.2
6	01/01/21	1.432	123489	5.2
7	01/01/21	1.568	134924	5.2
8	01/01/21	1.888	162636	3.33
9	01/01/21	2.816	242574	3.63
10	01/01/21	2.96	255256	3.86
11	01/01/21	3.184	274275	4.18
12	01/01/21	3.368	289809	4.47
13	01/01/21	3.48	299121	4.7
14	01/01/21	3.736	320775	4.97
15	01/01/21	3.912	335886	5.26
16	01/01/21	4	344191	5.46
17	01/01/21	4.296	370062	5.85
18	01/01/21	4.448	382739	6.15
19	01/01/21	4.592	395129	6.43
20	01/01/21	4.824	415092	6.7
21	01/01/21	4.896	421288	6.97
22	01/01/21	5.184	445584	7.35
23	01/01/21	5.28	454328	7.54
24	01/01/21	5.4	463644	7.79
25	01/01/21	5.576	479798	8.01
26	01/01/21	5.792	498382	8.31
27	01/01/21	5.92	508842	8.52
28	01/01/21	6.032	519034	8.74
29	01/01/21	6.152	528784	8.89
30	01/01/21	6.256	538307	9.06

```
import tkinter as tk
from tkinter import messagebox
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import math
from matplotlib.backends.backend_tkagg import FigureCanvasTkAgg
from sklearn.metrics import mean_squared_error
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, LSTM, Dropout

# Load the data from the Excel file
file_path = '/Users/ALMAS/Desktop/block1.xlsx'
df_performance = pd.read_excel(file_path)
df_performance.index = df_performance['Date']
df_performance = df_performance[['Rd', 'Exergy loss', 'I(kwh/m2)', 'PR']]

# Function to create trainable dataset
def create_trainable_dataset(dataframe, n_inputs, n_outputs):
    X, Y = [], []
    for i in range(len(dataframe) - n_inputs - n_outputs + 1):
        X.append(dataframe.iloc[i:(i + n_inputs), :].values)
        Y.append(dataframe.iloc[i + n_inputs:i + n_inputs + n_outputs, -1].values)
    return np.array(X), np.array(Y)

# LSTM Model for Performance Ratio Prediction
def build_lstm_model(X_train):
    model = Sequential()
    model.add(LSTM(256, input_shape=(X_train.shape[1], X_train.shape[2]),
return_sequences=True))
    model.add(Dropout(0.6))
```

```
    model.add(LSTM(256, return_sequences=True))
    model.add(Dropout(0.6))
    model.add(LSTM(256, return_sequences=True))
    model.add(Dropout(0.6))
    model.add(LSTM(256))
    model.add(Dropout(0.6))
    model.add(Dense(1))
    model.compile(loss='mae', optimizer='adam', metrics=['accuracy'])
    return model


train_loss = []
val_loss = []


def train_model():
    global model, X_train, Y_train, train_loss, val_loss
    model = build_lstm_model(X_train)
    history = model.fit(X_train, Y_train, epochs=500, batch_size=2,
validation_data=(X_test, Y_test), verbose=1, shuffle=True)
    train_loss = history.history['loss']
    val_loss = history.history['val_loss']


def predict_performance():
    global model, X_test, Y_test, rmse
    predicted_values = model.predict(X_test)
    rmse = np.sqrt(mean_squared_error(Y_test, predicted_values))
    message = f"Predicted Performance Ratio: {predicted_values[-1][0]}, RMSE:
{rmse}"
    messagebox.showinfo("Predicted Performance and RMSE", message)


def plot_actual_vs_train_PR():
    plt.plot(df_performance.index, df_performance['PR'], label='Actual PR')
    plt.plot(df_performance.index[:len(Y_train)], Y_train.flatten(), label='Train
PR')
```

```
    plt.xlabel('Date')
    plt.ylabel('Performance Ratio')
    plt.title('Actual vs Train PR')
    plt.legend()

def plot_actual_vs_train_PR_in_gui():
    fig = plt.figure()
    plot_actual_vs_train_PR()
    canvas = FigureCanvasTkAgg(fig, master=performance_frame)
    canvas.draw()
    canvas.get_tk_widget().pack(side=tk.TOP, fill=tk.BOTH, expand=1)

# Split data into train and test sets
n_input = 2
n_output = 1
train_size = 0.8

train_days = math.floor(len(df_performance) * train_size)
train_data, test_data = df_performance.iloc[0:train_days],
df_performance.iloc[train_days:len(df_performance)]

X_train, Y_train = create_trainable_dataset(train_data, n_input, n_output)
X_test, Y_test = create_trainable_dataset(test_data, n_input, n_output)

# Function to calculate exergy loss
def calculate_exergy_loss():
    try:
        # Get user input parameters
        wind_speed = float(wind_speed_entry.get())
        ambient_temp = float(ambient_temp_entry.get())
        voc = float(voc_entry.get())
        isc = float(isc_entry.get())
```

```
        vmpp = float(vmpp_entry.get())
        impp = float(impp_entry.get())
        latitude = float(latitude_entry.get())
        longitude = float(longitude_entry.get())
        length = float(length_entry.get())  # New entry for panel length
        width = float(width_entry.get())    # New entry for panel width
        G = float(irradiance_entry.get())    # New entry for irradiance

        # Known parameters
        NOCT = 41  # Nominal operating cell temperature (NOCT)
        sigma = 5.67e-8  # Stefan Boltzmann constant (σ)
        epsilon = 0.9  # Emissivity of the panel
        Ts = 5780  # Sun temperature

        # Calculate additional parameters
        Tsky = ambient_temp - 6
        Tm = ambient_temp + (NOCT - 20) * 800   # Change module_temp to Tm
        Tm = Tm / 1000  # Convert Tm to Kelvin

        # Calculate exergy loss
        A_PV = length * width  # Panel area
        A = A_PV * G * (1 - 4 / 3 * (ambient_temp / Ts) + 1 / 3 * (ambient_temp /
Ts) ** 4)
        hcon = 2.8 + 3 * wind_speed
        hrad = epsilon * sigma * ((Tsky + Tm) * (Tsky ** 2 + Tm ** 2))
        U = hcon + hrad
        Q = U * (Tm - ambient_temp)
        Exther = Q * (1 - ambient_temp / Tm)
        Exele = voc * isc * (vmpp - impp) /(voc-isc)
        FF = (vmpp - impp) / (voc - isc)
        Exele = voc * isc * FF
        Exout = Exele - Exther
```

```
        Exin = A * (1 - (ambient_temp / Tm))
        Exloss = Exin - Exout

        # Display the absolute value of the calculated exergy loss with unit
        exergy_loss_label.config(text=f"Exergy Loss: {abs(Exloss):.2f} W")

        # Display Tm with unit
        tm_label.config(text=f"Tm: {Tm:.2f} K")

        # Display FF
        ff_label.config(text=f"Fill Factor: {FF:.2f}")

    except ValueError:
        messagebox.showerror("Error", "Please enter valid numeric values.")

def plot_monthly_PR_variation_at_(t) ():
    monthly_PR = df_performance['PR'].resample('M').mean()
    fig = plt.figure()
    plt.plot(monthly_PR.index, monthly_PR.values)
    plt.xlabel('Month')
    plt.ylabel('Performance Ratio')
    plt.title('Monthly Variation of Performance Ratio')
    canvas = FigureCanvasTkAgg(fig, master=monthly_PR_canvas)
    canvas.draw()
    canvas.get_tk_widget().pack(side=tk.TOP, fill=tk.BOTH, expand=True)

def plot_PR_forecast_t1_t2_t3():
    import matplotlib.pyplot as plt

    # Prepare the input sequence (last 'n_input' entries from df_performance)
    input_seq = df_performance[['Rd', 'Exergy loss', 'I(kWh/m2)']].values[-
n_input:].reshape(1, n_input, -1)
```

```
    predicted_values = []
    input_temp = input_seq.copy()

    # Recursive forecast for t+1, t+2, t+3
    for _ in range(3):
        pred = model.predict(input_temp)[0][0]
        predicted_values.append(pred)

        # Create next input by appending predicted PR
        next_input = np.hstack([input_temp[0][-1], pred]).reshape(1, -1)  # Add
dimension
        # For simplicity, replicate last known inputs and append predicted PR
        next_features = input_temp[0][1:].tolist()
        next_features.append(input_temp[0][-1])  # reusing last timestep for now
        input_temp = np.array([next_features])

    # Create plot
    fig, ax = plt.subplots(figsize=(6, 4))
    ax.plot([1, 2, 3], predicted_values, marker='o', linestyle='--', color='purple',
label='Forecasted PR')
    ax.set_title('Predicted PR for (t+1), (t+2), (t+3)')
    ax.set_xlabel('Minute Steps Ahead')
    ax.set_ylabel('Performance Ratio (%)')
    ax.legend()

    # Embed in Tkinter GUI
    forecast_canvas = FigureCanvasTkAgg(fig, master=performance_frame)
    forecast_canvas.draw()
    forecast_canvas.get_tk_widget().pack(side=tk.TOP, fill=tk.BOTH,
expand=True)
```

```
def plot_exergy_loss_variation():
    monthly_exergy_loss = df_performance['Exergy loss'].resample('M').sum()
    fig = plt.figure()
    plt.plot(monthly_exergy_loss.index, monthly_exergy_loss.values)
    plt.xlabel('Month')
    plt.ylabel('Exergy Loss (W)')
    plt.title('Monthly Variation of Exergy Loss')
    canvas = FigureCanvasTkAgg(fig, master=monthly_exergy_loss_canvas)
    canvas.draw()
    canvas.get_tk_widget().pack(side=tk.TOP, fill=tk.BOTH, expand=True)


# Create GUI
root = tk.Tk()
root.title("Performance Ratio Prediction and Exergy Loss Calculator")


# Frame for performance ratio prediction
performance_frame = tk.Frame(root, bd=2, relief=tk.GROOVE)
performance_frame.pack(padx=10, pady=10, fill=tk.BOTH, expand=True)


train_button = tk.Button(performance_frame, text="Train Model",
command=train_model)
train_button.pack()


forecast_button = tk.Button(performance_frame, text="Plot PR Variation at (t+1),
(t+2), (t+3)", command=plot_PR_forecast_t1_t2_t3)
forecast_button.pack()
predict_button = tk.Button(performance_frame, text="Predict PR",
command=predict_performance)
predict_button.pack()


plot_button = tk.Button(performance_frame, text="Forecast PR",
command=plot_actual_vs_train_PR_in_gui)
plot_button.pack()
```

```
# Frame for exergy loss calculator
exergy_frame = tk.Frame(root, bd=2, relief=tk.GROOVE)
exergy_frame.pack(padx=10, pady=10, fill=tk.BOTH, expand=True)

# Labels and Entry widgets for input parameters
# (Entry widgets and labels for input parameters are here)
# Labels and Entry widgets for input parameters
wind_speed_label = tk.Label(exergy_frame, text="Wind Speed (m/s):")
wind_speed_label.grid(row=0, column=0, padx=5, pady=5)
wind_speed_entry = tk.Entry(exergy_frame)
wind_speed_entry.grid(row=0, column=1, padx=5, pady=5)

ambient_temp_label = tk.Label(exergy_frame, text="Ambient Temperature
(°C):")
ambient_temp_label.grid(row=1, column=0, padx=5, pady=5)
ambient_temp_entry = tk.Entry(exergy_frame)
ambient_temp_entry.grid(row=1, column=1, padx=5, pady=5)

voc_label = tk.Label(exergy_frame, text="Open-circuit Voltage (V):")
voc_label.grid(row=2, column=0, padx=5, pady=5)
voc_entry = tk.Entry(exergy_frame)
voc_entry.grid(row=2, column=1, padx=5, pady=5)

isc_label = tk.Label(exergy_frame, text="Short-circuit Current (A):")
isc_label.grid(row=3, column=0, padx=5, pady=5)
isc_entry = tk.Entry(exergy_frame)
isc_entry.grid(row=3, column=1, padx=5, pady=5)

vmpp_label = tk.Label(exergy_frame, text="Maximum Power Point Voltage
(Vmpp) (V):")
vmpp_label.grid(row=4, column=0, padx=5, pady=5)
```

```
vmpp_entry = tk.Entry(exergy_frame)
vmpp_entry.grid(row=4, column=1, padx=5, pady=5)

impp_label = tk.Label(exergy_frame, text="Maximum Power Point Current (Impp) (A):")
impp_label.grid(row=5, column=0, padx=5, pady=5)
impp_entry = tk.Entry(exergy_frame)
impp_entry.grid(row=5, column=1, padx=5, pady=5)

latitude_label = tk.Label(exergy_frame, text="Latitude:")
latitude_label.grid(row=6, column=0, padx=5, pady=5)
latitude_entry = tk.Entry(exergy_frame)
latitude_entry.grid(row=6, column=1, padx=5, pady=5)

longitude_label = tk.Label(exergy_frame, text="Longitude:")
longitude_label.grid(row=7, column=0, padx=5, pady=5)
longitude_entry = tk.Entry(exergy_frame)
longitude_entry.grid(row=7, column=1, padx=5, pady=5)

length_label = tk.Label(exergy_frame, text="Panel Length (cm):")
length_label.grid(row=8, column=0, padx=5, pady=5)
length_entry = tk.Entry(exergy_frame)
length_entry.grid(row=8, column=1, padx=5, pady=5)

width_label = tk.Label(exergy_frame, text="Panel Width (cm):")
width_label.grid(row=9, column=0, padx=5, pady=5)
width_entry = tk.Entry(exergy_frame)
width_entry.grid(row=9, column=1, padx=5, pady=5)

irradiance_label = tk.Label(exergy_frame, text="Irradiance (W/m^2):")
irradiance_label.grid(row=10, column=0, padx=5, pady=5)
irradiance_entry = tk.Entry(exergy_frame)
```

```
irradiance_entry.grid(row=10, column=1, padx=5, pady=5)

# Button to calculate exergy loss
calculate_button =tk.Button(exergy_frame, text="Calculate Exergy Loss",
command=calculate_exergy_loss)
calculate_button.grid(row=12, columnspan=2, padx=5, pady=10)

# Label to display exergy loss
exergy_loss_label = tk.Label(exergy_frame, text="")
exergy_loss_label.grid(row=13, columnspan=2, padx=5, pady=5)

# Label to display Tm
tm_label = tk.Label(exergy_frame, text="")
tm_label.grid(row=14, columnspan=2, padx=5, pady=5)

# Label to display fill factor
ff_label = tk.Label(exergy_frame, text="")
ff_label.grid(row=15, columnspan=2, padx=5, pady=5)

# Button to plot monthly variation of PR
monthly_PR_button = tk.Button(root, text="Plot Monthly PR Variation",
command=plot_monthly_PR_variation)
monthly_PR_button.pack(side=tk.LEFT)

# Button to plot monthly variation of exergy loss
monthly_exergy_loss_button = tk.Button(root, text="Plot Monthly Exergy Loss
Variation", command=plot_monthly_exergy_loss_variation)
monthly_exergy_loss_button.pack(side=tk.LEFT)

# Canvas for monthly PR variation plot
monthly_PR_canvas = tk.Canvas(root)
monthly_PR_canvas.pack(side=tk.RIGHT, padx=10, pady=10, fill=tk.BOTH,
expand=True)
```

```
# Canvas for monthly exergy loss variation plot

monthly_exergy_loss_canvas = tk.Canvas(root)

monthly_exergy_loss_canvas.pack(side=tk.RIGHT, padx=10, pady=10,
fill=tk.BOTH, expand=True)

root.mainloop()
```

```
Epoch 001/500

4/4 [==============================] - 0s 124ms/step - loss: 3.1801 -
accuracy: 0.0000e+00 - val_loss: 0.1654 - val_accuracy: 0.0000e+00

Epoch 002/500

4/4 [==============================] - 0s 82ms/step - loss: 3.9389 -
accuracy: 0.0000e+00 - val_loss: 0.1901 - val_accuracy: 0.0000e+00

.

.

.

Epoch 496/500

4/4 [==============================] - 0s 83ms/step - loss: 3.3280 -
accuracy: 0.0000e+00 - val_loss: 0.4126 - val_accuracy: 0.0000e+00

Epoch 497/500

4/4 [==============================] - 0s 82ms/step - loss: 3.1774 -
accuracy: 0.0000e+00 - val_loss: 0.3701 - val_accuracy: 0.0000e+00

Epoch 498/500

4/4 [==============================] - 0s 103ms/step - loss: 2.3155 -
accuracy: 0.0000e+00 - val_loss: 0.4234 - val_accuracy: 0.0000e+00

Epoch 499/500

4/4 [==============================] - 0s 113ms/step - loss: 1.7306 -
accuracy: 0.0000e+00 - val_loss: 0.3729 - val_accuracy: 0.0000e+00

Epoch 500/500

4/4 [==============================] - 0s 98ms/step - loss: 3.1963 -
accuracy: 0.0000e+00 - val_loss: 0.3553 - val_accuracy: 0.0000e+0
```