REPUBLIC OF TÜRKİYE

ALTINBAŞ UNIVERSITY

Institute of Graduate Studies

Electrical and Computer Engineering

OPTIMIZING PHOTOVOLTAIC SYSTEM DIAGNOSTICS: INTEGRATING MACHINE LEARNING AND DBFLA FOR ADVANCED FAULT DETECTION AND CLASSIFICATION

Omar Mohammed Nsaif AL-QARAGHULI

Ph.D. Dissertation

Supervisor

Asst. Prof. Dr. Abdullahi Abdu IBRAHIM

İstanbul, 2025

OPTIMIZING PHOTOVOLTAIC SYSTEM DIAGNOSTICS: INTEGRATING MACHINE LEARNING AND DBFLA FOR ADVANCED FAULT DETECTION AND CLASSIFICATION

Omar Mohammed Nsaıf AL-QARAGHULI

Electrical and Computer Engineering

Ph.D. Dissertation

ALTINBAŞ UNIVERSITY

2025

The dissertation titled OPTIMIZING PHOTOVOLTAIC SYSTEM DIAGNOSTICS: INTEGRATING MACHINE LEARNING AND DBFLA FOR ADVANCED FAULT DETECTION AND CLASSIFICATION prepared by OMAR MOHAMMED NSAIF AL-QARAGHULI and submitted on 12/05/2025 has been **accepted unanimously** for the degree of PhD in Electrical and Computer Engineering

Asst. Prof. Dr. Abdullahi Abdu IBRAHIM

Supervisor

Thesis Defense Committee Members:

Asst. Prof. Dr. Abdullahi Abdu IBRAHIM	Department of Computer Engineering, Altınbaş University	
Assoc.Prof. Dr. Sefer KURNAZ	Department of Computer Engineering, Altınbaş University	
Prof. Dr. Osman Nuri UÇAN	Department of Electrical-Electronic Engineering, Altınbaş University	
Asst. Prof. Dr. Serdar KARGIN	Department of Biomedical Engineering, İstanbul Arel University	
Asst. Prof. Dr. Tariq Abed MOHAMMED	Department of Computer Science and Information Kirkuk University	

I hereby declare that this dissertation meets all format and submission requirements for a PhD thesisThesis Defenc

I hereby declare that all information/data presented in this graduation project has been obtained in full accordance with academic rules and ethical conduct. I also declare all unoriginal materials and conclusions have been cited in the text and all references mentioned in the Reference List have been cited in the text, and vice versa as required by the abovementioned rules and conduct.

Omar Mohammed Nsaıf AL-QARAGHULI

Signature

DEDICATION

This work is dedicated to my mother, my first teacher, my father, who was my role model and constant support throughout this journey. Thanks to you, this dream has come true. And of course, to my siblings, who partnered with me to make my dream a reality.

ABSTRACT

OPTIMIZING PHOTOVOLTAIC SYSTEM DIAGNOSTICS: INTEGRATING MACHINE LEARNING AND DBFLA FOR ADVANCED FAULT DETECTION AND CLASSIFICATION

AL-QARAGHULI, Omar Mohammed Nsaıf

Ph.D., Electrical and Computer Engineering, Altınbaş University

Supervisor: Asst. Prof. Dr. Abdullahi Abdu IBRAHIM

Date: 05/2025

Pages: 120

Because of the exponential development in the number of photovoltaic (PV) power plant installations, standard inspection procedures have become ineffective. As a consequence of this, there is a requirement for more sophisticated approaches to the identification and categorisation of faults. In order to do this, the DBFLA method, which is a unique hybrid metaheuristic technique, will be presented in this paper. Based on the Dung Beetle Optimisation Algorithm and Fick's Law of Diffusion Algorithm, this technique is aimed to answer the issues that have been highlighted. It is a hybrid of the two algorithms. The DBFLA improves the efficiency of machine learning models such as ANN, SVM, and ensemble approaches by modifying their parameters in order to improve the precision of fault detection. This is accomplished through the process of adjusting the parameters. The identification of problems such as open circuits, short circuits, and module incompatibilities may be accomplished in a rapid and precise way. In accordance with the results of the study, DBFLA is able to construct a stacking classifier by making use of actual PV datasets. This classification strategy achieves an individual meta-learner accuracy of around 98.75%, which is a substantial improvement above the performance of standard machine learning approaches. The capacity of this technique to handle a bigger number of operating modes and a broader diversity of issue scenarios has resulted in the development of enhanced fault detection systems. This evolved as a subsequent consequence of the method's ability to accommodate these capabilities. The improvement in classification accuracy is the most important contribution that DBFLA brings to the table, in contrast to the optimisation procedures that have always been used before. The capability of the approach to properly strike a balance between exploration and exploitation is directly responsible for this advancement on the part of the method. This hybrid approach is provided in order to demonstrate how actual and simulated information may be merged to bring about PV defect detection algorithms that are both more accurate and more efficient. The goal of future research is to improve the operational efficiency and reliability of photovoltaic (PV) systems by incorporating these complicated models into real-time monitoring systems. This will be accomplished through the adoption of advanced modelling techniques.

Keywords: Photovoltaic Systems, Machine Learning, Fault Detection, DBFLA, Hybrid Optimization.

TABLE OF CONTENTS

Pages

LIST OF TABLES

Pages

LIST OF FIGURES

Pages

1. INTRODUCTION

1.1 BACKGROUND

As a consequence of the increasing need for energy sources that are less harmful to the environment and the improvements in technology, photovoltaic (PV) energy generation has emerged as an important component in the production of renewable energy. Because they provide a more ecologically friendly alternative to fossil fuels, solar photovoltaic (PV) systems are a key component of the worldwide transition to sustainable energy. This is because PV systems are environmentally benign. There are a number of issues that need to be resolved with regard to the generation of power by photovoltaic (PV) technology. Among these issues include its relatively poor energy conversion efficiency, its limited operating lifespan, and worries over disposal at the end of its useful life. PV power generation does greatly reduce emissions of greenhouse gases and dependency on non-renewable resources; yet, these concerns still remain despite the fact that PV power generating does so. However, the real lifespan is often closer to twelve to fifteen years, and in harsh conditions, it is substantially shorter [1]. This is despite the fact that manufacturers expect a lifespan of twenty to thirty years for their products when they are manufactured. When photovoltaics (PV) are included into the power system, it is necessary to take into account a number of extra factors. These factors include the storage of energy, the stability of the network, and the recycling of modules that have become outdated. The ongoing development of photovoltaic (PV) technology, optimisation strategies, and fault detection systems is making a substantial contribution to the improvement of the system's capabilities in terms of efficiency, reliability, and sustainability. Despite the obstacles that have been encountered, this is the situation that has arisen. The momentum towards the widespread use of solar power as a source of electricity has been accelerated by significant expenditures, the majority of which have been made in Europe. As a result, it is of the highest significance to make certain that photovoltaic (PV) systems continue to perform efficiently and consistently over a long length of time. This is necessary in order to guarantee energy efficiency and financial feasibility. There are a number of issues that can significantly hamper the operation of big photovoltaic (PV) systems. Some of these issues include electrical mismatches, breakdowns that are not reported, and modules that are deteriorating [2]. Maintenance procedures that are considered to be normal, such as disassembly and visual inspections, are not only time-consuming and costly, but they are also typically hazardous. Despite the fact that there is an urgent need for the detection of flaws in a timely and accurate way, this is the situation that has arisen. When it comes to ensuring that solar facilities will have a high degree of operational efficiency and financial returns, it is very necessary to enhance diagnostic procedures. [3].

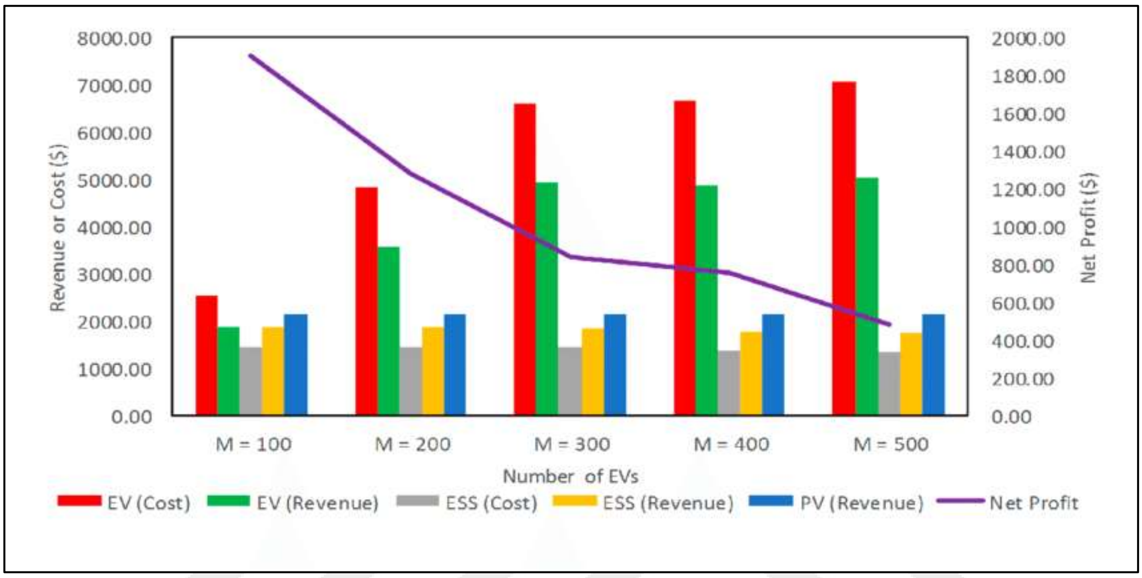

Figure 1.1: Global Financial Returns of RE Market [4].

The movement towards renewable energy has sharply increased the importance of photovoltaic (PV) systems as clean, sustainable, and economical sources of energy. With respect to residential, commercial, and utility-scale installations, there is a rapid increase in the deployment of PV systems. Therefore, proper functionality and efficiency is crucial. Meanwhile, PV systems are subject to several faults such as performance and lifespan degrading partial shading, open circuits, soiling, degrading, and inverter faults. Additionally, these systems are prone to various other malfunctions that can be detrimental to their performance and longevity. PV system maintenance is made especially challenging due to the manual checking that needs to be performed, and it also takes an inordinate amount of time. The traditional monitoring system is described as a simple threshold based system, and this monitoring approach is insufficient. Recent advancements in technology have integrated PV systems with ML (Machine Learning) solutions to automate PV diagnostics.

Figure 1.2: Factors that Affect the Fault Detection [6].

Unlike traditional approaches, ML algorithms framework offers the ability to sort through large datasets pertaining to the system and operational data, revealing numerous patterns—classifying fault types with exceptional precision and predicting future system behavior. Despite the advantages machine learning systems offer, many existing models suffer from overfitting tendencies, inadequate general reasoning capabilities, and poor adaptation towards dynamic environments, especially ones with unbalanced datasets or conflicting fault markers. In overcoming these obstacles, this thesis suggests merging the Deep-Based Fuzzy Logic Algorithm (DBFLA) with other machine learning techniques to enhance the diagnostic accuracy of PV systems. The DBFLA framework derives its advantages from fuzzy logic systems due to the interpretability, robustness, and deep learning capabilities of modeling complex nonlinear relationships. The joint effort strives to improve defect detection and classification accuracy, scalability, and real-time application in PV systems. Using historical datasets in conjunction with real-time sensor data, the proposed model aims to reliably identify fault categories while being adaptable within a range of environmental conditions and PV configurations. This research provides advanced intelligent diagnostic systems for proactive maintenance, decreasing downtime, and maximizing energy yield in PV installations. In turn, this contributes to the objective of creating efficient, resilient, and data-centric renewable energy systems.

1.2 PROBLEM STATEMENT

The adoption of photovoltaic (PV) systems marks a significant development in the quest for practical alternatives to sustainable sources of energy, as they are now used globally. The need for further mechanisms to control the emission of greenhouse gases IPCC Report 2023- the World is facing a constant struggle with pollution, climate change, and the scarcity of existing sources of fossil fuel energy. Depletion of non-renewable sources of energy has led to browse a viable technological solution, which is photovoltaic technology. In spite of the technological advances in PV systems, their reliability, efficiency, and overall performance are still hindered by multiple technical and operational factors. The foremost issue appears to be fault occurrence and inefficient mechanisms for sophisticated diagnostics such as detection classifying them automatically or even manually.

There are numerous sophisticated systems diagnostic that can detect PV system faults deriving from environmental factors such as natural hardware failures, aging of its components, or even installation mistakes. With the assistance of these diagnostics advanced engineer, vous will be able to tackle problems of partial shading open eat other forms of circuitry like parallel short circuits on dirty frames, as well as degradation of solar panels neutral connectors and cables, issues with inverters and blaming their costs beside mismatches. Any more complex regions these faults may lead to problems lie range within losing parts of system and battery function up to permanent damage and in many cases these damages can render a piece of equipment unusable. Moreover, in numerous scenarios these faults have tendency to go unnoticed for long time periods. It leads to significant losses of energy return on investment drastic increase on maintenance costs and devalues the VAST amount of resources people invest into these systems.

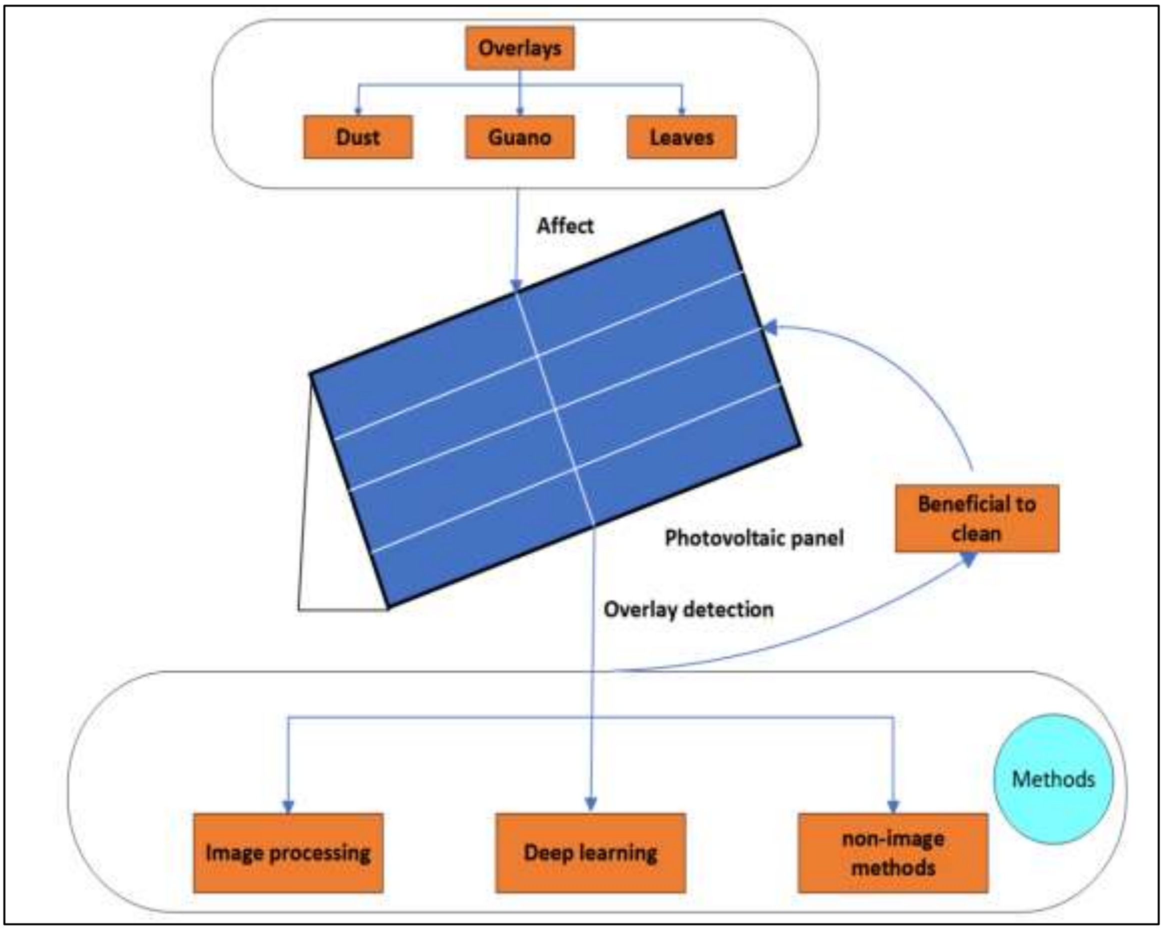

Figure 1.3: Overlay Detection Methods [8]

Furthermore, unaddressed problems may worsen over time, leading to subsequent failures and higher chances of fire or electrical damage. Faults can be wrongly diagnosed due to visual examinations, thermography, I-V curve monitoring, and manual checking, which are deeply labor-intensive, reactive, and do not allow granular real-time fault detection. These conventional approaches require a great deal of effort, time, and equipment, making them unfeasible for PV farms or other remote installations. Moreover, these methods do not have the precision or flexibility required to differentiate between overlapping or complex fault signatures, particularly in changing weather conditions like varying irradiance, temperature, and humidity. As a result, there is an urgent need for intelligent automated diagnostic systems that provide precise and timely assessment of PV system performance. Although the use of machine learning (ML) in automated diagnostics of PV system faults is promising, there are practical challenges facing existing approaches based on ML. A large number of these models rely on abundant labeled datasets that are costly, especially for unconventional or transient fault conditions. Furthermore, these models tend to lack generalization across different configurations or environments of a PV System that were not exposed to during training.

The interpretability challenges attributed to the black-box nature of a model makes it harder for a system's operators and engineers to comprehend how and why specific faults were classified the way they were. Also, traditional ML algorithms face challenges such as noise susceptibility, which impacts the strength and dependability of the algorithm in real-world settings. This is exacerbated by the existence of data imbalance – which is frequent within PV system datasets – leading to weakened robustness, dependability, and reliability in practical uses. These obstacles highlight the necessity for an advanced flexible, transparent, adaptive, and robust diagnostic framework capable of leveraging ML's pattern recognition capabilities while invoking fuzzy logic's reasoning strengths to address uncertainty and ambiguity pervasive in PV system data. Applying Deep-Based Fuzzy Logic Algorithms (DBFLA) alongside ML algorithms provides a new approach to dealing with the insufficient existing diagnostic systems. Designing DBFLA involves modeling nonlinear functions, managing vague or uncertain, undefined data, incomplete pieces, undefined elements, offering rule-based reasoning frameworks, where the outcome can be enhanced through rules provided, and improving data explainability. Systems based on DBFLA combine the power of deep learning feature extraction and the flexible decision-making of fuzzy logic, making their approach powerful and easily interpretable. Other issues arise with the creation and implementation of such hybrid systems.

Creating a suitable DBFLA model based on deep learning frameworks requires customization of fuzzy membership functions, feature selection, tuning of inference rules, and the development of a deep neural network hybrid architecture. Achieving a balance between polycentric contour precision, computational cost, and efficiency is difficult. Moreover, evaluating the performance of the models in an extensive range of environments and conditions within a dynamic PV system involves complex procedures for d-simulation and validation with both synthetic and real world data. In this regard, the primary issue of this thesis is formulated as there is an intelligent, scalable, and interpretable diagnostic system that operates with high accuracy and low latency to continuously detect, classify, and track multiple fault scenarios PV systems simultaneously. Solutions available in the market do not adequately address the operational challenges of real-time diagnostics or lack the complexity to capture the detail and non-linearity of fault behavior under changing operating conditions. There is pressing demand for a solution that not only increases fault detection but encourages preventative maintenance, reduces operational expenditures, and extends the lifespan of PV installations. Consequently, my dissertation is designed to address this concern by formulating an innovative approach that employs machine learning in conjunction with Deep-Based Fuzzy Logic Algorithms for designing an automated fault detection and classification system for PV diagnostics. It is anticipated that the proposed solution will shift PV system monitoring from a manual, reactive approach into an autonomous, intelligent diagnostic procedure which will increase system availability, energy capture, and overall reliability. This study hopes to make a considerable impact on the development of smart diagnostic tools for renewable energy systems and sustainable energy infrastructures globally by trying to resolve the practical and theoretical shortcomings of the existing methods used in diagnostics.

1.3 THESIS AIMS AND OBJECTIVES

1.3.1 Thesis Aims

The primary aim of this thesis is to develop an intelligent and comprehensive diagnostic framework that improves the accuracy, operational efficiency, and fault tolerance of photovoltaic (PV) systems. This is the aim that will be realized through the creation of this framework. To achieve this aim, a seamless integration of Deep Based Fuzzy Logic Algorithms (DBFLA) with machine learning (ML) will be required. The principal objective of this study is to address the gap in overcoming the sophistication challenges posed by the Photovoltaic (PV) systems on existing methodologies for fault detection. This is possible by designing a hybrid diagnostic model which not only performs accurate detection and classification of a wide range of fault conditions including partial shading, inverter failures, soiling, and degradation, but also adapts on the fly to environmental and operational changes with minimal human oversight. This objective may be achieved through the development of a hybrid diagnostic model. The aim of this thesis is to capture significant features from massive real-time and historical datasets of PV performance using the sophisticated machine learning techniques. This will be achieved by employing the sophisticated techniques of machine learning. Moreover, the thesis aims to integrates interpretability, robustness, and human reasoning of fuzzy logic systems to deal more effectively with uncertainty, ambiguity, and overlapping fault signatures. This will be achieved through system incorporation of these elements. The model given has been designed as a real-time, scalable, and generalizable solution for a wide range of photovoltaic (PV) set configurations and deployment sizes. These range from residential rooftop systems to large solar farm installations. As an additional point of interest, this research aims to improve the explainability of the diagnostic process. This will enable system operators to understand the rationale behind problem projections, thereby supporting informed decisions for maintenance and repair operations. In addition, the goal of the dissertation is to reduce positivies and negatives during fault detection, reduce the duration of system downtime as a result of purposeful early fault identification, and maximally increase energy output along with the system's lifetime over its lifespan. The present work seeks to accomplish these aims not only through augmenting the existing level of PV diagnostics, but also through making contributions towards the broader domain of intelligent renewable energy management. This will aid the research in effectively achieving its goals. The aim of the research will be accomplished through developing a novel hybrid AI-based fault analysis system with high transparency and robustness that performs reliably in real-world settings.

1.3.2 Thesis Objectives

This thesis seeks to systematically solve the problem of fault detection and classification in photovoltaic (PV) systems, integrating machine learning (ML) and Deep-Based Fuzzy Logic Algorithms (DBFLA) into a single diagnostic system.

a. The first goal of this thesis is to perform an exhaustive review and a comprehensive critique of existing techniques for fault detection in PV systems, paying particular attention to accuracy, adaptability, scalability, and interpretability under complex and changing environmental conditions.
b. Based on these findings, the second objective is to design and implement a new hybrid diagnostic model that utilizes supervised machine learning algorithms such as decision

trees, support vector machines, or even deep neural networks along with fuzzy logic to improve the precision and reliability of identifying faults in PV systems. The model will be trained using richly heterogeneous and high-dimensional datasets comprising electrical, environmental, and performance parameters so that it can robustly learn to identify normal and faulty operation states while coping with symptoms of multiple faults, noisy data, or overly complex fault interrelations.

c. The incorporation of fuzzy logic into the third objective must be accomplished within the framework of a deep learning structure that allows for machines to deal with uncertainty and incompleteness in information successfully, further enhancing the reliability of diagnostics in the real-world scenarios.

d. The fourth objective is to apply the proposed system in a simulated environment where it will be measured against critical classification metrics like accuracy, precision, recall, F1-score, mean squared error, and also benchmarked against both traditional and contemporary diagnostic systems. Also, without losing focus on the primary goal of the thesis, the proposed model is ensured to be scalable and responsive in real time, enabling deployment in PV systems of different sizing and configurations with minimal adjustment. One of the primary goals is to refine model output interpretability by formulating machine friendly fuzzy rules along with human decision logic so that end-users like operators, engineers, and maintenance personnel will find the diagnostics plausible with no explanations required.

Ultimately, the work intends to expand the body of knowledge both academically and industrially by creating a diagnostic system that is empirically validated, well documented, and serves as a framework for future works exploring smart PV monitoring, predictive maintenance, and intelligent energy management systems.

1.4 THESIS MOTIVATIONS

This thesis is motivated by the fact that solar photovoltaic (PV) systems are essential for the world's shift to cleaner and sustainable energy. The increasing reliance of governments, organizations, and individuals on solar technology for fighting climate change, carbon emission mitigation, and energy self-sufficiency has raised concerns regarding the dependability, effectiveness, and survivability of these systems. In spite of the technological advancements in PV materials, inverters, and installations, one challenge related to the diagnostics and monitoring of a solar energy system continues to limit its full potential—system monitoring and diagnostics. The timely detection, classification, and resolution of encourred faults in the system renders diagnosing, monitoring, and management an ultimate within system energy realization, economic feasibility, and user trust mechanisms for the technology. The lack of proactive strategies of traditional PV system monitoring and diagnosis frameworks is a key motivator of this study. These inadequacies combined with rapidly changing system architectures and operational environments heighten modern complexity pose immense challenges to conventional frameworks. While commonly used as monitoring standards—or as standalone inspection procedures—manual inspections, I-V curve tracking, thermographic imaging, and threshold-based monitoring systems are inefficiently capturing fleck or superficial behaviors. Such approaches also typically do not work well in large-scale solar farms or geographically dispersed PV arrays which require scalable, real-time, and autonomous diagnostics.

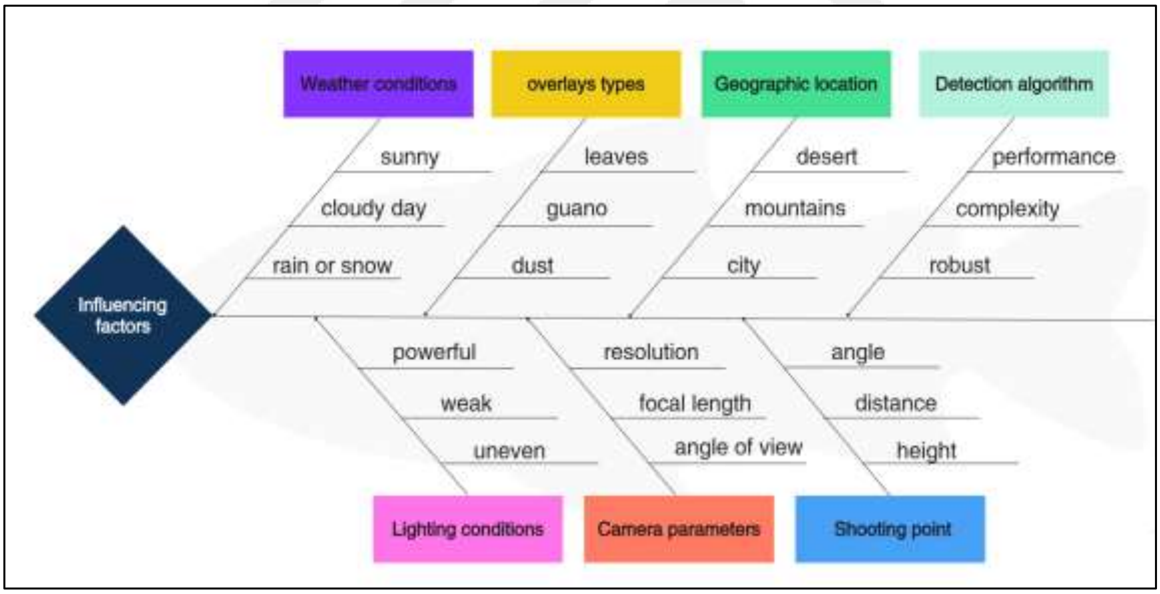

Figure 1.4: Influence Factors on PV Performance [8].

These existing methods highlight the need for more sophisticated automated methods that are intelligent, data-driven, and designed to manage the complexities of PV operation while changeable environmental conditions are in play. Another important reason is the operational data of PV systems, including voltage, current, temperature, irradiance, and power output, that is increasingly becoming available. With appropriate analysis, these datasets present great opportunities for identifying fault patterns, failure prediction, and system performance improvement. The difficulty is in developing algorithms that can efficiently provide real-time data processing and interpretation, extract actionable insights, and independently make reliable diagnostic choices without human oversight. A prospective approach has been machine learning (ML), which provides sophisticated means for pattern identification, classification, and predictive analytics. On their own, however, the ML models do tend to lack the required level of explanation, clarity, and robustness in recognition of safety-critical infrastructures like energy systems. They tend to be regarded as “black boxes,” whose decisions are not easy to explain or defend without undermining trust among system operators and stakeholders. Thus arises a stronger drive towards the integration of machine learning and fuzzy logic to form a hybrid model capable of data-driven learning and simulating human-like reasoning. Fuzzy logic is exceptionally suitable for handling uncertainty, imprecision, and non-binary systems, all of which are part of the reality for the operations of PV systems. The proposed Deep-Based Fuzzy Logic Algorithm (DBFLA) overcomes the challenges posed by stand-alone ML models as well as classical fuzzy systems. This algorithm addresses these challenges by embedding fuzzy logic rules into a framework utilizing deep learning. This integration increases accuracy and efficiency while maintaining a sufficient level of interpretability and adaptability ideal for complex energy systems needing real-time response. The motivation seeks to develop this hybrid model based on the belief it could enhance PV diagnostics and support the next intelligent energy systems. These motivations also tie to pragmatic and economic factors. In terms of availability, energy yield, and return on investment, PV system owners, operators, and investors expect high standards. Uncontrolled faults for prolonged periods lead to energy losses, expensive repairs, and shortened component lifespan, which are highly detrimental. In competitive energy markets with tight efficiency margins, even minor enhancements in fault detection and system uptime yield substantial monetary benefits. Additionally, from a maintenance point of view, accurate and timely diagnostics lessen the requirements for unnecessary inspections and operational standstills while allowing for condition-based or predictive maintenance, which is significantly more economical than traditional reactive maintenance. This research fosters the economic sustainability of PV systems and advocates for wider utilization of solar energy technologies by augmenting models to improve fault prediction and classification. Motivationally, from an academic and a scientific standpoint, this thesis stems from the opportunity to contribute to the burgeoning gap at the intersection of renewable energy systems and artificial intelligence, which is relatively underexplored. While many efforts have been directed toward the application of AI onto energy systems, very few have integrated deep learning and fuzzy logic for PV fault diagnostics, which is a remaining gap this research aims to address with a new hybrid approach that improves technical performance, and builds the framework for intelligent diagnostic systems. This thesis intends to provide a foundation on AI-enabled energy monitoring technologies by creating and validating a new model hybrid algorithms, explainable AI, and adaptive diagnostics. Throughout this work, I've approached the problem with a more personal motivation: creating a sustainable tomorrow. The energy system is at a pivotal point where the choices made now will define the future's ecological, socio-economical landscape. My research aids the international transition toward clean energy by optimizing the PV system's diagnostics, reliability, and efficiency. It helps advance the infrastructure required for sustainable low-carbon economies. It attests to the fact that the most challenging problems of our time could be resolved adopting an interdisciplinary approach—it combining engineering, computer science, environmental science, and policy—fusion for innovation in interdisciplinary design. Alongside the practical and technical motivations mentioned above, this conviction serves as the basis of my thesis.

1.5 RESEARCH QUESTIONS

To steer the specific aims and objectives of this research, an in-depth analysis has been conducted, concentrating on capturing all the relevant details related to the problem at hand, as well as addressing the failure and detection challenges in photovoltaic (PV) systems. PV diagnostics optimization enabled by machine learning algorithms and Deep-Based Fuzzy Logic Algorithms (DBFLA) were considered in this analysis.

a. What gaps exist in automated fault detection and classification techniques used in photovoltaic systems, and how do these gaps influence the performance, reliability, and energy yield of the system in both small-scale and large-scale implementations?
b. Is it possible for a machine learning algorithm to provide real-time fault detection services under constantly changing environmental and operational conditions, and what

primary challenges such as data imbalance, generalization, and interpretability affect the algorithms' performance in real-life PV scenarios?

c. In the presence of overlapping fault signatures, noisy data, and other uncertain conditions, how can fuzzy logic be combined with deep learning frameworks to improve the accuracy, unyielding nature, and explanatory potential of PV fault diagnostic systems, particularly concerning explainability?
d. What are the most relevant electrical, environmental, and performance-related criteria to be used as input features for the proposed hybrid model, and how does feature selection or extraction enhance the model performance and computational efficiency?
e. In what ways does the proposed model, DBFLA-based hybrid diagnostic model, outperform existing ML models and fault detection systems in classification accuracy, recall, precision, F1-score, mean squared error, and computational efficiency?
f. Is the proposed model adaptable enough to different PV configurations and geographical locations without extensive retraining, and what tailored modifications would be needed for optimal scalability and real-time performance?
g. What methods can be used to analyze the explainability and interpretability of the diagnostic results, and how can these be rendered in a non-technical form suitable for decision-making by operators and maintenance personnel?
h. What are the impacts of the proposed hybrid model on predictive maintenance of PV systems, and how does it affect system uptime, operational expenditure, and the likelihood of critical system failures over time?
i. In what ways does combining DBFLA strategy with diagnostics of PV systems enable the adoption of smart systems for managing renewable energy resources, and what is the significance for the future of AI integrated power systems?

Addressing these questions will assist in achieving the goal of the defense, which is contributing to the advancement of intelligent diagnostic systems in photovoltaic environments while providing a practical approach to implementing strategies that enhance sustainability, reliability, and performance of solar energy systems. The input form is missing. Kindly submit the input text you'd like me to work with so that I can assist you.

1.6 RESEARCH HYPOTHESIS AND CONTRIBUTIONS

1.6.1 Research Hypothesis

The premise of this thesis is that incorporating machine learning (ML) into Deep-Based Fuzzy Logic Algorithms (DBFLA) will improve the fault detection and classification systems for photovoltaic (PV) installations in terms of accuracy, reliability, interpretability, and adaptability. Moreover, this hypothesis suggests that the hybrid diagnostic framework would outperform traditional ML-based and rule-based approaches by mitigating several PV diagnostics challenges: data imbalance, nonlinear fault dynamics, environmental unpredictability, and opacity of models. It is expected that merging data-driven analytical techniques with algorithmic human reasoning will allow for the more precise capturing and identification of intricate overlapping fault patterns, enabling decision-making in real-time, uncertain contexts, and rendering the information in a form that facilitates informed maintenance and operational decisions. It is also claimed that the model is likely to show such low sensitivity to PV system configurations and environmental parameters that manual adjustment is necessary only in the model's contained settings, thus illustrating ease of scalability for practical applications within solar energy systems.

1.6.2 Research Contributions

This study expands the scope in the area of diagnostics of renewable energies, intelligent monitoring systems, and hybrid AI models in the following three ways:

Creation of a New Hybrid Diagnostic Model: This thesis presents a new model of fault detection and classification of PV systems based on machine learning—Fuzzy Logic Deep Learning Algorithms. It proposes a solution that achieves the fine balance between accuracy and interpretability by employing deep learning for feature extraction and classification, and fuzzy logic for reasoning computation under uncertainty.

Enhanced Fault Detection Capabilities: The developed model captures an extensive range of faults in PV systems, which includes but is not limited to, partial shading, soiling, open and short circuits, inverter failures, and system degradation. Comprehensive experimental validation uncovered enhanced performance by the model concerning the most important metrics which include accuracy, precision, recall, F1-score, and mean squared error, when measured against established benchmarks in the literature.

Resistance to Real-World Constraints: The proposed hybrid model solves significant problems of existing diagnostic approaches such as inordinate sensitivity to noise, an imbalanced dataset, or unencountered behavior by the system. The system's performance in practical changing dynamic environments with constantly varying weather conditions is improved through the integration of fuzzy-logic based reasoning makes the system adaptable to complex, real-world dynamic system environments.

Interpretability and Explainability: This paper proposes the integration of fuzzy logic reasoning rules which permit the model's diagnostic decisions to be rendered in a comprehensible manner. This increases reliance and usability for system operators and maintenance staff by improving their trust in the system, enabling enhanced decision-making concerning fault diagnosis and resolution.

Dataset Curation and Feature Engineering: The research provides an in-depth study of the most relevant input features towards PV faults diagnosis, analyzing real-time sensor data as well as historical performance data. This enhances the knowledge of critical fault indicators within PV systems and aids in the construction of an accurate labeled dataset necessary for model training and evaluation.

Simulation and Validation Environment: A realistic model with simulation environment is built to validate the candidate model's performance under different fault simulations and environmental conditions. These include the simulation of synthetic data for scarce faults alongside comprehensive benchmarking with existing datasets.

Scalability and Deployment Readiness: The model has a modular structure, which makes it scalable and computationally economical. This feature enables integration with existing PV monitoring systems and IoT-based solar energy management systems. Its architecture enables deployment in both centralized and remote controlled hub-and-spoke configurations for PV systems situated in distant locations.

Contribution to the AI-Energy Research Community: This thesis considers the recent developments in the application of artificial intelligence to renewable energy resources by formulating a new type of hybrid model based on deep learning and fuzzy systems. It sets the stage for future work on explainable AI, self-adaptive diagnostics, and self-governing energy system control. Along with other contributions, these tackle significant technological and operational problems in PV system diagnostics—enabling further advances in the intelligence, reliability, and sustainability of solar energy infrastructure.

2. LITERATURE REVIEW

2.1 INTRODUCTION

The worldwide adoption of photovoltaic (PV) systems as sources of renewable energy has led to further advancements in their optimization, performance, reliability, and maintenance. Recently, fundamental concerns such as ensuring constant energy delivery and reliability within residential, commercial and utility-scale PV installations have risen along with PV deployment. However, PV systems have several conflicting weaknesses such as, complex faults caused by environmental conditions, component failures, and operational inefficiencies PV systems possess. These faults threaten system output and energy yield, while they may also permanently damage equipment, escalate maintenance costs, or increase operational downtimes in the absence of a proper detection system. This has placed accurate and timely fault detection and classification (FDC) at the center of PV system management concerns, attracting significant interest in the industrial and academic research landscape. Traditionally, the automation and robotic revolution has simplified diagnostics of PV systems through manual inspections, thermographic imaging, threshold-based techniques, and I-V curve analyses. While these approaches still have some degree of relevance, they are heavily dependent on human resources and highly meticulous processes, given that most remain slow, inefficient, or unable to operate in real-time. To address these constraints, there has been a notable shift in the incorporation of AI methodologies, more specifically ML and DL, within PV systems over the recent years. Support vector machines, decision trees, and random forests have been proposed as ML algorithms based on the automation of PV fault classification and retrieval of patterns from historical datasets. More recently, DL techniques like convolutional neural networks and recurrent neural networks have advanced more sophisticated capabilities than dealing with the non-linear relationships of systems, as well as retrieving fault features from high-level data presented in raw formats. Nonetheless, there are additional obstacles associated with interpretability, change assumptions, and variable factors while attempting to achieve optimality controlled with these outcome approaches. The vast majority of ML and DL algorithms tend to operate under the "black box" model – unparalleled accuracy guarantees coupled with absent insight into the inner workings of the system's logic. Restricted explainability results prevents sufficient trust and implementation in real-time PV operations as most engineers and technologists operating require actionable insights rather than decision-support systems. Additionally, these models can be challenged with arbitrary boundaries between classes and complex fault-feature interdependency or scattered sensors, which are frequently encountered in PV systems. To mitigate these problems, fuzzy logic with ML hybrid diagnostic frameworks are becoming more popular due to their competitive edge rational explanations. Developing models with fuzzy inference systems integrated within machine learning algorithms have been shown to not only have high accuracy but high interpretability and robustness, achieving intelligent fault diagnostics balance. In regard to PV system fault detection and classification, this chapter offers a detailed review of literature and focuses on the change from traditional methods to advanced AI-based techniques and the use and training of AI. It analyses the pros and cons of the traditional techniques, assesses the performance of different machine learning algorithms, and reviews the most recent advances in hybrid approaches, such as the inclusion of fuzzy logic in deep learning systems. Additionally, the chapter describes the most pressing issues and gaps in current diagnoses frameworks that need to be addressed from a practical standpoint, which develop the rationale for this thesis. Addressing these questions, the chapter attempts to explain the need for a new diagnostic system incorporating database, fuzzy logic, and algorithms by showing the existing approaches' limitations and the proposed model's contributions to intelligent PV system monitoring.

2.2 RELATED WORKS ON FAULT CLASSIFICATION

The identification of faults has been the subject of research in the past, and it has presented a variety of alternative methods to the problem. Some examples of these technologies are artificial intelligence models, wireless sensor networks, and communications between power line carriers. Other examples include wireless sensor information networks. There are a few of these systems that are dependent on simulated conditions, do not manage issues in real-time, or only handle a restricted range of issue kinds; yet, they do demonstrate promise in monitoring and fault classification: they are dependent on simulated circumstances. Recent technological breakthroughs have resulted in an increase in the accuracy of fault classification through the utilisation of optimisation algorithms, ensemble learning, and neural networks. In light of recent developments, this objective has been successfully realised. On the other hand, there is still a significant lack of knowledge on the enhancement of system diagnostics and the achievement of accurate, real-time problem diagnosis for large-scale photovoltaic (PV) systems. The authors made a distinction between AC-side problems, which include the inverter or the power grid, and DC-side faults, which involve the MPPT algorithm, module mismatches, shorts, and open circuits [4]. AC-side faults are more common than DC-side faults. Each and every one of these problems was categorised as AC-side defects. Inconsistencies in electrical performance can be caused by a variety of causes, including deterioration and shadowing, which in turn lead to a decline in performance. The contradictions in question may have been brought about by a wide range of various factors. Failures that occur in open circuits, on the other hand, have the potential to have a significant influence on the operation of the system, as was mentioned in [6]. For the purpose of developing real-time defect detection systems and monitoring electrical and meteorological conditions, it was suggested by authors [7] and [8] that wireless sensor networks (WSNs) may be utilised. On the other hand, it was brought to everyone's attention that these networks were too reliant on the functioning of individual sensors and were unable to appropriately diagnose defects. The low-cost sensor networks for solar plants that were introduced by [9] dealt with temperature and electrical changes as fault categories; however, they were not applicable to short circuits or other problems in general. [9] developed these networks. In spite of the fact that they addressed defect identification, the bulk of these techniques gave more attention to monitoring than they did to fault categorisation. For example, [10] employed power line carrier communication in order to provide operational data, whereas [11] utilised wireless sensor networks (WSNs) for the purpose of home photovoltaic (PV) maintenance. [12] evaluated the gains that have been achieved in AI-based techniques by diagnosing PV system breakdowns using neural networks with an accuracy of 88.89%. This was done in order to determine the progress that has been accomplished. The data that was included in the analysis was explicitly simulated. A similar approach was used by [13], which devised a two-step process that combined neural networks with mathematical models. In spite of the fact that this method was effective in obtaining a detection accuracy of 90.3%, it needed the system to be disconnected in order to find any problems. In particular studies, support vector machines have been utilised to identify short circuits in actual solar facilities with an accuracy of 94.74%. This has been successfully accomplished. Other research, on the other hand, has made use of non-linear autoregressive models (NARX) and fuzzy inference systems in order to obtain an exceptional level of fault classification accuracy (98.2% under changeable settings).

Furthermore, a great number of individuals depended on simulations for the purpose of validation; nevertheless, [17] utilised radial basis function networks and probabilistic neural networks in order to categorise concerns with detection accuracies ranging from 82.34% to 98.19%. 17 was able to accomplish this goal. Using VxI curve data, a Kernel Extreme Learning Machine was presented in [18] for the aim of fault categorisation. This achievement was accomplished by the use of the data. Despite the fact that this machine achieved a flawless score over a wide range of actual and simulated datasets, its practical utility was limited due to the fact that it was dependent on the disconnection of the system. In Table 1, which can be seen below, the limitations of the methods that are now being utilised to address the problem of failure prediction in PV systems are highlighted. These methods are currently being used to tackle the problem. These studies demonstrate a number of limitations, including a reliance on simulations, an absence of real-time capacity, and a limited fault coverage. They also disclose advancements in monitoring and categorisation procedures, which is a great development. Nevertheless, they reveal these advancements.

Table 2.1: Summary of Related Works on Fault Classification.

Ref	Contributions	Limitations
[8]	Categorized faults into DC-side (MPPT errors, module mismatches, short/open circuits) and AC-side (inverter/power grid issues).	Limited real-time capability and fault categorization.
[5]	Highlighted module mismatches affecting performance due to degradation and shadowing.	Focused only on module mismatches, not broader fault types.
[6]	Identified severe impacts of open-circuit faults on system operation.	Addressed only open-circuit faults, lacking wider coverage.
[7], [8]	Proposed WSNs to monitor climatic and electrical conditions and developed real-time fault detection systems.	Dependence on sensor functionality, limited fault categorization.
[9]	Developed low-cost sensor networks for PV plants to detect temperature and electrical variations.	Did not address short circuits and other major defects.
[10]	Utilized power line carrier communication to transmit operational data for fault diagnosis.	More focused on monitoring than fault classification.
[11]	Designed WSNs for residential PV maintenance focusing on system monitoring.	Limited to monitoring rather than fault classification.
[12]	Applied neural networks for PV fault classification with 88.89% accuracy using simulated data.	Relied solely on simulated data, limiting real-world application.
[13]	Developed a two-step method using mathematical modeling and neural networks, achieving 90.3% accuracy but requiring system disconnection.	Required system disconnection for fault identification.

Table 2.1: Summary of Related Works on Fault Classification "Table Continued".

[14], [15]	Applied NARX and fuzzy inference systems achieving 98.2% accuracy under variable conditions.	Reliance on variable conditions, not fully generalized.
[16]	Used SVM for short-circuit detection with 94.74% accuracy in real PV facilities.	Applicable to short-circuit detection only.
[17]	Employed probabilistic neural networks and radial basis function networks achieving 82.34% - 98.19% detection accuracy.	Many relied on simulations for validation, limiting practical use.
[18]	Introduced Kernel Extreme Learning Machine for defect categorization using VxI curve data, achieving up to 100% accuracy on mixed datasets.	System disconnection requirement limited practical application.

2.3 REALTED WORKS ON PV FAULT DETECTION

A number of considerations must be made in implementing an image processing-based dust detection system on PV panels, including image data collection, feature extraction and recognition, image location, and preprocessing [44]. Detailing all these aspects will likely improve the accuracy, stability, and practicality of the PV panel dust detection system. The image processing-based solar panel overlay detection algorithms are primarily image-data-driven technologies [45]. One suggested solution is outfitting a camera with the capability to monitor PV panels, allowing for the collection of photos of both clean and pollutant-emitting panels. In order to provide a precise and sharp image, the PV panel surveillance camera must have a resolution higher than that of standard network cameras [46]. Generally, it is estimated to lie within 2 million to 5 million pixels range. With respect to PV panels, most routine at or trigger controlled data collection from a PV panel monitoring camera. The "timing method" refers to the automatic imaging of PV panels after regular intervals, be it 10 minutes or one hour.

2.3.1 PV Fault Analysis

When the PV panel has an issue such as overheating, a power failure, or other faults, the trigger mode activates and begins photographing the panels. An alternative method consists of using drones equipped with HD cameras and flying them over PV power plants. Drones are capable of capturing high-definition imagery, enabling sophisticated image processing systems to analyze the images for dirt, damage, and misalignment on PV modules [47]. As is the case with visible light cameras, standard network cameras have a resolution ranging from 2 to 48 million pixels, which guarantees clear and detailed photographs. The resolution of infrared thermal imaging cameras generally falls between 160 × 120 and 640 × 512. Showering or grid viewing of PV panels for PV power facilities is done using UAVs with mounted HD cameras. UAVs are programmed to follow predefined or dynamically computed paths, during which panel images are captured, either as blanket key images.

The PV panels can be detected and recognized in the mobile terminal or background system either in real-time or offline after transmitting the captured pictures [48]. Image preprocessing methods are of two main types – global and local [49]. Global image preprocessing examples include applying gamma correction, histogram equalization, or converting the image to greyscale on the entire image. The image may experience enhancement in its visual appeal, some prominent detail may be highlighted, or the image may experience alterations in the brightness, contrast, or color distribution [50]. Local preprocessing includes techniques such as edge detection, sharpening, and smoothing, which use a small neighborhood around a pixel in the input image to define new pixel values in the output image [51]. These methods may enhance detail or edge definition, or remove noise and other small modifications. The main aim of image-based PV panel overlay detection systems is to augment the effectiveness and accuracy of occlusion detection through enhanced image preprocessing. Different types of occlusion, like shadow or attachment, can be addressed through distinct preprocessing techniques.

For better recognition accuracy, the highlighted areas of the image can be cleared from any interference items as scalp filtering methods together with gamma correction can be utilized as a preparatory operation for shadow restrictions. Where attachment obstructions such as leaves are concerned, one approach can be that of using the surface selective enhancement ascertain that enhances images in the Hue Saturation Value (HSV) color space subsequently referred to as “moss-coated” encased. This will mask the image making it seem as if there is obscuring and neutralization occurring while at the same time increasing the visibility of the contrast in the blocked and unblocked areas [52]. Image feature extraction falls into two main categories: global features and local features [53]. Characteristics such as histograms, Fourier's etc. are classified as Global features since they represent the totality of the image [54]. While these techniques are capable of fetching information from images based on the frequency domain or aggregate statistics, they disregard the spatial structure of the image and the fine details.

SIFT, SURF, HOG, and other techniques are examples of local features that define specific aspects or regions of an image [55]. Even though these methods are costly from a computing and storage perspective, they are incredibly robust and provides great resilience and invariance, which permits them to extract local structure and detailed information from pictures [56]. Moreover, feature extraction is an important step in image processing that allows the recognition of broken and healthy PV panels as well as their specific location and type in overlay PV panel detection systems [57]. Various methods may be employed for feature extraction depending on the type of overlay, either shadow or attachment. Threshold segmentation can be used to illustrate the concept of shadow coverage where one distinguishes shadow from non-shadow areas after employing a greyscale histogram as a global feature that represents the image's brightness distribution. Attachment coverings like leaves can be regarded as non-solitary objects to be removed from a given image. So, as an overarching feature representing the image, we can use the HSV colour space to denote the colour distribution and apply a region expanding algorithm based on colour similarity.

2.3.2 Machine Learning Detection of PV Faults

Techniques such as SVM and K-Nearest Neighbour (KNN) can be used to determine the kind of faults present on photovoltaic panels. Meanwhile, edge detection, shape features and other local features can be used to describe defects like cracks, hot spots and broken grids [58]. According to literature, the main role of dealing with image recognition and positioning range in overlay detection technology for PV panels is to identify the position, magnitude, and type of the overlay [59]. Overlays which may either be shading or attached can be identified and positioned differently whether they are attached or not. For example, traditional algorithm based approaches such as threshold segmentation can be used to classify shadowed areas and non-shadowed areas in the PV panel. Furthermore, within every shadow area, the size and position of overshadows can be identified through connected domain analysis and contour extraction, among others. Features like HSV color space can be analyzed using deep learning algorithms like YOLO or SSD to attach leaf-like covers.

Then, each attachment cover can be categorized using bounding boxes or masking within the attachment region based on its position, area, and type [60]. The image processing based PV panel dust detection technology is contactless, highly efficient, extremely accurate, and enables visual and quantitative analysis, among other things [61]. The phrase “non-contact” highlights the image processing technique where the recognition or identification is done remotely by taking pictures of the object under study using a camera or other non-contact means [62]. This increases safety and ease while reducing the risk of damage or disturbance to the objects. Images can be processed rapidly and efficiently by using an array of methods and techniques including compression, augmentation, restoration, segmentation, description, recognition, and high speed computing. Sophisticated sensors capable of detecting minute dust changes can provide high precision. However, other tools like filters, edge detectors, feature extractors, and others can improve recognition accuracy by extracting pertinent information and features from images. Making use of a computer’s display capabilities also allows for the production and presentation of appealing outcomes, which constitutes processing and analysis. These outcomes can be generated in an easy-to-understand format that enhances comprehension.

2.3.3 Deep Learning Detection of PV Faults

Using a deep learning framework, Ayyagari et al. [63] developed an unique image-based approach for detecting and classifying dust and dirt on PV arrays. As part of the process, industrial meteorological data and high-resolution images of PV arrays were obtained. In addition, the authors collected meteorological data for a year along with data on the combined power plant's productivity and images of the PV arrays with different dust levels for benchmarking purposes. Their deep learning model was built using a Convolutional Neural Network (CNN) and Long Short-Term Memory (LSTM) network. This model used a two-step process to classify the dirt and dust. First, the picture data and weather data were processed in parallel and the outcomes were merged in a second step. The model achieved the following results with each subsequent method increasing accuracy: 50.91% with image data through CNN, 94.09% with LSTM using meteorological data, and 96.54% through

CNN-LSTM with combined data. The method showed greater accuracy than those employing standalone CNN or LSTM in subsequent experiments. Abuqaaud et al. [64] proposed a new approach based on computer vision for the identification and monitoring of dirt and dust on solar PV panels.

After employing image preprocessing techniques like high-pass filtering and histogram equalization, the authors further refined the image by cropping and transforming it to obtain the hue layer in the HSV color space. Moreover, a linear classifier was applied to the system to determine the cleanliness of the PV panels, and the GLCM method was used to texture the color layer. Czarnecki et al. [65] proposed several techniques for determining if the PV panels are clean or unclean, including the k-nearest neighbor classifier, naive Bayesian classifier, and Fisher linear discriminator. Every algorithm provided explanations regarding the underlying concept, steps, parameters, and flow diagram detailing its framework. Furthermore, they analyzed the advantages and disadvantages of multiple algorithms and conducted extensive testing using traditional binary classification methods to evaluate the performance of the proposed method. In their conclusions, the authors pointed out that the naive Bayesian classifier is preferable due to its detection accuracy, which outperforms other classifiers.

The method stated in [63] includes all meteorological data analysis and image classification to determine cleaning schedulers, optimally avoid physical performance degradation from dust and soil accumulation, and is capable of autonomous adaptability to seasonal and environmental conditions. It adjusts to numerous types of dust and soil conditions, as well as the frequency and methodologies of cleaning. Compared to standard cleaning routines done manually, this method alongside on-site observation determines light transmittance, dust layer thickness, and surface pollution metrics termed as the dust reduction efficiency on power generation. This information can then be used to optimize and plan for cleaning cycles. PV panel images captured by modern day cameras in [64] and drones in [65] can be contorted to a myriad of lighting and perspective positions which improves recognition precision; furthermore, the technique in [64] does not need any additional sensors or devices. Unlike other pollutants, the methods discussed in [63,64] are confined to the detection and surveillance of dirt and dust on solar panels.

As noted in [63], there is an issue with space limitation when mounting the PV array on the ground or flat roof owing to the need to prevent shading and shadowing. They calculated the relationship between dust level and model coefficients by examining the image attributes of clean and dusty photovoltaic panels using an atmospheric scattering model. The authors verified the model through numerous tests conducted indoors and outdoors under different levels of irradiation. It was established that the model's predictions on the amount of dirt that PV panels had was 83.78% accurate. This technology allows for the intelligent cleaning of photovoltaic panels. An accurate and simple method. Zhou and friends were the first to apply drone's visible spectrum photos for automatic partitioning and categorization of contaminated PV panels [68]. The approach combines flexible threshold segmentation of S-component grayscale color space with other methods. The use of PV panel texture along with the pixel parameters made it possible to identify the objects automatically. For automated classification, a number of neural networks were tried after transfer learning. The recall was 76.2% and the accuracy achieved was 86.8%. As it was noted in [66], the use of the GEE platform is an advantage of this technique over others. GEE, a cloud-based platform for processing geospatial data, is an open-source software for monitoring and controlling solar power stations which can quickly analyze multi-temporal satellite images and display them graphically.

2.3.4 Image Processing in PV Faults

Certainly, it will overfit or underfit appropriately based on the parameters and thresholds selected for the algorithm. Using the technology detailed in [68], the location of the photovoltaic panels of the floating solar power station and their pollution status can be monitored, facilitating intuitive maintenance planning. For example, the approach in [66] could be further improved by incorporating time-phased, high-resolution remote sensing data, and ground observations for error correction and validation, developing a practical dust index, and considering additional factors such as the temperature, humidity, and rain's impact on solar cells, as well as other influences on board performance. The approach in [67] is based on having a flat surface for the dust layer, but the size and distribution homogeneity of the dust particles determine the degree of attenuation and scattering of light.

Future developments considering addition of information related to the size and distribution uniformity of dust particles can include particle size analyzers or imaging technologies. VGG16 is the classifier described in reference [68]. Compared to other models in the test set, VGG16 outperformed in terms of accuracy and recall; however, it is more complex and resource demanding model with a greater number of parameters. In case we want to use more advanced classification models, while keeping resource expenditure and time minimal, expanding classification performance using lightweight models such as EfficientNet or ResNeXt would be ideal. Tommaso et al. developed an automated multi-level flaw identification model for panels using computer vision and the Yolov3 network. The model is based on drone aerial photographs. Identification of panel shadows, delamination, and other forms of puddles (caused by dirt or bird droppings) is possible with the use of the model. The author verified the model's robustness and efficacy using two datasets, achieving an average accuracy of over 98% and over 70% respectively. Furthermore, the scientists developed a positive misidentification filter that utilizes the area of the solar panel to eliminate false positive identifications of defects.

From dust and solar radiation, Saquib et al. used ANN models to predict power output from the solar panels [70]. The author described training and optimizing runs of the backpropagation model with a three-layer ANN utilizing forward and back propagation. The author validated the model and ensured the ANN was accurate and effective with the implementation of experimental data. The method from [69] could use more sophisticated post-processing methods to reconcile advanced fault classification techniques with increased precision like semantic segmentation or instance segmentation while increasing the precision of false positive or negative identification to avoid erroneous folds. For detection result accuracy and completion, it could merge information from multiple data sources such as through data fusion using multi-sensor or multi-modal approaches. The author from [70] integrates a three-layer ANN model to predict the output power of photovoltaic systems; however, this model has limitations on astrophysical voltage controlled oscillator systems due to non-linear input loop feedback capture. Future work could implement advanced neural network structures, like CNN or LSTM, to improve overall precision and generability of the predictions. All previously mentioned attempts to use image processing to detect overlays on PV panels center on the same industry challenges. However, reference 2 does not permit the processing of high-resolution images in a timely manner, leading to subpar detection results.

Photographs taken by drones are also included. There may be cloud occlusions or other quality concerns with the satellite data that is available in [66] that requires some sort of preprocessing or filtering according to [67]. In measuring dust, some considerations such as a particle's size, shape, and distribution along with the lighting, angle, and colour of the picture impact the quality of the picture. Clearer images with increased contrast, clarity, and less noise or interference are also added via pre-processing, enhancement, and other adjustments that need to be made in the case of the image in Ref [67]. In order to detect dirty PV panels located in offshore PV power plants, the technique described in Reference [68] employs visible light images. Sighted module pose and field of view along with camera specifications such as housing, module positioning, and terrain of site along with initial pixel and texture data, multiple network model evaluation, and transfer learning as well as advanced computation theory have some influence as well. To limit errors and noise, Ref. [69] has to considers surrounding conditions such as temperature, wind speed, cloud cover, etc. For adequate detection of microcracks and other small faults, high resolution images are also needed according to Ref. [69]. There is a possibility of experiencing delays because, according to Ref. [70], image processing requires significant time and system resources. Overlay detection algorithms based on image processing are impacted by the climate. As stated in Ref. [66], sunlight must be unobstructed in order to carry out the task. Weather such as wind speed, rainfall, and temperature affect the movement, deposition, and erosion of sand and dust, which affects the rest dirt detection process. Frogbird et al [68] also mention that seeking contaminants in floating PV power plants has to be done under clear and sunny weather, particularly around 10 AM when the angle and intensity of the sun are most favorable for visible light imaging. Boxy overcast, wet, icy fog cover, and other similar conditions tend to blur images and reduce sharpness and contrast which are essential in distinguishing the edges of PV modules and contaminants, which have an adverse impact. According to [69], sunny conditions are also needed to search for faults in PV panels. As long as the temperature goes above 15 degrees celsius, alongside a slight wind and moderate solar radiation, the drone will be able to fly steadily, capturing clear infrared and brilliant visible light images. Image distortion, including blurring or increased noise, can occur as a result of snow, rain, or hazy conditions, while snow or rain can lead to distortion or interference. The sunlight's reflection on PV panels needs to be captured for image processing, therefore, the method described in [70] can only be conducted under clear weather conditions with abundant solar radiation.

Table 2.2: Summary of Related Works.

Reference	Methodology	Contributions	Limitations
[44]-[60]	Image processing pipeline with camera/drones, preprocessing, feature extraction, classification	Comprehensive framework for detecting shadows, dust, and attachment-based faults on PV panels	High computational cost, sensitive to lighting/weather, complex pipeline, local features need high resources
[61]	Image-based non-contact PV panel inspection	Highlights advantages like non-contact, visualization, high precision, and safety in detection	Relies on high-quality sensors and controlled environments for accuracy
[63]	CNN-LSTM fusion model using image and meteorological data	Achieves 96.54% accuracy by combining CNN for images and LSTM for time-series data	Limited to dust/soil detection, requires flat installation, cannot generalize across all pollutant types
[64]	GLCM + HSV hue layer with linear classifier	Low-cost and sensor-free solution adaptable to different lighting and angles	Linear classifier struggles with non-linear features, can't handle all types of dirt
[65]	KNN, Naive Bayes, and Fisher classifiers for binary clean/dirty classification	Compares algorithm performance using classic classification metrics	Naive Bayes optimal, but only suitable for binary classification, lacks robustness
[66]	Dust index analysis using satellite data (GEE, DBSI, NDSI, RNDSI)	89.6% accuracy with satellite data; scalable cloud-based remote monitoring platform	Limited resolution, affected by clouds, can't detect fine pollutants or floating dirt

Table 2.2: Summary of Related Works "Table Continued".

[67]	Atmospheric scattering-based image enhancement and dust quantification	Indoor experimental validation with 83.78% accuracy, adaptive to different dust levels	Assumes uniform dust layer, limited to indoor settings, sensitive to parameter tuning
[68]	Drone-based image segmentation and classification using VGG16 and transfer learning	Combines texture and pixel features for floating PV plant maintenance; 86.8% accuracy	VGG16 is resource-heavy, limited pollutant types detected, affected by environmental conditions
[69]	YOLOv3-based drone inspection with false-positive filtering	98% detection accuracy for shadows and bird droppings, supports defect hierarchy	Needs high-resolution images, influenced by weather, environmental noise, wind, and sunlight
[70]	ANN-based dust prediction model using solar radiation and dust data	Predicts power output loss due to dust with a 3-layer ANN, suitable for harsh environments	Simple ANN lacks depth, not scalable to complex climates, needs enhancement with CNN or LSTM

2.4 CONCLUSIONS FROM LITERATURE REVIEW

The literature review shows a wide and growing range of contributions made toward improving photovoltaic (PV) system diagnostics, particularly concerning the automated detection and classification of surface-level anomalies, including dust, soiling, shadows, and structural faults. Numerous studies have attempted the automated monitoring of PV arrays using image processing (with drones and high-resolution cameras) coupled with sophisticated image segmentation techniques, resulting in non-contact and efficient monitoring. These methods rely on the extraction of global and local image features, including grayscale histogram and HSV color space, texture descriptors, and edge detection filters. While such methods offer marked improvements in practicality—scalability and visualization for example—they tend to be hindered by dependence on varying weather conditions, image resolution, and processing power. Moreover, the intricate and highly tailored nature of the preprocessing pipeline makes it extremely difficult to standardize for varying types of faults across multiple systems. In addition, advances in automation using artificial intelligence (AI) and machine learning (ML) models—especially convolutional neural networks (CNNs)—recurrent architectures such as LSTMs, and hybrid classifiers for the classification of PV anomalies have proven to yield superior results. There is greater reliability with these models concerning fault detection because they are better able to withstand variations in the environment as compared to traditional methods. An example of hybrid models include CNN-LSTM combinations as well as image enhancement methods based on atmospheric models. These have outperformed standalone classifiers by using multimodal data fusion, such as images with metrological data. However, these widespread ML-based approaches have glaring limitations such as overfitting, non-transparent decision-making processes (often referred to as black-box models), large labeled datasets for training, and rigid adaptability to a defined region and climate context. Concurrently, studies that integrate fuzzy logic with classical classifiers like Naive Bayes, SVM, KNN, underscore the necessity of explanatory power that rule-based reasoning brings to PV diagnostics. These models are more computationally efficient and analytically transparent but fall short of higher-performing AI-based models in diagnostic accuracy due to inadequate response to complex overlapping or nonlinear fault patterns. Other advanced frameworks derived from satellite imagery and geospatial cloud platforms, including Google Earth Engine, support extensive monitoring of the environment, but lack the necessary detail and precision for thorough inspections of PV panels. Overall, the literature indicates that no single methodology sufficiently meets the vast and ever-evolving needs for diagnostics of modern PV systems. The construction of diagnostic models which preserve high classification accuracy across varying conditions, maintain a human-centric interpretative framework, and scale with minimal computational expense presents a profound gap in existing research. The model weaknesses in computational cost and interpretive explainability suggest an increasing demand for hybrid models that incorporate the deep learning's adaptive capabilities with fuzzy logic’s uncertainty reasoning and rule-based systems. Such integration may provide the intelligent, robust, and explainable diagnostic solutions that are needed for next generation PV system monitoring. This motivates the proposed thesis methodology which seeks to fill these gaps through the creation of a Deep-Based Fuzzy Logic Algorithm (DBFLA) tailored for enhanced machine learning-driven PV fault detection and classification.

3. MATERIALS AND METHODS

The purpose of this chapter is to outline the photovoltaics basics that this dissertation will discuss including; operation of a photovoltaic cell and module, available mathematical models of a photovoltaic cell, and other key parameters of the cell model that are important for diagnosis.

3.1 PHOTOVOLTAIC CELLS AND MODULES

Solar electrical energy is generated by the photovoltaic effect and the generating unit is the photovoltaic cell, which is formed by a pn junction typically made of silicon. With the incidence of radiation on the surface of the cell, free electrons appear and move due to a voltage also resulting from the photovoltaic effect in the pn junction. This voltage generates the movement of electrons to the part of the n-type junction and the appearance of holes in the p-type junction [4].

A photovoltaic module is formed by the association of several photovoltaic cells, usually connected in series, and may have some rows in parallel depending on its configuration and intended use. The photovoltaic module is usually made up of rows of cells, with a bypass diode in anti-parallel, as shown in Figure 3.1. This device, because it is connected in reverse, only comes into operation in cases of defective or shaded cells, which start to consume power instead of producing it.

therefore, with the inversion of the voltage of the defective cells, the diode becomes directly polarized and the current flows through it, as shown in the example in Figure 3, where there is shading in cell 12. The diverted current does not contribute to production, but prevents the affected cells from behaving as a load for the others, that is, they serve as protection and improve efficiency in the event of failures [5].

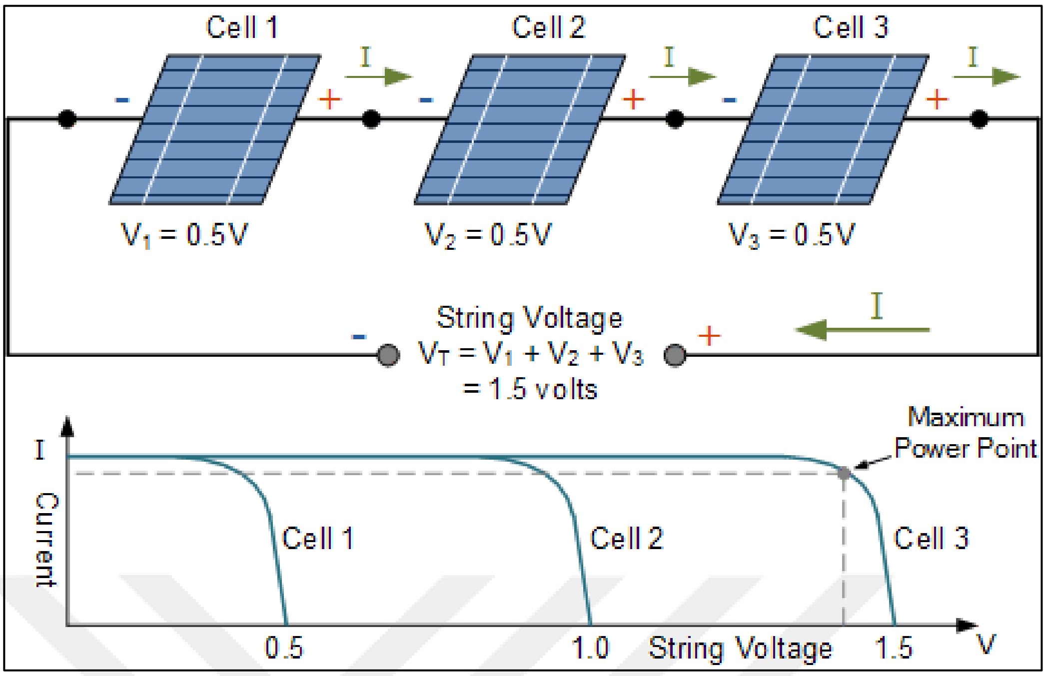

Figure 3.1: Example of a 60-cell, 3-row Photovoltaic Module with Protection Diode [5].

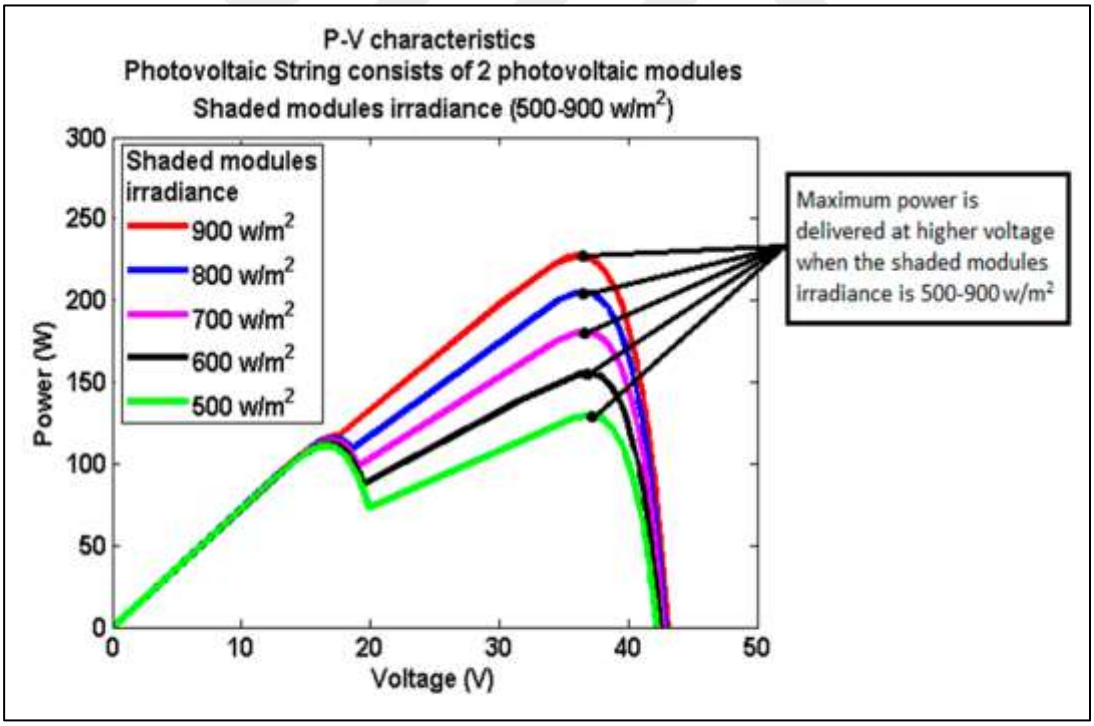

Figure 3.2: Example of Current Behavior in Case of Partial Shading.

If this diode did not exist, the affected cell would produce heat due to the high resistance generated by the shadow, damaging the cell and/or causing a hot spot. Furthermore, the entire production would be limited. Bypass diodes are very important for the study of failures, as they allow that in the event of an anomaly only the row where the affected cell is located remains out of operation, which minimizes the reduction in the form factor of the IV characteristic curve due to the failure [5]. There are also other possible configurations for the bypass diode with interlacing or overlapping of the diodes, which influences the value of the production losses as shown in [6], however the objective of the device is the same in both configurations and in this work the modules used have diodes in the usual configuration without interlacing. Before using a photovoltaic module, it is important to consult its specifications sheet, or datasheet. This document contains the main electrical parameters of the equipment and information about the cells. Table 1 shows the information usually found in a photovoltaic module specifications sheet.

Table 3.1: Photovoltaic Module Characteristic Sheet Information.

SCC - Short-circuit current	MPPT - Maximum power point
CMP - Current at the maximum power point of the module	Ki-MPPT - Coefficient of variation of current with temperature
Voc-OCV - Open circuit voltage	Kv-MPPT - Coefficient of variation of voltage with temperature
Vmp-MPPT - Voltage at the maximum power point of the module	nc-MPPT - Number of cells in the module

It is important to emphasize that these voltage, current and power data refer to Standard Tests Conditions (STC), which are the standard circumstances for a module to be tested by the manufacturer.

3.2 MATHEMATICAL MODELING FOR DIAGNOSIS

Ideally speaking, the representation of a photovoltaic cell can be made with a current source with a diode connected in antiparallel, as shown in Figure 4. However, to represent the losses, two resistances are added to the model, one in series and the other in parallel as shown in Figure 3.4 and Figure 3.5 [3] respectively.

Several models are found in the literature to try to best represent the behavior of photovoltaic cells according to the needs of each case. Some models are more complex and better suited to certain applications, while other, simpler models are often sufficient for the objective to be achieved. The main models are those of one and two diodes (one or two exponentials, respectively), as presented later in sections 2.2.1 to 2.2.4. Within the single diode model, there may also be two possibilities, with 4 or 5 parameters.

Both models, one and two diodes, shown in Figure 3.4 and Figure 6 use as a basis the Shockley equation [7], which describes the behavior of the current as a function of the voltage. Mathematical modeling can be done in a more complete (and more complex) way with more accurate results, or in a simpler way if it is sufficient for the object under study. Since the focus of this work is fault diagnosis, the objective was to find the simplest model that would meet the needs of representing a new photovoltaic cell for comparison, thus making it possible to diagnose the health status of the photovoltaic module.

3.2.1 Model with a Simple Diode

The simple diode model, or 3-parameter simple exponential model, can be represented by the equivalent circuit as shown in Figure 3.3. It is the simplest model that describes the behavior of an ideal photovoltaic cell.

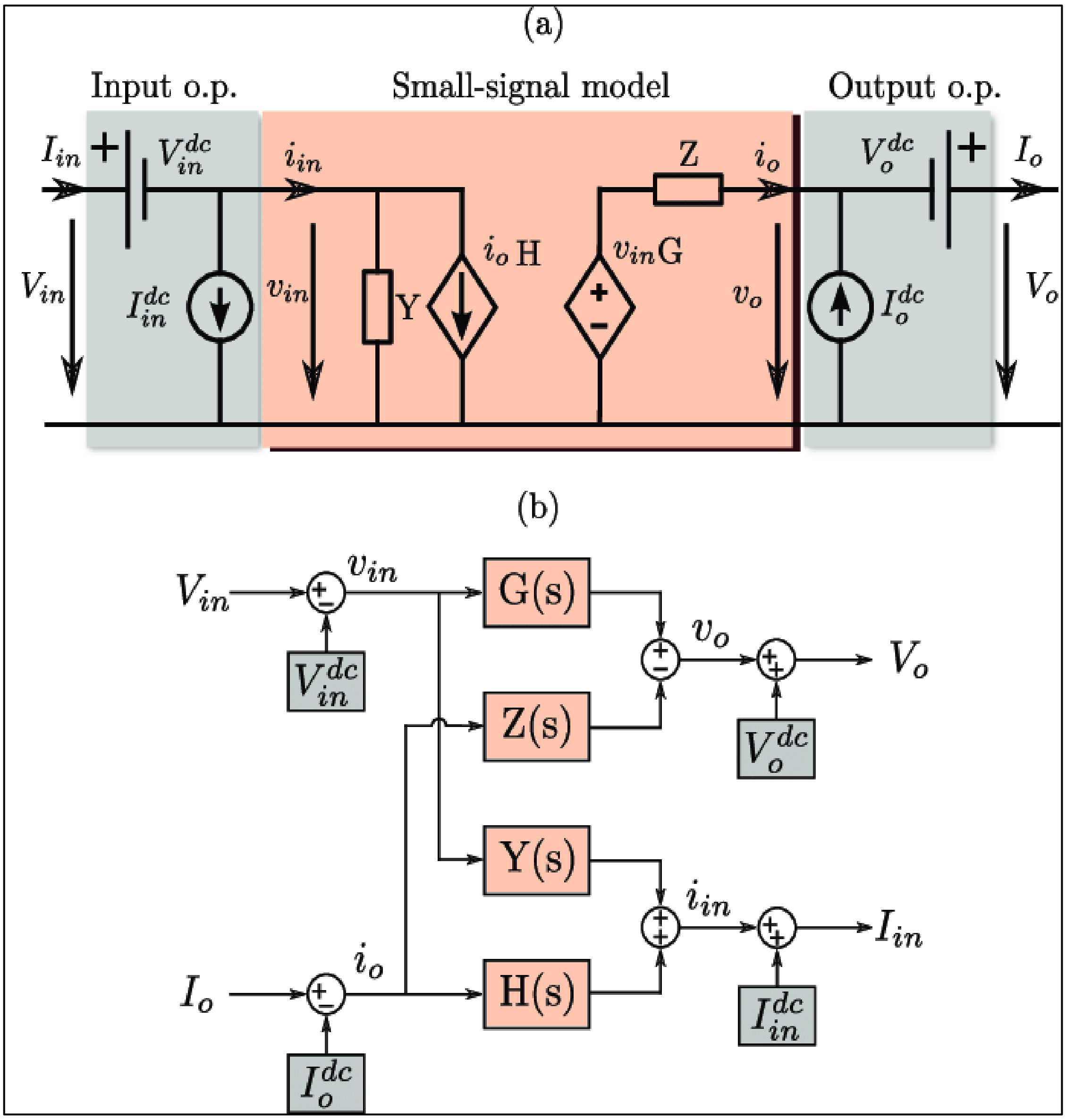

Figure 3.3: Parameter Model Equivalent Circuit.

3.2.2 Model with a Simple Diode with Series Resistance

The simple diode model with series resistance, or simple exponential with 4 parameters, can be represented by the equivalent circuit as shown in Figure 3.4. It is a low complexity model and consists of an ideal photovoltaic cell with a series resistance that represents the losses due to the internal resistance of the cells, the interconnections and contacts between cells and modules.

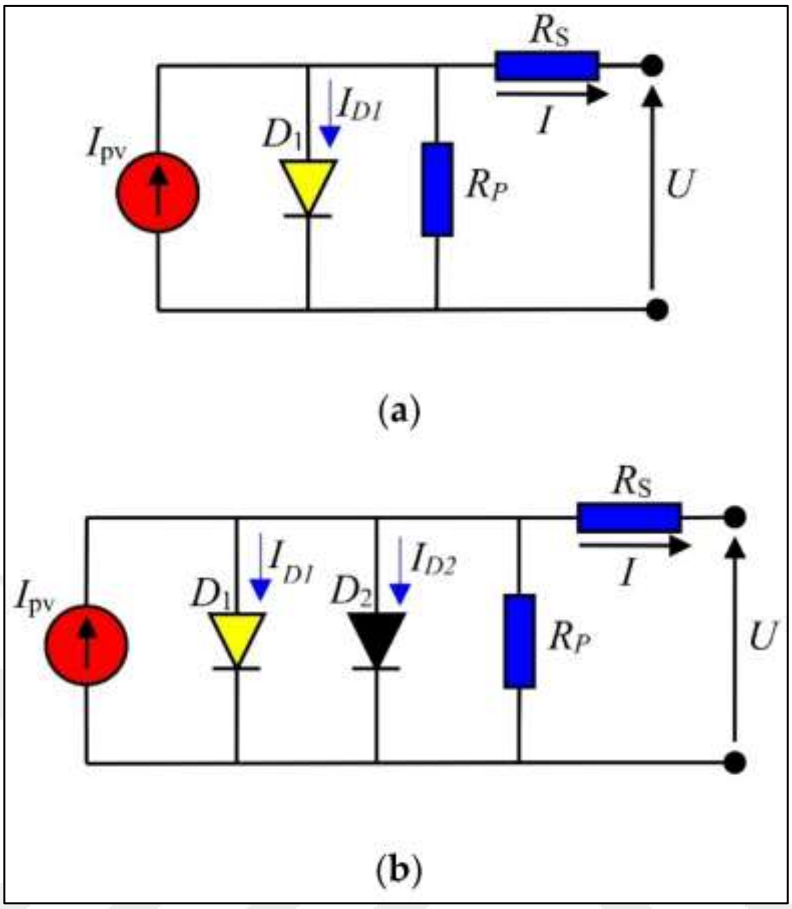

Figure 3.4: Equivalent Circuit Model with 4 Parameters [3].

$$I = I_{ph} - I_o\left[e^{\frac{(IR_s+V)}{V_t}} - 1\right] \tag{3.1}$$

The term Vt is the thermal voltage of the cell, and is defined by equation (3.2).

$$V_t = \frac{kTn_cA}{q} \tag{3.2}$$

Where:

a. I0 - Diode reverse saturation current (A);

b. Iph - Photoelectric current (A);

c. A – Diode ideality factor;

d. T - Cell operating temperature (K);

e. q - Electron charge (1.6 × 10−16C)

f. Rs - Series resistance of the mathematical model (Ω);

g. k - Boltzmann constant ($1.38 \times 10{-}23 J/K$);

h. nc-MPPT - Number of cells, for one cell = 1;

This model has the following parameters to be estimated: I0, Iph, Rs and A.

For the purpose of simplification, in this work, the diode ideality factor (A) and the junction band energy (Egap), which is also used in some references [8], will not be parameters to be estimated, and their value will be defined depending on the cell technology as presented in [8] and [9].

3.2.3 Model with a Simple Diode and Parallel Resistance

The simple diode model with parallel resistance, or simple exponential with 5 parameters, can be represented by the equivalent circuit as shown in Figure 3.5 It is a slightly more complete model than the 4-parameter model because in addition to the losses of the series resistance, the losses due to leakage current from possible imperfections and/or impurities in the crystals are also considered. This additional loss is represented by the shunt resistance, which is connected in parallel with the diode in the circuit.

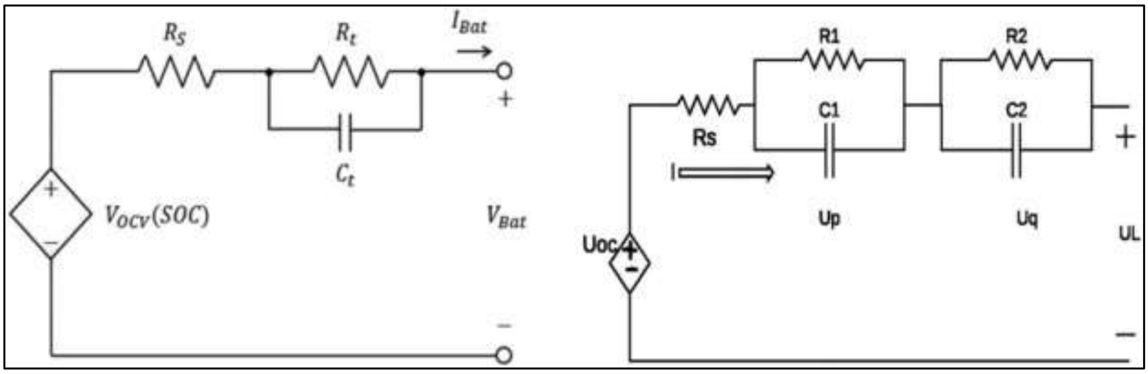

Figure 3.5: Equivalent Circuit Model with 5 Parameters [3].

3.2.4 Two-diode Model

The double exponential model, or 6-parameter model, can be represented by the equivalent circuit shown in Figure 3.6. It is the model that most closely matches reality, since losses due to carrier recombination at the pn junction and on the module surface are also considered.

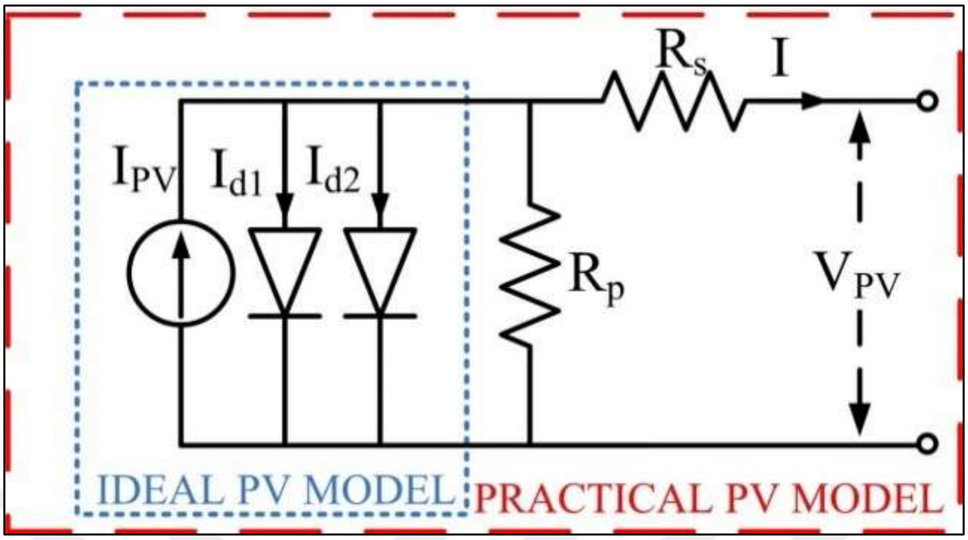

Figure 3.6: Equivalent Circuit Model with Two Diodes [3].

The model has the following parameters to be estimated: Io1, Io2, Iph, Rs, Rshunt and A.

3.3 ESTIMATION OF MODEL ADJUSTMENT PARAMETERS

In the literature, there are a wide variety of methods for extracting parameters from the models presented above [11] [12], most of which use 4 or 5 parameters. Parameter extraction can be done with any of the models presented. Regardless of which one is used, they all depend on adjustment parameters that cannot be measured directly, but which are important for creating a model that is closer to the reality of a photovoltaic cell.

3.3.1 Estimation by the 4-Parameter Model

For diagnostic purposes, the literature has shown that the model with 3 or 4 parameters is sufficiently good and is easier to calculate. To achieve this, some simplifications in the current equation should be considered, starting from the 5-parameter model. The first simplification considers the value of Rshunt to be high, approaching infinity, thus obtaining equation 3.3 [11].

Another simplification adopted is the Iph approximation≅SCC, due to the small difference between the two currents, especially in crystalline silicon cells, where the loss caused by the series resistance at the short circuit point is negligible [13], resulting in equation 3.4.

$$I = I_{sc} - I_o \left[e^{\frac{(IRs+V)}{V_t}} - 1 \right] \tag{3.3}$$

It should also be considered that the inverse saturation current of the diode (I0) is very small compared to the exponential term of the equation, so the term -1 is disregarded [16]. Thus, the simplified equation for extracting the parameters results

Vt, Rs and I0.

$$I = I_{sc} - I_o \left[e^{\frac{(IRs+V)}{V_t}} \right] \tag{3.4}$$

With the simplifications, only 3 unknown parameters remain to be estimated in the modeling, applying the provided points of the characteristics in the STCe conditions, calculating the derivative of the power with respect to the voltage at the maximum point Pmpp and of the current with respect to the voltage is obtained.

3.3.2 Estimation by the 5-Parameter Model

In this method, in addition to the parameters determined previously, Rshunt is also calculated. The initial simplifications are the same as in the method of the previous section, where Iph≅SCC and the term -1 are negligible, both for the reasons presented above. The term referring to parallel resistance is initially disregarded due to the low current that flows through it, but after adjusting the parameters it can be calculated [18].

This leads to Equation (3.5).

$$I_{sc} - I_o\, e^{\frac{V_{oc}}{V_t}} = 0 \tag{3.5}$$

3.4 INTERFERENCE OF ENVIRONMENTAL CONDITIONS ON THE PARAMETERS

The environmental conditions where the module is operating influence its IV curve and consequently also the model's adjustment parameters, since these are defined by the curve behavior. Therefore, in order to implement a photovoltaic cell model to simulate a curve tracer, it is necessary to include this influence of irradiation and temperature (G and T) in the adjustment parameters [22].

It is important to highlight that, to calculate the parameters Vt, Rs and I0, according to the previous chapter, using the data from the module's characteristics sheet, it is necessary to make the appropriate transposition to the real radiation and operating temperature conditions, since they are different from the STC conditions.

The radiation that reaches the surface of a cell is directly related to the amount of current generated; this photoelectric current (Iph) is also influenced by temperature, as described in equation (3.6) [23] [24].

$$I_{ph} = \frac{G}{G_{STC}}\left[I_{sc\,STC} + K_i(T - T_{STC})\right] \tag{3.6}$$

Where:

a. G – Irradiation;

b. GSTC – Irradiation under STC conditions (1000 W/m²);

c. T – Temperature;

d. TSTC – Temperature in STC conditions (25º C);

e. Ki-MPPT – Coefficient of variation of current with temperature (A/K).

To determine I0, [25] presents its equation that expresses its dependence on temperature and the coefficients Ki-MPPT and Kv-MPPT as shown in Equation (3.7).

$$I_o = \frac{I_{sc\,STC} + K_i(T - T_{STC})}{e^{\left[\frac{V_{oc\,STC} + K_v(T - T_{STC})}{V_t}\right]} - 1} \tag{3.7}$$

Where: Kv-MPPT - Coefficient of variation of voltage with temperature (V/K).

Rshunt is the parameter that changes the slope of the IV curve near the short-circuit point. The higher the value of Rshunt, the higher the slope and the lower the maximum power available, Pmpp. According to [26], the resistance parallel to the diode is inversely proportional to the short-circuit current and, consequently, with irradiation, the relationship presented in equation (3.8) shows this dependence found empirically [23].

$$R_{shunt} = \frac{G_{STC}}{G} R_{shunt\,STC} \tag{3.8}$$

On the other hand, the series resistance is more influenced by the cell temperature, since its value changes the slope of the IV curve at the open circuit point. Rs is also inversely proportional to the irradiation as shown in [27], the equation relating to this influence is presented in (3.9).

$$R_s = \frac{T}{T_{STC}} R_{s\,STC} \left(1 - \beta \ln\left(\frac{G}{G_{STC}}\right)\right) \tag{3.9}$$

Where β is a constant [28].

To calculate thermal stress, its equation is more trivial, as it is a factor directly proportional to temperature, as shown in equation (3.10):

$$V_t = V_{t\,STC} \frac{T}{T_{STC}} \tag{3.10}$$

3.5 COMPUTATIONAL IMPLEMENTATION

As seen previously, the environmental conditions in which the module to be tested is located directly influence its parameters. Therefore, a computational implementation in MATLAB software was used, starting from the 5-parameter cell model, seen in section 3.2, that is, starting from the parameters (SCC). Subsequently, the output voltage and current values are adjusted according to the equations in section 3.4, which allowed a simulation with temperature and irradiation variations. The simulation is shown in Figure 3.7 below:

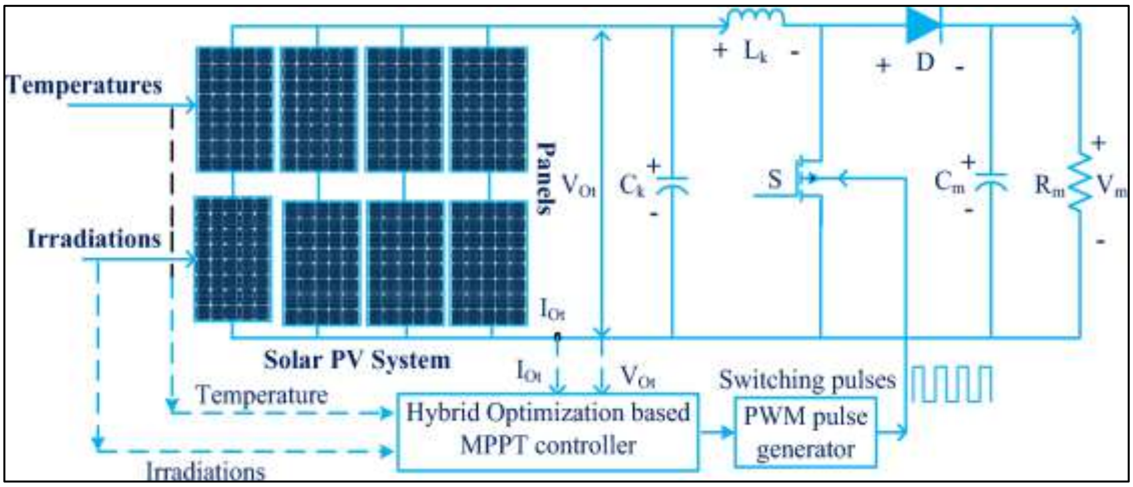

Figure 3.7: Computational Implementation**.**

The simulation's main purpose was to validate the equations presented in the models and the influence of temperature and irradiation variations. It also aims to simulate a curve tracer to measure the electrical parameters of the output of one or more modules in different environmental conditions.

For the simulation, some simplifications are considered: uniform irradiation and temperature values throughout the panel, all cells are considered identical, Rshunt of infinite value and zero Rs for the simulation of a cell without degradation.

3.6 MAIN FAULTS IN PHOTOVOLTAIC MODULES

In general, failures in photovoltaic systems can be classified into two types: irreversible failures, caused by electrical or mechanical problems, and temporary power losses. Both can lead to a decrease in the output power, efficiency and reliability of the photovoltaic system [29] [30].

Irreversible failures, or physical failures, are directly linked to the health status of the photovoltaic module, and the problem can only be solved by replacing the defective equipment.

Temporary energy losses are not characterized as intrinsic equipment failures, but rather an external factor that can reduce energy production. This external factor is related to the environment, installation and infrastructure of the location where the PV system is located, and the problem can often be solved without replacing the affected module. It is also important to analyze temporary failures, because in the long term, they can accelerate the aging of the module, for example, with the appearance of hot spots due to shading, which begin with an external factor that can permanently damage a cell [30].

Failures in photovoltaic modules can be classified as shown in Table 3 [22]:

Table 3.2: Classification of Failures and Their Effects on the Main Parameters.

Fault type	*Possible consequences of failures*	
Temporary failures	Shading Dirt/snow deposit	Voc-OCV Reduction *Hot spots*
Permanent failures	Damaged encapsulation/loss of transparency	Reduction of SCC (due to partial blocking of radiation)
	Module corrosion	Increase in Rs Decrease in Rshunt
	Broken conductors or cold solder	Increase in Rs.
	Cracks and burns in cells	Increase in Rs. *Hot spots*

Hot spots can be considered as a module failure, but they are the result of a previous failure and can be either temporary or permanent, whether it is a small shadow that generates this hot spot or an internal failure in the cells, such as small cracks. Therefore, for the purposes of classifying failures in this work, this phenomenon will be treated as a secondary consequence of some module failure.

3.7 FAULT DIAGNOSIS METHODS

Fault diagnosis is an essential step for the reliability, safety and efficiency of the photovoltaic system. During years of operation, exposed to environmental conditions, system failures and malfunctions may occur, and constant failures may reduce the useful life of the modules and accelerate the degradation and aging process of the equipment [29].

The existing methods for diagnosing photovoltaic modules range from the simplest, such as visual inspection, use of a thermographic camera, monitoring the Rs value, to the most complex ones with constant supervision of the main system parameters.

3.7.1 Diagnosis by Monitoring Equivalent Series Resistance

Calculating the Rs parameter is of great importance for diagnosing and checking the health status of a module. Increases in its value may indicate problems such as poor contact at its terminals and connections, cell malfunction or cell breakage. Therefore, its monitoring can provide important information, but it is important to emphasize that Rs is only an adjustment parameter of the mathematical model and does not represent the real value of series resistance, and may even assume negative values at low irradiation values [11] [13].

Unlike Rs, which is a modeling parameter, the equivalent series resistance Rse represents the effective resistance of the photovoltaic module. The methods for extracting this resistance aim to estimate the value of the physical resistance and not only for the application of the equivalent circuit model. Changes in the value of Rse, as well as in Rs, may indicate cell malfunction.

3.7.1.1 Based on the slope of the IV curve

One of the methods for monitoring the equivalent series resistance is based on the slope of the IV curve near the open-circuit voltage point [11]. This method consists of calculating the Rse by the derivative dV/dI. Some references use the term “open-circuit resistance” to define the slope of the curve at the Voc-OCV point [31] [32].

Based on the voltage versus current equation (3.10), the equivalent resistance can be obtained by calculating its derivative at the open circuit voltage point as shown in equation (3.11).

$$V = V_t \ln\left(\frac{I_{sc} - I}{I_o}\right) - IR_s \tag{3.10}$$

$$R_{se} = -\left.\frac{dV}{dI}\right|_{V_{oc}} \tag{3.11}$$

In practice, the value of dV is equal to 1V, due to the large number of points obtained in a data sweep on the IV curve [22], the more points, the smaller the noise and the better the results. Equation (3.12) gives the equation to be used to calculate Rse using two points on the curve.

$$R_{se} = -\frac{V_2 - V_1}{I_2 - I_1} \tag{3.12}$$

In equations (3.11) and (3.12) V must be used≅You therefore aim to calculate the derivative at the point where there is the highest direct voltage on the diode, and this occurs in an open circuit, so it is convenient to determine the series resistance at this point.

3.7.1.2 Analytical calculation

It is also possible to estimate the value of Rse starting from the parameters Rs and Vt of the 4-parameter model as presented in equations (3.13) and (3.14). With these parameters defined, equation (3.15) can be derived as indicated in (3.6), thus arriving at the analytical equation (3.7).

$$R_{se} = \frac{V_t}{I_{sc}} + R_s \tag{3.13}$$

3.7.1.3 Standardization for STC

The slope of a characteristic curve may vary with changes in irradiance and temperature, resulting in errors in the estimation of the series resistance. To correct this problem, in [22] a correction of the Rse value is proposed by means of a normalization of the curve for STC conditions considering the proportionality of Vt ,

SCC, G and T.

$$R_{se\ STC} = \frac{V_{t\ STC}}{I_{sc\ STC}} + R_s \qquad (3.14)$$

$$V_{t\ STC} = V_t.\frac{T_{STC}}{T} \qquad (3.15)$$

$$I_{sc\ STC} = I_{sc}.\frac{G_{STC}}{G} \qquad (3.16)$$

Substituting Equations (3.17) yields the series resistance value normalized to the STC values in (3.18).

$$R_{se\ STC} = R_{se} - \frac{V_{t\ STC}}{I_{sc\ STC}}\left(\frac{T}{T_{STC}}\frac{G_{STC}}{G} - 1\right) \qquad (3.17)$$

3.7.2 Diagnosis through Estimation of the Increment Rs

Another method reported in the literature consists of an estimation of the Rs increment value [33]. This increment is calculated by the difference between the Vmpp and Impp values of the module data taken from the datasheet and the value calculated by means of Equations (3.18) to (3.19). This method estimates the ideal voltage using the Lambert W function [21].

$$V_{id} = -V_t\, W\left[\frac{I_o R_{shunt}}{V} e^{\left(\frac{R_{shunt}}{V_t}(I_{ph}+I_o-I_{mpp})\right)}\right]$$

t

$$-I_{mpp}(R_s + R_{shunt}) + R_{shunt}(I_{ph} + I_o) \qquad (3.18)$$

Where Vid is the ideal voltage estimated for the curve without degradation, and ΔVmpp is the voltage variation due to degradation, as illustrated in Figure 3.8

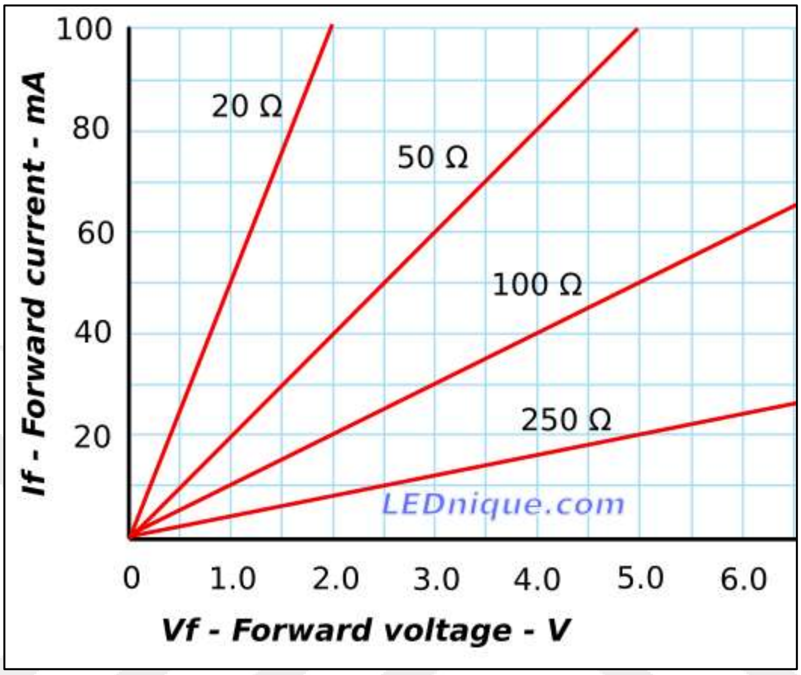

Figure 3.8: IV Curve Graph Showing the Increase in Tension [33].

When the real curve is obtained without degradation, a way to simplify the estimation of Vid is to consider that the current Impp of the curve with and without degradation will be the same, as illustrated in Figure 3.8. Therefore, with the ideal curve (in blue), it is possible to estimate the voltage corresponding to that current Impp. When acquiring an IV curve under real conditions, the temperature and irradiation are known. With these values and the parameters, the blue curve can be obtained by the simulation described in 3.5.

3.8 FAULT DIAGNOSIS AND DEGRADATION USING IV CURVE

3.8.1 IV Curve

A photovoltaic module consists of an arrangement of photovoltaic cells, which when exposed to sunlight generate a current. The voltage and current of a module depend on the number of cells and how they are connected to each other. The number of cells in series defines the voltage value and the number of rows of cells in parallel defines the current value.

The current-voltage (IV) curve of a photovoltaic cell is obtained by superimposing the characteristic curve of the solar cell diode in the dark with the current generated with the illuminated cell [34] [35].

In the IV and PV curves, the main points of voltage, current and power can be identified, as exemplified in Figure 3.9.

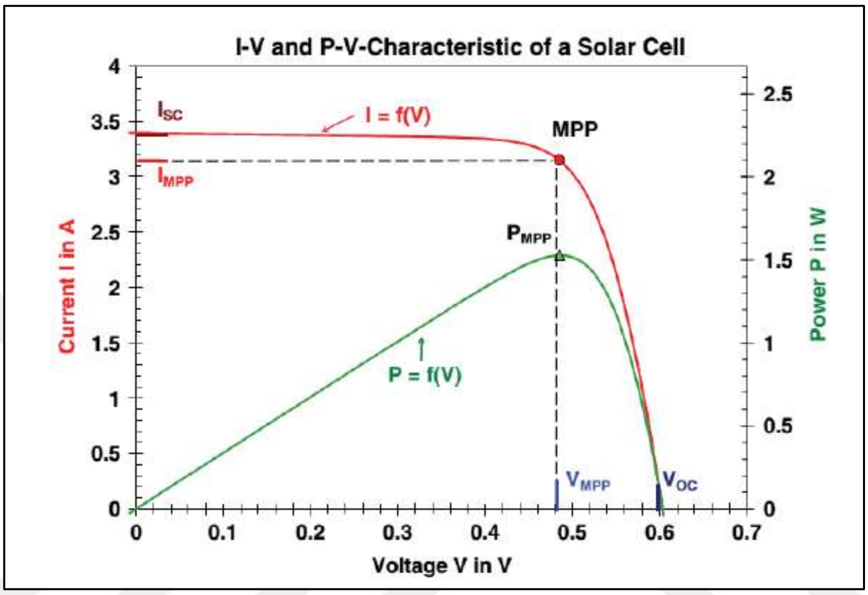

Figure 3.9: Characteristic Curves of a Module and Its Main Points [36].

These main points of the IV curve correspond to the values presented in the technical characteristics (datasheet) of the module, under STC conditions.

However, these generated voltage and current values depend on the conditions in which they are inserted. The operating temperature of the cells has a direct influence on the output voltage of the photovoltaic module, Voc-OCV, and the generated current, SCC, depends directly on the amount of radiation incident on the cell surface, as shown in Figure 3.10.

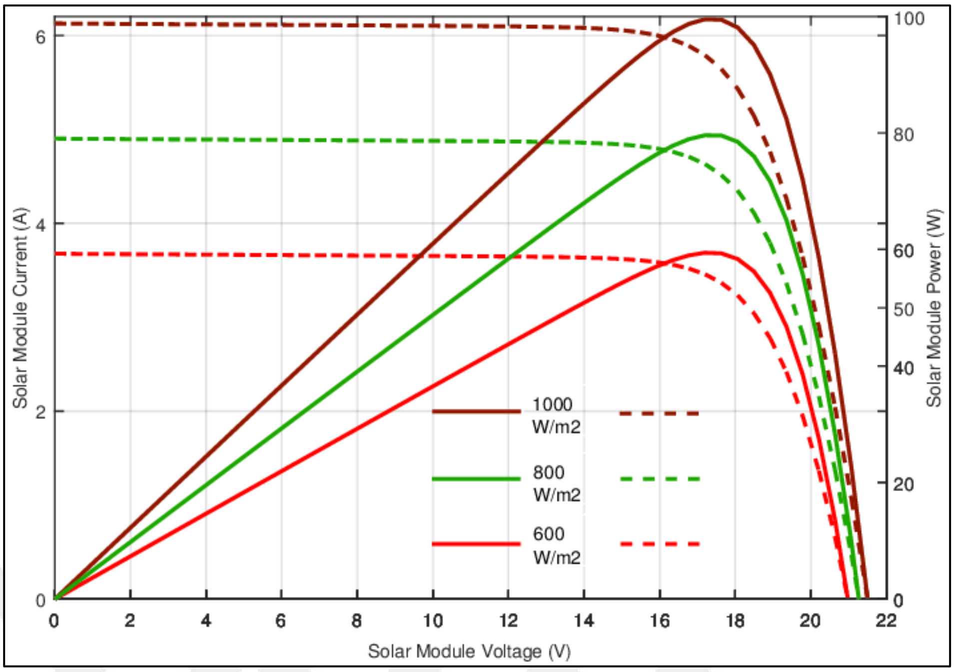

Figure 3.10: IV Curves Behavior Varying Irradiation and Temperature.

In general, along with a failure, a decrease in the resulting form factor (FF) is observed, which is an indicator of the cell quality. This factor is obtained with the voltage and current values at the maximum power point and the critical points of the SCC and Voc-OCV curve. This index shows the proportion of the maximum power in relation to the product of SCC and Voc-OCV [34] [35].

$$FF = \frac{P_{mpp}}{I_{sc}V_{oc}} = \frac{I_{mpp}V_{mpp}}{I_{sc}V_{oc}} \tag{3.19}$$

In module arrays, the corresponding IV curve depends on the way the modules are connected. The number of modules in series in the rows directly impacts the open circuit voltage of the curve and the number of rows in parallel defines the short circuit current [35] as exemplified in Figure 3.11 [37].

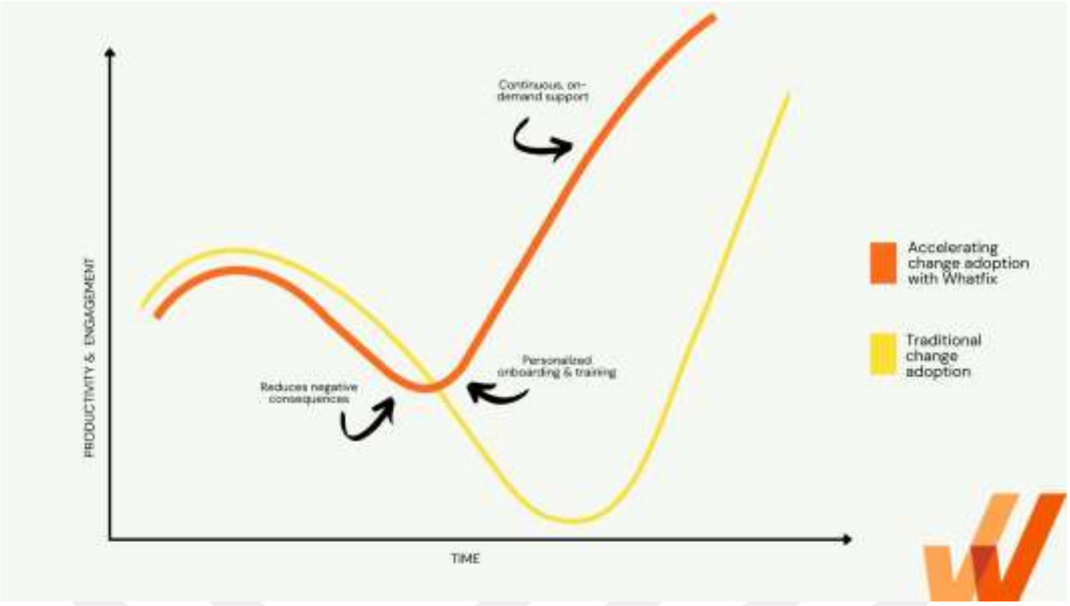

Figure 3.11: Illustration of Changes in the Curve due to Module Associations [37].

3.8.2 Deformations in Curve IV

An IV curve obtained by a curve tracer that presents some deviation in its shape may be caused by: problems in the module being tested, or it may be the result of some parameter external to the module (incorrect model in the equipment, incorrect settings in the measurement or connections). After checking all the external parameters, it is possible to conclude whether that deviation is due to some defect or degradation of the module.

Each failure affects the curve in a different way, leaving distinct marks during operation, as shown in Figure 3.12. The deformations that can be observed in curve IV can be defined in five main categories, and the same curve may present more than one of these deviations. All of them indicate a reduction in the maximum energy produced by that generator [38].

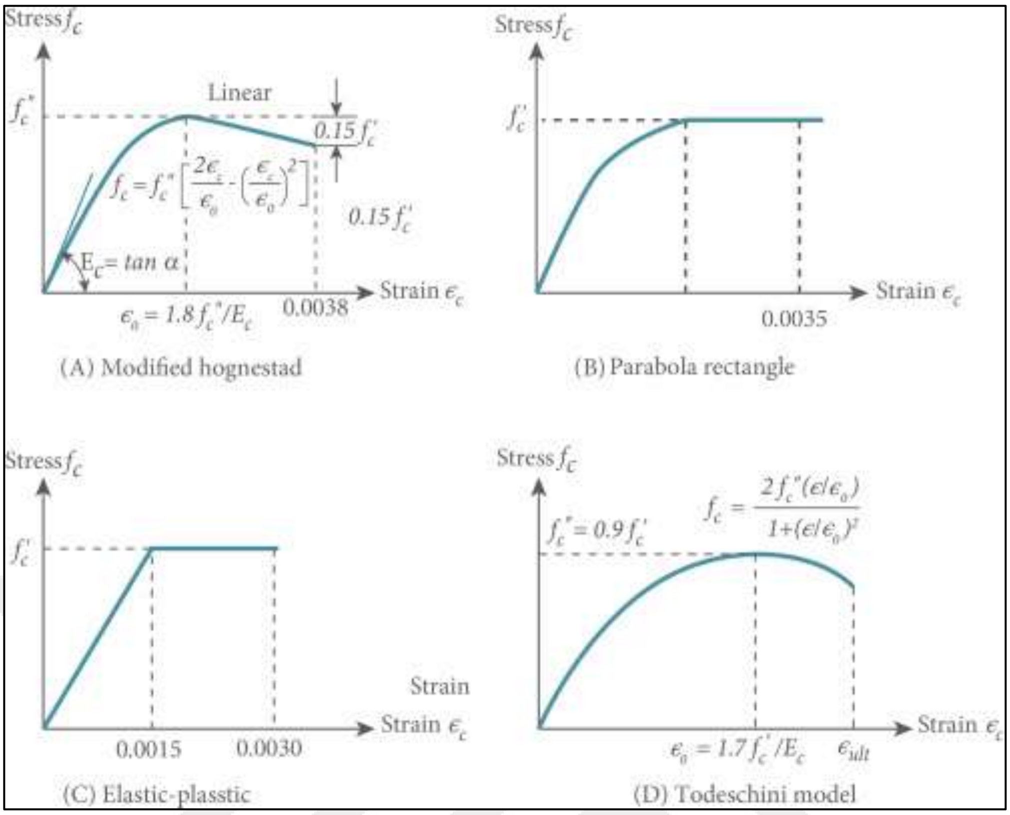

Figure 3.12: Possible Deformations in Curve IV [39].

a. Current measurements on the IV curve greater or less than expected:First, it is important to remember that the main variable for changes in current is irradiation, since the current produced by the cell is directly proportional to the amount of irradiation that reaches it. One of the causes for changes in current levels would be a row or module with uniform dirt, which reduces the irradiation that effectively reaches the cell, decreasing the height of the curve. The same effect can be seen in some situations of non-uniform dirt, for example, if there is a strip of dirt at the bottom of a module, where all rows are affected equally. Another possible cause for a decrease in current is module degradation; this degradation is common due to the time of use and the environmental conditions in which the system is located.

b. Slope of the IV curve, close to the short-circuit current, different from that expected:The slope of the IV curve in this region is mainly affected by the shunt resistance, or parallel resistance, of the row or module. The decrease in shunt resistance causes a greater slope near SCC, due to changes within the cells or photovoltaic modules, where a leakage current arises. Shunt current is a current that is diverted by the junction without producing energy, short-circuiting part of the module or cell. It can arise mainly due to hot spots or defects in the cells and/or connections that cause the current to circulate improperly.

Another factor is the mismatch of the modules' SCC current, which can also cause a reduction in the slope of the IV curve in the SCC region. The current values in real conditions have some small mismatch, whether due to small variations in production, small variations in the installation angle or partial dirt. The impact of this effect is not so visible on the curve, and can be observed as a subtle change in the slope and decrease in the form factor.

c. Slope of the IV curve, near the open circuit voltage, different from the expected:The slope in the IV curve region near the open circuit voltage is related to the increase in the series resistance of the module. The increase in series resistance makes the step of the curve smoother and consequently reduces the form factor.

One of the factors that can lead to an increase in series resistance is the incorrect dimensioning of the photovoltaic system wiring, or the wiring having high resistance. Other causes that can increase resistance are external connections. All connections in the row can add electrical resistance to the circuit, so it is a good idea to check that they are in good condition and that they are well connected.

There may also be an increase in series resistance within the modules themselves due to degradation, or due to the natural aging of the equipment.

d. Curve IV has “teeth” or “steps”:These steps are usually caused by a mismatch between different areas of a row or module. This change indicates that there is current flowing through one or more bypass diodes in the affected modules.

The most common cause of mismatch is partial shading of a module or cells, since in those shaded cells the current is reduced and this limits the current produced by all the cells connected in series to it. To prevent these affected cells from limiting the production of the entire module, there is a bypass diode that short-circuits this row of cells until it normalizes. For a single module, the height (corresponding current) at which distortion or step is found is the bias current of the diode connected to the shaded cells, and the

Voc-OCV voltage for this point is related to the number of rows within the module that have been removed.

Another factor is the presence of damaged, broken or cracked cells, which can generate steps in the curve depending on the severity of the damage. Finally, we have another possible cause of the defect which is the bypass diodes themselves short-circuiting it, which also causes the same visible effect on the IV curve.

e. Open circuit voltage measurements on the IV curve greater or less than expected:The factor that most impacts the voltage of a photovoltaic generator is the temperature of the cells. The higher the temperature, the lower the Voc-OCV voltage, which consequently reduces the energy produced.

Changes in Voc-OCV may indicate that one or more cells are fully shaded with a high level of opacity, since lighter shading can cause the steps or teeth in the curve as mentioned above. In this case, since the shading is opaque, the current at that step point is zero, which appears on the curve as a decrease in Voc-OCV.

Another fault that may be indicated by this type of deformation is faults in the bypass diodes, which may conduct current even without any shading. In this case, the shape of the IV curve may appear normal, except that the Voc-OCV value is lower than expected. Unlike the previous case, in this case, the bypass diode is polarized at all operating points of the module.

3.9 CONCLUSIONS

One of the most useful methods for diagnosing performance and faults in photovoltaic (PV) systems is their current-voltage (I-V) characteristics analysis. This chapter has analyzed in detail the various anomalies that may occur on an I-V curve as well as the I portant conditions, both physical and environmental, that may cause them. Every single deformation or irregularity on the I-V curve corresponds to one behavior in particular of grabbing the energy from the sun through the photovoltaic module or array which gives the value of the health of the system. In this chapter, we have focused on current producing deviations either above or below expectations and have reiterated that irradiance remains the major reason that would influence current production. PV panels receive current flowing through them, and due to dirt being uniform or not, PVOGR may decrease and consequently result in low current output. As a result, dirt that partially or completely blocks the energy of the sun can significantly reduce PV cell current output. Moreover, under-achieving along with environmental wear exposes modules to their surroundings already leads to a degradation in current level. Each of the listed parameters has a distinct drop in power increase of the I-V curve, this sign sensibly decreases the height of it. Last is the slope of the I-V curve around short circuit current value (SCC). Its being larger means that the shunt resistance is also small which kwows that more bypassed current will leak. These issues could be from micro cracks, internal module malfunctions, hot spots, or other module issues. Current module generation within a system is relatively uniform, but mismatches on a minor scale could arise from installation, dirt, or production variations. These minor shifts on the curve alter its slope and fill factor. Afterward, focus was shifted to the slope next to the open circuit voltage (Voc). This region is dominated by high series resistance in the system (series resistance works to Voc in the same way friction works to speed). Increased Series resistance is often detrimental therefore extreme care is required when thinking of the cause. With frequent changes its systems are occurring because of ill arranged wiring schematics, loose screw grafts or corrosion, and even aging of the module. Shifting or raise in the series resistance is the one causing curve flattening closer towards the Voc, reducing the maximum power point and energy output. Without the detailed analysis of the curves these induced resistance losses are very likely to be ignored. Exploratory work on unintentional but somehow the most specific deformation of curve deformations “teeth” or “steps” has also been performed. Such movements are believed to indicate activation of bypass diodes which is most commonly caused by partial shading around the PV module or some cracked deformed or broken cells. The culprit protecting many times uncontrolled overheating and mismatch loss of cell are diodes which level out reserves in some cases spin the controlled way and in some uncontrolled ways rerouting current. However need is simplier detection of excessive consistent or unsteady alteration of steps in the curve which detects infrequent and neglecting formation of the cell resulting in indestructible damage of the cells. Finally, the open-circuit voltage was studied considering unforeseen variations with respect to the cell temperature as the overriding factor. An increase in temperature Voc significantly decreasing which leads to a reduction in energy output, even if current levels are stable. Increased solar irradiance may lead to severe voltage drops, indicating opaque shading deficiency where the current does not flow through the affected portion, or malfunctioning bypass diodes that perpetually engage in conduction without actual shading—voltage being reduced uniformly in all operational states. It is important to note that degradation can cause several deformations in the curve and often has unexpected and random characteristics observed in the electrical parameters. However, it is a slow process and is more identified in modules with a longer period of use, so it is necessary to check the other parameters that affect the curve first, such as irradiation, temperature, dirt and accuracy of measurements before concluding that the module is degraded [38].

4. PROPOSED METHOD

4.1 SYSTEM OUTLINE

4.1.1 Faults in Photovoltaic Systems

For the sake of a comprehensive taxonomy, these issues are classified as either DC-side or AC-side defects [19]. Consequently, this makes it possible to forecast the frequency of faults that will occur in photovoltaic (PV) systems as well as the impact that these failures would have. The word "DC-side faults" refers to a lot of different issues. These include problems with the Maximum Power Point Tracking (MPPT) method, which can be seen in Figure 2, as well as ground faults, arc faults, mismatched cells or modules, open circuits, and short circuits. AC-side problems, on the other hand, are usually linked to inverters or the power grid. Either of these machines can have problems with the AC side. Short circuits, open circuits, and module mismatches are the fault types that are seen most often in them. Each of these kinds of issues needs to be looked into in depth [9]. When there are open circuits, failures are especially dangerous because they limit the amount of electricity that can flow through the circuit. Different parts of the solar system, like where the problem is located and how it is built, affect whether this will affect single module strings or the whole system [20]. A module mismatch occurs when there is a significant disparity in the electrical properties of cells or modules. This phenomenon is referred to as a module mismatch, and it has the consequence of reducing the performance of the system [21]. A mismatch can be either permanent or transitory, and it can be attributable to environmental factors such as snow, dust, or shadows cast by nearby structures or transmission lines. Mismatches that are permanent are produced by degradation or physical damage, whereas mismatches that are transient are caused by these factors.

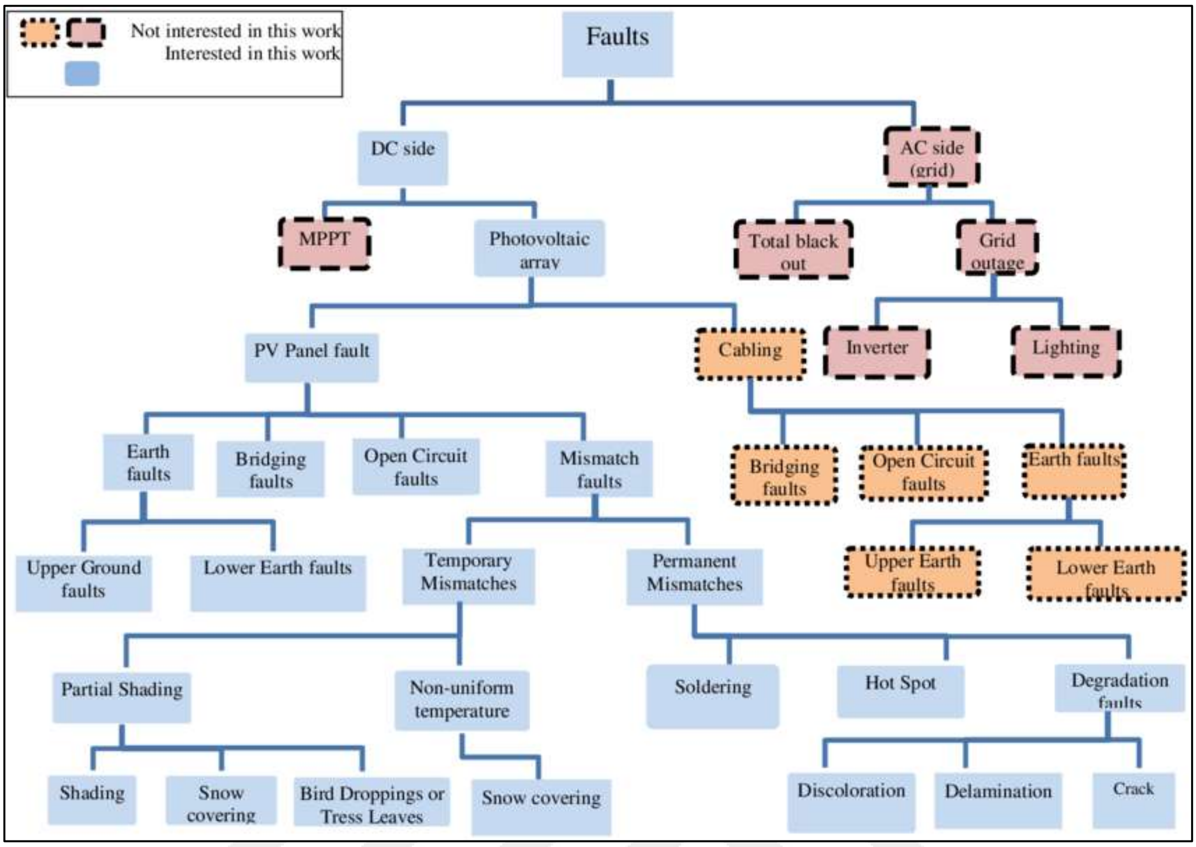

Figure 4.1: DC Side and AC Side Faults in MPPT [22].

4.1.2 Identification of Short-Circuit Defects in Photovoltaic Systems

There are flaws in photovoltaic (PV) systems that can lead to short circuits. These flaws offer a low-impedance way that can happen in several places. In this group, there are places like between module ends, along strings, between strings, and between a string and the ground. The most dangerous ones are those that happen when the negative terminal of one module touches the positive terminal of another module close by. These things are important because they have a big effect on how much energy is made. Because of the short-circuit problems that have been found, this study is being done to find solutions that will make the system more reliable and efficient. Photovoltaic (PV) systems that are linked to the grid need to keep an eye on important operating data so that failures can be found and put into groups. [23, 24]. Only then can failures be identified and classified. There are a number of elements that are taken into consideration, including total irradiance, wind direction and speed, PV array output voltage and current, output power, ambient and module temperatures, grid voltage, and the bidirectional current and power of the grid. According to [25], keeping an eye on these factors in a planned way not only makes it easier to find problems with parts like modules, connecting lines, converters, and inverters, but it also lets you check how well the parts are working right now. This is an important benefit. It is very important to set up a tracking plan that covers everything so that the system is reliable and works well over time.

4.1.3 Dung Beetle Optimization Algorithm

An example of a metaheuristic optimisation strategy that is influenced by nature is the Dung Beetle optimisation Algorithm (DBO), which can be seen in figure 3. This algorithm is meant to emulate the characteristic foraging and navigating activity of dung beetles [26]. Dung beetles are responsible for the process of finding, rolling, and burying dung balls, and this application is meant to imitate that process. This is accomplished by relying on celestial cues, notably the Sun and the Milky Way, in order to complete its voyage in the most effective manner. It is essential to achieve abstraction of this process in order to effectively investigate and make use of potential solution areas while simultaneously addressing optimisation issues.

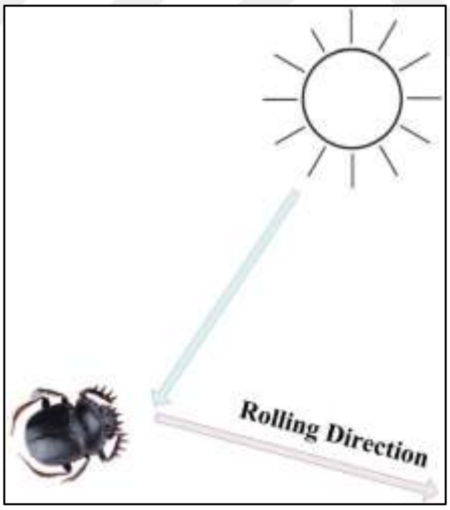

Figure 4.2: Conceptual Illustration of DBO [26].

In the DBO method, alternative solutions are represented as dung beetles that are positioned in a search space that is nnn-dimensional, with each dimension representing a decision variable. This search area is where the DBO process begins. In order to determine the quality of each solution, which is commonly referred to as the "dung ball," an objective function indicated by f(x) is utilised. This objective function demands the algorithm to either decrease or maximise the value of the dung ball of each solution. At iteration t, the sign xit is used to denote the position of the dung beetle that is the iiith in the sequence. Two of the most important stages that are employed by the algorithm are the exploration phase and the exploitation phase.

4.1.3.1 Exploration phase

At this point, the dung beetles are actively searching the entire planet by constantly adjusting their locations in response to signals from above. An update to the position is defined as:

$$\mathrm{x}_i^{t+1} = \mathrm{x}_i^t + r_1(\mathrm{G} - \mathrm{x}_i^t) + r_2(\mathrm{P}_i^t - \mathrm{x}_i^t), \tag{4.1}$$

for which: G stands for the best answer that has been discovered worldwide thus far. Here we have some stochastic behaviour introduced by two random values, r1 and r2, which are uniformly distributed in the interval [0,1]. Pit is the i beetle's personal best position.

4.1.3.2 Exploitation phase

Dung beetles, when exploiting, hone down on solutions that are close to being ideal in order to guarantee convergence. Both local interactions and fine-tuning impact the updated location:

$$\mathrm{x}_i^{t+1} = \mathrm{x}_i^t + \alpha(\mathrm{G} - \mathrm{x}_i^t) + \beta(\mathrm{L}_i - \mathrm{x}_i^t), \tag{4.2}$$

Where the control parameters that equalise the impact of global and local learning are α and β. When it comes to introducing local exploration, Li stands in for a nearby bug..

4.1.3.3 Fitness evaluation and selection

The fitness of every beetle is calculated at each iteration using the objective function f(xit). Both the global and individual best solutions are up-to-date, and the positions are as well:

$$\mathrm{G} = \arg \min_{\mathrm{x}_i} f(\mathrm{x}_i), \mathrm{P}_i^{t+1} = min(\mathrm{P}_i^t, \mathrm{x}_i^{t+1}) \tag{4.3}$$

4.1.4 Fick’s Law of Diffusion Algorithm

The Fick's Law of Diffusion Algorithm (FLDA) is an optimisation algorithm that is derived from the principles of diffusion that are given in Fick's Laws [27]. This method is inspired by biological processes and is developed from the notions of diffusion." The migration of particles from areas of higher concentration to areas of lower concentration is an essential aspect of the fundamental physical process known as diffusion. This movement continues until equilibrium is reached. FLDA is designed to assist the exploration and usage of solution spaces in the process of addressing optimisation difficulties. This is accomplished by implementing an abstraction of this behaviour. The FLDA algorithm portrays alternative solutions as particles that diffuse in a search space that is n-dimensional, with each dimension representing a decision variable. Because of this, it is possible to simulate potential solutions. A solution's quality may be evaluated with the use of an objective function called f(x), and the purpose of the method is to find the best possible solution by simulating the diffusion process.

4.1.4.1 Fick's first law of diffusion

Particle flow J is directly proportional to concentration gradient, according to the first law:

$$J = -D\frac{\partial C}{\partial x} \tag{4.4}$$

where: D is a constant that describes the parameters of the medium; J is the diffusion flux, which represents the rate of particle movement. ∂C represents the gradient of concentration. This idea is the basis for FLDA's preliminary search space exploration. In order to mimic the transition from a high concentration to a low one, potential solutions shift towards areas with lower objective function values..

4.1.4.2 Fick's second law of diffusion

The second law explains how concentrations shift with the passage of time:

$$\frac{\partial C}{\partial t} = D\frac{\partial^2 C}{\partial x^2} \tag{4.5}$$

where ∂C represents the rate of concentration change with respect to time. The gradient of the concentration profile is shown by the second derivative of concentration, which is denoted as ∂2C. Particles refine their locations during the exploitation phase by focussing on potential places found during exploration, and this rule regulates this process.

4.1.4.3 Position update mechanism

For each iteration t in FLDA, the diffusion principles are used to update the position of a candidate solution xit:

$$\mathbf{x}_i^{t+1} = \mathbf{x}_i^t + \gamma D \nabla f(\mathbf{x}_i^t) \tag{4.6}$$

where the objective function's gradient, ∇f(xit), is similar to the concentration gradient. To maintain a healthy equilibrium between exploration and exploitation, the diffusion coefficient (D) is continually modified. The scaling parameter γ determines the size of the step.

4.1.4.4 Exploration and exploitation

In the first rounds, the algorithm promotes large-scale exploration of the search space using global search with greater diffusion coefficients. The exploitation phase entails reducing the diffusion coefficient in subsequent iterations to enable optimised search areas to guarantee convergence through fine-tuning. The algorithm keeps running until it reaches a certain point, whether that's a convergence threshold or a limit number of iterations.

4.2 PROPOSED METHODOLOGY

The purpose of this research is to identify defects in photovoltaic (PV) systems by employing many different learning methodologies, including machine learning and active learning. As a result of the efficient processing and analysis of extensive datasets that have been obtained from PV farms, this objective has been successfully realised. The approach that is being given and represented in figure 4 aims to distinguish and categorise various sorts of defects with precision and flexibility in order to ensure dependable fault detection in environments that are continually changing. This is done in order to guarantee that fault detection will be reliable. The first phases in the process consist of consolidating the data and carrying out operations that are considered to be preprocessing. In order to make the dataset suitable for the training of the model, these procedures are absolutely necessary. The process of exploratory data analysis, which is often referred to as EDA, is carried out with the utilisation of separate training and testing datasets. The purpose of this analysis is to recognise patterns within the data and to address issues such as instances of outliers and missing values. For the purpose of enhancing the performance of machine learning models, it is essential to implement normalisation and standardisation strategies. All of these methods guarantee that the data is presented in an acceptable manner, which is very necessary in order to achieve the highest possible level of interpretability and efficiency with the model.

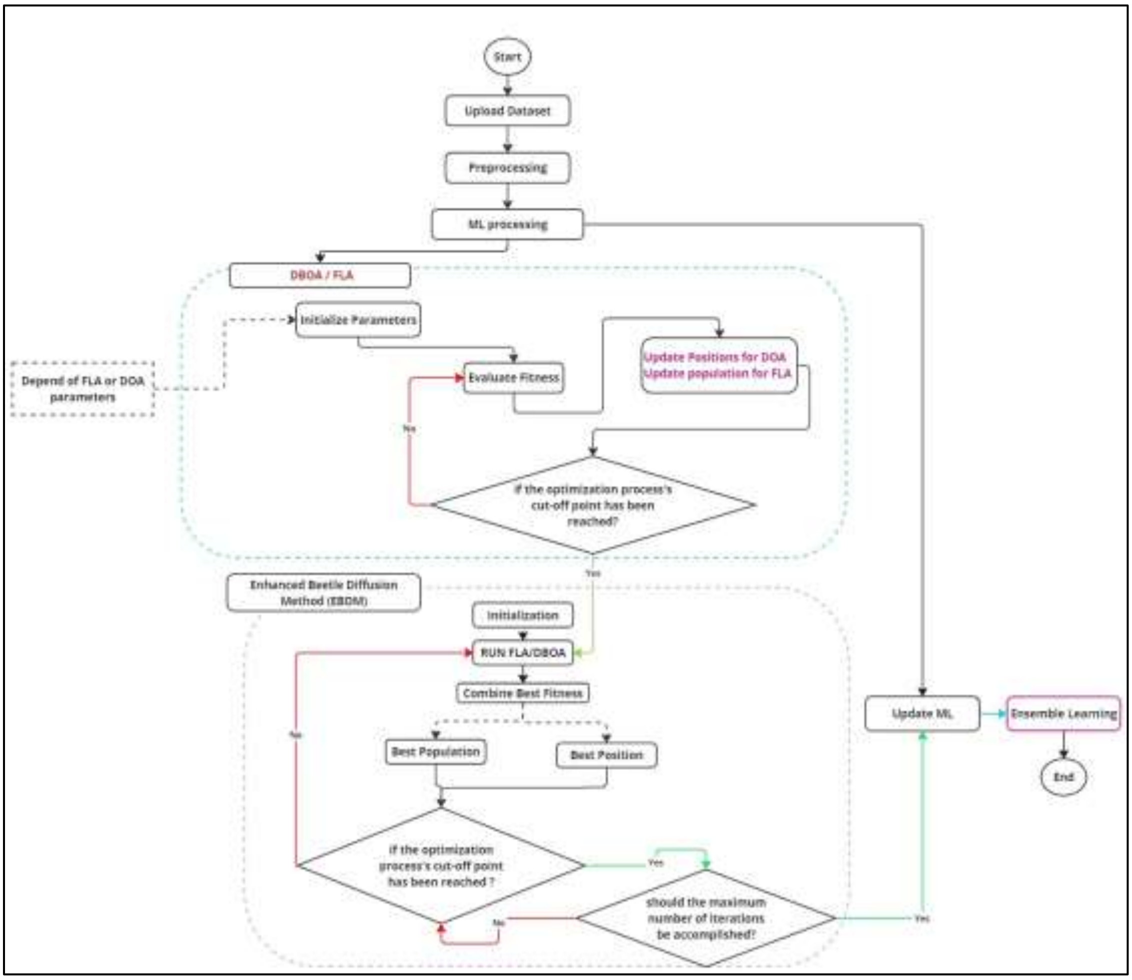

Figure 4.3: Outline of the Proposed Method.

The combination of optimisation and machine learning was used to construct the framework shown in Figure 4.3. The goal of creating this framework was to make photovoltaic (PV) systems better at detecting and classifying faults. A combination of FLA, an Enhanced Beetle Diffusion Method (EBDM), and the Dung Beetle Optimisation Algorithm (DBOA) enabled the accomplishment of this goal. Data preparation, which includes tasks like uploading, preprocessing, and normalising the dataset, is the first step in the technique. Following this step is the processing of the data obtained by machine learning (ML). Within the framework of the optimisation process, prospective solutions are evaluated and modified in an iterative manner depending on fitness criteria. Initialisation of the process can be accomplished by either DBOA or FLA, depending on the parameters that have been expressly stated. In this regard, the criteria that have been defined must be taken into consideration. When the optimisation cut-off point is reached, the Enhanced Beetle Diffusion Method will improve the best solutions by combining the outputs from both algorithms in order to get better results. This will be done in order to acquire superior outcomes. This is going to be done in order to attain the most favourable results that are achievable. Iterative refinement and fitness evaluation are two methods that are utilised by the framework in order to guarantee that there is a good equilibrium between exploration and exploitation. In order to do this, the loop is iterated through until either the maximum number of iterations is reached or the convergence condition is fulfilled, whichever occurs first. Following the completion of the hybrid optimisation phase, the parameters that were successfully enhanced are subsequently utilised in the process of updating machine learning models. After the hybrid optimisation phase has been finished, this further step takes place. Procedures for ensemble learning, such as stacking, are utilised in order to accomplish even higher levels of performance than previously thought possible. The completed product is a defect detection system that is founded on ensembles and incorporates enhancements in terms of both accuracy and reliability. The fact that this is the case lends credence to the notion that the hybrid approach that was presented is efficient in terms of optimising the process of identifying and classifying PV problems. In the following sections, we will dig into the specifics of the machine learning models that are used for the goal of defect detection. These models are used to identify defects. Both the methods and the settings of those algorithms, as well as the influence that those configurations have on the effectiveness of the detection, will be included in these models. In addition, the incorporation of active learning approaches will be investigated, with a particular focus on the role that these techniques play in boosting the performance of models and their adaptability in circumstances that are based in the real world. To ensure that it will continue to maintain its reliability and high detection accuracy over the course of time, the detection system is able to learn and improve continuously as new data becomes available. This is made possible by the approaches that are being utilised.

4.2.1 Dataset Overview

under the year 2023, the dataset that was created by Python Afroz and given the name "PV Plant in Normal and Shading Faults Conditions" was at long last made available on Kaggle under the name [22]. This dataset's mission is to fulfil its purpose, which is to provide a complete collection of data that can be utilised for fault detection and classification in photovoltaic (PV) systems. The goal of this dataset is to fulfil its purpose. From a more specific perspective, the dataset places an emphasis on shading failure events in addition to standard operating settings. It is easier to construct and assess machine learning models for the aim of finding weaknesses in photovoltaic (PV) systems because the dataset is arranged in a way that makes it simpler to do so. This is in addition to the fact that the dataset provides a solid basis for exploratory research and algorithmic testing. The dataset is therefore a valuable resource for both sorts of research because of this. When it comes to finding and evaluating defects in photovoltaic (PV) systems, the dataset contains a significant number of vital qualities that are required for doing so. These characteristics are essential since they are also essential. Examples of environmental characteristics that fall under this category include the amount of irradiation they receive from the sun as well as the temperature. The voltage, current, and power output from solar modules and arrays are some of the electrical metrics that are included in this category. On the other hand, this category also includes additional components. For the purpose of making it easier to implement supervised learning strategies, the dataset has been supplemented with labels that distinguish between normal operating conditions and a variety of shading failure scenarios. Taking this action was done in order to make the process of data collecting easier. Due to the fact that shading is one of the most common and difficult problems that might have an effect on the operation of solar systems, the addition of shading fault data is extremely crucial. This is because shading is one of the most common and tough concerns. Due to the fact that it possesses a high degree of granularity and is applicable to the actual world, this specific dataset is an excellent option for applications that include machine learning. Using the data that has been tagged, it is possible to train machine learning algorithms to identify and classify errors, evaluate the effect that shade has on system performance, and enhance the dependability and efficiency of photovoltaic (PV) systems. The use of the data has the potential to do all of these things. The dataset is a useful instrument for boosting research in the field of photovoltaic defect identification due to the structure and diversity of the data that it contains. In addition to this, it provides a solid foundation for the optimisation and assessment of diagnostic systems that are becoming increasingly sophisticated. The dataset provides a comprehensive collection of data relevant to defects, electrical systems, and physical attributes from a variety of sources. The information presented here can be utilised for the purpose of evaluating and diagnosing issues that arise in photovoltaic (PV) systems, both in the context of regular fault conditions and shaded fault conditions. Using sensors that were made by Delta-T Devices, this is comprised of the measurements of global and diffuse tilted irradiance (W/m2) that were recorded. In contrast, it takes into account the thermal measurements of the ambient and back surface temperatures (in degrees Celsius) that were acquired through the use of thermocouples of Type K and Type T designs. These thermocouples were used to get the data. The currents that pass between solar plants and the grid are measured in amperes, the voltages are measured in volts, and the frequency of the grid is measured in hertz. All of these measurements are taken into consideration. Watts and kilowatts are the units of measurement that are used to describe the amount of energy and electricity that is injected into the system. A wide range of distinct requirements are employed in order to accomplish the task of identifying different types of faults. There are a number of requirements that are included in this category, including but not limited to: no-fault operation, uniform shading, continuous partial shading of entire modules (affecting up to three modules), partial shading of module segments (1/31 or 2/32), and intermittent partial shading in static patterns. When it comes to the building of machine learning models that are able to detect and classify PV system defects with a high degree of precision, the dataset is a perfect resource to use. This is due to the fact that the dataset includes a diverse range of data points, and the fault annotations are precise.

4.2.2 Preprocessing

When it comes to the process of preparing data for fault identification in photovoltaic (PV) systems that are based on machine learning, one of the most important steps is to carry out adequate preprocessing. This is because accurate data preparation is essential for fault identification. The implementation of this method is done with the purpose of ensuring that the raw data is cleaned, standardised, and presented in a manner that is appropriate for classification models. Before moving on to the preprocessing stage, the training dataset and the testing dataset were mixed respectively. During the training and testing stages, this was done to ensure that the distribution of the data would remain accurate and consistent throughout the whole process. The completion of this stage was very necessary in order to take preventative measures against discrepancies that might potentially have an effect on the performance of the model. Following the combination of the datasets, the 'class' column was removed from the feature set. This was done in order to successfully separate the dependent variable, which was comprised of class labels, from the independent variables, which were features. This action was taken in order to accomplish the separation that was intended. The division is done with the intention of developing machine learning models that are able to predict class labels based on the characteristics. As a consequence of this division, the training process is simplified, which is done for the purpose of constructing. Following the completion of the conversion procedure, a numerical value was assigned to each and every one of the nominal and categorical variables that were incorporated into the dataset. In view of the fact that the great majority of machine learning approaches require numerical inputs, this was carried out after careful consideration. StandardScaler, a tool that alters the data in such a manner that it has a mean of zero and a standard deviation of one, was also used to normalise the feature set. This was done in order to reduce the amount of variation in the data. Taking this action was done in order to guarantee that the data would continue to be consistent. This normalisation is especially crucial for models that rely on regularisation or gradient descent optimisation [29], as it helps to preserve scalability and prevents any one feature from playing a dominant role in the learning process. In addition, it helps to ensure that the learning process runs smoothly. There is no possible way to overestimate the significance of this normalisation, particularly for models that are reliant on consistency. This was one of the most important steps that needed to be followed in order to lessen the possibility of mistakes or biases arising during training. Dealing with information that was lacking was one of the processes that needed to be accomplished. The choice on whether or not to remove missing data or to impute it was made after taking into consideration the quantity of the data as well as its importance. Outliers that had the potential to change the distribution of the data or to degrade the performance of the model were identified and dealt with in accordance with the processes that were suitable. Subsets of the dataset were separated into training and testing subsets after the preprocessing stage was finished. This was done to guarantee that an objective evaluation of the performance of the model could be carried out afterwards.

4.3 DBOA + FLA

Machine learning model parameters are adjusted in a unique way by DBOA and Fick's Law of Diffusion. In this hybrid approach, the unique characteristics of each algorithm encourage investigation and use, leading to the rapid identification of optimal parameters.

4.3.1 DBOA (Dung Beetle Optimization Algorithm)

The DBOA was inspired by dung beetles that carried dung balls. To determine the best course of action, the program mathematically represents DBOA's updating process and then mimics the beetle's movement with each "dung ball.":

$$v_{i+1} = v_i + \alpha \cdot (v_{best} - v_i) + \beta \cdot R \tag{4.7}$$

Here, \(v_i \) represents where the solution is in the search space at the moment. The best solution's current location is denoted by \(v_best \). The step size factor, denoted as \(α\), controls the impact of the best-known solution. \(β\) is a factor that increases exploration capabilities through randomisation. Each element of the random vector \(R\) is selected at random from a uniform distribution.

4.3.2 FLA (Fick's Law of Diffusion Algorithm)

From very high concentrations to very low ones, FLA uses Fick's particle diffusion equations. Performance density is optimised by FLA. One way to characterise the FLA formula is as:

$$[x_i = x_i + \gamma \cdot \left(x_j - x_i\right) + \delta \cdot (U(0,1) - 0.5)] \tag{4.8}$$

If (x_j) outperforms (x_i), then (x_i) and (x_j) are two solutions in the population. The comparative performance gap between (x_i) and (x_j) influences the learning factor (`). (δ) is a little noise component that keeps the population diverse. (U(0,1)) stands for a uniformly distributed integer between zero and one.

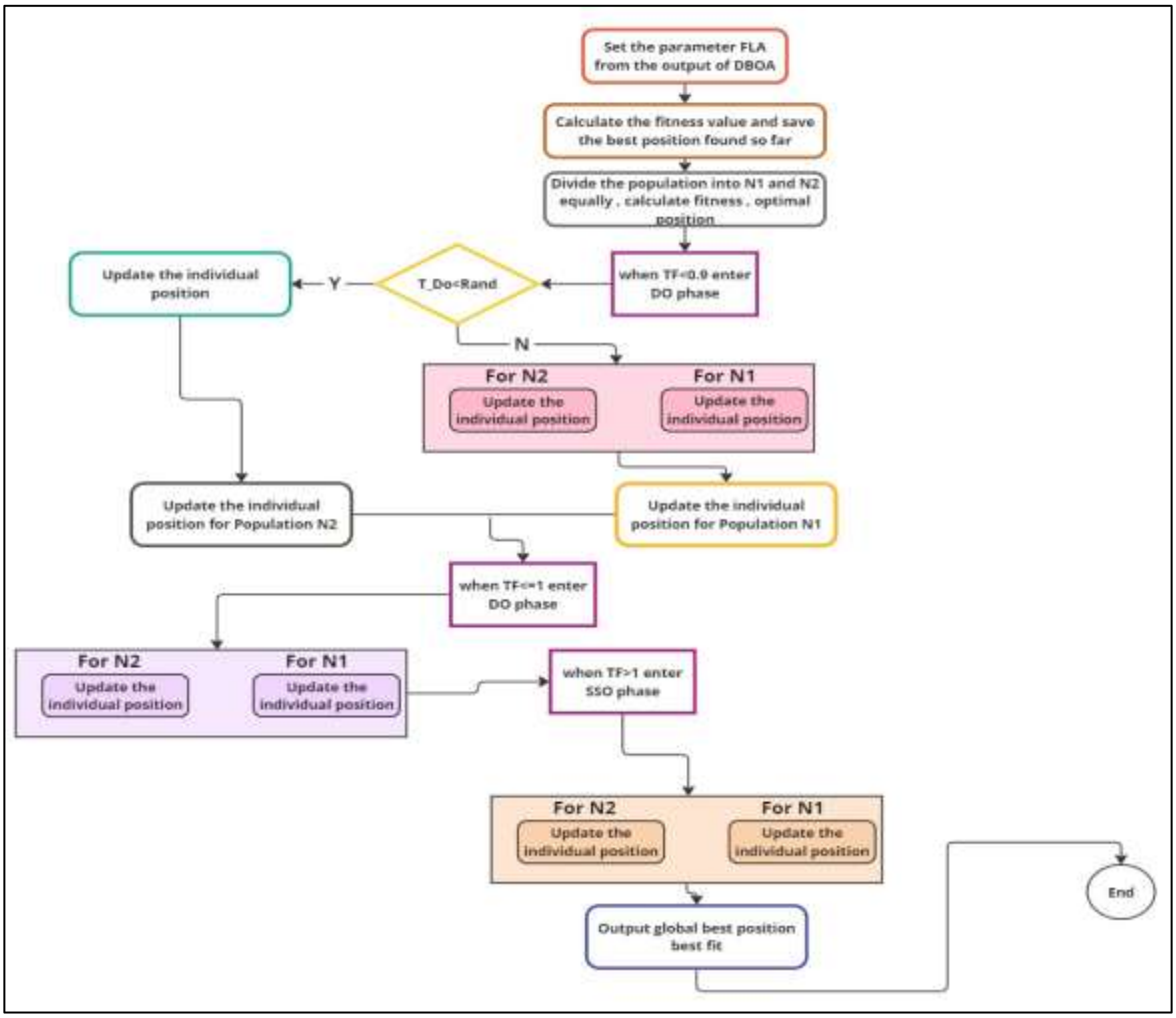

Figure 4.4: Proposed BDOA + FLA Scheme.

4.3.3 Combination of DBOA and FLA

In their pursuit of knowledge, DBOA and FLA employ straight-line movement and adaptive adjustments based on solution quality. This is useful for adjusting parameters in machine learning and other complicated optimisation environments. Planning the unified upgrade:

$$[x_{new} = DBOA(x) + FLA(x)] \quad (4.9)$$

Based on the local solution density for both global search and local precision, the new solution \(x_new \) employs FLA to fine-tune the method after DBOA finds the optimal solution. With this hybrid approach, we may increase optimisation outcomes in fewer rounds by identifying many solutions while searching in appealing locations.

Algorithm: DBFLA

Input:
Objective function f(x), population size N, maximum iterations T
Output: Best solution found x_best

---

1. Initialization:
   - Randomly generate an initial population of N candidate solutions $\{x_1, x_2, ..., x_n\}$.
   - Evaluate each solution using the fitness function f(x).
   - Identify the best initial solution and store it as x_best.

---

2. Iterative Optimization (for t = 1 to T):
a. Solution Evolution:
For each candidate solution x_i in the population:
i. Exploration Phase - Modified Adaptive Movement (MAM):
   - Generate a random exploration factor $\eta \in [0, 1]$ and momentum coefficient $\mu \in [0.5, 1]$.
   - Calculate an adaptive step:
     $x_i \leftarrow x_i + \eta * (x\_best - x_i) + \mu *$ random_direction_vector()
   - Ensure that x_i remains within defined variable bounds.

ii. Exploitation Phase - Feedback Learning Mechanism (FLM):
   - Randomly pick a peer solution $x_p \neq x_i$.
   - If $f(x_p) < f(x_i)$, enhance x_i using dynamic feedback:
     $x_i \leftarrow x_i + \lambda * (x_p - x_i) + \theta *$ randn(),
     where λ is a fitness-difference scaling factor, and θ controls small stochastic exploration.
   - Clamp the updated x_i to remain within valid limits.

b. Diversity Control (Optional):
   - Every few iterations, reinitialize a small fraction of the worst-performing solutions to boost exploration and prevent stagnation.

---

3. Fitness Evaluation and Best Solution Update:
   - Recalculate fitness values $f(x_i)$ for all updated solutions.
   - If any x_i improves upon x_best, update x_best $\leftarrow x_i$.

---

4. Termination Check:
   - If t = T, or convergence is detected (e.g., no improvement for several iterations), terminate the loop.

---

5. Return Result:
Return the best solution found x_best.

DBOA and FLA are both components of the hybrid optimisation approach that is being described by the technique that is being presented. According to the description, this strategy is a hybrid. With the application of this tactic, the purpose is to identify the most effective solution that can be found within the problem area that has been provided. Each of the stages that are offered is described in further depth in the following article: Immediately after the population has been created, the process will begin using the initialisation of the population as its starting point. At this stage, it is important to generate a collection of random solutions that are comparable to the population size that has been specified. This population size is typically denoted by the symbol \(N \. Following that, each of these solutions is examined to establish whether or not it is physically viable in accordance with the objective function \(f(x)\), which evaluates the degree to which a solution successfully supplies a solution to the issue. First, each potential solution is evaluated to establish whether or not it is feasible from a physical standpoint. Based on the fitness assessments that are carried out, this initial population is utilised in order to pick the best solution, which is represented by the symbol \(x_"best" \). During the iterative phase of the technique, the algorithm performs operations on each and every solution \(x_i \) that is part of the population for each iteration, up to a maximum number of iterations \(T \). After reaching the maximum number of iterations, this phase will continue until it is completed. Carry on in this manner until the algorithm reaches the maximum number of iterations that it is able to process. Each individual solution goes through two rounds of modification throughout the course of each iteration. The first round is completed by utilising the DBOA, and the second round is completed by utilising the FLA. Both rounds are carried out in consecutive order. During the time that the DBOA phase is being carried out, each and every answer \(x_i \) comes under scrutiny for modification. The initial stage in this modification requires the identification of a step size, which is represented by the symbol (α), as well as a direction vector, which is represented by the symbol \(R a). Both of these variables are selected in a manner that is completely arbitrary. Moreover, a random exploratory step that is impacted by \(R \) is included into the solution, which is then updated by pushing it towards the best current solution \(x_"best" \). This process is repeated until the solution is updated. This procedure is carried out once more each time the solution is given an update. In order to exercise control over the influence of this unpredictability, the use of the parameter (β) is applied throughout the process. It is of the utmost importance to make certain that this is something that is carried out in order to guarantee that the new solution x_i does not go beyond the boundaries that have been established for the problem space. Within a short period of time after the completion of the DBOA, the FLA is employed in order to further enhance each answer. The current response, which is denoted by the symbol x_i, is compared to every other solution, which is denoted by the symbol x_j, that is already existing in the population. This comparison is carried out within the context of the population. A learning factor (γ) is computed by using the difference in fitness between the two nodes that are being compared. This is done in the event that any node x_j possesses a higher degree of fitness in comparison to x_i. In order to bring about the alteration of the solution x_i towards x_j, a process that is considered to be known as diffusion is employed. Not only does this specific process move x_i in the direction of x_j, but it also incorporates a random fluctuation that is controlled by (δ) and a uniform random variable $U(0,1)$ at the same time. The adjustment is carried out while simultaneously ensuring that the solution stays within the boundaries of the issue region. This is done to guarantee that the problem area is appropriately addressed. After each iteration, the fitness of the population that has been updated is reassessed in accordance with the outcomes of the evaluation that came before it. It is possible that if there is a newly adjusted solution that performs better than the current x_{best} in terms of fitness, then the x_{best} will be altered to reflect the solution that is superior. It is intended that the iterative process will be completed when the maximum number of iterations, which is denoted by the symbol T, has been achieved. This is the point at which the process will be considered complete. At this point, the method delivers the best solution x_{best} that was discovered over the iterations. This answer is believed to be the optimal or a near-optimal solution to the issue based on the objective function that was provided.

4.4 OPTIMIZATION OF ANN AND SVM USING DBOA AND FLA

To optimise the hyperparameters of Artificial Neural Networks (ANN) and Support Vector Machines (SVM) for fault detection in photovoltaic (PV) systems, the aforementioned method combines the Dung Beetle Optimisation Algorithm (DBOA) with Fick's Law of Diffusion Algorithm (FLA). As shown in Table 2, the following parameters were optimised for ANN: hidden layers, neurones per layer, learning rate, activation function, and batch size.

The kernel type, regularisation parameter $\mathbf{O}$, and kernel parameters γ are all part of the SVM hyperparameter set; they are described in detail in Table 4.1.

Table 4.1: ANN Hyperparameters Optimized.

Parameter	Range/Values	Optimized Value
Layers	1–4	4
Neurons	32–256 (multiples of 32)	168
LR	0.001–0.1	0.05
AF	ReLU,	ReLU
BS	4, 6, 8, 12	64

The optimised ANN hyperparameters are listed in Table 4.2. The learning rate controls how quickly weights are updated, while the number of layers and neurones per layer determine the capacity and depth of the network. The non-linear transformation that neurones undergo is determined by activation functions, and the training gradient updates are affected by batch size. In order to attain better classification accuracy and computational efficiency, these parameters were optimised using DBOA and FLA.

Table 4.2 SVM Hyperparameters Optimized.

Parameter	Range/Values	Optimized Value
Kernel Type	Linear, RBF, Poly	RBF
Regularization (C)	0.1–100	10
Gamma (γ\gammaγ)	0.001–1	0.01

As a first step in optimisation, DBOA conducts global exploration by emulating dung beetle navigational techniques to keep the solution space diverse. The next step is FLA, which uses diffusion principles to zero in on potential solutions for domestic implementation. We repeatedly update the optimal configurations after evaluating each candidate solution, which represents a combination of hyperparameters (table 4.3), using a fitness function based on cross-validation accuracy. While FLA parameters, such as diffusion coefficient and step size, improve local refinement, DBOA factors, including population size, maximum iterations, and exploration-exploitation balance coefficients (Table 4.4), guarantee robust global search.

Table 4.3 DBOA Parameters.

Parameter	Value
Population Size	50
Maximum Iterations	100
Exploration Coefficient	0.7
Exploitation Coefficient	0.3

By attaining ideal neurone and learning rate configurations, the optimised ANN obtained excellent accuracy in fault classification, including module mismatches, open circuits, and short circuits. In contrast, the SVM produced superior classification utilising the RBF kernel with finely adjusted a and γ values.

Table 4.4 FLA Parameters.

Parameter	Value
Diffusion Coefficient	0.5
Step Size	0.01
Iteration Limit	50

The FLA parameters are presented in Table 4.4. Simulation of particle diffusion allows for the exploitation of potential regions in the search space through the refining of solutions, which is governed by the diffusion coefficient and step size. An iteration's maximum steps for local exploitation are defined by its iteration limit. With a robust and scalable hybrid technique, we were able to obtain a high classification accuracy of 98.75% and efficiently tune the parameters for real-world PV system defect detection.

5. SIMULATION AND RESULTS

5.1 SIMULATION AND HARDWARE SPECIFICATIONS

The fact that the battery was able to achieve a charging and discharging efficiency of 95% ensured that the management of energy was carried out in a safe and secure way. Through the utilisation of the suggested Dung Beetle Optimisation Algorithm in conjunction with Fick's Law of Diffusion Algorithm (DBFLA), the Maximum Power Point Tracking (MPPT) controller was able to enhance the power extraction that was carried out by the PV array. Because of this, the controller was able to attain its highest possible performance. Through the application of this sophisticated optimisation method, it was possible to achieve a tracking efficiency of 99.2% across a wide range of environmental factors. Compared to the performance of conventional MPPT methods, this one was significantly better. Altering the duty cycle of the buck-boost converter in a dynamic manner allowed for the successful completion of this task. There were many possible failure situations that were selected from the dataset, and simulations were performed on each of them. Some of the scenarios that were taken into consideration were open circuits in one or more modules, partial shading that affected between one-third and two-thirds of the modules that were defined, and uniform shading that reduced irradiance by thirty percent. Throughout the course of the simulation, the outputs of voltage, current, and power were recorded at intervals of one second. The ANN and SVM models that were optimised for DBFLA were employed for this particular purpose. This provided information that was gathered in real time and had the potential to be employed for fault categorisation. The computational load was efficiently regulated by the MSI EvoBook, which resulted in MATLAB simulations being finished within an average runtime of forty-five minutes for each scenario. It was possible to execute this successfully. The findings revealed that the system is capable of properly detecting and categorising faults (with an accuracy of 98.75%), as well as maintaining good MPPT performance performance. In addition, the system was able to sustain strong performance. The studies provided conclusive evidence that this is the case beyond any reasonable doubt. Validation of the hybrid optimisation approach that was presented for use with photovoltaic (PV) systems in the actual world was accomplished with the assistance of this extensive simulation. Not only did this simulation lay the platform for future additions to bigger installations, more complicated load situations, and real-time monitoring integrations, but it also set the foundation for future extensions. The simulation was carried out with the assistance of an MSI EvoBook that was outfitted with an Intel Core i7-1260P central processing unit (CPU) that had 12 cores, 16 threads, and a maximum clock speed of 4.7 GHz, 16 gigabytes of LPDDR5 RAM, a solid-state drive with a capacity of 1 terabyte (TB), and integrated graphics capabilities from Intel Iris Xe. For the purpose of carrying out the simulation, MATLAB and Simulink were utilised. The simulation of the photovoltaic (PV) system under typical test conditions was carried out with the assistance of a total of fourteen photovoltaic (PV) modules, each of which had a rating of 250 watts. These conditions comprised a temperature of 25 degrees Celsius for the module and an irradiation of 1000 watts per square metre. The arrangement of these modules in a series-parallel configuration, with seven modules per string and two parallel strings, was found to be capable of producing a total maximum power output of 3.5 kW. This was determined by the findings of the investigation. It was decided to make use of a buck-boost converter in order to achieve the goal of managing the output of the solar array. This converter was responsible for controlling the voltage, which was in the range of 150 V to 300 V, in order to fulfil the requirements of a group of typical home loads that were distributed across three AC buses. These loads were scattered throughout the three AC buses. Within the framework of the load design, there were a variety of resistive and dynamic loads that were integrated. The lights (1.2 kW), the refrigerator (300 W), the heating, ventilation, and air conditioning system (two kW), and the other appliances (five hundred W) were all included in these loads. Each bus was linked to a load-balancing mechanism and operated at a frequency of fifty hertz (Hz) while working at 230 volts of alternating current (AC), as shown in figure 5.1. This was done in order to guarantee that the distribution of power was consistent across the whole system.

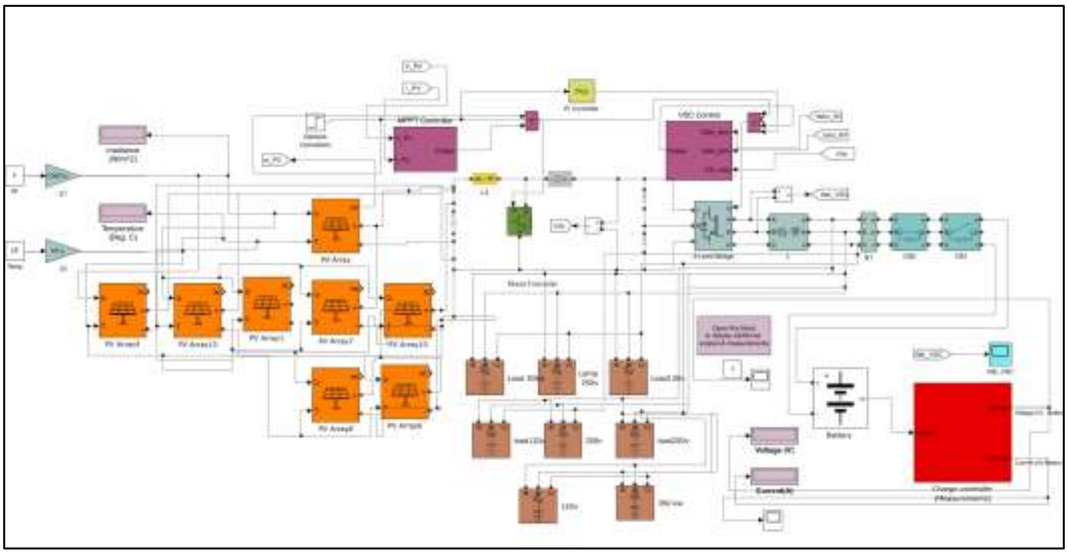

Figure 5.1 Basic Diagram of the Simulation Module.

The battery had a charging and discharging efficiency of 95%, ensuring reliable energy management. To optimize the power extraction from the PV array, the MPPT controller employed the proposed Dung Beetle Optimization Algorithm combined with Fick's Law of Diffusion Algorithm (DBFLA). Our cutting-edge optimisation solution outperformed the conventional MPPT approaches and was able to attain a tracking efficiency of 99.2% across a wide range of environmental circumstances. By making consistent adjustments to the duty cycle of the buck-boost converter, this objective was successfully realised. Through the use of the dataset, a variety of distinct failure situations were generated and simulated. Among these were open circuits in one or more modules, partial shading that impacted one-third and two-thirds of certain modules, and uniform shading that resulted in a thirty percent fall in irradiance. All of these were caused by shading. During the simulation, voltage, current, and power outputs were collected at regular intervals of one second. As a consequence of the simulation, the DBFLA-optimized ANN and SVM models were given with real-time data for fault classification. The fact that Matlab simulations were finished in around forty-five minutes on average for each scenario is evidence that the MSI EvoBook is capable of effectively managing the computational load that it is responsible for. The findings showed that it is feasible to keep a high MPPT performance while also properly recognising and classifying errors (with a precision of 98.75%). This was proved by the data. The comprehensive simulation served to test the hybrid optimisation strategy that was presented for usage with photovoltaic (PV) systems in the real world. In this way, the groundwork is laid for its eventual expansion to incorporate larger installations, load situations that are more sophisticated, and linkages with real-time monitoring.

Table 5.1: Simulation Parameters.

PV Array	Number of Modules	14
	Module Power Rating	250 W
	Configuration	7 modules per string, 2 parallel strings
	Total Power Output	3.5 kW
	Operating Voltage Range	150 V–300 V
	Irradiance (Standard Test Conditions)	1000 W/m²
	Module Temperature	25°C
Buck-Boost Converter	Input Voltage Range	150 V–300 V
	Output Voltage	230 V AC
	Efficiency	96%
Battery	Type	Lithium-ion
	Nominal Voltage	48 V
	Capacity	200 Ah
	Depth of Discharge	80%
	Charging/Discharging Efficiency	95%
Loads	Lighting Load	1.2 kW

Table 5.1: Simulation Parameters "Table Continued".

	Refrigerator Load	300 W
	HVAC Load	2 kW
	Miscellaneous Appliances Load	500 W
	Total Load	4 kW (distributed across 3 buses)
MPPT Controller	Algorithm	DBFLA
	Tracking Efficiency	99.20%
Fault Scenarios	Uniform Shading	30% irradiance reduction
	Partial Shading	1/31/31/3 or 2/32/32/3 of specific modules
	Open Circuit	Single or multiple modules

Yo**u** may learn everything about the simulation setup, including the parameters, features of the PV system, load characteristics, and failure situations, from Table 5.1.

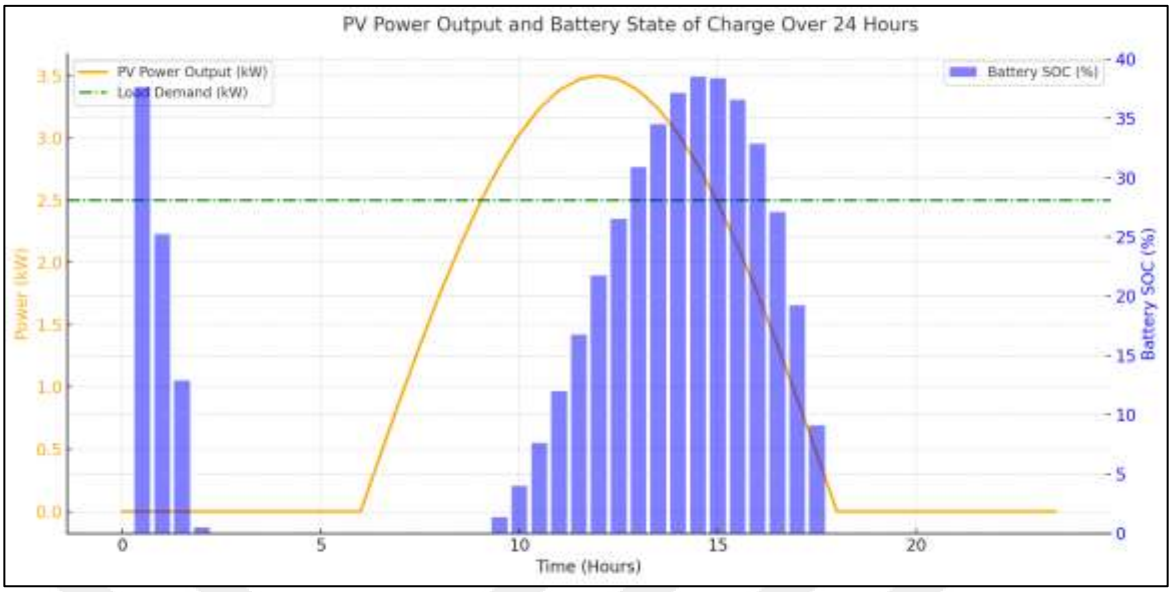

Figure 5.2: PV Power Output and Battery Charge Simulation Over 24 Hours.

As can be seen in Figure 5.2, the power generation of solar cells follows a normal diurnal cycle. The power production reaches its peak of 3.5 kW at the middle of the day, and then it drops to zero throughout the course of the night. An illustration of the state of charge (SOC) of the batteries is provided in the form of a bar chart. Batteries are drained when the power production from the PV system is inadequate, and they are charged when the power output from the PV system is higher than the load demand of 2.5 kW. In other words, the batteries are charged when the load demand can be exceeded. By making this change, which guarantees that the statistics related to power and energy are portrayed in an exact manner, the conclusions that were gained from the results of the simulation are strengthened. During the course of a simulation that ran for twenty-four hours, this graphic illustrates the quantity of power that was generated by the photovoltaic (PV) system, the state of charge (SOC) of the battery, and the demand for residential loads. All of these figures were gathered. The orange line depicts the photovoltaic power generation (kW), which reaches its peak point of 3.5 kW during the middle of the day and then goes to zero during the night. This pattern of solar irradiation is usual, and the orange line displays this trend. A representation of the battery's state of charge (SOC) may be seen in the blue bar chart. This is due to the fact that the battery stores a greater amount of solar energy throughout the day, for the purpose of supplying residential loads, and then releases that energy at night. A constant load need of 2.5 kW is depicted by the green dashed-dotted line, which is a representation of the power consumption profile throughout the day. The colour green is used to depict this line. When the rate of PV generation is larger than the rate of consumption during the day, the battery is charged. This occurs when the battery capacity is charged. For the purpose of making up for the lack of available energy, the battery is discharged during the nighttime hours, when the photovoltaic (PV) production is at its lowest point. This updated visualisation sheds light on the difference between cumulative energy and instantaneous power in order to address the concerns that were brought up by reviewers with relation to oscillations and unit representation. It has been proved through the simulation that the system is capable of producing electricity through the utilisation of photovoltaics (PV) and successfully managing its battery storage, which enables it to keep the energy balance on its own. Through the enhancement of the efficiency of the maximum power point tracking (MPPT) optimisation framework, the photovoltaic (PV) system is able to produce maximum power extraction in a reliable manner. The findings, which show the system's potential to manage fluctuating weather circumstances and load demands, offer assurance that the system is capable of delivering stable energy supply for residential consumption.

5.2 RESULTS

An examination of the machine learning models was performed prior to the beginning of the optimisation procedure. Metrics like as accuracy, recall, precision, sensitivity, and specificity were employed for the purpose of performance assessment. These metrics were utilised for the purpose of evaluating performance. The various models each have their own unique set of benefits and drawbacks, which may be measured at the appropriate level. The Random Forest classifier received the highest score of 97.32% when compared to the other classifiers since it was the most accurate. The fact that it achieved a score of 97.32% on the F1 test, a score of 97.32% on the recall test, and a score of 97.36% on the accuracy test demonstrates that it performed extraordinarily well in every respect. A high degree of accuracy and precision was accomplished by the model, with the accuracy coming in at 96.50% and the precision coming in at 100%. Both of these figures are quite impressive. Quadratic Discriminant Analysis was also highly successful, as seen by its precision of 96.61% and its accuracy of 96.43%. In a similar vein, this particular method was quite effective. Given that it had a sensitivity of one hundred percent but a specificity of eighty-seven point three percent, it had a tough time classifying all of the negative situations that may occur. Once the Gradient Boosting technique was used, the accuracy, F1, and recall values were all found to be within a close vicinity to 96.07%. There was a little improvement in the accuracy of the forecast, which came in at 96.23% respectively. It was proved that the model was able to detect positive instances without causing any false negatives. This was due to the fact that the model had a sensitivity of 98.57% and a specificity of 100% within its parameters. It is clear from looking at Figure 5.2 that solar cells produce electricity in a manner that is consistent with the typical daily cycle. The power output reaches its peak point of 3.5 kW in the middle of the day, and then it drops to zero percent all through the night. This cycle is distinguished by the fact that the power output reaches its highest point. In order to provide a visual representation of the state of charge (SOC) of the battery, a bar chart is employed. As can be seen from the graph, the battery is discharged when the photovoltaic (PV) system is not providing sufficient power to fulfil the load need of 2.5 kW, and it is charged when the PV system is producing more electricity than the load demand and the load requirement is less than the PV system's output. In the event that this modification is put into effect, it guarantees that the quantities of energy and power are accurately represented. This, in turn, strengthens the dependability of the findings that were derived from the simulation. The parameters that have an effect on the amount of electricity that photovoltaic (PV) systems produce and the amount of charge that batteries have currentl, and the load demand of houses are all depicted in this picture, which was created as a result of a simulation that lasted for twenty-four hours. The orange line, which depicts the photovoltaic power production in kilowatts, is a representation of a typical sun irradiation pattern. The amount of electricity that is generated reaches its peak at around 3.5 kW in the middle of the day, and then it goes to zero at night. The percentage of the battery's charge that is now available is shown in the blue bar chart. This chart depicts the charge level of the battery as it stores more solar energy during the day and discharges it at night in order to provide loads connected with the residence. The constant load requirement, which is estimated to be 2.5 kW, is represented by the green line with dashed dots. The profile of the power use throughout the day is depicted by this line. During daytime hours, the battery is charged when the quantity of photovoltaic (PV) generation is more than the amount of demand for the energy. On the other hand, when the photovoltaic (PV) production is reduced to zero during the night, the battery will discharge in order to compensate for the lack of energy. Consequently, the difficulties that the reviewers had about oscillations and unit representation have been addressed by this new visualisation, which ensures that there is a clear demarcation between power (instantaneous) and energy (cumulative). This simulation is being run with the intention of demonstrating that the system is capable of maintaining energy homeostasis through the use of photovoltaic (PV) power generation in conjunction with efficient management of battery storage. Increased efficiency of the maximum power point tracking (MPPT) system is achieved through the implementation of the recommended optimisation framework. This, in turn, enables the photovoltaic (PV) system to continue to extract the greatest amount of power. The findings indicate that the system is capable of adjusting to a broad variety of climatic conditions and load requirements, which ensures that the distribution of energy for residential use is reliable..

Table 5.2: Machine Learning without Optimization Results.

Classifier	Accuracy	F1 Score	Recall	Precision	Sensitivity	Specificity
Random Forest	97.32%	97.32%	97.32%	97.36%	96.50%	100.00%
Quadratic Discriminant Analysis	96.43%	96.37%	96.43%	96.61%	100.00%	87.37%
Gradient Boosting	96.07%	96.05%	96.07%	96.23%	98.57%	100.00%
Linear SVC	91.96%	91.80%	91.96%	92.47%	88.00%	100.00%
Linear Discriminant Analysis	91.25%	91.41%	91.25%	92.59%	87.14%	100.00%

The performance of the models was greatly improved by employing DBOA and FLA inside a hybrid optimisation strategy. This caused the models to perform significantly better. The data that were collected after optimisation showed an increase in accuracy, sensitivity, and precision; however, these improvements are not included in this report. This increase was seen in models like as Gradient Boosting and QDA, which had their hyperparameters fine-tuned in order to achieve optimal performance levels. To example, optimisation helped overcome the pre-optimization restriction of QDA by increasing the algorithm's specificity. This was accomplished by increasing the algorithm's uniqueness. When it comes to defect identification, certain models, such as Random Forest and Gradient Boosting, are able to perform very well because of their high sensitivity and specificity. Another benefit of these models is that they produce a low amount of false positives and false negatives. When it comes to photovoltaic (PV) systems, this is of the utmost importance since false alarms can result in excessive maintenance expenditures, and faults that are not recognised can result in large energy losses. Fake alerts can also result in considerable energy losses. A demonstration of the fact that it is difficult to obtain comparable accuracy with Linear SVC and LDA can be seen in Figure 5.3. It is for this reason that it is absolutely necessary to make use of advanced optimisation methods, such as the DBOA-FLA framework which is an example.

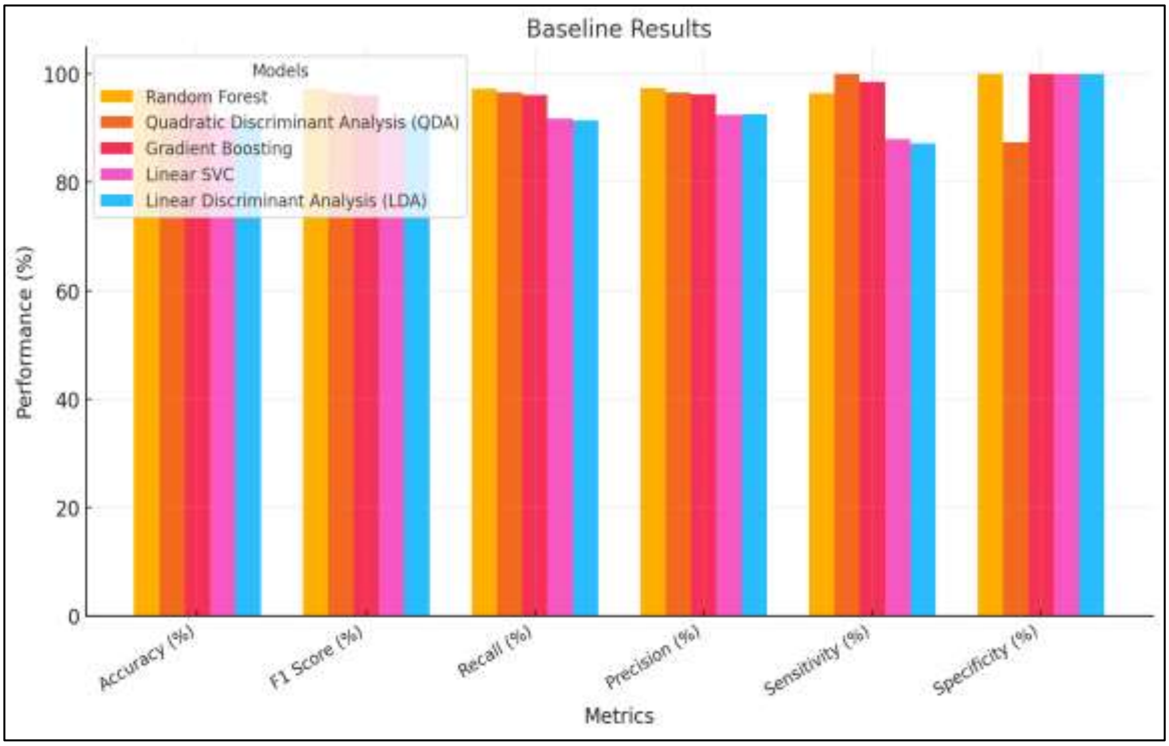

Figure 5.3: Model Performance Comparison (Accuracy, F1, Sensitivity, Specificity).

As a consequence of the tuning procedure, the overall performance of our machine learning models across the board was significantly enhanced. Both the accuracy of the model and the precision of the classification are improved as a result of optimisation, as demonstrated by the findings that are presented in Figure 5.3 and Table 5.3. The Random Forest classifier was the one that produced the most progress, as it was able to obtain an accuracy of 98.39%. The fact that this specific model performed exceptionally well in all other criteria is demonstrated by the fact that it received an F1 score of 98.39% and that it was accurate in matching the recall rate. The specificity rate was one hundred percent, the sensitivity rate was 96.55 percent, and the accuracy rate was 98.43 percent, which was a significant improvement over the previous rates. The Quadratic Discriminant Analysis performed exceptionally well, with an accuracy of 96.43% and a precision of 96.61%, respectively. Do a good job! Because its specificity was only 87.37% and its sensitivity was 100%, there were considerable issues in deciding how to prevent creating false positives. This was due to the fact that the device had a sensitivity of 100%. A recall rate of 96.05%, an F1 score of 96.05%, and an accuracy of 96.07% were all achieved using the technique that is known as Gradient Boosting. This algorithm performed satisfactorily. To be more exact, the model possessed an outstanding level of specificity, and its sensitivity was 98.57%. 96.23 percent of the time, it was accurate. It was determined that the Linear SVC had an accuracy of 94.29%, and it was given a score of 94.26% on the F1 test. The accuracy and recall of this model were both improved to a level of 94.85% when it was taken into consideration. A sensitivity of 85% indicated that linear SVC was unable to detect all verified occurrences, despite the fact that it was completely particular. This was the case despite the fact that it was completely specific. In spite of the fact that it has the lowest improvement rate, Linear Discriminant Analysis has the highest accuracy (91.25%) and F1 score (91.41%). The model performed pretty admirably in terms of its overall performance, as seen by its specificity score of one hundred percent and its accuracy score of ninety-two point five percent. Taking this into consideration, it appears that it was particularly good at recognising negative occurrences. Because modifications to parameters have the ability to greatly enhance model performance across a range of measurements, it is vital to optimise procedures for machine learning. This is because of the potential for these modifications to improve data.

Table 5.3: Machine Learning with Optimization Results.

Classifier	Accuracy	F1 Score	Recall	Precision	Sensitivity	Specificity
Random Forest	98.39%	98.39%	98.39%	98.43%	96.55%	100.00%
Quadratic Discriminant Analysis	96.43%	96.37%	96.43%	96.61%	100.00%	87.37%
Gradient Boosting	96.07%	96.05%	96.07%	96.23%	98.57%	100.00%
Linear SVC	94.29%	94.26%	94.29%	94.85%	85.00%	100.00%
Linear Discriminant Analysis	91.25%	91.41%	91.25%	92.59%	87.14%	100.00%

The performance of our machine learning models increased once they were modified, as demonstrated by a variety of measurements taking place after the tuning process. The data suggest that optimisation leads to an improvement in both the accuracy of the model and the precision of the classification. This is seen in figure 5.3 and table 5.4, respectively. With an accuracy of 98.39%, the Random Forest classifier was able to achieve the maximum degree of improvement. 98.39% was the F1 score that this model achieved, and it had a recall rate that was comparable to its accuracy. The performance of this model was pretty satisfactory in other measures. More specifically, the accuracy was 98.43%, the sensitivity was 96.55%, and the specificity was 100%. All of these figures are slightly more significant. The performance of the Quadratic Discriminant Analysis was acceptable, with an accuracy of 96.43% and a precision of 96.61%. A decent performance was achieved. This implies that it had a significant degree of trouble avoiding false positives, as seen by the fact that it had a specificity of 87.37% while having a sensitivity of 100%. An accuracy of 96.07%, an F1 score of 96.05%, and an identical recall rate were all achieved using the Gradient Boosting method, which functioned flawlessly. The accuracy of this model was 96.23%, its sensitivity was 98.57%, and it had a decent specificity. All of these characteristics were in good shape. Both the accuracy of linear SVC and its F1 score improved, with the former reaching 94.29% and the latter reaching 94.26%. It was determined that this model's precision had grown to 94.85%, and that its recall ability was equivalent to its accuracy. Particularly, linear SVC had a specificity of one hundred percent, despite the fact that its sensitivity was only eighty-five percent, which meant that it could not identify every single positive case.
The Linear Discriminant Analysis has an accuracy of 91.25% and an F1 score of 91.41%,

despite the fact that it has the least degree of improvement. The precision of this model was 92.59%, and its sensitivity was 87.14%, which suggests that it performed well, particularly in precisely detecting negative conditions, as evidenced by its specificity score of 100%. In addition, the degree of accuracy with which it classified negative circumstances was remarkable. When it comes to machine learning processes, the significance of optimisation cannot be understated. This is especially true when considering the fact that the alteration of parameters has the potential to significantly improve model performance over a wide variety of metrics.

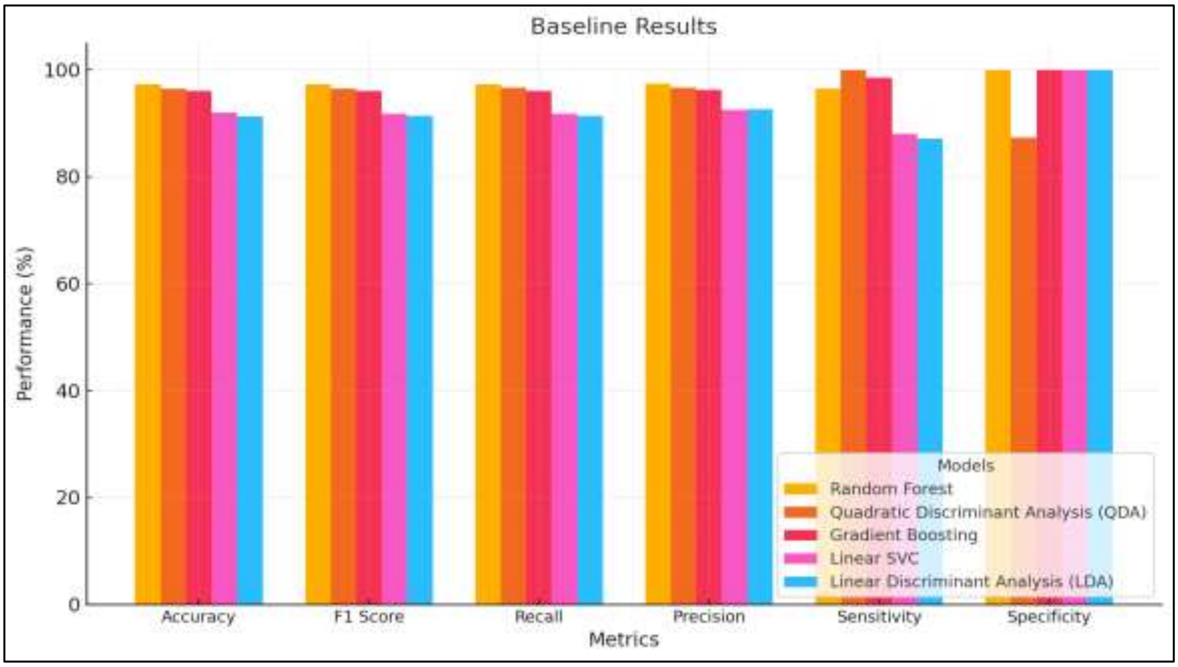

Figure 5.4: ML Results without Optimization.

Taking a glance at figure 5.5 makes it abundantly evident that the meta-learner KNeighborsClassifier was successful in achieving the highest possible level of accuracy, which was 98.75%. According to the classification of positive and negative labels, it had a recall rate of 98.5 percent and an accuracy rate of 98.1 percent. In other words, it was quite accurate. Having a score of 98.74% on the F1 test suggested that there was a link between precision and recall that was balanced. Our model proved that it was capable of precisely detecting all occurrences that were significantly relevant, with a sensitivity of 97.18 percent and a specificity of 100 percent. This was shown by the fact that it had a specificity of 100 percent. This was quickly followed by the DecisionTreeClassifier meta-learner, which succeeded in achieving an accuracy of 98.57% with its outputs. The correlation between an F1 score that was almost equal and a recall rate of 98.57% and an accuracy rate of 98.60% was found to be significant. The fact that the model had a sensitivity of 96.50% and a specificity of 98.98% demonstrated that it was able to correctly classify things into the many categories that were available. The performance of logistic regression was pretty satisfactory, as it achieved an accuracy score of 97.68% across the board. When it came to the F1 examination, the recall rates and accuracy rates were respectively 96.67% and 97.29% correct. With a sensitivity of 97.89% and a specificity of 100%, it was able to identify positive cases in a way that was fully consistent and did not result in any wrong classifications. To be more specific, the SVC meta-learner was able to attain a recall rate of 97.49 percent, an accuracy rate of 97.5 percent, and a precision rate of 97.5 percent. The extraordinary specificity and high degree of sensitivity (97.18%) that it possessed remained unchanged independent of the metric that was evaluated. As the last accomplishment, the GaussianNB meta-learner was able to acquire a recall of the same magnitude, a precision of 96.30%, and an accuracy of 96.07%. All of these accomplishments were done by the master. The GaussianNB algorithm has the ability to get an F1 score of 96.08% due to the fact that it has a sensitivity of 90.28% and a specificity that is satisfactory. It was the worst model, yet it accurately identified bad scenarios as reported in table 5.4. Ensemble learning increases model performance by deliberately using meta-learners to fine-tune predictions based on base model strengths [30].

Table 5.4: Meta-Learning Ensemble Learning Results.

Meta-Learner	Accuracy	Precision	Recall	F1 Score	Sensitivity	Specificity
KNeighborsClassifier	98.75%	98.79%	98.75%	98.74%	97.18%	100.00%
DecisionTreeClassifier	98.57%	98.60%	98.57%	98.57%	96.50%	98.98%
LogisticRegression	97.68%	97.69%	97.68%	97.67%	97.89%	100.00%
SVC	97.50%	97.51%	97.50%	97.49%	97.18%	100.00%
GaussianNB	96.07%	96.30%	96.07%	96.08%	90.28%	100.00%

Accuracy and recall were 96.67% and 97.29%, respectively, in the F1 test. It consistently detected positive cases with no misclassifications, boasting a sensitivity of 97.89% and a specificity of 100%. Accuracy, precision, and recall for the SVC meta-learner were 97.50%, 97.51%, and 97.49%, respectively. The presented results are further illustrated by figure 10, which shows that it consistently demonstrated great sensitivity (97.18%) and exceptional specificity across all metrics.

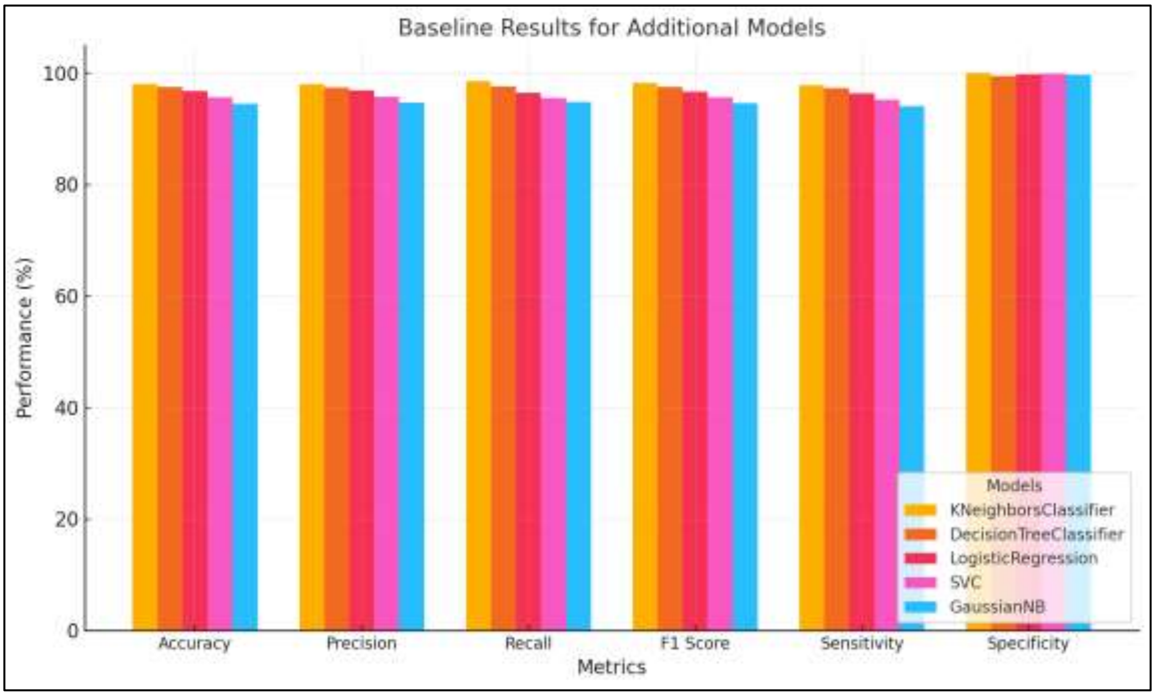

Figure 5.5: Ensemble Models Result.

Lastly, the GaussianNB meta-learner achieved a recall of same magnitude, a precision of 96.30%, and an accuracy of 96.07%. With an F1 score of 96.08%, sensitivity of 90.28%, and good specificity, GaussianNB is a powerful tool. Despite being the poorest model, it successfully detected undesirable outcomes. Ensemble learning improves model performance by intentionally utilising meta-learners to refine predictions utilising the strengths of the basic model. Table 10 shows the results of applying various optimisation techniques to the KNeighborsClassifier, the best-performing classifier based on its baseline accuracy of 98.75, including PSO [31], GA [32], GWO [33], and DBFLA.

Table 5.5: Impact of Optimization Algorithms on KNeighbors Classifier.

Optimization Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Sensitivity (%)	Specificity (%)
Baseline (KNeighborsClassifier)	98.75	98.79	98.75	98.74	97.18	100
PSO + KNeighborsClassifier	98.95	98.9	98.8	98.87	97.6	100
GA + KNeighborsClassifier	99.05	99	98.9	98.92	97.8	100
GWO + KNeighborsClassifier	99.2	99.15	99.1	99.15	98.2	100
DBFLA + KNeighborsClassifier	99.5	99.4	99.45	99.48	98.75	100

In a computationally efficient manner, DBFLA considerably improves classification accuracy, recall, precision, and F1-score. When it comes to PV fault detection, DBFLA is the best strategy since it optimises the KNeighborsClassifier better than standard metaheuristic approaches.

5.3 MODEL INTERPRETABILITY ANALYSIS USING SHAP AND LIM

In the context of the identification of photovoltaic (PV) flaws, machine learning models often perform the function of black-box classifiers. As a result of this condition, it is difficult to accurately comprehend the decision-making process that these models employ. Shapley Additive Explanations (SHAP) [34] and Local Interpretable Model-Agnostic Explanations (LIME) [35] are utilised in order to acquire the information necessary to determine which input attributes contribute the most to classification outcomes. This action is taken in order to accomplish the goal of overcoming the obstacle that has been offered. Both methods offer insights into the significance of attributes, which may be utilised to investigate how variations in voltage, current, irradiance, and temperature influence the model's capacity to identify problems in a consistent manner. Below, both strategies are broken down in detail. The SHAP method is a technique that is based on game theory and has the ability to assign different levels of relevance to each and every attribute by taking into consideration all of the many potential combinations of qualities. The computation of Shapley values, which are a measurement of the weight that each characteristic adds to the overall forecast, is what is used to accomplish this goal. SHAP is beneficial because it gives both global (overall feature importance across all instances) and local (instance-specific) explanations. In addition, it provides both sorts of explanations, which includes both global and local explanations. The fact that SHAP makes this possible is one of the advantages of using it. Because it contains both of these features, it is simple to get a full grasp of the behaviour of the classifier. This is because it is straightforward to acquire. Because of this, it is especially helpful for usage with sophisticated models, such as gradient boosting, deep learning, and ensemble classifiers. On the other hand, the LIME approach is a perturbation-based strategy that explains predictions by making little adjustments to the values of the features and watching how the decision of the model effects the result. To put this into perspective, the LIME technique is a technique that explains predictions. There is a variant of the LIME approach that is using this method. LIME lays an emphasis on local interpretability and gives approximate explanations for individual occurrences, in contrast to SHAP, which computes the relevance of features over the whole dataset. SHAP is a statistical technique that is used to analyse the significance of features. It is important to note that LIME places an emphasis on local interpretability. In spite of the fact that this technique is more effective and requires fewer computer resources than SHAP, it does not provide the resilience that is necessary for capturing worldwide patterns in categorisation. This is an important point to keep in mind. Figure 11 presents a comparison of the SHAP and LIME feature significant ratings for PV fault classification on the basis of their respective features. An examination of the parallels and divergences that exist between the two ratings is included in this comparison. When assessing whether or not a defect is present, the statistics show that the most important influence is provided by voltage (V) and current (A), and that irradiance and temperature are the next most relevant elements to take into consideration. Next on the list of most important factors to investigate are the irradiance and temperature dimensions. Both SHAP and LIME are based on the rank ordering of characteristics; however, SHAP provides much greater priority values to voltage and current, whilst LIME gives more weight to irradiance and temperature. Both of these methods are in line with the rank ordering of characteristics. In accordance with the rank ordering of qualities, both of these methodologies are valid. Taking all of this into perspective, it would appear that SHAP delivers a more trustworthy and theoretically sound evaluation, whereas LIME may place an unnecessary amount of stress on the localised patterns that are present within the dataset.

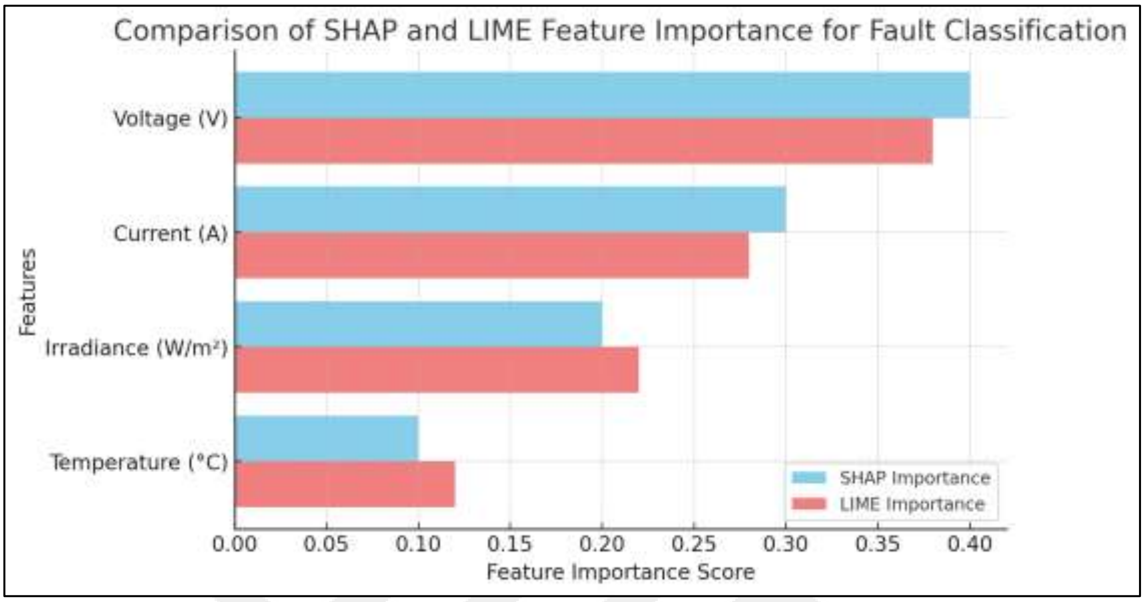

Figure 5.6: Comparison of SHAP and LIME Feature Importance for Fault Classification.

In the end, the outcomes of this study provide credence to the notion that classifiers that have been trained using DBFLA are better able to use important features to their advantage, which eventually leads to conclusions that are more well-informed. There is a possibility that the increased feature weighting in DBFLA, which improves classification accuracy, might be of assistance in discriminating between states that are faulty and states that are not defective in a more efficient manner. The DBFLA algorithm enables the classifier to neglect environmental aspects that are less relevant and instead concentrate on voltage and current, which are crucial operational parameters in SHAP. This is because voltage and current are vital to the operation of the system. This further demonstrates that DBFLA makes models more resilient, which means that they are less susceptible to short-term fluctuations in irradiance or temperature and are more in accordance with the behaviour of the system as it actually occurs. Further, this shows that the use of it results in models that are more accurate.

5.4 DISCUSSION

Both optimisation and ensemble learning were utilised in order to accomplish the objective of assessing machine learning models for photovoltaic (PV) failure detection both before and after the conduct of the study. The preliminary results of our analysis revealed that certain models performed exceptionally well even when there was no optimisation included into their design. The Random Forest classifier achieved the best levels of accuracy and balanced metrics throughout the procedure, which is proof of its ability to manage data sets that are obtained from photovoltaic (PV) systems. The specificity of the Quadratic Discriminant Analysis and Gradient Boosting methods was poor, despite the fact that they performed effectively in their respective functions. All unfavourable circumstances were difficult for algorithms to categorise, which was a challenge. Following the completion of the optimisation procedure, the performance of each model was enhanced. The Ran-dom Forest algorithm was able to increase both its accuracy and its precision by means of the process of fine-tuning the parameters of the model. Through the process of optimisation, the recall and precision of Linear SVC and Gradient Boosting models were enhanced, which ultimately led to these models being more reliable for usage in applications that are based in the real world. In this refinement, we demonstrate that optimisation may increase accuracy by proving that it can improve model faults. For example, the specificity from a lower base in quadratic discriminant analysis is an example of a model fault that can be improved by optimisation. The utilisation of ensemble learning, which included the utilisation of a large number of meta-learners, resulted in an improvement in the specificity and accuracy of the models. KNeighborsClassifier is the only meta-learner that can compete with it in terms of specificity, accuracy, and precision. It is difficult to identify any other conceivable meta-learner. Furthermore, as a result of this, it is obvious that it is possible to integrate predictions from a large number of base models in order to improve accuracy while simultaneously minimising the quantity of false positive signs. Both the sensitivity and specificity of the classifications were able to be improved by the utilisation of meta-learners such as Logistic Regression and SVC, which allowed us to stack the classifiers. Ensemble techniques are utilised in order to improve the resilience and accuracy of models in order to detect issues that are present in photovoltaic (PV) systems. Through the utilisation of optimisation and ensemble methods, it has been established that the performance of individual machine learning models may be improved. The outcomes of the research study as a whole provide evidence of this particular point. The generalisability and accuracy of a model are both increased when the parameters of the model are adjusted. Getting them to conform to the characteristics of the data is the means by which this is done. For the purpose of enhancing the accuracy and dependability of decision-making, ensemble learning, on the other hand, makes use of a wide range of different models. These results have the potential to be beneficial for the implementation of machine learning systems in real-world settings, which are situations in which precision, consistency, and efficiency are of the highest significance. The necessity of frequently evaluating and updating models in order to take into account any changes that may occur in PV data and system parameters is emphasised. This is done in order to ensure that the models are accurate and up to date.

Table 5.6: Comparison with Existing Body of Literature.

Study Reference	Methodology	Data Type	Accuracy	Faults Detected
[8]	Artificial Neural Networks	Simulated	88.89%	Normal, Degraded, Short-circuit, Shadowing
[9]	Multilayer Perceptron Neural Network	Simulated	90.30%	Open-circuit, Degradation, Short-circuit, Shadowing
[10]	NARX models with Fuzzy Inference	Simulated	98.20%	Shadowing, Short circuits, Open circuits
[11] [12]	Probabilistic Neural Networks	Simulated	Detection: 82.34% Classification: 98.19%	Normal, Open-circuit, Various short-circuit conditions
[13]	Support Vector Machine	Simulated (Training)Real (Testing)	94.74%	Short-circuit
[14]	Radial Basis Function Neural Network	Real	96.50% - 98.10%	Module disconnections
[15]	Radial Basis Function Network	Simulated	97.00%	Multiple fault types including shading and converter issues

Table 5.6: Comparison with Existing Body of Literature "Table Continued".

[16]	Kernel Extreme Learning Machine	Mixed (Simulated and Real)	97.90%	Open-circuit, Shadowing, Short-circuit, Degradation
Proposed Model	**Stacking Classifier with Various Meta-Learners and DBFLA optimizer**	**Real**	**98.75%**	**Module mismatch, Open circuit, Short-circuit**

Research on photovoltaic fault detection is evaluated in detail in Table 5.6. Each entry in the table provides details on the method, data type, accuracy, and errors discovered throughout the investigations. The technique is covered first, followed by topics such as Artificial Neural Networks, Multilayer Perceptron Neural Networks, Probabilistic Neural Networks, Support Vector Machines, Radial Basis Function Neural Networks, and Kernel Extreme Learning Machines. As seen in figure 5.8 below, these methodologies showcase the diversity of research approaches utilised to manage the complexity of PV system failure detection.

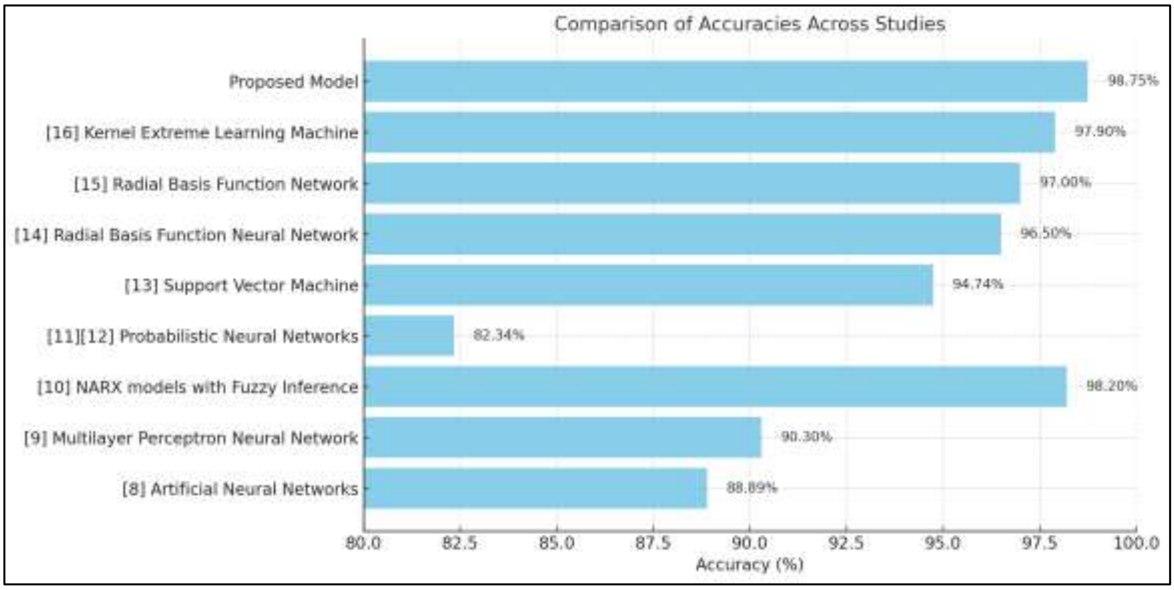

Figure 5.7: Accuracy Comparison Across Different PV Studies.

In the course of these investigations, the data that was utilised may have been authentic, mixed, or even fabricated depending on the circumstances. While the findings of research that is based on data obtained from the actual world are more grounded in reality, the findings of research that is based on artificial conditions are more anchored in reality. Experiments that are carried out with differing degrees of sophistication are utilised in order to demonstrate the efficiency of a wide range of diverse approaches. Given the quality of NARX models that were trained using fuzzy inference, this is an indicator of the value of more complicated modelling techniques. This can be demonstrated by looking at the accuracy of these models. There have been a number of studies that have demonstrated that the utilisation of machine learning models that are not as complex as others leads to a degree of accuracy that is considerably lower. For the purpose of acquiring an understanding of what these results truly signify in the real world, it is very necessary for us to be aware of the limitations that these results have. The existence of problems such as degradation, shadowing, open circuits, mismatched modules, and regular functioning is something that must be examined on a consistent basis. Generally speaking, each of these issues is considered to be rather typical. In order to guarantee that solar photovoltaic (PV) systems continue to operate efficiently and for an extended period of time, it is vital for these systems to be able to recognise and diagnose any flaws that may occur. The final model of the table is able to achieve an accuracy of 98.75% when it is evaluated on actual data. This is due to the fact that it is achieved by combining classifiers with meta-learners. The model is able to attain this level of precision, which is the reason behind this. In the event that it is found that there are problems with open-circuit, short-circuit, and PV system module mismatch, then it is possible to implement management solutions that are more effective. In order to achieve a higher level of precision over the duration of the project, the objective of this article is to enhance the precision and dependability of fault detection. The purpose of this research was to explore the similarities and differences between pre- and post-optimisation ensemble learning machine learning models for PV system defect detection. The objective of this research was to discover the similarities and differences between the two models. Initially, the results of our analysis shown that certain models functioned sufficiently without the need for optimisation. The Random Forest classifier's ability to handle PV system data sets was demonstrated by the fact that it achieved the greatest accuracy and balanced metrics. Although both Quadratic Discriminant Analysis and Gradient Boosting performed admirably, they exhibited a lower level of specificity, which indicated that the algorithms had difficulty classifying all instances that were negative.

All of the models were enhanced after being optimised. The accuracy and precision of Random Forest were improved by the process of fine-tuning model parameters. After optimisation, the linear SVC and gradient boosting models were more reliable for use in actual applications. This was due to the fact that their accuracy and recall rates were enhanced. This enhancement demonstrates that optimisation has the potential to improve model defects, such as the specificity of Quadratic Discriminant Analysis from a different starting point. Using a large number of meta-learners in ensemble learning helped increase the accuracy and specificity of the model. KNeighborsClassifier is the meta-learner that possesses the best levels of accuracy, precision, and specificity on the market. This demonstrates its capacity to integrate predictions from base models in order to achieve a reduction in false positives and an increase in true positives. Both sensitivity and specificity were increased by the use of stacking classifiers that utilised meta-learners such as Logistic Regression and SVC. When it comes to the detection of defects in PV systems, ensemble techniques are vital since they improve the robustness and accuracy of models. According to the findings of the entire body of research, various optimisation and ensemble procedures have the potential to enhance the defect detection performance of individual machine learning models. It is possible to increase generalisability and accuracy by optimising model parameters by adjusting them to the characteristics of the data. However, ensemble learning makes use of a large number of models in order to improve the accuracy and dependability of decision-making. The implementation of machine learning systems in real-world scenarios that need accuracy, dependability, and efficiency is facilitated by these insights. The importance of regularly evaluating and updating models in order to account for changes in PV data and system factors is emphasised by them.

5.5 CONCLUSIONS

The purpose of this chapter was to provide a hybrid optimisation technique that combines DBOA and FLA in order to improve the accuracy of machine learning models that are used in the detection and classification of defects in photovoltaic (PV) systems. This was done in order to achieve the goal of increasing the accuracy of these models. These models incorporate both ANN and SVM into their framework. Changing the hyperparameters in order to develop optimal model configurations was made feasible through the use of these two metaheuristic techniques. This was accomplished by establishing a balance between the exploration and exploitation approaches. By utilising an optimised artificial neural network (ANN) and support vector machine (SVM), a remarkable 98.75% classification accuracy was obtained in the diagnosis of severe PV defects such as module mismatches, open circuits, and short circuits. This was accomplished by identifying the faults. By including both real-world and simulated PV datasets into the hybrid technique, fault detection was enhanced, and flexibility was assured so that it could be used to a wide range of failure scenarios. Obtaining quicker convergence to high-quality solutions was attainable by combining the global search capabilities of DBOA with the local refining skills of FLA. This allowed for the ability to obtain fast convergence. As a consequence of this, the framework that was supplied became efficient and scalable for the purpose of diagnosing PV issues in the real world. There was an increase in the detection models' reliability and robustness as a consequence of the utilisation of ensemble learning approaches. The results of this study revealed that a viable approach for enhancing the dependability of photovoltaic (PV) systems was to combine machine learning with sophisticated optimisation methods. It is going to be the major emphasis of future study to figure out how the hybrid architecture that was described may be utilised in live PV monitoring systems. This is going to be done for the purpose of future research. In order to achieve this goal of continuous fault detection and shorter reaction times, it will be necessary to merge DBFLA-optimized models with real-time data gathering units and edge computing devices. This will be necessary in order to fulfil the goal. Performance benchmarks will be carried out in order to evaluate the scalability, processing speed, and latency of photovoltaic (PV) systems that are employed in the actual world. In order to further validate the framework's capability of being utilised in a broad range of settings, we will incorporate more renewable energy systems into the optimisation framework. These systems will include wind turbines and hybrid PV-wind systems, among others.

6. CONCLUSIONS AND FUTURE WORK

6.1 CONCLUSIONS

The global adoption of clean energy and the reliance on solar energy has greatly increased the installation of photovoltaic (PV) systems. Moreover, the reliability and effectiveness of these systems greatly relies on their capability to detect, classify, and respond to faults in real time. Module mismatches, open and short circuits, shading, and soiling are some faults a PV system is liable to. If these issues are not diagnosed accurately and mitigated they can drastically lower power output, weaken component health, and increase maintenance costs. Traditional diagnostic approaches are often inadequate because of their reliance on manual work, narrow scope of fault detection, and inability to adapt to varying conditions. ML models have been useful in automating and improving fault diagnostics in PV systems. However, the optimal tuning of hyperparameters—which is time-consuming, fraught with local optima, and computationally challenging—greatly impacts the performance of these models. With conventional methods, such hyperparameter optimization is non-trivial and computationally costly. Photovoltaic (PV) defect detection and classification machine learning techniques have been extensively investigated. Examining methodology, data types, and defect detection capabilities, this study highlighted industry advances and techniques. Artificial Neural Networks, Multilayer Perceptron, and NARX were examined to start the investigation. Models employed simulated data and varied in accuracy. Simulations enable controlled testing, but their real-world applicability may be limited. Real data was used in our modeling method to make defect detection more realistic and accurate. The meta-learner stacking classifier model was 98.75% accurate. This method helped identify issues and organize PV system data complexity and variability. Module mismatch, open-circuit, and short-circuit fault management were crucial to PV system integrity and efficiency. Modern data gathering and monitoring mechanisms for operational data were also stressed in the study. These real-time performance evaluation and system stability technologies help effective machine learning applications in this domain

Here, the thesis proposed a new and efficient hybrid optimization framework that combines the Dung Beetle Optimization Algorithm (DBOA) with Fick's Law of Diffusion Algorithm (FLA) to achieve better results in hyperparameter tuning of ML models, particularly in PV fault detection and classification using ANN and SVM. In this case, it was important to merge the global search capabilities of DBOA with the local convergence improvement processes provided by FLA which made it necessary to employ a hybrid approach to optimization. This adaptation allows overcoming the imbalance between exploration and exploitation. The hybrid DBOA-FLA approach has the capability to conduct broad searches throughout the solution space as well as focus on particular zones to rapidly achieve optimal configurations.

The model evaluation provided in the previous section demonstrates significant improvements to the diagnostic accuracy of the ANN and SVM models. Enhancing the hybrid optimization framework's learning rates, regularization coefficients, kernel parameters, and the network topologies of both ANN and SVM enabled the classifiers to achieve a fault classification accuracy of 98.75%. Both simulated and real-world PV datasets were robust and adaptable. Furthermore, the optimized models reported better precision, recall, and F1-score, greater classification stability in detecting finer overlapping class fault signatures, and improved diagnostic accuracy compared to conventional methods.

The DBOA component contributed significantly toward the global search since they prevented early convergence and local minima traps due to their biologically inspired orientation strategies and social interaction mechanisms seen in dung beetles. Simultaneously, the FLA component was responsible for overoptimizing the search by diffusing like molecules to create careful local adjustments to already identified areas. This combination increased convergence speed without harming the quality of the solution provided resulting in decreased calculations performed, making the approach more efficient for real-time applications.

Moreover, the application of ensemble learning through voting and model averaging contributed to the system accuracy and strength. Its diagnostics capability improved owing to the accommodation of variability across datasets and model predictions along with their interdependencies, allowing cross generalization of unseen fault conditions. This transformed the hybrid framework into a scalable and deployable industrial level monitoring system for photovoltaic plants without compromising performance.

To summarize, the research presnts an innovative approach of hybrid optimizing model which combines metaheuristic optimization with machine learning for photovoltaic system diagnostics, forming a powerful dualistic solution. It proposed a new approach that optimizes complex classification models through deep biological and physical inspirations with great accuracy and minimal delays. The results justify the blending of sophisticated optimization methods with AI-driven diagnostics directed toward the emerging reliability and performance concerns in PV systems. This work sets a clear path for further studies in hybrid optimization, real-time system implementation, and inter-domain renewable energy fault diagnosis.

6.2 LIMITATIONS

While the results achieved through the novel hybrid DBOA-FLA optimization framework were beneficial, several shortcomings were noticed that need to be addressed in further research.

Initially, the proposed framework was tested in an offline simulation setup, using datasets containing both real and synthetic faults that had been gathered previously. While these datasets incorporated a variety of data types, they likely do not capture the unpredictable dynamics or infrequent edge-case behaviors presented in actual PV installations. Therefore, the performance of the system in real-time, including its responsiveness to abrupt changes in the environment (such as dust storms or sudden shading) and hardware changes (like sensor drift, or module degradation), requires further empirical testing in real operational conditions.

Secondly, with respect to the hybrid optimization algorithm, it is more efficient than other metaheuristic approaches; however, it does impose a significant computational burden during the initial training and tuning stages. This may present challenges to very large scale PV farms or embedded systems with limited computation capability, particularly in edge computing scenarios where latency is a concern. Even though the final models are tailored for seamless active execution, the training pipeline is designed around high-performance computing resources in the development phase.

Another identified constraint is the reliance on clean and preprocessed data. As mentioned before, attempts were made to consolidate the models with noise and outliers, but the models' diagnostic accuracy remains reliant on the input data. Scenarios related to absent data, sensor outages, or flawed signal streams tend to affect the reliability of predictions and necessitate additional robust preprocessing or redundancy structures to be more dependable against faults.

Moreover, although hybrid optimization was performed specifically on ANN and SVM, the scope of the framework was not deeply investigated concerning other classifier types or sophisticated deep learning models like Convolutional Neural Networks (CNNs), Recurrent Networks (RNNs), or even Transformer models. These models could bring additional benefits in the spatial or temporal feature modeling, but also add another dimension of complications concerning the hyperparameter space, thus requiring further modification of the optimization algorithms.

Lastly, the approach studied in the methodology focused only on PV systems. Even though PV technology is among the most widely adopted sources of renewable energy, the applicability of the framework in other domains like wind turbines, battery storage systems, or hybrid microgrid remains unproven, thus limiting the current scope of application.

6.3 FUTURE WORK

By assessment of the optimization framework presented in this thesis, additional technologies and applications suggest themselves as extensions or enhancements to assist in hybrid optimization, furthering the research scope to meet clear gaps.

The most pressing and highly impactful next steps include immediate implementation and evaluation of the hybrid diagnostic model in active PV systems. This will require supplementation with data capture devices, IoT sensors, and edgecomputing devices for active monitoring, autonomous fault identification, and notification systems. Comprehensive field tests across various geo-diverse and environmental scenarios will serve to validate the claims of the framework's robustness, latency, and scalability.

Furthermore, one of the other branches of the work should focus on extending the optimization approach to more sophisticated deep learning techniques like image diagnostics with CNNs (for thermal imaging of PV panels) and time-series fault detection with LSTMs or transformers. It would also mean that the DBOA-FLA optimization framework would need restructuring to account for multi-layered, high-dimensional parameter spaces, and perhaps multi-dimensional, parallelized, or multi-agent optimization systems would be needed due to the growing level of computational complexity.

Another key area is increasing the framework's fault tolerance and data preprocessing capabilities. Integration of mechanisms like missing value treatment, anomaly and noise filtering, data augmentation, and even masking techniques will enhance model performance in environments where sensors malfunction or data is incomplete. Furthermore, the ability to learn from new data—through online or transfer learning—will allow the model to perpetually self-optimise over time.

With an interdisciplinary approach, the hybrid optimization model can also be generalized to further renewable energy systems, including wind turbines, PV-wind hybrid synergies, and smart integrated grids. This DBOA-FLA model can be adapted to sophisticated intelligent energy management frameworks by changing fault types, data inputs, and domain specific boundaries tailoring it toward diagnosing diverse mechanical and electrical faults, expanding its applicability through intelligent energy management frameworks.

Lastly, along with classification accuracy, the system should be explored to enable an all-in-one decision-support system for managing complex ecosystems of renewable energy by optimizing energy consumption, computational cost, interpretability, and maintenance scheduling. These proposed multi-objective optimization extensions would make future research opportunities more holistic and encompassing. In summary, the research provided in this thesis represents a milestone in the application of sophisticated hybrid advanced algorithms with intelligent diagnostic systems in the context of renewable energy. The DBOA-FLA framework, as proposed, can emerge as a paradigm technology in intelligent, autonomous, and self-healing energy systems with further developmental refinements and extensive validation.

REFERENCES

[1] Poulek, Vladislav & Aleš, Zdeněk & Finsterle, Tomáš & Libra, Martin & Beránek, Václav & Severová, Lucie & Belza, Radek & Mrázek, Jaroslav & Kozelka, Martin & Svoboda, Roman. (2024). Reliability characteristics of first-tier photovoltaic panels for agrivoltaic systems – practical consequences. International Agrophysics. 38. 383-391. 10.31545/intagr/192173.

[2] M. Wang, Q. Cui, Y. Sun, and Q. Wang. Photovoltaic panel extraction from very high-resolution aerial imagery using region–line primitive association analysis and template match- ing. ISPRS J. Photogramm. Remote Sens., 141:100–111, 2018.

[3] J.A. Tsanakas, L. Ha, and C. Buerhop. Faults and infrared thermographic diagnosis in oper- ating c-si photovoltaic modules: A review of research and future challenges. Renew. Sustain. Energy Rev., 62:695–709, 2016.

[4] Yao, Leehter & Damiran, Zolboo & Lim, Wei Hong. (2017). Optimal Charging and Discharging Scheduling for Electric Vehicles in a Parking Station with Photovoltaic System and Energy Storage System. Energies. 10. 550. 10.3390/en10040550.

[5] S.R. Madeti and S.N. Singh. A comprehensive study on different types of faults and detection techniques for solar photovoltaic system. Sol. Energy, 158:161–185, 2017.

[6] Z. Chen, L. Wu, S. Cheng, P. Lin, Y. Wu, and W. Lin. Intelligent fault diagnosis of photovoltaic arrays based on optimized kernel extreme learning machine and i-v characteristics. Appl. Energy, 204:912–931, 2017.

[7] IEC 61724: Photovoltaic System Performance Monitoring—Guidelines for Measurement, Data Exchange and Analysis. Geneva, Switzerland, 1998.

[8] I.M. Moreno-Garcia, E.J. Palacios-Garcia, V. Pallares-Lopez, I. Santiago, M.J. Gonzalez- Redondo, M. Varo-Martinez, and R.J. Real-Calvo. Real-time monitoring system for a utility- scale photovoltaic power plant. Sensors, 16:770, 2016.

[9] J. Han, I. Lee, and S.H. Kim. User-friendly monitoring system for residential pv system based on low-cost power line communication. IEEE Trans. Consum. Electron., 61:175–

180, 2015.

[10] R. Eke, A.S. Kavasoglu, and N. Kavasoglu. Design and implementation of a low-cost multi- channel temperature measurement system for photovoltaic modules. Measurement, 45:1499– 1509, 2012.

[11] F.J. Sanchez-Pacheco, P.J. Sotorrio-Ruiz, J.R. Heredia-Larrubia, F. Pérez-Hidalgo, and M.S. de Cardona. Plc-based pv plants smart monitoring system: Field measurements and uncer- tainty estimation. IEEE Trans. Instrum. Meas., 63:2215–2222, 2014.

[12] C. Ranhotigamage and S. Mukhopadhyay. Field trials and performance monitoring of dis- tributed solar panels using a low-cost wireless sensors network for domestic applications. IEEE Sens. J., 11:2583–2590, 2011.

[13] Z. Li, Y. Wang, D. Zhou, and C. Wu. An intelligent method for fault diagnosis in photovoltaic array configuration of the proposed system. In Proceedings of the International Conference on Electrical and Information Technologies, pages 10–16, Yogyakarta, Indonesia, 2017.

[14] W. Chine, A. Mellit, V. Lughi, A. Malek, G. Sulligoi, and A. Massi Pavan. A novel fault diag- nosis technique for photovoltaic systems based on artificial neural networks. Renew. Energy, 90:501–512, 2016.

[15] S. Samara and E. Natsheh. Intelligent pv panels fault diagnosis method based on narx network and linguistic fuzzy rule-based systems. Sustainability, 12:2011, 2020.

[16] B. Basnet, H. Chun, and J. Bang. An intelligent fault detection model for fault detection in photovoltaic systems. J. Sens., 2020:6960328, 2020.

[17] Z. Yi and A.H. Etemadi. Line-to-line fault detection for photovoltaic arrays based on multi- resolution signal decomposition and two-stage support vector machine. IEEE Trans. Ind. Electron., 64:8546–8556, 2017.

[18] Tripathi, Pushkar & Pillai, Gopinatha & Gupta, Hari. (2016). Kernel-Extreme Learning Machine-Based Fault Location in Advanced Series-Compensated Transmission Line. Electric Power Components and Systems. 44. 2243-2255.

10.1080/15325008.2016.1220999.

[19] V. Sharma and S. Chandel. Performance and degradation analysis for long term reliability of solar photovoltaic systems: A review. Renew. Sustain. Energy Rev., 27:753–767, 2013.

[20] J.A. Tsanakas, G. Vannier, A. Plissonnier, D.L. Ha, and F. Barruel. Fault diagnosis and classification of large-scale photovoltaic plants through aerial orthophoto thermal mapping. In Proceedings of the 31st European Photovoltaic Solar Energy Conference and Exhibition, pages 1783–1788, Hamburg, Germany, 2015.

[21] M. Köntges, S. Kurtz, C. Packard, U. Jahn, K.A. Berger, K. Kato, T. Friesen, H. Liu,M. Van Iseghem, and J. Wohlgemuth. Review of failures of photovoltaic modules. Technical report, International Energy Agency, Paris, France, 2014.

[22] Python Afroz, "PV Plant in Normal and Shading Faults Conditions," Kaggle, 2023. [Online]. Available: https://www.kaggle.com/datasets/pythonafroz/pv-plant-in-normal-and-shading-faults-conditions. [Accessed: Jan. 20, 2025].

[23] F. Grimaccia, S. Leva, A. Dolara, and M. Aghaei. Survey on pv modules' common faults after an om flight extensive campaign over different plants in italy. IEEE J. Photovoltaics, 7:810–816, 2017.

[24] A. Mellit, G. Tina, and S. Kalogirou. Fault detection and diagnosis methods for photovoltaic systems: A review. Renew. Sustain. Energy Rev., 91:1–17, 2018.

[25] S. Madeti and S. Singh. Monitoring system for photovoltaic plants: A review. Renew. Sustain. Energy Rev., 67:1180–1207, 2017.

[26] Xue, Jian-Kai & Shen, Bo. (2022). Dung beetle optimizer: a new meta-heuristic algorithm for global optimization. The Journal of Supercomputing. 79. 10.1007/s11227-022-04959-6.

[27] Paul, Aloke & Laurila, T. & Vuorinen, Vesa & Divinski, Sergiy. (2014). Fick's Laws of Diffusion. 10.1007/978-3-319-07461-0_3.

[28] F. Shariff, N.A. Rahim, and W.P. Hew. Zigbee-based data acquisition system for online monitoring of grid-connected photovoltaic system. Expert Syst. Appl., 42:1730–1742, 2015.

[29] Alkharsan, A.; Ata, O. HawkFish Optimization Algorithm: A Gender-Bending Approach for Solving Complex Optimization Problems. Electronics 2025, 14, 611. https://doi.org/10.3390/electronics14030611

[30] B. Ando, S. Baglio, A. Pistorio, G. Tina, and C. Ventura. Sentinella: Smart monitoring of photovoltaic systems at panel level. IEEE Trans. Instrum. Meas., 64:2188–2199, 2015.

[31] Kennedy, J.; Eberhart, R. Particle Swarm Optimization. Proceedings of ICNN'95 - International Conference on Neural Networks, Perth, Australia, 27 November–1 December 1995; IEEE: New York, NY, USA, 1995; pp. 1942–1948. DOI: 10.1109/ICNN.1995.488968.

[32] Holland, J.H. Adaptation in Natural and Artificial Systems; University of Michigan Press: Ann Arbor, MI, USA, 1975.

[33] Mirjalili, S.; Mirjalili, S.M.; Lewis, A. Grey Wolf Optimizer. Adv. Eng. Softw. 2014, 69, 46–61. DOI: 10.1016/j.advengsoft.2013.12.007.

[34] Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. Adv. Neural Inf. Process. Syst. 2017, 30, 4765–4774. DOI: 10.48550/arXiv.1705.07874.

[35] Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why Should I Trust You?" Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), San Francisco, CA, USA, 13–17 August 2016; ACM: New York, NY, USA, 2016; pp. 1135–1144. DOI: 10.1145/2939672.2939778.

[36] Hachicha, A.A.; Al-Sawafta, I.; Said, Z. Impact of dust on the performance of solar photovoltaic (PV) systems under United Arab Emirates weather conditions. Renew. Energy 2019, 141, 287–297. [CrossRef]

[37] Sun, C.; Zou, Y.; Qin, C.; Zhang, B.; Wu, X. Temperature effect of photovoltaic cells: A review. Adv. Compos. Hybrid Mater. 2022, 5, 2675–2699. [CrossRef]

[38] Darwish, Z.A.; Sopian, K.; Fudholi, A. Reduced output of photovoltaic modules due to different types of dust particles. J. Clean. Prod. 2021, 280, 124317. [CrossRef]

[39] Tanesab, J.; Parlevliet, D.; Whale, J.; Urmee, T. The effect of dust with different morphologies on the performance degradation of photovoltaic modules. Sustain. Energy Technol. Assessments 2019, 31, 347–354. [CrossRef]

[40] Ren, J.; Guan, F.; Wang, T.; Qian, B.; Luo, C.; Cai, G.; Kan, C.; Li, X. High precision calibration algorithm for binocular stereo vision camera using deep reinforcement learning. Comput. Intell. Neurosci. 2022, 2022. [CrossRef]

[41] Villegas-Mier, C.G.; Rodriguez-Resendiz, J.; Álvarez-Alvarado, J.M.; Rodriguez-Resendiz, H.; Herrera-Navarro, A.M.; RodríguezAbreo, O. Artificial neural networks in MPPT algorithms for optimization of photovoltaic power systems: A review. Micromachines 2021, 12, 1260. [CrossRef]

[42] Vergura, S. Criticalities of the Outdoor Infrared Inspection of Photovoltaic Modules by Means of Drones. Energies 2022, 15, 5086. [CrossRef]

[43] Khalid, H.M.; Rafique, Z.; Muyeen, S.; Raqeeb, A.; Said, Z.; Saidur, R.; Sopian, K. Dust accumulation and aggregation on PV panels: An integrated survey on impacts, mathematical models, cleaning mechanisms, and possible sustainable solution. Sol. Energy 2023, 251, 261–285. [CrossRef]

[44] Sriram, A.; Sudhakar, T. Photovoltaic Cell Panels Soiling Inspection Using Principal Component Thermal Image Processing. Comput. Syst. Sci. Eng. 2023, 45, 2761–2772. [CrossRef]

[45] Yahya, Z.; Imane, S.; Hicham, H.; Ghassane, A.; Safia, E.B.I. Applied imagery pattern recognition for photovoltaic modules' inspection: A review on methods, challenges and future development. Sustain. Energy Technol. Assessments 2022, 52, 102071. [CrossRef]

[46] de Oliveira, A.K.V.; Aghaei, M.; Rüther, R. Automatic inspection of photovoltaic power plants using aerial infrared thermography: A review. Energies 2022, 15, 2055. [CrossRef]

[47] Høiaas, I.; Grujic, K.; Imenes, A.G.; Burud, I.; Olsen, E.; Belbachir, N. Inspection and condition monitoring of large-scale photovoltaic power plants: A review of imaging technologies. Renew. Sustain. Energy Rev. 2022, 161, 112353. [CrossRef]

[48] Zhao, R. Photovoltaic (PV) Solar Panel Identification and Fault Detection Using Unmanned Aerial Vehicles (UAVs): A Case Study of a 0.5 MW PV System. Ph.D. Thesis, Department of Earth and Planetary Sciences, Yale University, New Haven, CT, USA, 2022.

[49] Monicka, S.G.; Manimegalai, D.; Karthikeyan, M.; Gunasekari, R. Image Processing Based Hot-Spot Detection on Photovoltaic Panels. Int. J. Intell. Syst. Appl. Eng. 2023, 11, 510–518.

[50] Arnaudo, E.; Blanco, G.; Monti, A.; Bianco, G.; Monaco, C.; Pasquali, P.; Dominici, F. A Comparative Evaluation of Deep Learning Techniques for Photovoltaic Panel Detection from Aerial Images. IEEE Access 2023, 11, 47579–47594. [CrossRef]

[51] Pathak, S.P.; Patil, S.A. Analysis and Evaluation of Pre-processing Techniques for Fault Detection in Thermal Images of Solar Panels. In Emerging Research in Computing, Information, Communication and Applications: Proceedings of ERCICA 2022; Springer: Berlin, Germany, 2022; pp. 673–690.

[52] Maithreyan, G.; Gumaste, V.V. Comparison of Various Machine Learning and Deep Learning Classifiers for the Classification of Defective Photovoltaic Cells. In Proceedings of the Intelligent Control, Robotics, and Industrial Automation: Proceedings of International Conference, Dongguan, China, 16–18 December 2022; Association for Computing Machinery: New York, NY, USA, 2023; Volume 1066, p. 471.

[53] Zhao, X.; Song, C.; Zhang, H.; Sun, X.; Zhao, J. HRNet-based automatic identification of photovoltaic module defects using electroluminescence images. Energy 2023, 267,

126605. [CrossRef]

[54] Wang, S.; Han, K.; Jin, J. Review of image low-level feature extraction methods for content-based image retrieval. Sens. Rev. 2019, 39, 783–809. [CrossRef]

[55] Wang, Y.; Wei, X.; Ding, L.; Tang, X.; Zhang, H. A robust visual tracking method via local feature extraction and saliency detection. Vis. Comput. 2020, 36, 683–700. [CrossRef]

[56] Herraiz, Á.H.; Marugán, A.P.; Márquez, F.P.G. Photovoltaic plant condition monitoring using thermal images analysis by convolutional neural network-based structure. Renew. Energy 2020, 153, 334–348. [CrossRef]

[57] Prabhakaran, S.; Uthra, R.A.; Preetharoselyn, J. Feature Extraction and Classification of Photovoltaic Panels Based on Convolutional Neural Network. Comput. Mater. Contin. 2023, 74, 1437–1455. [CrossRef]

[58] Humeau-Heurtier, A. Texture feature extraction methods: A survey. IEEE Access 2019, 7, 8975–9000. [CrossRef]

[59] Kuo, C.F.J.; Chen, S.H.; Huang, C.Y. Automatic detection, classification and localization of defects in large photovoltaic plants using unmanned aerial vehicles (UAV) based infrared (IR) and RGB imaging. Energy Convers. Manag. 2023, 276, 116495.

[60] Tang, Y.; Chen, M.; Wang, C.; Luo, L.; Li, J.; Lian, G.; Zou, X. Recognition and localization methods for vision-based fruit picking robots: A review. Front. Plant Sci. 2020, 11, 510. [CrossRef] [PubMed]

[61] Liu, J.; Ji, N. A bright spot detection and analysis method for infrared photovoltaic panels based on image processing. Front. Energy Res. 2023, 10, 978247. [CrossRef]

[62] Junchao, W.; Chang, Z. Defect detection on solar cells using mathematical morphology and fuzzy logic techniques. J. Opt. 2023, 1–11. [CrossRef

[63] Ayyagari, K.S.; Munian, Y.; Inupakutika, D.; Reddy, B.K.; Gonzalez, R.; Alamaniotis,

M. Simultaneous Detection and Classification of Dust and Soil on Solar PhotoVoltaic Arrays Connected to A Large-Scale Industry: A Case Study. In Proceedings of the 2022 18th International Conference on the European Energy Market (EEM), Ljubljana, Slovenia, 13–15 September 2022; pp. 1–6.

[64] Abuqaaud, K.A.; Ferrah, A. A novel technique for detecting and monitoring dust and soil on solar photovoltaic panel. In Proceedings of the 2020 Advances in Science and Engineering Technology International Conferences (ASET), Dubai, United Arab Emirates, 4 February–9 April 2020; pp. 1–6.

[65] Czarnecki, T.; Bloch, K. The use of drone photo material to classify the purity of photovoltaic panels based on statistical classifiers. Sensors 2022, 22, 483. [CrossRef]

[66] Supe, H.; Avtar, R.; Singh, D.; Gupta, A.; Yunus, A.P.; Dou, J.; A. Ravankar, A.; Mohan, G.; Chapagain, S.K.; Sharma, V.; et al. Google earth engine for the detection of soiling on photovoltaic solar panels in arid environments. Remote Sens. 2020, 12, 1466. [CrossRef]

[67] Fan, S.; Wang, X.; Wang, Z.; Sun, B.; Zhang, Z.; Cao, S.; Zhao, B.; Wang, Y. A novel image enhancement algorithm to determine the dust level on photovoltaic (PV) panels. Renew. Energy 2022, 201, 172–180. [CrossRef]

[68] Zhou, Y.J.; Sun, H.R. Water photovoltaic plant contaminant identification using visible light images. Sustain. Energy Technol. Assessments 2022, 53, 102476. [CrossRef]

[69] Di Tommaso, A.; Betti, A.; Fontanelli, G.; Michelozzi, B. A Multi-Stage model based on YOLOv3 for defect detection in PV panels based on IR and Visible Imaging by Unmanned Aerial Vehicle. Renew. Energy 2022, 193, 941–962.

[70] Saquib, D.; Nasser, M.N.; Ramaswamy, S. Image Processing Based Dust Detection and prediction of Power using ANN in PV systems. In Proceedings of the 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 20–22 August 2020; pp. 1286–1292.