Proceedings of 14th IOE Graduate Conference
Peer Reviewed
Year: 2023 Month: December Volume: 14
ISSN: 2350-8914 (Online), 2350-8906 (Print)

Machine Learning Approach for Fault Detection and Diagnosis of PV Modules

Kabina Maharjan [a], Arbin Maharjan [b], Laxman Motra [c],

[a, c] *Department of Mechanical and Aerospace Engineering, Pulchowk Campus, IOE, Tribhuvan University, Nepal*
[b] *Department of Electrical Engineering, Pulchowk Campus, IOE, Tribhuvan University, Nepal*
[a] imkabinaa@gmail.com, [b] arbinmaharjan@ioe.edu.np, [c] laxman.motra@pcampus.edu.np

Abstract

Fault analysis in solar photovoltaic (PV) arrays is a crucial aspect that helps to increase PV system's efficiency, reliability, and safety and, if not detected, it not only compromises the system's generation but also accelerates system aging and also jeopardizes the functionality of the overall system. Currently, the solution involves manual monitoring by system operators, but this approach is time-consuming, prone to inaccuracies, and poses safety risks. Therefore, it is imperative to implement automatic fault detection and diagnosis methods to ensure the PV systems' safety and reliability. Existing techniques either lack the precision to provide detailed fault information or are overly complex. This research introduces a fault detection and diagnosis method in solar PV systems using a machine learning approach. The research extends to defining normal conditions and four distinct fault categories for the proposed fault detection and classification algorithm. A forecasting model is prepared using a machine learning approach to forecast the DC output power. The Multilayer Neural Network (MNN) model is found to have the Root Mean Square Error (RMSE) of 6.74 and 6.11 for the training and validation sets. By analyzing the difference between the power predicted by the MNN model and the actual PV system power, the predefined fault types in the PV modules are detected. The proposed approach offers rapid detection and high accuracy. Simulation results demonstrate the effectiveness of this method in identifying and diagnosing open-circuit string, module, Maximum Power Point Tracking (MPPT), and partial shading faults in PV systems.

Keywords

Photovoltaic system, Fault detection and diagnosis, Machine learning approach, Maximum Power Point Tracking, Root Mean Square Error

1. Introduction

Fossil fuels have numerous harmful environmental effects, including climate change, greenhouse gas emissions, global warming, and air and water pollution. Experts predict that fossil fuel reserves will be depleted by 2030. To promote sustainability, researchers are exploring eco-friendly, cost-effective, and practical energy production methods. Renewable energy, particularly solar photovoltaic (PV) systems, is emerging as a crucial solution for addressing climate change and achieving sustainable development. PV systems are popular due to their environmental benefits and the ability to harness clean, endless solar energy. Many countries are adopting strategic energy plans to facilitate the installation of large-scale PV systems, aiming to reduce reliance on traditional fossil fuels, lower electricity costs, and leverage the unique advantages of PV systems [1].

As the global shift towards renewable energy gains momentum, grid-connected solar PV systems are gaining prominence as a major contributor to large-scale electricity production. Nepal possesses substantial potential for implementing a solar energy infrastructure. The Government of Nepal's 2018 white paper outlines a vision for renewable energy integration in the national energy matrix to ensure energy security, targeting 5-10 percent contribution from renewable sources in the power generation mix. The promotion of renewable energy in Nepal has played a pivotal role in extending clean energy access to its populace. As of now, approximately 55 MW of electricity has been generated from mini/micro-hydro and solar energy, facilitated by the Alternative Energy Promotion Center (AEPC), providing clean electricity solutions to 3.6 million households. This achievement has reached 18 percent of the total population and has fostered the creation of 30,000 jobs within this sector. According to DoED, till June 2023 many licenses had been issued for solar projects, where for construction 21 projects had a total capacity of 133.56 MW, and for survey 44 projects had a total capacity of 747.6 MW [2].

However, boosting the performance of solar PV systems requires addressing various performance issues. One approach to achieve this is by implementing a Monitoring System (MS) within the PV plant. This system monitors system variables, identifies discrepancies and errors, and reports the status and benchmarking data to the grid operator through a communication system. However, relying solely on the MS is insufficient to completely address the issue, as PV faults require specific techniques for detection and classification based on monitored data. Lately, machine learning-based techniques have shown improved classification performance in different scenarios, particularly in dealing with shadowing and PV module degradation [3].

Fault detection can be done by scrutinizing the I-V and P-V curves of each PV array, employing the Artificial Neural Network (ANN) algorithm. This novel approach excels in detecting faults with remarkable accuracy and possesses the capability to identify the specific fault type. The findings validate that the effectiveness of this proposed algorithm improves with the expansion of distinguishing data points, offering substantial benefits to the Solar PV industry [4].

This research work deals with the fault detection and diagnosis of photovoltaic systems. The real-time data from the 120kW grid-tied solar PV plant located at Time Pharmaceutical Factory in Gaidakot, Nawalpur, Nepal is collected. Firstly, the prediction model is developed using a multilayer neural network model by using the meteorological data. The forecasted power is then compared with the PV array output power of the real system to detect the presence of faults. Then, fault diagnosis is carried out by comparing the different electrical parameters, which are calculated using the analytical equations. The proposed model is then verified through simulation results.

2. Faults in Photovoltaic Systems

In general, faults in a PV system can occur on two sides: the DC side and the AC side, with the DC/AC inverter serving as the interface connecting them to the grid. To enhance system efficiency, it's crucial to implement an internal maximum power point tracker algorithm. On the AC side of the PV system, two common types of faults may arise; total blackouts, which are considered external faults like lightning or unbalanced voltage, and grid-related issues like switch failures, overcurrent, or overvoltage. This study, however, does not delve into the latter type of defect. Furthermore, it assumes that the PV array is the primary source of faults, as most PV inverters feature transformers that provide effective galvanic isolation between PV arrays and utility grids, along with robust electrical protections[5].

To effectively evaluate fault occurrence and its consequences, it is essential to examine, at the very least, the most prevalent faults in photovoltaic systems. In this context, an extensive study has categorized these faults into two groups: those occurring on the direct current (DC) side and those on the alternating current (AC) side. AC side faults are typically linked to issues with the system's inverter or the power grid itself. On the other hand, DC side faults encompass a wider range of problems, including faults in the Maximum Power Point Tracking (MPPT) algorithm, issues with bypass diodes, ground faults, arc faults, cell or module mismatches (which can be temporary or permanent), open circuits, and short circuits. For this study's purposes, the focus is primarily on the most common faults, namely module mismatches, open circuits, and short circuits [6].

3. Advanced Fault Identification Methods

Researchers have been investigating various approaches to identify and diagnose faults in PV systems. Techniques incorporating machine learning have been extensively examined as machine learning algorithms are powerful tools and provide an alternative approach to addressing intricate issues. Some of these methods focus on early fault detection, employing predictive measures to prevent significant power losses and damages to PV systems. Various approaches have been used in fault classification, including visual methods[7], thermographic image analysis [8], and mathematical methods [9] that employ theoretical and simulated models of PV plants. More recently, machine learning-based techniques have been proposed, which have shown improved classification performance in different scenarios, particularly in dealing with shadowing and PV module degradation. Different faults in the PV systems can be identified using a combination of different machine learning techniques with high accuracy. The combined sugeno fuzzy logic and artificial neural network method proposed in [10] showed an accuracy of 99.43% and 99.28% in detecting disconnected PV strings and short-circuited PV modules respectively. In [11] Multilayer Perceptron (MLP) and Radial Basis Function (RBF) have been implemented to evaluate their effectiveness in identifying the open-circuited strings. The RBF is found to have 97.9% accuracy. The existing literature concerning fault detection methods based on machine learning demonstrates the considerable reliability, efficiency, and accuracy of the suggested systems.

4. The Proposed Fault Detection Technique

This research presents a method for detecting and diagnosing faults in solar PV systems. Initially, a multilayer neural network model estimates optimal power based on environmental conditions using measured irradiance and temperature data. Subsequently, analytical equations are applied to characterize the nature of the faults. The proposed MNN model is trained and validated using real-time dataset from a solar power plant installed at the Time Pharmaceutical Factory in Gaidakot, Nawalpur, Nepal. To validate the proposed fault detection technique, a SIMULINK model is designed and it is injected diverse fault scenarios. For simplicity, the PV array connected to one of the inverters is considered. The figure 1 illustrates the overall methodology of the proposed research

Furthermore, the power plant's performance has been continuously tracked over the years using the SOLARMAN PRO online portal and retrieved via a data logger. Comprehensive site data, energy generation statistics, and technical details were gathered through both physical site inspections and the online portal. Key attributes of the panel are outlined in the table 1.

Table 1: Technical specifications of PV modules installed in site

Parameter	Value @ STC
Manufacturer	Vikram Solar
Model	Eldora VSP.72.AAA.05
Maximum Power, Pmax [Wp]	325 W
Maximum Voltage, Vmp [V]	37.8 V
Maximum Current, Impp [A]	8.61 A
Open circuit voltage, Voc [V]	46.2 V
Short circuit current, Isc [A]	9.13 A
Efficiency (%)	16.78%
No. of Cells	72
Temperature Coefficient of	
1. Isc	0.057%/°C
2. Voc	-0.29%/°C
3. Voc	-0.38%/°C

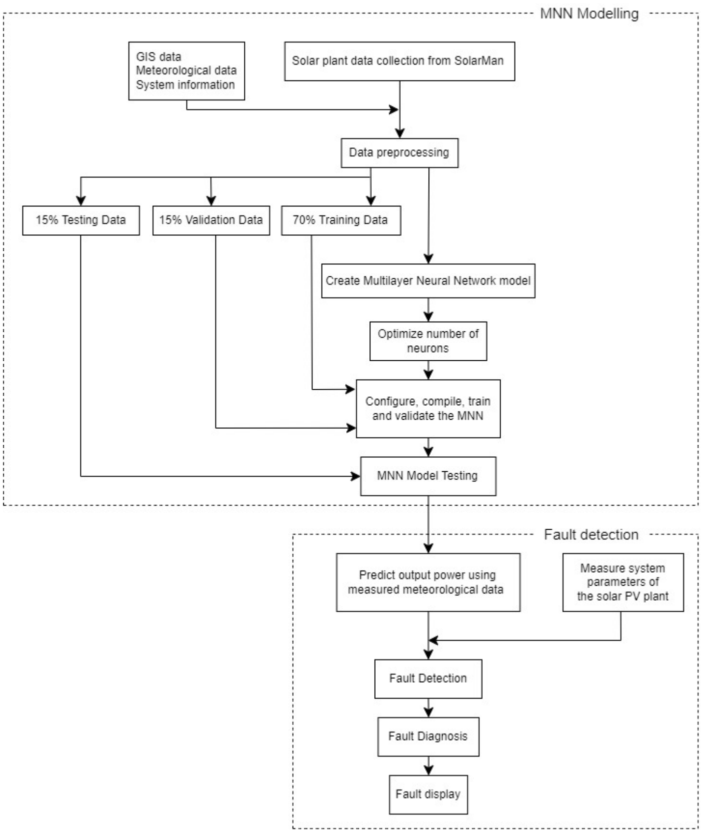

Figure 1: Research flowchart

4.1 Data Collection and Preprocessing

The historical time series data obtained through ground-based stations or satellite-based data set are used to prepare forecast models for solar output power. Given the limited availability of ground-based stations, this research opted to utilize satellite-based data from NASA's POWER database. Hourly solar irradiance and temperature data, encompassing All Sky Surface Shortwave Downward Irradiance (Wh/m^2) and Temperature at 2 Meters (°C), were collected for each hour [12].

Before inputting the collected variables into the artificial neural network, the collected datasets are preprocessed. Initially, missing values are removed, and non-relevant columns are dropped. Additionally, a preprocessing step involved filtering out data recorded between 6 p.m. and 6 a.m. daily since during these nighttime hours, PV systems typically do not generate significant energy, resulting in computational values near zero. To streamline the data and reduce redundancy in relationships, normalization is applied, thereby eliminating unnecessary features like duplications and anomalies.

4.2 Multilayer Neural Network-Based Solar PV Power Modelling

The Multilayer Neural Network (MNN) falls under the category of artificial neural networks, characterized by multiple interconnected layers of nodes or artificial neurons. As a feedforward neural network, it processes information in a unidirectional manner—from the input layer through the hidden layers to the output layer. MNNs find applications in various tasks, including pattern recognition, classification, and regression [13]. Each node in an MNN incorporates an activation function, such as the sigmoid, hyperbolic tangent (tanh), or Rectified Linear Unit (ReLU), governing its output based on the input.

Figure 2: Google Earth Map of Time pharmaceutical factory

The Node Output Equation with node i and output y_i in an MNN is given by eq. 1.

$$y_i = f\left(\sum_{j=1}^{n}(w_{ij}x_j + b_i)\right) \tag{1}$$

MNN training involves adjusting weights and biases to minimize the difference between forecasted and actual outputs, typically using optimization algorithms like gradient descent. The loss function quantifies the difference and aims to minimize it. The loss function, L, measures the difference between actual and forecasted outputs, to minimize:

$$L = \frac{1}{2}\sum_{k=1}^{n}(y_k - \hat{y}_{ik})^2 \tag{2}$$

In this research, the data format considered training the proposed model comprises of 4468 individual data sets pertaining to data of every hour of May. The data hour is considered from 6 pm to 6 am. Each data set comprises 2 data points about different parameters being employed in the forecast system. The pre-processed data is comprised of 240 input and output samples. The inputs are solar irradiance and ambient temperature of every hour of every day of May, 2023. This will allow the model to imitate the real system in any given scenario.

- Inputs = 4468 samples
- Targets = 4468 samples
- Training = 70%
- Validation = 15%
- Testing = 15%
- No. of neurons in the hidden layer = 20

The ANN is trained for different values of neurons in the hidden layer ranging from one to sixty, to ensure the accuracy of the forecast model. To analyse the neural network's performance, regression analysis is conducted between the network response and the corresponding targets. For each neuron count, the RMSE of the model is computed. The trained ANN with the lowest error for the given day is selected to forecast the progression of PV power for that specific day.

4.3 Fault Detection

In this research, the Artificial Neural Network (ANN) serves as an estimator, predicting expected power based on measured irradiance and temperature. By comparing the estimated power (P_pred) to the measured power (P_{PV}), the fault detection process is triggered, and the fault is categorized and displayed accordingly. The fault detection algorithm is shown in figure 3.

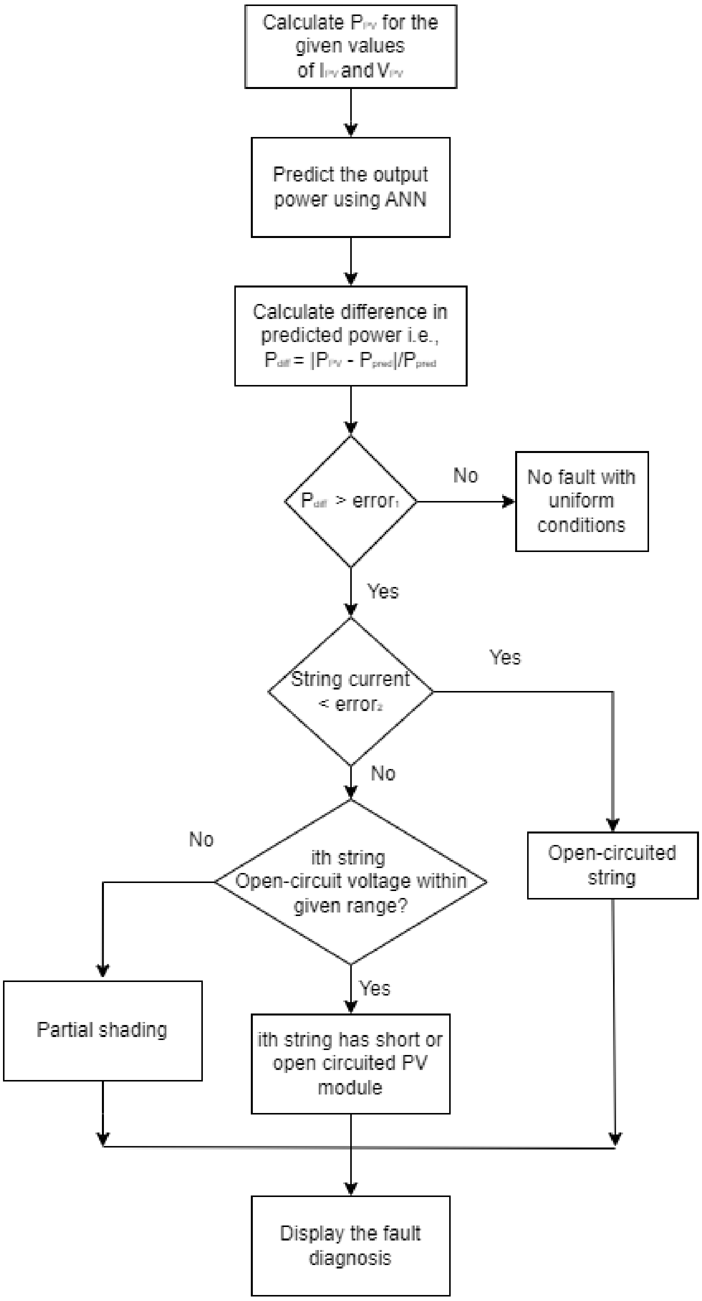

Figure 3: Fault detection and diagnosis model flowchart

5. Results and Discussion

5.1 Multilayer Neural Network Model

The MNN Model is trained with the optimal neurons in the hidden layer, iterating until the desired level of accuracy is achieved. During this training process, the model not only learns how external factors affect the PV output but also begins to replicate the losses occurring in the specific power plant configuration, incorporating all relevant details.

After multiple rounds of training and using different datasets, the optimal configuration identified consists of a network with a single hidden layer containing an ideal number of neurons as shown in figure 4.

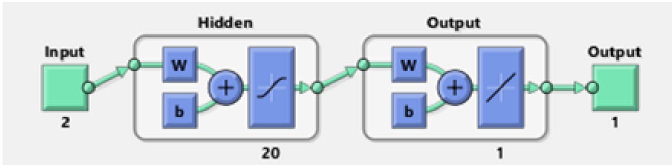

Figure 4: Trained MNN Model configuration

To validate the trained model, an independent test was conducted using a random month of observation, specifically June 2023, which was not part of the model's training/validation set. This test evaluates the model's forecasting accuracy by subjecting it to varying irradiance and temperature values over an entire month. The forecasted outcomes are then compared to measurements of the real system to gauge the accuracy of the ANN model, as shown in figure 5.

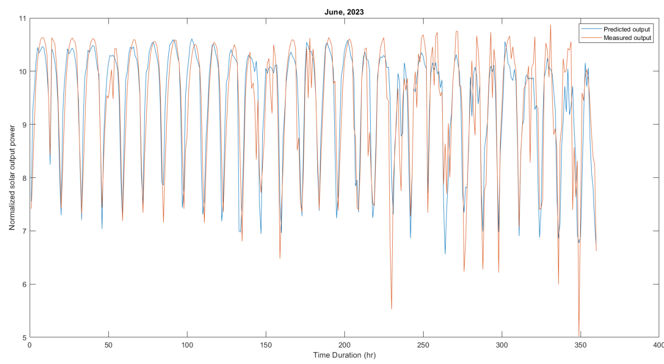

Figure 5: Full-month comparison between PV output forecasting using MNN Model and measured PV output (June, 2023)

The results from the tests reveal that the ANN has successfully learned the operational modes of the photovoltaic system based solely on solar radiation and measured temperatures. The final results were highly promising, with the ANN providing accurate estimations of output power for the specified days, as evidenced by very low measured errors. It's noteworthy that the inputs introduced were entirely new to the model, not part of the training/validation set, yet the ANN successfully aligned with the measured values.

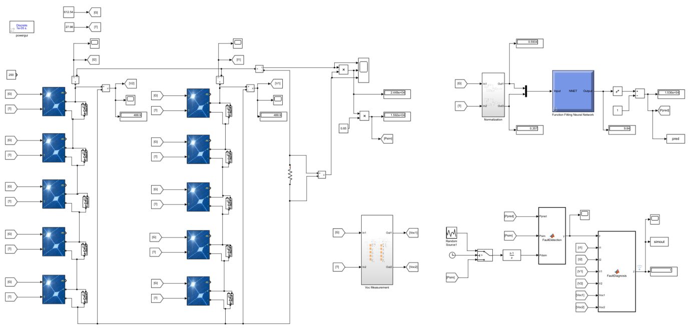

Figure 6: Fault detection model

5.2 Fault Detection

To test the proposed fault detection and diagnosis method, a Simulink model replicating the PV array installed at the Time Pharmaceutical Factory is designed as shown in figure 6. For simplicity, the model only incorporates the PV arrays connected to one of the 50 kWp inverters. This model is designed based on the PV string configuration of a reference PV system installed at site. It comprises a total of 180 solar modules, arranged with 20 panels in series and 9 strings in parallel.

The fault detection mechanism is proposed to detect four possible faults, which are MPPT fault, open-circuited PV string, open-circuited/short-circuited PV module and shaded solar module(s). Using the Simulink model, the proposed fault detection algorithm was subjected to four different types of faults. Each fault consists of a gradual percentual degradation in power. Each fault scenario introduced a gradual percentage degradation in power. The algorithm relies on detecting changes in system parameters to classify and identify faults accurately.

Codes are assigned to the different types of faults as shown in the table 2 such that the respective code is displayed in the output terminal.

Table 2: Fault Cases and Codes

Cases	Fault Code	Fault Type
Case 0	0	No fault
Case 1	1	MPPT fault
Case 2	2	Open-circuited string 1
Case 3	3	Open-circuited string 2
Case 4	4	Open/short-circuited module in string 1
Case 5	5	Open/short-circuited module in string 2
Case 6	6	Partial shading

For the given irradiance and temperature, the forecasted output power is used as a base to test the condition of the solar PV model undergoing different faults. For each case, the solar PV system is modeled to represent the respective fault. The fault classification algorithm is devised to pinpoint the specific string affected by a fault. In testing, identical faults are introduced simultaneously to different strings. The results demonstrate that the corresponding fault codes are accurately displayed for the respective strings experiencing faults, as illustrated in Cases 2 and 3 in figure 7. However, for clarity regarding the string number experiencing the fault, distinct

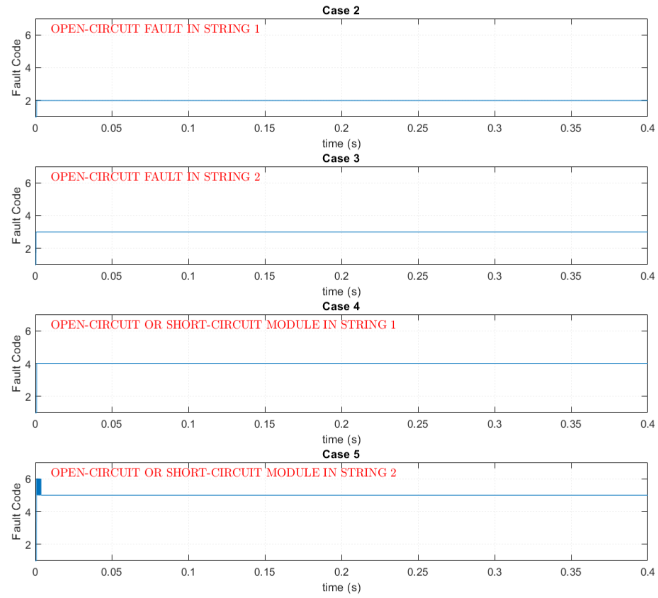

Figure 7: Fault indicators for open-circuited string and module fault

fault codes are assigned to individual strings, as depicted in the figure 7. Similarly, for MPPT fault and partial shading, the fault codes indicated in the table 2 are displayed by the proposed model.

6. Conclusion

As the renewable energy sector continues to expand, the importance of reliable fault detection mechanisms becomes crucial. The application of advanced machine learning techniques for fault detection represents a promising approach for optimizing the performance of solar energy systems. This study embarked on a fault detection mechanism of solar modules, employing Multilayer Neural Network as a powerful tool for enhancing the reliability and efficiency of solar energy systems. A total of 240 datasets were used to train, validate and test the MNN model and rigorous testing under varied meteorological conditions validated the model's effectiveness, demonstrated by its accurate output power predicting over the entire month. The accuracy of the trained model is found to be 96.1% and 94.9% for May and June respectively. The normal condition and four different fault conditions are defined for the proposed fault detection and diagnosis algorithm. The proposed is found to be significantly more accurate in fault identification than the existing traditional methods.

Acknowledgments

The authors are thankful to Mr. Narayan Dhakal, Factory Operation Director at Time Pharmaceutical Factory, for generously providing data and resources essential for the success of this research. The authors also express their sincere gratitude to the program coordinator Assoc. Prof. Dr. Hari Bahadur Darlami of Master of Science in Renewable Energy Engineering, for his invaluable guidance and inspiration throughout this research.

References

[1] Yassine Chouay and Mohammed Ouassaid. An intelligent method for fault diagnosis in photovoltaic systems. In *2017 International Conference on Electrical and Information Technologies (ICEIT)*, pages 1–5. IEEE, 2017.

[2] Annual Progress Report - AEPC — aepc.gov.np. https://www.aepc.gov.np/documents/annual-progress-report-aepc. [Accessed 09-09-2023].

[3] Elias Roumpakias and Tassos Stamatelos. Health monitoring and fault detection in photovoltaic systems in central greece using artificial neural networks. *Applied Sciences*, 12(23):12016, 2022.

[4] Ahmed A Al-Katheri, Essam A Al-Ammar, Majed A Alotaibi, Wonsuk Ko, Sisam Park, and Hyeong-Jin Choi. Application of artificial intelligence in pv fault detection. *Sustainability*, 14(21):13815, 2022.

[5] Hee-Won Lim, Il-Kwon Kim, Ji-Hyeon Kim, and U-Cheul Shin. Simulation-based fault detection remote monitoring system for small-scale photovoltaic systems. *Energies*, 15(24):9422, 2022.

[6] Siva Ramakrishna Madeti and SN Singh. A comprehensive study on different types of faults and detection techniques for solar photovoltaic system. *Solar Energy*, 158:161–185, 2017.

[7] C. Packard, J. Wohlgemuth, and S. Kurtz. Development of a Visual Inspection Data Collection Tool for Evaluation of Fielded PV Module Condition. aug 1 2012.

[8] J. Tsanakas, D. Chrysostomou, P. Botsaris, and A. Gasteratos. Fault diagnosis of photovoltaic modules through image processing and Canny edge detection on field thermographic measurements. *International Journal of Sustainable Energy*, 34(6):351–372, aug 14 2013.

[9] S. R. Madeti and S. Singh. Monitoring system for photovoltaic plants: A review. *Renewable and Sustainable Energy Reviews*, 67:1180–1207, 1 2017.

[10] R. G. Vieira, M. Dhimish, F. M. U. de Araújo, and M. I. S. Guerra. Pv Module Fault Detection Using Combined Artificial Neural Network and Sugeno Fuzzy Logic. *Electronics*, 9(12):2150, dec 15 2020.

[11] M. Hussain, M. Dhimish, S. Titarenko, and P. Mather. Artificial neural network based photovoltaic fault detection algorithm integrating two bi-directional input parameters. *Renewable Energy*, 155:1272–1292, 8 2020.

[12] Paul Stackhouse. NASA POWER | Data Access Viewer — power.larc.nasa.gov. https://power.larc.nasa.gov/data-access-viewer/. [Accessed 09-09-2023].

[13] K. El Bouchefry and R. S. de Souza. Learning in Big Data: Introduction to Machine Learning. *Knowledge Discovery in Big Data from Astronomy and Earth Observation*, pages 225–249, 2020.