Fault Diagnosis of SPV Power Plant Based On Real-Time Data

Laxmi Kant Sagar
Department of Computer Science & Engineering, G.L.Bajaj Institute of Technology & Management

D Bhagwan Das
Department of Electrical Engineering, Dayalbagh Educational Institute

https://doi.org/10.5109/7160903

出版情報：Evergreen. 10 (4), pp.2277-2286, 2023-12. 九州大学グリーンテクノロジー研究教育センター
バージョン：

Fault Diagnosis of SPV Power Plant Based On Real-Time Data

Laxmi Kant Sagar[1], D Bhagwan Das[2]
[1]Department of Computer Science & Engineering, G.L.Bajaj Institute of Technology & Management, Greater Noida, India
[2]Department of Electrical Engineering, Dayalbagh Educational Institute, Agra, India

*Author to whom correspondence should be addressed:
E-mail: laxmi.sagar@gmail.com

(Received December 1, 2022; Revised September 15, 2023; accepted November 27, 2023).

Abstract: To guarantee maximum performance and reliability of solar plants, the probability of occurrence of improper operations should be very low. For that, a fault diagnosis expert system can be one of the solutions that can figure out the improper operations very precisely and suggest corrective actions. In this study, we will use a hybrid approach for fault detection, i.e., a knowledge-based and data-driven approach, i.e., fuzzy logic and machine learning. A fuzzy model will sense the real-time data and classify it into good or faulty. The fuzzy simulation model, thus, developed, will be used to generate a large amount of training data required for machine learning methods. Further clustering techniques are applied to the available database to determine which improper operation (fault) has the highest probability of occurrence.

Keywords: fuzzy logic; machine learning; clustering; classification.

1. Introduction

Fault diagnosis is becoming an area of prime focus as it can detect early stages of improper operations and prevent converting the fault into failure. Technological advancements, especially IoT and A.I., have provided a much-needed pace. Real-Time monitoring systems are being designed to collect real-time data from plants to inspect any plant's current state of health. Data collection is speedy, making it impossible for any human operator or expert to analyze and draw inferences. Plotting graphs to study the changes or variations in prime attributes or parameters can be one of the options to analyze the behavior of solar plants but still, human interpretation and continuous monitoring. The only solution can be designing a method based on human experience and expert knowledge that can use human intelligence and make automated and intelligent decisions. Thus, expert fault diagnosis has become a hot topic of research.

Fault diagnosis systems started in the early 1970s, from quantitative/qualitative models to today's artificial approach (rule base or data-driven). Deyin et al.[1)] propose a Multi backpropagation expert system for fault diagnosis in transformers. Results show BPNN is better than BPEG. Shin et al.[2)] focused on the predictive maintenance of wind farms with the help of A.I. assistants. Wang et al.[3)] propose an enhanced fault detection technique based on the Bayesian network and expectation-maximization algorithm. The algorithm imputes missing data with high accuracy. Then these imputed data sets are processed to get parameters for the Bayesian network. Hocine et al.[4)] propose an Observing Degradation System that automatically detects any anomaly based on P-V and I-V characteristics. Abu-Rub et al.[5)] discussed the machine learning approach for solving various challenges posed by a power electronics-dominated grid. Nan et al.[6)] developed a knowledge-based fault diagnostic system for a micro steam plant with two essential elements: process trend recognition and a fuzzy logic system. A GDA application was constructed utilizing the G2 real-time expert system shell to implement both components. Heung et al.[7)] presented an online fuzzy expert system that diagnoses faults in substations and transmission networks. Monsef et al.[8)] demonstrated a component-oriented fuzzy fault diagnosis expert system for the power system. The expert system handles uncertain input data and processes it to give information about the faulted component. Honglu et al.[9).] Using the Spatio-Temporal Distribution Characteristics of Photovoltaic Array Output, a new fault diagnosis method is proposed. Under various fault scenarios, features of the temporal fluctuation and spatial distribution of P.V. array output are examined. Houssein et al.[10)] proposed software for fault diagnosis, the software diagnosis panel faults, and also various zones of failure. The fault diagnosis is based on comparing production estimated with production measured. Dhimish and Holmes[11)] proposed a t-test statistical technique-based fault detection algorithm for grid-connected photovoltaic plants. The algorithm could detect various GCPV faults, such as faulty maximum power point tracking. Ali et al.[12)] developed a real-time monitoring and fault diagnosis technique. Fault diagnosis is made based on each fault's normal and fault threshold values. Chine et al.[13] presented an artificial neural network approach for fault diagnosis of P.V. arrays using various parameters like solar irradiance, the P.V. array's current and voltage, and the P.V. module's temperature. Furthermore, this fault diagnosis is also implemented in Field Programmable Gate Array (FPGA) to prove its effectiveness. Silvestre et al.[14] presented a method for grid-connected P.V. systems to automatically monitor and diagnose faults. In the case of free fault system operation, the fault detection threshold is established based on the discrepancy between the simulated and measured capture losses. Berawi et al.[15] employed Decision Tree and Fuzzy Logic to rank earthquake victims based on their needs. Gupta et al.[16] and Kou et al.[28] proposed carbon management techniques for low-carbon supply chain management by integrating fuzzy DEMATEL and fuzzy TOPSIS techniques. Zohedi et al.[17] and Mohd et al.[29] proposed a fuzzy logic for modeling a remotely operated vehicle. According to the results, SIFLC is superior to Proportional, Integral, and Derivative (PID) controllers. Choirunisa et al.[18] modeled an M.R. Damper utilizing an Adaptive Neuro-Fuzzy Inference System with Gaussian and generalized bell membership functions. Results indicate that the Gaussian function is more precise than the generalized bell membership function. Sagar and Das[19] proposed a fuzzy expert system for determining the state of solar photovoltaic plants based on real-time data. Yang et al.[20] proposed a novel technique for fault detection in components of wind turbines based on a reconstruction model using SVR. The results show proposed methodology performance is much better than SCADA. Bode et al.[21] analyzed the working of the heat pump with fault laboratory data in real-time conditions. The results show that the trained algorithm performs better than real-time data. Murphey et al.[22] proposed a model-based diagnostic system for detecting multiple faults in electric drive inverters. The approach was very effective and detected nine types of faults with an accuracy of 98%. Zhao et al.[23] proposed a graph-based semi-supervised learning method for fault detection in photovoltaic systems. The proposed models not only detects faults but also suggests fault type. Momeni et al.[24] proposed a graph-based semi-supervised learning model with extended class labels to identify, classify and locate faults in solar photovoltaic arrays. The results show that the current method is better than the previous GBSSL algorithm. Heo and Lee[25] used deep neural networks for fault classification and detection. The results obtained outperform other data-driven methods.

Many methods are used for fault diagnosis and predictive maintenance, which include model-based, Knowledge-Based, and data-driven methods[26]. Liao and Felix[30] proposed a hybrid framework to improve prediction accuracy. The hybrid framework consists of a Bayesian model along with two data-driven approaches. Shiekh et al.[31] used a data-driven approach for monitoring batteries' state charge and health. A proposed machine learning algorithm extracts feature from the discharge curve to estimate SOH and SOC. Hui et al.[32] used fuzzy and machine learning to improve the classical BN-S algorithm to analyze price fluctuations in the stock exchange. Results show that the refined BN-S model is better than the classical BN-S model in terms of long-term dependence. Fan et al.[33] proposed a fuzzy-based Gaussian Error Relative Support vector machine to predict body fat. Experimental results show that FWRESVM outperforms other conventional machine-learning algorithms

This study will focus on a knowledge-based and data-driven approach for dynamic systems such as solar plants. These models can be easily used where explicit mathematical models or equations cannot be derived. Knowledge-based expert diagnosis analyzes real-time data and compares it with predefined rules derived from past human experience. The fuzzy logic will deal with uncertainty due to stochastic[27] behavior of climatic conditions and machine learning algorithms will classify and cluster the incoming real-time data. In this study, we will consider only four types of fault- Grid Outage, Excessive Battery charging, Problem with Grid Export Settings, and Dusty Panels. All these faults can easily be interpreted by analyzing data.

2. Experimental Section

Smart Micro grid at DEI, Agra forms the basis of this study. Inverter data collected through this smart micro grid for last 10 years is used as reference. The real time data is fed to fuzzy expert system as input variables. For simulation, matlab is used. Further for classification and clustering, spyder software is used.

3. Proposed System for Fault Diagnosis

3.1 System Methodology

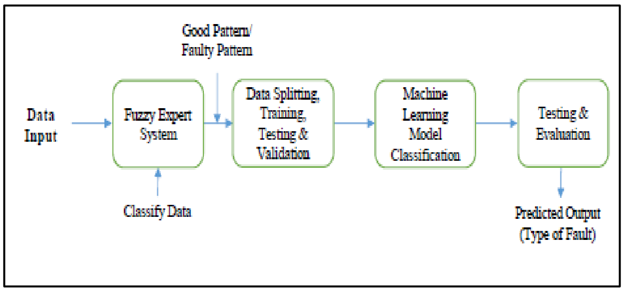

Fig 1. System Methodology

The suggested methodology is broken down into four steps. Real-time data collection comes first in the process. The second stage is to integrate a fuzzy model with live-streaming data. The fuzzy model's output is stored as the third stage. The fourth stage is to train the model using machine learning techniques on saved data collected from fuzzy diagnosis systems to improve the system's intelligence.

3.2 Data Collection

Data Acquisition Software is specially designed to collect data from an inverter connected to a solar panel, and it also provides a graphical user interface to monitor streaming data. Streaming data comes at 3 secs and is stored on a dedicated server through MySQL database. The setup has been done at Dayalbagh Educational Institute, Agra (India). The streaming data coming from Data Acquisition System is fed to the fuzzy model as input to the fuzzy system.

3.3 Designing Fuzzy Model

3.4 Input Parameters

Data acquisition software was used to store historical data, which was analyzed to determine the data range for each input parameter. The stored data consists of thirty-six parameters, but only four, i.e., Src_Kw, solar_irrad, battvpc, and load_kW, are considered necessary for fault detection. The range for all input parameters is selected from the previous ten years' stored data.

Table 1. Various SPV parameters, along with their range

PARAMETERS	***RANGE***
Battery VPC	1.5V-2.3V
Irradiation	0-1332 W/m2
Load_kW	1kW-32kW
Source_kW	-28.5kW to + 36kW

The proposed fuzzy fault diagnosis system consists of four linguistic inputs, i.e., Battery_ VPC (voltage per cell), "load," "irradiance," and "source." Membership functions are used to represent these fuzzy variables described by linguistic terms. The membership functions translate the fuzziness level between 0 and 1. Triangular-shaped membership function has been taken into consideration for this study. The range of input variables is taken from table 1.

The four input variables are as follows:

Battery VPC- The battery voltage per cell (VPC) indicates the battery's charging status. The range of Battery VPC is between 1.5V to 2.3V. The mathematical and graphical representation is shown in fig (4), where $f_l(x)$ =low, $f_m(x)$ =medium, and $f_h(x)$ =high.

$$f_l(x) = \begin{cases} 0, \quad x \leq 1.529, x \geq 1.916 \\ \dfrac{(x-1.528)}{(1.744-1.528)}, 1.528 < x \leq 1.744 \\ \dfrac{(1.916-x)}{(1.916-1.745)}, 1.745 < x < 1.916 \end{cases}$$

$$f_m(x) = \begin{cases} 0, \quad x \leq 1.837, x \geq 2.095 \\ \dfrac{(x-1.837)}{(1.965-1.837)}, 1.837 < x \leq 1.965 \\ \dfrac{(2.095-x)}{(2.095-1.965)}, 1.965 < x < 2.095 \end{cases}$$

$$f_h(x) = \begin{cases} 0, \ x \leq 2.047, x \geq 2.393 \\ \frac{(x-2.047)}{(2.21-2.047)}, 2.047 < x \leq 2.21 \\ \frac{(2.393-x)}{(2.393-2.21)}, 2.21 < x < 2.393 \end{cases}$$

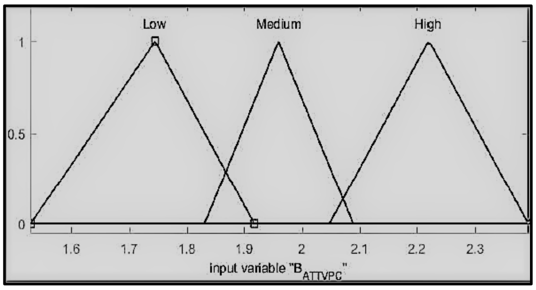

Fig 2. M.F.s for Batt_VPC

Source_Kw defines the power import/export from the grid. Negative value. i.e., "-ve" power is supplied to the grid, and the positive value is "+ve" power is taken from the grid. The range of Source_kW is between -28.5 to 36.0 kW. Mathematical and graphical representations are shown in fig (5) where $f_{vn}(x)$ =very negative, $f_n(x)$ =negative, $f_z(x)$ =zero and $f_p(x)$ =positive, and $f_{vp}(x)$ =very positive

$$f_{vn} = \begin{cases} 0, \quad x \leq -40.03, x \geq -15.4 \\ \dfrac{(x - -40.03)}{(-27.63 - -40.03)}, 9.51 < x \leq -27.63 \\ \dfrac{(-15.4-x)}{(-15.4 - -27.63)}, -27.63 < x < -15.4 \end{cases}$$

$$f_n(x) = \begin{cases} 0, \quad x \leq \ -17.39, x \geq -4.254 \\ \dfrac{(x - \ -17.39)}{(-10.02 - \ -17.39)}, -17.39 < x \leq -10.02 \\ \dfrac{(-4.254-x)}{(-4.254 - -10.02)}, -10.02 < x < -4.254 \end{cases}$$

$$f_z(x) = \begin{cases} 0, \quad x \leq -5.14, x \geq 5.064 \\ \dfrac{(x - -5.14)}{(0.757 - -5.14)}, [-5.14 < x \leq 0.757 \\ \dfrac{(5.064-x)}{(5.064-0.757)}, 0.757 < x < 5.064 \end{cases}$$

$$f_p(x) = \begin{cases} 0, \quad x \leq 3.33, x \geq 13.2 \\ \dfrac{(x-3.33)}{(8.995-3.33)}, 3.33 < x \leq 8.995 \\ \dfrac{(13.2-x)}{(13.2-8.995)}, 8.995 < x < 13.2 \end{cases}$$

$$f_{vp}(x) = \begin{cases} 0, \quad x \leq 11.97, x \geq 34.57 \\ \dfrac{(x-11.97)}{(22.17-11.97)}, 11.97 < x \leq 22.17 \\ \dfrac{(34.57-x)}{(34.57-22.17)}, 8.995 < x < 34.57 \end{cases}$$

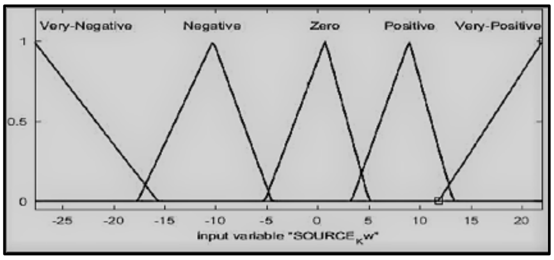

Fig 3. M.F.s for Input "Source_kW"

Irradiance is described as how much sunshine hits the solar panel. The collected data ranges between 0 and 132 W/m2, even though the irradiance range is between 0 and 1300 W/m2 because it has been scaled down by one-tenth of its original value**.** The mathematical and graphical representation is shown in fig (3), where f_l(x) =lowf_m(x)=medium, and f_h(x) =high.

$$f_l(x) = \begin{cases} 0, \quad x \leq 0, \\ \dfrac{(x-0)}{44.5}, 0 < x \leq 44.5 \\ 1, \quad x \geq 44.5 \end{cases}$$

$$f_m(x) = \begin{cases} 0, \; x \leq 42.3, x \geq 89.6 \\ \frac{(x-42.3)}{(65.6-42.3)}, 42.3 < x \leq 65.6 \\ \frac{(89.6-x)}{(89.6-65.6)}, 65.6 < x < 89.6 \end{cases}$$

$$f_h(x) = \begin{cases} 0, \quad x \leq 87, \\ \dfrac{(x-87)}{132-87}, 87 < x \leq 132 \\ 1, x \geq 132 \end{cases}$$

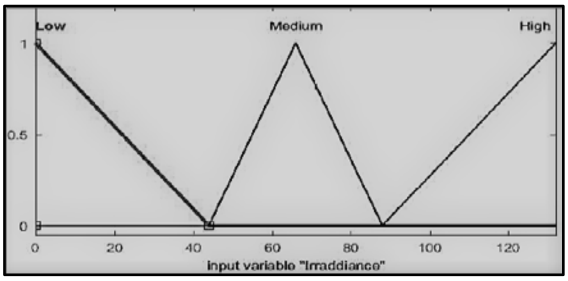

Fig 4. M.F.s for Input "Irradiance"

Load_kW represents load of a building or unit. A high load necessitates a high power consumption. The Load_kW range was between 1kW and 30kW. The mathematical and graphical representation is shown in fig. (4) where f_l(x)=low, f_m(x)=medium, and f_h(x)=high.

$$f_l(x) = \begin{cases} 0, \quad x \leq 0, \\ \dfrac{(x-0)}{9.9}, 0 < x \leq 9.9 \\ 1, x \geq 9.9 \end{cases}$$

$$f_m(x) = \begin{cases} 0, \quad x \leq 9.53, x \geq 20.5 \\ \dfrac{(x-9.53)}{(14.9-9.53)}, 9.53 < x \leq 14.9 \\ \dfrac{(20.5-x)}{(20.5-14.9)}, 14.9 < x < 20.5 \end{cases}$$

$$f_h(x) = \begin{cases} 0, \; x \leq 19.7, \\ \frac{(x-19.7)}{29.7-19.7}, 19.7 < x \leq 29.7 \\ 1, x \geq 29.7 \end{cases}$$

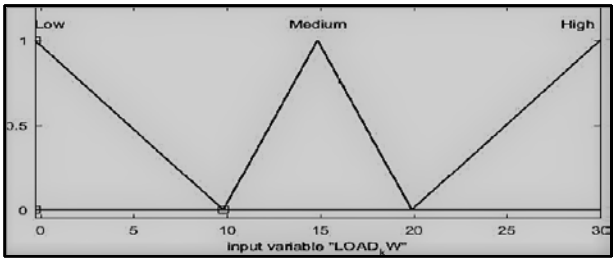

Fig 5. M.F.s for Input "Load_kW"

3.5 Output Parameter

The output parameters of the fuzzy fault diagnosis system are "Faulty Data" or "Good Data," defined over universal discourse. In case of faulty data, corrective action is also suggested.

3.6. Rule Base

Decisions are based on knowledge represented by the rule base. Simple if-then statements represent the rule base in fuzzy systems. The Mamdani model was used to create the fuzzy inference system in this study, but only 45 of the rules are represented in Table 2. which is faulty.

3.7. Defuzzification

Because the output of a fuzzy inference system is not a crisp value, defuzzification is a technique used to convert fuzzy variables into crisp values. There are various approaches to defuzzification. The centroid approach was employed for defuzzification in this study. Formula to defuzzify the output -

$$Centroid\ \bar{X} = \frac{\sum_i^n x\mu(x)}{\sum_i^n \mu(x)}$$

where $\bar{X}$ = crisp output

$\mu(x)$=aggregated membership function

x=output variable

3.8 Faulty Patterns in Data

Data gathered using data capture software is examined to look for patterns that do not match our assumptions. These unexpected patterns are then converted into forms of rules to design a fuzzy expert system. These faulty patterns are based on facts and rules.

Some of the rules include the following:

- No electricity can be sent to the grid when the battery is low.
- Power is exported to the grid during periods of high irradiation and low load.
- Load and battery charging needs will be met mainly through irradiance (solar) rather than the grid.

This study will focus on the prime attributes important in finding faulty patterns. These attributes are "battery_vpc," "load," "source," and "irradiance." The range of each parameter was calculated from the last ten years of collected data. The calculated range is further divided into three categories, i.e., "low," "medium," and "high" for "battery_vpc," "load," and "irradiance." To get more precise results, the "source" range is divided into five categories "very negative," "negative, "zero," "positive," and "very positive." These parameters will be input parameters to fuzzy fault diagnosis expert system and the categories "low," "high," "medium" will contribute to membership functions of input parameters.

Based on predefined rules. Some of the faulty patterns include:

1. When the "irradiance" value is medium, the value of "load" is low. The battery is fully charged, i.e., the value of battery_vpc is "high." The value of "source" (i.e., grid) should be negative (i.e., electricity is exported), but if the value is positive (i.e., electricity is being imported) or zero (i.e., electricity is neither imported nor exported). In this case, the fault will be because "Grid outage/battery charging through the grid is disabled, " which must be checked.
2. When the "irradiance" value is high, the value of "load" is low. If the battery is fully charged, i.e., the value of battery_vpc is "high." The value of the "source" (i.e., grid) should be negative or very negative (i.e., electricity should be exported), but if the value is zero or positive. In this case, the fault will be because "Grid outage/battery charging through the grid is disabled, " which must be checked.
3. When the "irradiance" value is low, the value of "load" is low. The battery is not fully charged, i.e., the value of batt_vpc is low. The value of "source" (i.e., grid) should be positive (i.e., electricity is imported), but if the value is zero, then the fault will be because of "Grid outage/battery charging through the grid is disabled," which has to be checked.
4. When the "irradiance" value is medium, the " load " value is low, and battery_vpc, i.e., is low. The value of "source" (i.e., grid) should be positive (i.e., electricity is imported), but if the value is zero, then the fault will be because of "Grid outage/battery charging through the grid is disabled," which has to be checked.
5. When the "irradiance" value is low, the value of "load" is low, and battery_vpc is low. The value of "source" (i.e., grid) should be positive (i.e., electricity is imported), but if the value is zero, then the fault will be because of "Grid outage/battery charging through the grid is disabled," which has to be checked.
6. When the "irradiance" value is low, the " load " value is medium, and battery_vpc, i.e., is medium. The value of "source" (i.e., grid) should be positive (i.e., electricity is imported), but if the value is zero, then the fault will be because of "Grid outage/battery charging through the grid is disabled," which has to be checked.
7. When the "battery_vpc" value is high, the value of "irradiance" is low, and the "load" value is low then the value of "source" (i.e., grid) should be zero; if the value is very positive, then the fault will be because of "Excessive battery charging through grid software," which has to be checked.
8. When the "battery_vpc" value is high, the value of "irradiance" is low, and the "load" value is low. The value of "source" (i.e., grid) should be zero; if the value is positive, then the fault will be because of "Excessive battery charging through grid software," which has to be checked.
9. When the "battery_vpc" value is high, the value of "irradiance" is low, and the "load" value is low, then the value of "source" (i.e., grid) should be zero if the value is very negative or negative then the fault will be because of "Problem with grid export settings."
10. When the "battery_vpc" value is low, the value of "irradiance" is medium, and the "load" value is low, then the value of "source" (i.e., grid) should be zero; if the value is very negative or negative, then the fault will be because of "Problem with grid export settings."
11. When the "battery_vpc" value is low, the value of "irradiance" is medium, and the "load" value is low, then the value of "source" (i.e., grid) should be zero; if the value is very negative or negative, then the fault will be because of "Problem with grid export settings."
12. When the "battery_vpc" value is low, the value of "irradiance" is high, and the "load" value is low, then the value of "source" (i.e., grid) should be negative if the value is positive. The fault will be "Dusty panels / faulty connectors/ MPPT problem / Shadows."

13. When the "battery_vpc" value is low, the value of "irradiance" is high, and the "load" value is medium. The value of "source" (i.e., grid) should be negative; if the value is very positive, then the fault will be because of "Dusty panels / faulty connectors/ MPPT problem / Shadows."
14. When the "battery_vpc" value is low, the value of "irradiance" is high, and the "load" value is medium. The value of "source" (i.e., grid) should be negative; if the value is positive, then the fault will be because of "Dusty panels / faulty connectors/ MPPT problem / Shadows."

Above are some faulty data patterns obtained from the data acquisition system. These faulty patterns are almost forty-five, but only a few are listed to show how the data is analyzed. The fuzzy expert system will convert these faulty conditions into fuzzy rules.

Table 2. Fuzzy Rules for Fault Detection System

INPUT					OUTPUT
R NO	VPC	Rad	Load	Source	FAULTY DATA
1	L	L	L	Z	FAULTY DATA
2	L	L	L	VN	FAULTY DATA
3	L	L	M	Z	FAULTY DATA
4	L	L	M	N	FAULTY DATA
5	L	L	H	Z	FAULTY DATA
6	L	M	L	VN	FAULTY DATA
7	L	M	M	VP	FAULTY DATA
8	L	M	M	N	FAULTY DATA
9	L	H	L	VP	FAULTY DATA
10	L	H	L	P	FAULTY DATA
11	L	H	M	VP	FAULTY DATA
12	L	H	M	P	FAULTY DATA
13	L	H	M	N	FAULTY DATA
14	L	H	M	VN	FAULTY DATA
15	L	H	H	Z	FAULTY DATA
16	M	L	L	Z	FAULTY DATA
17	M	L	M	Z	FAULTY DATA
18	M	L	M	N	FAULTY DATA
19	M	L	M	VN	FAULTY DATA
20	M	L	H	Z	FAULTY DATA
21	M	L	H	N	FAULTY DATA
22	M	M	L	N	FAULTY DATA
23	M	M	M	Z	FAULTY DATA
24	M	M	M	N	FAULTY DATA
25	M	M	H	Z	FAULTY DATA
26	M	H	L	VP	FAULTY DATA
27	M	H	L	P	FAULTY DATA
28	M	H	M	VP	FAULTY DATA
29	M	H	H	Z	FAULTY DATA
30	M	H	H	N	FAULTY DATA
31	H	L	L	P	FAULTY DATA
32	H	L	M	N	FAULTY DATA
33	H	L	M	VN	FAULTY DATA
34	H	L	H	N	FAULTY DATA
35	H	L	H	VN	FAULTY DATA
36	H	M	L	P	FAULTY DATA
37	H	M	L	Z	FAULTY DATA
38	H	M	M	VP	FAULTY DATA
39	H	M	H	VP	FAULTY DATA
40	H	H	L	VP	FAULTY DATA
41	H	H	L	P	FAULTY DATA
42	H	H	L	Z	FAULTY DATA
43	H	H	M	VP	FAULTY DATA
44	H	H	M	P	FAULTY DATA
45	H	H	H	P	FAULTY DATA

3. Result of Fuzzy Approach

The simulation of the fuzzy system is done in MATLAB. For this study, the SPV plant of 40 kWp, installed at Dayal Bagh Educational Institute, Arts Department, Agra, is considered. The Fuzzy Diagnostics system's results (Fig. 6) are very convincing, according to professionals overseeing the power plant's operations for the past ten years.

In Fig.6, n1 is the variable containing the dataset. The fuzzy system classifies the real-time streaming data into faulty and good data patterns, and in case of faulty data, corrective actions are also suggested.

For Fault_Case1- the data pattern is of a good pattern, so no rule fires.

For Fault_Case2- the data pattern is of **faulty pattern, i.e.,** [Low Low Low Zero] as per **rule 1** as per table 3. The corrective action is- "Grid Outage/ battery charging disabled."

Fault_Case3- the data pattern is faulty**,** i.e. [Low Low Medium Zero] as per **rule 3** as per table 3. The corrective action is- "Problem with grid export settings."

Fault_Case4- the data pattern is faulty, i.e. [Medium Low Low Zero] as per **rule 16** as per table 3. The corrective action is-"Grid Outage/ battery charging disabled."

Fault_Case5- the data pattern is faulty, i.e. [Low High Medium Positive] as per **rule 12** as per table 3. The corrective action is- "Dusty Panels/MPPT/ Shadows/ Faulty Connectors."

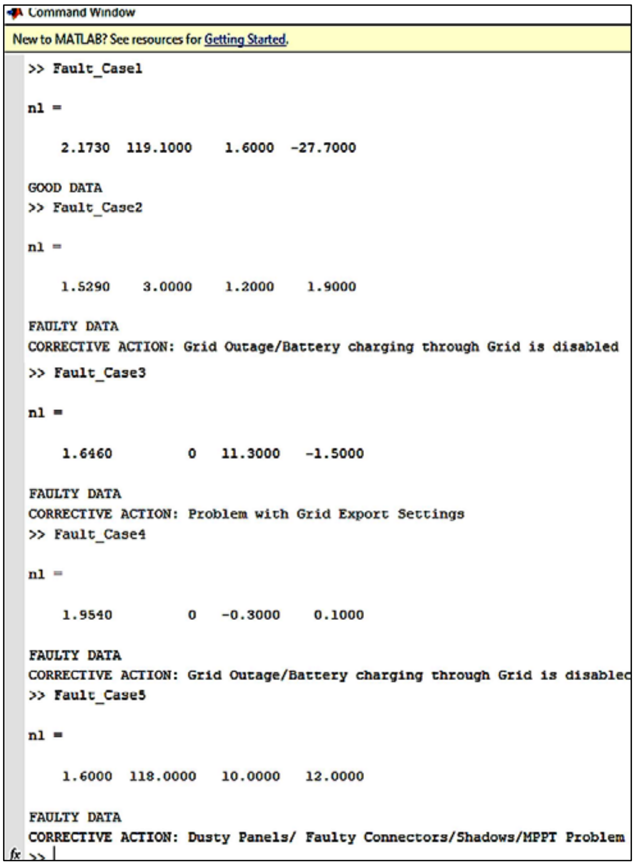

Fig 6. Some test cases and their outputs

The results from the fuzzy diagnosis system will be the dataset for training the model using machine learning algorithms to make the system more intelligent and faster.

battvpc	solar_irr	load_kW	Src_kW	Fault
2.168	81.5	0	0	Grid Outage
2.167	81.5	0	0	Grid Outage
2.167	81.8	0	0	Grid Outage
2.167	82.5	0	0	Grid Outage
2.166	83.2	0	0	Grid Outage
2.167	83.6	0	0	Grid Outage
2.168	84.3	0	0	Grid Outage
2.148	19.7	7.8	4.2	Excessive Battery Charging through Grid- Software setting
2.149	19.7	7.7	4.1	Excessive Battery Charging through Grid- Software setting
2.148	19.8	7.9	4.2	Excessive Battery Charging through Grid- Software setting
2.154	19.4	7.8	4.1	Excessive Battery Charging through Grid- Software setting
2.154	19.1	7.8	4.2	Excessive Battery Charging through Grid- Software setting
2.154	19	7.8	4.2	Excessive Battery Charging through Grid- Software setting
2.154	18.8	7.4	4.4	Excessive Battery Charging through Grid- Software setting
2.177	36.7	0.4	-8.8	Problem with grid export settings
2.177	37.1	0.4	-9.2	Problem with grid export settings
2.177	35.5	0.4	-8.6	Problem with grid export settings
2.177	42	0.4	-8.8	Problem with grid export settings
2.176	40.7	0.4	-8.7	Problem with grid export settings
2.177	33.5	0.4	-8.4	Problem with grid export settings
2.177	32.7	0.4	-8.1	Problem with grid export settings
2.319	88.1	4.3	5.2	Dusty Panels/MPPT/ Shadows/
2.324	88.2	4.3	6.1	Dusty Panels/MPPT/ Shadows/
2.32	88.2	4.7	6.9	Dusty Panels/MPPT/ Shadows/
2.327	88.3	4.7	8.1	Dusty Panels/MPPT/ Shadows/
2.321	88.4	4.5	9.3	Dusty Panels/MPPT/ Shadows/
2.324	88.5	4.6	10	Dusty Panels/MPPT/ Shadows/
2.332	88.6	4.5	9.4	Dusty Panels/MPPT/ Shadows/

Fig 7. Snapshot of the obtained database

4. Training

4.1 Feature Engineering:

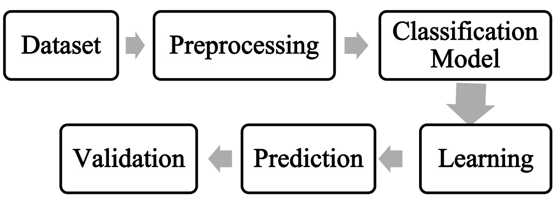

Fig 8. Various phases of Machine Learning

Preprocessing Data. The data in our dataset are on different scales. It is essential to scale them, or else the attributes with a larger number of digits will dominate the algorithm reducing the accuracy of the final output. Data Mining performs better and faster when the input variables are on a moderately comparative scale.

For this purpose, we have used the Standard Scaler method to scale data on the same scale. It facilitates the transformation of data having a zero mean-variance and one standard deviation. It does so by subtracting the mean from every data and then scaling it by dividing all the values by the standard deviation. The standard score is found with the help of raw score x, mean, and standard deviation:

$$\mathbf{z = (x - u) / s} \qquad \mathbf{(1)}$$

Here,
x = raw score;
u = mean of the data;
s = standard deviation of the data

Encoding Categorical Variables. Most of the algorithms are unable to handle categorical variables. So we convert them to numerical values. Our dataset contains categorical Data, i.e., fault. We have used One Hot Encoder, which maps each value to a vector 1 or 0, which denotes the presence or absence of that feature.

battvpc	solar_irr	load_kW	Src_kW	Fault
2.168	81.5	0	0	1
2.167	81.5	0	0	1
2.167	81.8	0	0	1
2.167	82.5	0	0	1
2.166	83.2	0	0	1
2.167	83.6	0	0	1
2.168	84.3	0	0	1
2.148	19.7	7.8	4.2	2
2.149	19.7	7.7	4.1	2
2.148	19.8	7.9	4.2	2
2.154	19.4	7.8	4.1	2
2.154	19.1	7.8	4.2	2
2.154	19	7.8	4.2	2
2.154	18.8	7.4	4.4	2
2.177	36.7	0.4	-8.8	3
2.177	37.1	0.4	-9.2	3
2.177	35.5	0.4	-8.6	3
2.177	42	0.4	-8.8	3
2.176	40.7	0.4	-8.7	3
2.177	33.5	0.4	-8.4	3
2.177	32.7	0.4	-8.1	3
2.319	88.1	4.3	5.2	4
2.324	88.2	4.3	6.1	4
2.32	88.2	4.7	6.9	4
2.327	88.3	4.7	8.1	4
2.321	88.4	4.5	9.3	4
2.324	88.5	4.6	10	4
2.332	88.6	4.5	9.4	4

Fig 9. Dataset after Encoding

4.2 Classification Models

Classifying a given collection of data in groups is known as classification. It approximates the mapping function from input variables to discrete output values. Identifying the data category was the key objective. It begins by training the classifier. The classifier uses the fit method. This type of learning is called supervised learning.

Decision Tree- a binary tree that recursively separates the data set until we are left with pruned leaf nodes or data with only one class type. Decision nodes and leaf nodes are the two types of nodes. The decision node contains a split decision, and the leaf node assists in determining the class of the new data point. A decision tree is a collection of if-then rules. These rules can be easily converted from decision trees. The model must learn which features to take and optimally split the data. The model chooses to split, which maximizes information gain.

$$Gain = Info - Info_A \tag{2}$$

Here,
Info = original information required
$Info_A$ = new information required after partitioning on A

Random Forest is a typical ensemble method that combines the findings from various predictors. Additionally, the random forest uses the bagging technique, which enables each tree to be trained on a random sample of the initial dataset and obtain the consensus of the trees.

Support Vector Machine- It is a method for dividing data into linear and nonlinear categories. It helps solve both classification and regression types of problems. In classification, SVM separates the classes using a hyperplane and marginal distance. The hyperplane divides the classes. The larger the marginal space, the more generalized the model

$$M(x^i, x^j) = f(x^i)^r g(x^j) \tag{3}$$

KNN- Classifiers compare a given test tuple to similar training tuples, a method known as learning by analogy. We have to consider the "k" value. It uses two parameters, i.e., Euclidian distance and Manhattan distance.

$$\text{Euclidean distance} = \sqrt{\sum_{i=1}^{n}(pi - qi)^2} \tag{4}$$

Data sets from the fuzzy fault diagnosis system are used to train the system using various classification algorithms, and results are compared in Fig 8.

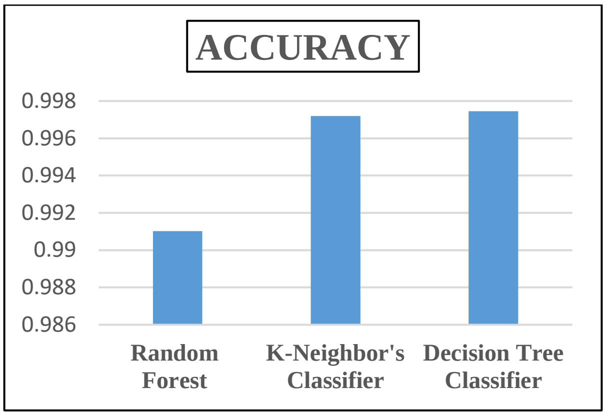

Fig 10. Comparison of Accuracy

Furthermore, 297082 faulty data pattern has been analyzed by using a k-means clustering algorithm

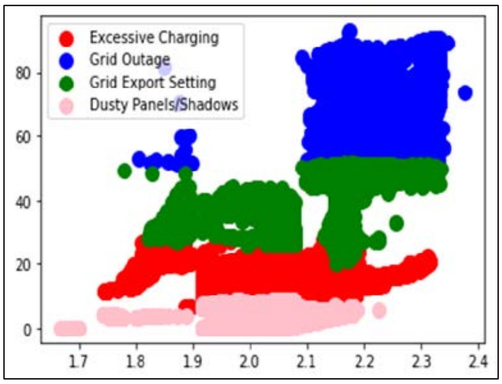

Fig 11. Clustering Diagram

The result of the analyzed data is as follows:

Fault Type	Number of Data Patterns
Grid Outage	238728
Excessive Charging	16006
Grid Export Settings	40289
Dusty Panels/Shadows	35

The results show that in this study, the frequent fault is "Grid Outage/ battery charging through the grid is disabled." Fault due to "Shadow/Dusty Panels" occurrence is very low as most of the P.V. panels are rooftop installed.

5. Conclusion

The proposed hybrid-based approach-based Fault Diagnosis System can detect and classify the fault as per the framed rules. The Fault Diagnosis System will detect faults and suggest corrective actions. The obtained results are quite convincing when the fault diagnosis system is tested both in "offline" and real-time "online" mode. The decision to implement the fault diagnosis system through fuzzy logic is also correct, as, for dynamic systems, no mathematical model can be framed. Thus, the objective of implementing a fuzzy and machine learning approach-based fault diagnosis system is successful as fuzzy logic will deal with uncertainty associated with irradiance. Machine learning in a real environment will become faster as training data is generated by a fuzzy model where uncertainties are taken care of, and data sets are generated. In this study, only four-fault are considered, but more faults can be incorporated to make the fault diagnosis expert system more robust, thus preventing fault from failure. The results from the fuzzy system are used to train the model to make the system faster. The comparison table shows that the results are nearly 99 percent accurate. Moreover, using the clustering method, the maximum occurrence of any fault can be calculated, thus helping administrators focus on that particular fault.

References

1) Ma, Deyin, Yanchun Liang, Xiaoshe Zhao, Renchu Guan, and Xiaohu Shi. "Multi-BP expert system for fault diagnosis of powersystem." *Engineering Applications of Artificial Intelligence 26*, no. **3** (2013):937-944. https://doi.org/10.1016/ j.engappai. 2012.03.017
2) Shin, Won, Jeongyun Han, and Wonjong Rhee. "AI-assistance for predictive maintenance of renewable energy systems." *Energy* **221** (2021): 119775. https://doi.org/10.1016/j.energy.2021.119775
3) Wang, Zhanwei, Lin Wang, Yingying Tan, and Junfei Yuan. "Fault detection based on Bayesian network and missing data imputation for building energy systems." *Applied Thermal Engineering* **182** (2021): 116051. https://doi.org/10.1016/j.applthermaleng. 2020.116051.
4) Hocine, Labar, Kelaiaia Mounia Samira, Mesbah Tarek, Necaibia Salah, and Kelaiaia Samia. "Automatic detection of faults in a photovoltaic power plant based on the observation of degradation indicators." *Renewable Energy* **164** (2021): 603-617.. https://doi.org/ 10.1016/ j.renene. 2020.09.094
5) Abu-Rub, Omar H., Amin Y. Fard, Muhammad Farooq Umar, Mohsen Hosseinzadehtaher, and Mohammad B. Shadmands. "Towards intelligent power electronics-dominated grid via machine learning techniques*." IEEE Power Electronics Magazine 8*, no. **1** (2021): 28-38., doi: 10.1109/MPEL.2020.3047506.
6) Nan, Cen, Faisal Khan, and M. Tariq Iqbal. "Real-time fault diagnosis using knowledge-based expert system." *Process safety and environmental protection 86*, no. **1** (2008): 55-71., https://doi.org/ 10.1016/ j.psep.2007.10.024.
7) Heung-Jae Lee, Deung-Yong Park, Bok-Shin Ahn, Young-Moon Park, Jong-Keun Park and S. S. Venkata, "A fuzzy expert system for the integrated fault diagnosis," in *IEEE Transactions on Power Delivery*, **vol. 15**, no. 2, pp. 833-838, April (2000), doi: 10.1109/61.853027.
8) Monsef, Hassan, Ali Mohammad Ranjbar, and S. Jadid. "Fuzzy rule-based expert system for power system fault diagnosis." *IEE Proceedings-Generation, Transmission and Distribution 144*, no. **2** (1997): 186-192. https://doi.org/10.1049/ip-gtd:19970799
9) Zhu, Honglu, Haizheng Wang, Dahai Kang, Lizhong Zhang, Lingxing Lu, Jianxi Yao, and Yang Hu. "Study of joint temporal-spatial distribution of array output for large-scale photovoltaic plant and its fault diagnosis application." *Solar Energy* **181** (2019): 137-147. Energy. 181.137-147. 10.1016/ j.solener.2019.01.083.
10) A. Houssein, N. Héraud, I. Souleiman and G. Pellet, "Monitoring and fault diagnosis of photovoltaic panels," *IEEE International Energy Conference*, (2010), pp. 389-394, doi: 10.1109/ ENERGYCON .2010.5771711.
11) Dhimish, Mahmoud, and Violeta Holmes. "Fault detection algorithm for grid-connected photovoltaic plants." *Solar Energy* **137** (2016): 236-245. https://doi.org/10.1016/j.solener. 2016.08.021.
12) Ali, Mohamed Hassan, Abdelhamid Rabhi, Ahmed El Hajjaji, and Giuseppe M. Tina. "Real time fault detection in photovoltaic systems." *Energy Procedia* **111** (2017): 914-923. https://doi.org/ 10.1016/j.egypro.2017.03.254.
13) Chine, W., A. Mellit, Vanni Lughi, A. Malek, Giorgio Sulligoi, and A. Massi Pavan. "A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks." *Renewable Energy* **90** (2016): 501-512. https://doi.org/10.1016/j.renene. 2016.01.036
14) Silvestre, Santiago, Aissa Chouder, and Engin Karatepe. "Automatic fault detection in grid connected P.V. systems." *Solar energy* **94** (2013): 119-127. https://doi.org/10.1016/j.solener. 2013.05.001
15) Berawi, Mohammed Ali, Sutan Akbar Onggar Siahaan, Perdana Miraj, and Pekka Leviakangas. "Determining the prioritized victim of earthquake disaster using fuzzy logic and decision tree approach." *Evergreen* ***7(2):*** 246-252, (2020) doi.org/10.5109/4055227
16) Gupta, Vivek, and Arvind Jayant. "A novel hybrid MCDM approach followed by fuzzy DEMATEL-ANP-TOPSIS to evaluate Low Carbon Suppliers." *Evergreen* ***8(3)***, 544-555, (2021). https://doi.org/10.5109/4491640.
17) Zohedi, Fauzal Naim, Mohd Shahrieel Mohd Aras, Hyreil Anuar Kasdirin, and Mohd Bazli Bahar. "A new tuning approach of single input fuzzy logic controller (siflc) for remotely operated vehicle (rov) depth control." *Evergreen* ***8(3)***, 651-657, (2021) https://doi.org/10.5109/4491657
18) Choirunisa, Ikhtiar, Fitrian Imaduddin, Elliza Tri Maharani, Gigih Priyandoko, and Saiful Amri Mazlan. "M.R. Damper Modeling using Gaussian and Generalized Bellof ANFIS Algorithm." *Evergreen* ***8(3)*** 673-685, (2021). https://doi.org/10.5109/ 4491844
19) Sagar, Laxmi Kant, and D. Bhagwan Das. "Fuzzy Expert System for Determining State of Solar Photo Voltaic Power Plant Based on Real-Time Data." *EVERGREEN Joint Journal of Novel Carbon Resource Sciences & Green Asia Strategy*, **Vol. 9(3)**, 870-870 (2022). https://doi.org/10.5109/4843118
20) Yang, Chunzhen, Jingquan Liu, Yuyun Zeng, and Guangyao Xie. "Real-time condition monitoring and

fault detection of components based on machine-learning reconstruction model." *Renewable Energy* **133** (2019): 433-441. https://doi.org/10.1016/ j.renene.2018.10.062

21) Bode, Gerrit, Simon Thul, Marc Baranski, and Dirk Müller. "Real-world application of machine-learning-based fault detection trained with experimental data." *Energy* **198** (2020): 117323. https://doi.org/ 10.1016/ j.energy.2020.117323.
22) Yi Lu Murphey, M. A. Masrur, ZhiHang Chen and Baifang Zhang, "Model-based fault diagnosis in electric drives using machine learning," *IEEE/ASME Transactions on Mechatronics*, vol. **11**, no. 3, pp. 290-303, (2006), doi: 10.1109/TMECH.2006.875568..
23) Y. Zhao, R. Ball, J. Mosesian, J. -F. de Palma and B. Lehman, "Graph-Based Semi-supervised Learning for Fault Detection and Classification in Solar Photovoltaic Arrays," *IEEE Transactions on Power Electronics*, vol. **30**, no. 5, pp. 2848-2858, (May 2015), doi: 10.1109/TPEL.2014.2364203.
24) H. Momeni, N. Sadoogi, M. Farrokhifar and H. F. Gharibeh, "Fault Diagnosis in Photovoltaic Arrays Using GBSSL Method and Proposing a Fault Correction System," *IEEE Transactions on Industrial Informatics*, vol. **16,** no. 8, pp. 5300-5308, Aug. (2020), doi: 10.1109/TII.2019.2908992.
25) Seongmin Heo, Jay H. Lee, "Fault detection and classification using artificial neural networks, *IFAC-PapersOnLine*, **Volume 51**, Issue 18, (2018), Pages 470-475, ISSN 2405-8963. https://doi.org /10.1016/j.ifacol. 2018.09.380.
26) Bousdekis, Alexandros, Katerina Lepenioti, Dimitris Apostolou, and Gregoris Mentzas. "A review of data-driven decision-making methods for industry 4.0 maintenance applications." Electronics 10, no. 7 (2021): 828. https://doi.org/10.3390/ electronics 10070828
27) Fahim, P., Vaezi, N. (2022). Data-Driven Techniques for Optimizing the Renewable Energy Systems Operations. In: Fathi, M., Zio, E., Pardalos, P.M. (eds) Handbook of Smart Energy Systems. Springer, Cham. https://doi.org/10.1007/978-3-030-72322-4_60-1
28) Kuo, R. J., C. W. Hsu, and Y. L. Chen. "Integration of fuzzy ANP and fuzzy TOPSIS for evaluating carbon performance of suppliers." *International Journal of Environmental Science and Technology* 12, no. **12** (2015): 3863-3876.
29) Aras, Mohd Shahrieel Mohd, Shahrum Shah Abdullah, Siti Yasmin Binti Othman, Marizan Sulaiman, Mohd Farriz Basar, Mohd Khairi Mohd Zambri, and Muhammad Nizam Kamarudin. "FUZZY LOGIC CONTROLLER FOR DEPTH CONTROL OF UNDERWATER REMOTELY OPERATED VEHICLE." *Journal of Theoretical & Applied Information Technology* 91, no. **2** (2016).
30) Liao, Linxia, and Felix Köttig. "A hybrid framework combining data-driven and model-based methods for system remaining useful life prediction." *Applied Soft Computing* **44** (2016): 191-199. https://doi.org/10.1016/j.asoc.2016.03.013.
31) Sheikh, Shehzar Shahzad, Mahnoor Anjum, Muhammad Abdullah Khan, Syed Ali Hassan, Hassan Abdullah Khalid, Adel Gastli, and Lazhar Ben-Brahim. "A battery health monitoring method using machine learning: A data-driven approach." *Energies* 13, no**.** **14** (2020): 3658. doi:10.3390/en13143658.
32) Hui, Xianfei, Baiqing Sun, Hui Jiang, and Indranil SenGupta. "Analysis of stock index with a generalized BN-S model: an approach based on machine learning and fuzzy parameters." *Stochastic Analysis and Applications* (2022): 1-20. https://doi.org/10.1080/ 07362994.2022.2094960.
33) Fan, Z., Chiong, R. & Chiong, F. A fuzzy-weighted Gaussian kernel-based machine learning approach for body fat prediction. *Appl Intell* **52**, 2359–2368 (2022). https://doi.org/10.1007/s10489-021-02421-3