Real Time Analysis of Faults in Power-Grid Integrated PV Systems Based Upon Sensor Data Analysis

Sonali Negi[1],Dr Bhupal Singh[2], Dr S. S. Dhillon[3]
[1]Research Scholar, Ajay Kumar Garg Engineering College
[2,3]Professor, Ajay Kumar Garg Engineering College

Article History: Received: 11 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published Online: 4 June 2021

ABSTRACT
With the increase in the attention towards renewable sources of energy, the integration of solar PV Systems with Power Grids is increased substantially. This is of particular importance in those areas which receives good temperature and sunlight for most of the days in the year, like the states of northern India. As the power grid is catering the power need of a number of industries including the critical ones like healthcare and production, it is critical to have an accurate forecast of the solar energy output. The parameter that affects the output to greatest extent is the failure rate. In this research paper, high frequency sensor data is analyzed for Fault Detection (FD). Data consisting of 2.2X106 measurements from a GPV (Grid Connected PV), both under Maximum PPT (MPPT) and Intermediate PPT (IPPT) modes is considered. The seven types of faults which are considered in the dataset include:open circuit, voltage sags, partial shading, inverter, current feedback sensor, and MPPT/IPPT controller in boost converter faults. The regression model isproved to be computationally efficient and very accurate for successful FD under large temperature and irradiancevariations with noisy measurements.

Keywords: Solar PV, Power Grid integration, Failure Rates, Data Analysis

1Introduction

It is one of the major concerns of the modern world to start moving towards renewable sources, thereby replacing the conventional sources of energy like fossil fuels. A large number of studies indicate the ill effects that had already been created, with the increasing consumption of fossil fuels, and much more would happen if it is not controlled now. Climate change, global warming, acid rains, etc., are the consequence that had happened with the increasing consumption of conventional energy sources. Undoubtedly, the solar energy is considered as the most promising solution as a renewable source of energy. It is estimated that in the year 2018, the earth received approximately 1.2 × 1017 kWh ofsolar energy, but the estimated energy consumption in the same year was only 1.6 × 1011 kWh [1]. Thus, the same can be harnessed to a much greater extent to meet the demands of the modern world. This incredible finding means that the solar energy matches the load requirement by 12500 times per minute. Recent studies estimate that in the next two years, the share of photovoltaic power generation will reach to 23 percent of the overall power generation in the world. Most of the solar PV sites are increasing connected through power grids so that the overall generation and transmission can be managed in a proper manner [2,3].

In this research, the focus has been made over the regression analysis of reliability of the solar PV systems from grid connected sensor data. Section 2 outlines the literature review overtraditional mechanisms usually considered as the causes of the failure of the PV Systems.In traditional approaches, considering the range of probabilities of failure of devices, one can infer the cumulative reliability analysis. The traditional models used for failure detection uses the MTTF (Mean Time to Failure) of the individual components. However, there is a significant difference in the failure parameters of PV components as compared to their counterparts in microelectronics. Most of the PV failure relates to the manual steps in manufacturing that corresponds to steps done manually. It also depends upon the quality of the installation, packaging and other aspects [4-7]
A very few researches had focused on real time data analysis of line voltage/current/phase with a machine learning classifier for detection of faults.In this research, the binary classification using a logistic regression model over the fault data is framed and the relationship of independent variables- including voltage, current and phase is investigated over Fault / Normal output labels.

2. Common Failure Mechanisms in PV Systems

Generating detailed reliability for different types of components under different deployment conditions is a difficult task for PV Systems. However, a stochastic modeling can be performed which can fairly predict the failure rate and the consequences over actual power production. The common failure mechanism considered in this analysis is:

1. Broken Interconnects
2. Broken Cells in Panels
3. Bypass Diodes / Hot Spots
4. Corrosion
5. Delamination
6. Encapsulant Discoloration
7. Junction Box Failure.

The main cause of broken interconnect is due to thermo-mechanical impact. Robustness against broken interconnect requires rigorous testing. This is majorly dependent on environmental conditions / climate of the region of deployment. This failure mechanismcan be identified as dark regions in the electroluminescence image where the failed interconnect wouldotherwise be collecting carriers.
Broken or cracked cells are usually the results of storms, heavy wind and long term mechanical stress. The exact reason which is applicable in all cases under consideration is difficult to outline. However, this is one major parameter which substantially contributes towards the failure of the PV panel.
Failed bypass diodes can hinder the performance of PV modules. Bypass diodes are placed in parallelwith PV cells in the opposite direction of the cells' p-n junctions. Their purpose is to dissipate the reversebias current and voltage stress that otherwise occurs when a subset of cells are shaded orunderperforming. Bypass diodes that fail short can be identified using electroluminescence. If thebypass diode fails open, the reverse bias stress from partial shading can lead to significant heatdissipation and the formation of hot spots [8]. These hot spots can be identified through IR imaging.Prolonged operation in this condition can result in further solder or back-sheet damage or arc faults.

One of the most prominent reasons for corrosion is the moisture aggregation over the surfaces of the component.
Delamination is a common effect which happens when the product is exposed to sunlight for large time durations. With delamination, the corrosion increases with a rapid rate.It plays an especially important role in thinfilm photovoltaic where PV cell materials are susceptible to corrosion [9]. This can be identified through visual inspection.
One of the primary cause of degradation in the efficiency is solar PV panels is encapsulate discoloration. The primary purpose of encapsulate is the same as that in microelectronics, that is, to prevent the PV cell from moisture. With this effect, there might not be the sudden failure but the efficiency of the PV cell deteriorates over time.
Junction Boxes are usually placed at the back of the panel and provide connection between PV module cells and the PV System. These boxes are also susceptible to moisture and corrosion. Degradation of junction boxes is critical because there is a presence of high potential inside these boxes as compared to the individual panels.

2.1 Traditional Reliability model of Grid Connected PV Systems

Reliability, availability, andmaintainability (RAM) is an engineering tool used to address operational and safety issues of systems.It aims to identify the weakest areas of a system which will improve the overall system reliability.Generally, reliability is defined as the probability of system, subsystem, or even sub-assembliesto perform its required function adequately. The reliability function of a system is the probability ofsuccessfully operating the system within a given time, t. The reliability or survivor function equationfor a system can be written as:

$$R(T) = P(T > t)$$

The cumulative distribution function (CDF), denoted F(t), is called failure probability orunreliability. It interprets the probability of the system’s success, which can be given by:

$$F(t) = 1 - R(T) = P(T < t)$$

The probability density function (PDF), denoted f(t), indicates the distribution of the failure over the entire time range. Equations (1) and (2) can be expressed with the density function f(t) as:

$$R(T) = \int_{t}^{\infty} f(t)dt$$

$$F(T) = \int_{-\infty}^{t} f(t)dt$$

The mean time to failure (MTTF) for the sub-assembly, which expresses the expected life for the sub-assembly, represents the most common method for specifying reliability of non-repairable items.
The solar-PV systems are complex and contain a large number of sub-assemblies that may beconnected in series, in parallel or even a combination of series and parallel. When the sub-assembliesconnected in series, the overall system will be interrupted in case of failure of one sub-assembly. On theother hand, all subassemblies must fail in order to interrupt the overall system in the parallel system.
According to Boolean techniques, the reliability performance for a non-repairable system containsindependent series n subassemblies can be calculated by:

$$R_{system} = \prod_{i=1}^{n} R_i$$

where R_i is the reliability of the sub-assembly i.
If the system contains x series units with M parallel subassemblies, the system reliability can be obtained using:

$$R_{system} = 1 - (1 - R^x)^M$$

3. Proposed Model

The Grid-connected PV System Faults (GPVS-Faults) data is used for the detection of faults in a PV micro-grid system. The data set is provided by [10], in the form of CSV files in which the files are labeled as “Fxy”, where:

x∈{0,1,…,7} represents the fault scenario:

'0' is a fault-free experiment.

'1',…,'7' are the 7 types of faults.

y∈{'L','M'} is the operation mode:

'L' is Limited power mode (LPPT)

'M' is Maximum power mode (MPPT)

e.g. “F4M” is fault F4 in MPPT mode, “F1L” is fault F1 in IPPT mode.
The seven different types of faults are : open circuit, voltage sags, partial shading, inverter,current feedback sensor, and MPPT/IPPT controller in boost converter faults.
Each data file includes the following columns:

Time: Time in seconds, average sampling T_s=9.9989 μs.
Ipv: PV array current measurement.
Vpv: PV array voltage measurement.
Vdc: DC voltage measurement.
ia, ib, ic: 3-Phase current measurements.
va, vb, vc: 3-Phase voltage measurements.
Iabc: Current magnitude.
If: Current frequency.
Vabc: Voltage magnitude.
Vf: Voltage frequency.

The proposed model for fault identification is depicted as shown in the figure 3.1

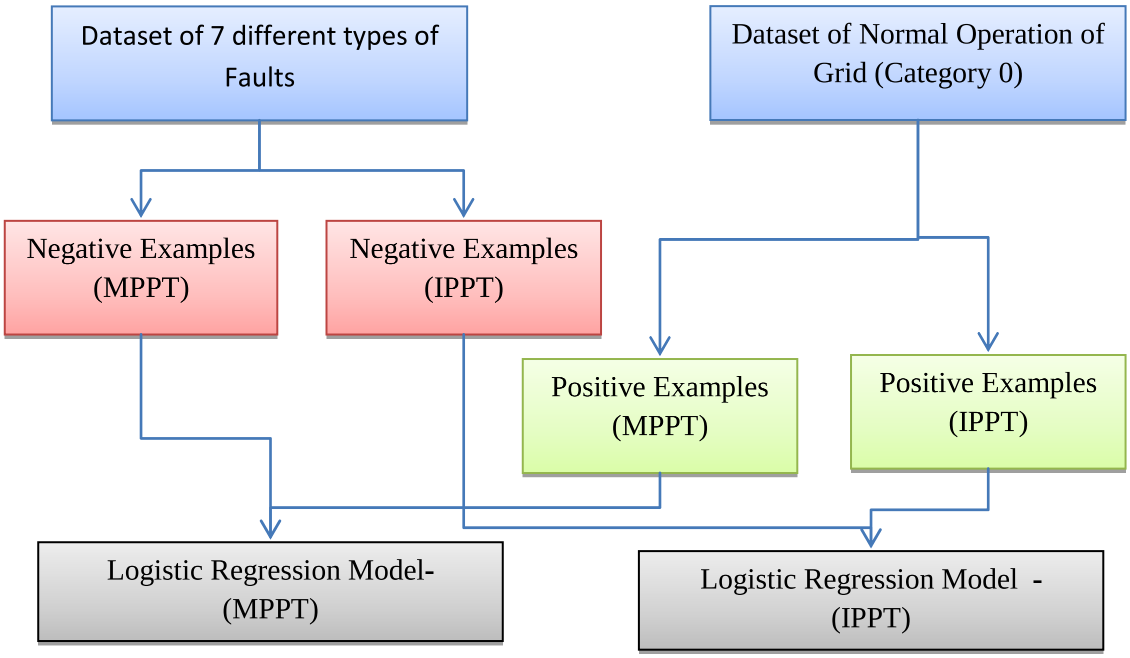

Fig 3.1 Proposed Real Time fault Detection Model for GPV

4. Results

The characteristic of 13 different parameters for fault and non-fault types of cases is depicted in terms of Box and whisker plot in figure 4.1. The plot consists of box in which the middle 50% of the data values exists. The upper and the lower horizontal lines indicate the upper 25% and the lower 25% of the entries in the dataset. The values outside these horizontal lines indicate the outliers. The solid line in between the box indicates the mean values of the parameter under consideration.

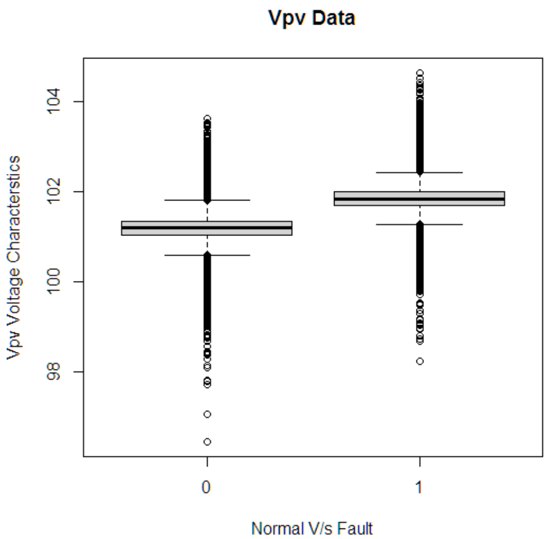

Fig 4.1.1 Vpv Box Plot for Normal(0) and Fault(1) States

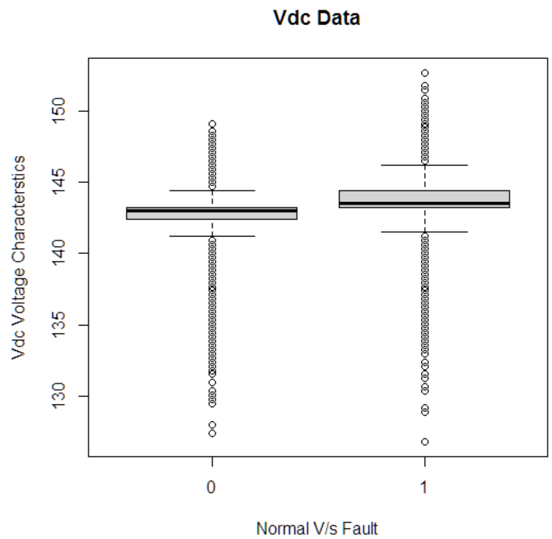

Fig 4.1.2 Vdc Box Plot for Normal(0) and Fault(1) States

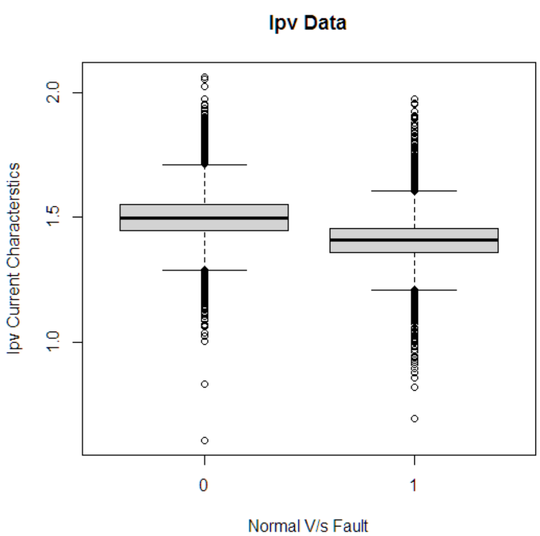

Fig 4.1.3 Ipv Box Plot for Normal(0) and Fault(1) States

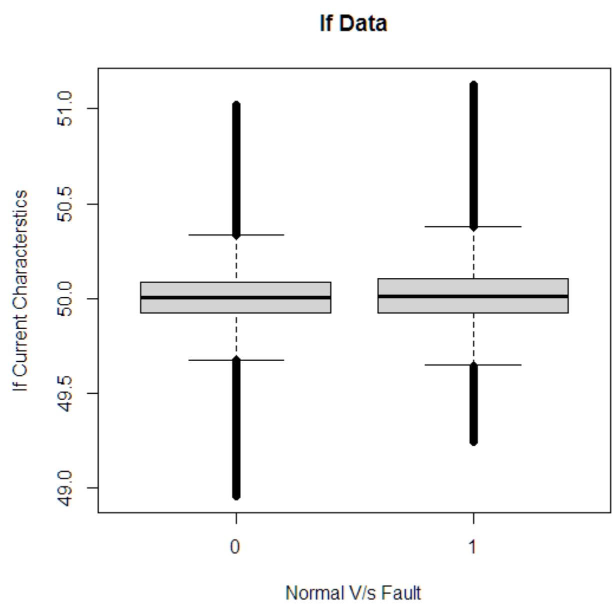

Fig 4.1.4 If Box Plot for Normal(0) and Fault(1) States

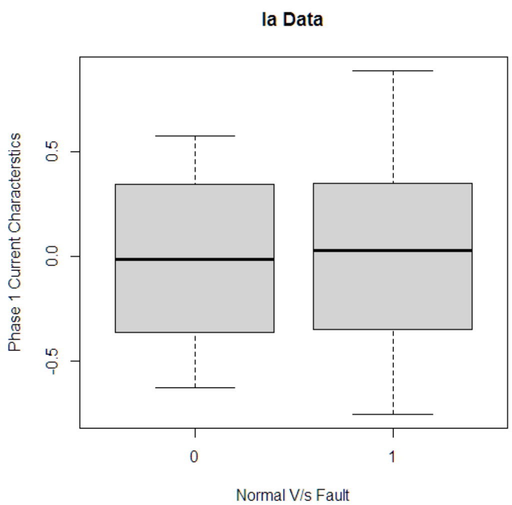

Fig 4.1.5 Phase “a” current Box Plot for Normal(0) and Fault(1) States

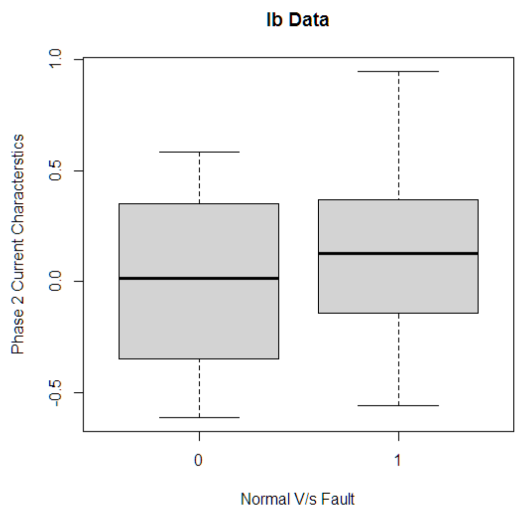

Fig 4.1.6 Phase “b” current Box Plot for Normal(0) and Fault(1) States

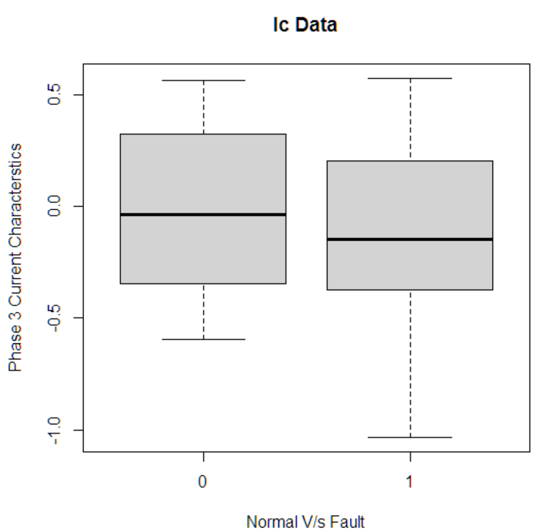

Fig 4.1.7 Phase “c” current Box Plot for Normal(0) and Fault(1) States

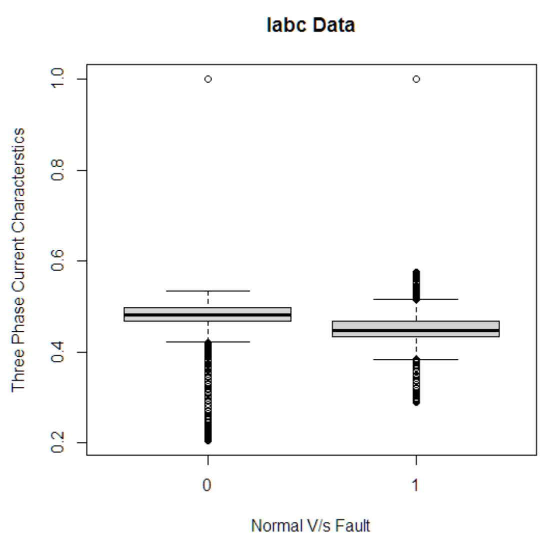

Fig 4.1.5 Aggregate of “a-b-c” current Box Plot for Normal(0) and Fault(1) States

The above plot tabulates the values of most significant parameters under MPPT and IPPPT modes. As indicated in the figure 4.1.1 (a), there is significant difference in the mean values of Vpv under fault and normal operations, which indicates that the parameter has significant impact over the values. The same can also be concluded from hypothesis testing wherein it can be tested with p values under 5% confidence interval. Other parameters which are having significant difference in mean values under fault and no-fault condition includesIpv and Iabc. The parameters vdc, If, Ia, Ib and Ic does not contribute much to the machine learning model and can be skipped safely.

The confusion matrix of the logistic regression analysis for all the 7 types of faults is shown pictorially as shown in the figure 4.2. The similarly between actual and predicted similarity is shown in green while the disagreements between actual and predicted are shown in red. One can view the components of the graph in clockwise direction as : (1) True Positives (2) False Positives (3) True Negatives and (4) False Negatives.

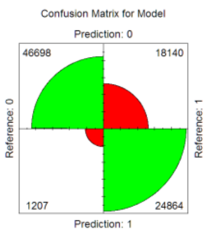

Fig 4.2.1: Confusion Matrix-F1

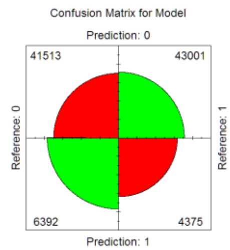

Fig 4.2.2: Confusion Matrix-F1

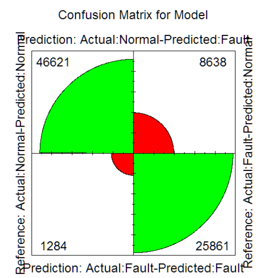

Fig 4.2.3: Confusion Matrix-F1

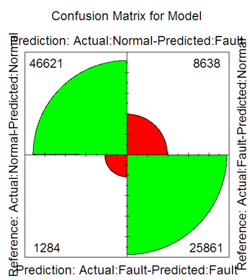

Fig 4.2.4: Confusion Matrix-F4

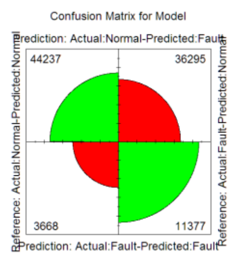

Fig 4.2.5: Confusion Matrix-F5

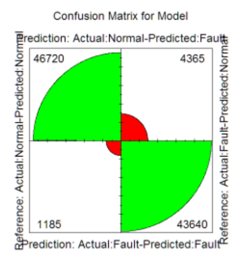

Fig 4.2.6: Confusion Matrix-F6

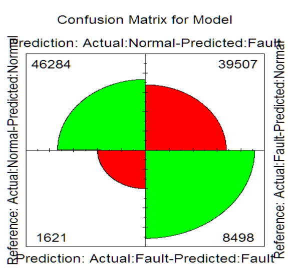

Fig 4.2.7: Confusion Matrix-F7

The logistic regression model is developed over the entire dataset has the following summarization for both MPPT and IPPT scenarios.

Classification Accuracy		
	IPPT	MPPT
Fault-1	93	94
Fault-2	95	94
Fault-3	96	96.5
Fault-4	95	94.5
Fault-5	93	93.5
Fault-6	94	95
Fault-7	95	96

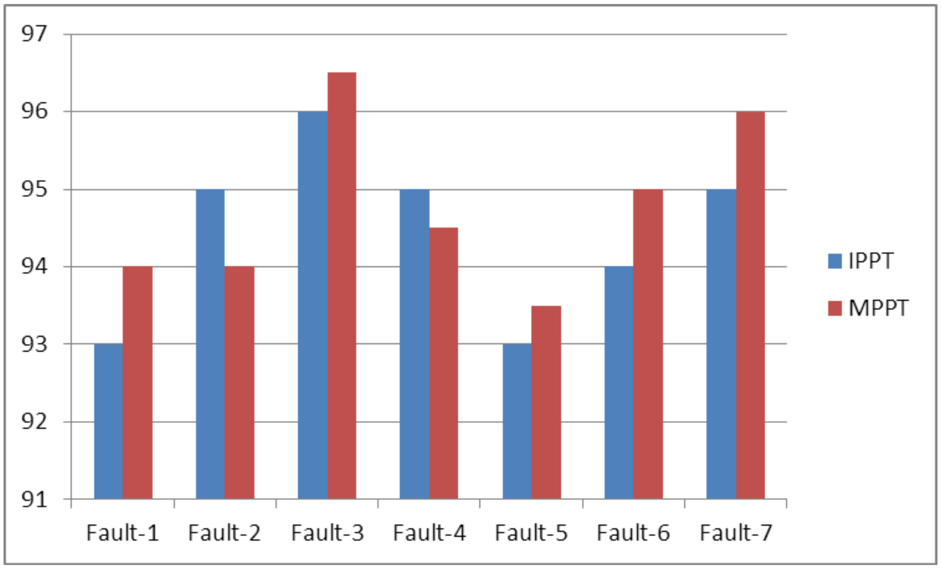

Figure 4.3 Accuracy of the logistic regression model over the seven parameters w.r.t IPPT and MPPT values.

Conclusion and Future Scope

It turns out that the classification accuracy for faults using linear regression model over MPPT and IPPT data gives fairly good accuracy. However one can uses other machine learning models also over the dataset from sensor-net or SCADA systems. As the future scope of the work, we shall be considering the ensemble learning techniques in which the results from different models are critically analyzed and the one with minimum error rate is considered for subsequent processes.

REFERENCES

[1] C. Pascal and C. Morgan, "Global Energy Trends a New Historic High in Energy Consumption and CO2 Emissions," Enerdata-2018.

[2] M. Drechsler, J. Egerer, M. Lange, F. Masurowski, J. Meyerhoff, andM. Oehlmann, "Efficient and equitable spatial allocation of renewable power plants at the country scale," *Nature Energy*, vol. 2, no. 9, p. 17124, 2017.

[3] L. Sussams and J. Leaton, "Expect the unexpected: The disruptive power of low-carbon technology," *Carbon Tracker Initiative (CTI), London, United Kingdom*, 2017.

[4] N. Shahidirad, M. Niroomand, and R.-A. Hooshmand, "Investigation of PV Power Plant Structures Based on Monte Carlo Reliability and Economic Analysis," IEEE Journal of Photovoltaics, vol. 8, no. 3, pp. 825-833, 2018.

[5] FIDES Guide 2009, Edition A: Reliability Methodology for Electronic Systems, 2010.

[6] S. E. De León-Aldaco, H. Calleja, and J. A. Alquicira, "Reliability and mission profiles of photovoltaic systems: A FIDES approach, "IEEE Transactions on Power Electronics, vol. 30, no. 5, pp. 2578-2586, 2014.

[7] S. de Leon, H. Calleja, and J. Mina, "Reliability of photovoltaic systems using seasonal mission profiles and the FIDES methodology," Microelectronics Reliability, vol. 58, pp. 95-102, 2016.

[8] Spagnolo, G. Schirripa, et al. "A review of IR thermography applied to PV systems." *2012 11th International Conference on Environment and Electrical Engineering*. IEEE, 2012.

[9] Carlson, D. E., et al. "Corrosion effects in thin-film photovoltaic modules." *Progress in Photovoltaics: Research and applications* 11.6 (2003): 377-386.

[10] Bakdi, Azzeddine, et al. "Real-time fault detection in PV systems under MPPT using PMU and high-frequency multi-sensor data through online PCA-KDE-based multivariate KL divergence." International Journal of Electrical Power & Energy Systems 125 (2021): 106457.