International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

Intelligent Fault Diagnosis of Solar PV Systems Using Machine Learning Techniques

G Venkata Suresh Reddy[1], Immanuel Anupalli[2], Dr.P.Sudheer[3]
[1]Research scholar, BEST Innovation University, guddetivenkatasuresh.reddy@gmail.com
[2]Professor of EEE, BEST Innovation University, anupalliimmanuel@gmail.com
[3]Professor of EEE, Sreenivasa institute of technology and management studies, psudheer35@gmail.com

Abstract:
Solar photovoltaic (PV) systems need effective fault detection system in order to sustain effective energy generation under different environmental and operating conditions. The proposed study is a data-driven study on fault categorization using machine learning methodologies of detecting faults in PV systems under nonlinear and noisy conditions. Electrical data of PV panels such as current voltage (I V) and power voltage (P V) of the panels were measured in three operating conditions: Healthy, Shading, and Open-Circuit. Important electrical characteristics that included open-circuit voltage, short-circuit current, maximum power point voltage and current, fill factor, and some statistical indicators were determined to describe the behavior of the system.Naive Bayes and k-Nearest Neighbors (KNN) are two supervised learning algorithms that were used to identify faults. All of these models were optimized with the help of hyperparameter tuning and k-fold cross-validation to provide accurate performance. Comparative analysis revealed that the KNN classifier had greater detection capacity along with an accuracy of 95.47, whereas the Naive Bayes classifier had stable and consistent performance and an accuracy of 94.13, which proved to be robust in the nonlinearity and noise in measurements.On the whole, the findings suggest that Naïve Bayes and KNN models can be utilized in real-time detection of PV faults with a help of the easily measurable electrical parameters. The suggested solution is applicable in the practical monitoring of PV systems and it can endorses proactive maintenance measures, enhance reliability in the systems, and minimize energy waste.

Keywords- Solar PV systems, fault detection and classification, machine learning, Naïve Bayes, k-Nearest Neighbors, logistic regression, data-driven diagnosis, nonlinear operating conditions, and predictive maintenance

I.FAULT DETECTION PROCESS IN PHOTOVOLTAIC SYSTEM

We made a PV module model in MATLAB to see how different kinds of faults affect the module's output when it is tested under standard test conditions (STC). The characteristics given to this simulated module were chosen to be the same as those of real PV modules used in the field (Table 1). The power output of a PV module is usually based on the conditions it gets, such as the amount of sunlight it gets and the temperature of the cells, as well as the module's own settings. The PV datasheet[4] from the manufacturer lists these important electrical and thermal parameters. Also, you can add a few more things to the model to make it a more accurate picture of how the module really works.

Table.1.PV Module specification

Parameter	**Value**
Pmax	245
Isc	8.58
Voc	37.80
Impp	7.94

International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

Parameter	Value
Vmpp	30.85
Cells per module	60

Data collection is the initial stage in utilizing AI to identify issues with photovoltaic (PV) panels. Here you may see readings for the temperature and sun irradiation, in addition to the electrical parameters of the PV system, such as power, current, and voltage. After that, the data is cleaned up by eliminating noise, adding missing values, and standardizing its interpretation. Afterwards, the most important characteristics that can indicate PV panel problems are extracted or selected by feature extraction. Afterwards, you can use these features to teach an artificial intelligence model—such as a decision tree, neural network, or support vector machine—to distinguish between normal and broken working situations. The AI model[5] has been programmed to continuously monitor the PV system, scanning for specific types of failures as they occur. So that it may be fixed promptly, it can send alerts or initiate maintenance activities when it finds one. Because of this, the PV installation is now more efficient and dependable.

II. AI-INSPIRED WAYS TO FIND FAULTS IN PV PANELS

A) Naive Bayes Method

Methodologically, making a synthetic dataset is the initial stage. This involves making 4,500 nonlinear and noise-rich I–V curves that show different PV operating conditions. In particular, 2,000 curves show healthy modules, 1,500 show partial shading faults, and 1,000 show open-circuit conditions. Every group has realistic electrical changes: Healthy curves have random nonlinear exponents, Gaussian noise, and degradation effects. Shading curves have much lower short-circuit current (Isc), clear partial-shading steps, and noise. Open-circuit curves have an exponential decay profile with very low Isc. The dataset's overall nonlinearity and randomness come from changes in Isc and Voc, an exponential factor taken from a normal distribution N(1.2,0.52), adding Gaussian noise to current values, and using a stochastic degradation multiplier. This makes sure that the dataset that was created closely matches how PV works in the real world when conditions are different and unpredictable.

Gaussian Naive Bayes classifier trained with 5-fold cross-validation to make sure that the Naive Bayes classification model works well on the whole dataset. The model learns the statistical distribution of the six features for each class in order to handle three types of faults: Healthy, Shading, and Open-Circuit. The algorithm calculates the posterior probability of each class based on the input feature vector during classification, following the Naive Bayes assumption of conditional independence among all features. The classifier uses the relation to mathematically figure out the posterior.

$$P(\text{class} \mid \text{features}) \propto P(\text{class}) \prod_{i=1}^{6} P(x_i \mid \text{class})$$

Where P(class) is the prior probability of each condition, and $P(x_i|\text{class})$ is the Gaussian likelihood of the iii-th feature for that class. The class with the highest posterior probability is picked as the final predicted condition. This makes it possible to quickly and accurately find faults in PV systems[6].

International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

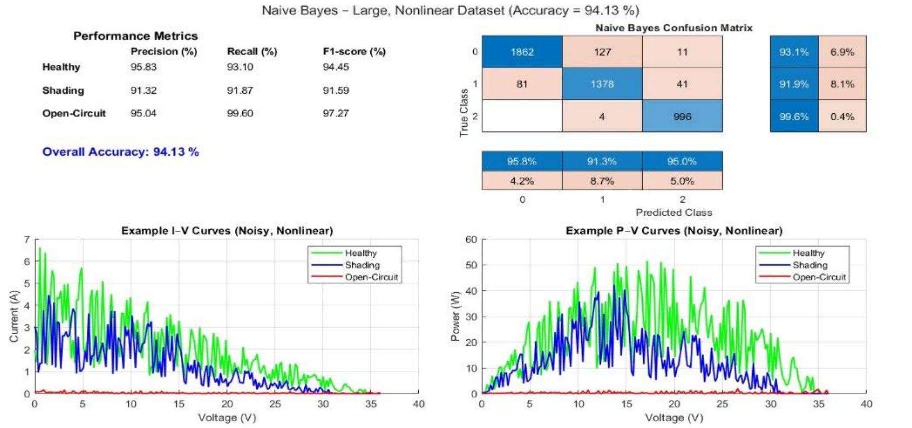
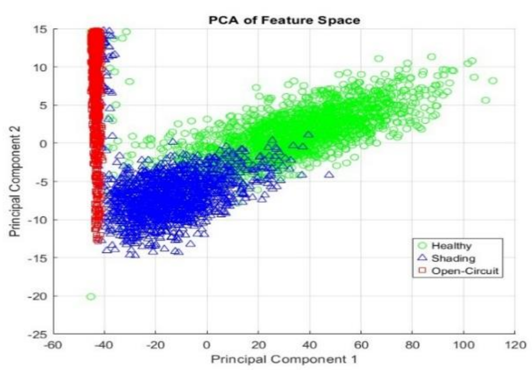


Fig. 1. Performance metrics for a fault detection system based on Naive Bayes

The combined Fig. 1. summarizes both quantitative performance and qualitative examples. The confusion matrix shows that open-circuit faults are found with almost perfect recall, while partial-shading samples show the most confusion between classes, especially with healthy samples. The synthetic noise and nonlinear behaviors used for feature extraction (Voc, Isc, Vmp, Imp, Pmax, FF) are shown in the Example I–V and P–V curves.

The feature set in PCA Fig. 2 shows that it works well but not perfectly for separating things. Open-circuit samples form a separate group (low Isc, different profile), while Healthy and Shading overlap a lot along the main axes. This is why there is a higher rate of confusion between these two groups.



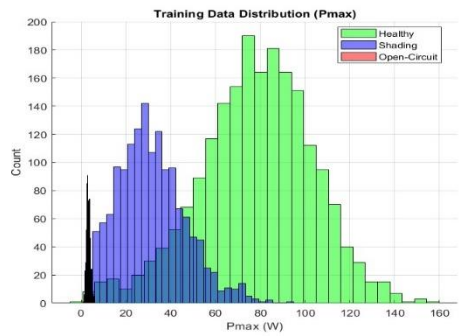


Fig. 2 Analysis of principal components

Fig.3. Distribution of power in training data

The overlaid histograms in Fig. 3 of Pmax clearly show the different ways power is distributed for each operating condition. The maximum power values for healthy modules are the highest, with the most common range being between 70 and 90 W. When shaded, modules usually have Pmax values that peak around 20–50 W, but when they are open-circuit, they always have values that are close to zero. These distributions for each class show that Pmax is a very good way to tell the difference between things. The synthetic dataset (4500 samples) showed that the Naive Bayes classifier was 94.13% accurate overall. The F1 scores for each class were 94.45% for Healthy, 91.59% for Shading, and 97.27% for Open-Circuit. The confusion matrix shows that most mistakes happen between Healthy and Shading. This is because the Pmax and Isc ranges for lightly shaded conditions overlap. PCA and

International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

Voc–Isc scatter plots confirm these results: open-circuit samples create a separate group (with very little Isc), while shading and healthy samples mix in the main component space. The noisy and nonlinear shapes used for training are shown in Example I–V and P–V curves. The open-circuit curves show an exponential current falloff and a Pmax that is close to zero, which explains why the classifier did so well on that class.

B) The K-Nearest Neighbors (KNN) Classification Model

From Fig.4.The KNN classifier is highly good at detecting the Healthy class, as seen by the confusion matrix. This is because it doesn't overlap much with the other classes, which means it has good precision and recall. The Shading class has a somewhat lower recall rate (approximately 92%), which means that a few shaded samples are put in the wrong category. This is most likely due to the fact that healthy and partially damaged curves have very similar features[9]. Conversely, with a rate of around 99.4 percent, the Open-Circuit class boasts the best recall. The reason behind this is its extremely unique behaviour, characterized by practically negligible current values. The model can easily distinguish it from the other classes because of this. Even on a large synthetic dataset with additional noise, the KNN model consistently and effectively performs. Out of all the classes, 95.47 percent are correct. That it is effective in locating PV system problems is evident from this.

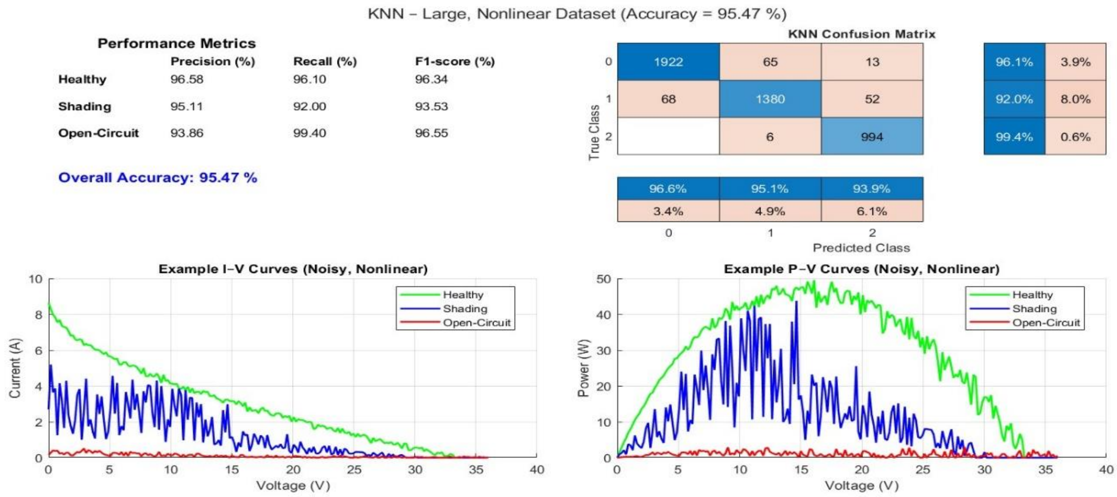


Fig. 4. KNN-based fault detection system performance metrics

International Journal of
AI ELECTRICAL CIVIL AND MECHANICAL ENGINEERING
Peer Reviewed, Referred & Indexed Journal
ISSN: 3070-0434 www.zesterapublications.com Original Research Paper

By displaying the number of samples that were appropriately and incorrectly assigned to each class, the confusion matrix illustrates the efficacy of the categorization. There were a total of 1922 accurately identified samples in the Healthy class; however, 65 were incorrectly labeled as Shading and 13 as

Fig. 5. Distribution of training data

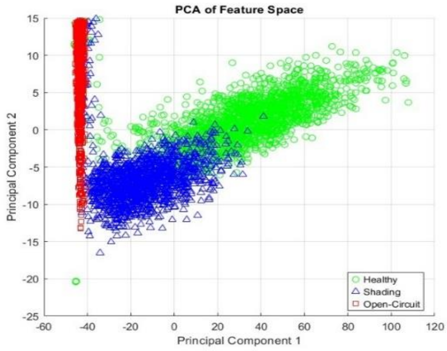


Fig. 6. Analysis of the main components

Open-Circuit. Thirteen hundred and eighty-one predictions were correct in the Shading class. On the other hand, 52 were marked as Open-Circuit and 68 were named Healthy erroneously[10]. The Open-Circuit class did an excellent job of keeping things apart.It made 994 correct predictions, none of which were wrongly labeled as Healthy, and only 6 were wrongly labeled as Shading. This means that the Healthy and Open-Circuit classes are very different from each other in the feature space. The Shading class is harder to work with because of the different shading patterns, which can cause some Healthy curves to overlap with Shading curves. The normalized confusion matrix shows that the model is very reliable because it gives class-wise accuracies of 96.1% for Healthy, 92.0% for Shading, and 99.4% for Open-Circuit. This is in line with the overall metrics table[1].We utilized Principal Component Analysis (PCA) Fig. 6 to combine six features into two primary components that were visible. The Healthy samples are all grouped together with positive PC1 values, which means they all have the same features. Shading samples have a lot of spread and overlap with other classes, which shows how partial shading patterns can modify things. Clustered on the left side with large negative PC1 values, the Open-Circuit examples stand out from the rest of the classes. In most cases, principal component analysis shows that feature extraction is an effective method for classifying input curves[12] for use in machine learning. Figure 5 shows the distribution of the training data across classes in the histogram of maximum power (Pmax). The maximum power that healthy panels can handle varies greatly, ranging from about 60 W to 120 W. These numbers might fluctuate due to noise. Pmax values vary from 10 W to 55 W, indicating a significant difference between healthy panels and shaded samples. Since the Pmax value of open-circuit samples is often around zero, they can be easily categorized into a small subset. Particularly in KNN-based models, the large interclass variation in Pmax highlights the importance of this parameter for classification[13].

Table.1.Comparison of Naïve bayes and KNN methods

Classifier	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Remarks
Naïve Bayes	94.13	H=95.83, S=91.32, O=95.04	H=93.10, S=91.87, O=99.60	H=94.45, S=91.59, O=97.27	Lowest performance; independence assumption fails for PV data
K-Nearest Neighbors	95.47	H=96.58, S=95.11, O=93.86	H=96.10, S=92.00, O=99.40	H=96.34, S=93.53, O=96.55	Good but sensitive to noise; shading misclassifications higher

V. CONCLUSIONS

On comparing two machine-learning classifiers to detect photovoltaic faults, it is clear that different algorithms are sensitive to nonlinear operating conditions to varying extents. The Naïve Bayes classifier was identified as having an accuracy of 94.13, which is relatively low because of the high feature-independence assumption that the coupled electrical nature of PV systems is not well reflected in their feature independence. However, it also exhibited constant performance and resilience when subjected to measurement noise, which is why it is a lightweight and computationally efficient alternative.KNN classifier demonstrated a better ability to capture nonline aspects of I-V and P-V features space and the accuracy of the detection was improved to 95.47%. Nonetheless, its performance was a little impaired in very noisy and partially shaded environments where there is overlap in the class boundaries causing misclassification is sometimes experienced.

In general, the findings show that both Naive Bayes and KNN can be used to successfully differentiate among open-circuit, shaded and healthy operating states of PVs using statistical electrical measurements. Naive bayes is simpler and more robust whereas KNN is more effective in terms of nonlinear discrimination. Future research can be aimed at develops increased robustness in the face of complicated shading conditions with hybrid or ensemble learning techniques.

REFERENCES

[1] A. Skoczek, T. Sample, and E. D. Dunlop, “The results of performance measurements of field-aged crystalline silicon photovoltaic modules,” *Prog. Photovolt. Res. Appl.*, vol. 17, no. 4, pp. 227–240, 2009.

[2] M. Köntges et al., “Review of failures of photovoltaic modules,” *IEA PVPS Task 13 Report*, 2014.

[3] P. Guerriero, F. Di Napoli, S. Daliento, and M. Pecchia, “PV module faults: classification and detection techniques,” *Solar Energy*, vol. 211, pp. 539–551, 2020.

[4] S. Silvestre, A. Boronat, and A. Chouder, “Study of bypass diodes configuration on PV modules,” *Appl. Energy*, vol. 86, no. 9, pp. 1632–1640, 2009.

[5] D. Sera, R. Teodorescu, and P. Rodriguez, “PV panel model based on datasheet values,” in *Proc. IEEE ISIE*, 2007, pp. 2392–2396.

[6] S. Vergura, G. Acciani, V. Amoruso, G. Patrono, and F. Vacca, “Descriptive and inferential statistics for supervising and monitoring the operation of PV plants,” in *Proc. IEEE IEMDC*, 2009.

[7] S. Mekhilef, R. Saidur, and A. Safari, "A review on solar energy use in industries," *Renew. Sustain. Energy Rev.*, vol. 15, no. 4, pp. 1777–1790, 2011.

[8] A. Chouder and S. Silvestre, "Automatic supervision and fault detection of PV systems based on power losses analysis," *Energy Convers. Manage.*, vol. 51, no. 10, pp. 1929–1937, 2010.

[9] A. Safari and S. Mekhilef, "Simulation and hardware implementation of incremental conductance MPPT with direct control method," *IEEE Trans. Ind. Electron.*, vol. 58, no. 4, pp. 1154–1161, 2011.

[10] A. Mellit and S. A. Kalogirou, "Artificial intelligence for photovoltaic systems: A review," *Prog. Energy Combust. Sci.*, vol. 34, no. 5, pp. 574–632, 2008.

[11] J. Bishop, "Computer simulation of the effects of electrical mismatches in photovoltaic cell interconnection circuits," *Solar Cells*, vol. 25, pp. 73–89, 1988.

[12] Y. Zhao et al., "Decision tree–based fault detection and classification in solar PV arrays," in *Proc. IEEE PES General Meeting*, 2012.

[13] G. Makrides, B. Zinsser, M. Norton, and G. E. Georghiou, "Performance of PV systems under different climatic conditions," *Prog. Photovolt.*, vol. 20, pp. 636–648, 2012.