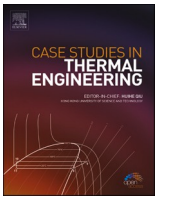
Case Studies in Thermal Engineering 25 (2021) 100980

Contents lists available at ScienceDirect

Case Studies in Thermal Engineering

journal homepage: http://www.elsevier.com/locate/csite



Defining the best-fit machine learning classifier to early diagnose photovoltaic solar cells hot-spots

Mahmoud Dhimish

University of Huddersfield, Laboratory of Photovoltaics, Huddersfield, HD1 3DH, UK

ARTICLE INFO

Keywords:
Photovoltaics
Hot-spots
Machine learning
Artificial intelligence
Classification

ABSTRACT

Photovoltaic (PV) hot-spots is a reliability problem in PV modules, where a cell or group of cells heats up significantly, dissipating rather than producing power, and resulting in a loss and further degradation for the PV modules' performance. Therefore, in this article, we present the development of a novel machine learning-based (ML) tool to diagnose early-stage PV hot-spots. To achieve the best-fit ML structure, we compared four distinct machine learning classifiers, including decision tree (DT), support vector machine (SVM), K-nearest neighbour (KNN), and the discriminant classifiers (DC). Results confirm that the DC classifiers attain the best detection accuracy of 98%, while the least detection accuracy of 84% was observed for the decision tree. Furthermore, the examined four classifiers were also compared in terms of their performance using the confusion matrix and the receiver operating characteristics (ROC).

1. Introduction

Nowadays, photovoltaic (PV) module's reliability and durability became a vital determinant to utilize the leading cause of PV degradation, failure-rates, and mismatching conditions. PV installations often experience partial shading conditions [1], while this would typically create an uneven increase in the cells' temperature, causing what is known by "PV hot-spots" [2]. The rising in the hot-spotted cells' temperature is caused by the reversed biasing of the output current. As proven by Ref. [3], hot-spotted cells will dissipate rather than generate power, while the total loss in the PV system's yield energy is expected to drop by up to15%. Not only shading creates PV hot-spots, but also, it was evident by Ref. [4] that there is a correlation between the existence of PV cracks "snail-trail or micro-cracks" and the presence of the hot-spots.

Most reliable PV technologies in today's market are equipped with bypass diodes, as explained in early 1986 [5]. Unfortunately, multiple studies, including [6,7], prove that bypass diodes do not overcome the PV hot-spotting events.

Current studies such as [8–10] show assuring techniques to avoid hot-spots in the PV modules. Those techniques rely on the switching mechanism of the MOSFETS that are integrated into parallel with the PV modules. They show robust results; however, they do not detect hot-spots in the PV modules; all are dependent on thermal imaging as a pre-stage functionality.

In practice, thermal imaging cameras are not available for nearly all PV user's, domestic-wise. In comparison, commercial PV operators in today's market intend to use thermal drones to detect the hot-spots; this procedure is expensive to operate and requires legal authorisation to do so. Furthermore, a limited number of robust methods can diagnose hot-spots in PV modules, particularly early-stage hot-spots, where the PV module is only affected by either one, two, or at most five hot-spotted solar cells. Therefore, in our study, we present the development of a machine learning-based tool to diagnose early-stage hot-spots. Before moving ahead to the

E-mail address: m.a.dhimish@hud.ac.uk.

https://doi.org/10.1016/j.csite.2021.100980
Received 8 February 2021; Received in revised form 28 March 2021; Accepted 30 March 2021
Available online 7 April 2021

methodology, the following section presents current hot-spots detection algorithms and their main limitations.

1.1. Literature review: existing hot-spots detection algorithms and their limitations

Thermal imaging cameras most frequently inspect the detection of hot-spots in PV modules. However, in the last couple of years, promising yet not entirely accurate techniques have been developed to diagnose PV hot-spots. H. Chen et al. [11] proposed a data-driven feature extracting method to analyse PV hot-spots. The main limitations of this technique, including that all results were obtained using shaded PV cells, while in theory, this is a fair experiment, while in practice, many studies show that only applying shade does not necessarily correlate with the amount of voltage and current loss compared with actual hot-spotted PV modules. Besides, this technique depends on adding an extra capacitive load to the PV module; hence, the method's physical application is challenging.

J. Gosumbonggot & G. Fujita [12] introduced a method to detect PV hot-spots adopting the characteristics of the power-voltage (P–V) curve of the PV modules. This method can only be employed with stand-alone PV modules. It only defines whether the PV modules have hot-spots, but it inevitably fails to identify the nature "type" of the hot-spots.

M. Dhimsih et al. [13] suggested a suitable algorithm for detecting PV hot-spots based on the analysis of 2580 PV modules. To detect the hot-spots accurately, this algorithm uses the cumulative density function analysis (CDF). The maximum attained detection accuracy is equal to 80%, while this algorithm is only appropriate if an extensive data set of hot-spotted PV modules are accessible.

Other methods use artificial intelligence (AI) algorithms to identify hot-spots in PV systems. A very recent study by K. Niazi et al. [14] revealed a machine learning-based algorithm to diagnose PV hot-spots. The proposed algorithm requires the thermal images of the tested PV modules to aid the texture and histogram of gradient features of each thermal image. Naive Bayes (nBayes) classifier is then used to classify the type of hot-spots with a maximum recognition rate of 94.1%; experimented on 375 different samples. This algorithm still needs the thermal imaging procedure (as a pre-stage action), and the detection accuracy significantly diminishes as the resolution of the thermal camera is dropped.

A similar study recommended by G. Ngo et al. [15] used K-means colour quantisation for pre-processing and density-based spatial clustering of applications with noise for processing in the images captured by an infrared camera. The same technique was further applied by Ref. [16], although the infrared images' transformation was done using a Faster-RCNN machine learning algorithm. The detection accuracy of both [15,16] is limited to 90%.

A photovoltaic hot-spots fault detection method using a fuzzy-inference system developed by Ref. [17] shows another promising solution to detect hot-spots without thermal imaging cameras. This method uses a Mamdani-type fuzzy controller, which entails three different parameters, including the output voltage, current and power. The maximum attained detection accuracy was equal to 96.7%. This method's principal drawback is that the fuzzy controller requires a large set of hot-spotted PV data for training and validation purposes. Also, the type of hot-spots can be accurately detected; however, PV modules affected by one or two hot-spots are challenging to distinguish, making this method unsuitable for early-stage PV hot-spots detection. Comparatively, a deep learning-based process presented in Ref. [18] can identify hot-spots and micro-cracks in PV modules. This technique requires electroluminescence (EL) imaging cameras to pre-process the information "images" into a deep learning algorithm. Practically speaking, this tool is quite challenging to operate as outdoor EL imaging systems are expensive and requires a suitable outdoor space which is not the case in most residential roof-topped PV installations.

1.2. Contribution to knowledge and article organization

As discussed in the previous section, a limited number of methods are used to diagnose PV hot-spots without using the pre-processing procedure that requires the input of either thermal, infrared or EL images of the examined PV system. Furthermore, most recent algorithms, including [11–18], achieve a low detection accuracy of the PV hot-spots, mainly when utilised for early-stage hot-spotting scenarios.

Our main contributions are (i) the development of a machine learning tool that can detect hot-spots in PV modules; consequently, avoid using thermal imaging systems for early hot-spots detection, (ii) the machine learning tool is implemented and assessed using three different data setups. Hence, comparing different methodologies and, ultimately, finding the excellent data requirement for PV hot-spots detection, and last (iii) use four different classification algorithms to train and validate the machine learning tool, algorithms include decision tree (DT), support vector machine (SVM), K-nearest neighbour (KNN), and the discriminant classifier. These algorithms will be compared in terms of their performance using the confusion matrix and the receiver operating characteristics (ROC).

The following sections of the article are organised as follows: Section II introduces the examined PV installation and the data normalization process, whereas Sections III and IV explain the analysis of different machine learning classifiers and the overall results. Finally, Section V includes the discussion details, whereas Section VI illustrates this work's innovative conclusions.

2. Overall examined PV installation

This paper has implemented a machine learning tool that could be used to diagnose early-stage PV hot-spots. Practically, not all types of hot-spots are categorised as early-stage hot-spots; therefore, we have categorised the hot-posts conditions as follows:

1) PV module affected by one hot-spotted solar cell
2) PV module affected by two hot-spotted solar cells
3) PV module affected by $\geq$ three hot-spotted solar cells

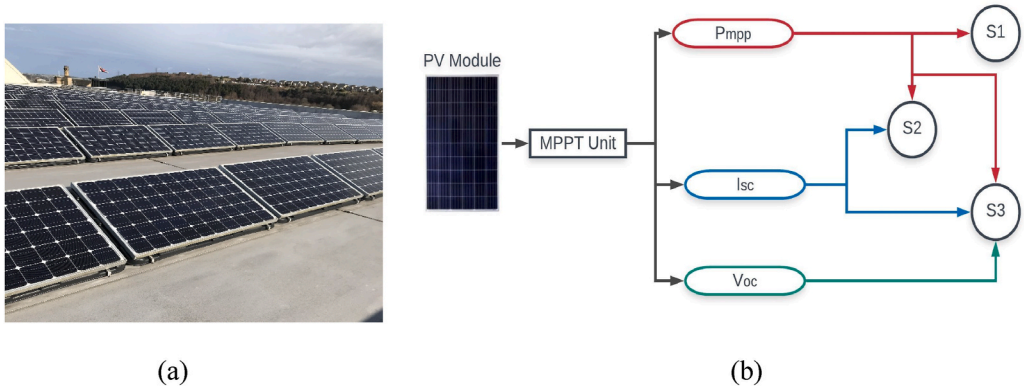


Fig. 1. (a) Examined PV installation, (b) Schematic of the parameters used to test the effectiveness of the machine learning algorithms.

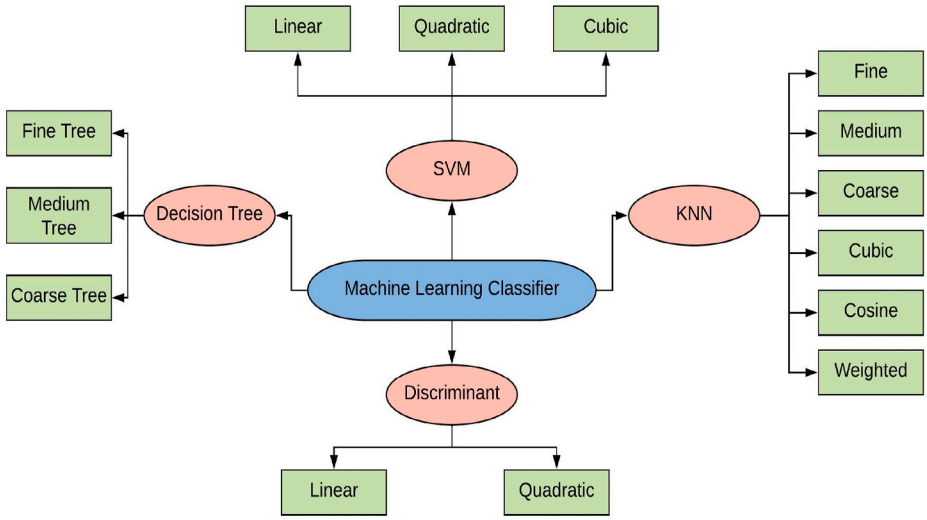


Fig. 2. List of examined classification algorithms.

To analyse PV modules' performance affected by these hot-spotting categories, we have examined a PV system shown in Fig. 1(a). Each PV module's peak output power is 220W, I_{sc} and V_{oc} are equal to 8.18A and 36.7V, respectively.

Before the machine learning tool's employment to diagnose PV hot-spots, it is essential to define the data setup(s), which will be used later by the classification algorithms. Therefore, three different data setups were identified; these setups are presented as S1 to S3 in Fig. 1(b). Each setup consists of an additional input parameter, resulting in a generic overview of the minimum data requirement, as a system point-of-view, to implement a robust hot-spotting detection algorithm, including an accurate detecting accuracy.

Another reason for selecting different data setups, as some PV installations are fitted with MPPT units that only have the acquisition of the output power at maximum power point (P_{mpp}), while some other advanced MPPT units, widely available in todays' market, offer a wide range of data assets including P_{mpp}, I_{sc} and V_{oc}.

Therefore, the first setup (S1) only contains the data of the P_{mpp}, while S2 consists of two parameters, Pmpp and I_{sc}. The last setup, S3, comprises three inputs, P_{mpp}, I_{sc}, and Voc. Hence, S3 would necessitate additional computational time to process the classification model compared with the previous two setups according to another input parameter requirement.

Before the implementation of the machine learning tool, the data of each examined PV module was normalized. Without the normalization process, and as the PV modules' acquired data are of different scale, it is therefore required to normalize the dataset using (1) [19].

$$X_{new} = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

where X_{new} is the new normalized data point, X is the actual measured data point from the PV module, X_{min} and X_{max} are the minimum and maximum observed values, respectively.

The normalization of the data samples plays a strong impact, especially when it comes to validating the machine learning model.

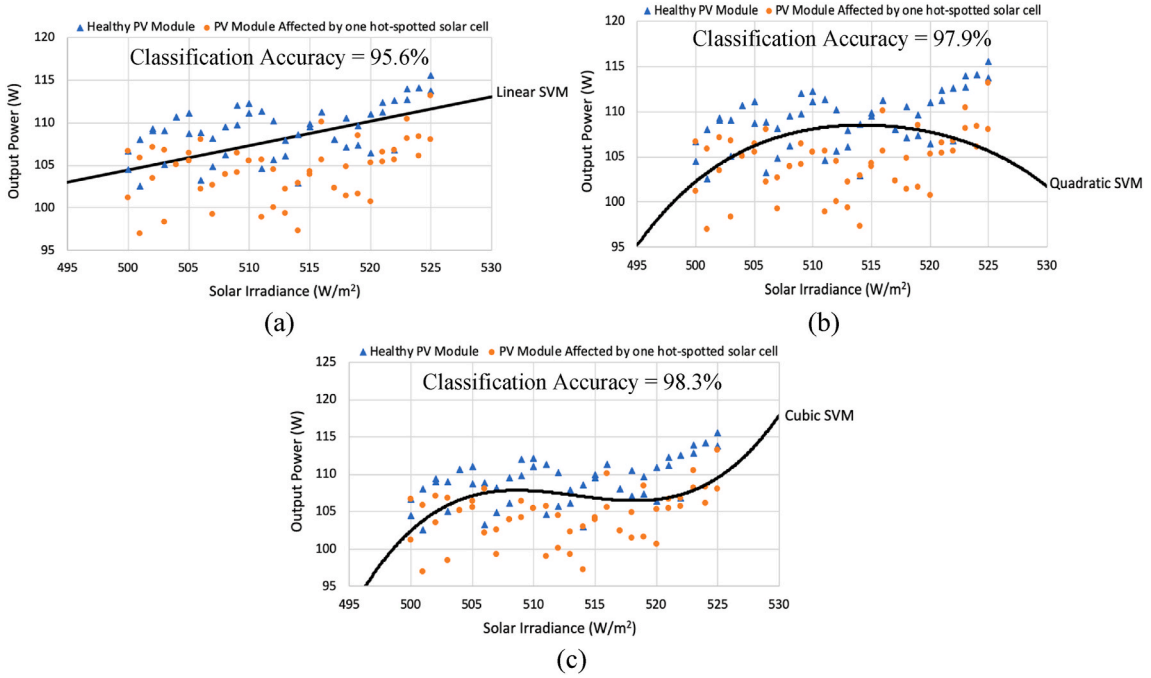


Fig. 3. Results of the SVM algorithms using datapoints of solar irradiance vs output power from 500 to 525 W/m^2. (a) Linear SVM, (b) Quadratic SVM, (c) Cubic SVM.

Consequently, models without normalized data (when the data samples are with different ranges) habitually fail to attain a high detection accuracy rate.

3. Machine learning tool

This section presents all classification algorithms, shown in Fig. 2, that were used to validate the proposed PV hot-spots detection tool's accuracy. For each classifier, three different data setups were testified.

3.1. Decision tree

DTs are a type of supervised machine learning where the data is continuously split according to a specific parameter [20]. There are two types of DTs, classification, and regression. The main difference between both types that the regression gives an output as a number, while the classification indicates the output according to the input used for training and validation process.

Three different DTs algorithms were tested. The fine tree uses many leaves (categories) that makes many refined distinctions between the observed classes; max distinctions of the observations could go up to 100. While the only difference from other types of DTs, including the Medium Tree and Coarse Tree, is that these DTs utilize fewer distinctions between the observations. The medium tree typically uses up to 40, and Coarse Tree is limited to 4. The data classification is strongly dependent on the observations in terms of the quality of the data (noiseless or noisy) and the data setup.

3.2. Support vector machine

SVM is a supervised machine learning algorithm that can be used for classification data sets. In this algorithm, we plot each observation as a point in n-dimensional space, where n is defined as the number of features. Therefore, according to our data, n is equal to 4, representing four different features including, a healthy PV module, a PV module affected by one hot spotted solar cell, a PV module affected two hot spotted solar cells, and a PV module affected by $\geq$ three hot spotted solar cells.

We have examined the performance of three different SVM algorithms, including Linear, Quadratic and Cubic. Linear SVM makes a linear separation between the observations, and it makes it the easiest SVM to develop. Quadratic and Cubic SVMs use a second and third-order polynomial kernel [21], calculated using (2) for two-classes $\vec{x}$ and $\vec{z}$.

$$k(\vec{x}, \vec{z}) = (\vec{z}^{T}\ \vec{x} + c)^{n} \tag{2}$$

where n is the "order" of the kernel, and c is a constant that allows to trade off the influence of the higher-order and lower order terms.

When the solar irradiance increase, the output power increases; therefore, it is highly expected that the ideal classification would be

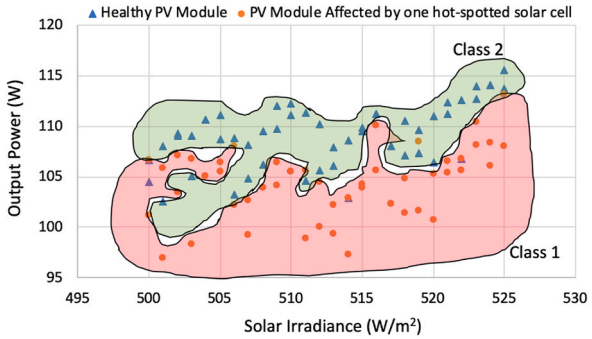


Fig. 4. KNN algorithm class boundaries for class1 and class2; data of class 1 corresponds to a PV module affected by one hot-spotted solar cell, while data of class 2 corresponds to a healthy PV module.

the Linear SVM. To understand whether this is true, we have plotted the data (solar irradiance ranges from 500 to 525 W/m2) of a healthy PV module against a PV module affected by one hot-spotted solar cell; results of the SVM classification is shown in Fig. 3. As expected, the linear SVM attains the highest classification accuracy of 98.3% compared with the Quadratic and Cubic SVMs that have 95.6% and 97.9%, respectively.

The Cubic SVM algorithm is typically not suitable to use along with large data sets and would not perform well in overlapping classes [22], such as for PV hot-spotting scenarios. Although, the Cubic SVM algorithm is well suited for extreme case binary classification. On the other hand, the Quadratic SVM algorithm performs similarly to the Linear SVM. In contrast, this algorithm's main limitation is that it requires a significant amount of time to process the classification, typically twice the time of the linear SVM algorithm.

In today's deployed machine learning models, practically speaking, data with overlapped classes are usually classified using KNN algorithms. The performance of six different KNN algorithms will be discussed in the next sub-section.

3.3. K-nearest neighbour

KNN algorithms are easy to implement, fast and reliable, predominantly with small to medium-sized data sets. In practice, implementing a KNN algorithm takes less computational time compared with DT and SVM algorithms [23].

In this article, we have evaluated the performance of six different KNN algorithms, described as follows:

1) Fine KNN classifier makes exquisitely clear distinctions between the classes, with the number of neighbors set to 1.
2) Medium KNN classifier makes fewer number of distinctions between the classes, with the number of neighbors set to 10.
3) Coarse KNN classifier makes coarse distinctions between the classes, with several neighbors set to 100.
4) Cosine KNN classifier uses the cosine distance metric between the classes. The cosine distance between two n-dimensional vectors u and v is defined using (3).

$$Cosine\ Distance = 1 - \frac{u \,.\, v}{|u| \,.\, |v|} \tag{3}$$

5) Cubic KNN classifier uses the cubic distance metric between the classes. The cosine distance between two n-dimensional vectors u and v is defined using (4).

$$Cubic\ Distance = \sqrt[3]{\sum_{i=1}^{n} |u_i - v_i|^3} \tag{4}$$

6) Weighted KNN classifier uses the weights of the distances between the classes. The weighted distance between two n-dimensional vectors u and v is defined using (5); where w_i is the actual weights of the complete classification process, where $\sum_{i=1}^{n} w_i = 1$.

$$Weighted\ Distance = \sqrt{\sum_{i=1}^{n} w_i (u_i - v_i)^2} \tag{5}$$

To visualize the classification of a KNN-based algorithm, we have used data of healthy vs PV module affected by one hot-spotted solar cell, as presented in Fig. 4. The classification algorithm set to Fine KNN. It is evident that the KNN algorithm divided the data set into two classes; in this case, the number of neighbors set to 1; hence, inaccurate overall classification of the dataset is resolute.

Nevertheless, according to the PV modules data set, it is well suited to use either the cubic or the weighted KNN algorithm to classify the classes because both of these algorithms define each class's distance (hot-spotting scenarios) against their overall number of

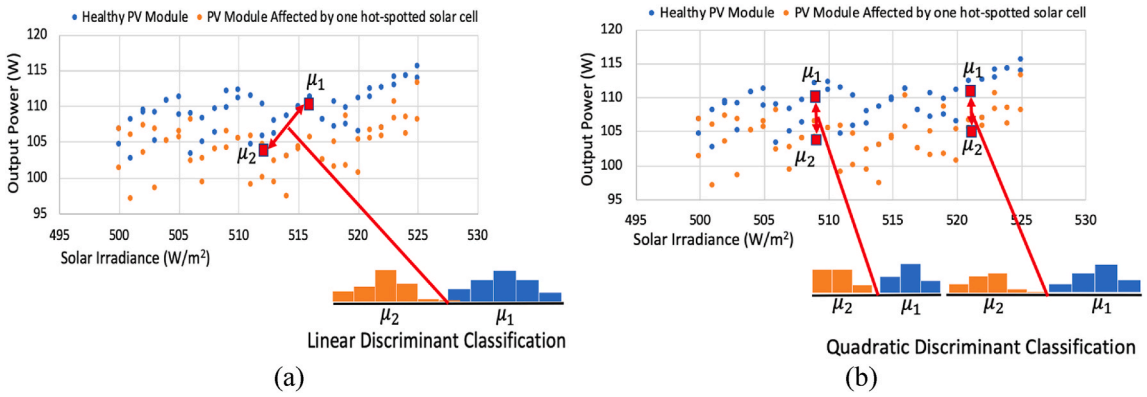


Fig. 5. Output results obtained using discriminant classifier. (a) Linear discriminant, (b) Quadratic discriminant.

Table 1
Results of the accuracy determined using all machine learning classification algorithms.

Classifier	Classification Algorithm	Accuracy (%) using data setup 1	Accuracy (%) using data setup 2	Accuracy (%) using data setup 3
DT	Fine	25	83	86
	Medium	27	87	88
	Coarse	26	80	84
SVM	Linear	44	79	93
	Quadratic	42	77	94
	Cubic	42	77	92
KNN	Fine	51	85	93
	Medium	54	86	94
	Coarse	43	86	94
	Cosine	49	77	86
	Cubic	56	87	97
	Weighted	52	86	96
Discriminant	Linear	49	87	95
	Quadratic	52	90	98

observations, rather than only selecting a specific number of neighboring (i.e., using fine, medium, and coarse KNN) to adjust the classification of the data set.

3.4. Discriminant classifiers

In theory, there are two discriminant classifiers, Linear and Quadratic. The primary purpose of a discriminant classifier is to map the data of different classes into a new dimensional axis. For example, if data contains two different PV hot-spotting scenarios, it is only required for one new dimensional axis. In comparison, three new dimensional axes will be required to have four different strategies [24].

The first criteria are to maximize the distance between the means m of the data sets. The second criteria are to minimize the variation, which the linear discriminant classifier calls 'scatter' and represented by s2 with each data set. In contrast with the above criteria, the discriminant classifier's distance is defined using (6). There is only one axis that will be generated using the linear discriminant classifier. However, two axes are created using the quadratic discriminant.

$$Disriminant\ Classifier\ Distance = \frac{(\mu_1 - \mu_2)^2}{s^2{}_1 + s^2{}_2} \tag{6}$$

Results of the discriminant classifiers are shown in Fig. 5. As shown in Fig. 5(a), the linear discriminant classifier generates a single x-axis that consists of a scatter (histogram) of each data set. As there is no significant difference between both identified classes ($\mu_1$and μ_2), the results of the classifier for both new classes are overlapping; hence, the accuracy of this classifier would be improved using the quadratic classification.

According to the quadratic discriminant classifier results, presented in Fig. 5(b), the data set was classified into two distinct classes. In contrast, the distances between the data points were measured using a singular scatter, but two different scatter plots were identified for each data set. They were resulting in improved separation of the acknowledged classes.

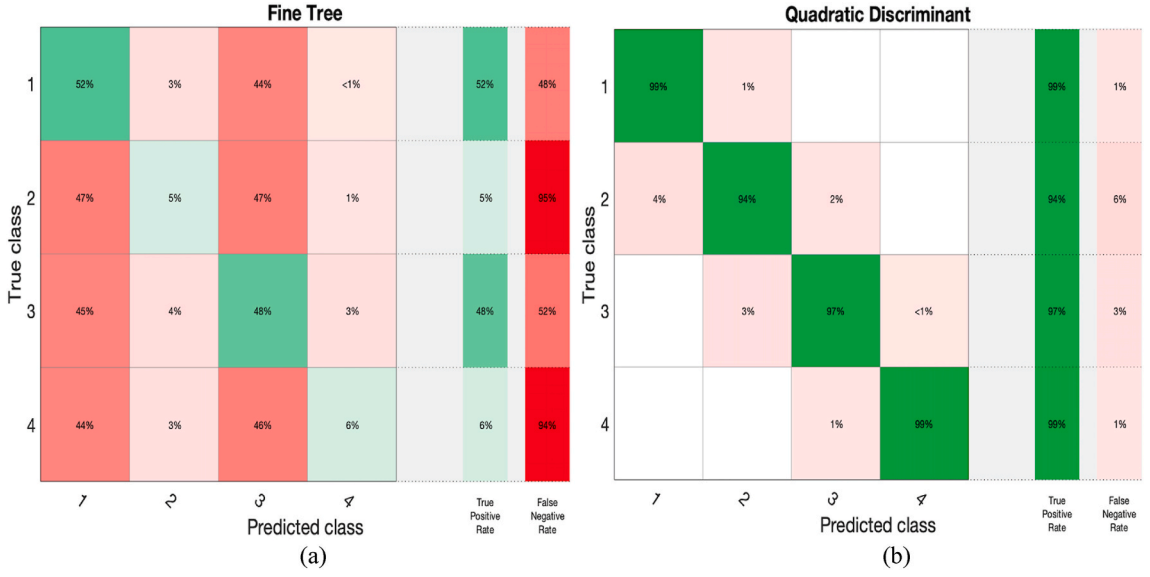


Fig. 6. Output confusion matrix of the best and the worst classifier. (a) Fine decision tree using data setup 1, (b) Quadratic discriminant using data setup 3.

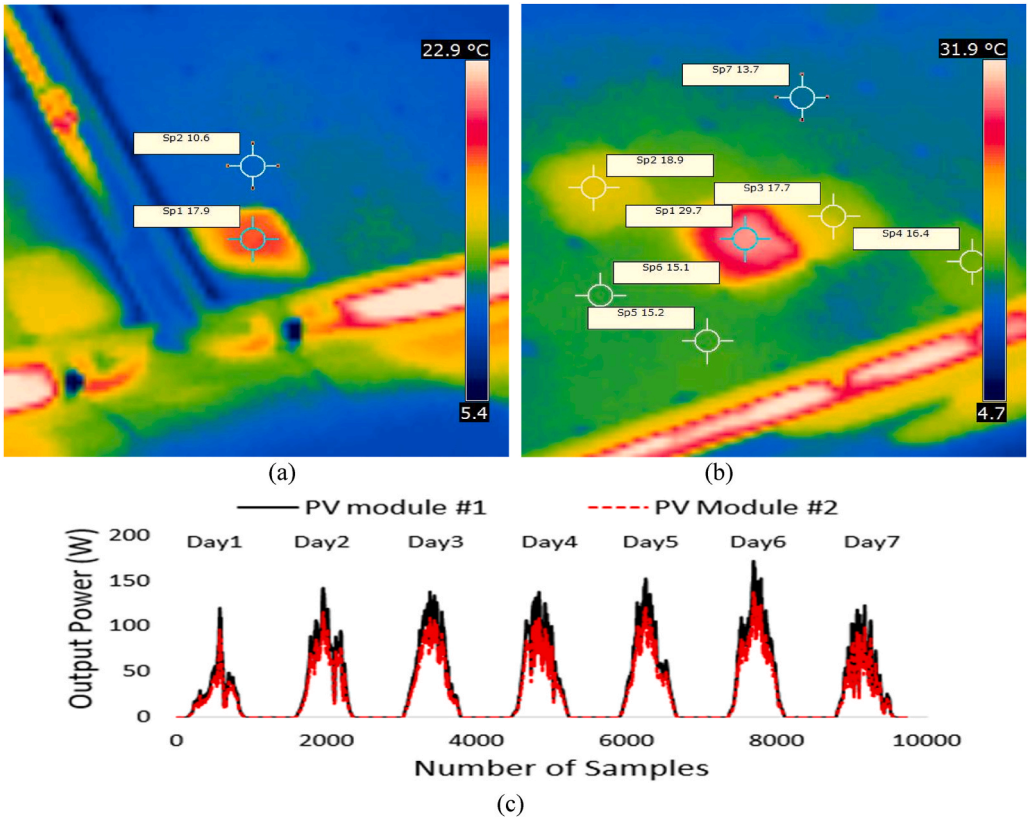


Fig. 7. Examined PV modules for further validation. (a) Thermal image of PV module #1, (b) Thermal image of PV module #2, (c) Measured output power for both PV modules over 7 days.

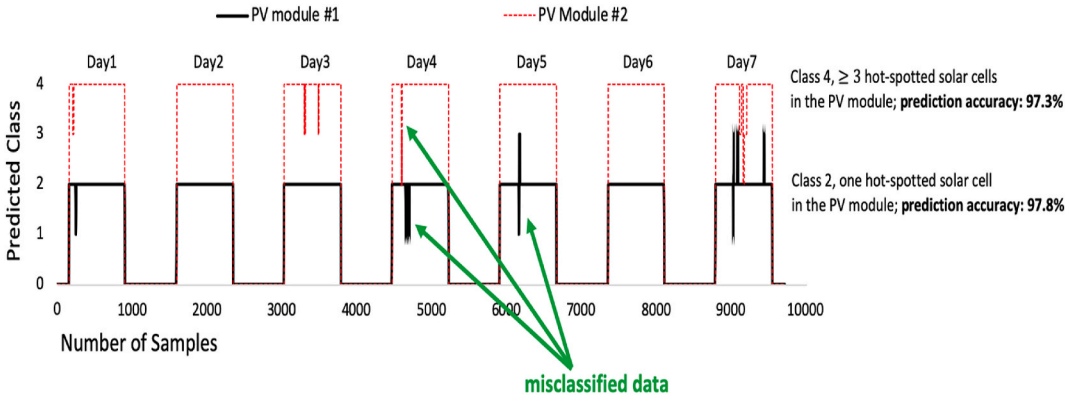


Fig. 8. Output results obtained using discriminant classifier on two hot-spotted PV modules.

4. Results

In this section, the analysis and the performance of all examined classification algorithms will be compared using real-time long-term PV data measurements. In addition, to validate the machine learning tool, two PV modules affected by different hot-spots will be assessed.

4.1. Training and validation performance

To ensure that the implemented detection tool can also be used with PV modules under normal conditions, a data set of a healthy PV module has also been interposed in the training and validation process. In addition, as discussed earlier in the previous section, three different data setups were compared during the training and validation phase. To analyse the output of each ML algorithm, MATLAB software was used for this purpose.

Obtained results using all classification algorithms are presented in Table 1. Using data setup 1, all algorithms attain low detection accuracy, where the maximum is equal to 56% achieved by cubic KNN. The second and third data setups show an accurate hot-spots classification in 77%–98%.

The overall minimum accuracy of 25% is determined for the fine decision tree using data setup 1, while the maximum of 98% is observed for the quadratic discriminant classifier using data setup 3. The confusion matrices of both classifiers are shown in Fig. 6. Classes 1 to 4 are healthy PV module, one hot-spotted solar cell, two hot-spotted solar cells, and three or more hot-spotted solar cells, respectively.

According to Fig. 6(a), the fine tree decision can precisely classify 52% of all class 1 samples (obtained using the healthy PV module). In contrast, in all other classes, the classifier fails to categorize the hot-spot conditions accurately. However, according to the quadratic discriminant classifier's output confusion matrix presented in Fig. 6(b), an accurate classification of the hot-spots was achieved. The minimum classification of 94% for class 2 (PV module affected by one hot-spotted solar cell) is determined because the classification in this class is highly overlapping with class 1 and class 2. The measured data of Isc and Voc show a high degree of similarity in both cases.

4.2. Further validation of the accuracy of the proposed machine-learning tool

To further validate the best-fit machine learning model, using data setup three input for a quadratic discriminant classifier, we have tested two different PV modules, presented as PV module #1 and #2 as in Fig. 7(a) and (b), respectively.

PV module #1 is affected by one hot-spotted solar cell, while PV module #2 is suffering from ≥3 hot-spotted solar cells. The measured output power of both PV modules for a period of 1-week has been recorded and presented in Fig. 7(c). As PV module #2 is affected by a larger number of hot-spots compared with the PV module #1, it is evident that its actual output power loss is higher.

After reprocessing both PV modules' data, the output predicted class, shown in Fig. 8, of the quadratic discriminant classifier, shows a significant prediction accuracy for the hot-spot stirring type in the PV modules.

According to PV module #1, 97.8% was predicted that this PV module is affected by one hot-spotted solar cell, as the predicted class is unvarying at 2. In comparison, PV module #2 is predicted to have ≥3 hot-spotted solar cells since the predicted class is 4, with a prediction accuracy of 97.3%.

As labelled in Fig. 8, for both predicted classes, there are several misclassified data points, this is due to several reasons, including, (i) the imperfection of the classification algorithms and their deployment, (ii) The variations of the temperature, shading, and solar irradiance can reduce the type of hot-spot type, and (iii) dataset used to train and validate the classification algorithms indubitably has some imprecise data points. In contrast with the above results, this experiment shows the importance, significance, and accuracy of the machine learning tool proposed using the quadratic classifier for predicting early-stage hot-spots in PV modules.

Table 2
Comparative Study Of Our Proposed Tool vs Recent published work in Ref. [14] and [[30,31]].

Reference	Number of Required Input Parameters	Best-Fit Machine Learning Classification Model	PV Hot-spots Detection/Classification Accuracy
[14]	4	Naive Bayes classifier	94.1%
[29]	5	Mamdani Fuzzy logic classifier	95.27%
[30]	4	SVM classifier	92%
[31]	5	Mamdani and Sugeno Fuzzy logic classifier	92.1%
This work	3	Quadratics Discriminant classifier	98%

5. Discussion

In our work, we have presented the development and the analysis of multi-ML algorithms that can be used to detect hot-spots in PV modules. We found that the discriminant classifier has the highest detection accuracy of all other testified ML algorithms. This result has also been suggested in a different scientifical field of study, such as in water level classification (ocean engineering) [25] or even in power machines (power engineering) [26]. The advantage to detect early hot-spots can lead to a significant contribution to our primary mission of "zero-carbon" cities. This detection would improve the quality and reliability of the PV systems and increase their yield annual energy.

We have profound solid work that can lead to a significant understanding of hot-spots' behavior and detection. Our progressive approach can be further embedded with suitable maximum power point tracking units or dc-ac inverters, of course, with the integration of other sophisticated microcontrollers such as FPGAs suggested by D. T. Nguyen et al. [27] that can process the data samples within micro-to-milliseconds.

The advantage to detect early hot-spots can lead to a significant contribution to our primary mission of "zero-carbon" cities. This detection would improve the quality and reliability of the PV systems and increase their yield annual energy. Early diagnosis of hot-spots can also be helpful to reduce the impact of micro-cracks in solar cells. For example [28], suggests that micro-cracks can lead to hot-spots, and the annual energy loss of the impacted PV modules could be as low as 25%. The most popular micro-crack detection system is electroluminescence (EL); in contrast, these systems are expensive (>£10,000) and must follow a specific experimentation procedure, including running the PV module at it is short-circuit current to obtain good quality EL images. However, in our proposed model, we can detect such events in PV modules without any additional equipment. It could simply function simultaneously while connecting the PV system into the load or, in-case off-grid scenario, to the battery storage.

A comparative study of our presented work vs recent work published by Refs. [14,29–31] is demonstrated in Table 2. As can be noticed, that different classifiers have been widely stated, while the best-fit has not been yet fully identified. Our proposed model has only three input parameters; whistle all other work has at least four inputs to perform the ML algorithm.

In our work, we have checked against 15 different ML classifiers, and we have found that the quadratic discriminant scores the best-fit classification for hot-spots PV modules with an average detection accuracy of 98%; way above all recent studies such as the maximum 95.27% up-to-date published by Ref. [26]. We have also confirmed no requirement to use complex algorithms for data processing, such as using fuzzy logic-based classifiers as performed by Refs. [30,31].

6. Conclusion

In this article, we have experimented with 15 different machine learning classifiers to detect and classify early hot-spots in PV modules. We have found that the quadratic discriminant classifier is the best-fit machine learning to early diagnose PV hot-spots with an average detection accuracy of 98% when using appropriate data samples, including the short circuit current, the open-circuit voltage, and the maximum output power. It is also understood that the decision tree classifiers are not suitable to apply as they depend on the ratio of the samples' average; when used with PV hot-spots detection and classification, their detection accuracy is consistently below 90%. It was also concluded that KNN classifiers, particularly Cubic-KNN, is appropriate for PV hot-spots classification as they resemble detection accuracy in the range of 86%–97%.

Author statement

M Dhimish (Conceptualization; Data curation; Formal analysis; Funding acquisition; Investigation; Methodology; Project administration; Roles/Writing – original draft; Writing – review & editing).

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] A. Riad, M.B. Zohra, A. Alhamany, M. Mansouri, Bio-sun tracker engineering self-driven by thermo-mechanical actuator for photovoltaic solar systems, in: Case Studies in Thermal Engineering, vol. 21, Oct. 2020, 100709, https://doi.org/10.1016/j.csite.2020.100709.

[2] M. Dhimish, Thermal impact on the performance ratio of photovoltaic systems: a case study of 8000 photovoltaic installations, in: Case Studies in Thermal Engineering, vol. 21, October 2020, 100693, https://doi.org/10.1016/j.csite.2020.100693.
[3] M. Dhimish, P. Mather, V. Holmes, Evaluating power loss and performance ratio of hot-spotted photovoltaic modules, 12, in: IEEE Transactions on Electron Devices, vol. 65, Dec. 2018, pp. 5419–5427, https://doi.org/10.1109/TED.2018.2877806.
[4] M. Dhimish, G. Badran, Current limiter circuit to avoid photovoltaic mismatch conditions including hot-spots and shading, in: Renewable Energy, vol. 145, Jan. 2020, pp. 2201–2216, https://doi.org/10.1016/j.renene.2019.07.156.
[5] R.A. Hartman, Bypass Diode Assembly for Photovoltaic Modules, 1986 (No. US 4577051), US Patent, 1986.
[6] R. G. Vieira, F. M. de Araújo, M. Dhimish, and M. I. Guerra, "A comprehensive review on bypass diode application on photovoltaic modules.," In Energies, vol. 13, no. 10, pp. 2472, doi: 10.3390/en13102472.
[7] M. Dhimish, P. Mather, Ultrafast high-resolution solar cell cracks detection process, 7, in: IEEE Transactions on Industrial Informatics, vol. 16, July 2020, pp. 4769–4777, https://doi.org/10.1109/TII.2019.2946210.
[8] S. Ghosh, V.K. Yadav, V. Mukherjee, A novel hot spot mitigation circuit for improved reliability of PV module, 1, in: IEEE Transactions on Device and Materials Reliability, vol. 20, March 2020, pp. 191–198, https://doi.org/10.1109/TDMR.2020.2970163.
[9] M. Dhimish, V. Holmes, B. Mehrdadi, M. Dales, P. Mather, Detecting defective bypass diodes in photovoltaic modules using Mamdani fuzzy logic system, Global J. Res. Eng. vol. 17 (2017) 33–44.
[10] M. Dhimish, 70% decrease of hot-spotted photovoltaic modules output power loss using novel MPPT algorithm, 12, in: IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 66, Dec. 2019, pp. 2027–2031, https://doi.org/10.1109/TCSII.2019.2893533.
[11] H. Chen, H. Yi, B. Jiang, K. Zhang, Z. Chen, Data-driven detection of hot spots in photovoltaic energy systems, 8, in: IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 49, Aug. 2019, pp. 1731–1738, https://doi.org/10.1109/TSMC.2019.2896922.
[12] J. Gosumbonggot and G. Fujita, "Global maximum power point tracking under shading condition and hotspot detection algorithms for photovoltaic systems," in Energies, vol. 12, no. 5, pp. 882, doi: 10.3390/en12050882.
[13] M. Dhimish, P. Mather, V. Holmes, Novel photovoltaic hot-spotting fault detection algorithm, 2, in: IEEE Transactions on Device and Materials Reliability, vol. 19, June 2019, pp. 378–386, https://doi.org/10.1109/TDMR.2019.2910196.
[14] K. Niazi, W. Akhtar, H.A. Khan, Y. Yang, S. Athar, Hotspot diagnosis for solar photovoltaic modules using a Naive Bayes classifier, in: Solar Energy, vol. 190, Sep. 2019, pp. 34–43, https://doi.org/10.1016/j.solener.2019.07.063.
[15] G.C. Ngo, E.Q.B. Macabebe, Image segmentation using K-means color quantization and density-based spatial clustering of applications with noise (DBSCAN) for hotspot detection in photovoltaic modules, in: IEEE Region 10 Conference (TENCON), Singapore, 2016, 2016, pp. 1614–1618, https://doi.org/10.1109/TENCON.2016.7848290.
[16] M. Dhimish, P. Mather, Development of novel solar cell micro crack detection technique, 3, in: IEEE Transactions on Semiconductor Manufacturing, vol. 32, Aug. 2019, pp. 277–285, https://doi.org/10.1109/TSM.2019.2921951.
[17] M. Dhimish, G. Badran, Photovoltaic hot-spots fault detection algorithm using fuzzy systems, 4, in: IEEE Transactions on Device and Materials Reliability, vol. 19, Dec. 2019, pp. 671–679, https://doi.org/10.1109/TDMR.2019.2944793.
[18] C. Henry, S. Poudel, S.W. Lee, H. Jeong, Automatic detection system of deteriorated PV modules using drone with thermal camera, in: Applied Sciences, vol. 10, May 2020, p. 3802, https://doi.org/10.3390/app10113802.
[19] C. Chen, K. Li, A. Ouyang, Z. Tang, K. Li, GPU-accelerated parallel hierarchical extreme learning machine on flink for big data, 10, in: IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 47, Oct. 2017, pp. 2740–2753, https://doi.org/10.1109/TSMC.2017.2690673.
[20] Z. Guo, Y.Y. Haimes, Exploring systemic risks in systems-of-systems within a multiobjective decision framework, 6, in: IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 47, June 2017, pp. 906–915, https://doi.org/10.1109/TSMC.2016.2523918.
[21] J. Wang, D. Yang, W. Jiang, J. Zhou, Semisupervised incremental support vector machine learning based on neighborhood kernel estimation, 10, in: IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 47, Oct. 2017, pp. 2677–2687, https://doi.org/10.1109/TSMC.2017.2667703.
[22] M. Su, Z. Zhang, Y. Zhu, D. Zha, Data-driven natural gas spot price forecasting with least squares regression boosting algorithm, in: Energies, vol. 12, March 2019, p. 1094, https://doi.org/10.3390/en12061094.
[23] N. Ali, D. Neagu, P. Trundle, Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets, in: SN Applied Sciences, vol. 1, Nov. 2019, pp. 1–15, https://doi.org/10.1007/s42452-019-1356-9.
[24] K.S. Gyamfi, J. Brusey, A. Hunt, E. Gaura, Linear classifier design under heteroscedasticity in linear discriminant analysis, in: Expert Systems with Applications, vol. 79, Aug. 2017, pp. 44–52, https://doi.org/10.1016/j.eswa.2017.02.039.
[25] B. Mohammadi, Y. Guan, P. Aghelpour, S. Emamgholizadeh, R. Pillco Zolá, D. Zhang, Simulation of titicaca lake water level fluctuations using hybrid machine learning technique integrated with grey wolf optimizer algorithm. Water, 11, in: Water, vol. 12, July 2020, p. 3015, https://doi.org/10.3390/w12113015.
[26] I. Rahman, M. Kuzlu, S. Rahman, Power disaggregation of combined HVAC loads using supervised machine learning algorithms, in: Energy and Buildings, vol. 172, Aug. 2018, pp. 57–66, https://doi.org/10.1016/j.enbuild.2018.03.074.
[27] D.T. Nguyen, T.N. Nguyen, H. Kim, H. Lee, A high-throughput and power-efficient FPGA implementation of YOLO CNN for object detection, 8, in: IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 27, Aug. 2019, pp. 1861–1873, https://doi.org/10.1109/TVLSI.2019.2905242.
[28] M. Dhimish, V. Holmes, Solar cells micro crack detection technique using state-of-the-art electroluminescence imaging, 4, in: Journal of Science: Advanced Materials and Devices, vol. 4, Dec. 2019, pp. 499–508, https://doi.org/10.1016/j.jsamd.2019.10.004.
[29] M. Dhimish, V. Holmes, B. Mehrdadi, M. Dales, P. Mather, Photovoltaic fault detection algorithm based on theoretical curves modelling and fuzzy classification system, in: Energy, vol. 140, Dec. 2017, pp. 276–290, https://doi.org/10.1016/j.energy.2017.08.102.
[30] M.U. Ali, H.F. Khan, M. Masud, K.D. Kallu, A. Zafar, A machine learning framework to identify the hotspot in photovoltaic module using infrared thermography, in: Solar Energy, vol. 208, Sep. 2020, pp. 643–651, https://doi.org/10.1016/j.solener.2020.08.027.
[31] M. Dhimish, V. Holmes, B. Mehrdadi, M. Dales, Comparing Mamdani Sugeno fuzzy logic and RBF ANN network for PV fault detection, in: Renewable Energy, vol. 117, March 2019, pp. 257–274, https://doi.org/10.1016/j.renene.2017.10.066.