Artificial Neural Network based photovoltaic fault detection algorithm integrating two bi-directional input parameters

Muhammed Hussain, Mahmoud Dhimish, Sofya Titarenko, Peter Mather

University of Huddersfield, Laboratory of Photovoltaics, Huddersfield, HD1 3DH, United Kingdom

Abstract— In this paper, a fault detection algorithm for photovoltaic systems based on artificial neural networks (ANN) is proposed. Numerous literatures can be found on the topic of PV fault detection through the implementation of artificial intelligence. The novel part of this research is the successful development, deployment and validation of a fault detection PV system using radial basis function (RBF), requiring only two parameters as the input to the ANN (solar irradiance and output power). The results obtained through the testing of the developed ANN on a PV installation of 2.2 kW capacity, provided an accuracy of 97.9%. To endorse the accuracy of the newly developed algorithm, the ANN was tested on another PV system, installed at a remote location. The total capacity of the new system was significantly higher, 4.16 kW. A vital part of the test was to see how the proposed ANN would perform with 'scaled-up' input data, during normal operation as well as partial shading scenarios. The validation process provided an overall fault detection accuracy of above 97%. The decrease in accuracy was due to the varying nature of the two systems in terms of total capacity, number of samples and type of faults.

1. Introduction

1.1 Research Background

Technological advancements in both hardware and software have enhanced the monitoring and analysis of grid-connected photovoltaic (GCPV) systems, for optimal energy harvesting along with reliable power production. The increase in the installation of PV due to its numerous benefits, means researchers are actively looking into the development of diagnostic methods for fault detection in PV plants.

PV systems process monitoring is based on a distributed sensor network (DSN), for analysis of the system and performance reviews. Corresponding time series analysis of obtained data is vital for statistical analysis of PV systems. The implementation of fault detection in PV systems can become complex depending on the variety of data being logged. A DSN may consist of several variables such as current, voltage, irradiance and temperature [1-2]. Therefore, the complexity of the process may hinder enterprises from implementing fault detection in their systems.

33 Researchers have proposed numerous methods to ease the complexity of failure detection systems
34 by decreasing the number of input variables required, along with more sophisticated statistical
35 analysis [3-4]. This paper proposes the use of Artificial Neural Networks (ANN) for this purpose.

36 Various fault detection techniques for PV systems are widely available in the literature, with
37 varying accuracy levels, detection speed and algorithm complexity. Prediction of faults in PV
38 systems through the unitisation of metrological and satellite data is one of the techniques used [5-
39 6]. Whilst other fault detection algorithms for PV systems do not require any climate data [7].

40 Fault detection in PV systems can be split into three categories visual, thermal and electrical [8].
41 Before looking deeper into the electrical category, it is important to mention another technique
42 which is also applicable, known as Electro Luminescence Imaging (EL). This technique is based
43 on the solar module being supplied with external excitation current through its metal contacts,
44 acting as a light emitting diode. A sensitive Si-CCDs camera can then take an image of the emitted
45 photons at a wavelength greater than 850nm.

46 The electrical category is further divided into sub-categories consisting off:

47 • Methods that do not require any climate data (solar irradiance, temperature). An example
48   of this approach is the Time-Domain Reflectometry (TDR) proposed in [9] for detection
49   of disconnection of a PV string.
50 • Methods based around the analysis of the current and voltage characteristics. S. Silvestre
51   *et al.* [10] calculates Series Resistance (R_s), Fill Factor (FF) and Shunt Resistance (R_{sh})
52   based on the I-V characteristics leading on to performance indicators. Fault detection for
53   PV systems based on the evaluation of current and voltage indicators.
54 • Methods based around Maximum Power Point Tracking (MPPT). X. Li *et al.* [11] Proposes
55   an automatic supervision and fault detection method based on power loss analysis. The
56   approach led to the identification of faults including faulty module, faulty string and faults
57   linked to partial shading, MPPT failure and ageing.
58 • Methods based on Artificial Intelligence (AI) techniques. Authors' in [12] look at the
59   effectiveness of BP neural network for fault diagnosis in PV systems, comparing it to Fuzzy
60   Logic. The author concludes BP neural network as the solution to most limitation faced
61   through the implementation of Fuzzy Logic in fault detection of PV systems.

62 Artificial neural networks (ANNs) are mathematical tools, imitating biological human neural
63 networks, learning from experience and generalizing previous behaviour as characteristics [13].
64 The ANN architecture consists of an input layer, one or more hidden layers and an output layer.
65 ANN's ability to process information in non-linear, high-parallelism, fault and noise environments
66 makes it of considerable interests to researchers in many fields [14-16]. In comparison to
67 traditional model-based methods, ANN's are data-driven, self-adaptive methods learning from
68 examples whilst picking-up subtle and hidden functional relationships that are otherwise unknown
69 or hard to describe. In addition, ANNs are suitable for solving problems where explicit knowledge
70 is difficult to identify, but a vast amount of data is available [17-19]. A. Lapedes *et al*. [20],

71 demonstrates how backpropagation neural networks surpass by order of magnitude any of the
72 conventional linear and polynomial methods dealing with chaotic time series of data. In addition,
73 A. Millit *et al*. [21] demonstrates ANN networks as a solution for the modelling and estimating of
74 output power for PV systems. Whereas, F. Polo *et al*. [22] proposes a failure mode prediction and
75 energy harvesting of PV systems to support dynamic maintenance tasks using ANN-based models.
76 Ultimately, it looks to analyse data and disregard the erroneous prediction of faults in a PV system.
77 The paper implements a back-propagation network, trained on historical data consisting of past
78 five years of an inverter used in the PV system. Highlighting the nature of failures under
79 consideration, the paper describes the faults as a result of equipment deterioration and useful life
80 reduction owing to operational and geographical features. The author in his conclusion claims that
81 the proposed methodology could further improve its performance if 'enough data' is available for
82 significant training of the ANN.

83 Yasuhiro Yagi *et al.* [23] proposes a learning method based on expert systems for the identification
84 of two types of faults (shading effect and inverters failure). The main advantage of this technique
85 is that it relies on simple and reprogrammable ANN network, but, on the other hand, the proposed
86 technique cannot identify faulty conditions occurring in PV systems such as PV short circuit failure
87 conditions and PV String failure. An advanced ANN network proposed by [24] demonstrates the
88 detection of faults in the DC part of a PV system such as faulty bypass diodes, faulty PV modules,
89 and faulty PV string. The paper proposes two algorithms based on MLP and RBF networks. The
90 confusion matrix shows the accuracy of the MLP (90.3%) was higher than that of RBF (68.4%).
91 It is a well-known fact that MLP networks, have a much higher accuracy as compared to RBF and
92 are more commonly used. Especially in networks requiring more hidden layers. However, it is also
93 important to note that MLP networks are more demanding in terms of computational time.
94 Therefore, if the use of multiple hidden layers is not required, then an RBF network can be
95 implemented saving computational time and appealing to a broader audience. However, the data
96 set consists of a modest number of samples (775). Also, the data set does not consist of real-time
97 data from a PV-system, but rather it is simulated using MATLAB/Simulink. Conversely, the data
98 sample used in our proposal consists of 97200 samples over a 10-week, obtained from a live
99 installation, refer to section 2.

100 Yuchuan Wu *et al.* [25] looks at the limitations of Fuzzy Logic used in early stages of fault
101 detection in PV systems. It highlights barriers such as the process for obtaining fuzzy rules and
102 membership function, along with the constancy of fuzzy systems. This method is implemented
103 using BP neural networks for fault detection in PV systems, resolving the issues faced with Fuzzy
104 logic through its ability to better self-learning, self-adaptability and non-linearity pattern
105 recognition.

1.2 Contribution and Paper Organization

107 The main contribution of this work is to present a novel algorithm that can carry out fault detection
108 in a PV system, to a high degree of accuracy, requiring only two inputs. This is done through the

109 implementation of an RBF network for the classification of the faults presented in section 3.2. The
110 rationale for the selection of RBF over Multilayer Perceptron Networks (MLP) was due to the
111 network only requiring a single hidden layer but more importantly due to RBF's robustness to
112 adversarial samples in the data set. As proven in section 3.1, the ANN architecture accuracy was
113 compared with a varying number of hidden layers. The results showed that a single hidden layer
114 was the most optimal solution providing an accuracy of almost 99% while consuming the least
115 amount of computational time. As a result of this an RBF network was selected rather than an
116 MLP. Although MLP can also be used in a single hidden layer configuration, it demands more
117 computational power, further discussed in section 3.8.

118 Rest of the article is organized as follows; Section 2 presents the examined PV installations. In
119 section 3 we discuss in detail the rationale for selecting RBF over MLP and look at the structure
120 of the proposed network along with the four different ANN-based methodologies to detect faults
121 in PV systems. Section 4, looks at the results of the network. In section 5, we compare our
122 developed ANN network with recent ANN-based models available in present literature. Finally,
123 sections 6 and 7 present the conclusion and reference list, respectively.

2. Examined PV system

125 The overall system design is shown in Figure 1. The PV plant consists of 10 PV modules set-up
126 in string topology, irradiance sensor, MPPT unit and DC- load. The input/output pins of the MPPT
127 unit are linked via Ethernet-capable to a personal computer (PC) to facilitate real-time data
128 monitoring. The proposed ANN algorithm for fault detection of the PV modules is developed in
129 MATLAB software.

130 As shown in Figure 2, the PV plant consisting of 10-polycrystalline silicon PV modules, with a
131 nominal power of 220 W (per module), the electrical parameters under 'standard test conditions'
132 (STC) of the PV modules are shown in Table 1; STC of the PV modules at solar irradiance = 1000
133 W/m^2, module temperature = 25 °C, spectral distribution of the incident light according to AM 1.5
134 and irradiation perpendicular onto the receiving plane. The Maximum Power Point Tracker
135 (MPPT) has an output efficiency of not less than 95.0%. Internal sensors within the MPPT are
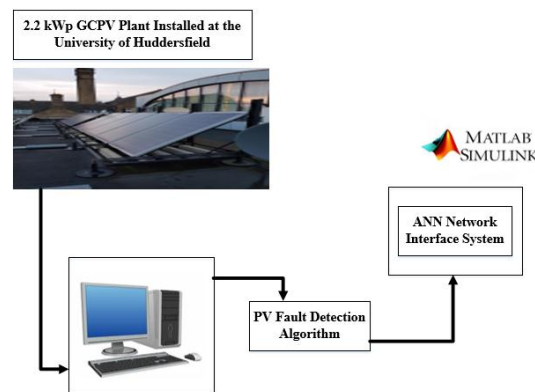136 used to measure the DC current and voltage.

Figure 1. Overall System Architecture Design for the Examined PV Plant

Table 1. Electrical characteristics of SMT6 (60) P PV module

Solar Panel Electrical Characteristics	Value
Peak Power	220 W
Voltage at maximum power point (V_{mpp})	28.7 V
Current at maximum power point (I_{mpp})	7.67 A
Open Circuit Voltage (V_{oc})	36.74 V
Short Circuit Current (I_{sc})	8.24 A
Number of cells connected in series	60
Number of cells connected in parallel	1
Series resistance R_S	0.53 Ω
Parallel resistance R_{sh}	1890 Ω

137 Davis weather station measures the global solar irradiance, which is passed onto the monitoring
138 unit connected to the PC for data recording and monitoring. A Hub 4 communication manager
139 facilitates acquisition of modules temperature via the Davis external temperature sensor, as well
140 as the electrical data for each photovoltaic string.

141 The weather station is located on the same level and position of the PV modules, as presented in
142 Figure 2(a). The weather station is mounted near to the examined PV system, the solar irradiance
143 is measured using a pyranometer which has a resolution of ±2 W/m^2, while the angle of incident
144 is set the same as the PV modules inclination of 37 degrees.

145 If the angle of inclination is changed, therefore, the results of the solar irradiance would typically
146 impact the ANN detection accuracy, since the solar irradiance against the output predicted power
147 would be expected to be inaccurate. For generalization purposes, it would be more appropriate to
148 use a mathematical modeling for the solar irradiance which can predict the output irradiance on a
149 particular location including the inclination of a typical PV system, this was not the case in our
150 model, as we ensure that the pyranometer is on the same inclination as the examined PV
151 installations.

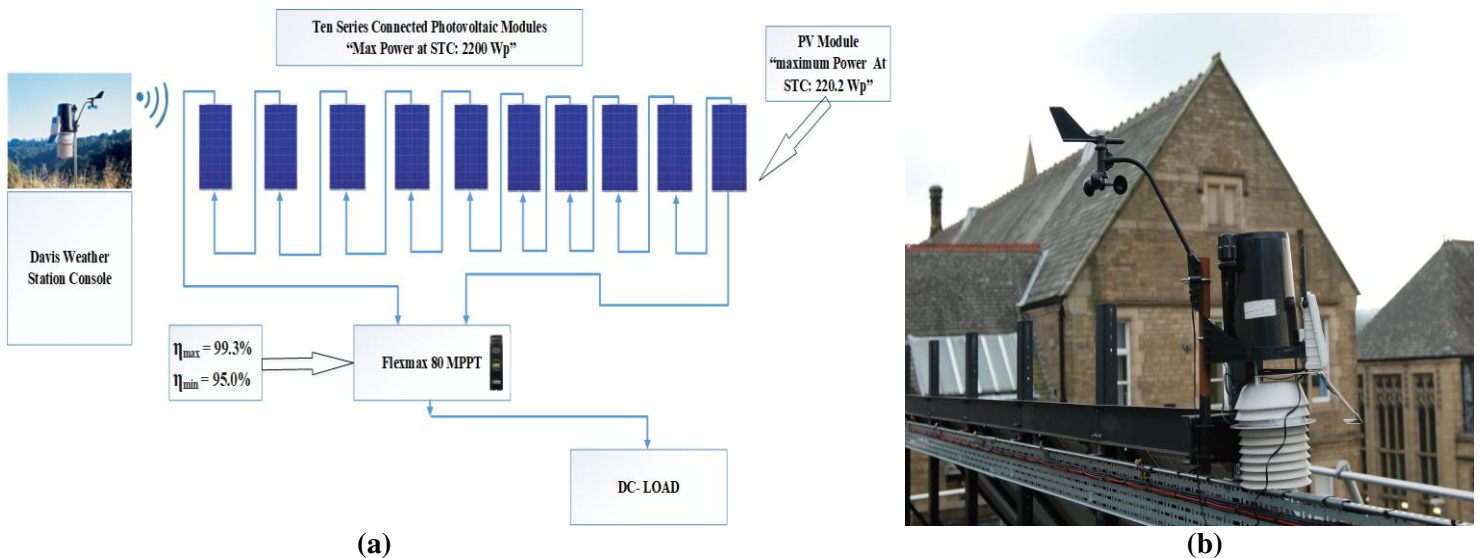(a)   (b)

Figure 2. **(a)** Examined PV System layout including the weather station and DC-Load, **(b)** weather station mounted in the examined PV installation

In this article, we have taken a data samples of the PV installation captured over a commencing period of November 2019 to February 2020. As a result, it might be useful to outline that during summer period, when the irradiance and ambient temperature is expected to have a high peak, the prediction and accuracy of any typical ANN network might differ. Hence, in the following section, we have used four different data input setups to overcome this issue.

3. Methodology

Researchers have demonstrated various methodologies for data normalisation, training, validation and testing of ANN networks. To the best of our understanding, the majority of the implemented methodologies found in recent literature are based on data sets, consisting of several inputs to the network [5, 21 & 24]. The following sub-sections demonstrate in detail the four methodologies that were implemented on the sample data set consisting of solar irradiance and total output power. The data consisted of 10 weeks, one week worth of data for every fault, starting with normal operation (NO). The ANN was trained with each methodology and the overall detection accuracy for the four methodologies is used as a reference to the success of each.

3.1 ANN Structure

The purpose of the ANN was to detect faults in PV modules, as shown in Figure 3. Before the ANN could be trained with a sample data set, the first step was the formation of a strategy, to test the optimal process for selecting the sample data to train the ANN. For this critical task, four different methodologies for data processing were investigated:

1. Methodology 1 (M1): solar irradiance and output power was randomly selected, including all off-state data (where power is equal to zero), no normalisation process was involved.
2. Methodology 2 (M2): solar irradiance and output power was randomly selected, all off-state data was dismissed, no normalisation process was implemented.
3. Methodology 3 (M3): solar irradiance and output power was randomly selected and normalised using the max-min normalisation technique.
4. Methodology 4 (M4): solar irradiance and output power was randomly selected and normalised. In addition, solar irradiance of 0 - 1000 W/m^2 was mapped along with the output power.
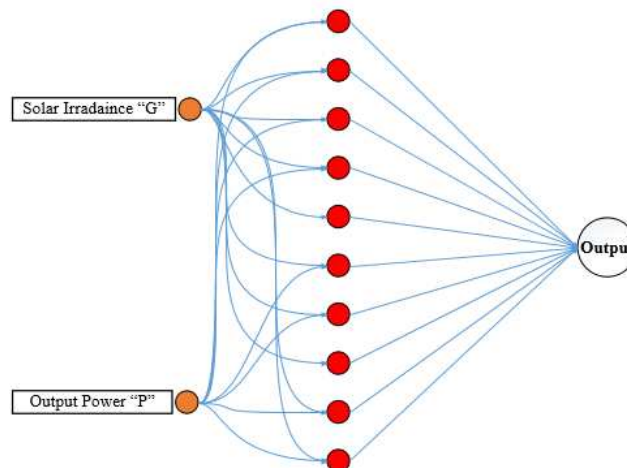

Figure 3. The proposed RBF architecture

180 The architecture of the ANN was based on the Radial Basis Function (RBF). The network was
181 made up of three parts (input, hidden layer and output). For this research, a single hidden layer
182 was implemented, for all investigative methods, due to the linear nature of the response, high
183 accuracy and minimal computational time. The hidden layer can be increased for further accuracy
184 and computational performance, depending on the application under consideration. Increasing the
185 number of hidden layers would convert the network into an MLP, increasing the computational
186 time. The developed ANN architecture using one hidden layer achieved a high rate of detection
187 accuracy, almost equal to 99%; this will be discussed in the following sections.

188 In principle, each neuron takes a formed linear combination of the outputs of previous neurons.
189 This linear combination is weighted through the strength between the neurons (w_{ij}) and multiplied
190 by the input (x_j). Further, the activation threshold (w_{j0}) is also assigned to each neuron. This
191 process is expressed using (1). Note: i is equal to number of hidden neurons (1, 2, 3 … 10), j is
192 equal to number of inputs (1 and 2).

193
$$\sum_{j=1}^{n}(w_{ij}x_j + w_{j0}) \tag{1}$$

194 Next, the weighted activation process is then multiplied by the non-liner function f_1 as shown in
195 (2), this is usually a sigmoid function (MLP) or Euclidian function (RBF). Finally, the output value
196 of the hidden layers y_i is expressed by (3).

197
$$f_1 \times \sum_{j=1}^{n}(w_{ij}x_j + w_{j0}) \tag{2}$$

198
$$y_i = f(u) = \frac{1}{1 + e^{-\sum_{j=1}^{n}(w_{ij}x_j + w_{j0})}} \tag{3}$$

199 in order to enhance the performance of the selected ANN structure, we have used the quick
200 prorogation method, expressed by (4), where the tanning iteration number is equal to 1000.

201
$$\Delta w(s) = \frac{\nabla\delta|w(s)}{\nabla\delta|w(s-1) - \nabla\delta|w(s)} \Delta w(s-1) \tag{4}$$

202 where δ is the error function of the simulation process, w is the vector of the weights for the
203 developed ANN network shown in Figure 3, and s is the iteration number. The best layout of the
204 developed ANN network is also shown in Figure 4.

205 The developed structure of the ANN network is shown in Figure 4(a). The RBF network with two
206 inputs, one hidden layer and 10 hidden neurons is selected. In fact, the selection of the inputs was
207 obtained using the available parameters from dataset, including the solar irradiance (G) and the
208 output power (P). The selection of the hidden layers is obtained using an extensive simulation
209 from 1 to 100 hidden layers; as a result, ten hidden layers were selected due to its optimum
210 performance. The results of the ANN network accuracy vs the number of hidden neurons used for
211 the considered methodologies (M1, M2, M3, and M4) is shown in Figure 4(b). It is noticed that
212 the last adopted methodology achieved the highest ANN detection accuracy of 98.6% using 10
213 hidden neurons. The minimum ANN accuracy of 40.2% is observed for the first methodology
214 using 90-100 hidden neurons.

Input Layer   Hidden Layer   Output Layer

w_{ij}
$i = 1, 2, ..., 10$

x_j
$j = 1, 2$

$j = 1, 2$

y_i
$i = 1, 2, 3, ..., 10$

Output

Solar Irradaince "G"

Output Power "P"

Output

Outputs

$$y_i = f(u) = \frac{1}{1 + e^{-\Sigma_{j=1}^{n}(w_{ij}x_j + w_{j0})}}$$

$$\sum_{j=1}^{n}(w_{ij}x_j + w_{j0})$$

$$f_1 \times \sum_{j=1}^{n}(w_{ij}x_j + w_{j0})$$

Neuron

Activation Function

(a)

Maximum

98.3 98.1 98.6 98.3 98.1 98 98 98.1 97.7 97.7 97.6 97.8 97.2 97 97 97 97.3 97.1 97.6 97.8 98.1

M1
M2
M3
M4

ANN Accuracy (%)

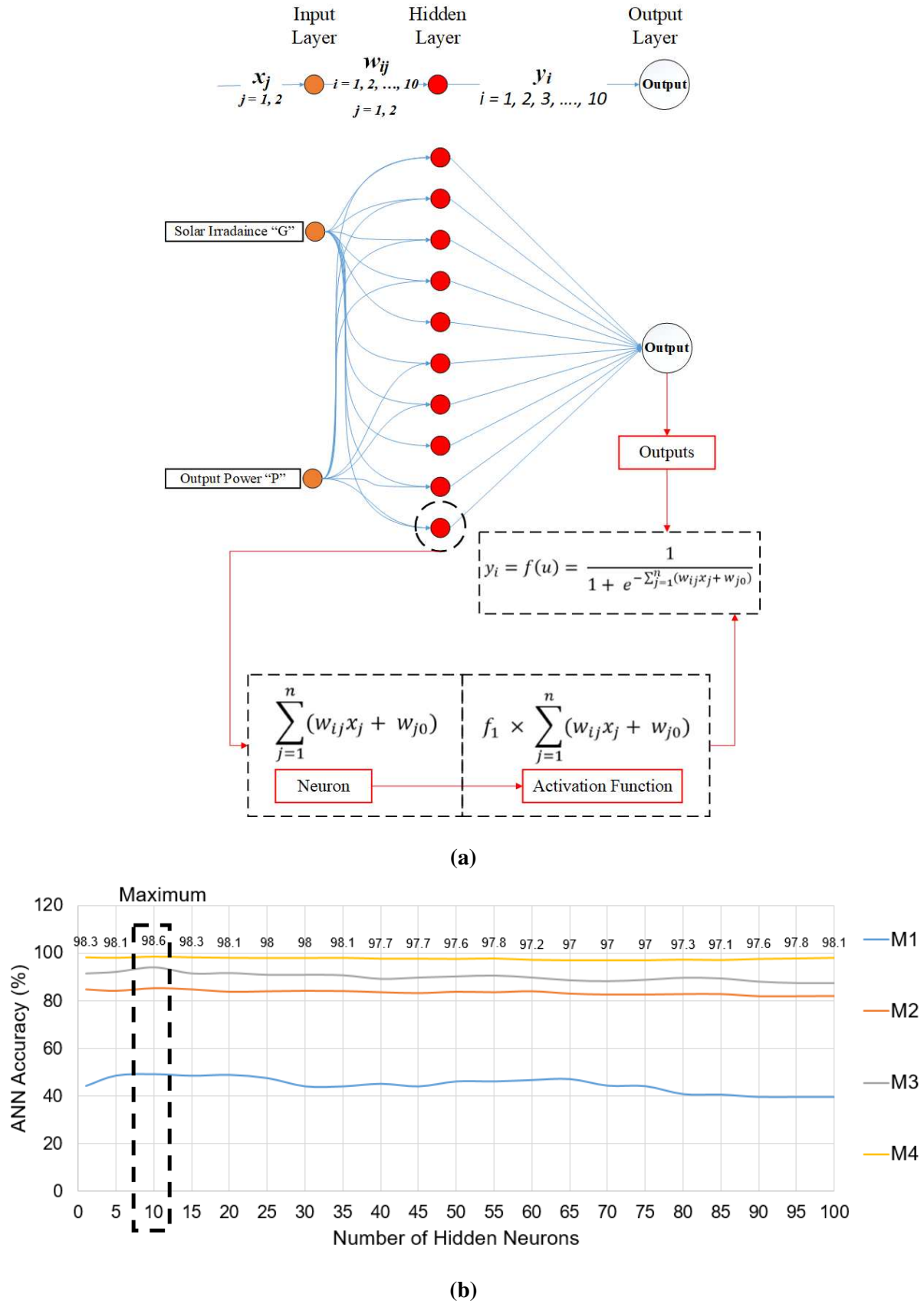Number of Hidden Neurons

(b)

Figure 4. **(a)** Details of the proposed ANN network architecture, **(b)** ANN accuracy vs number of hidden neurons

215

8

3.2 ANN Network Training and Validation

Before the training and validation process began, the faulty conditions considered for the ANN to detect had to be identified. In this article, 10 different scenarios have been taken into consideration, presented as follows:

- Case1: Normal operation mode, where no faults were applied to the PV string
- Case 2: 1 Fault applied to the system; 1 PV module disconnected from the PV string
- Case 3: 2 Faults applied to the system; 2 PV modules disconnected from the PV string
- Case 4: 3 Faults applied to the system; 3 PV modules disconnected from the PV string
- Case 5: 4 Faults applied to the system; 4 PV modules disconnected from the PV string
- Case 6: 5 Faults applied to the system; 5 PV modules disconnected from the PV string
- Case 7: 6 Faults applied to the system; 6 PV modules disconnected from the PV string
- Case 8: 7 Faults applied to the system; 7 PV modules disconnected from the PV string
- Case 9: 8 Faults applied to the system; 8 PV modules disconnected from the PV string
- Case 10: 9 Faults applied to the system; 9 PV modules disconnected from the PV string

It is worth noting that the partial shading considered in this research is when a PV module is affected by either shading caused by moving clouds or when an overcasting weather condition is arisen. Hence, no applied shading was practiced as we tried the best to comply with real shading scenarios rather than using opaque objects like some research do. For that reason, we took long-term data measurements of the PV installation over a period of 10 weeks, rather than simply reliant on simulated or applied shading conditions which can be operated over couple hours.

Figure 5 shows a flowchart of the proposed fault detection architecture. Initially, the measured output power of the PV string was attained using the MPPT unit. If the output power was greater than zero, the measured power was passed into the developed ANN network. In case the output was equal to zero, verification of the measured voltage had to be carried out to decide whether the PV string was faulty (voltage > 0), or in sleep mode (V = 0).
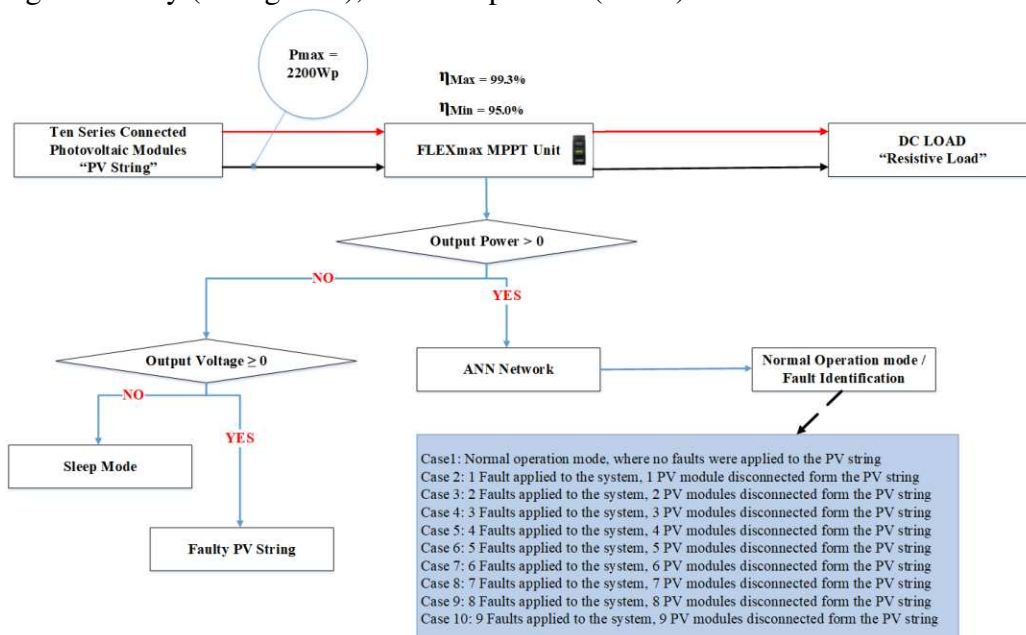

Figure 5. Flowchart of the proposed fault detection algorithm

241   In order to practically testify the examined faulty scenarios, we have used the junction-box placed
242   next to the PV installation in which we can apply any faulty condition to the PV string, and this
243   junction-box can be configured manually using a switch-connection where a PV module(s) can be
244   connected or disconnected from the PV string.

245   For instance, Figure 6 shows three case scenarios applied to the PV installation including case 1
246   "normal operational mode", case 2 "one PV module is disconnected from the PV string", and case
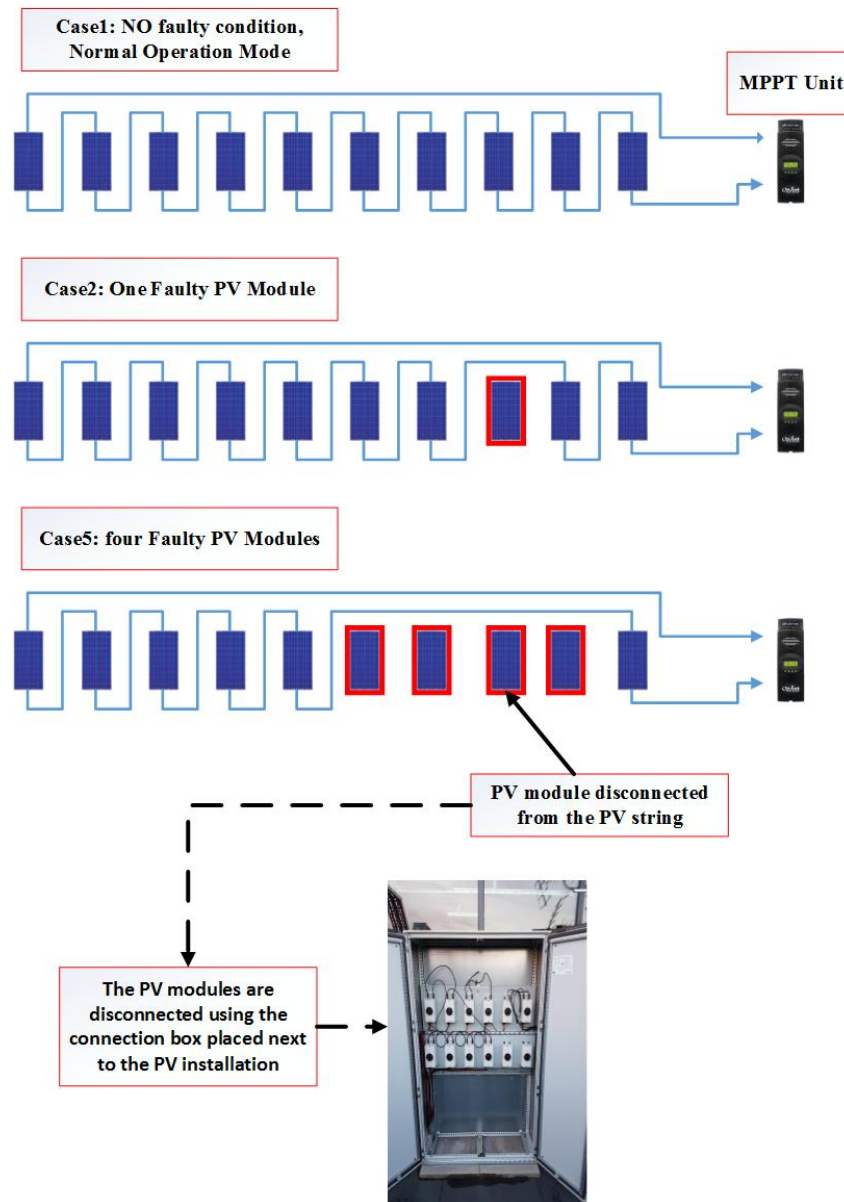247   5 "four PV modules are disconnected from the PV string".

Figure 6. Schematic shows two applied cases on the examined PV installation

248　The data set used for training the ANN was recorded from the experimental PV setup shown
249　previously in Figure 2. During the experiments, the value of solar irradiance and total power were
250　logged, with the PV modules' temperature between 9.8 – 24.6 ^{0}C. As the proposed ANN model
251　does not require the temperature of the solar modules; hence, this value was not taken into
252　consideration. The data set shown in Figure 7 of the solar irradiance and the output power, consists
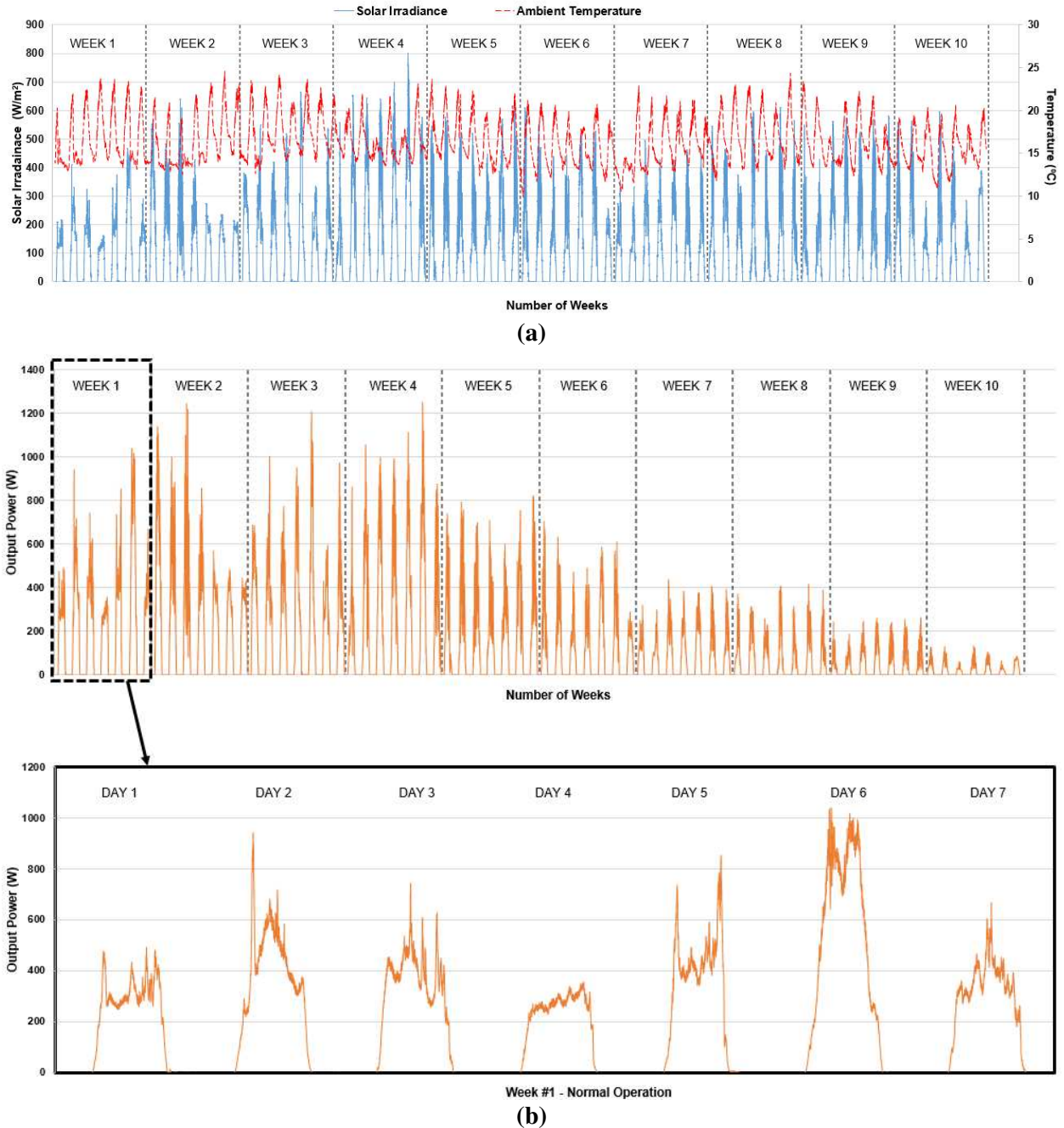253　of 97200 measurements captured for 10 weeks. Each week corresponds to a different condition.

(a)

(b)

Figure 7. Data set used for training purposes; 97200 samples, each scenario has 9720 samples gathered over ten weeks. **(a)** Solar irradiance, **(b)** Output PV power

254 Various methodologies available in literature [5, 21, 24, and 26] used training samples for one to
255 three days for faulty conditions. While in our case, the data was recorded over a duration of one
256 week for every faulty condition. Another limitation of the recent work, 2019, conducted by [27
257 and 28] is that the ANN networks were trained using various inputs such as PV string voltage,
258 current, solar irradiance, power and ambient temperature. Whereas our proposed ANN network
259 only requires solar irradiance and output power as an input for the network. Note, a brief
260 comparison of the accuracy will be discussed later in section 5.

261 It is worth noting as the number of faults increases on a week-by-week basis, the total output power
262 measured for the PV system deteriorated, as shown in Figure 8.
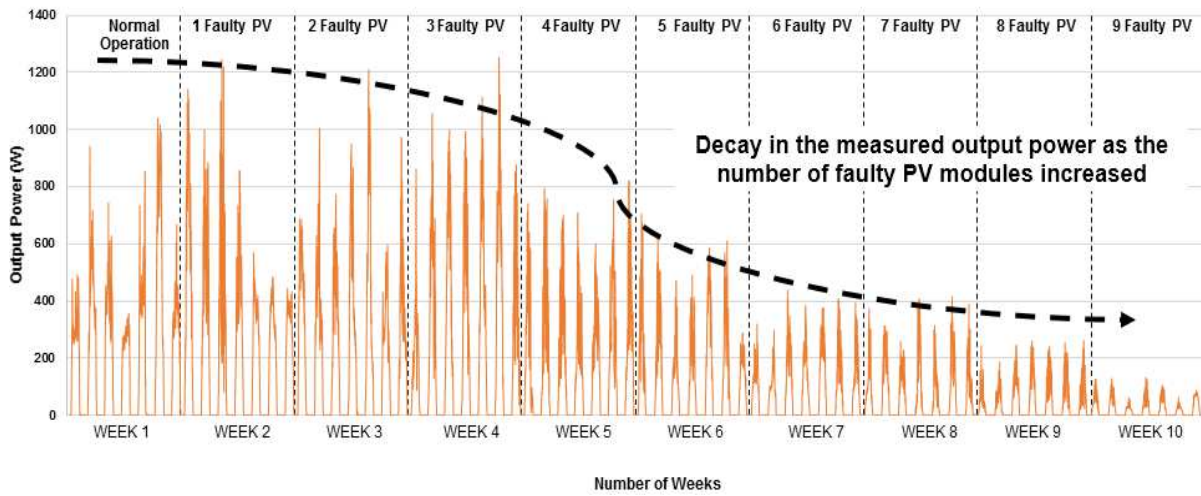
Figure 8. Flowchart of the proposed fault detection algorithm

263 As part of some methodologies (M3 and M4) which will be discussed in the following sub-
264 sections, the normalisation process of the input data for the solar irradiance and output power had
265 to be carried out.

266 The standard input parameters used to configure all tested ANN models are the solar irradiance
267 (G) and total output power (P). The Data set (input variables) are normalised using the max-min
268 normalisation technique, within the range of 0 and +1 using (5).

269
$$y = \frac{(y_{max} - y_{min})(x - x_{min})}{(x_{max} - x_{min})} + y_{min} \tag{5}$$

270 where x ∈ $\{xmin, xmax\}$ is the original data value, and y ∈ $\{ymin, ymax\}$ is the corresponding
271 normalised value with $ymin = 0$ and $ymax = +1$."

3.3 Implementation and Validation of the First Methodology (M1)

The ANN network was trained with randomly selected solar irradiance and output power, including non-representative data (represented by a zero for power), no normalisation of sample data was carried out in this approach. The ANN achieved an overall detection accuracy of 49.4%, refer to Figure 9(a). Many factors could have contributed to the low accuracy of the network. One possible factor was the inclusion of 'non-representative' data denoted by a zero, either for solar irradiance or PV output power.

3.4 Implementation and Validation of the Second Methodology (M2)

This methodology looks to improve the accuracy of the ANN by addressing the possible cause highlighted in M1. The input parameters (solar irradiance and output power) were randomly selected, but all 'non-representative' data was removed from the sample set which was to be used for training the network. The ANN network achieved an improved overall accuracy of 85.4%, refer to Figure 9(b). However, this is still considered as low accuracy for an ANN. This could have been due to no normalisation of data being carried out on the selected sample set.

For ANN networks training and validation, every dataset does not require normalisation. It is required only when features have different ranges. In our case, normalisation process would be expected to enhance the accuracy of the ANN as both ANN inputs, solar irradiance and output power, have a divergent range, i.e. the solar irradiance ranges from 0-1000 W/m^2, while the output power ranges from 0 to 2200 W.

The green and red cells of the matrix represent the number of correct and incorrect classifications by the ANN, respectively. The grey cells represent the total detection accuracy with respect to each row and column. The number 1, represents 1- fault (F1) and so on, ending at 10- faults (F10).
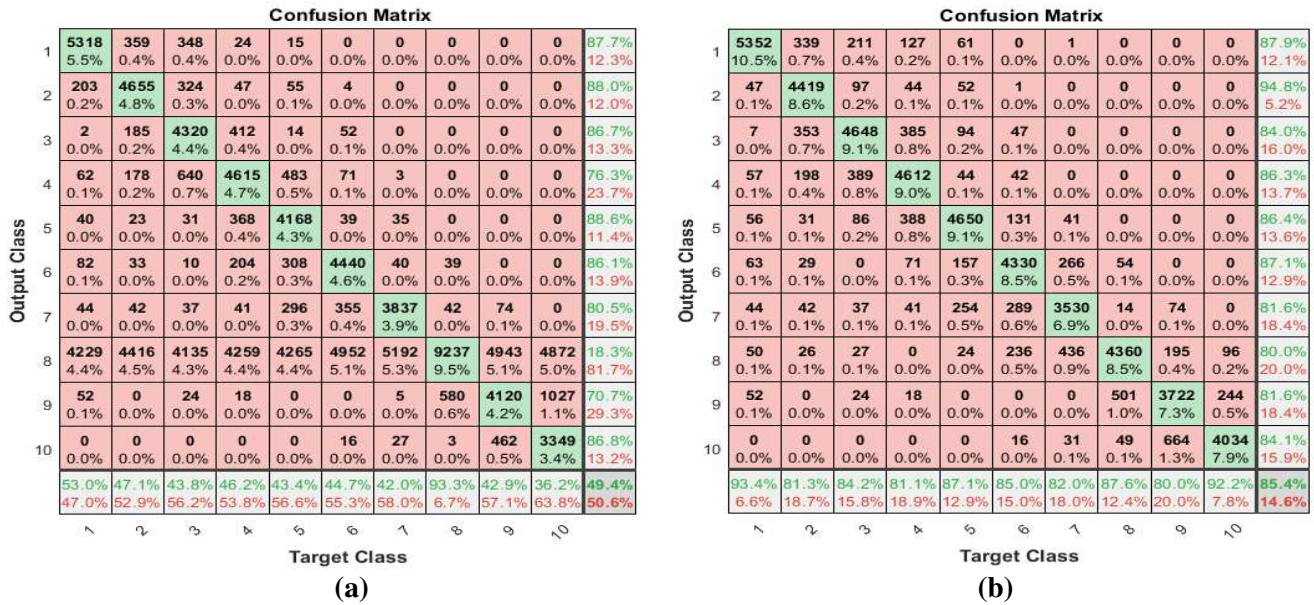
Confusion Matrix (a) — M1

Output Class (rows 1–10), Target Class (columns 1–10):

	1	2	3	4	5	6	7	8	9	10	
1	5318 / 5.5%	359 / 0.4%	348 / 0.4%	24 / 0.0%	15 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	87.7% / 12.3%
2	203 / 0.2%	4655 / 4.8%	324 / 0.3%	47 / 0.0%	55 / 0.1%	4 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	88.0% / 12.0%
3	2 / 0.0%	185 / 0.2%	4320 / 4.4%	412 / 0.4%	14 / 0.0%	52 / 0.1%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	86.7% / 13.3%
4	62 / 0.1%	178 / 0.2%	640 / 0.7%	4615 / 4.7%	483 / 0.5%	71 / 0.1%	3 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	76.3% / 23.7%
5	40 / 0.0%	23 / 0.0%	31 / 0.0%	368 / 0.4%	4168 / 4.3%	39 / 0.0%	35 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	88.6% / 11.4%
6	82 / 0.1%	33 / 0.0%	10 / 0.0%	204 / 0.2%	308 / 0.3%	4440 / 4.6%	40 / 0.0%	39 / 0.0%	0 / 0.0%	0 / 0.0%	86.1% / 13.9%
7	44 / 0.0%	42 / 0.0%	37 / 0.0%	41 / 0.0%	296 / 0.3%	355 / 0.4%	3837 / 3.9%	42 / 0.0%	74 / 0.1%	0 / 0.0%	80.5% / 19.5%
8	4229 / 4.4%	4416 / 4.5%	4135 / 4.3%	4259 / 4.4%	4265 / 4.4%	4952 / 5.1%	5192 / 5.3%	9237 / 9.5%	4943 / 5.1%	4872 / 5.0%	18.3% / 81.7%
9	52 / 0.1%	0 / 0.0%	24 / 0.0%	18 / 0.0%	0 / 0.0%	0 / 0.0%	5 / 0.0%	580 / 0.6%	4120 / 4.2%	1027 / 1.1%	70.7% / 29.3%
10	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	16 / 0.0%	27 / 0.0%	3 / 0.0%	462 / 0.5%	3349 / 3.4%	86.8% / 13.2%
	53.0% / 47.0%	47.1% / 52.9%	43.8% / 56.2%	46.2% / 53.8%	43.4% / 56.6%	44.7% / 55.3%	42.0% / 58.0%	93.3% / 6.7%	42.9% / 57.1%	36.2% / 63.8%	49.4% / 50.6%

Confusion Matrix (b) — M2

Output Class (rows 1–10), Target Class (columns 1–10):

	1	2	3	4	5	6	7	8	9	10	
1	5352 / 10.5%	339 / 0.7%	211 / 0.4%	127 / 0.2%	61 / 0.1%	0 / 0.0%	1 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	87.9% / 12.1%
2	47 / 0.1%	4419 / 8.6%	97 / 0.2%	44 / 0.1%	52 / 0.1%	1 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	94.8% / 5.2%
3	7 / 0.0%	353 / 0.7%	4648 / 9.1%	385 / 0.8%	94 / 0.2%	47 / 0.1%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	84.0% / 16.0%
4	57 / 0.1%	198 / 0.4%	389 / 0.8%	4612 / 9.0%	44 / 0.1%	42 / 0.1%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	86.3% / 13.7%
5	56 / 0.1%	31 / 0.1%	86 / 0.2%	388 / 0.8%	4650 / 9.1%	131 / 0.3%	41 / 0.1%	0 / 0.0%	0 / 0.0%	0 / 0.0%	86.4% / 13.6%
6	63 / 0.1%	29 / 0.1%	0 / 0.0%	71 / 0.1%	157 / 0.3%	4330 / 8.5%	266 / 0.5%	54 / 0.1%	0 / 0.0%	0 / 0.0%	87.1% / 12.9%
7	44 / 0.1%	42 / 0.1%	37 / 0.1%	41 / 0.1%	254 / 0.5%	289 / 0.6%	3530 / 6.9%	14 / 0.0%	74 / 0.1%	0 / 0.0%	81.6% / 18.4%
8	50 / 0.1%	26 / 0.1%	27 / 0.1%	0 / 0.0%	24 / 0.0%	236 / 0.5%	436 / 0.9%	4360 / 8.5%	195 / 0.4%	96 / 0.2%	80.0% / 20.0%
9	52 / 0.1%	0 / 0.0%	24 / 0.0%	18 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	501 / 1.0%	3722 / 7.3%	244 / 0.5%	81.6% / 18.4%
10	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	0 / 0.0%	16 / 0.0%	31 / 0.1%	49 / 0.1%	664 / 1.3%	4034 / 7.9%	84.1% / 15.9%
	93.4% / 6.6%	81.3% / 18.7%	84.2% / 15.8%	81.1% / 18.9%	87.1% / 12.9%	85.0% / 15.0%	82.0% / 18.0%	87.6% / 12.4%	80.0% / 20.0%	92.2% / 7.8%	85.4% / 14.6%

(a)   **(b)**

Figure 9. Output confusion matrix. **(a)** M1, **(b)** M2

3.5 Implementation and Validation of the Third Methodology (M3)

295 For the development of the third methodology, we have considered the previous methodologies
296 (M1 and M2), but also included the normalisation of the sample set data before training the ANN
297 network. The overall detection accuracy of the ANN using this approach was 94.2%, refer to
298 Figure 10. An improvement can be seen as compared to the above approaches due to the
299 normalisation of data before training the ANN. However, normalisation does not cater for missing
300 data samples between two data points. For example, the normalised data of the solar irradiance
301 may include values from 500 and 505 but not the values in between. This issue is further explored
302 in the implementation of the fourth ANN architecture.

Figure 10. Output confusion matrix obtained using M3

3.6 Implementation and Validation of the Fourth Methodology (M4)

304 The normalised data of the solar irradiance or output power may include value, for example, 500
305 and 505 but not the values in between. This method implemented the concept of 'mapping' sample
306 data set, before feeding it into the ANN for training, validating and testing purposes. Through the
307 implementation of mapping input parameters, all values within the defined data points are taken
308 into account. The ANN was trained using randomly selected data set, no "off-state" data was
309 involved, normalisation process had been implemented, and finally the novel part of this approach,
310 the mapping feature of the solar irradiance against the PV system output power was implemented.

311 The solar irradiance 0 - 1000 W/m^2 was mapped along with the corresponding output power as
312 shown in Figure 11(a), with a step size of 1 W/m^2. The accuracy of the ANN improved
313 significantly (98.6%) compared to the prior methods, as presented in Figure 11(b).

As suggested, the rationale behind the mapping of the inputs was to try and obtain a complete dataset before training the network. This was made possible due to only having two inputs. However, datasets containing multiple variables such as temperature, wind speed, voltage along with irradiance and power would make the task more difficult.
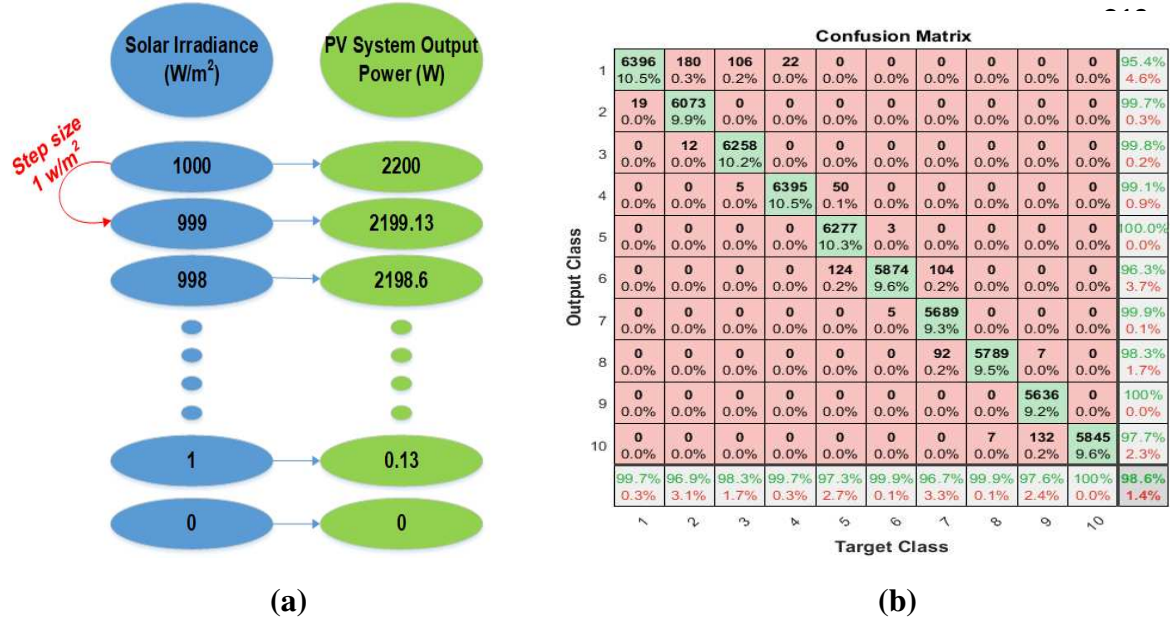

Figure 11. **(a)** Mapping solar irradiance and PV system output power, **(b)** Confusion Matrix for M4

3.7 Evaluation of the Four Developed Methodologies

The overall detection accuracy of the ANN based on the four methodologies is presented in Table 2. A key observation to take away from Table 2, is the importance of data normalisation before proceeding with the training of the ANN. M1, had the least detection accuracy (49.4%). This was primarily because the raw data extracted from the PV setup shown in Figure 2, was directly used to train the ANN without filtering noisy data. The dismissal of 'off-state' data (M2), significantly improved the accuracy of the network (85.4%). Normalisation (M3), further improved the detection accuracy (92.2%) thanks to the sync of the input range, with (M4) achieving the highest detection accuracy due to the mapping of inputs to cater for missing data points.

Table 2. Evaluating all developed methodologies

Methodology No.	Training data set including solar irradiance and output power	Training Data does not include "off-state" samples	Normalisation Process of the training data set	Mapping 0-1000 W/m^2 into the actual output power	ANN network detection accuracy (%)
M1	✓	✗	✗	✗	49.4
M2	✓	✓	✗	✗	85.4
M3	✓	✓	✓	✗	92.2
M4	✓	✓	✓	✓	98.6

338 The critical element in increasing the ANN-network detection accuracy is principally due to the
339 increase of the ANN receiver operating characteristics (ROC) during the training, validation and
340 testing stages. The ROC is a graph showing the performance of a classification model at all
341 classification scenarios (i.e. class 1 corresponds to case1 "normal operation mode", while class 10
342 corresponds to case 10 "9 faulty PV modules").

343 As can be noticed in Figure 12(a), during the training and validation of methodology 1, the ROC
344 tends to have a high false-positive rate which at the end reduces the overall detection accuracy of
345 the ANN network. Comparatively, Figure 12(d) shows the ROC for the last methodology, while it
346 is evident that the true-false rate has significantly decreased, there is an increase in the true-positive
347 rate due to the impact of the normalisation and the mapping procedure implemented during the
348 data handling processing stage, resulting in the highest rate of detection accuracy of 98.6%.
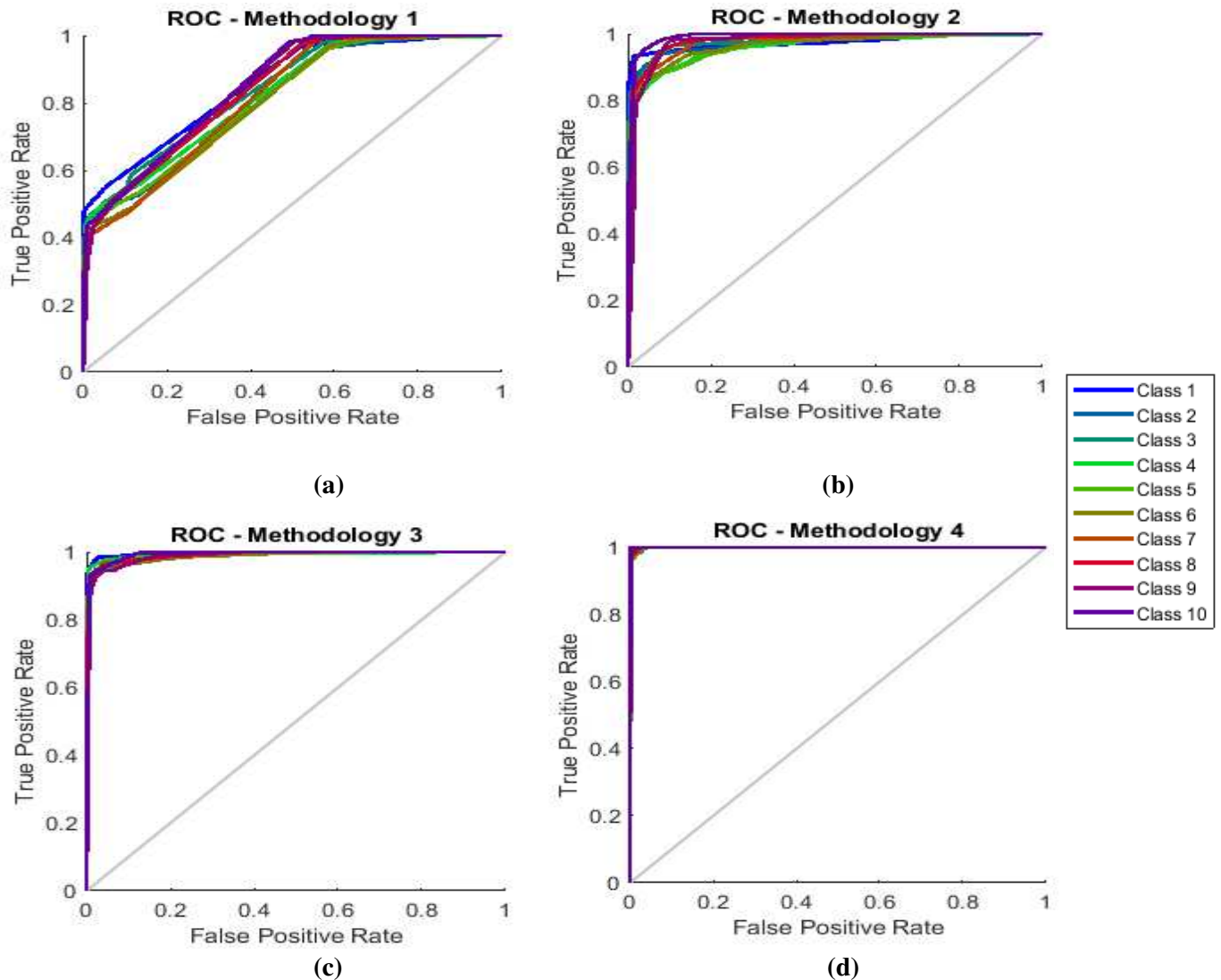
Figure 12. ROC performance of each ANN methodology. **(a)** M1, **(b)** M2, **(c)** M3, **(d)** M4

16

349 According to what has been discussed so far, it is worth noting that the rationale for the selection
350 of RBF over Multilayer perceptron networks (MLP) was due to the network only requiring a single
351 hidden layer but more importantly due to RBF's robustness to adversarial samples in the data set.
352 As proven in this section, the ANN architecture accuracy was compared with a varying number of
353 hidden layers. The results showed that a single hidden layer was the most optimal solution
354 providing an accuracy of almost 99%. As a result of this an RBF network was selected rather than
355 an MLP. Although MLP can also be used in a single hidden layer configuration, it demands more
356 computational power.

357 The first novelty this research brings to the field of PV fault detection through ANN, is the use of
358 only two inputs for training the network, i.e. solar irradiance and output power. While the authors
359 acknowledge that the introduction of smart meters provide an effective platform for obtaining
360 various inputs used for training ANN, without having major implications in terms of more
361 hardware, there are other issues which are introduced as a result. The use of, for example, a five
362 inputs ANN-based fault detection system rather than two means more time must be spent on data
363 processing. It is, in fact, this stage of the process in the development of an ANN which directly
364 impacts the overall accuracy of the network. If the network requires five inputs, the smart meter
365 will be able to provide this data, but the problem occurs when the data must be processed for
366 training the network as additional inputs result in more time required for data censoring and higher
367 chances of non-representative data making its way into the network due to human error. Whereas,
368 by limiting the inputs to only two, as proposed in this article, less time and effort is required for
369 data processing, less chances of noisy data leaking into the training stage and ultimately a higher
370 performing network, as evident from section 4.

371 In addition, as we have used solar irradiance data, academics, industry and national interest alike
372 are demanding this data for various reasons such as predicting the ambient temperature, techno-
373 economic analysis of the heating systems, transportation sector, etc. Therefore, we do not consider
374 acquiring solar irradiance is a major drawback of this research, since, as an example, the UK Met-
375 office[1] have a live grid climate variable system to download live solar irradiance as well as
376 forecasted data; this approach of making live data for public use is now widely held in various
377 counties such as Spain, Italy, and Singapore.

378 Another novel part of this research (principally using the fourth methodology) is the mapping of
379 the inputs after removing non-representative data and carrying out data normalisation. A
380 fundamental component within this methodology which plays a vital role in achieving high
381 accuracy is the selection of the 'step size' for the mapping. A step size of 'one' was selected as
382 shown in Figure 11. The rationale for not selecting a step size of '10' was due to under-fitting. As
383 the gap between the data points would make it difficult for the network to be able to generalize.
384 Conversely, a step size of less than one, for example '0.5', would force the network to accurately
385 map the data points without generalizing, resulting in a network that is unable to accurately classify
386 new input data, rendering it as an 'over-fit' network.

3.8 RBF vs. MLP

RBF and Multi-Layer Perceptron (MLP) are both neural networks as shown in Figure 13. Both networks have some common features but also significant differences in the way they operate, which makes one network more suited for certain applications as compared to the other. The well-known difference between the two networks is that RBF only uses a single hidden layer, whilst MLP can accommodate multiple hidden layers.

Both RBF and MLP networks can be used for regression and classification problems. Also, both networks have the capability to approximate complicated functions. However, RBF's have explicit local representations as each neuron represents a specific region of the input space. Conversely, in an MLP network, each neuron tries to capture a specific feature from the training set. Early neuron layers capture low-level features and as the information propagates through the hidden layers the feature extraction matures. This can optimize the overall accuracy of the network but at the expense of increased computational power, a trade-off which comes down to the objectives of the project. As there is only one hidden layer in an RBF network, each neuron captures the similarity between the whole training set and the center of the Euclidian.
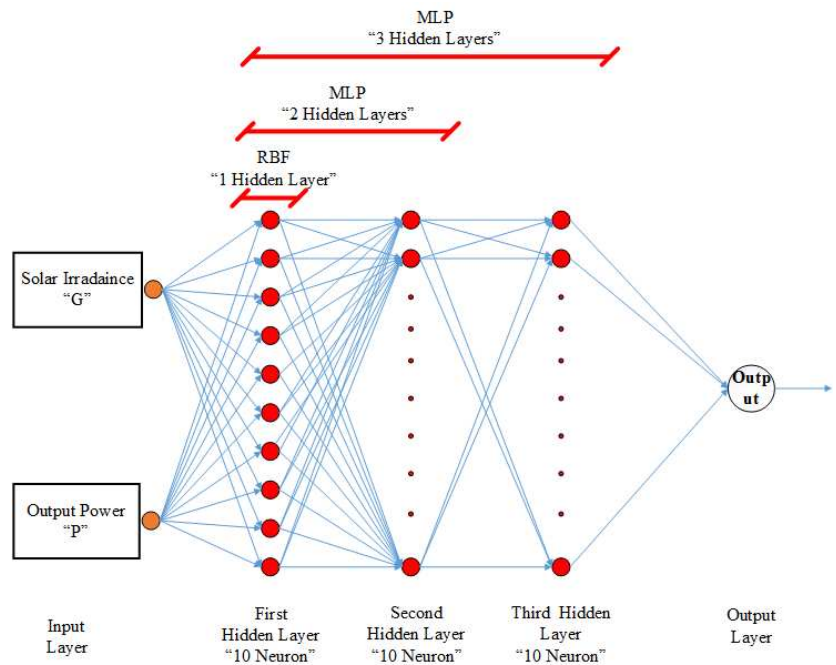
Figure 13. RBF vs. MLP

Moreover, the RBF is more adequate compared with the MLP is due to the digital nature of the failure associated with the PV system; the individual PV modules are completely "ok" or completely "off", where no "intermediate" failure occurs.

The selection of the ANN network was based on the results obtained from testing and comparison of the two networks in terms of the overall detection accuracy and required computational time. Figures 14 (a) and (b), show the overall detection accuracy of two MLP networks with 2 and 3 hidden layers, respectively. There is only a small difference in the accuracy (+0.2%) whereas the computational time increased by 155ms, refer to Figure 14 (c).

When comparing an RBF with an MLP network consisting of 3 hidden layers the detection accuracy increased by a small margin (+0.5%) whilst the corresponding computational time

increased considerably from 45ms to 290ms. Consequently, increasing the time of which a PV fault can be detected as well as the overall energy consumption of the microcontroller power unit.
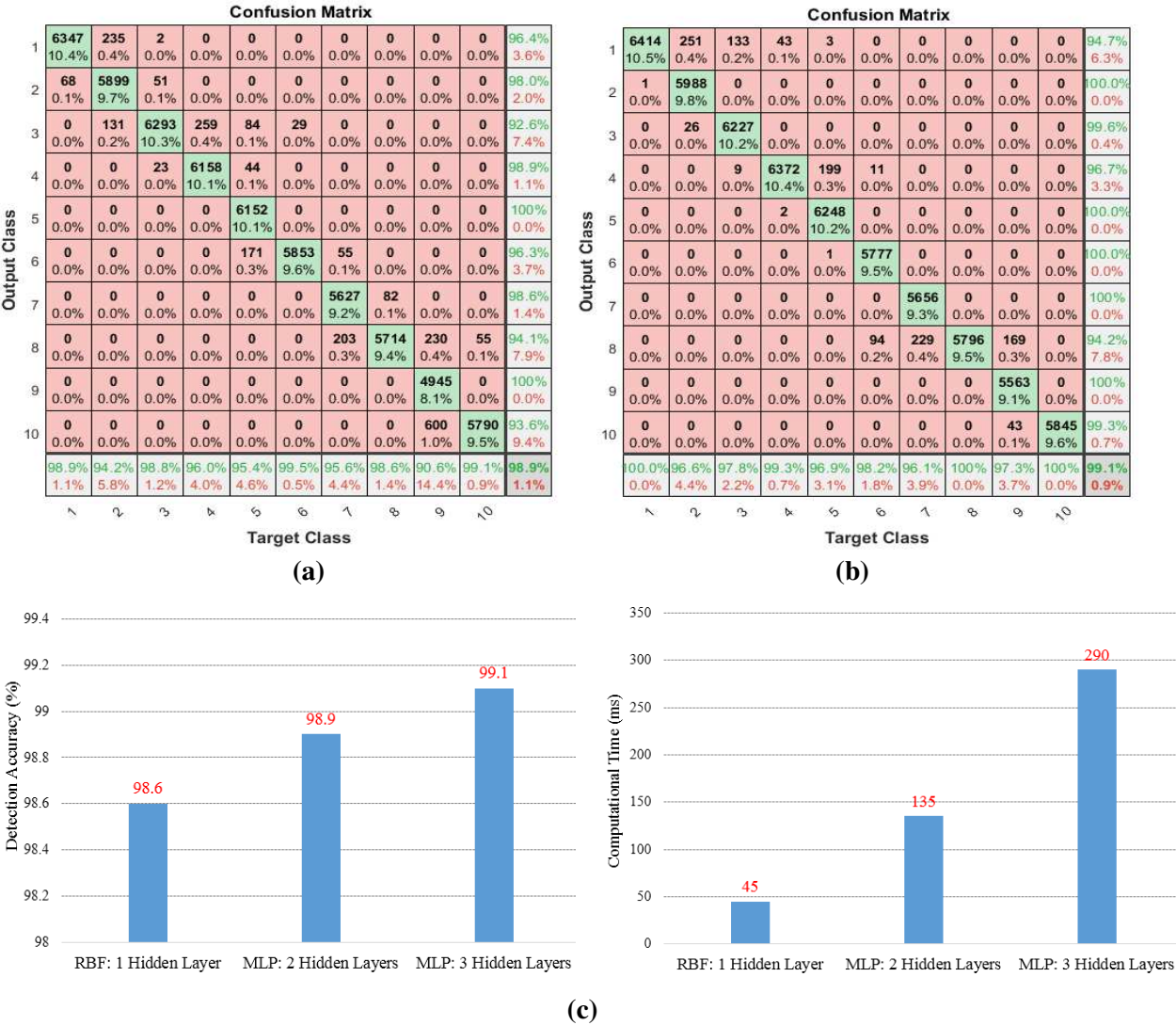

Figure 14. Output detecting accuracy of the MLP networks vs. RBF network. (a) 2 hidden layers, (b) 3 hidden layers, (c) Detection accuracy and the minimum computational time difference

3.9 t-test statistical analysis

In practice, comparing the competence of ANN models with conventional statistical techniques is critical, as uncomplicated statistical models could result the same detection accuracy compared with complicated ANN models. Therefore, we have used the well-established t-test technique that uses the mean value of two tests (in our case normal operation vs number of faulty PV module) and resulting a t-test value, this test has been verified using Minitab software.

With a confidence of 99%, in theory, if the t-test value is greater than 2.58, therefore, there is a significant difference in the two tested value, therefore a fault is ascertained in the PV system.

19

433 During no shading conditions, two tests were performed, including one faulty and nine faulty PV
434 modules in the PV string, results are shown in Figures 15(a) and (b). It is clearly notable that the
435 t-test is beyond the limit of 2.58, therefore a fault is detected in the PV string.

436 Whistle testing the same faulty conditions under partial shading condition, the result of the t-test
437 does not show a significant difference, as the t-test value is still within the theoretical threshold of
438 2.58 as can ben seen in Figures 15(c) and (d). This result confirms that even sophisticated statistical
439 technique such as t-test can only be used in some cases to detect the faults in the PV string, while,
440 for example, under shading conditions, it fails to determine the difference between the normal
441 operation and one faulty or nine faulty PV modules.

442 In conclusion, ANN model is therefore having a benefit over statistical-based techniques, as ANN
443 models can determine the difference of the power and irradiance levels and has the possibility to
444 distinguish the variations of these parameters including shading and non-shading conditions.
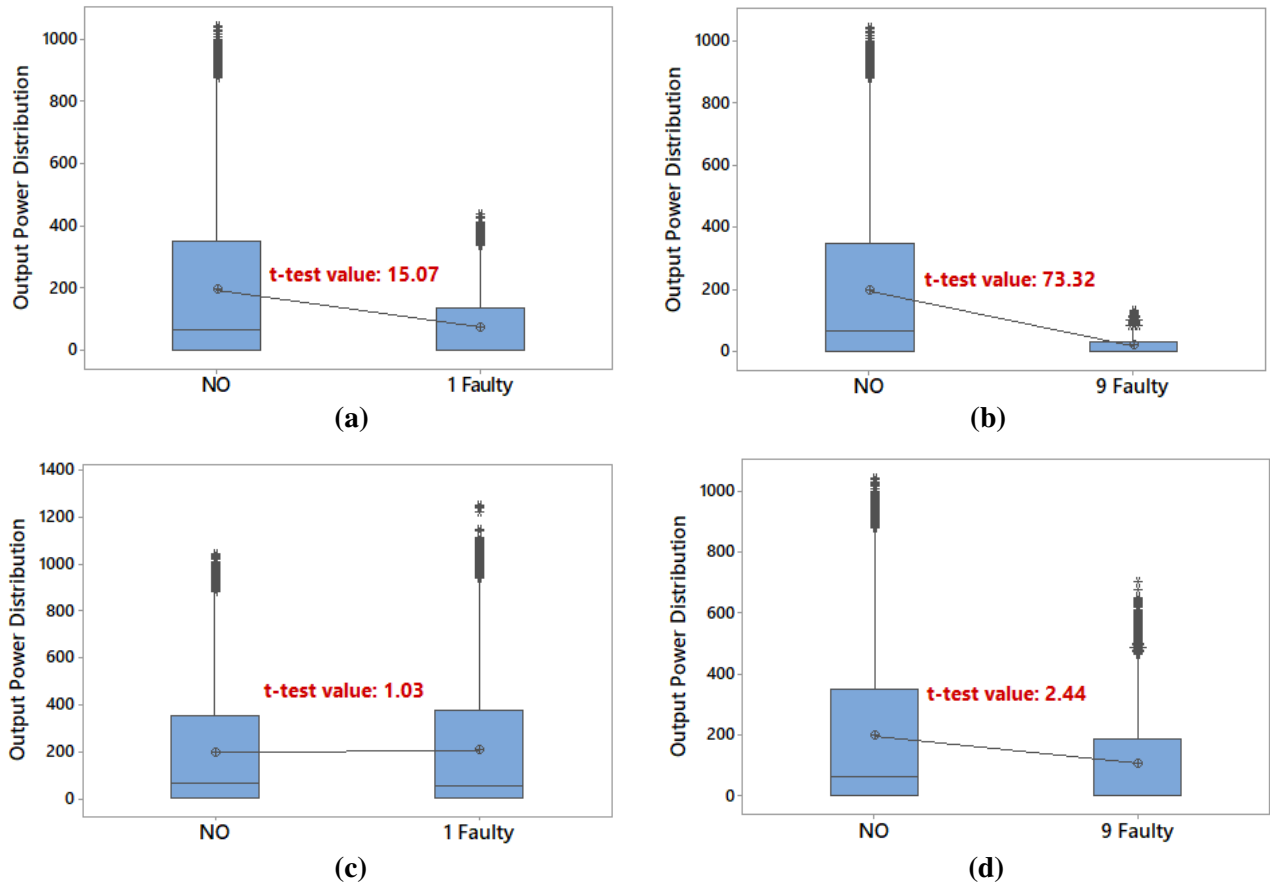
Figure 15. Output t-test value. (a) Normal operation vs one faulty PV module, without partial shading "sunny day", (b) Normal operation vs nine faulty PV modules in the PV string, without partial shading, (c) Normal operation vs one faulty PV module, data captured under partial shading scenario, (d) Normal operation vs nine faulty PV modules in the PV string, without partial shading

445   ## 4. Results

446   This section reports on the accuracy of the ANN through the implementation of the selected
447   methodology (M4) from section 3.6, as this methodology had achieved the highest fault detection
448   accuracy. Furthermore, the developed ANN using M4 is trained with a 'scaled-up' PV system and
449   reduced data set, refer to section 4.3. Note that the data presented in this section for testing the
450   actual accuracy of the ANN has not been previously used for training the ANN network, previously
451   discussed in section 3, reiterating the authenticity and integrity of our ANN.

452   ### 4.1. Partial shading results

453   The developed fault detection algorithm was subject to various experiments, in order to validate
454   its resilience, robustness and accuracy. The sample-set was collected over two weeks with the first
455   week testing under partial shading conditions and the latter under overcast conditions. Each
456   scenario persists for an entire day with a different fault applied to the examined PV system
457   illustrated previously in Figure 2.

458   The first week was based on validating the accuracy of the network based on the data collected on
459   partial shading. The solar irradiance for week-one is represented in Figure 16.
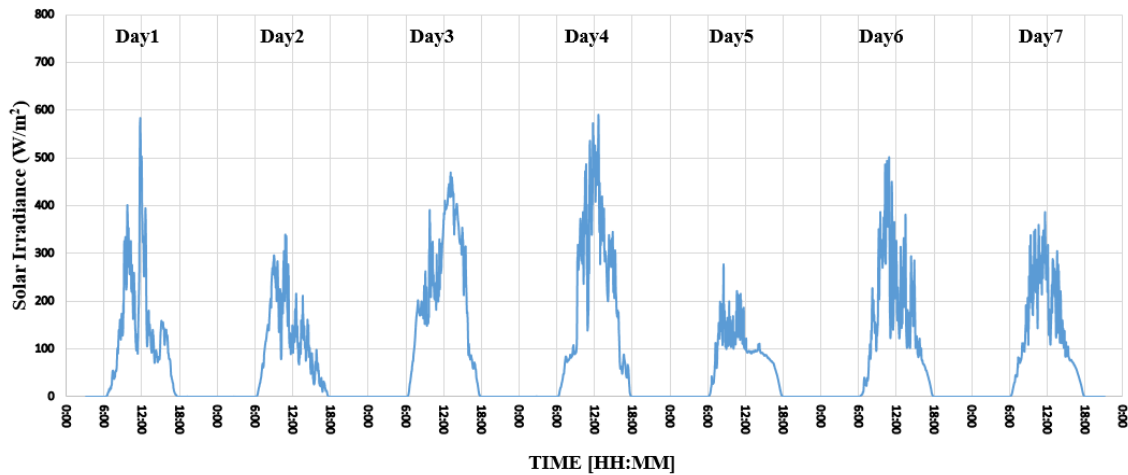
Figure 16. Solar irradiance of PV system for week one under partial shading conditions

460   The total output power of the system in question under different test conditions is represented in
461   Figure 17. The system is operating without any applied faults on the first day; hence, the total
462   output power is at its peak. During the duration of the week, as various faults are applied on a day
463   to day basis, the total output power decays. With the system generating its lowest output power on
464   day 6 with 7 faults applied. Note day-five has a low output power even though the system is
465   running without any applied faults, and this is due to the corresponding solar irradiance being at
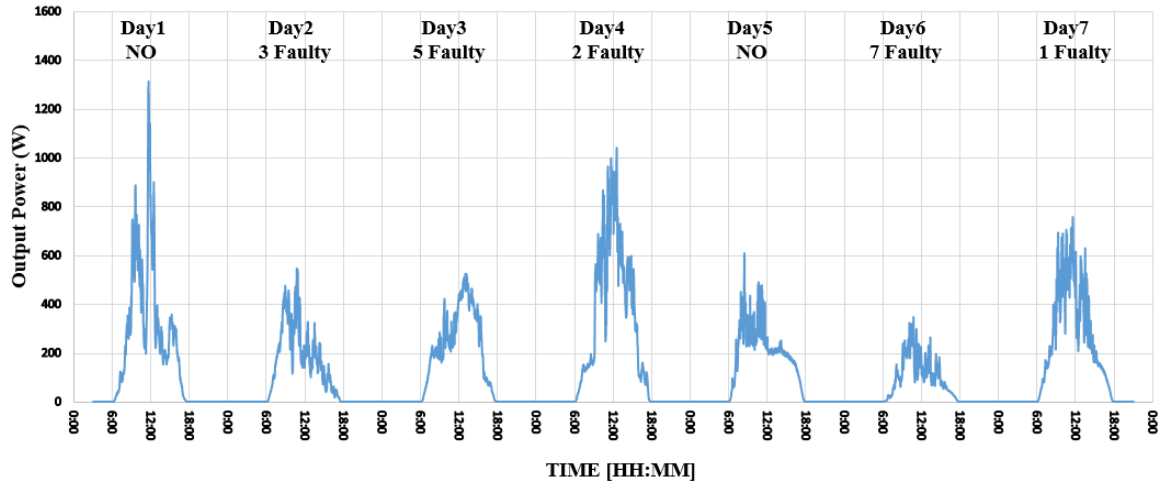466   its lowest for that day, represented in Figure 16.

21

Figure 17. Total output power of PV system for week one under partial shading conditions

467 The authentication of the accuracy of the developed ANN under partial shading conditions over
468 the week is represented by the output classification matrices, shown in Figure 18. The green and
469 red cells of the matrix represent the number of correct and incorrect classifications by the ANN,
470 respectively. The grey cells represent the total detection accuracy with respect to each row and
471 column.

472 The overall accuracy of the system under partial shading conditions was 97.9%, as shown in Figure
473 18. Out of 1413 samples, there are 1373 samples for NO (normal operation) correctly classified,
474 whereas 40 samples are misclassified as F2, F3 or F4, this corresponds to 2.8% faulty
475 classifications during NO and shading conditions. On the other hand, 666 samples are correctly
476 classified as F2, while some samples are misclassified as either F1 or F3, this is due to the change
477 of the solar irradiance affecting the PV system during the fourth examined day.
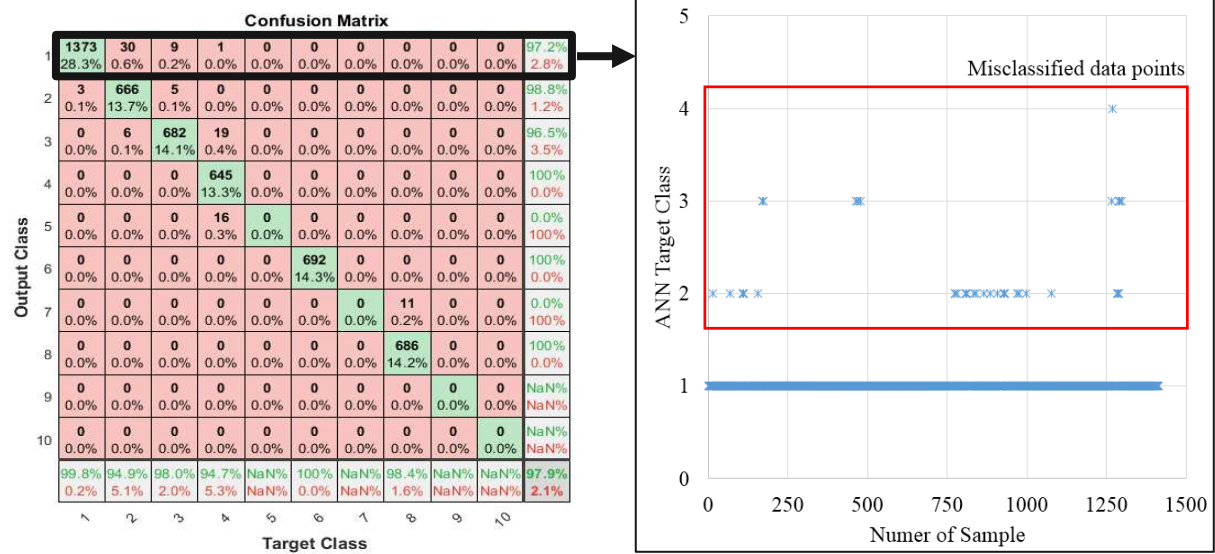
Figure 18. Confusion Matrix for ANN under partial shading conditions

478 **4.2. Overcast Results**

479 The second week was based on validating the accuracy of the network with regards to the data
480 collected for overcast conditions (partially cloudy and overcast). The solar irradiance for week one
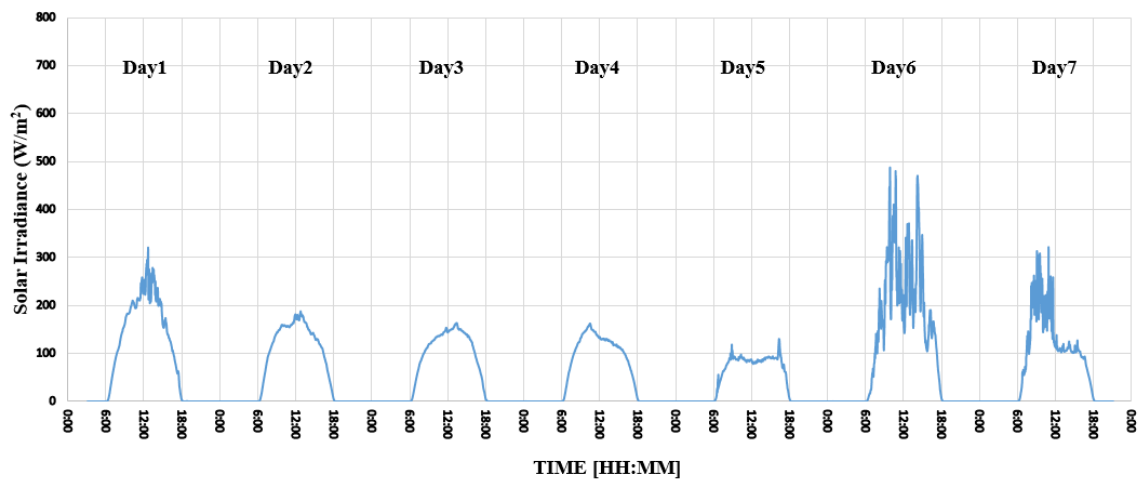481 is shown in Figure 19.

Figure 19. Solar irradiance of PV system for week two under overcast conditions

482 The total output power of the system in question under different test conditions is represented in
483 Figure 20. The system is operating without any applied faults (normal operation) on the first day,
484 but the total output power is not at its peak, whereas for the sixth day 1-fault is applied but it has
485 the highest total output power for the week, this is due to the solar irradiance being significantly
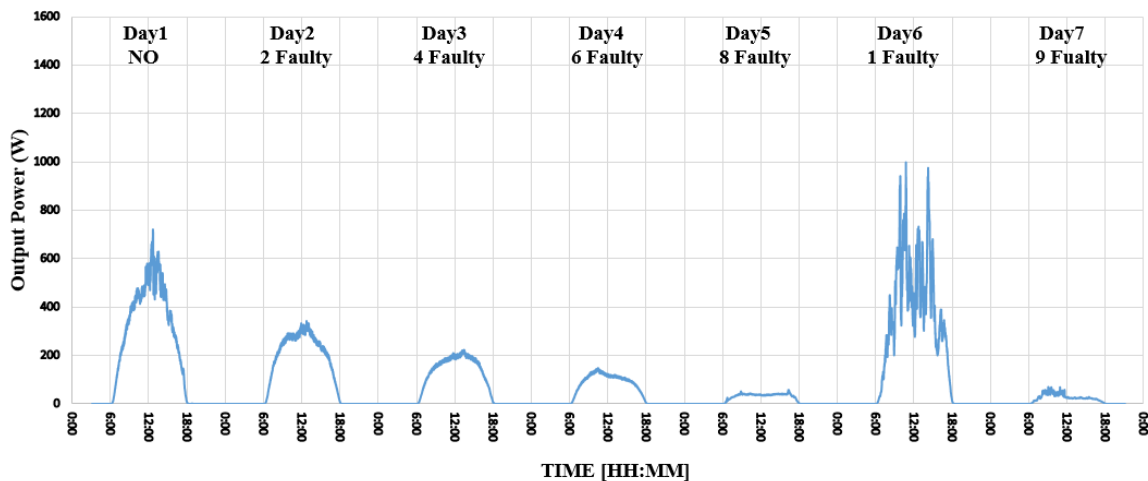486 higher on day-six as compared to any other day of the week, including day-one.

Figure 20. Total output power of PV system for week two under overcast conditions

As shown in Figure 21, the overall accuracy of the system for overcast conditions was 96.5%. Total of 710 samples for normal operation "NO" are correctly classified. This corresponds to 14.2% of all tested samples. Similarly, 661 samples are correctly classified as F2, corresponding to 13.2% of all samples. In row 1, 34 samples of NO are incorrectly classified as F1 corresponding to 0.7%. This was because the output power during normal operation mode and 1-faulty PV module are similar, particularly during partial shading conditions.

Figure 21. Confusion Matrix for ANN under overcast conditions

The ANN network performed to a higher accuracy in partial shading conditions as compared to overcast conditions. However, the overall accuracy of both systems was over 95% with a difference of 1.4% in terms of accuracy, between the two networks. The overcast conditions did not have a significant impact on the accuracy of the ANN, demonstrating the robustness of the system. It is important to note that although the sample data from both conditions was the same, the actual content of the data varied (the solar irradiance was much lower for overcast conditions as compared to partial shading). The developed ANN was able to handle this variance in the input data, through its training and validation, testifying its effectiveness through the overall detection accuracy for both conditions, as shown in the respective confusion matrices, Figures 18 and 21.

4.3. ANN results with different PV system

517   To validate the detection accuracy of the developed ANN, the network was tested with a different
518   PV plant (refer to Figure 22). The plant contained four PV strings, each consisting of 8 PV panels
519   with all strings connected in series. The switches on the inverter allowed the switching on-off of a
520   whole string. However, only one string was to be switched off at any given time representing 8
521   faults. If two strings were taken off-line, the occurred faults would equal 16, this was not within
522   the scope of this research. The total capacity of the examined PV system is equal to 4.16 kW.
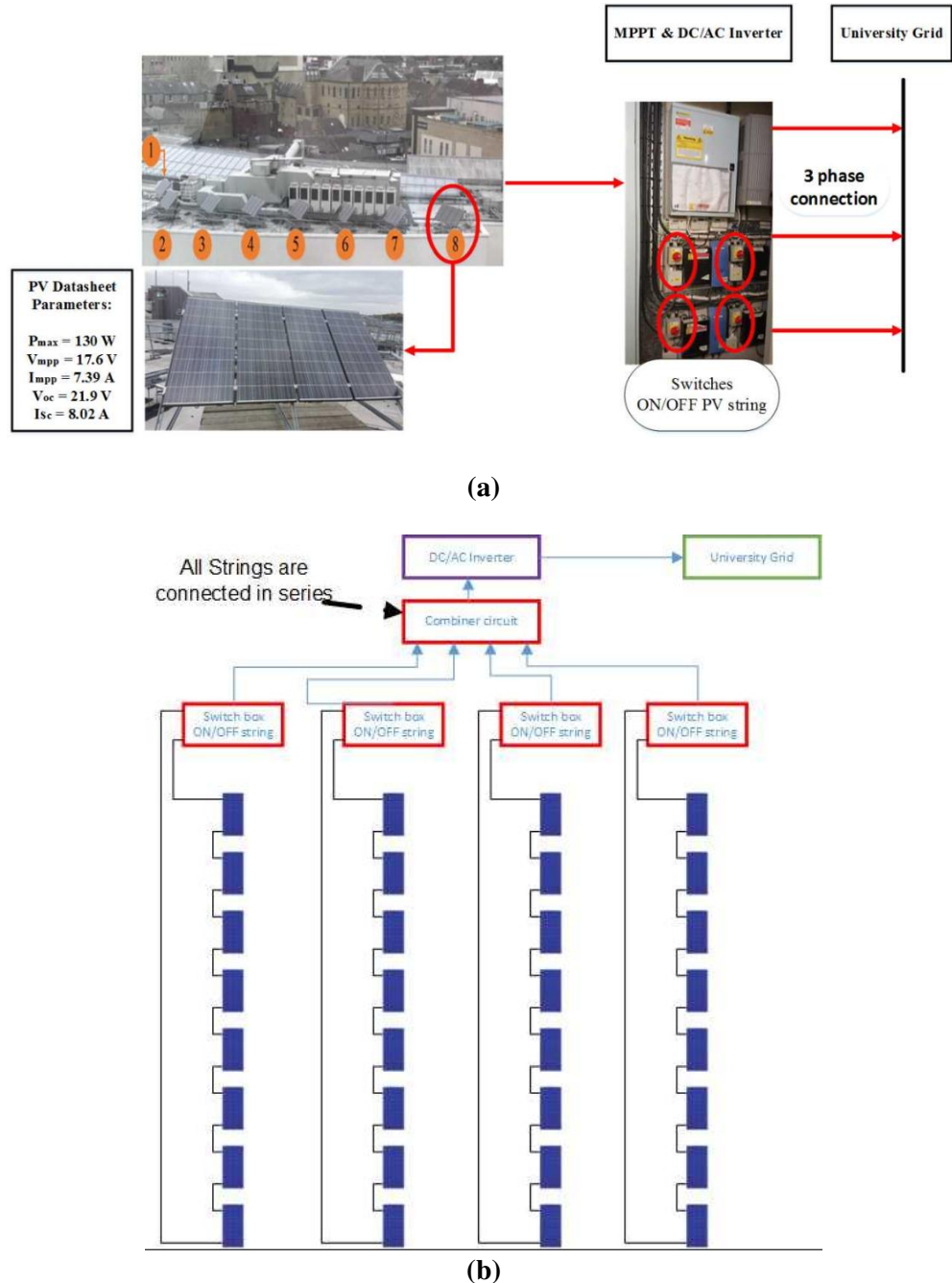

(a)

(b)

Figure 22. **(a)** Actual setup of PV system, **(b)** PV system schematic for test case study

523 The test procedure involved inducing 8-faults by switching off a complete string, followed by
524 normal operation for the next day, refer to Table 3. The solar irradiance and total power for the
525 duration of the testing period is represented in Figure 23; comprising of overcast and partial
526 shading conditions.

Table 3. Methodology for test case

No. Days	Condition Applied
1	NO
2	1st string is disconnected
3	NO
4	2nd string is disconnected
5	NO
6	3rd string is disconnected
7	NO
8	4th string is disconnected
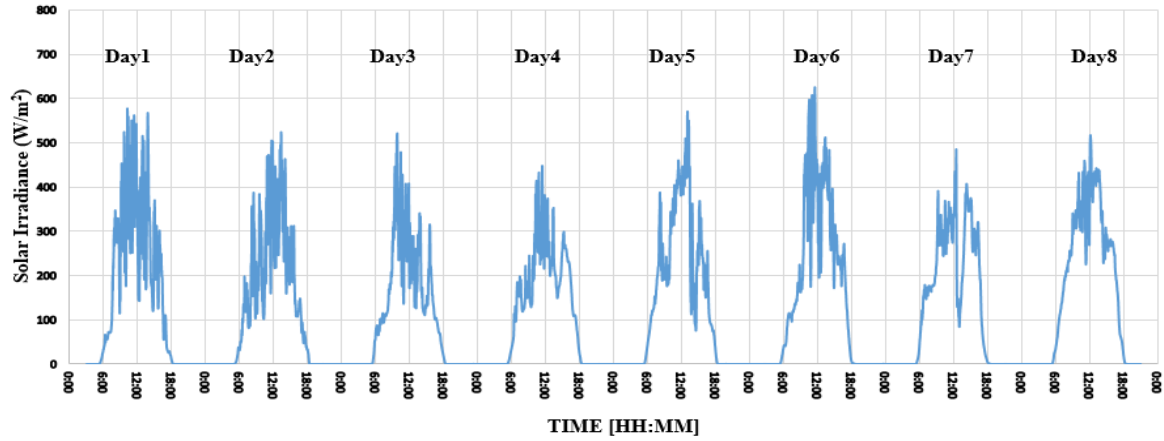
(a)
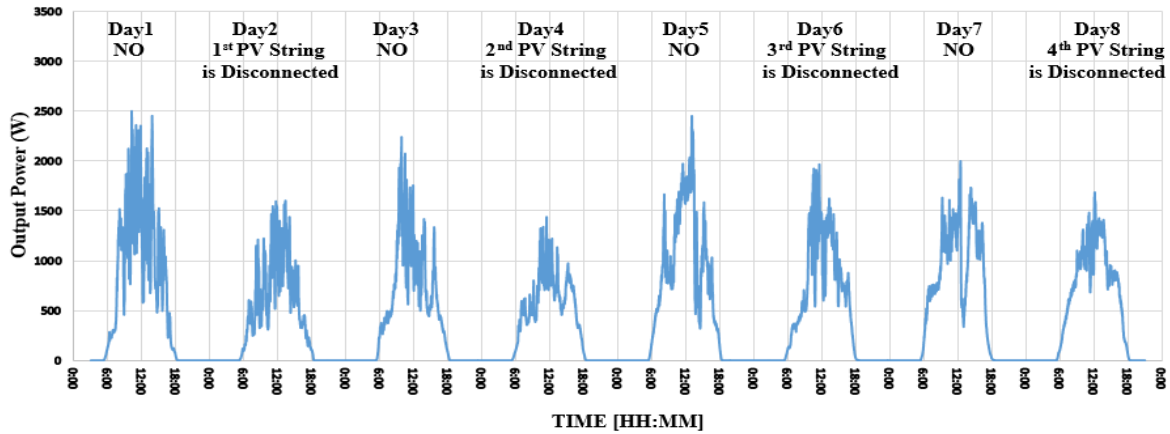
(b)

Figure 23. **(a)** Solar irradiance, **(b)** Output power for test case

527 Figure 24 illustrates the overall detection accuracy of the developed ANN for the test case PV
528 system shown in Figure 22. Before extracting any conclusions on the accuracy of the ANN, based
529 on the comparison of the two PV systems, key factors need to be considered. The first important
530 factor is the significant variation in the total power capacity of the two systems (2.2 kW and 4.16
531 kW). Achieving an accuracy of over 96% for both PV systems, the ANN has shown it is highly
532 adaptable to various capacity of PV systems. Another critical factor was the substantial difference
533 of the sample data for training the ANN. The first PV system provided the ANN with 'rich' training
534 data consisting of every hour of the day, for a duration of 10-weeks. Whereas, the latter system
535 reported in this sub-section decreased the amount of training data to 8-days. Nevertheless, the
536 overall detection accuracy of the ANN for both PV systems boasted results of over 96%. This
537 again testifies to the reliability, adaptability and successful implementation of our proposed ANN.

538 In order to validate the latter ANN network, an evaluation of the total detection accuracy was
539 carried out every two days interval. The reason for the selection of 'two-day' intervals was due to
540 the first day consisting of normal operation "NO" followed by 8-faulty "F8" (induced faults by
541 switching of a whole string, refer to, Figure 22(b)) for the second day. The ANN network original
542 MATLAB code is shown in Appendix A.

543 Hence, the results of the ANN network during the first two days including NO "as in first day"
544 and 8-faulty PV modules "as in the second day" is shown in Figure 24(a). The attained detection
545 accuracy is equal to 97.4%. The main factor that the detection accuracy slightly dropped during 8-
546 faulty "F8" mode, was due to 26 samples being incorrectly classified as either 9-faulty or 10-faulty
547 PV modules, resulting in a 2.1% decrease in accuracy.

548 The following 2-days (days 3-4), shown in Figure 24(b), experienced an increase in the overall
549 detection accuracy of the ANN at 98.0%. The improvement in the overall accuracy was due to the
550 ANN misclassifying 16 samples as 9-faulty or 10-faulty, when the actual fault induced was 8-
551 faulty "F8" relating to an error of 1.3%.

552 In contrast to the above pattern of gradual increase in the overall detection accuracy as the week
553 progressed, day 5 and 6, shown in Figure 24(c) showed signs of deterioration in the over accuracy
554 of the ANN (96.6%). A total of 43 samples were incorrectly classified as 9-faulty or 10-faulty
555 instead of 8-faulty.

556 The last two days (days 7 and 8), shown in Figure 24(d) indicated towards an improvement in the
557 overall detection accuracy of the network, with a total accuracy of an impressive 98.1%. Similarly,
558 to the first two days (days 3 and 4), 15 samples were incorrectly classified as either 9-faulty or 10-
559 faulty PV modules, whereas the induced fault consisted of 8-faulty, this led to the reduction in the
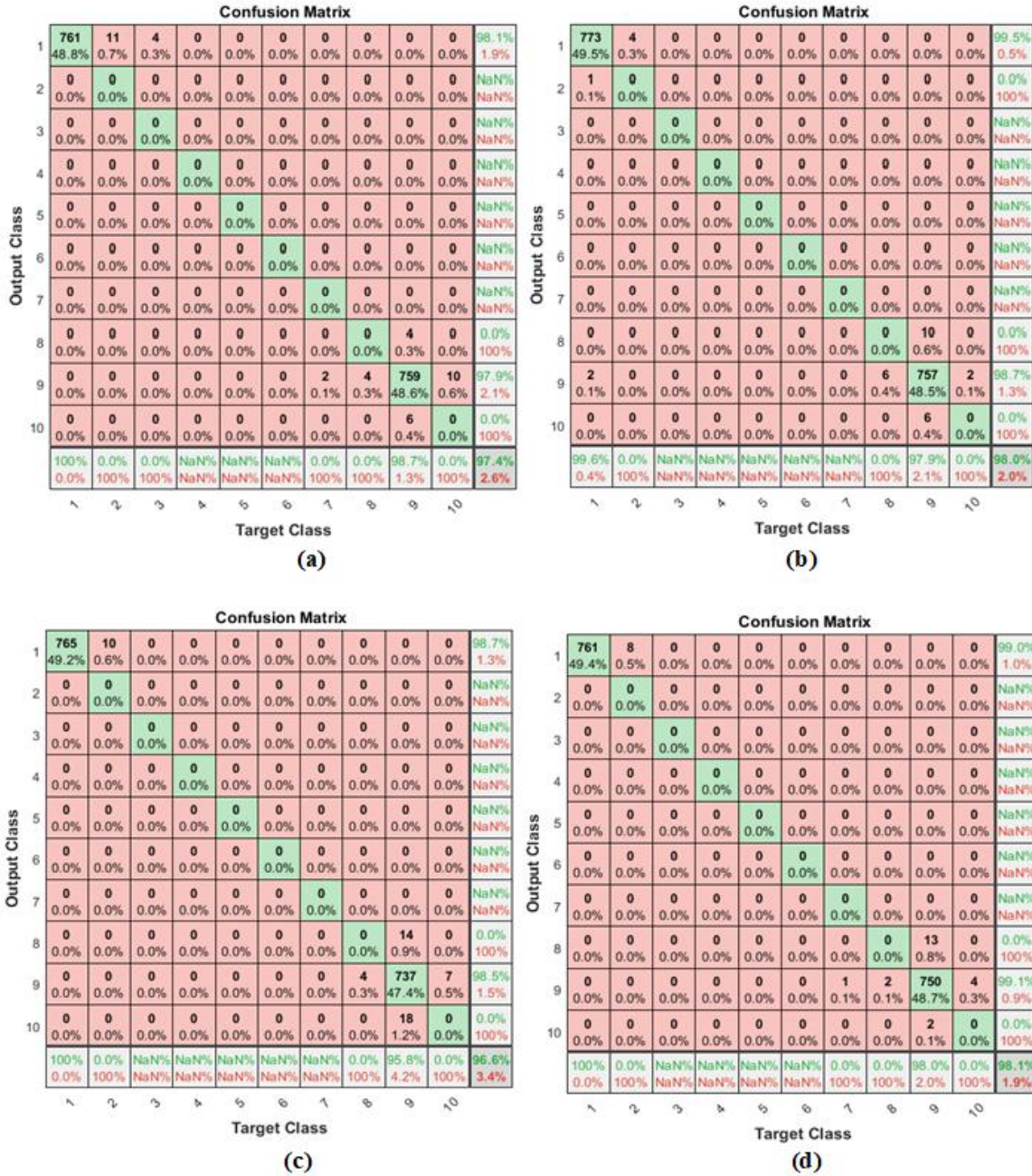560 overall accuracy by about 0.9%.

Figure 24. Classification Confusion Matrices for the ANN based on the test case PV system shown previously in Figure 22. **(a)** Accuracy is 97.4% for the first 2 days, **(b)** 98.0% for days 3 and 4, **(c)** 96.6% for days 5 and 6, **(d)** 98.1% for the last 2 days (day 7 and 8)

5. Comparative Study

561

562   This section evaluates and compares the new research outcomes [24, and 26-28] in the field of
563   ANN PV fault detection systems against our proposed methodology. A comprehensive evaluation
564   of the methodologies is summarized in Table 4.

565   As illustrated, all recent ANN-based PV fault detection systems require a large number of input
566   parameters for the ANN network to operate; hence, the algorithm becomes more complex in terms
567   of the practicality as well as the required historical data for the ANN training/validation process.

568 However, in this article, the proposed ANN network only requires two input parameters in order
569 to activate, namely the irradiance and the output power, while there is no need of any other PV
570 parameters such as the V_{mpp} and I_{mpp} as appear in all other recent algorithms [24, and 26-28].

571 The other advantage of our proposed algorithm is that the ambient temperature has not been
572 included in the ANN architecture. Hence, temperature sensors are no longer required in the PV
573 installation, as this is required by the algorithms proposed by [24] and [28].

574 According to our method, the ANN detection accuracy is ranging from 96.5%~98.1% in normal
575 operation (NO) and partial shading conditions, respectively. So far, the obtained PV fault detection
576 accuracy is considered the highest, as illustrated in Table 4.

577 Based on the examined PV installations shown earlier in Figures 2(a) and 22(a), the data
578 acquisition of the produced power occur at the DC side after the MPPT unit is placed. This place
579 of sampling data can generally not be taken for granted to be expandable to other PV systems since
580 not many commercial inverters offer this signal to be extracted. Taking this restraint into
581 consideration, this becomes the main limitation of the proposed model developed in this article.

Table 4. Comparitive study of recent ANN-based PV fault detection algorithms [24, and 26-28] and our proposed method

Ref.	Year	Number of examined PV modules "connected in series"	ANN Methodology – required input parameters								ANN detection accuracy (%)	
			G	T_{amb}	P	I_{sc}	V_{oc}	I_{mpp}	V_{mpp}	No. of required parameters	Normal Operation	Partial Shading
[24]	2016	Four 25% each setp	✓	✓	✓	✗	✗	✓	✓	5	90.3~97	max:90.3
[26]	2017	Up to Five 20% each step	✓	✗	✗	✓	✓	✓	✓	5	96.5	n/a
[27]	2018	Five 20% each step	✗	✗	✓	✓	✓	✓	✓	4	97.1~95.3	87.3~92.1
[28]	2019	Up to four 25% each step	✓	✓	✓	✗	✗	✓	✓	5	87.4~98.5	max: 66.45
Our proposed algorithm	2020	Ten 10% each step	✓	✗	✓	✗	✗	✗	✗	2	98.1%	96.5%

6. Conclusion

In this article, we have presented four different methodologies to detect PV faults based on two inputs parameters; solar irradiance and output power. It was found that the ANN accuracy increased up to 98.1% based on the implementation of the fourth methodology consisting of data normalisation as well as mapping of solar irradiance against output PV power. Hence, this methodology has been tested on two different PV installations, with significantly different electrical parameters.

Results show that the developed ANN network accurately detected PV faults in the range of 96.5-98.1% during normal operational mode, where no shading/overcast is present. Whereas, during partial shading conditions, the minimum ANN network accuracy deteriorated to was 96.2%. This outcome is based on the evaluation of our ANN on two differing PV installations, demonstrating the robustness and adaptability of the proposed network architecture.

In future, it is intended to enhance the accuracy of the developed ANN network through the evaluation of different methodologies comprising of different ANN input parameters as well as using deep learning methods to enhance the capabilities of the ANN neurons during training/validation.

7. References

1. Di Piazza, M. C., Viola, F., & Vitale, G. (2018). Evaluation of ground currents in a PV system with high frequency modeling. International Journal of Renewable Energy Research, 8(3), 1770-1778.
2. Dhimish, M., Holmes, V., Mehrdadi, B., Dales, M., Chong, B., & Zhang, L. (2017). Seven indicators variations for multiple PV array configurations under partial shading and faulty PV conditions. Renewable Energy, 113, 438-460.
3. Mansouri, M., Hajji, M., Trabelsi, M., Harkat, M. F., Al-khazraji, A., Livera, A., ... & Nounou, M. (2018). An effective statistical fault detection technique for grid connected photovoltaic systems based on an improved generalized likelihood ratio test. Energy, 159, 842-856.
4. Dhimish, M., & Holmes, V. (2016). Fault detection algorithm for grid-connected photovoltaic plants. Solar Energy, 137, 236-245.
5. Tadj, M., Benmouiza, K., Cheknane, A., & Silvestre, S. (2014). Improving the performance of PV systems by faults detection using GISTEL approach. Energy conversion and management, 80, 298-304.
6. Mellit, A., & Pavan, A. M. (2010). A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy. Solar Energy, 84(5), 807-821.
7. Takashima, T., Yamaguchi, J., Otani, K., Oozeki, T., Kato, K., & Ishida, M. (2009). Experimental studies of fault location in PV module strings. Solar Energy Materials and Solar Cells, 93(6-7), 1079-1082.
8. Dhimish, M., Mather, P., & Holmes, V. (2018). Evaluating power loss and performance ratio of hot-spotted photovoltaic modules. IEEE Transactions on Electron Devices, 65(12), 5419-5427.
9. L. Schirone, F.P. Califano, M. PastenaFault detection in a photovoltaic plant by time domain reflectometry Prog. Photovolt. Res. App, 2 (1994), pp. 35-44.

620   10. S. Silvestre, M.A. da Silva, A. Chouder, D. Guasch, E. Karatepe New procedure for fault detection in grid
621   connected PV systems based on the evaluation of current and voltage indicators Energy Convers. Manag., 86
622   (2014), pp. 241-249.

623   11. Li, X., Wen, H., Hu, Y., & Jiang, L. (2019). Drift-free current sensorless MPPT algorithm in photovoltaic
624   systems. Solar Energy, 177, 118-126.

625   12. W. Yuchuan, L. Qinli, S. YaqinApplication of BP neural network fault diagnosis in solar Photovoltaic System
626   Proceedings of the IEEE International Conference on Mechatronics and Automation, Changchun, China (2009),
627   pp. 9-12

628   13. W.T. Miller III, F.H. Glanz, L.G. Kraft IIICmas: an associative neural network alternative to backpropagation
629   Proc IEEE, 78 (1990), pp. 1561-1567

630   14. Ashok, V., Yadav, A., & Naik, V. K. (2019). Fault Detection and Classification of Multi-location and Evolving
631   Faults in Double-Circuit Transmission Line Using ANN. In Soft Computing in Data Analytics (pp. 307-317).
632   Springer, Singapore.

633   15. Pan, T., Chen, J., Zhou, Z., Wang, C., & He, S. (2019). A Novel Deep Learning Network via Multi-Scale Inner
634   Product with Locally Connected Feature Extraction for Intelligent Fault Detection. IEEE Transactions on
635   Industrial Informatics.

636   16. Belaout, A., Krim, F., Mellit, A., Talbi, B., & Arabi, A. (2018). Multiclass adaptive neuro-fuzzy classifier and
637   feature selection techniques for photovoltaic array fault detection and classification. Renewable energy, 127,
638   548-558.

639   17. Belaout, A., Krim, F., Mellit, A., Talbi, B., & Arabi, A. (2018). Multiclass adaptive neuro-fuzzy classifier and
640   feature selection techniques for photovoltaic array fault detection and classification. Renewable energy, 127,
641   548-558.

642   18. Malik, H., & Yadav, A. (2018). EMD and ANN based intelligent model for bearing fault diagnosis. Journal of
643   Intelligent & Fuzzy Systems, (Preprint), 1-12.

644   19. Javed, W., Chen, D., Farrag, M. E., & Xu, Y. (2019). System Configuration, Fault Detection, Location, Isolation
645   and Restoration: A Review on LVDC Microgrid Protections. Energies, 12(6), 1001.

646   20. A. Lapedes, R. FarberNonlinear signal processing using neural networks (1987)

647   21. Mellit, A., Sağlam, S., & Kalogirou, S. A. (2013). Artificial neural network-based model for estimating the
648   produced power of a photovoltaic module. Renewable Energy, 60, 71-78.

649   22. Polo, F. A. O., Bermejo, J. F., Fernández, J. F. G., & Márquez, A. C. (2015). Failure mode prediction and
650   energy forecasting of PV plants to assist dynamic maintenance tasks by ANN based models. Renewable Energy,
651   81, 227-238.

652   23. Yagi, Y., Kishi, H., Hagihara, R., Tanaka, T., Kozuma, S., Ishida, T., ... & Kiyama, S. (2003). Diagnostic
653   technology and an expert system for photovoltaic systems using the learning method. Solar energy materials
654   and solar cells, 75(3), 655-663.

655   24. Chine, W., Mellit, A., Lughi, V., Malek, A., Sulligoi, G., & Pavan, A. M. (2016). A novel fault diagnosis
656   technique for photovoltaic systems based on artificial neural networks. Renewable Energy, 90, 501-512.

657   25. Yuchuan Wu, Qinli Lan and Yaqin Sun, "Application of BP neural network fault diagnosis in solar photovoltaic
658   system," 2009 International Conference on Mechatronics and Automation, Changchun, 2009, pp. 2581-2585.

659   26. Karmacharya, I. M., & Gokaraju, R. (2017). Fault location in ungrounded photovoltaic system using wavelets
660   and ANN. IEEE Transactions on Power Delivery, 33(2), 549-559.

661   27. Dhimish, M., Holmes, V., Mehrdadi, B., & Dales, M. (2018). Comparing Mamdani Sugeno fuzzy logic and
662   RBF ANN network for PV fault detection. Renewable energy, 117, 257-274.

663   28. Fadhel, S., Delpha, C., Diallo, D., Bahri, I., Migan, A., Trabelsi, M., & Mimouni, M. F. (2019). PV shading
664   fault detection and classification based on IV curve using principal component analysis: Application to isolated
665   PV system. Solar Energy, 179, 1-10.

Appendix A

667 ANN Matlab Code:

668 `function [y1] = myNeuralNetworkFunction(x1)`

669 `% Input 1`
670 `x1_step1.xoffset = [0;0];`
671 `x1_step1.gain = [2;2];`
672 `x1_step1.ymin = -1;`

673 `% Layer 1`
674 `b1 = [-0.53379789162008717263;-5.6867963876778855337;10.177896012639834566;-2.5991370178094546084;-`
675 `0.9599606523864749852;2.9461760878212293058;4.3468884838846104657;-1.6827716763193365512;-`
676 `15.784457032110500663;18.014608760107652333];`

677 `IW1_1 = [10.739580122585271837 -11.159248213784534798;2.5403215144122435198 -`
678 `0.56388784665404756424;-15.530906692487940646 25.724533995852432611;2.6108726483375148675`
679 `2.8644466106623927004;2.3492627965250321154 -3.9135393841679908533;2.5621241841492614633`
680 `1.3108436436418966498;-18.156042599742011845 22.592529645218835554;-2.2026222558534906959`
681 `4.5796412164194322258;10.73037657323408034 -26.46367903771555774;-4.6700671639881869979`
682 `22.673036080767015932];`

683 `% Layer 2`

684 `b2 = [-0.6079524630706741487;0.52920177750850172504;0.87317087958293260197;-`
685 `0.77428730832701464504;0.80856340475453847283;-0.98776905950334870088;-`
686 `0.038084773428153848029;0.17631768142364898089;0.1177441519341838605;-1.2286308806154839779];`
687
688 `LW2_1 = [-16.218127732052877832 -0.61842039136869231264 13.65432479360931417`
689 `0.0279526869588864052 -1.5457829250363672724 -4.5249730228320954595 29.031811827208226617`
690 `1.3545922631578695139 -7.4451922992102366194 2.1032270854732173504;-1.8610677877454995244 -`
691 `0.13584261850222981161 14.599061786278145547 0.98142879231714974519 0.023068305715024127467`
692 `1.2729092739892586827 17.352259289313366253 -1.7969598569705769187 -7.5890283129856896949`
693 `5.3777628244524660062;0.017690508580209889197 -1.7872590063111608583 14.453052562066568854 -`
694 `0.28584189135600646114 -2.2943096899107273678 4.6075393554886687753 2.5732414558365808155 -`
695 `1.3164653264959451651 -9.2113122319186704345 3.9165651867640507433;0.78870635459690852098`
696 `0.45655538810872159372 15.726589320682379025 1.2188914303559659214 1.5572518953867149349 -`
697 `0.74512350082267531093 -12.537359667453197076 0.55369386754372895698 -8.5220919780286497058`
698 `3.4338668969512400331;2.0317783279326593338 -1.1142040895525726629 0.31670104901807094588`
699 `0.9784865679835351715 0.76022477810981570201 3.7280444386583488914 -9.8244365967376374726`
700 `0.52810244259625593877 -11.604561582830564603 3.3371779646911083894;1.7526128243198901835`
701 `0.18517512931874907656 -15.506889259234823086 0.96516150852280624406 0.37972493906172594125 -`
702 `1.1950923243193563028 -5.2382773533629025664 0.77967472280614491531 -11.860434415464807145`
703 `5.1071998684124046974;2.1381733507637172842 0.056596023834588657375 -13.298494169561720923`
704 `0.11021173484657678654 1.2536020059998913556 1.7661104686248074724 -5.3361035540167647985`
705 `0.23036465741826600562 3.3863911471888354932 8.9226575295582417624;3.6723972390395536181`
706 `0.87284720834851448057 -8.4578343073060278101 -0.044134247790334736605 0.37388197719519100648 -`
707 `1.0196077545498973826 -5.8551819327672580684 -0.11479275509819439338 17.324612120669645066`
708 `6.7927117492973421164;6.6441975543224023326 0.44978145240560007956 -8.0408855433529229373 -`
709 `0.81189573510752477414 0.7060671169563150773 1.493396637034419383 -5.7368690002926161497 -`
710 `0.8718663367356225689 16.908781753661273228 -9.5834534191820850424;3.0982763799864270204`
711 `0.49939749005411848692 -8.9893465264005989468 0.36043960291007642871 0.22646282549337756751 -`
712 `4.3925487668578648837 -5.6103623569920149095 0.51747719198805008922 18.039045221338205064 -`
713 `27.296588790470323715];`

```matlab
714    % ===== SIMULATION ========
715    Q = size(x1,1); % samples

716    x1 = x1'; % Input 1
717    xp1 = mapminmax_apply(x1,x1_step1);

718    a1 = tansig_apply(repmat(b1,1,Q) + IW1_1*xp1); % Layer 1

719    a2 = softmax_apply(repmat(b2,1,Q) + LW2_1*a1); % Layer 2

720    y1 = a2; % Output 1
721    y1 = y1';
722    end

723    function y = mapminmax_apply(x,settings) % Map Minimum and Maximum Input Processing Function
724    y = bsxfun(@minus,x,settings.xoffset);
725    y = bsxfun(@times,y,settings.gain);
726    y = bsxfun(@plus,y,settings.ymin);
727    end

728    function a = softmax_apply(n,~) % Competitive Soft Transfer Function
729    if isa(n,'gpuArray')
730       a = iSoftmaxApplyGPU(n);
731    else
732       a = iSoftmaxApplyCPU(n);
733    end
734    end

735    function a = iSoftmaxApplyCPU(n)
736    nmax = max(n,[],1);
737    n = bsxfun(@minus,n,nmax);
738    numerator = exp(n);
739    denominator = sum(numerator,1);
740    denominator(denominator == 0) = 1;
741    a = bsxfun(@rdivide,numerator,denominator);
742    end

743    function a = iSoftmaxApplyGPU(n)
744    nmax = max(n,[],1);
745    numerator = arrayfun(@iSoftmaxApplyGPUHelper1,n,nmax);
746    denominator = sum(numerator,1);
747    a = arrayfun(@iSoftmaxApplyGPUHelper2,numerator,denominator);
748    end
749    function numerator = iSoftmaxApplyGPUHelper1(n,nmax)
750    numerator = exp(n - nmax);
751    end
752    function a = iSoftmaxApplyGPUHelper2(numerator,denominator)
753    if (denominator == 0)
754       a = numerator;
755    else
756       a = numerator ./ denominator;
757    end
758    end
759    function a = tansig_apply(n,~) % Sigmoid Symmetric Transfer Function
760    a = 2 ./ (1 + exp(-2*n)) - 1;
761    end
```