Failure diagnosis in photovoltaic systems: a pattern recognition approach based on artificial neural networks

Francisco Costa Santos, Instituto Superior Técnico, Lisboa
francisco.costa.santos@tecnico.ulisboa.pt

***Abstract*— This thesis targets a method of diagnosing behaviours in photovoltaic (PV) systems using artificial intelligence technique. It uses a known model, the one diode and five parameters, and simulates the panel as set of solar cells, in order to represent mismatch faults in the panel. This simulation model is employed to create a database of characteristic curves that include the response under standard condition, but also two faulty conditions: short circuited cells and partial shading. From each curve, a few key informations are extracted: the voltage at maximum power point, the current at maximum power point, the panel temperature and the irradiance. These informations are used as inputs for training an artificial neural network (ANN) to classify the original PV panel condition. Applying normalization to the inputs increased the convergence and classification of the ANN. The trained ANN was tested with real outdoor measurements of a PV panel under all described conditions. The results showed an accuracy of 95% in the detection of a faulty condition and an accuracy of 73% in the diagnosis of the behaviour. Post conclusion several recommendations are made to improve and develop this method of diagnosis.**

***Index Terms*— Failure diagnosis, Photovoltaic systems, Mismatch faults, artificial neural networks**

I. INTRODUCTION

Along the years there has been a shift of paradigms in ways of producing energy, relocating the non-renewable fonts of energy to renewables. Solar energy has seen increased investment in the last few years. It has some unique perks, installation in almost every place due to its flexibility in size, and it is flexible in power sizing, noise-free [1]. A decrease in photovoltaic panels and equipment costs, special remunerations to renewable energies, more efficient panels all brought investment to this technology. Some disadvantages persist being availability one of the main problems. With all the developments made in new photovoltaic panels, the efficiency of converting energy has increased. However, these were not tested in real conditions, only in ideal ones and not for an extended period of such as their lifespan (25 years). These require a continuous operation to keep availability and compete with other forms of producing energy. When a fault occurs in a system, less energy is produced. Some faults may even damage the panels permanently, and all these flaws reduce the income. Also, there may be abnormal behaviour which is only temporary; however, if not detected, can lead to underperformance and in the future can develop into a fault. To maintain the performance, to the maximum, monitoring of the system is required [2].

Every system now and in the last few years has a digital inverter making it simple to collect the electrical parameters. Panels change their behaviour with irradiance and temperature as such these will also enter in the fault detection method to be developed. Both are critical in determining variations in power delivered by the panels. Collecting these weather parameters may resort to weather data servers, depending on the country, or to sensors installed near the system. Collecting electrical parameters and weather data over long periods requires storage and analysis of the data. To determine if the system is working as it should a data processing method will compare typical values of power to the ones observed. Even though processes like these exist they prove not very effective. In the detection of fault, the methods are more developed and prove to be more efficient, however in diagnosis there are still many nuances and details that are not detected and prove hard to pinpoint.

In this thesis, a method of detection and diagnosis is proposed. It monitors the electrical parameters (voltage, current) and in case the output power is not in the expected level, it will determine the more probable fault. In order to develop this method, a model capable of simulating standard and faulty behaviours in a PV array is used. The values obtained are used for training of an artificial neural network (ANN) to identify and classify the faulty behaviours in a photovoltaic system.

The method proposed should be able to determine if the system, knowing its topology and specifications, has faults, temporary abnormal behaviour or is working as intended. If the system is faulty, the approach will provide a probability of which fault may have occurred.

II. I-V CURVES

A usual way of analyzing the behaviour of a photovoltaic system is to trace its I-V curve. From this curve, three significant points are retrieved and identified, the short circuit current (I_{sc}), the open circuit voltage (V_{oc}) and the maximum power point (V_{mpp}, I_{mpp}). Figure 1 gives an example of an I-V curve with these points. There are two factors that influence the I-V curve, the temperature and the irradiance, meaning for each irradiance and temperature there will be a different I-V curve.

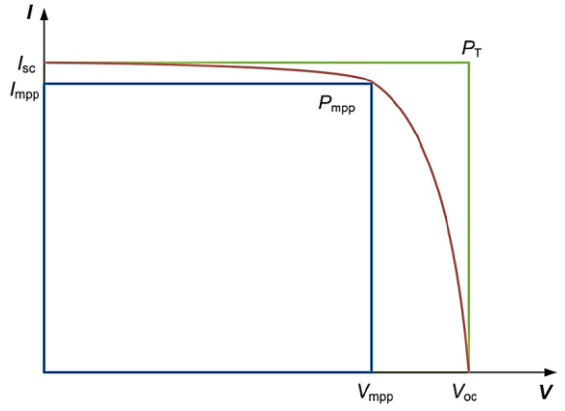

Figure 1 - I-V curve with the most important parameters: I_{sc}, I_{mpp}, V_{oc}, V_{mpp}, P_T, P_{mpp} [3]

III. ARTIFICIAL NEURAL NETWORKS AND PV APPLICATION

An artificial neural network can be described as a massively parallel combination of simple processing units which can acquire knowledge from environment through a learning process and store it in its connections [4]. From inputs and outputs gathered it can create a mathematical model which can classify new inputs into known outputs. Papers have already started to introduce artificial neural networks to fault prediction methods in photovoltaic systems. In [5] the authors explore the utilization of this tool to build a detection network. Evaluation of the ANN's performance is done using the mean square error (MSE). Their system is composed by a photovoltaic panel connected to a DC/DC boost converter with P&O algorithm for maximum power point tracker (MPPT) control. In this study five faults were tested, demonstrated in Table 1. It was subdivided in normal operation, one inverse module where the current passes through the bypass diode, two inversed modules, partial shadow in two or three modules and intense shadow effect. To construct the artificial neural network, they used a Matlab based model and configured the inputs as voltage and power of the PV system, the targets were set as the codes shown in Table 1. Data set for this problem was split into two subsets, 70% was used to train the gradient and to readjust the bias and weights. The other 30% were samples to validate the model. MSE allowed for a more accurate test of the model as it utilized as an error the difference between the targets and the outputs obtained in the training process. A normal measure to apply to these types of networks.

Table 1 - Classification of PV array faults

Category	Fault type	Symbol	Code
1	Normal operation	H	[1;0; 0]
2	Inversed module: one module	F1	[0;1; 0]
3	Inversed modules: two modules	F2	[1;1; 0]
4	Shading 1: two modules shadowed	F3	[1;0; 1]
5	Shading 2: three modules shadowed	F4	[0;1; 1]
6	Shading 3: intense shading effect	F5	[1;1; 1]

Another type of fault detection using ANN is proposed in [6]. First simulations were performed based on the system to identify normal operation and define a threshold limit for it. After simulations and already an established set of faults, starts the first part of the schematic designed by the authors. Values are compared with simulations to detect if the threshold limits, previously stated, are in check or not. Then it leads, based on the previous identification, to an attribute's selection. This was defined in the simulations, where from the study of the characteristics of the system a relation could be made with the electric parameters. The attributes can relate to the number of faults and their type of flaw. Moving further, knowing the attributes, the authors differentiated faults with two algorithms, number one isolating the faults when they have different combinations. Here is where the simulated and measured are calculated, and their relative difference is compared with threshold values. All of them are determined by the measurement noise and the model's uncertainty. Then the second algorithm distinguishes faults that have the same attributes. It uses an ANN to choose which fault is affecting the system.

In this thesis it is sought to simulate chosen faults, especially mismatch, and create a database from that. Then applying ANNs, such as previous works stated, and determine if the process is reliable and can be implemented in PV systems.

IV. SIMULATION MODEL

The model developed by Eduardo Sarquis [7], was chosen to create a database with standard behaviour and abnormal behaviours. This model uses the same methodology of the one diode and five parameter model to represent the equivalent solar cell circuit. However instead of having a single equivalent circuit it represents the panel as strings of cells and each of those cells has its own equivalent circuit.

It allows for a better simulation of mismatch faults. The basic model for the equivalent cell circuit is the California energy commission (CEC) model, a variation of the one diode and five parameters model [8]. From equation (1) the parameters can be obtained.
Where I_{ph} is the photodiode current, I_s is the saturation current, n is the ideality factor, V_t is the thermal voltage.

$$I = I_{ph} - I_s\left(e^{\frac{V+I\cdot R_s}{n\cdot V_t}} - 1\right) - \frac{V + I\cdot R_s}{R_{sh}} \qquad (1)$$

However, CEC uses another parameter, *adjust*, to adapt the temperature coefficient of the short circuit (μ_{Isc}) and open circuit voltage (β_{Voc}). Both equations are presented in (2,3).

$$\mu_{Isc} = \alpha_{sc,ref}\left(1 - \frac{adjust}{100}\right) \qquad (2)$$

$$\beta_{Voc} = \beta_{oc,ref}\left(1 + \frac{adjust}{100}\right) \qquad (3)$$

At the website [9] is available a database of photovoltaic panels parameters ($R_{s,ref}$, $R_{sh,ref}$, n, I_s, I_{ph}, *adjust*) at STC [10]. These parameters can be applied for each cell and their respective condition. Then the simulation model will run and determine the state of the system.

V. EXPERIMENTAL PROCEDURE

An experimental procedure for analysis of the I-V characteristic curve of healthy and damaged solar cells in PV panels was carried out, these are presented below in Figure 2. In Figure 3, Figure 4, Figure 5, respectively broken cell, broken glass and healthy panel cell, are represented the I-V curves obtained for each radiation (230, 400, 600, 800, 1000 W/m2), where each colour is bond to a radiation value. It is shown all the ten tests in each radiation, creating a cloud of results for all I-V curves. These tests were run to understand the characteristics of the I-V curves. It determined which variables were essential for the creation of the database. Also, from the results it was possible to comprehend the difficulty in evaluating the consequences of the broken cell and broken glass effect in the system electrical parameters. Going further, the plots referring to the experimental procedure will be using the healthy panel. The basic unit defined in the model was tested and produced similar results to the experimental, proceeding with an experiment on the whole panel. The same experiment as the one done for the cell was applied to the whole panel. However, it was done outdoors.

Figure 2 - The two panels tested in the laboratory: left side (broken glass and a broken cell), right side (healthy panel)

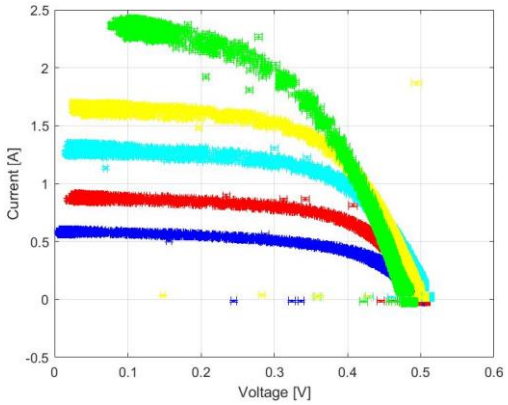

Figure 3 - Broken cell I-V curve for the following irradiances: 230 (blue), 400 (red), 600 (cyan), 800 (yellow), 1000 (green) [W/m^2]

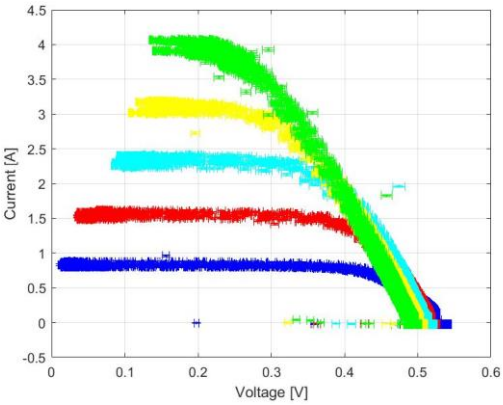

Figure 4 - Broken glass I-V curve for the following irradiances: 230 (blue), 400 (red), 600 (cyan), 800 (yellow), 1000 (green) [W/m^2]

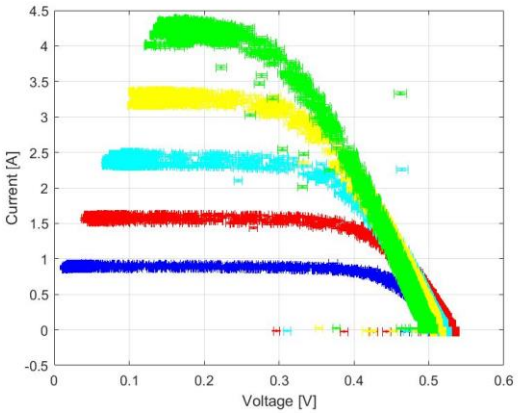

Figure 5 - Healthy panel I-V curve for the following irradiances: 230 (blue), 400 (red), 600 (cyan), 800 (yellow), 1000 (green) [W/m^2]

VI. DATABASE CONSTRUCTION

To create a database, an understanding of the faults produced by the model was necessary. For this, it was decided to plot specific abnormal behaviours and a standard one to compare with each other and comprehend them hence choosing the best way to form a database. A standard mode, a short circuit of one cell, a short circuit of a substring, a short circuit of two substrings and all three short-circuited substrings. These form a general perspective of a panel and carry general options that can happen. For shading, broader subtypes were simulated. It was due to having more factors that could influence this type of power loss. Such as the number of cells being shadowed, their location and the percentage of shading being applied. In Figure 6 can be seen a plot comparison between standard and short circuit behaviour.

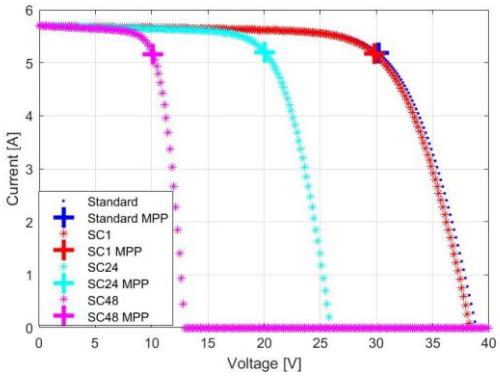

Figure 6 - Standard and Short circuit, both with an irradiance of 1000 W/m^2 and temperature of 70ºC

An introduction of the behaviours applied to the panel and their simulation characteristics are shown in Figure 7. On the left of the figure are visually represented the electrical equivalent circuits in those different conditions and on the right the necessary inputs to produce their respective simulations.

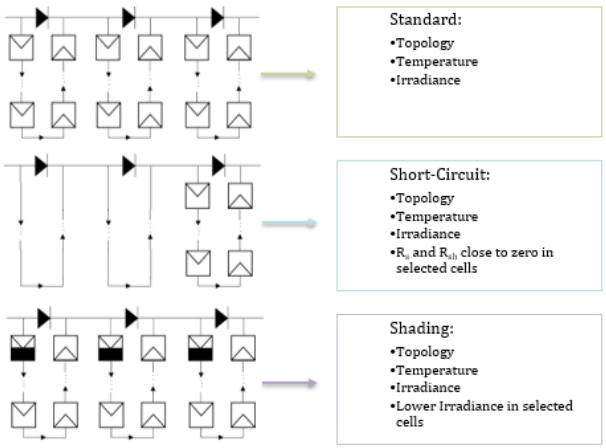

Figure 7 - Methods applied to simulate chosen faults and visual representation of said faults

Having a defined structure with all variables used stored was very important. It allowed to define and identify specific subtypes of faults; in this matter, it is presented a vector of the final structure containing all the parameters inputted and the outcome variables. As mentioned before the first values are the voltage and then the current from the I-V curve tracing, necessary to observe if anything appears to be out of order. Subsequently come the temperature, the irradiance, the voltage of maximum power, the current of the maximum power and the maximum power. Later the subtypes specific variables appear, the number of short circuits (Nsc), the number of shaded cells (Nsh) and the percentage of shade (Psh). The final three columns represent the output code, defining the behaviour. Over in Table 2, an example of a row of the matrix, for one panel, is shown, it is represented as a column since it provides a better portrayal.

Table 2 - Example of a row in a database, with the number of components below their respective names

V (300)	I (300)	T (1)	G (1)	V_{mpp} (1)	I_{mpp} (1)
P_{mpp} (1)	Nsc (1)	Nsh (1)	Psh (1)	OC (3)	

After concluding the random database had better results in artificial neural network training as it had less issues with overfit, however another issue arose from the same theme. The number of entries in the system could cause overfitting. So, to understand how many entries the database should have the maximum number of possibilities were thought out, considering only one panel. In Table 3, it is presented all the possible cases.

Table 3 - Total combinations possible using the random database

Standard	Short-Circuit	Shading
Variables: T, G	Variables: T, G, Number of Short circuits (Nsc)	Variables: T, G, Number of shaded cells (Nsh), percentage of shading (Psh)
T = [0,90]; G = [200,1000]	T = [0,90]; G = [200,1000]; Nsc = [1,72]	T = [0,90]; G = [200,1000]; Nsh = [1,3]; Psh = [20,80]
Total Standard = 91·800 = 72.800	Total SC = 91·800·72 = 5.241,600	Total SH = 91·800·3·60 = 13.104,000
Total = 72,800 + 5.241,600 + 13.104,000 = 18.418,400		

A 10000 entries database was constructed. It was decided to be this size due to computational time, and as can be seen, it is not a number that will cause overfitting for an ANN as it represents less than 1% of the total possible cases. To give an insight into the time spent on this process, 10000 entries took approximately 12 hours to complete, in a student's laptop. However, when measuring in a real system, there will be oscillations in values that cannot occur in a simulation. To create this effect in the database, a random matrix, with the same size of the dataset, was manufactured. It had values, in percentage, between -5 and +5 which were multiplied with the dataset, causing a variation in all variables, creating noise in the results. After finalizing both matrices, the simulated and the noise were put together creating a 20000 matrix database. A flowchart, of all process behind the creation of the random database, is presented in Figure 10.

VII. ARTIFICIAL NEURAL NETWORK DEVELOPMENT

To start an ANN the inputs must be chosen, and in this case, there are four essential variables: the cell temperature, the irradiance, the voltage at maximum power point and the current at maximum power point. These are normalized for a better convergence in the training to the networks' outputs. This process is made by dividing the input by, for the weather variables, their respective maximum values, for the voltage the open circuit value and for the current the short circuit value. For the hidden layer of the network, after testing exhaustively, the one with the best results was with one hidden layer and five nodes in that layer. Two datasets are tested, one without noise and the other one with added noise. This was done by increasing or decreasing five percent of the values on each entry. It was to create a measuring error threshold. The performance graphs are presented in Figure 8 and Figure 9. In Figure 8 there is a lower error, however the stages of training, validation and testing do not converge to the same value. This shows a problem of overfitting. In Figure 9 the error increases compared to the previous, this is due to the added noise in the database. However, all stages converge to the same point showing there is no issue with overfitting. The first performance graph shows worse results to new inputs.

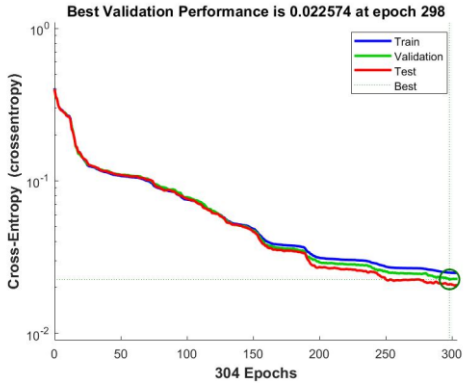

Figure 8 - Performance graph of the 10000 database with no noise added

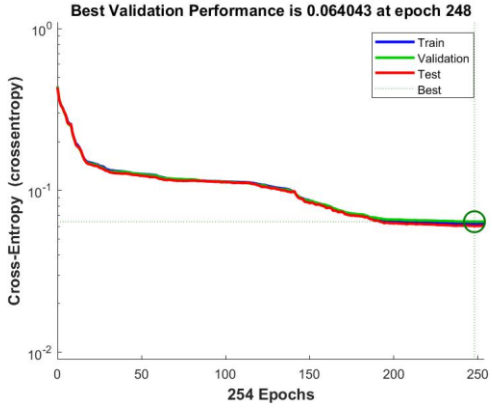

Figure 9 - Performance graph of the 20000 database with noise added

The second tool to access the correct construction of a well-defined network is the confusion matrix. In Figure 11 and Figure 12 can be seen the confusion matrices for the network without noise and for the one with added noise, respectively. It shows overfitting in the no noise database as the test stage has better accuracy than the training stage. As expected, the noise database produced worse accuracy, however still very high, guaranteeing reliable variations of experiments and measurements.

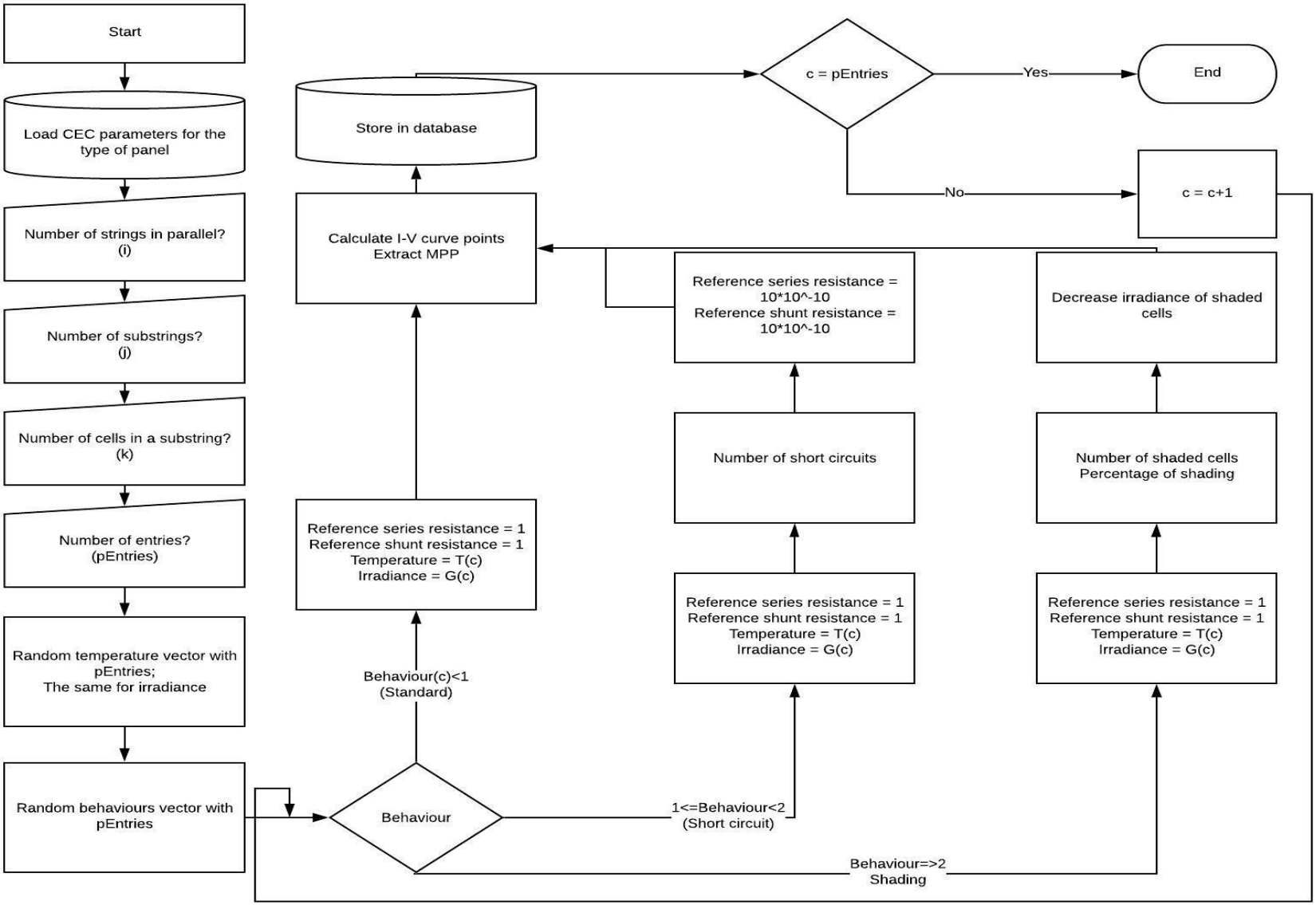

Figure 10 - Flowchart of database creation

Figure 11 - Confusion matrices of all steps in a learning process of an ANN, in this specific case 10000 normalized entries with no noise added

Figure 12 - Confusion matrices of all steps in a learning process of an ANN, in this specific case 20000 normalized entries with noise added

VIII. TESTING THE ARTIFICIAL NEURAL NETWORK

An outdoor experiment was tested with the known behaviours described in Table 4. For this test, a broad range of faults was tested. First every test was run five times, to guarantee precision and avoid errors. Secondly the temperature and irradiance values were stored. Thirdly the tests were divided in different categories and subcategories. Standard operation, without any purposely fault. Short circuit, which was divided into three: one cell, twenty four cells and forty eight cells. And finally shading which required to two levels of separation: first in the number of cells being shaded and then on the percentage of shading being applied. In Table 4 can be seen all the behaviours tested outdoor.

Table 4 - Behaviours tested outdoor

Behaviour:	Standard	Short Circuit	Shading
Subtypes of behaviour	Standard operation	One cell	One cell-shaded (in one substring)
			25%, 50%, 75%, 100%
		Twenty-four cells	Two cells shaded (one in each substring)
			25%, 50%, 75%, 100%
		Forty-eight cells	Six cells shaded (two in each substring)
			25%, 50%, 75%, 100%

And so, an example of the results outdoor are shown in Figure 13 and Table 5. It demonstrates a clear I-V curve.

Table 5 - Irradiances and temperatures related to Figure 13

Tests	Test 1	Test 2	Test 3	Test 4	Test 5
G [W/m²]	641	636	638,5	641,5	643
T [ºC]	42	42	43	42,5	42

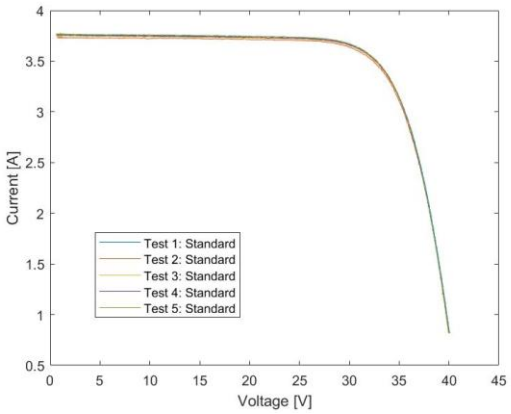

Figure 13 - Healthy panel in standard operation

The results were inserted in the ANN and the output from it is presented in Figure 14. It correlates the known behaviour with the one outputted by the network. The first name tag indicates the targeted behaviour and the second indicates the one found by the model. For example, in Standard → Short circuit (SC) or shading (SH), the panel is in standard behaviour, however the model indicates a short circuit operation. In the graph presented, it is a defined behaviour when the output grants more than 50% of a mode. In a first look at the failure detection, caring only if it managed to know if the system was working without faults or not, it showed a performance of 95% efficiency. It indicated 81 correct outputs, 3 false

alarms and 1 unnoticed failure. From the analysis of the graph, the standard operating condition is confused with short circuit only. It is caused by a low difference in voltage drop, mainly in the test with only a short-circuited cell, a decrease in 0.5 volts. There is also a case in the short circuit of 24 cells in a substring that caused an error in the classification. It chose standard instead of short circuit due to the claw. The one used to short circuit the whole substring by connecting it with both terminals of the bypass diode, not providing contact to the metal.

In Figure 15, the I-V curves of that sample of the test are drawn, and it shows the inefficiency of the used method to cause short circuit in the blue curve, where it fails to represent the loss of approximately 12 volts. The previous analysis gathered a 73% accuracy however here lies an associated error from the simulation model and from the trained neural network results. Using the variables gathered in the outdoor results (temperature and irradiance), a simulation was made, and those values were inputted in the artificial neural network to demonstrate this propagated error.

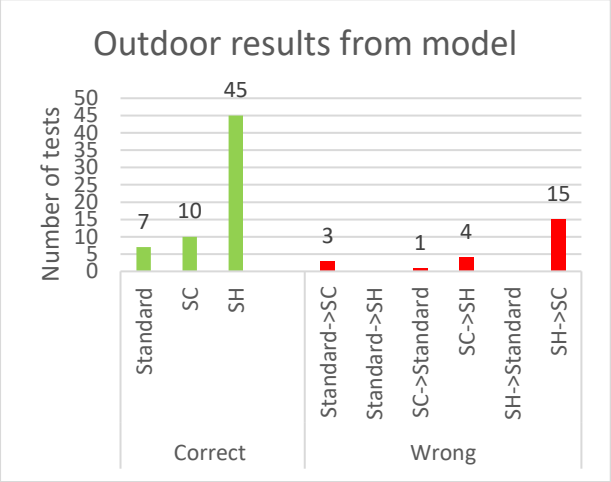

Figure 14 - Outdoor results after passing through the trained model

In Figure 16 the results can be seen. As explained, there is also a percentage error in the artificial neural network and the simulation model which will increase the error in the outdoor test. In the database results, there is a clear area where the trained model has difficulty accessing. It is the short circuit boundaries. If the short circuit is low, it struggles to decide between standard and short circuit. If the short circuit is equal to a substring or more than one, it coincides with the shading operation. From a total of 85 simulated tests, the trained model deduced correctly 70. It portraits an accuracy of 82,35%. In the outdoor results, a more different situation is met, there are more imprecise outputs. Standard and short circuit get mistaken between each other with a low-level drop of voltage.

The measurement error rate can even mistake a short circuit when standard behaviour is applied. Moreover, for shading and a short circuit occurs the misinterpretation stated in the above paragraph. However, there is an indication in the probabilities, showing the feasibility of another behaviour.

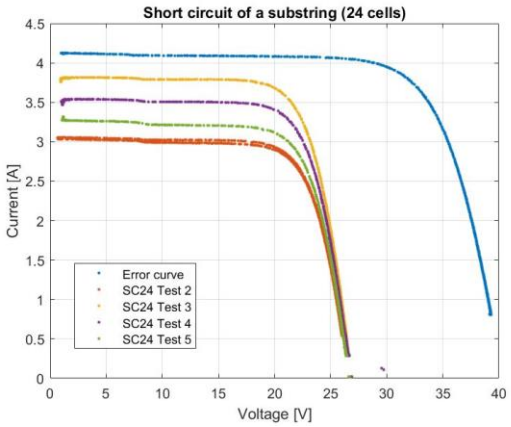

Figure 15 - I-V curves of a short circuit in a substring (24 cells) of the outdoor experiments

A total of 85 outdoor tests were performed, and from that sample, 62 were correctly diagnosed, giving an accuracy of 72,94%. A decrease in performance was expected, considering the measurement errors of all the equipment utilized.

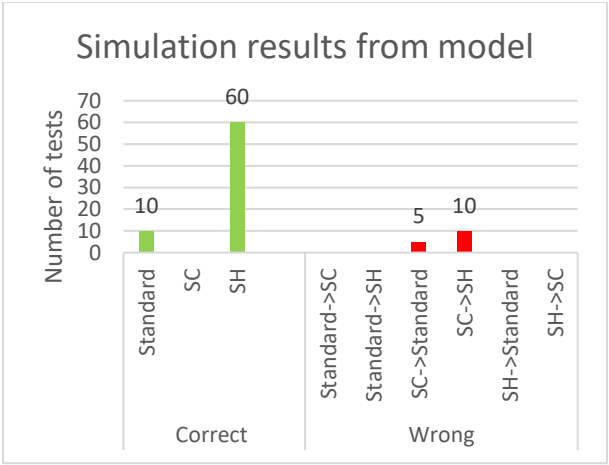

Figure 16 - Simulation results after passing through the trained model

IX. CONCLUSIONS

In this thesis, known behaviours of photovoltaic systems were studied, namely, the standard operating condition, short circuit and shading. Open circuits were not developed in this work as there was no way

of properly test them. However, this fault has a pattern equivalent to a high percentage of shading when the bypass diode activates. After studying and learning about those faults, a knowledge of the one diode five parameter model was gathered to understand the model used to simulate cell behaviour. Then the simulation model was optimized to be able to create the best reference values of each behaviour. First a static database was created, however it created overfitting in the artificial neural network training. A random database proved to be more efficient for ANN training as it did not overfit. Normalization of the input values established better results. However, only when normalized values were approximated to the interval [0;1]. Reducing the number of inputs hindered the convergence process in the ANN. The four inputs cannot be taken out, they are all essential for an accurate detection and diagnosis. Even when one input was normalized using another it proved to complicate the analysis of the artificial intelligence method. Many networks were created with different hidden layer nodes however the ones with five showed the best results. The addition of noise occurred with the intention of creating a threshold in the values, to create the measurement error threshold in the network. It demonstrated overall positive effects, also in removing the overfitting. Then the outdoor test in a photovoltaic panel was done and compared the accuracy of the trained model of the ANN in outdoor circumstances and the accuracy of the simulation model. The primary observation made was of the detection method, where it achieved a performance of 95%. In the diagnosis the accuracy was 73%. In the detection department it displays a very solid achievement. Regarding diagnosis performance, it decreased due to measurement errors and error propagation of all applied methods and techniques, however it showed promising results, proving it is possible and achievable.

From collected results, there are improvements recommend such as an insertion of the topology in the process. Testing as a system and complement the process adding topology to all the steps of the method. Increasing the number of faults tested, with one more, open circuit. If shading is identified, then it can run the program in different hours along the day and if it persists consider an open circuit.

To guarantee reliable accuracy in other systems it should be tested with different panels and other system configurations. This can lead to improvements and more guarantees about this procedure.

X. REFERENCES

[1] S. R. Madeti and S. N. Singh, “Monitoring system for photovoltaic plants: A review,” *Renew. Sustain. Energy Rev.*, vol. 67, pp. 1180–1207, 2017.

[2] M. Green, *Improving Efficiency of PV Systems Using Statistical Performance Monitoring, IEA International Energy Agency, IEA PVPS Task 13, Subtask 2 Report IEA-PVPS T13-07:2017*.

[3] M. Köntges *et al.*, *Performance and reliability of photovoltaic systems subtask 3.2: Review of failures of photovoltaic modules : IEA PVPS task 13 : external final report IEA-PVPS*. 2014.

[4] Abu-Mostafa, Y. S. (1992). Neural networks and learning. Institute of Physics Conference Series (Vol. 127).

[5] S. Laamami, M. Benhamed, and L. Sbita, “Artificial neural network-based fault detection and classification for photovoltaic system,” *Int. Conf. Green Energy Convers. Syst. GECS 2017*, no. 1, 2017.

[6] W. Chine, A. Mellit, V. Lughi, A. Malek, G. Sulligoi, and A. Massi Pavan, “A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks,” *Renew. Energy*, vol. 90, pp. 501–512, 2016.

[7] E. A. S. Filho, F. F. Costa, A. P. N. Tahim, and A. C. C. De Lima, “Photovoltaic panel simulation based on individual cell condition,” *ECCE 2016 - IEEE Energy Convers. Congr. Expo. Proc.*, 2016.

[8] P. Gilman, “SAM Photovoltaic Model Technical Reference SAM Photovoltaic Model Technical Reference,” *Sol. Energy*, vol. 63, no. May, pp. 323–333, 2015.

[9] “Advancing the Science of Solar Data | National Solar Radiation Database (NSRDB).” [Online]. Available: https://nsrdb.nrel.gov/. [Accessed: 15-Oct-2019].

[10] M. Sengupta, Y. Xie, A. Lopez, A. Habte, G. Maclaurin, and J. Shelby, “The National Solar Radiation Data Base (NSRDB),” *Renew. Sustain. Energy Rev.*, vol. 89, no. January 2018, pp. 51–60, 2018.