Enhancing Solar Cell Classification using Mamdani Fuzzy Logic over Electroluminescence Images: A Comparative Analysis with Machine Learning Methods

Hector Felipe Mateo-Romero[1][0000−0002−5569−3532], Mario Eduardo Carbonó dela Rosa[2][0000−0002−6741−6145], Luis Hernández-Callejo[1][0000−0002−8822−2948], Miguel Ángel González-Rebollo[1][0000−0001−9130−7339], Valentín Cardeñoso-Payo[1][0000−0003−1460−158X], Victor Alonso-Gómez[1][0000−0001−5107−4892], and Sara Gallardo-Saavedra[1][0000−0002−2834−5591]

[1] Universidad de Valladolid, Spain,
H.F.M.R. hectorfelipe.mateo@uva.es
L.H.C. luis.hernandez.callejo@uva.es
M.A.G.R. mrebollo@eii.uva.es
V.C.P. valentin.cardenoso@uva.es
V.A.G. victor.alonso.gomez@uva.es
S.G.S. sara.gallardo@uva.es
[2] Universidad Nacional Autónoma de México mecr@ier.unam.mx

Abstract. This work presents a Mamdani Fuzzy Logic model capable of classifying solar cells according to their energetic performance. The model has 3 different inputs: The proportion of black pixels, gray pixels, and white pixels. One additional output for informing of possible bad inputs is also provided. The three values are obtained from an Electroluminescence image of the cell. The model has been developed using cells whose performance has been obtained by measuring the Intensity-Voltage Curves of the cells. The performance of the model has been shown by testing it with a validation set, obtaining a 99.0% of accuracy, when other methods such as Ensemble Classifiers and Decision Trees obtain a 97.7%. This shows that the presented model is capable of solving the problem better than traditional Machine Learning methods.

Keywords: Fuzzy Logic · Photovoltaic · Electroluminescence · Machine Learning.

1 Introduction

A number of different issues (energy crisis, climate change, wars, etc.) are reducing the use of traditional energies in favor of more clean and accessible sources such as renewable energies [1]. This change is also important in Smart Cities since it provides cheaper and cleaner energy. Among the different types of renewable energy, solar energy is one of the most important ones for its facility to be installed in the urban area.

Photovoltaic (PV) modules are composed of a high amount PV cells. These small units can suffer from different problems (mechanical, thermal, or artificial) which can reduce their performance, the amount of energy provided. It is extremely important to verify the conditions of the solar cells in order to optimize production and avoid possible security threats.

Traditionally, the maintenance of PV installations was made by human labor but this is not the best alternative in urban areas or in big installations. To solve this issue, Artificial Intelligence (AI) techniques are being used, helping to optimize the production and to monitor the conditions of the modules [2, 3].

Checking the production of the PV modules is one of the most frequently addressed problems. Different works [4, 5] propose systems to detect defects in the surface of the PV modules. The majority of these methods use a technique known as Electroluminescence (EL) [6] to capture the light emitted by the PV cells/modules when they are injected with electric current, this technique makes visible more kinds of defects than direct visual inspection. These images are used in different AI methods, being Convolutional Neural Networks (CNNs) [7] the method that produces the best results. However, CNN-based methods have some limitations: they need a large amount of data to find patterns, they are highly computer-demanding and their training can be slow.

Other articles have tackled the idea of using fuzzy logic-based models to classify PV cells. The work presented in [8] is applied to detect microcracks in Electroluminescence images, obtaining an efficient system. Another proposal [9] combines fuzzy logic with mathematical morphology to classify the defects from PV cells using photography of the PV. Another work presented in [10] tackles this issue at plant level, comparing the performance of Neural Networks with the performance of Fuzzy Logic Models. Fuzzy logic has been also used in other PV problems such as Max Power Point Tracking [11] or Modelling of PV systems [12]. More works can be found in reviews about the topic [13, 14] but any of them tackles the issue of classifying the PV cells using their EL image in terms of their performance.

This paper presents a new way of analyzing the state of photovoltaic cells, using not only the information about the surface of the cell with the EL images but also the information about the energetic production of the cell, obtained by measuring the Intensity-Voltage (IV) Curve. Another innovation of this paper is that it proposes a Fuzzy Logic (FL) [15] algorithm for solving this problem by analyzing the histogram of the EL images. The advantages that FL provides are that is a not computer-demanding algorithm and it can produce knowledge comprehensible to humans, which is extremely important to understand the effects of the defects in the performance of the cell.

The rest of the paper is organized as follows: Section 2 explains the basis of fuzzy logic, Section 3 explains the methodology used, Section 4 shows the results and findings that can be observed from them, finally Section 5 presents the conclusions of the paper.

2 Introduction to Fuzzy Logic

The term fuzzy represents values that are not clear. Fuzzy Logic [16] is an extension of the traditional logic [17] where the truth value of a variable is a real number between 0 and 1, instead of the traditional values of true or false. It can be applied to models that use imprecise information and for dealing with uncertainty in decision-making.

The most important concept of Fuzzy Logic is the membership function, which defines the degree of membership of the variable to a certain set or category. The membership is a function that can provide any value between 0 and 1, being 0 non-membership and 1 full-membership. FL systems are tolerable to errors or noise in the input data.

Among the different kinds of fuzzy logic systems, the Mandami systems are the most used FL Inference Systems [18], their most important features are to following:

- They are more intuitive and have easier-to-understand rules.
- Each Output has a corresponding membership function
- The surface of the output is discontinuous
- High Expressive Power and Interpretable
- Less Flexibility in the system design

3 Methodology

This section will explain the different processes followed in the creation of the models which include the gathering, preprocessing, and labeling of the data, the creation of the rules of the fuzzy model, and its optimization.

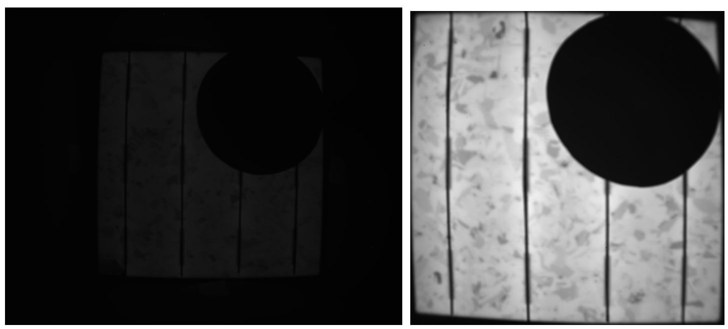
3.1 Data gathering

Image acquisition was not a trivial process, since it was necessary to obtain two different things: The Electroluminescence (EL) image of each cell using a EL camera and the Intensity-Voltage Curve (IV) using an IV-tracer [19] which will provide the information about the energetic performance of the cells; more details about the processes of gathering can be found in [20].

The dataset is composed of the original measurements presented in [20] with some additional data that was obtained exclusively for this work, resulting in 666 different images and their IV curve.

3.2 Image Preprocessing

The preprocessing of the images was performed using the same procedure as in other works: Removal of dead pixels and luminous noise, fixing the scale of lighting of the images, removing the black surrounding contours of the images, and fixing the perspective. Fig 1 shows an example of an image after each of the processes. More information can be found in [20] and was performed using Python.



(a) Original Image. This image needs to be preprocessed.

(b) Image after the preprocessing

Fig. 1: A sample image before and after preprocessing.

3.3 Maximum Power Normalization

As explained before, the IV curve of each PV cell provides information about the energetic production of the cell, but it is necessary to perform certain steps to make the data useful for a model. Two different techniques (Z-score normalization and Min-Max normalization) are used together to obtain a normalized variable with values between 0 and 1. The following process is used:

1. Computation of the Power-Voltage curve of each cell using the information about the IV curve.
2. Calculation of the maximum value of power (Maximum Power Point) for each of the curves.
3. The cells are divided into six different groups, depending on the irradiance that was used to obtain the measures. For each group:
   (a) The mean value and the standard deviation of maximum power are computed.
   (b) A Z-score normalization is performed, using the computed values of mean and standard deviation, this results in a variable with a mean of 0 and std of 1, with values between -2 and 2.
   (c) The maximum value and minimum value of the obtained variable are computed for each group.

   (d) A Min-max normalization is performed on each group, using the computed values of the maximum and the minimum. This results in a variable with values between 0 and 1, a mean of 0.5, and std of 0.2.
4. This value measures the relative performance of a cell, high values correspond with cells that have high energy production, and low values with cells that are not producing as much as they should.
5. After that, the data was divided into 3 groups, according to their value:
   (a) Class 0 ($0.81 <= X$) represents the cells that are in good condition since their performance is near the expected value.
   (b) Class 1 ($0.572 <= X < 0.81$) represents the cells whose performance is enough but not as high as it should be.
   (c) Class 2 ($X < 0.572$) represents the cells that do not have enough performance, due to their defects or other problems.

3.4 Feature Extraction

Traditional AI methods can not deal directly with images, so it is necessary to obtain manageable characteristics from the images for these methods.

Fig. 2 presents an image and its intensity histogram. It can be seen how the histogram has three different regions: The area of the first peak corresponds with pixels with low-intensity values (Black or dark), and these pixels correspond with zones where the cell is not emitting light in response to the electric current. The second peak corresponds with the zones where the cell is active since it is producing light in response to the electric current. The area after the second peak corresponds with areas where energy production is extremely high.

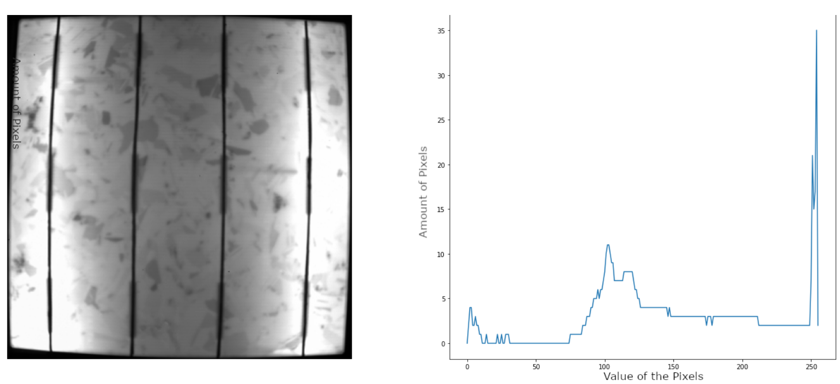


Fig. 2: Sample of Image and its histogram

After analyzing all the images, it can be seen that these aspects appear in most of the images of the dataset, the number of dark areas is directly connected with the performance of the cells (more dark areas imply less production).

Taking these facts into consideration, the final features selected were the proportion of dark pixels, the proportion of gray pixels, and the proportion of white pixels. The following process was used:

- The data is divided into two different sets: Training (80%, 532 samples) and Validation (20%, 134 samples). The training set is composed of 92 images of Class 0, 228 of Class 1, and 212 of Class 2. The validation set is composed of 22 images of Class 0, 53 of Class 1, and 58 of Class 2. The following steps are repeated for each set.
- For each image, the intensity histogram is computed and all the histograms are accumulated into a summary histogram, normalized between 0 and 1.
- The separation intensity value between the black and the gray zone is calculated using the intensity at the minimum between the two first peaks. The resulting value after rounding was 0.35.
- The point to divide the gray area and the white area is calculated using the minimum point between the two peaks. The resulting value after rounding was 0.70.
- For each image, the amount of pixels for each group is computed and divided by the number of total pixels of the image to get the ratio of black, gray, and white pixels which provide the features to characterize the image.

3.5 Model

As it has been discussed before, the main objective of this article is to create a mode capable of classifying PV cells in terms of their performance. Fuzzy Logic models have some great benefits, that are extremely interesting for this problem: FL is a symbolic method, which means that the knowledge that provides can be easily understood by humans. This is an extremely important quality since it can help to find new patterns that are not visible to the human mind. Moreover, FL algorithms are not computer-demanding and they can be run on almost every kind of device nowadays, which makes them extremely useful in a lot of different areas.

The presented model is a Fuzzy Logic Model based on the Mandami Inference System, which provides more intuitive and easier-to-understand rules and other Inference Systems, it has been implemented using Matlab with the application of Fuzzy Logic Designer. The design parameters of the proposed membership functions for each input were set according to the expert's experience and based on the behavior of the model's input variables at each actual classification level. Moreover, other design parameters such as the shape of the membership functions, the degree of membership, and the range of the output membership functions were set based on the statistical error between true and predicted classification.

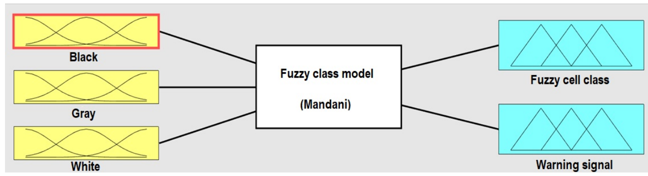
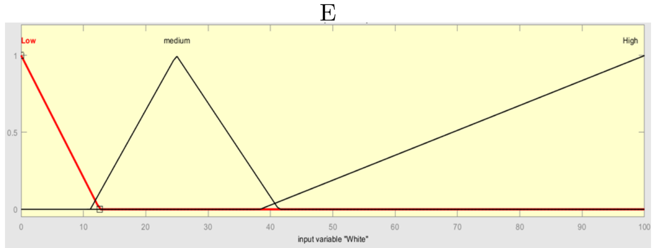


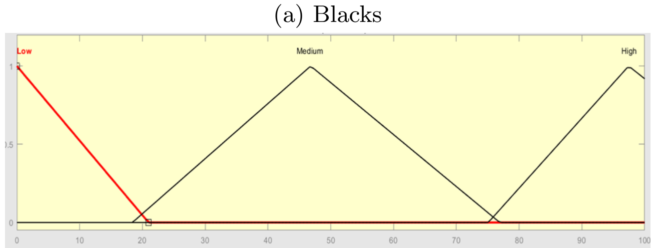
Fig. 3: Structure of the FL model: 3 inputs and 2 outputs

The model (see Fig. 3) has been designed with 3 different inputs: The proportion of dark pixels, the proportion of gray pixels, and the proportion of white pixels in the image. It also has two different outputs: The condition of the cell and a warning that indicates inputs that should be checked due to a possible problem in the image (proportions not summing up 100% or extreme values such as black 100%).

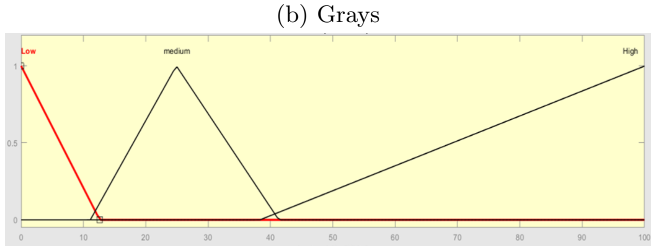
E



(a) Blacks



(b) Grays



(c) Whites

Fig. 4: Membership functions of the inputs

Membership Functions The inputs (see Fig. 4) have three different membership functions each one: Low, Medium, and High which corresponded directly with the proportion of pixels of that particular input.

The output of the classification (see Fig. 5) has also three different membership functions which correspond with each class: High Performance (0), Medium Performance (1), and Low Performance (2).

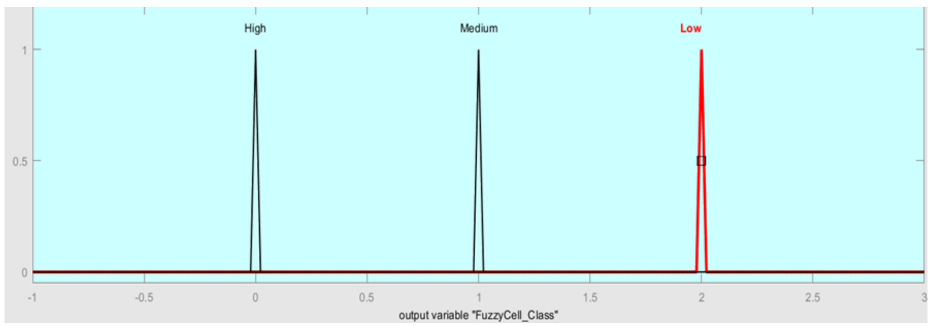


Fig. 5: Membership function of the classification output

The warning has two different membership functions: Negative and Positive (Fig. 6).

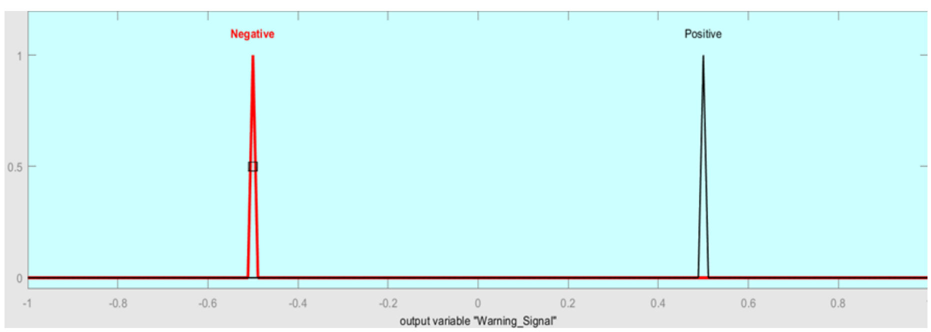


Fig. 6: Membership function of the warning

Rules The final rules of the models can be seen in Table 1. These rules were obtained using the knowledge of experts of the domain. Their knowledge about the effects of defects in the performance of cells combined with information about the output power obtained from the IV curve was used to create rules that classified the images in their corresponding class, trying to maximize the accuracy of the classification output in the training set. The output of the warning is not taken into account, as it is only used to detect when the inputs are not valid in order to warn the users that they should check that input. This process of parameter fitting is completely manual, in contrast with the training phase of Machine Learning algorithms. The validation test has not been considered in this modeling process. 27 rules were obtained by combining all of the possible states of the three inputs ($3^3 = 27$).

Rule	Black	Gray	White	Classification Output	Warning Signal
1	Low	Low	Low	High	Positive
2	Low	Low	Medium	Low	Negative
3	Low	Low	High	Low	Negative
4	Low	Medium	Low	Low	Negative
5	Low	Medium	Medium	Low	Negative
6	Low	Medium	High	Low	Negative
7	Low	High	Low	Medium	Negative
8	Low	High	Medium	Low	Negative
9	Low	High	High	Low	Positive
10	Medium	Low	Low	High	Positive
11	Medium	Low	Medium	Medium	Negative
12	Medium	Low	High	Medium	Negative
13	Medium	Medium	Low	Medium	Negative
14	Medium	Medium	Medium	Medium	Negative
15	Medium	Medium	High	Medium	Positive
16	Medium	High	Low	High	Negative
17	Medium	High	Medium	High	Positive
18	Medium	High	High	High	Positive
19	High	Low	Low	High	Negative
20	High	Low	Medium	High	Negative
21	High	Low	High	High	Negative
22	High	Medium	Low	High	Negative
23	High	Medium	Medium	High	Negative
24	High	Medium	High	High	Negative
25	High	High	Low	High	Positive
26	High	High	Medium	High	Negative
27	High	High	High	High	Positive

Table 1: Fuzzy Rules of the model

Fig. 7 presents the surface 3D diagram of the classification output. The diagram represents the knowledge of the model, and how the inputs blacks and grays modify the output depending on their values. It can be seen how low values of black implies an output of 0. Class 1 only appears when black is around 20%-30% and grays are less than 60%. Class 2 is selected in the other cases.

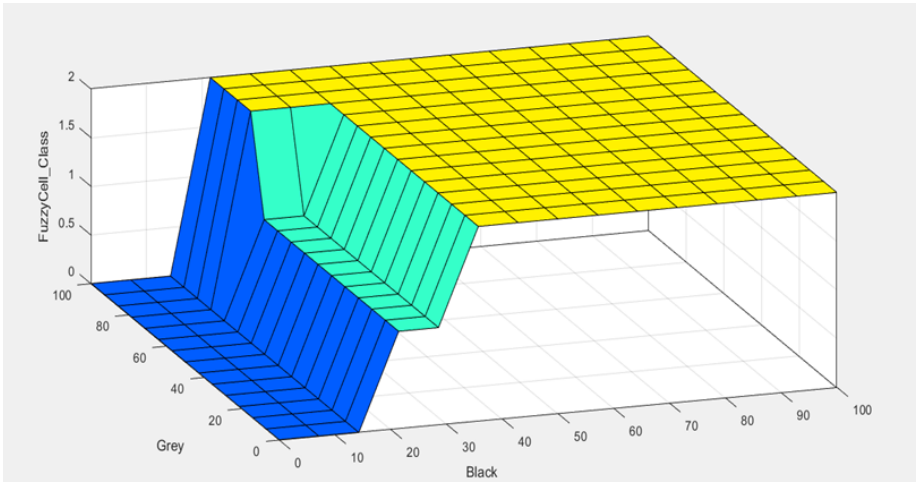


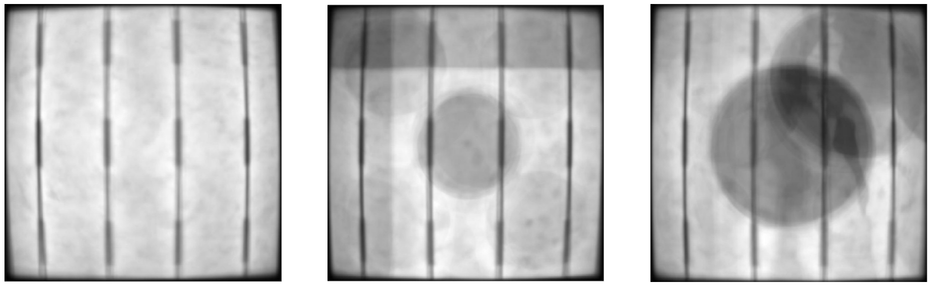
Fig. 7: 3D diagram for the classification output for two inputs: Blacks and Grays

4 Results

This section assesses the quality of the model by showing its performance in the validation set and compares the performance with other methods. The other methods have been implemented using the application Classification Learner from Matlab.

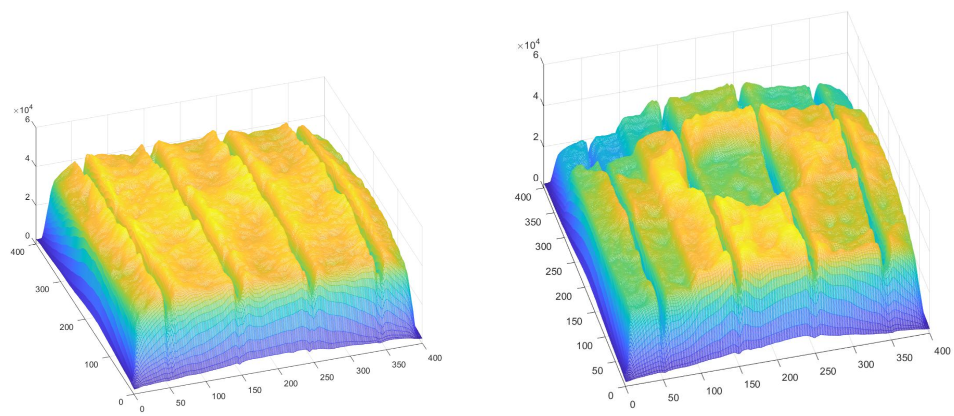
As explained before, the dataset was composed of 666, divided into two sets: Training (80%, 532 samples) and Validation (20%, 133 samples).

Figs. 8 and 9 present the distribution of the validation dataset, which can be seen in Fig. 8a how the mean of all of the images of Class 0 represents a cell in good condition, 8a also show this fact, with the stacking of all of the images of this class. Similar reasoning can be used with the images of class 1 (Figs. 8b and 9b), since they present minor defects that do not cover a high amount of the surface of the cell. Finally, Figs. 8c and 9c show how the images in class 2 have shadows that cover a high amount of the surface of the cell.

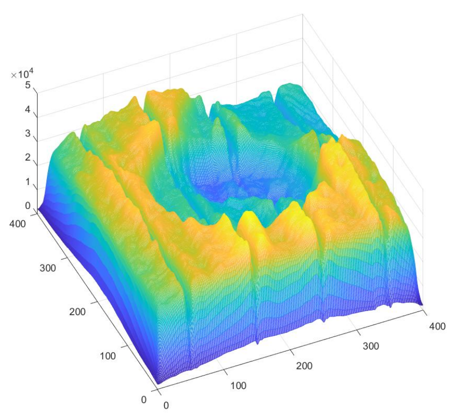


(a) Class 0 (b) Class 1 (c) Class 2

Fig. 8: Image obtained after making the mean of all of the images of each class from the validation set.



(a) Class 0 (b) Class 1



(c) Class 2

Fig. 9: 3D Diagram obtained after stacking the images of the same class of the validation set.

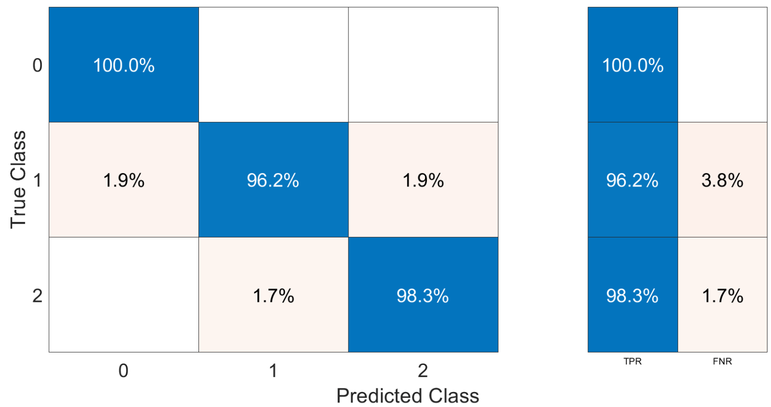
The model obtained a 99% of accuracy in the Validation set . Fig. 10 presents the results of the classification of this set, with information about the confusion matrix and the accuracy for each class. It can be seen how the performance in the three classes is quite similar, with a slight decrease in class 1. It can also be seen that the incorrect classification appears between adjacent classes, there are not any mistakes between class 0 and class 2.

True Class		Confusion Matrix			*TPR*	*FNR*
	0	21	1	0	**99,09%**	**0,91%**
	1	1	52	2	**98,12%**	**1,88%**
	2	0	2	53	**99,31%**	**0,69%**
		0	1	2		
			Predicted Class			

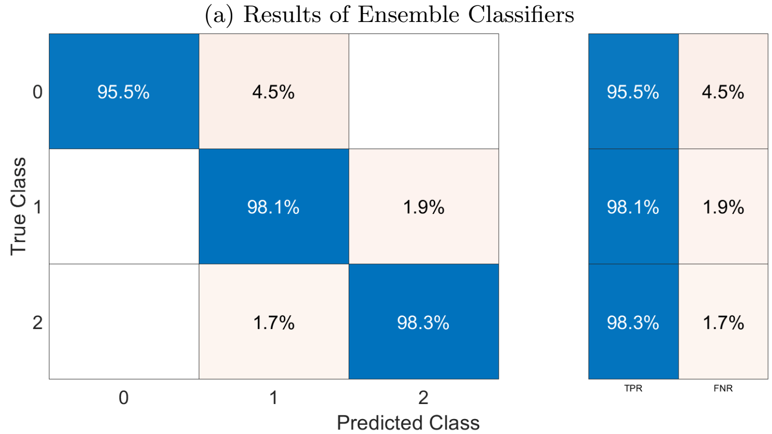
Fig. 10: Results of the classification of the proposed fuzzy model on the Validation Set. TPR: True Positive Rate, FNR: False Negative Rate.

Different methods were chosen for comparison with the presented method, all of them can be found in the application Classification Learner of Matlab. The selection of methods was composed of Decision Trees, Discriminant Analysis, Logistic Regressions Classifiers, Naive Bayes classifiers, Support Vector Machines, Nearest Neighbor Classifiers, and Ensemble Classifiers. Decision Trees and Ensemble Classifiers obtained the best performance with a 97.7% of accuracy in both of them.

Fig. 11 presents the classification matrix for both methods, they provide a good classification but the results are a bit lower than the proposed method, as can be seen in their accuracy. This is clear evidence of the importance of applying fuzzy logic to solve this problem.



(a) Results of Ensemble Classifiers



(b) Results of Decision Trees

Fig. 11: Results of the classification of other models on the Validation Set. TPR: True Positive Rate, FNR: False Negative Rate

5 Conclusions and future work

The detection of the conditions of the solar cells is a really important problem since it provides information that is vital in the optimization of photovoltaic production. The introduction of fuzzy logic to solve this problem is innovative since few works have tried this approach. The presented model has been tested with a 99% of accuracy as opposed to the 97.5% that obtains other models such as Ensemble Classifiers and Decision Trees.

The method has some flaws that need to be addressed to improve it. First of all, the creation of the rules has been made manually which can produce a certain bias, even if the knowledge of experts has been used to ensure the quality of the rules. To solve this different measures would be necessary: The inclusion of a new dataset of images, completely different from the images of training of validation, to verify the performance of the model in completely foreign conditions. Another important improvement would be the automatization of the creation of the rules, using Machine Learning to optimize this process.

6 Acknowledgements

The Universidad de Valladolid supported this study with the predoctoral contracts of 2020, co-funded by Santander Bank. This work has been financed also by the Spanish Ministry of Science and Innovation, under project PID2020-113533RB-C33. The Universidad de Valladolid also supported this study with ERASMUS+ KA-107. Finally, we have to thank the MOVILIDAD DE DOCTORANDOS Y DOCTORANDAS UVa 2023 from the University of Valladolid. We also appreciate the help of other members of our departments.

References

1. REN21. *Renewables 2022 Global Status Report.* REN21, 2022.
2. Hector Felipe Mateo Romero, Miguel Angel Gonzalez Rebollo, Valentin Cardenoso-Payo, Victor Alonso Gomez, Alberto Redondo Plaza, Ranganai Tawanda Moyo, and Luis Hernandez-Callejo. Applications of artificial intelligence to photovoltaic systems: A review. *Applied Sciences*, 12(19), 2022.
3. Adel Mellit and Soteris A. Kalogirou. Artificial intelligence techniques for photovoltaic applications: A review. *Progress in Energy and Combustion Science*, 34(5):574–632, 2008.
4. Dhanup S. Pillai, Frede Blaabjerg, and Natarajan Rajasekar. A comparative evaluation of advanced fault detection approaches for pv systems. *IEEE Journal of Photovoltaics*, 9(2):513–527, 2019.
5. Ying-Yi Hong and Rolando A. Pula. Methods of photovoltaic fault detection and classification: A review. *Energy Reports*, 8:5898–5929, 2022.
6. Luis Hernández-Callejo, Sara Gallardo-Saavedra, and Víctor Alonso-Gómez. A review of photovoltaic systems: Design, operation and maintenance. *Solar Energy*, 188:426–440, aug 2019.
7. Laith Alzubaidi, Jinglan Zhang, Amjad J Humaidi, Ayad Al-Dujaili, Ye Duan, Omran Al-Shamma, J Santamaría, Mohammed A Fadhel, Muthana Al-Amidie, and Laith Farhan. Review of deep learning: concepts, cnn architectures, challenges, applications, future directions. *J Big Data*, 8:53, 2021.
8. Rashmi Chawla, Poonam Singal, and Amit Kumar Garg. A mamdani fuzzy logic system to enhance solar cell micro-cracks image processing. *3D Research*, 9:1–12, 9 2018.
9. Wei Junchao and Zaibin Chang. Defect detection on solar cells using mathematical morphology and fuzzy logic techniques. 1:3.
10. Mahmoud Dhimish, Violeta Holmes, Bruce Mehrdadi, and Mark Dales. Comparing mamdani sugeno fuzzy logic and rbf ann network for pv fault detection. *Renewable Energy*, 117:257–274, 2018.
11. F. Chekired, C. Larbes, D. Rekioua, and F. Haddad. Implementation of a mppt fuzzy controller for photovoltaic systems on fpga circuit. *Energy Procedia*, 6:541–549, 2011. Impact of Integrated Clean Energy on the Future of the Mediterranean Environment?

12. Kunal Sandip Garud, Simon Jayaraj, and Moo-Yeon Lee. A review on modeling of solar photovoltaic systems using artificial neural networks, fuzzy logic, genetic algorithm and hybrid models. *International Journal of Energy Research*, 45(1):6–35, 2021.
13. M Sridharan. Short review on various applications of fuzzy logic-based expert systems in the field of solar energy. *International Journal of Ambient Energy*, 43(1):5112–5128, 2022.
14. Nebiyu Kedir, Phuong H. D. Nguyen, Citlaly Pérez, Pedro Ponce, and Aminah Robinson Fayek. Systematic literature review on fuzzy hybrid methods in photovoltaic solar energy: Opportunities, challenges, and guidance for implementation. *Energies*, 16(9), 2023.
15. Guanrong Chen, Trung Tat Pham, and NM Boustany. Introduction to fuzzy sets, fuzzy logic, and fuzzy control systems. *Applied Mechanics Reviews*, 54:B102–B103, 11 2001.
16. Vilem Novak, Irina Perfiljeva, and Jiri Mockor. *Mathematical Principles of Fuzzy Logic*. 01 1999.
17. Stewart Shapiro and Teresa Kouri Kissel. Classical Logic. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2021 edition, 2021.
18. E. H. Mamdani. Applications of fuzzy algorithms for control of a simple dynamic plant. *Proceedings of the IEEE*, 121:1585–1588, 1974.
19. José Ignacio Morales-Aragonés, Víctor Alonso Gómez, Sara Gallardo-Saavedra, Alberto Redondo-Plaza, Diego Fernández-Martínez, and Luis Hernández-Callejo. Low-cost three-quadrant single solar cell i-v tracer. *Applied Sciences*, 12(13), 2022.
20. Hector Felipe Mateo-Romero, Luis Hernandez-Callejo, Miguel Angel Gonzalez Rebollo, Valentín Cardeñoso-Payo, Victor Alonso Gomez, Hugo Jose Bello, Ranganai Tawanda Moyo, and Jose Ignacio Morales Aragones. Synthetic dataset of electroluminescence images of photovoltaic cells by deep convolutional generative adversarial networks. *Sustainability*, 15(9):7175, Apr 2023.