



Using Computer Vision with Edge Machine Learning to Recognize Dirt on Photovoltaic Modules

Raimundo Eider Figueredo^{1*}, João Lucas de Souza Silva², João Frederico Souza de Paula², Marcel Veloso²,
Marcelo Vinicius de Paula^{2,3}, Nurhayati Nurhayati⁴, Arnaldo de Carvalho jr.⁵, Alexandre Maniçoba de
Oliveira⁵, Auzuir Ripardo de Alexandria^{5,6}

¹ Technological Federal University of Paraná (UTFPR), Brazil

² State University of Campinas (UNICAMP), Brazil

³ College of Mechanical Engineering - UNICAMP, Brazil

⁴ University Negeri Surabaya, Indonesia

⁵ Maxwell Laboratory of Microwaves and Applied Electromagnetism – IFSP, Brazil

⁶ Federal Institute of Ceará (IFCE), Brazil

*Correspondence: E-mail: raimundoeider@alunos.utfpr.edu.br

ARTICLE INFO

Article History:

Submitted/Received 28 January 2025

First Revised 21 May 2025

Accepted 26 May 2025

First Available Online 11 June 2025

Publication Date 11 June 2025

Keyword:

TinyML. Soiling Photovoltaic Modules.
Edge Computing. Embedded Intelligence
in Microcontrollers. Machine Learning at
the Edge.

ABSTRACT

This paper explores the guidelines of edge computer vision, using the TinyML technique, to remove dirt from photovoltaic modules, to optimize photovoltaic energy generation and extend the lifespan of the modules. The research used two photovoltaic cells: one clean with and without shading and the other with dust. The images were collected with an Esp-32 Cam microprocessor and preprocessed on the Edge Impulse platform, which uses TensorFlow Lite to create edge Machine Learning (ML) models. The trained model, which can be embedded in the microcontroller, performs inference through computer vision to consider clean cells (PV_Clean), dirty cells (PV_Dirt), and partial shading (PV_Shading). The results showed 100% accuracy in detecting the three classes, evidencing the effectiveness of the model in distinguishing between cell states. Inference tests with a smartphone camera confirmed the accuracy of the model. In short, the study demonstrated TinyML's potential in identifying dirt on photovoltaic modules, which can help with the efficiency of solar energy generation and operation and maintenance of the levels of dirt considered to be damage to the photovoltaic system. It can also help identify critical levels of soiling on photovoltaic modules, enabling the creation of customized cleaning schedules for each location and season. This allows the schedule to be activated only when fouling is affecting energy production, optimizing resources and reducing cleaning maintenance costs.

1. INTRODUCTION

Climate change is mainly caused by gases with intense thermal effects and rainfall patterns, mainly causing periods of prolonged drought. Since Brazil's main energy matrix is hydroelectric power, accounting for 55% of its total, there is a need to diversify its electricity generation matrix. Given this, the Brazilian northeast has been standing out in generating wind and solar energy due to its climate conditions. Photovoltaic energy already represents more than 20 % of the energy matrix for the diversification of the electrical matrix, this includes centralized, mini, and micro generation [1]. Its advantages are low maintenance costs, guaranteed efficiency with regular cleaning of the modules for more than 25 years and the ability to be installed in remote locations without access to energy distributed by concessionaires. Its biggest disadvantage is that it does not provide energy at night and requires high investment for storage due to the high cost of batteries when considering off-grid installation.

The use of solar energy through photovoltaic modules, with the aim of producing electricity, is considered one of the promising markets for renewable energy. With the prospect of rapid growth and high levels of investment, the market for this type of electricity generation is increasingly competitive around the world [2]. In Brazil, according to the SIGA system (ANEEL Generation Information System) of National Electric Energy Agency (ANEEL), there are 146.695 GW of power granted for the construction of photovoltaic generation units, and it is observed that it is the source with the most power granted in relation to the others. Currently, photovoltaic plants represent 6.28% and are the fourth in centralized generation [1], which shows that there may be a large expansion in photovoltaic generation, which will also require technologies that help reduce costs in operation and maintenance.

Photovoltaic solar energy is an energy source that converts sunlight into electricity without emitting gases or noise [3], which uses the photovoltaic effect, observed in semiconductor materials, for example, silicon, which enables the creation of a crystalline network of photovoltaic cells used in photovoltaic modules [2].

The cells of photovoltaic modules use semiconductors, as described above, which absorb the sun's irradiance and convert it into electrical energy. Therefore, besides intrinsic material properties, the module's output power depends on the solar irradiance and the cell temperature, i.e., low irradiance leads to low power, and high temperature also causes a reduction in the output power. Another factor that must be taken into account is shading, which even if slight affects the current supplied by the photovoltaic module, but the voltage remains the same [4].

Regarding the photons from the sun that hit the surface of the module, they are not all converted into energy due to the limitations of the module technology. However, we can also mention environmental factors, such as the accumulation of dirt on the surfaces of the modules (soiling), which affect the maintenance and electricity production costs of a photovoltaic system, mainly due to the reduction in the conversion of photons into electricity. In addition to irradiance, latitude, longitude, angle of inclination and orientation are variables that affect the operation of a photovoltaic solar system precisely about the accumulation and deposition of dirt on the surface of the modules [5].

Dirt on modules is divided into three factors that influence the deposition of dirt on the surface, namely: environment, dust types, and installations. Environmental factors are related to wind speed and direction, temperature, irradiation, air pressure, polluted air, sandstorms, and humidity. Dust/dirt types are related to nature and are clay, sand, bacteria, carbon, and animal excrement. Installation types are defined as installation locations, module types, orientation, latitude and longitude, orientation angle, height, exposure time, and module surface type [5].

Still speaking about dirt, the size of dust particles and structural components change from one location to another [6-7], as well as the increase in dust density affects the productivity of the photovoltaic module, causing it as explained in the work of [8]. The size of the dust particle also interferes with the handling of the modules, as [9] reported that the accumulation of fine dust particles on the surface of the modules resulted in greater handling performance for modules subjected to the accumulation of larger dirt particles.

The study by [10] defined that rainfall with a minimum of 20 mm has the capacity to clean the surface of the modules, however, in lower rainfall, the water droplets combine with the dust in the air deposited on the surface of the module, causing a higher level of dirt.

Since photovoltaic modules suffer from thermal stress caused by dirt, degrading production and efficiency in the production of energy from sunlight through the photovoltaic effect, the benefits of washing the modules are doubled by acting on the cooling and cleaning of photovoltaic modules [11]. The use of automated intelligent cleaning systems with water reuse and capture mainly for regions with little rainfall, such as the Brazilian northeastern interior. According to SIGA, this region has 86,920 GW of granted power, representing 59.25% of the total of Brazil [1].

Within the context described, the problem that needs to be solved is the dirt that degrades the productivity of the module. This problem directly impacts the useful life of the module due to heating, which can create micro-cracks, resulting in a decrease in energy and lifespan [10], in addition to causing non-uniformity of the incidence of irradiance in the module plane, aggravating mismatch losses in the module [11]. For this purpose, the use of embedded computer vision at the edge, which has low energy and processing consumption, can help with the problem of dirt at a low cost. Therefore, it is important to develop research such as the one proposed in this work.

That said, several applications use computational intelligence to assist in production forecasting, control, and management of renewable energy. Computational intelligence methods are known as tools that improve the performance of energy systems in production and transfer. Deep learning algorithms have been applied to different fields of renewable energy, providing high prediction accuracy for solar and wind energy compared to traditional methods [12].

Computational intelligence has four main components, namely: processing, training, decision, and result. One type of machine intelligence that this work intends to study and apply is edge neural networks. Conventional artificial neural networks have several layers and are trained in such a way as to achieve high performance, being very efficient, especially in the application of computer vision. It is a method that requires a high computational cost. On the other hand, it needs a large data set to be an efficient method. This conventional method mentioned above requires a high graphics processing unit due to the nature of the data required. However, Tiny Machine Learning (TinyML) is a technology that is applied at the edge and uses low computational processing, enabling edge computing that uses low latency and high availability of various network recognition services, better privacy, security, and reliability for users, in addition to low energy consumption and connectivity in Internet of Things (IoT) devices, making it intelligent [13].

The application of TinyML is important for the development of intelligent IoT applications due to its possibility of use in speech and image recognition, sign language prediction, hand gestures, body pose estimation, keyword localization in images, voice activation, respiratory symptoms related to cough, ecological conservation, autonomous vehicle, and anomaly detection [13].

2. FAULTS AND ANOMALIES IN PHOTOVOLTAIC SYSTEMS

Faults in photovoltaic systems can arise on the Direct Current (DC) side, which ranges from the photovoltaic modules to the inverter input. Therefore, data analysis is a necessary tool to

understand the performance of photovoltaic systems. The accuracy of this data depends on the quality of the sensors and models used, as well as the general condition of the set. The better the data, the easier it is to identify faults and anomalies [14].

The most common faults on the DC side are incompatibilities, short circuits, open circuits, line-line or line-ground. Such faults occur in components considered to be part of the balance of the photovoltaic system, namely: cables, fuses, switches, and inverters. Now, considering only the photovoltaic modules, they can present faults and anomalies such as dirt, hot spots, breakage, and delamination during operation [15]. Faults on the Alternating Current (AC) side include consequences of blackouts, in addition to abnormalities in the network, which are frequently caused by unbalanced voltage and atmospheric discharge events [16]. Regarding the classification of failures in photovoltaic systems, three types are defined: physical, environmental, and electrical [17].

The faults reported in the literature may be categorized as physical failures refer to problems in the panels, such as internal damage to the photovoltaic cells, microcracks in the modules, defects in bypass diodes, and degradation. Environmental failures are related to shading caused by bird droppings, dust accumulation, cloud movement, and tree shadows. Electrical failures are related to problems in the Maximum Power Point Tracked (MPPT), open circuits, grounding, faults between lines, short circuits, electric arcs, and islanding mode operation situations [18].

The types of failures classified in the photovoltaic solar energy generation system are shown in Figure 1.

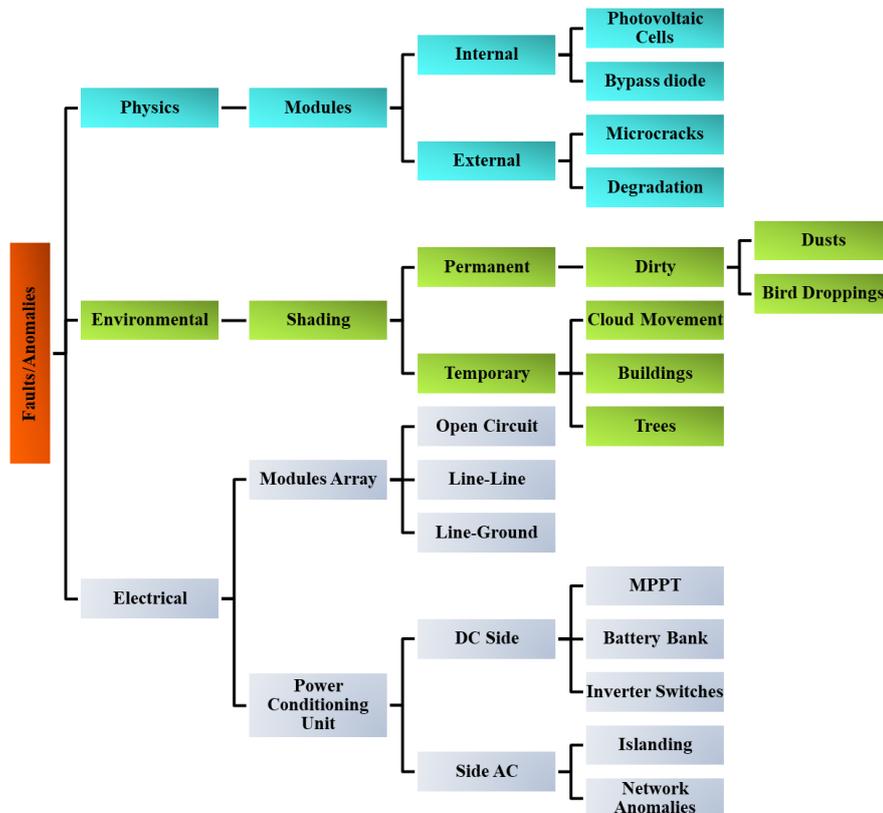


Figure 1. Classification of types of faults existing in the photovoltaic system [14].

Dirt on photovoltaic modules causes increasing energy losses resulting from the accumulation of snow, dirt, dust, bird droppings, and other particles that accumulate and cover the surface of the photovoltaic module, causing shading that can cause hot spots and increasing

of mismatch losses, affecting the overall energy supplied by the module [4]. In addition to reducing generation efficiency due to interference from the incidence of solar irradiance, causing the appearance of hot spots that accelerate the handling process of the modules [19].

Different factors can contribute to the deposition of dirt, some are environmental, such as wind speed and direction, relative humidity, air pollution, and ambient temperature. Others are related to the construction of the system, such as the characteristics of the module surface and the orientation angle of the modules. Dirt has a direct impact on cleaning and production costs in energy generation [19].

Dirt is not only related to the number of particles accumulated on the surface of the modules but the physical-chemical composition must be taken into account. Therefore, chemical properties, such as salinity, conductivity, and pH, and physical characteristics, such as color, texture, size, and shape, are relevant to the influence of dirt on photovoltaic systems. The work in [19] also argues that the morphological analysis of dirt should be considered in the impact that photovoltaic systems suffer from it.

As the location is a factor that affects and has particularities in the type of dirt, the work [4] demonstrates the levels of dust intensity in the world, in which Brazil is in Zone 2 with levels between 16-19. Uruguay, Paraguay, and several countries on the African continent have maximum dust intensity levels (Zone 4: 65-96).

That said, it is important to use mechanisms to identify the level of dirt since the impact on photovoltaic systems has costs in terms of energy generation losses and accelerates module degradation. This can minimize losses. Therefore, the need for this work to study ways to identify this dirt arose.

3. METHODS

The development process went through four steps: (i) data collection, (ii) preprocessing, (iii) machine learning, and (iv) model testing. Figure 2 presents the workflow of the work development in the dirt recognition process.

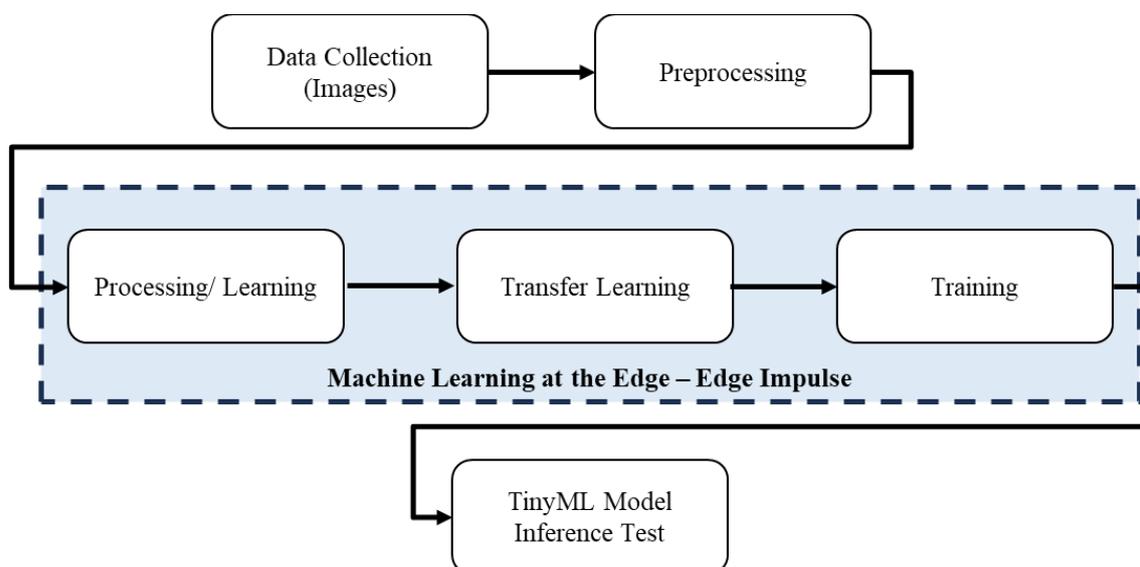


Figure 2. Development flow for training for skill recognition by computer vision.

In the data collection phase, images were taken using two photovoltaic cells. The first cell for the images is clean and shaded by a physical barrier using a plant. In the second, for the images with dust, an indeterminate dust was applied. With the Esp-32 Cam microprocessor

that has a 2 MP OV-2640 camera, images of the clean photovoltaic cell were collected, then those of cells with shadowing and, finally, images of the cell with dirt. The collected images are shown in Figure 3 and are pixelated due to the low quality of the Esp-32 Cam camera and because microcontrollers have low storage capacity memories, which is why low-quality images were chosen.

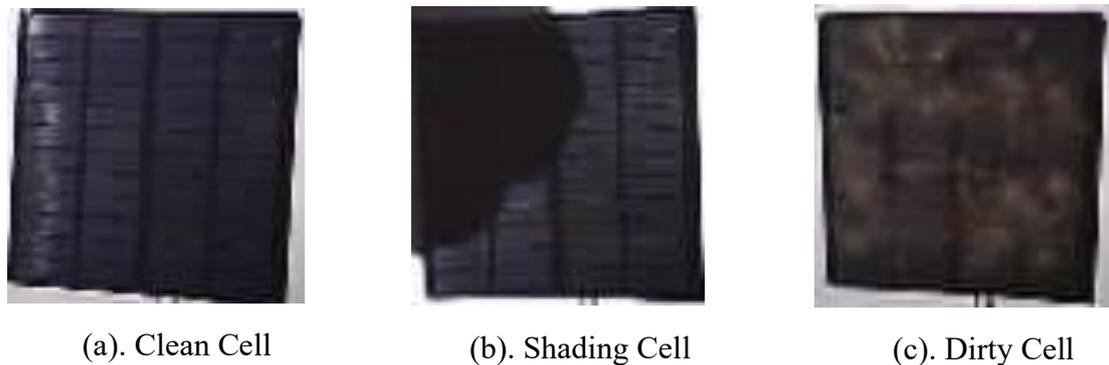


Figure 3. Images collected for training cell recognition (a) clean, (b) shaded, and (c) dirty.

The second stage involved preprocessing the images, which were separated and selectively labeled as PV_Clean for clean cells, PV_Shading for shaded cells, and PV_Dirt for dirty cells. Of these, 21 images were from the shading class, 9 from the clean class, and 14 from the dirty class, for a total 44 of 75 images. The images were divided into 59% (44 images) for training and 41% (31 images) for testing/validation.

The next stage, which deals with edge machine learning, used the Edge Impulse platform [20]. It is a cloud platform for developing ML on edge devices. Edge Impulse uses a set with TensorFlow Lite to create ML models and load them onto microcontroller boards [21]. This stage is divided into three substages within the platform, as it processes the transported images. Here, grayscale images were chosen to reduce processing on microcontrollers. Then, transfer learning was performed, and then training was performed, in which the parameters shown in Table 1 were chosen.

Table 1. Training parameters chosen on the Edge Impulse platform.

Training Parameters	
Training cycles	20
Learning rate	0,0005
Validation rate	0,20
Neurons	8

Still dealing with the training sub-stage configurations, the MobileNet V2 convolutional network was selected, which separates the convolutional action according to depth. For this, a division is performed in two stages, the first of which is known as deep convolution, while the second is known as point convolution. In addition, MobileNet uses residual blocks, in which each block has three distinct layers, called: (i) expansion, (ii) depth convolution, and (iii) projection [22].

The last stage is model inference, in which the trained and limited model of the neural network (TinyML) on the platform is embedded in the microcontroller, and inference is

performed through computer vision (camera) to recognize the clean photovoltaic cell (PV_Clean), with dirt (PV_Dirt) and with partial shading (PV_Shading).

Figure 4 shows the operating flow of the intelligence embedded in microcontrollers in the recognition of the three different classifications of photovoltaic cells of the system proposed in this work. The recognition begins with the visualization of the cell, through the camera, then the image processing is performed and will be used in the ML embedded in the edge to perform the recognition, according to the learning carried out in Edge Impulse. Finally, the response that can be according to the classification defined in the training, which in the case of this work, can be: shading (PV_Shading), dirt (PV_Dirt) or cleaning (PV_Clean).

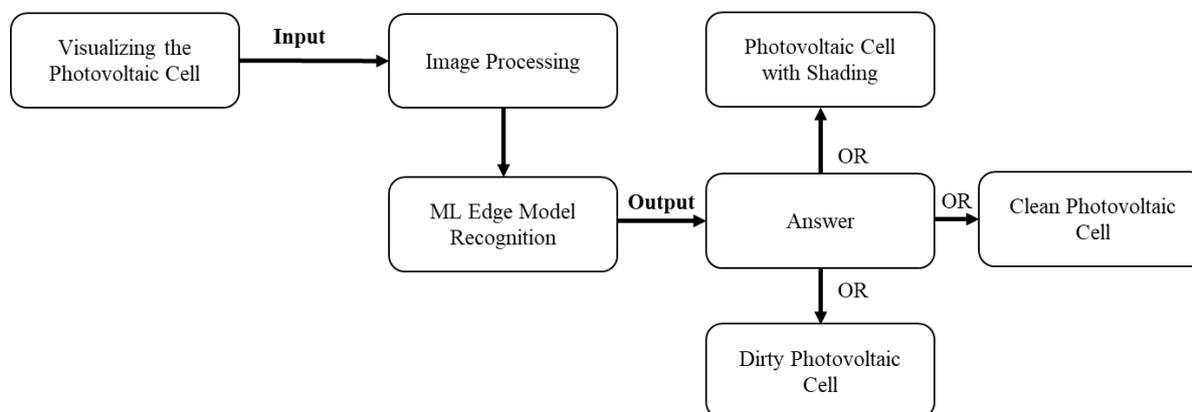


Figure 4. Flow of how the recognition of dirt, shading and cleaning of photovoltaic cells is performed by the microcontroller.

3. RESULTS AND DISCUSSION

Figure 5 shows the feature explorer of the three classes of photovoltaic cells before training. This analysis of the features of the dataset makes it possible to identify that part of the representative data for PV_Clean and PV_Dirt are in the same region, that is, mixed, while PV_Shading is more spread out. This may indicate the degree of difficulty in training the ML, as there is data that can be confused by the ML at the time of training and recognition.

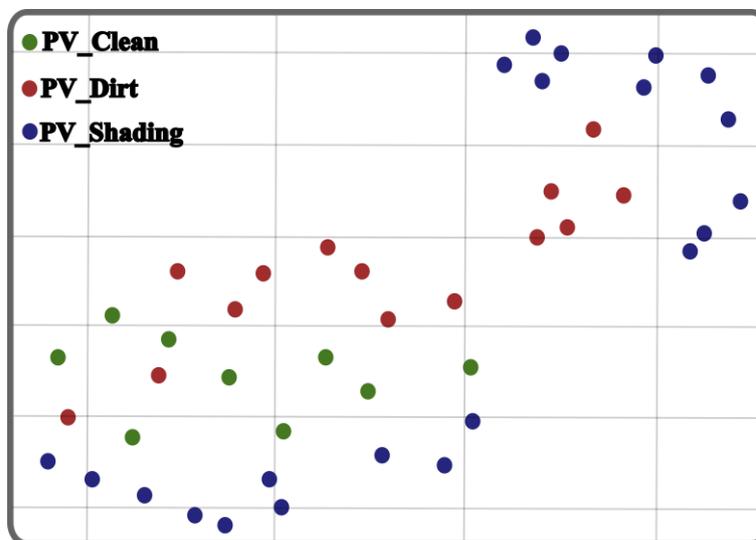


Figure 5. Feature explorer of classified dataset.

Post-training classification is presented in Figure 6, as can be seen, there was recognition of each class without errors, and the arrangement of the classes indicates that ML obtained a learning result that can distinguish even the similarities between the classes after training.

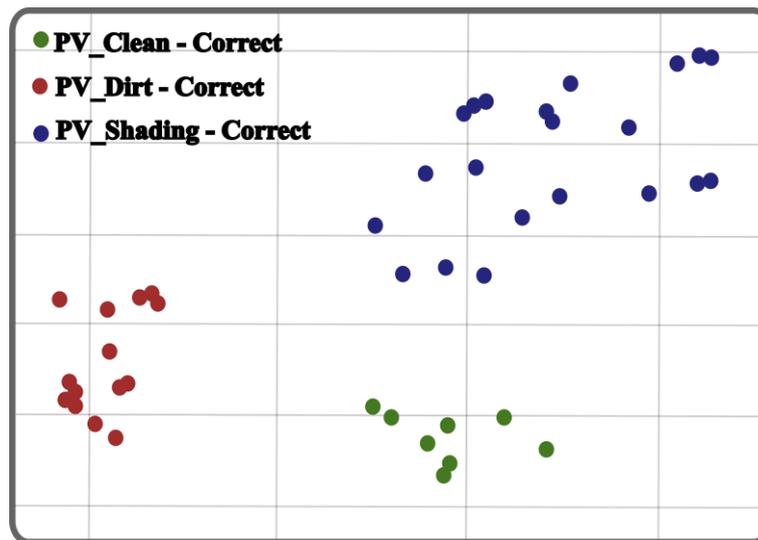


Figure 6. Classification of PV_Clean, PV_Dirt, and PV_Shading classes after training.

Figure 7 shows the confusion matrix of the trained model (the results are equivalent in the model test). It is possible to note the 100% accuracy of the model, so there are no false positives, false negatives, true positives, and true negatives. This result can be attributed to the fact that the images were collected in a controlled laboratory environment, without the variations of different environments, lighting conditions throughout the day, and cell installation angles, as well as the use of a limited number of images for each class.

	PV_Clean	PV_Dirt	PV_Shading
PV_Clean	100%	0%	0%
PV_Dirt	0%	100%	0%
PV_Shading	0%	0%	100%

Figure 7. Confusion matrix of the trained and tested model.

In order to present the results, the Receiver Operating Characteristic (ROC), Precision, Recall, and F1 Score metrics were considered and are presented below.

The ROC curve illustrates the relationship between the true positive rate (sensitivity) and the false positive rate (specificity) for varying classification thresholds. In this instance, the result was 1, indicating that only true positives were identified, and all were correctly classified. Precision is a metric that gauges the model's capacity to accurately distinguish between positive and negative examples. In this analysis, the precision value was 1, signifying a 100% precision rate with no classification errors. Recall is analogous to sensitivity and measures the capacity of the model to identify all positive examples, which also resulted in 1, indicating 100% capacity. While the F1-score deals with the harmonic mean between precision and recall, the resulting value of 1 indicates that the ROC, precision, recall, and F1-score all reached the maximum value for differentiating between the varying degrees of dirt, shading, and cleanliness observed in the cell. These results are for both training and testing, therefore for each class (PV_Clean, PV_Dirt, and PV_Shading) in the study.

In order to verify the recognition results of the three classes, an inference time of 3 milliseconds was observed for each, as demonstrated in Figure 8. This was achieved through the utilisation of the smartphone camera, which was employed for the recognition tests. On the other hand, the images used for inference were taken from Google searches of the classes.



Figure 8. Model inference test in recognizing clean (PV_Clean), dirty (PV_Dirt) and shading (PV_Shading) cells.

4. CONCLUSION

The study demonstrated the viability of edge computer vision, using the TinyML technique, as a solution for recognizing dirt on photovoltaic modules. This technology has the potential to optimize photovoltaic energy generation, reducing losses due to dirt and extending the useful life of the modules. Despite the 100% accuracy results, the study has some limitations such as the limited dataset, as the model was trained with a dataset with few images, which may affect its generalization to different types of dirt and lighting conditions. Another limiting factor is that the types of dirt were not considered, so it is necessary to evaluate the performance of the model with other types of dirt, such as sand, bird droppings, and leaves.

However, proposals for future work to overcome the limitations and improve the system are suggested as follows: collect a more comprehensive data set, including different types of dirt and lighting conditions, to improve the robustness and generalization of the model; implement a control system to integrate the dirt recognition system with an automated control system for cleaning the modules, triggering cleaning mechanisms when necessary; another suggestion is to develop an alert system that notifies users about the presence of dirt in the modules, allowing decision-making and corrective measures promptly. The research presented in this article can be shipped in drones, facilitating inspection in difficult places, such as coverage of buildings and large areas covered by photovoltaic plants.

The implementation of these proposals can contribute to the development of a more robust, efficient and reliable dirt recognition system in photovoltaic modules, optimizing solar energy generation and promoting sustainability.

5. AUTHORS' NOTE

The authors declare that there is no conflict of interest regarding the publication of this article. The authors confirmed that the paper was free of plagiarism.

7. AUTHOR'S CONTRIBUTION

Raimundo Eider Figueredo: Conceptualization, Methodology, Writing the Original Draft, Research; Nurhayati Nurhayati: Formal Analysis, Writing the Original Draft, Supervision; João Marcelo Vinicius de Paula, Marcel Veloso, João Lucas de Souza Silva, Arnaldo de Carvalho Jr., Alexandre Maniçoba de Oliveira, Auzuir Ripardo de Alexandria, and João Frederico Souza de Paula: Data Curation, Review and Editing, Research.

6. REFERENCES

- [1] Sistema de Informações de Geração ANEEL – SIGA (portuguese). Accessed on: <https://app.powerbi.com/view?r=eyJrIjoiNjc4OGYyYjQtYWM2ZC00YjllLWJlYmEtYzdkNTQ1MTc1NjM2IiwidCI6IjQwZDZmOWI4LWVjYTctNDZhMi05MmQ0LWVhNGU5YzAxNzBIMSIsImMiOjR9>. Accessed on: Oct 8, 2023.
- [2] P. G. V. Sampaio, M. O. A. González, “Photovoltaic solar energy: Conceptual framework”. *Renewable and Sustainable Energy Reviews*, v. 74, p. 590-601, 2017.
- [3] J. Mundo-Hernández *et al.*, “An overview of solar photovoltaic energy in Mexico and Germany”. *Renew Sustain Energy Rev* 2014; 31:639–49.
- [4] M. R. Maghami *et al.*, “Power loss due to soiling on solar panel: A review”. *Renewable and Sustainable Energy Reviews*, v. 59, p. 1307-1316, 2016.
- [5] G. Vedulla, A. Geetha, R. Senthil, “Review of strategies to mitigate dust deposition on solar photovoltaic systems”. *Energies*, v. 16, n. 1, p. 109, 2022.
- [6] M. Aghaei *et al.*, “Review of degradation and failure phenomena in photovoltaic modules”. *Renew. Sust. Energy Rev.* 2022, 159, 112160.
- [7] A. H. Alami *et al.*, “Management of potential challenges of PV technology proliferation”. *Sustain. Energy Technol. Assess.* 2022, 51, 101942. *Energies* 2023, 16, 109 22 of 28.
- [8] R. J. Mustafa *et al.*, “Environmental impacts on the performance of solar photovoltaic systems”. *Sustainability* 2020, 12, 608.
- [9] R. Kohli, K. L. Mittal, “Methods for Removal of Particle Contaminants. In *Developments in Surface Contamination and Cleaning*”, William Andrew Publishing: Norwich, UK, 2011; Volume 3.
- [10] A. Kimber *et al.*, "The effect of soiling on large grid-connected photovoltaic systems in California and the southwest region of the United States". In: 2006 *IEEE 4th World Conference on Photovoltaic Energy Conference*. IEEE, 2006. p. 2391-2395.
- [11] C. Deline *et al.*, "Estimating and parameterizing mismatch power loss in bifacial photovoltaic systems". *Progress in Photovoltaics: Research and Applications*, v. 28, n. 7, p. 691-703, 2020.
- [12] B. Sivakumar *et al.*, "Experimental study on the electrical performance of a solar photovoltaic panel by water immersion". *Environ. Sci. Pollut. Res.* 2021, 28, 42981–42989.
- [13] P. P. Ray, "A review on TinyML: State-of-the-art and prospects", *Journal of King Saud University-Computer and Information Sciences*, v. 34, n. 4, p. 1595-1623, 2022.
- [14] B. Taghezouit *et al.*, "Model-based fault detection in photovoltaic systems: A comprehensive review and avenues for enhancement". *Results in Engineering*, v. 21, p. 101835, 2024.
- [15] M. Köntges *et al.*, "Review of Failures of Photovoltaic Modules", *International Energy Agency - Photovoltaic Power Systems (IEA-PVPS)*, p 1-140, ISBN 978-3-906042-16-9, 2014.
- [16] A. Livera *et al.*, "Recent advances in failure diagnosis techniques based on performance data analysis for grid-connected photovoltaic systems", *Renewable energy*, v. 133, p. 126-143, 2019.

- [17] D. S. Pillai, N. Rajasekar, "A comprehensive review on protection challenges and fault diagnosis in PV systems". *Renewable and Sustainable Energy Reviews*, v. 91, p. 18-40, 2018.
- [18] R. Ghosh, S. Das, C. K. Panizrahi, "Classification of different types of faults in a photovoltaic system", In: 2018 *International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC)*. IEEE, 2018. p. 121-128.
- [19] TAKEUCHI, Renan de Oliveira Alves et al. "Physicochemical Evaluation of Dirt Deposited on Photovoltaic Modules Installed in Climatic Zones of Paraná, Brazil". *Revista Brasileira de Energia Solar*, v. 13, n. 1, p. 83-91, 2022.
- [20] V. Janapa Reddi *et al.*, "Edge impulse: An mlops platform for tiny machine learning. Proceedings of Machine Learning and Systems", *Proceedings of Machine Learning and Systems* , v. 5, p. 254-268, 2023.
- [21] R. Sanchez-Iborra, A. F. Skarmeta, "Tinym1-enabled frugal smart objects: Challenges and opportunities". *IEEE Circuits and Systems Magazine*, v. 20, n. 3, p. 4-18, 2020.
- [22] A. M. Arcanjo, A. L. C. Ottoni, "Analysis of Deep Learning Models for Embedded Systems applied to Crack Classification in Buildings". *XVI Brazilian Conference on Computational Intelligence (CBIC)*. 2023.