Integrated Thermal Monitoring System for Solar PV Panels: An Approach Based on TinyML and Edge Computing

Andrés David Suárez-Gómez*, Jorge Orlando Bareño Quintero

Universidad Nacional Abierta y a Distancia, Tunja, Colombia

Abstract

This paper presents an integrated system for thermal monitoring and anomaly detection of solar pv panels using TinyML and Edge Computing. The proposed system employs a low-resolution thermal sensor (MLX90640) in conjunction with embedded machine-learning techniques to perform early anomaly detection and preventive maintenance. This research seeks to address current challenges in the efficient management of photovoltaic installations by proposing a holistic solution that promises to significantly improve the performance and longevity of solar systems. The proposed approach aims to the processing of thermal data locally, reducing latency and improving energy efficiency. Four TinyML models were developed and compared using Edge Impulse, with the most successful model achieving 87.70% accuracy in anomaly detection. The study highlights the potential of this technology to improve the efficiency, reliability, and cost-effectiveness of PV installations, while also recognizing limitations and challenges in large-scale implementation. It highlights key areas for future studies, such as the integration of data from multiple sensors and the development of more advanced algorithms, highlighting the potential of this technology to drive the adoption of renewable energy and effectively combat climate change.

Keywords

TinyML, Solar PV panels, EdgeAI, Thermal Imaging

1. Introduction

Solar PV has established itself as one of the most promising and fast-growing renewable energy sources worldwide. Its ability to generate electricity in a clean and sustainable manner positions it as a key pillar in the transition to a greener energy future [1, 2]. However, the efficiency and longevity of photovoltaic (PV) systems are highly dependent on their proper operation and maintenance. One of the critical factors affecting the performance of solar panels is their operating temperature [3, 4].

High temperatures can significantly reduce the conversion efficiency and accelerate the degradation of PV modules. Recent studies have shown that thermal variations on the surface of solar panels can lead to the formation of hot spots, which not only decrease the overall system efficiency but can also cause irreversible damage to the modules [5, 6]. This thermal sensitivity not only affects short-term energy production but also has significant implications for the long-term degradation of PV systems [7, 8].

In this context, accurate and real-time thermal monitoring of solar panels becomes crucial to optimize system performance, prevent failures, and extend the lifetime of PV installations [4, 9]. However, traditional monitoring methods have several limitations:

1. High costs: Manual or drone-based thermographic inspections require costly equipment and specialized personnel [10, 11].
2. Low monitoring frequency: Due to logistical and economic constraints, these inspections are often performed sporadically, which can result in late detection of problems [2, 9].
3. Centralized processing: Conventional monitoring systems often rely on continuous data transmission to central servers, which can result in high communication costs and latency in anomaly detection [9, 12].

ICAIW 2024: Workshops at the 7th International Conference on Applied Informatics 2024, October 24–26, 2024, Viña del Mar, Chile
*Corresponding author.
andresd.suarez@unad.edu.co (A. D. Suárez-Gómez)
0000-0002-3576-4428 (A. D. Suárez-Gómez); 0000-0002-4821-8194 (J. O. Bareño Quintero)

4. Limited scalability: As solar plants grow, centralized approaches face challenges in terms of data processing and management of the communication infrastructure [13, 14].

The advent of Machine Learning (ML) and Edge Computing technologies offers new possibilities to address these challenges [15, 16]. TinyML, a branch of ML designed for resource-constrained devices, allows artificial intelligence algorithms to be implemented directly on low-cost microcontrollers and sensors [17, 18]. This capability, combined with the advantages of edge computing, such as reduced latency and improved data privacy, presents a unique opportunity to revolutionize the monitoring and optimization of solar systems [14, 19].

The integration of TinyML into solar panel monitoring systems offers several significant advantages:

1. Local processing: by running ML algorithms directly on the sensing devices, the amount of data that needs to be transmitted is drastically reduced, resulting in lower latency and more efficient use of bandwidth [20, 21].
2. Energy efficiency: TinyML models are optimized to run on low-power devices, enabling energy-efficient, stand-alone monitoring systems to be implemented [22, 23].
3. Enhanced privacy and security: Sensitive data is processed locally, reducing the risks associated with transmission and storage in the cloud.
4. Offline operation: Systems can operate autonomously, even in the absence of internet connectivity.
5. Reduced operational costs: Minimizes the need for network infrastructure and cloud storage.
6. Adaptability: TinyML models can be updated and adapted to changing PV system conditions, improving the accuracy of anomaly detection over time [23, 24].
7. Scalability: The distributed architecture inherent in TinyML-based systems allows for easy expansion as PV installations grow [13, 25]

However, the implementation of TinyML in solar thermal monitoring systems also presents unique challenges. The limited processing and memory capacity of edge devices requires careful optimization of ML models [26, 27]. In addition, the variability in environmental and operating conditions of PV systems requires robust and adaptive models [28, 29].

In this context, this paper presents a thermal monitoring system for solar panels based on TinyML and Edge Computing. The proposed system uses low-cost thermal sensors and microcontrollers with ML capabilities to perform real-time analysis of the thermal distribution of solar panels [12, 30]. By using platforms such as Edge Impulse for the development and deployment of ML models, this research contributes to the emerging field of TinyML application in renewable energy systems, proposing an innovative solution that combines the accuracy of machine learning with the efficiency and scalability of edge computing to significantly improve the maintenance, performance, and longevity of PV installations.

The rest of the article is organized as follows: Section 2 presents a comprehensive review of the state-of-the-art in thermal monitoring of solar panels, ML applications in PV systems, and advances in TinyML and Edge Computing. Section 3 describes in detail the design and architecture of the proposed system. Section 4 focuses on the development and optimization of the TinyML model. Section 5 presents the experimental results and a discussion of these. Finally, Section 6 concludes the paper and proposes future research directions.

2. State of the art

2.1. Thermal Monitoring of Solar Panels

Thermal monitoring of solar panels has been the subject of numerous studies in recent decades, given its importance for the performance and reliability of PV systems. Buerhop et al. [2] provide a comprehensive review of infrared imaging techniques for the inspection of PV modules. The authors highlight the importance of efficient measurement strategies for large solar plants and discuss the challenges in assessing thermal anomalies. Traditional thermal monitoring techniques include:

1. Hand-held thermographic cameras: used for periodic inspections, they offer high resolution but require time and skilled personnel [11].
2. Drone-mounted thermal imaging systems: Enable faster inspections of large areas, but face challenges in terms of flight regulations and processing of large volumes of data [10].
3. Fixed thermal sensors: Permanently installed on selected panels, they provide continuous monitoring but with limited coverage [12].

Demir et al. [10] demonstrated the effectiveness of using drones equipped with thermal cameras for the detection and diagnosis of faults in photovoltaic systems. Their approach, based on machine learning, achieved high accuracy in identifying various types of defects in solar panels.

Recent advances in thermal image processing have led to the development of automatic anomaly detection techniques. For example, Oulefki et al. [31] proposed an approach based on unsupervised detection algorithms and 3D augmented reality to identify damaged areas in PV modules. Such approaches improve the objectivity and efficiency of thermal data interpretation.

Wang et al. [9] developed an automatic anomaly detection system for PV systems using thermographic imaging and low-rank matrix decomposition. Their method proved to be effective for the online detection of various types of faults in solar panels.

2.2. Machine Learning Applications in Photovoltaic Systems

Machine Learning has gained ground in various applications related to photovoltaic systems, from energy production prediction to fault detection and performance optimization.

2.2.1. Fault Detection and Diagnosis

Mellit et al. [32] demonstrated the use of TinyML for PV module fault diagnosis using the Edge Impulse platform. Their approach achieved high accuracy in classifying common faults such as hot spots, cracks, and encapsulant degradation.

Jaybhaye et al. [8] proposed a method for solar panel damage detection and localization using thermal imaging and image processing techniques. Their approach combined thermal image analysis with segmentation algorithms to accurately identify and locate damaged areas on solar panels.

Pamungkas et al. [28] introduced a novel approach for efficient solar panel fault classification using a deep neural network architecture called coupled UDenseNet. Their method demonstrated a significant improvement in classification accuracy compared to other deep learning models.

Hassan and Dhimish [7] conducted a comprehensive review of convolutional neural network (CNN) based approaches for crack detection in photovoltaic modules. Their study highlighted the effectiveness of CNNs in identifying various types of defects in solar panels and discussed current trends and future directions in this field.

2.2.2. Performance Prediction and Optimisation

Bhattacharya and Pandey [33] developed a TinyML model optimized for soil quality monitoring and management in agriculture, which could be adapted for applications in photovoltaic systems. Their approach, based on sidechain and energy efficiency testing, demonstrated significant improvements in power consumption and latency.

Hayajneh et al. [34] investigated the role of TinyML in improving solar energy yield predictions. Their study compared several modern machine learning models and highlighted the potential of TinyML to provide intelligent and efficient forecasts in solar energy systems.

2.2.3. Real-Time Monitoring and Control

Cardinale-Villalobos et al. [6] presented an artificial intelligence-based IoT system for hot spot detection in PV modules. Their approach, which operates over a wide range of irradiances, proved to be effective for real-time monitoring and early detection of thermal anomalies.

Hidalgo et al. [13] proposed an irrigation control system based on TinyML and Edge Computing for smart agriculture scenarios. Although their application is focused on agriculture, the approach of using TinyML for real-time control is highly relevant for the optimization of photovoltaic systems.

2.3. TinyML and Edge Computing

TinyML has emerged as a promising technology for implementing ML algorithms on resource-constrained devices. The papers [19, 35], provide an overview of the challenges and opportunities in the field of TinyML, highlighting the importance of several topics like model optimization and energy efficiency.

2.3.1. Platforms and Tools

Several platforms and tools have emerged to facilitate the development and deployment of TinyML models:

1. Edge Impulse: An end-to-end platform for the development and deployment of TinyML models, offering tools for data collection, model training, and optimized code generation [32].
2. TensorFlow Lite for Microcontrollers: An optimized version of TensorFlow designed specifically for embedded systems [17, 18].
3. Arduino TinyML Kit: A toolkit that facilitates the implementation of TinyML models on Arduino boards [20, 30].

2.3.2. Model Optimization

Model optimization is crucial for the effective deployment of TinyML. Common techniques include:

1. Quantization: Reducing the precision of model weights from floating point to integers, resulting in smaller, computationally efficient models [27].
2. Pruning: Removal of redundant or unimportant connections and neurons in neural networks [27, 36].
3. Model compression: Techniques to reduce model size without significantly sacrificing accuracy [37, 38].

Liu et al. [38] proposed TinyTS, a memory-efficient TinyML model compilation framework for microcontrollers. Their approach demonstrated significant improvements in memory usage and execution time for machine learning models on resource-constrained devices.

2.3.3. Applications in Solar Energy and Related Fields

In the context of solar applications, Gruosso and Gajani [39] performed a comparison of ML algorithms for performance evaluation of PV power forecasting and management in the TinyML framework. Their results demonstrated the feasibility of implementing complex prediction models on edge devices.

Oliveira and Moreira [11] developed an edge AI system using a thermal camera for industrial anomaly detection. Although their application is not specifically focused on solar panels, the techniques used for thermal image processing and anomaly detection are highly relevant to our field of study.

Wardana et al. [30] demonstrated the application of TinyML models for low-cost air quality monitoring devices. Their approach of using low-cost sensors combined with optimized ML models is directly applicable to solar panel monitoring.

2.4. Challenges and Opportunities

Despite significant advances in thermal monitoring of solar panels and ML applications in photovoltaic systems, several challenges remain:

1. Balance between accuracy and efficiency: Implementing complex ML models on resource-constrained devices requires a careful balance between model accuracy and computational efficiency [15, 23].
2. Adaptability to varying conditions: PV systems operate in dynamic environments with seasonal and daily variations. ML models must be able to adapt to these changing conditions [29, 34].
3. Integration of multiple data sources: Combining thermal data with other sources (electrical, meteorological, etc.) can improve the accuracy of predictions and optimizations but poses challenges in terms of data fusion and efficient processing [6, 9].
4. Scalability and maintenance: As solar plants grow, in size and complexity, scalability of monitoring systems and efficient management of large fleets of edge devices become critical [13, 14].
5. Security and privacy: The implementation of ML on edge devices poses new challenges in terms of data security and protection against malicious attacks [17, 18].

These limitations and challenges present significant opportunities for research and development of innovative solutions that combine advances in TinyML, edge computing, and sensor technologies to create more efficient, scalable, and adaptive monitoring and optimization systems for PV installations [24, 35].

3. System Design

3.1. General Architecture

The proposed system is based on a three-tier architecture that combines the capabilities of TinyML and Edge Computing to provide a scalable and efficient solution for the thermal monitoring of solar panels. This architecture is specifically tailored to the needs of photovoltaic systems. The main components of the system are:

1. 100 W monocrystalline and polycrystalline solar panels 2.
2. MLX90640 sensor.
3. Microcontrollers with TinyML capabilities such as Arduino Nano 33 BLE Sense.
4. Additional sensors for ambient temperature, humidity, and solar radiation.
5. Local server based on Raspberry Pi 5 for data visualization and sending to the cloud.

3.2. Sensor level

In this initial testing phase, we use a public dataset that simulates the data that would be captured by thermal sensors in real solar panels. This approach allows us to validate the effectiveness of the system before its implementation on real hardware. The dataset [40] is selected because it is the one with twenty thousand data classified into 12 groups. Ten thousand data correspond to panels in normal conditions and the other ten thousand data are divided into anomalies such as hot spots, shadows, ruptures, etc. The files are in JPG format and have a resolution of 24x40 pixels, which is close to the 24x32 pixels of the MLX90640.

3.3. Microcontroller level

The model used in the Edge Impulse software is set for an Arduino Nano 33 BLE Sense. This device is fully supported in the software, is low cost and its processing capabilities, power consumption and size make it an ideal device for our application.

4. Development of the TinyML Model

4.1. Data Collection and Preparation

For this initial phase and as mentioned above, a public dataset [40] containing thermal measurements of solar panels in different conditions and with twelve labels or classes is selected. A description of each class is given in Table 1. Figure 1 shows random data for each class.

Table 1
Description of the dataset classes

Class Name	Description
Cell	Hot spot occurring with a square geometry in a single cell.
Cell-Multi	Hot spots occurring with a square geometry in multiple cells.
Cracking	Module anomaly caused by cracking on the module surface.
Hot-Spot	Hot Spot on a thin film module.
Hot-Spot-Multi	Multiple hot spots on a thin film module.
Shadowing	Sunlight obstructed by vegetation, man-made structures, or adjacent rows.
Diode	Activated bypass diode, typically 1/3 of the module.
Diode-Multi	Multiple activated bypass diodes, typically affecting 2/3 of the module.
Vegetation	Panels blocked by vegetation.
Soiling	Dirt, dust, or other debris on the surface of the module.
Offline-Module	Entire module is heated.
No-Anomaly	Nominal solar module.

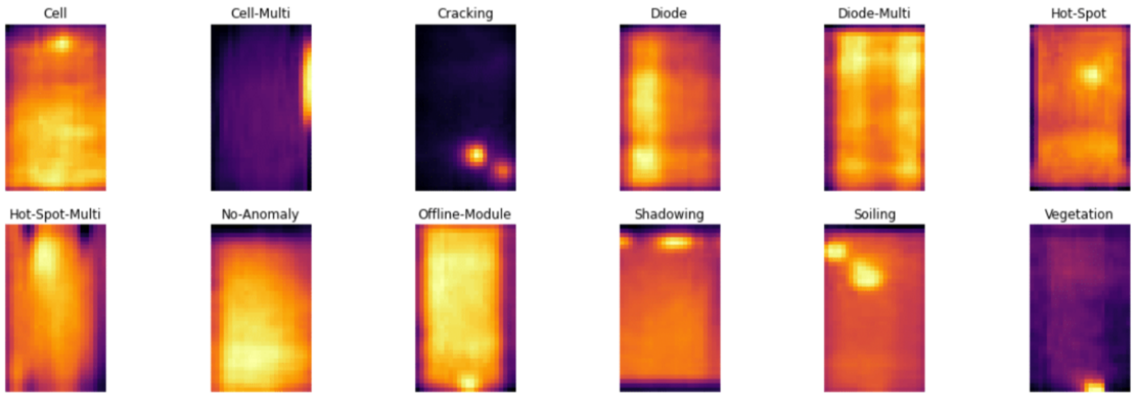

Figure 1: Example of data for each of the labels of the selected dataset.

Before using the dataset in Edge Impulse, preparation is made because the images are not classified in different folders, all the label information for each image is in a single JSON file. Taking this into account, a Python script is created to create a folder for each label and include the corresponding images. This makes uploading data to Edge Impulse much easier. By doing this preparation, the amount of data for each label is found (see Figure 2).

According to Figure 2, the No-Anomaly label has half of the data, while the other eleven labels contain the anomalies. A large imbalance can be observed in the dataset for the amount of data in each label.

4.2. Model development in Edge Impulse

Edge Impulse is used for the development and training of the TinyML model, taking advantage of its optimization capabilities for edge devices. The process includes:

1. Importing the prepared dataset into Edge Impulse.
2. Separating the training and test data, using a ratio of 80/20 for each label.

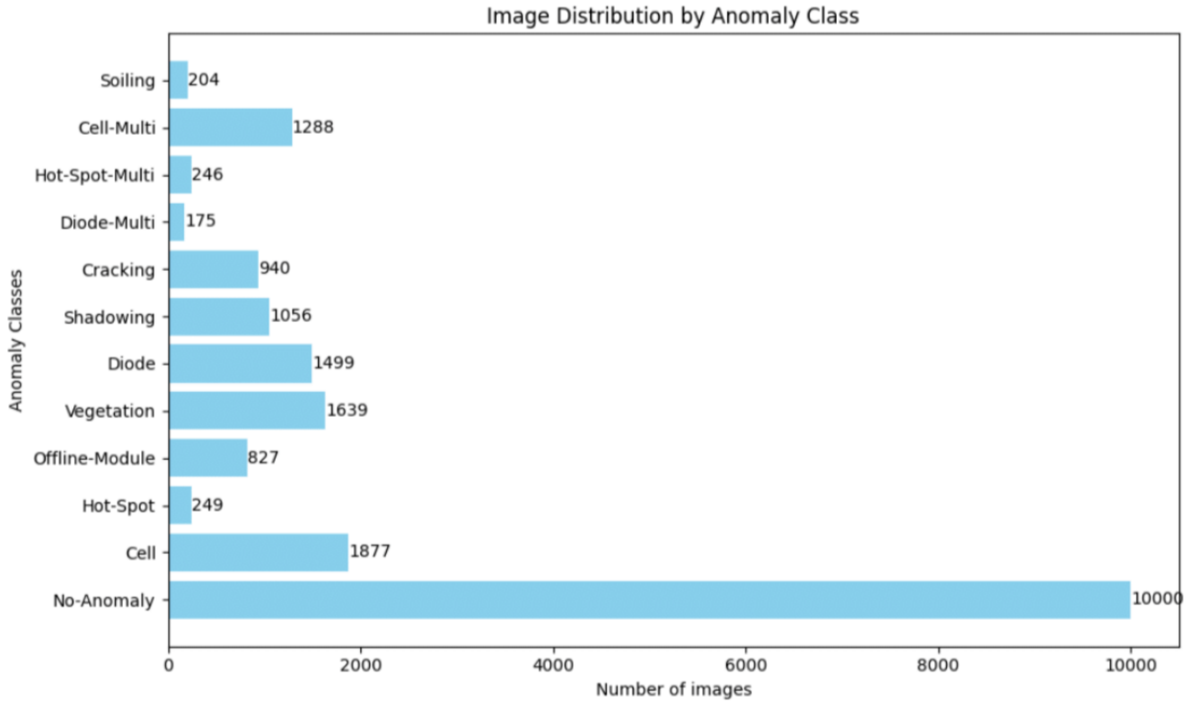

Figure 2: Distribution of the data of each label in the selected dataset.

3. Design of the model architecture, the classification block is used.
4. Model training.
5. Model evaluation by analyzing metrics such as accuracy, recall, and F1-score.

4.3. Realization of different models in Edge Impulse

Considering the characteristics of the selected dataset and the imbalance between the amount of data in the twelve labels, it was decided to develop four models in Edge Impulse and perform the respective comparison to determine which one best fits the application developed in this paper. The models developed are the following:

1. Model 1a: Model using all twelve labels, 20 training cycles, and the predetermined architecture for the neural network. The predetermined architecture consists of a 2D conv/pool layer (16 filters, 3 kernel size, 1 layer), a 2D conv/pool layer (32 filters, 3 kernel size, 1 layer), a flattened layer, and a Dropout (rate 0.25).
2. Model 1b: Model using all twelve labels, 30 training cycles and with the following architecture: a 2D conv/pool layer (16 filters, 3 kernel size, 1 layer), a 2D conv/pool layer (32 filters, 3 kernel size, 1 layer), a 2D conv/pool layer (64 filters, 3 kernel size, 2 layers), a 2D conv/pool layer (128 filters, 3 kernel size, 2 layers), a Flatten layer, a Dense layer (128 neurons), a Dropout (rate 0.25) and a Dense layer (12 neurons).
3. Model 2a: Model using only two labels, normal and abnormal, using the No-Anomaly label as normal and the other eleven labels as abnormal. This model uses 20 training cycles and with the same architecture as Model 1a.
4. Model 2b: Model using only two labels, same as Model 2a. This model uses 20 training cycles and with the following architecture: a 2D conv/pool layer (16 filters, 3 kernel size, 1 layer), a 2D conv/pool layer (32 filters, 3 kernel size, 1 layer), a 2D conv/pool layer (64 filters, 3 kernel size, 2 layers), a 2D conv/pool layer (128 filters, 3 kernel size, 2 layers), a Flatten layer, a Dense layer (128 neurons), a Dropout (rate 0.25) and a Dense layer (2 neurons).

5. Results and Discussion

The results of the Edge Impulse software for Model 2b, which uses only the Normal and Abnormal classes and the custom architecture, are presented below. Figure 3 shows some of the most relevant data produced by the Edge Impulse software for Model 2b. The results of the other models follow the same structure.

5.1. Model Performance

Evaluating the performance of the four TinyML models developed in Edge Impulse, the following is obtained:

Table 2
Global metrics of the four developed models

Metric	Model 1a	Model 1b	Model 2a	Model 2b
Accuracy (float32)	55.7%	69.0%	75.3%	87.7%
Accuracy (int8)	55.5%	68.8%	74.7%	87.7%
ROC AUC	0.81	0.90	0.75	0.88
Loss (float32)	1.42	0.99	0.52	0.33
Loss (int8)	1.43	1.01	0.53	0.34

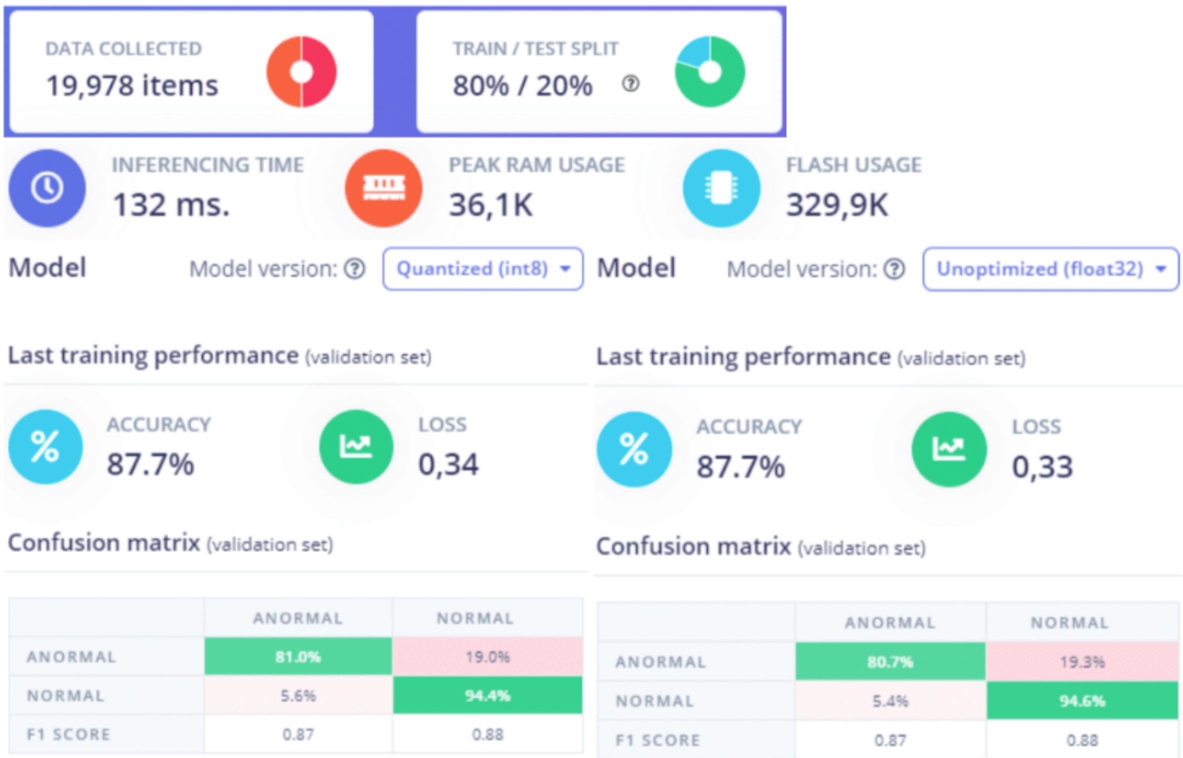

Figure 3: Snippets of data provided by Edge Impulse software for Model 2b.

5.2. Computational Efficiency

Analyzing the computational efficiency of the model, the following is determined:

Table 3
Computational metrics of the four developed models for the Quantized (int8) model

Metric	Model 1a	Model 1b	Model 2a	Model 2b
Inferencing Time	101 ms	190 ms	69 ms	132 ms
Peak RAM Usage	21.6 K	36,1 K	21.6 K	36.1 K
Flash Usage	57.8 K	331,3 K	39.0 K	329.9 K

5.3. Results Interpretation

According to the data shown in Tables 2 and 3, the results can be interpreted as follows:

- The best performing TinyML model (Model 2b) achieved an accuracy of 87.70% in detecting thermal anomalies. To enhance this accuracy, future work could explore techniques such as data augmentation for underrepresented anomaly classes, implementing ensemble learning methods, or applying advanced model optimization techniques. Also, data augmentation could greatly improve the performance of Models 1a and 1b.
- The float32 and int8 models have similar accuracy, which demonstrates the feasibility of using Edge devices in this type of application.
- Models 2a and 2b have a lower loss, which is expected since it is a simpler problem with only two classes.
- There is a trade-off between the accuracy, which was achieved with a more complex architecture, and the Computational metrics. This is expected, as a more complex architecture requires more memory and processing power. However, this trade-off is not as significant as expected, with about two times the amount of Inferencing Time and six to 9 times the amount of flash usage.
- Running ML algorithms directly on edge devices allows for more detailed, real-time analysis of thermal data. As noted by Pamungkas et al [29], early and accurate fault detection can significantly reduce downtime and associated maintenance costs. In addition, the ability to distinguish between different types of thermal anomalies, such as hot spots, cracks, or encapsulant degradation, allows for a more targeted and effective response to each problem [7, 8]
- By performing TinyML model inference directly on edge devices, the need for data transmission is drastically reduced. This approach not only saves energy in transmission but also reduces the load on the network infrastructure.
- Quantization techniques significantly reduced the computational requirements of the model. This optimization is crucial for implementation on resource-constrained microcontrollers.
- The energy efficiency achieved not only reduces operating costs but also opens the possibility of implementing stand-alone solar-powered monitoring systems. This feature is especially valuable for PV installations in remote areas or areas with limited access to the grid.
- The ability to respond quickly to thermal anomalies can prevent the formation of hot spots and associated damage to PV modules. As noted by Kirubakaran et al [5], rapid identification and mitigation of hot spots are crucial to prevent irreversible damage to solar panels.
- Low latency allows dynamic adjustments in PV system operation, maximizing efficiency in different environmental conditions.

5.4. Constraints and Challenges

Despite the promising results, it is important to recognize several limitations and challenges of our study:

5.4.1. Validation with Real Data

The use of a public dataset, while useful for the initial validation of the concept, presents significant limitations that must be addressed in future research:

- **Representativeness:** The current dataset, though diverse, may not fully capture the complexity and variability of thermal conditions in solar panels across different real-world scenarios. Real PV installations are subject to a wide range of environmental factors, installation configurations, and operational conditions that may not be adequately represented in the public dataset.
- **Parameter mismatch:** The data used, although close, does not have the exact pixel size of the data to be captured with the MLX90640 thermal sensor. This discrepancy could lead to potential inaccuracies when transitioning from the model trained on the dataset to real-world applications.
- **Lack of temporal data:** The current dataset likely consists of static images, whereas real-world thermal patterns in solar panels evolve over time. This temporal aspect is crucial for understanding the progression of anomalies and for developing more accurate predictive models.

There is a critical need for validation with real data to address these limitations. A thorough validation with real field data is required to confirm the effectiveness of the system under actual operating conditions. This involves collecting thermal data from a diverse range of operational solar installations and capturing data across different times of day, seasons, and weather conditions. It is imperative to establish a protocol for continuous, long-term data collection from multiple PV installations for the analysis of thermal patterns over extended periods, the study of the relationship between thermal anomalies and long-term panel degradation, and the development of more robust and adaptable models. Also, it is important to validate the thermal anomalies detected by the system against actual, physically confirmed faults in solar panels. Furthermore, it is necessary to conduct extensive testing with the actual MLX90640 sensor to be used in the final system, ensuring that the model's performance translates accurately to the specific characteristics of this sensor and that any discrepancies between the training data and real sensor output are identified and addressed. Finally, regarding scalability, it is important to implement the system across multiple panels and arrays to identify any unforeseen challenges in managing a network of edge devices.

5.4.2. Hardware Implementation Challenges

Although our study focused purely on the use of Edge Impulse, implementation on real hardware presents several challenges:

- **Resource constraints:** microcontrollers have significant limitations in terms of memory and processing power. On a positive note, Edge Impulse is fully integrated with the Arduino device to be used, so estimates are very close to reality.
- **Reliability and durability:** Devices deployed in the field will be exposed to harsh environmental conditions. Ensuring their long-term reliability and durability represents a significant challenge.

5.4.3. Scalability Challenges

Large-scale implementation of the system presents additional challenges:

- **Device management:** Managing a large fleet of edge devices presents logistical and technical challenges, especially in terms of firmware and model updates.
- **Data aggregation:** Effectively integrating data from multiple devices for plant-level analytics requires advanced data aggregation and fusion strategies.
- **Security and privacy:** The distributed nature of the system poses challenges in terms of data security and protection against malicious attacks.

5.5. Practical Implications and Future Directions

5.5.1. Improving the Efficiency and Longevity of Photovoltaic Installations

The proposed system has the potential to significantly improve the efficiency and longevity of PV installations:

- **Early failure detection:** The ability to identify thermal anomalies at early stages can prevent further damage and extend the lifetime of solar panels. This capability is particularly valuable considering the impact of temperature on panel performance.
- **Dynamic optimization:** Continuous, real-time monitoring allows for dynamic adjustments in system operation, maximizing efficiency in different environmental conditions.
- **Predictive maintenance:** The ability to predict failures more accurately and in advance allows for more efficient maintenance planning, reducing downtime and associated costs.

5.5.2. Scalability and Adaptability

The system design, based on Edge Computing and TinyML principles, offers significant advantages in terms of scalability and adaptability:

- **Flexible deployment:** The distributed architecture allows for flexible and scalable deployment, suitable for both small installations and large solar plants.
- **Adaptation to local conditions:** The ability to retrain models locally allows adaptation to site-specific conditions, improving the accuracy and relevance of detections.
- **Integration with existing systems:** The system can be integrated with existing monitoring and control infrastructures, providing an additional layer of intelligence and optimization.

5.5.3. Future Research Endearvors

Based on the identified results and limitations, we propose the following directions for future research:

- **Validation with Real Data:**Conduct extensive field testing with real thermal data from solar panels under various operating and environmental conditions. This is thoroughly discussed on section 5.4.1.
- **Integration of Multiple Data Sources:**Expand the model to incorporate meteorological data (e.g., ambient temperature, humidity, solar irradiance) to improve contextual understanding of thermal patterns. Integrate electrical performance data (e.g., voltage, current, power output) to correlate thermal anomalies with electrical behavior. Explore the inclusion of visual inspection data to complement thermal analysis, potentially using multi-modal machine learning approaches.
- **Developing More Advanced Models:**Explore techniques for continual learning and model adaptation to allow the system to improve its performance over time without requiring complete retraining. Develop ensemble methods that combine multiple lightweight models to improve overall accuracy and robustness.
- **Scalability and Network Optimization:**Research efficient methods for managing and updating large fleets of edge devices in distributed solar installations. Develop advanced data aggregation and compression techniques to minimize bandwidth usage while maintaining high-fidelity analytics at the system level. Investigate peer-to-peer communication protocols that allow edge devices to share insights and anomaly detections without relying on central servers.
- **Enhanced Anomaly Detection and Classification:**Develop more granular classification models that can distinguish between different types of thermal anomalies (e.g., hot spots, bypass diode failures, cell cracks) with high accuracy. Explore unsupervised and semi-supervised learning techniques to identify novel or previously unseen types of thermal anomalies.
- **Long-Term Reliability and Degradation Studies:** Conduct studies to evaluate the long-term impact of using TinyML-based thermal monitoring on PV system performance and longevity. Develop models that can predict long-term degradation patterns based on thermal and operational data collected over extended periods. Investigate the reliability and durability of edge devices and sensors in harsh environmental conditions typical of solar installations.

- **Economic and Environmental Impact Analysis:**Conduct comprehensive cost-benefit analyses of implementing TinyML-based monitoring systems at various scales of PV installations. Evaluate the potential reduction in carbon footprint achieved through improved PV system efficiency and reduced maintenance requirements. Explore the broader implications of widespread adoption of this technology on the solar energy industry and renewable energy adoption rates.

6. Conclusion

In this paper, we establish the significant potential of thermal monitoring systems based on TinyML and Edge Computing to transform the operation and maintenance of PV installations. The results obtained in terms of anomaly detection accuracy are promising and suggest that this approach can effectively address many of the current challenges in the solar industry. However, it is crucial to recognize the limitations of our study, particularly regarding the use of public datasets and the need for validation under real operating conditions. Future research should address these aspects, as well as explore proposed directions for system development and refinement. Successful implementation of this technology has the potential to significantly improve the efficiency, reliability, and cost-effectiveness of PV installations, thus contributing to the acceleration of the global transition to renewable energy sources. However, it is important to carefully address ethical and social considerations to ensure that the benefits of this technology are distributed in an equitable and sustainable manner. Ultimately, the continued development and adoption of advanced technologies such as TinyML in the solar energy sector not only promises technical and economic improvements but can also play a crucial role in combating climate change and building a more sustainable energy future.

7. Acknowledgements

This work is supported by the Universidad Nacional Abierta y a Distancia through project grant PGDT4602ECBTI2024.

References

[1] G. F. Al-Doori, R. A. Mahmood, A. Al-Janabi, A. M. Hassan, G. T. Chala, Impact of surface temperature of a photovoltaic solar panel on voltage production, in: Energy and Environment in the Tropics, Springer, 2022, pp. 81–93.

[2] C. Buerhop, L. Bommes, J. Schlipf, T. Pickel, A. Fladung, I. M. Peters, Infrared imaging of photovoltaic modules: a review of the state of the art and future challenges facing gigawatt photovoltaic power stations, Progress in Energy 4 (2022) 042010.

[3] W. Ahmed, A. Hanif, K. D. Kallu, A. Z. Kouzani, M. U. Ali, A. Zafar, Photovoltaic panels classification using isolated and transfer learned deep neural models using infrared thermographic images, Sensors 21 (2021) 5668.

[4] S. P. Pathak, S. Patil, S. Patel, Solar panel hotspot localization and fault classification using deep learning approach, Procedia Computer Science 204 (2022) 698–705.

[5] V. Kirubakaran, D. Preethi, U. Arunachalam, Y. K. Rao, M. K. Gatasheh, N. Hoda, E. M. Anbese, Infrared thermal images of solar pv panels for fault identification using image processing technique, International Journal of Photoenergy 2022 (2022) 6427076.

[6] L. Cardinale-Villalobos, E. Jimenez-Delgado, Y. García-Ramírez, L. Araya-Solano, L. A. Solís-García, A. Méndez-Porras, J. Alfaro-Velasco, Iot system based on artificial intelligence for hot spot detection in photovoltaic modules for a wide range of irradiances, Sensors 23 (2023) 6749.

[7] S. Hassan, M. Dhimish, A survey of cnn-based approaches for crack detection in solar pv modules: Current trends and future directions, 2023.

[8] S. Jaybhaye, O. Thakur, R. Yardi, V. Raut, A. Raut, Solar panel damage detection and localization of thermal images, Journal of Failure Analysis and Prevention 23 (2023) 1980–1990.

[9] Q. Wang, K. Paynabar, M. Pacella, Online automatic anomaly detection for photovoltaic systems using thermography imaging and low rank matrix decomposition, Journal of Quality Technology 54 (2022) 503–516.

[10] B. Kaycı, B. E. Demir, F. Demir, Deep learning based fault detection and diagnosis in photovoltaic system using thermal images acquired by uav, Politeknik Dergisi 27 (2023) 91–99.

[11] V. M. Oliveira, A. H. Moreira, Edge ai system using a thermal camera for industrial anomaly detection, in: International Summit Smart City 360°, Springer, 2021, pp. 172–187.

[12] V. Chidurala, X. Li, Detection of moving objects using thermal imaging sensors for occupancy estimation, Internet of Things 17 (2022) 100487.

[13] C. H. Hidalgo, A. González-Vidal, A. F. Skarmeta, Edge computing in smart agriculture scenario based on tinyml for irrigation control, in: 2023 IEEE 9th World Forum on Internet of Things (WF-IoT), IEEE, 2023, pp. 01–08.

[14] V. Tsoukas, A. Gkogkidis, A. Kakarountas, Internet of things challenges and the emerging technology of tinyml, in: 2023 19th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT), IEEE, 2023, pp. 491–495.

[15] N. Tekin, A. Aris, A. Acar, S. Uluagac, V. C. Gungor, A review of on-device machine learning for iot: An energy perspective, Ad Hoc Networks (2023) 103348.

[16] P. P. Ray, A review on tinyml: State-of-the-art and prospects, Journal of King Saud University-Computer and Information Sciences 34 (2022) 1595–1623.

[17] Y. Abadade, A. Temouden, H. Bamoumen, N. Benamar, Y. Chtouki, A. S. Hafid, A comprehensive survey on tinyml, IEEE Access (2023).

[18] L. Dutta, S. Bharali, Tinyml meets iot: A comprehensive survey, Internet of Things 16 (2021) 100461.

[19] M. Shafique, T. Theocharides, V. J. Reddy, B. Murmann, Tinyml: Current progress, research challenges, and future roadmap, in: 2021 58th ACM/IEEE Design Automation Conference (DAC), IEEE, 2021, pp. 1303–1306.

[20] J. Botero-Valencia, C. Barrantes-Toro, D. Marquez-Viloria, J. M. Pearce, Low-cost air, noise, and light pollution measuring station with wireless communication and tinyml, HardwareX 16 (2023) e00477.

[21] T. Medeiros, M. Amaral, M. Targino, M. Silva, I. Silva, E. Sisinni, P. Ferrari, Tinyml custom ai algorithms for low-power iot data compression: A bridge monitoring case study, in: 2023 IEEE International Workshop on Metrology for Industry 4.0 & IoT (MetroInd4. 0&IoT), IEEE, 2023, pp. 54–59.

[22] Q. Sun, C.-H. Tu, Energy aware collaborative machine learning on energy-harvesting devices, in: 2023 International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan), IEEE, 2023, pp. 179–180.

[23] R. Kallimani, K. Pai, P. Raghuwanshi, S. Iyer, O. L. López, Tinyml: Tools, applications, challenges, and future research directions, Multimedia Tools and Applications 83 (2024) 29015–29045.

[24] D. A. N. Gookyi, F. A. Wulnye, E. A. E. Arthur, R. K. Ahiadormey, J. O. Agyemang, K. O.-B. O. Agyekum, R. Gyaang, Tinyml for smart agriculture: Comparative analysis of tinyml platforms and practical deployment for maize leaf disease identification, Smart Agricultural Technology (2024) 100490.

[25] S. B. Lakshman, N. U. Eisty, Software engineering approaches for tinyml based iot embedded vision: A systematic literature review, in: Proceedings of the 4th International Workshop on Software Engineering Research and Practice for the IoT, 2022, pp. 33–40.

[26] J. Shiraishi, M. Thorsager, S. R. Pandey, P. Popovski, Tinyairnet: Tinyml model transmission for energy-efficient image retrieval from iot devices, IEEE Communications Letters (2024).

[27] Y.-Y. Liu, H.-S. Zheng, Y. F. Hu, C.-F. Hsu, T. T. Yeh, Tinyts: Memory-efficient tinyml model compiler framework on microcontrollers, in: 2024 IEEE International Symposium on High-Performance Computer Architecture (HPCA), IEEE, 2024, pp. 848–860.

[28] R. F. Pamungkas, I. B. K. Y. Utama, Y. M. Jang, A novel approach for efficient solar panel fault classification using coupled udensenet, Sensors 23 (2023) 4918.

[29] M. S. H. Onim, Z. M. M. Sakif, A. Ahnaf, A. Kabir, A. K. Azad, A. M. T. Oo, R. Afreen, S. T. Hridy, M. Hossain, T. Jabid, et al., Solnet: a convolutional neural network for detecting dust on solar panels, Energies 16 (2022) 155.

[30] I. N. K. Wardana, S. A. Fahmy, J. W. Gardner, Tinyml models for a low-cost air quality monitoring device, IEEE Sensors Letters (2023).

[31] A. Oulefki, Y. Himeur, T. Trongtirakul, K. Amara, S. Agaian, S. Benbelkacem, M. A. Guerroudji, M. Zemmouri, S. Ferhat, N. Zenati, et al., Detection and analysis of deteriorated areas in solar pv modules using unsupervised sensing algorithms and 3d augmented reality, Heliyon 10 (2024).

[32] A. Mellit, N. Blasuttigh, A. M. Pavan, Tinyml for fault diagnosis of photovoltaic modules using edge impulse platform, in: 2023 11th International Conference on Smart Grid (icSmartGrid), IEEE, 2023, pp. 01–05.

[33] S. Bhattacharya, M. Pandey, Deploying an energy efficient, secure & high-speed sidechain-based tinyml model for soil quality monitoring and management in agriculture, Expert Systems with Applications 242 (2024) 122735.

[34] A. M. Hayajneh, F. Alasali, A. Salama, W. Holderbaum, Intelligent solar forecasts: Modern machine learning models & tinyml role for improved solar energy yield predictions, IEEE Access (2024).

[35] H. Oufettoul, R. Chaibi, S. Motahhir, Tinyml applications, research challenges, and future research directions, in: 2024 21st Learning and Technology Conference (L&T), IEEE, 2024, pp. 86–91.

[36] C. Srinivasan, P. Sridhar, S. Swathi, et al., A tinyml based residual binarized neural network for real-time image classification, in: 2022 6th International Conference on Electronics, Communication and Aerospace Technology, IEEE, 2022, pp. 1117–1124.

[37] P. Esmaili, F. Cavedo, P. Esmaili, M. Norgia, Tinyml anomaly detection in portable cutting tools, in: 2023 IEEE International Conference on Metrology for eXtended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE), IEEE, 2023, pp. 747–751.

[38] J. Da Silva, T. Flores, S. Júnior, I. Silva, Tinyml-based pothole detection: A comparative analysis of yolo and fomo model performance, in: 2023 IEEE Latin American Conference on Computational Intelligence (LA-CCI), IEEE, 2023, pp. 1–6.

[39] G. Gruosso, G. S. Gajani, Comparison of machine learning algorithms for performance evaluation of photovoltaic energy forecasting and management in the tinyml framework, IEEE Access 10 (2022) 121010–121020.

[40] M. Millendorf, E. Obropta, N. Vadhavkar, Infrared solar module dataset for anomaly detection, 2020.