PVPerformance
MODELING COLLABORATIVE

Deep SVDD-Driven Anomaly Detection and Multiclass Classification for Predictive Maintenance of Photovoltaic Systems

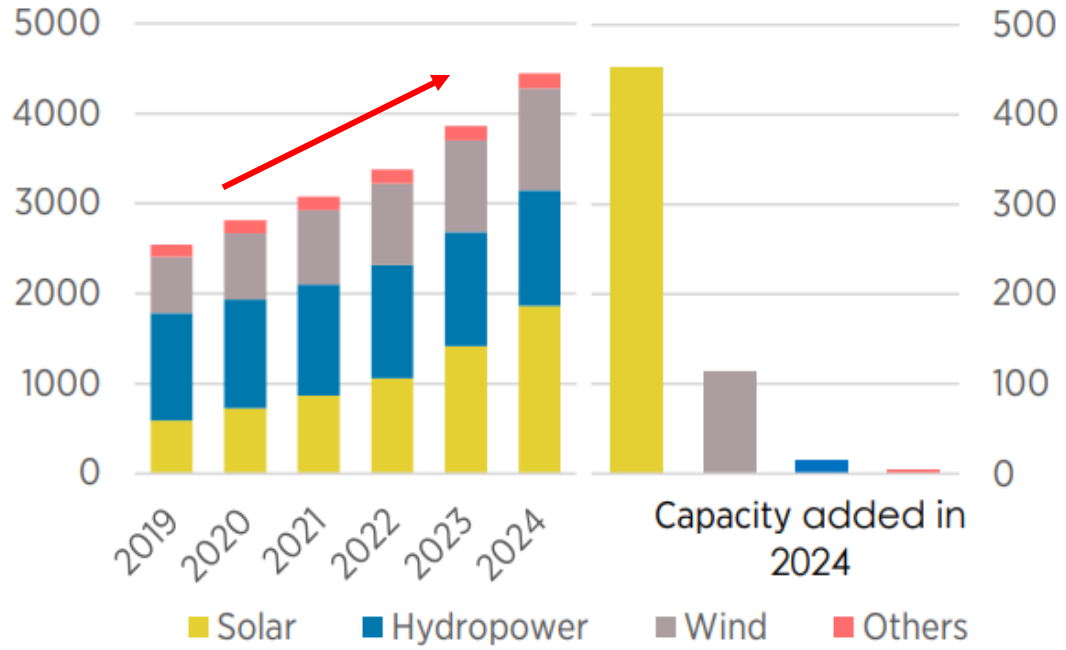
Imane FLOUCHI*, Yasmine El Mrabet, Said ElHamaoui , Khadija EL Ainaoui, Abdellatif Ghennioui, Khalid Bouabid, Hassan Nounah



flouchi@greenenergypark.ma

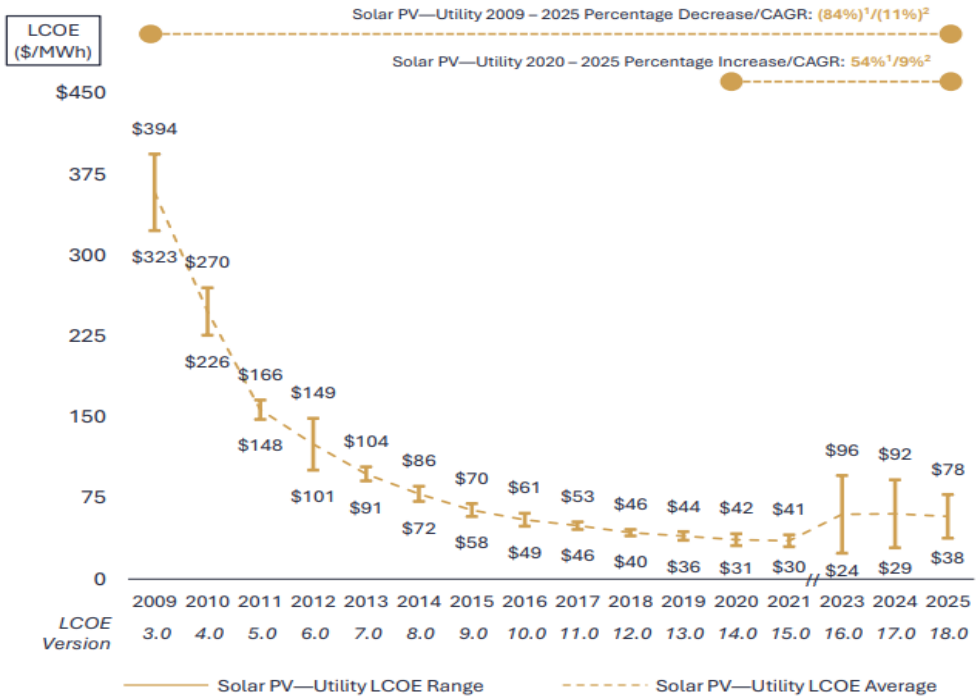
The Global Rise of Solar Energy

Renewable power capacity growth (GW)



+440 GW Solar PV accounted for 75% of all new renewable capacity added in 2023-2024.

The Economic Driver: Cost Reduction



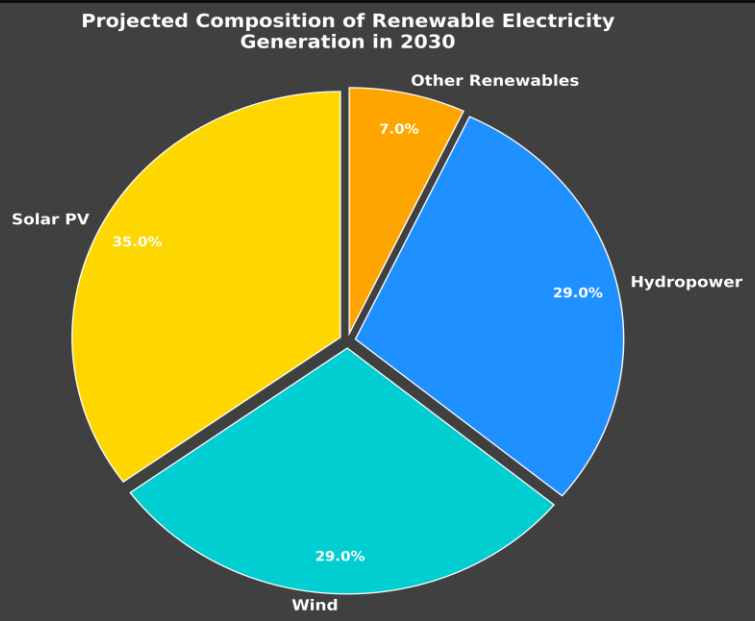
-85% The cost of utility-scale solar electricity has plummeted by over 85% in a decade.

Solar has become the cheapest source of electricity in history → **This global revolution is the engine of Morocco's energy strategy and creates critical new challenges for large-scale operation and predictive maintenance.**

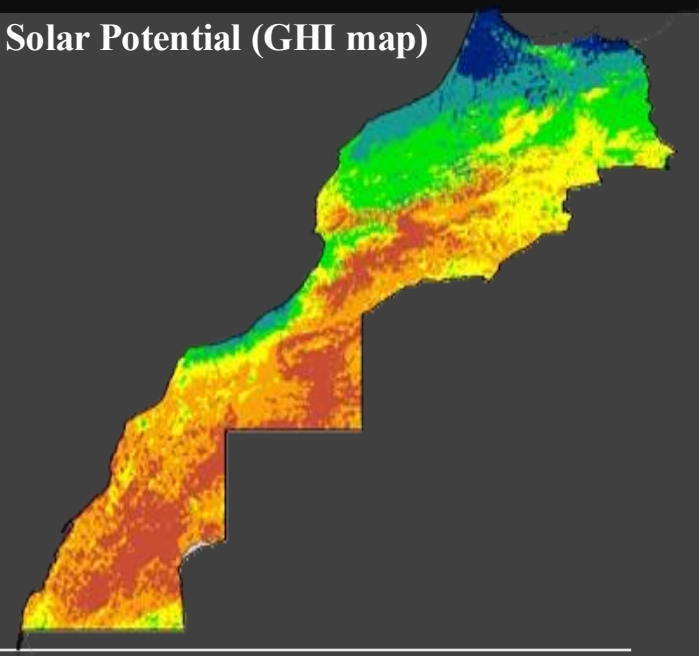
IRENA, "Renewable Capacity Statistics 2024".
Lazard, "Levelized Cost of Energy+ v18.0", 2024..

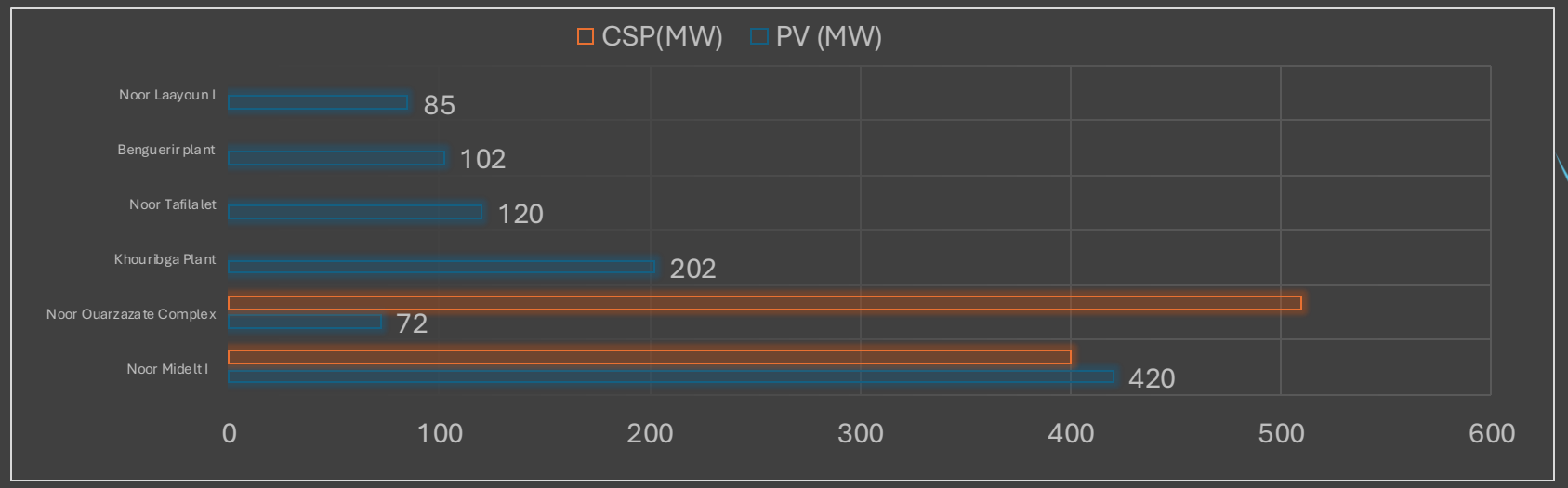
1 Introduction & Context

National Strategy



Morocco's national strategy targets 52% of installed electrical capacity from renewable sources by 2030, with solar energy as a central pillar.





Sources (2025)

IRESEN
Institut de Recherche en Energie Solaire et en Energies Nouvelles

MINISTÈRE DE LA TRANSITION ÉNERGÉTIQUE ET DU DÉVELOPPEMENT DURABLE

masen
force inépuisable de développement

Référence: Green Energy Park, Benguerir, 2025

Challenges & Objectives

The Challenges of PV Maintenance

Production & Revenue Losses

Undetected faults reduce energy yield and profitability.

Premature Degradation

Certain faults can damage components and reduce the system's lifetime.

Complex Diagnostics

Systems often flag an issue without identifying the root cause.

The Need: To shift from **reactive** to **predictive maintenance**.

A Unified Framework

DETECT

Enable early detection of any deviation from normal operation (anomalies).

CLASSIFY

Determine the exact nature of the fault to guide targeted maintenance actions.

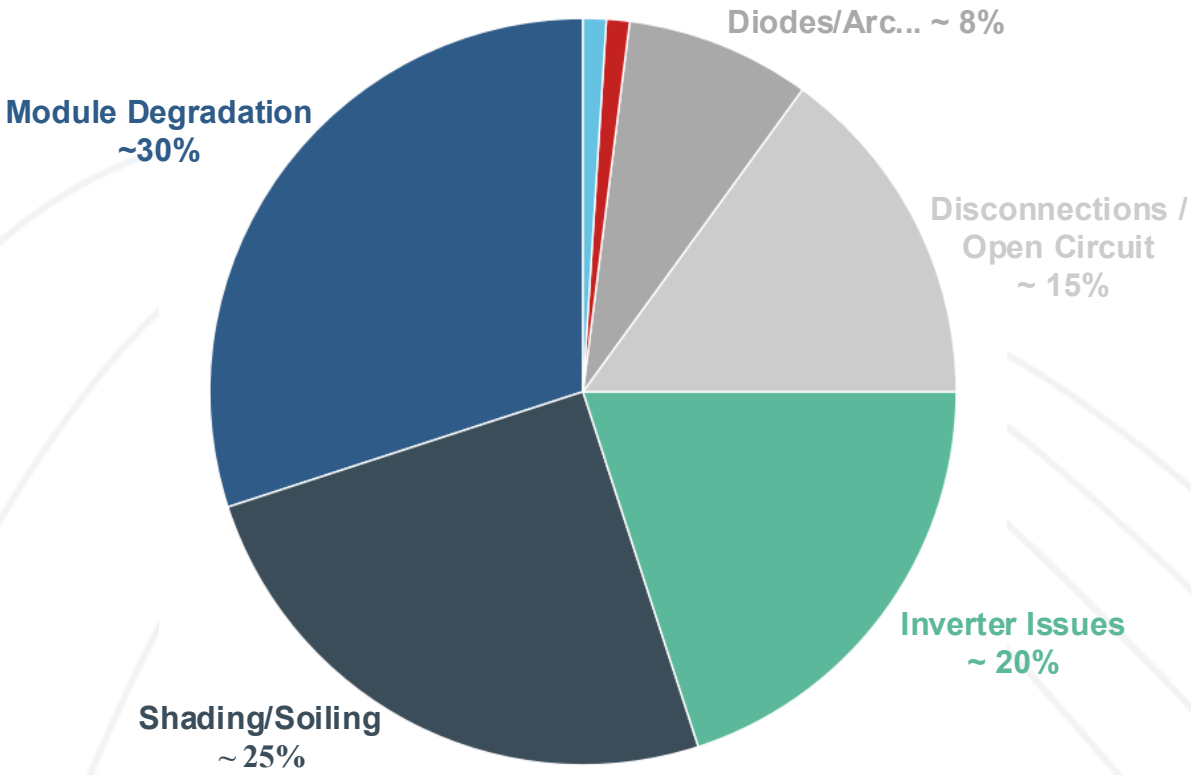
ACCELERATE

Provide real-time analysis compatible with the demands of field monitoring.

The **Solution**: A single model that rapidly **detects & classifies faults.**

The Cost of Inaction: The Quantified Impact of PV Faults

Distribution of Common Faults and Failures



NREL Reliability Workshop, 2022

Financial and Operational Consequences

Up to 10-30%
ANNUAL PRODUCTION LOSS
Due to undetected faults, degradation, and failures.

2x to 3x MORE EXPENSIVE
COST OF CORRECTIVE MAINTENANCE
Compared to a planned predictive approach.

INCREASED RISK
CRITICAL FAULTS (ELECTRICAL ARCS)
Can lead to safety risks and irreversible damage..

Early and precise detection is not an option…it's an economic and safety necessity.

From Data to Actionable Diagnosis

Phase 1 :Acquisition & Encoding

1.Raw Data Input

Irradiance, Voltage, Current, Temperature

2. Preprocessing & Feature Extraction

•Time Windowing: Data is segmented into sequences.

•1D-CNN Encoder: Extracts relevant patterns from each sequence.

3.Latent Vector (z)

- A compact digital "health signature'

Phase 2 :Detection

ANOMALY DETECTION

- Method: Deep SVDD

Is the latent vector (z) inside the learned normality boundary?

YES

Normal Operation
(Continuous Monitoring)

NON

Phase 3: Classification

FAULT CLASSIFICATION

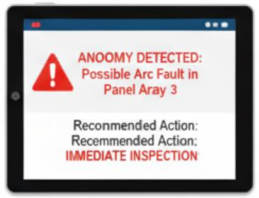
Method: One-vs-Rest (OvR) SVDD

Determines the specific nature of the fault:

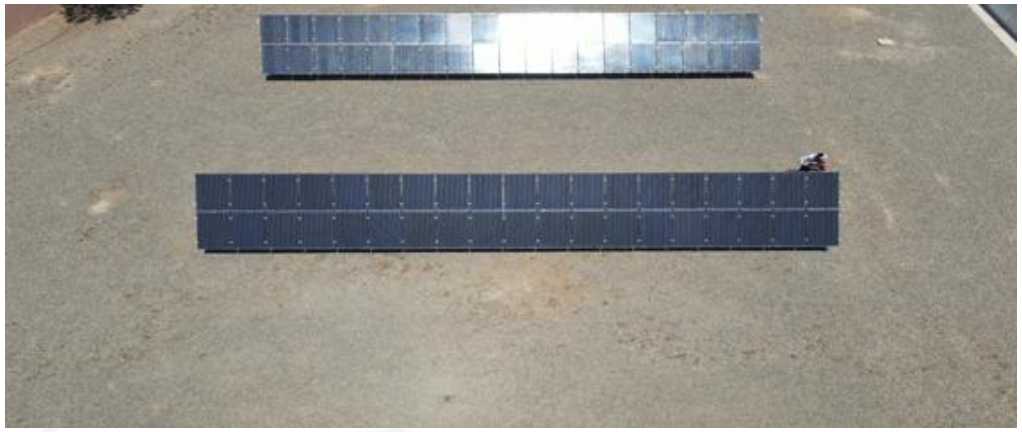
- Short-Circuit
- Open-circuit
- Partial Shading
- Degradation
- Arc Fault

Phase 4: Actionable Result

DIAGNOSIS & ALERT



Experimental Protocol: From On-Field Data Generation to Model Validation





•**LOCATION:** PV test platform at the Green Energy Park (GEP), Benguerir, Morocco.

•**OBJECTIVE:** To create a rich and realistic dataset, including both baseline normal operation and controlled fault conditions

PV Array Specifications:

Peak Power: 22.23 kWp

Technology: Monocrystalline Silicon

Modules: Trina 195W

Configuration: 6 parallel strings, 19 modules per string

Inverter Specifications:

Model: SMA STP 20000TLEE-10

Max Power: 20 kW

Monitoring System:

Equipment: Gantner Q.series Data Acquisition System

Parameters: String Currents (I_1-I_6), Voltage (Voc), Irradiance (POA), Temperature.

Acquisition Frequency: 1 Hertz (1 sample/second)

Our Guiding Principles for Data Generation

- **Realism:** To simulate plausible and realistic fault conditions observed in the field.
- **Severity Diversity:** To create faults with low, medium, and high intensity for each failure type.
- **Location Variability:** To apply faults at different locations within the system (on a single module, across multiple modules, on one string, across multiple strings).

Short circuit



Degradation



Partial Shading



Open circuit



Fault Type	Severity Group	Simulation Procedure	Weather Conditions	Duration	Repetition	Programmed Time
Open-Circuit Faults	0-25%	Disconnect 1 PV string	Clear, Cloudy	15 min	3	10:00 - 10:15, 12:00 - 12:15, 14:00 - 14:15
	25-50%	Disconnect 2 PV strings	Clear, Cloudy	15 min	3	10:30 - 10:45, 12:30 - 12:45, 14:30 - 14:45
Short-Circuit Faults	0-25%	Short-circuit 1 PV module	Clear, Cloudy	15 min	2	11:00 - 11:15, 13:00 - 13:15
	25-50%	Short-circuit 2 PV modules	Clear, Cloudy	15 min	2	11:30 - 11:45, 13:30 - 13:45
	50-75%	Short-circuit 2 PV strings (LL fault)	Clear, Cloudy	15 min	2	12:00 - 12:15, 14:00 - 14:15
	75-100%	Short-circuit 1 PV string (Line to Ground fault)	Clear, Cloudy	15 min	4	12:30 - 12:45, 14:30 - 14:45
Degradation	0-25%	Add resistors (10 ohms) in series with 1 PV string	Clear, Cloudy	15 min	2	10:00 - 10:15, 12:00 - 12:15
	25-50%	Add resistors (10 ohms) in series with 2 PV strings	Clear, Cloudy	15 min	2	10:30 - 10:45, 12:30 - 12:45
	50-75%	Add resistors (20 ohms) in series with 1 PV string	Clear, Cloudy	15 min	2	11:00 - 11:15, 13:00 - 13:15
	75-100%	Add resistors (20 ohms) in series with 2 PV strings	Clear, Cloudy	15 min	2	11:30 - 11:45, 13:30 - 13:45
Partial Shading	0-25%	Shade 1 PV module	Clear, Cloudy	15 min	2	12:00 - 12:15, 14:00 - 14:15
	25-50%	Shade 2 PV modules	Clear, Cloudy	15 min	2	12:30 - 12:45, 14:30 - 14:45
	50-75%	Shade 2 PV modules in 2 different PV strings	Clear, Cloudy	15 min	2	10:00 - 10:15, 12:00 - 12:15
	75-100%	Shade 3 PV modules in 2 different PV strings	Clear, Cloudy	15 min	2	10:30 - 10:45, 12:30 - 12:45
	0-25%	Shade 3 PV modules	Clear, Cloudy	15 min	2	11:00 - 11:15, 13:00 - 13:15
	25-50%	Shade 4 PV modules in 2 different PV strings	Clear, Cloudy	15 min	2	11:30 - 11:45, 13:30 - 13:45
	50-75%	Shade 4 PV modules	Clear, Cloudy	15 min	2	12:00 - 12:15, 14:00 - 14:15

By applying faults at different locations, we force the model to learn complex electrical signatures and not overfit to a single configuration, making it far more robust.



Our High-Resolution Dataset: The Foundation of Our Approach

1. Our Dataset

- **Key Indicators**

Volume: 1.2 Million data points

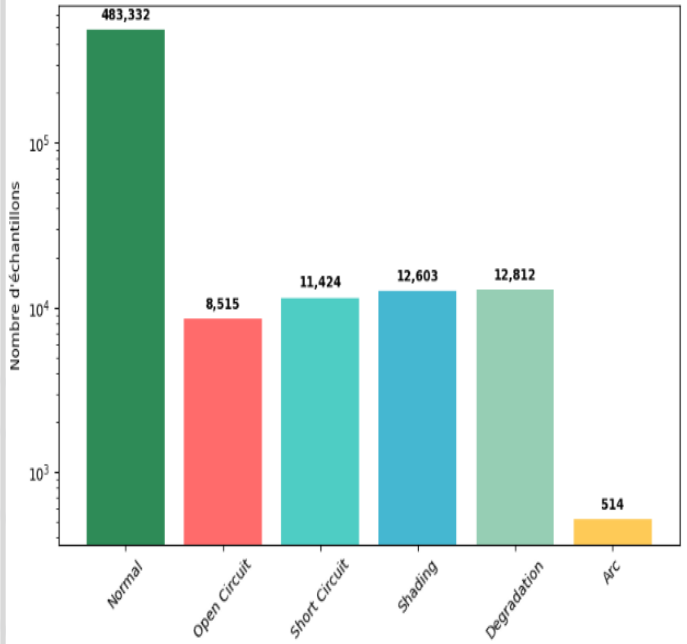
Period : 15-day intensive Campaign

Frequency: 1 Hertz (1 sample/second)

- **Detailed Distribution**

Class	Number of Samples	Percentage
No.Fault	1129240	94,51549967
Open circuit	12165	1,018190158
Short circuit	16320	1,36595671
Shading	18004	1,506904694
Degradation	18303	1,531930494
Arc	735	0,061518271

2. The Main Challenge

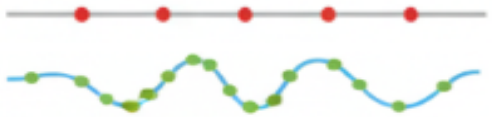


The Challenge: Extreme Class Imbalance.
Faults are rare events, which makes their detection difficult for a standard model.

3. Our Strategic Advantage

- **The Power of High Resolution**

•Low-Resolution Data



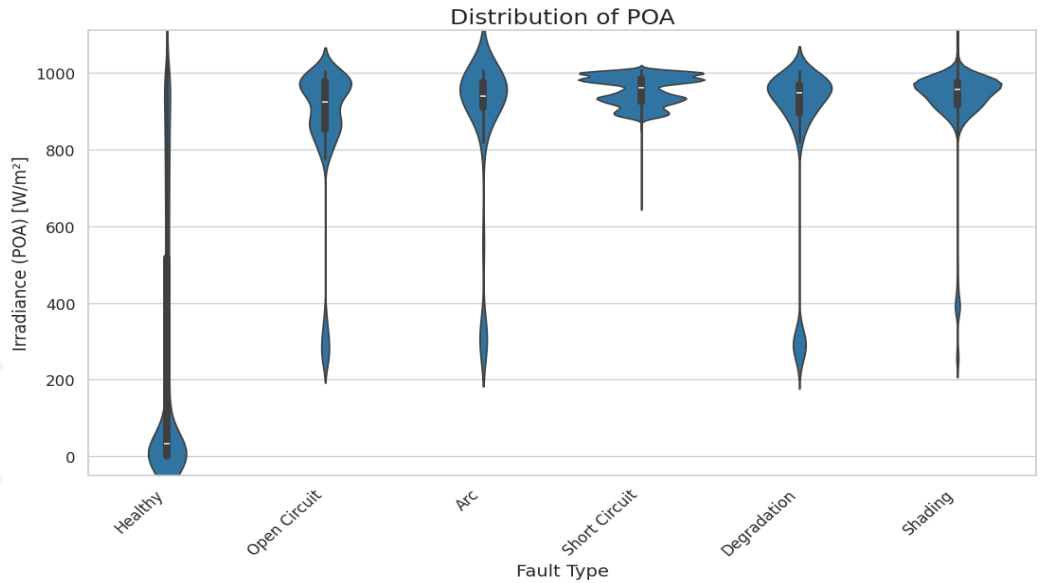
•1 Hz Data (Ours)

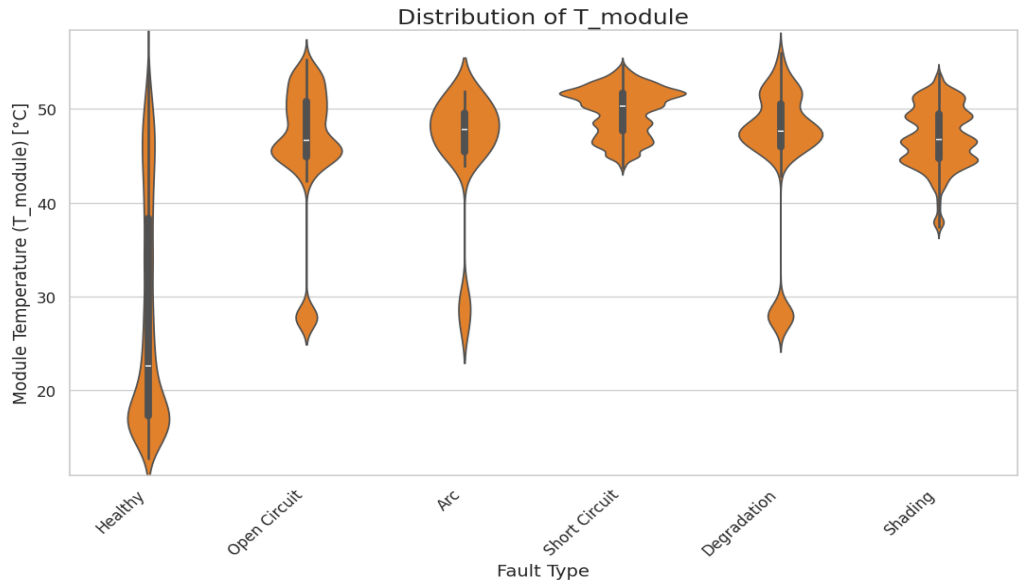
The Justification for Our Model:

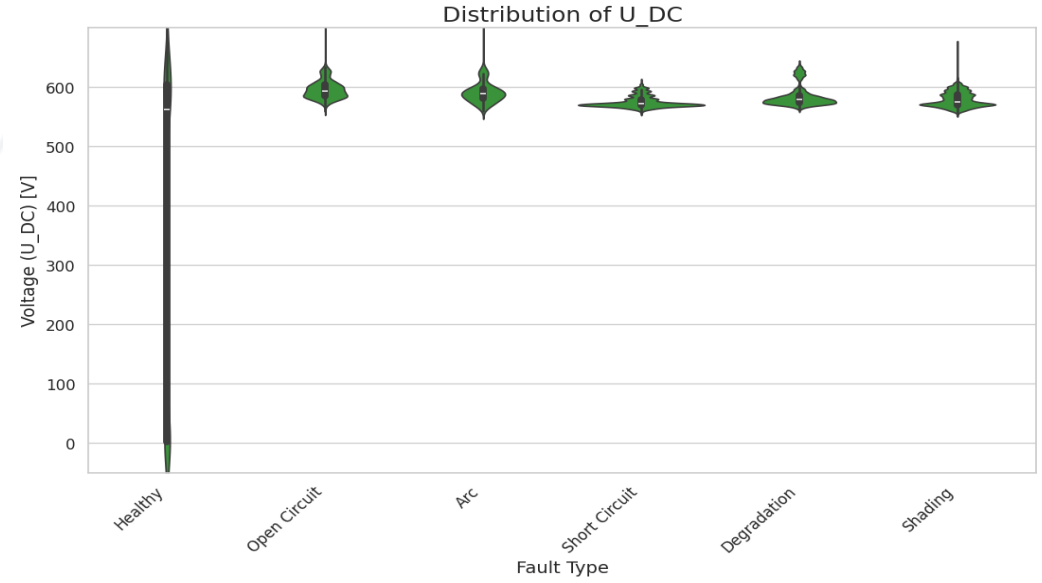
•The **1 Hertz frequency** allows us to capture the complete **dynamic signature** of a fault (its "shape").

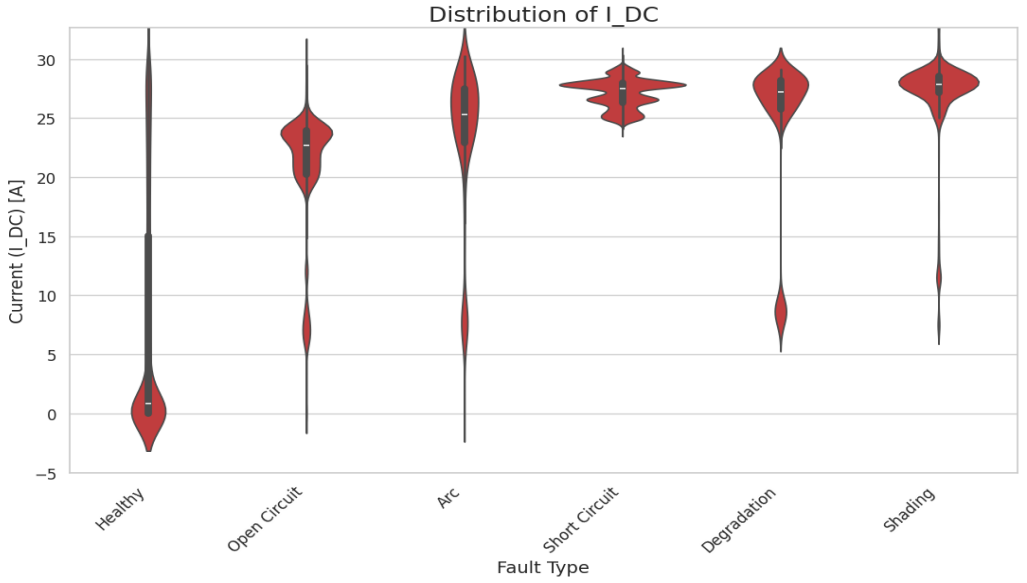
•These high-resolution sequences are the ideal input for a **1D-CNN**, which is an expert at recognizing these complex shapes.

Statistical Distribution of Key Features by Fault Type

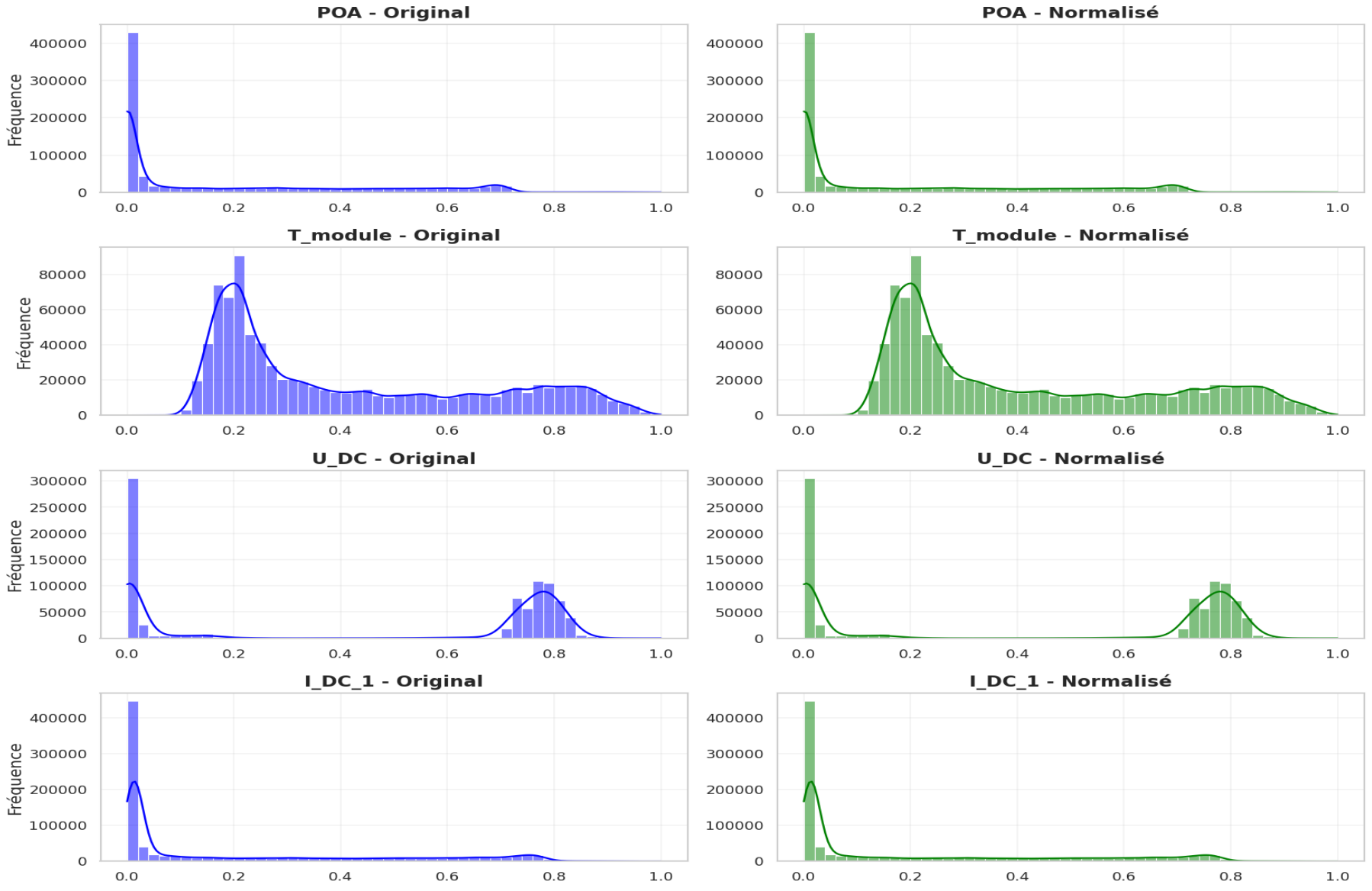




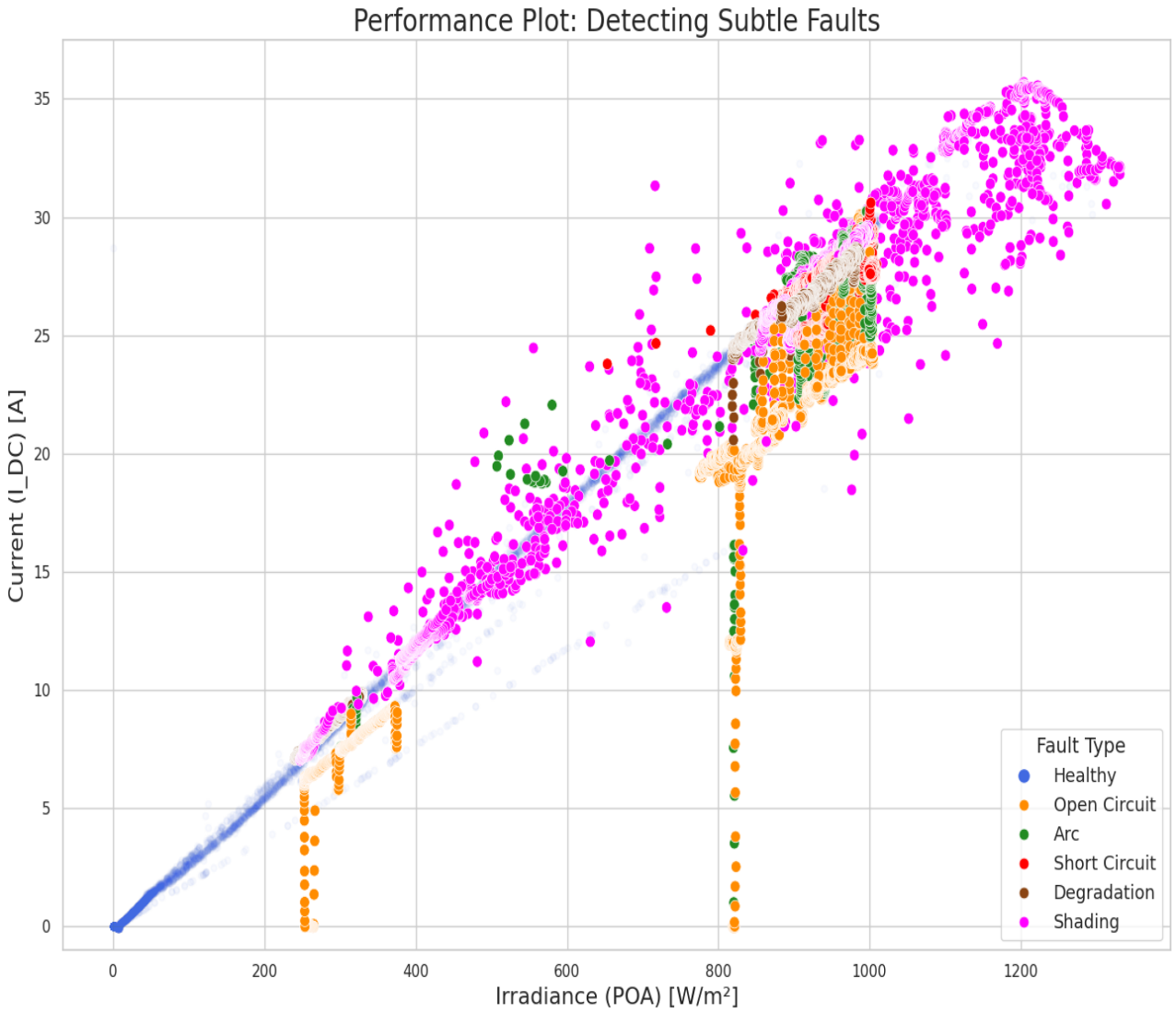


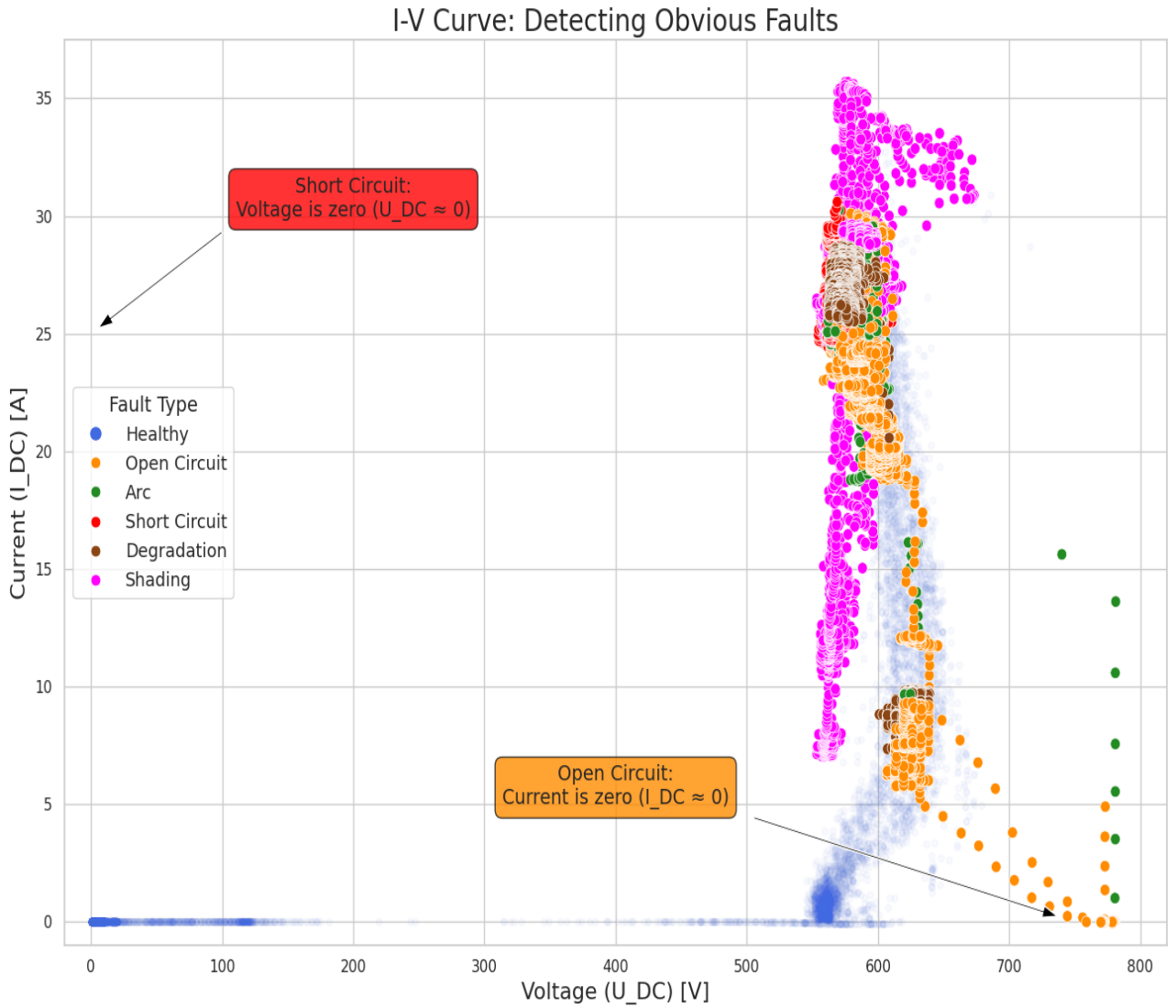


Visual Analysis of Fault Signatures

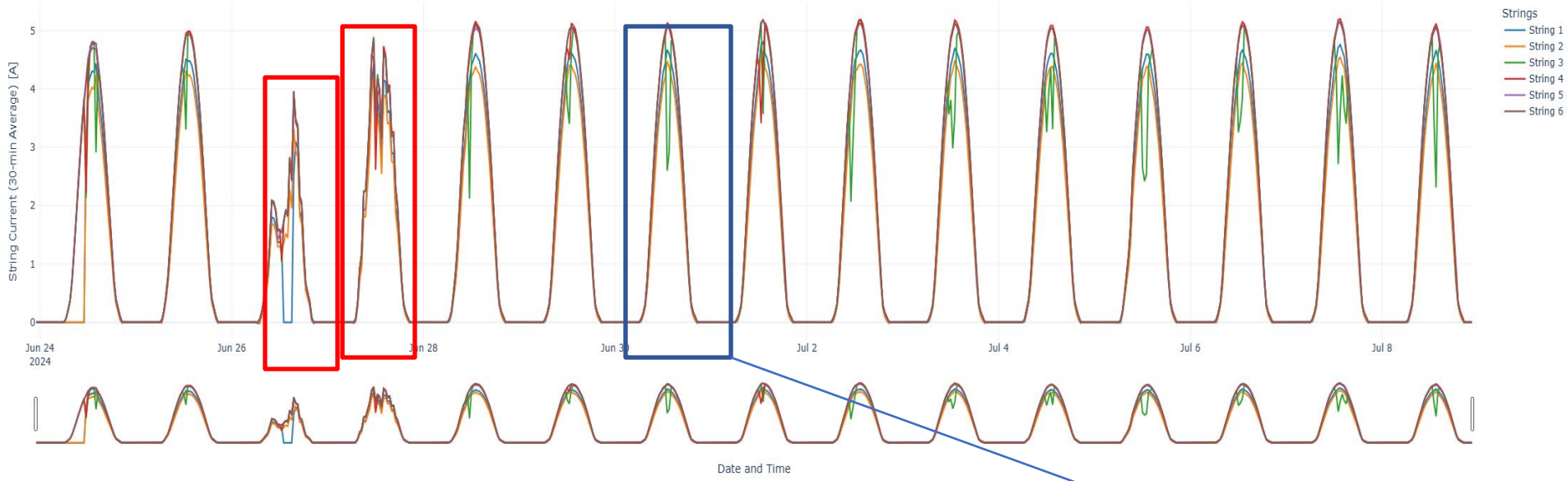


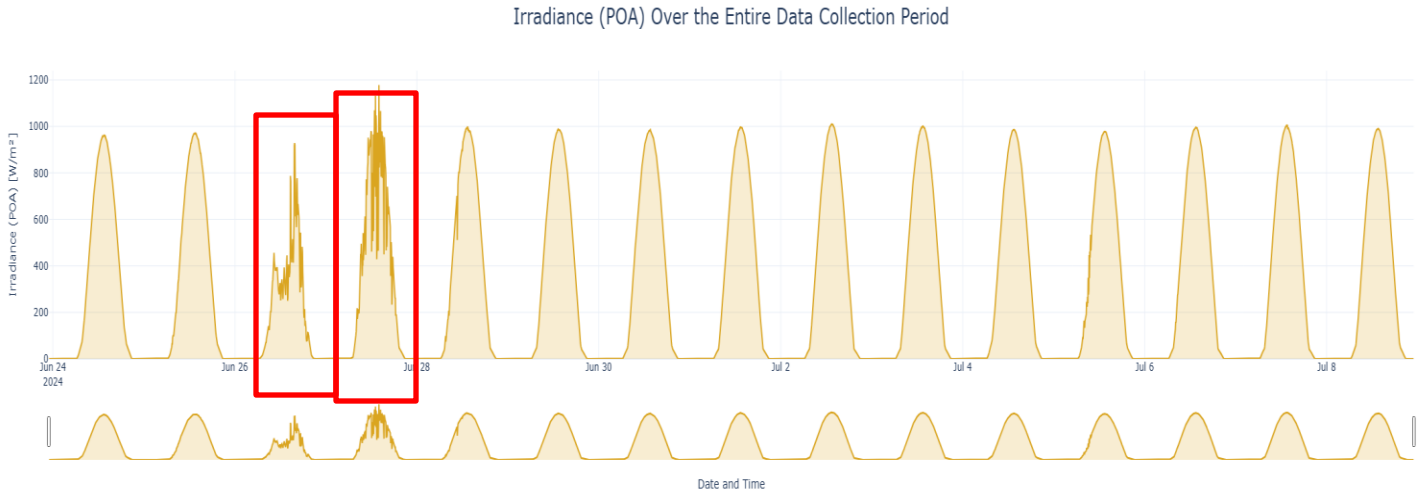
Visual Analysis of Fault Signatures

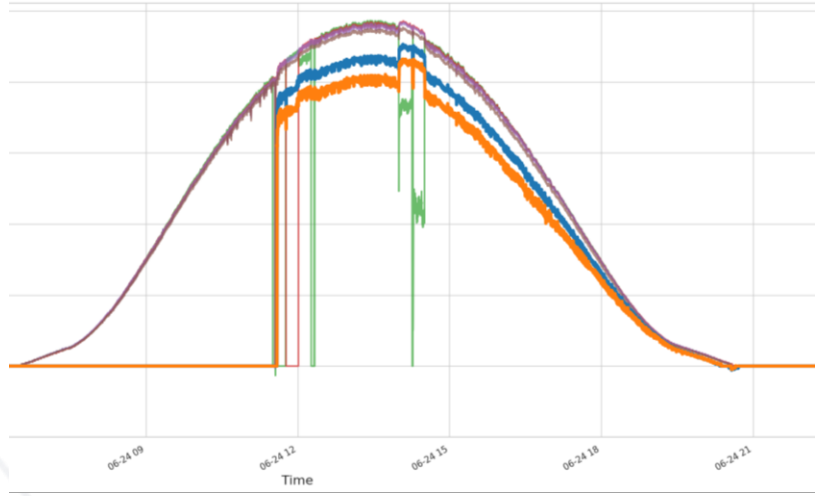




Visual Analysis of Fault Signatures

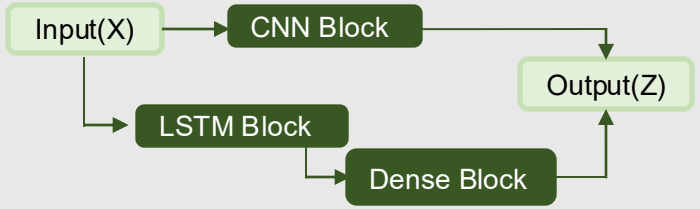






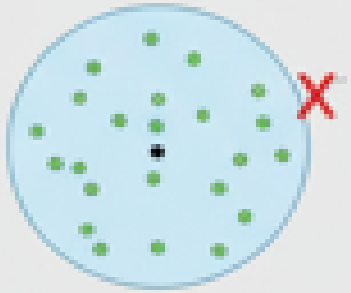
Detailed Methodology: The Mathematical Framework of the Hybrid Model

Block 1: Feature Extraction (Encoder Φ)



1. Convolution (Pattern Detection): h_{CNN} = ReLU(Conv1D(X))
2. Temporal Analysis (Memory): h_{LSTM}= LSTM(h_{CNN})
3. Final Projection (Signature): z = Dense(h_{LSTM})

Block 2: Anomaly Detection (Deep SVDD)



- Training Objective (Loss Function):

$$\underbrace{\min_{W,c} \frac{1}{n}\sum_{i=1}^{n}\|\Phi(X_i;W)-c\|^2}_{\text{Objective Term}} + \underbrace{\frac{\lambda}{2}\sum_{l}\|W_l\|_F^2}_{\text{Regularization Term}}$$

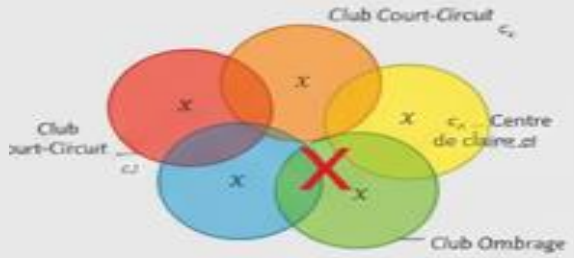
- Decision Rule:

Anomaly if : $\|\Phi(X_i;W)-c\|^2 > R^2$

- In simple terms:

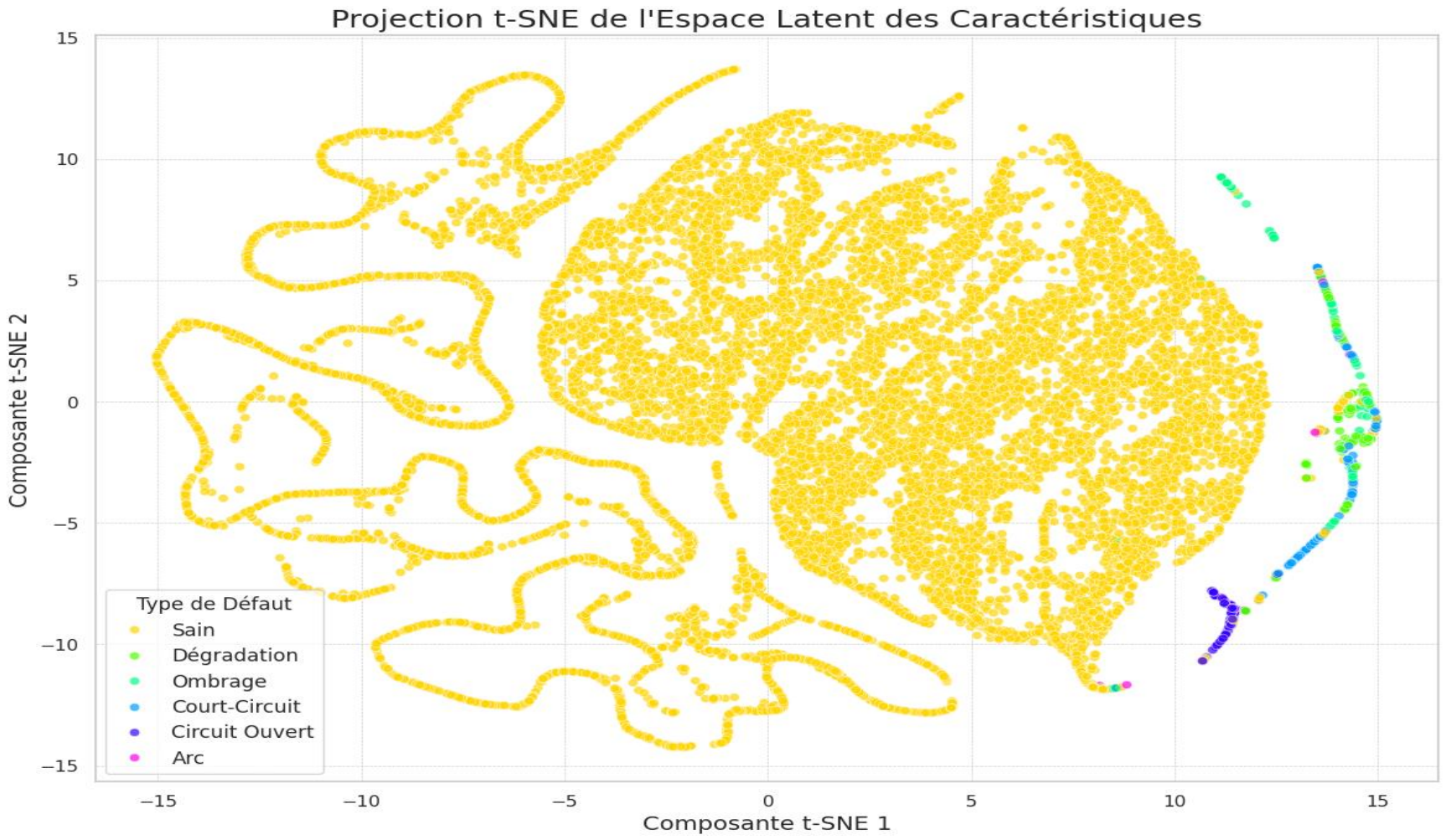
Training: The model learns to enclose all "healthy" signatures within the smallest possible sphere.
Decision: If a new signature falls outside this sphere, it's an anomaly.

Block 3: Fault Classification (One-vs-Rest)



- Principle: 5 independent expert models (Φ_k, C_k), one for each fault class k.
- Final Decision Rule: Predicted Class= $\arg\min_K \|\Phi_k(x,w_k)-c_k\|^2$
- In simple terms:
  Intuition: The diagnosis is given by the expert that finds the anomaly most "familiar" (i.e., the one for which the distance is the smallest).

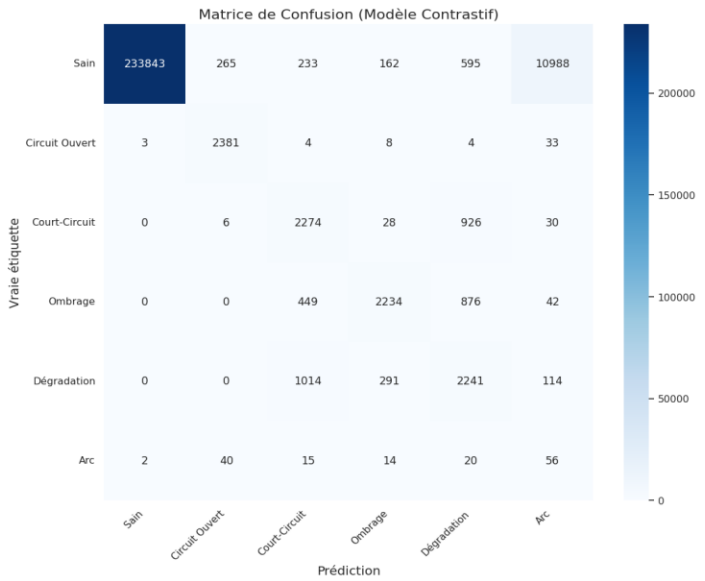
Visual Analysis of Fault Signatures



Performance of the hybrid Model (CNN-LSTM + Deep SVDD)

Detailed Classification Report

Class	F1-Score
Normal	0.98
Open Circuit	0.92
Short-Circuit	0.89
Shading	0.91
Degradation	0.88
Arc Fault	0.87



- **Quantitative Analysis:**

-The model achieves high overall performance (F1-Score > 0.90), which validates the robustness of the Deep SVDD approach.
-However, slightly lower scores are observed for certain classes, indicating some difficulties in distinguishing between them.

- **Qualitative Analysis:**

-The strong diagonal confirms the model's high overall accuracy.
-Arc Fault shows some confusion with Short-Circuit, as their initial electrical signatures can be similar.
-Degradation is sometimes confused with Shading, as both manifest as a reduction in current.

The baseline Deep SVDD model provides a solid foundation for diagnosis. To resolve the observed confusions and improve the distinction between faults, we explored a more advanced feature learning strategy: **Contrastive Learning**.

Impact & Future Work

Impact

✓ Enables **early detection** of anomalies.

✓ Provides **precise diagnosis**.

✓ **Optimizes maintenance costs** and **maximizes energy production**.

Perspectives

→ **Deployment on an embedded system.**

→ **Validation on other sites and climates.**

→ **Integration into a full monitoring platform.**

Thank you for your attention!

Imane Flouchi

Email :Flouchi@greenenergypark.ma
Tél. : (+212) 633527617