

Article

Hybrid Fault-Space Restructuring for Machine Learning-Based Fault Diagnosis in Power Electronic Converters

José M. García-Campos ^{1,*}, Abraham M. Alcaide ^{2,3}, Alejandro Letrado-Castellanos ², Ramon Portillo ²
and Jose I. Leon ^{2,3}

¹ Department of Telematics Engineering, University of Sevilla, 41092 Sevilla, Spain

² Department of Electronic Engineering, University of Sevilla, 41092 Sevilla, Spain; amalcaide@us.es (A.M.A.); aleletcas@alum.us.es (A.L.-C.); ramonpg@us.es (R.P.); jileon@us.es (J.I.L.)

³ ENGREEN Laboratory of Engineering for Energy and Environmental Sustainability, University of Sevilla, 41092 Sevilla, Spain

* Correspondence: jgarcia111@us.es

Abstract

Fault diagnosis in power electronic systems is challenging when fault categories geometrically overlap within the measurement space, limiting class separability and introducing classification ambiguity. This work proposes an edge-oriented hybrid fault-space restructuring methodology that utilizes UMAP-based embeddings and hierarchical clustering to group overlapping fault conditions into robust hybrid representations. Subsequently, supervised machine learning models execute the final classification over this optimized space. Validation was conducted using a large-scale synthetic dataset generated via real-time hardware-in-the-loop (HIL) simulation, evaluating electrical measurements from three-dimensional RMS values to 60-dimensional instantaneous waveforms. Tested with Decision Tree and Random Forest algorithms, the restructuring strategy significantly improves robustness under geometric ambiguity compared to conventional classification without space restructuring. Specifically, low-dimensional measurements achieved F1-score improvements of approximately 72% and 46% for the Decision Tree and Random Forest algorithms, respectively, while high-dimensional measurement configurations still exhibited significant improvements of 36% and 52%. Consequently, these results confirm that the combined restructuring and classification pipeline is highly effective across the analyzed measurement dimensionalities, establishing a dependable cluster-based diagnostic strategy that enhances classification robustness while accepting a trade-off in individual fault-isolation granularity. Finally, hardware deployment experiments on a Raspberry Pi 4 platform demonstrated the feasibility of executing the trained classifiers for real-time inference under constrained computational environments. The experimental evaluation validated real-time execution capabilities, achieving sub-millisecond inference latencies (as low as 0.32 ms), a memory footprint under 0.14 MB, and processing rates exceeding 2600 inferences per second using lightweight Decision Tree classifiers. Ultimately, these findings indicate that the proposed strategy improves fault detection across the evaluated measurement configurations while ensuring a highly viable execution on resource-constrained devices once the classifiers are trained.

Keywords: power electronics; fault diagnosis; UMAP; geometric fault analysis; machine learning; edge deployment



Academic Editor: Pedro J. Villegas

Received: 2 June 2026

Revised: 3 July 2026

Accepted: 6 July 2026

Published: 9 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article

distributed under the terms and

conditions of the [Creative Commons](#)

[Attribution \(CC BY\)](#) license.

1. Introduction

Power electronic converters play a critical role in modern energy systems, including renewable energy generation, electric mobility, and distributed energy resources. Their reliability directly impacts system availability, safety, and operational costs. Consequently, condition monitoring and predictive maintenance strategies have become essential to ensure the continuous and efficient operation of these systems.

In industrial environments, monitoring infrastructures are typically built upon Supervisory Control and Data Acquisition (SCADA) systems and related supervisory platforms. These solutions provide reliable plant-level supervision, long-term data storage, and centralized control capabilities. However, they are mainly designed for steady-state operation and low-frequency data acquisition [1], which limits their ability to capture the fast electrical and thermal dynamics associated with degradation processes in power electronic components. As a result, relevant information for early fault detection and accurate State-of-Health (SoH) estimation may remain hidden within high-frequency signals that are not fully exploited [2].

At the same time, the increasing availability of embedded sensing and high-resolution measurements in power converters has led to a significant growth in the volume of operational data [3]. While these data streams contain valuable information about the internal behavior of the system, a key challenge lies in transforming raw measurements into actionable knowledge. In current practice, this transformation is often performed offline, separating data acquisition from data interpretation and limiting the ability to react in real time to abnormal conditions [4], as condition monitoring approaches frequently involve post-processing of acquired data, with limited capability for real-time interpretation [5].

In response to these limitations, recent research has explored distributed monitoring architectures based on IoT and microservices paradigms to improve the scalability and deployment flexibility of monitoring systems [6]. These architectures enable the integration of multiple data sources and facilitate the management of complex monitoring environments. However, despite these architectural advancements, existing solutions often emphasize data collection and storage functionalities, while the integration of advanced analysis and decision-making within the real-time monitoring loop remains limited [7].

Nevertheless, the fault diagnosis problem in power electronic systems remains particularly challenging due to the existence of partially overlapping operating conditions within the measurement space [8]. Different fault categories may exhibit highly similar electrical behavior, especially under noisy or dynamic operating conditions [9], reducing fault separability and increasing inter-fault confusion during supervised classification.

In this context, machine learning techniques have demonstrated strong capabilities for identifying complex patterns and extracting discriminative information from high-dimensional measurement data in power electronic systems [10]. However, their application is commonly restricted to offline environments or centralized computing infrastructures, which introduces latency and reduces system autonomy [11].

Edge computing has emerged as a promising approach to address these limitations by enabling local data processing and supporting near-real-time decision-making at the monitoring node. Within the paradigm of edge intelligence [12], deploying artificial intelligence models on resource-constrained devices, such as Raspberry Pi platforms, enables data processing closer to the source, reducing response times and mitigating the impact of communication constraints [13,14]. However, integrating edge-based intelligence into distributed monitoring systems for power converters, while maintaining scalability and data reliability, remains a challenging task, as it involves balancing computational constraints, data management requirements, and timing constraints inherent to power electronic systems.

Although machine learning has significantly improved fault diagnosis capabilities, different fault categories may still exhibit overlapping geometric distributions within the measurement space. In such cases, increasing model complexity alone may not resolve the ambiguity, since the limitation arises from the separability of the measurements rather than from the learning algorithm itself. Consequently, improving diagnostic performance may require adapting the diagnostic label space to the information actually contained in the measurements before supervised learning is performed. This paper investigates a geometry-aware diagnostic methodology that restructures the diagnostic label space according to the geometric relationships identified within the selected measurement space. The proposed approach combines fault-space restructuring based on UMAP projections and hierarchical clustering with machine learning inference evaluated under edge-oriented deployment conditions. Additionally, the methodology decouples model training and inference, allowing computationally intensive learning tasks to be performed offline, while inference feasibility is analyzed on resource-constrained embedded platforms.

The main contributions of this work are as follows:

- A geometry-aware fault-space restructuring methodology, based on UMAP and hierarchical clustering, that identifies fault categories with limited geometric separability and automatically restructures the diagnostic label space before supervised learning.
- An edge-oriented diagnostic strategy where the optimized fault-space representations enable lightweight machine learning classifiers to be trained offline and subsequently deployed for real-time inference on resource-constrained hardware.
- A comprehensive experimental validation of the proposed methodology using a large-scale hardware-in-the-loop (HIL) synthetic dataset, evaluating its diagnostic performance across three measurement dimensionalities and verifying its real-time execution viability on an embedded hardware platform.

The rest of this manuscript is organized as follows. Section 2 reviews related work on condition monitoring, machine learning-based fault diagnosis, distributed monitoring architectures, and edge computing in power electronic systems. Section 3 describes the real-time simulation environment and dataset generation process. Section 4 presents the proposed hybrid fault-space restructuring methodology. Section 5 provides the experimental validation, including geometric fault-space analysis, classification performance evaluation, and edge-oriented inference experiments on a Raspberry Pi 4 platform. Section 6 discusses the limitations of the proposed approach. Finally, Section 7 summarizes the main conclusions and future research directions.

2. Related Works

Condition monitoring and fault diagnosis in power electronic converters have been extensively studied due to the critical role of these systems in modern energy applications [15–17]. Traditional approaches are typically based on physics-based models and the observation of key electrical and thermal variables to estimate degradation and predict failures [18–20]. These methods rely on predefined failure models and often require detailed knowledge of device behavior, which can be difficult to obtain in complex and dynamic operating conditions [21,22].

With the increasing availability of sensing capabilities, data-driven approaches have gained significant attention in recent years [23,24]. Machine learning techniques have been successfully applied to condition monitoring tasks, including fault detection [25], classification [26], and remaining useful life estimation [27] in power electronic systems. These approaches are particularly effective in capturing nonlinear relationships and extracting patterns from high-dimensional data. However, most existing implementations [28]

are designed for offline analysis or centralized processing environments, limiting their applicability in real-time scenarios.

Despite the promising performance of machine learning techniques, fault diagnosis in power electronic systems remains challenging when different operating conditions exhibit partially overlapping measurement distributions. Under these scenarios, geometrically similar fault categories may generate highly correlated electrical responses, reducing fault separability and increasing inter-fault confusion during supervised classification [8,9]. Conventional dimensionality reduction techniques, such as Principal Component Analysis (PCA) or linear projections, often fail to resolve this intrinsic geometric ambiguity, leaving residual misclassifications that compromise diagnostic reliability [29]. Consequently, improving fault separability within the measurement space has become an important challenge for robust data-driven fault diagnosis.

In parallel, the development of Industrial IoT technologies has enabled the deployment of distributed monitoring systems capable of collecting data from multiple sources. Microservices-based architectures [30] have been proposed to improve scalability, modularity, and flexibility in monitoring platforms. These architectures decouple data acquisition, processing, and storage components, facilitating the integration of heterogeneous devices and communication protocols. Nevertheless, they are often primarily oriented toward management, while advanced analytics and decision-making are typically performed in external or offline processing layers.

More recently, large-scale industrial models and large language models (LLMs) have attracted increasing attention in fault diagnosis research [31]. These approaches aim to exploit the reasoning and knowledge-integration capabilities of foundation models to improve diagnostic adaptability under variable operating conditions. For example, adaptive industrial LLMs have been explored for mechanical fault diagnosis [32], providing a promising direction for knowledge-assisted and context-aware industrial diagnosis.

However, these methods usually involve computational requirements and deployment assumptions that differ from the edge-oriented scenario considered in this work. In contrast, the present study focuses on lightweight supervised classifiers trained offline and deployed on resource-constrained hardware, while diagnostic robustness is improved through fault-space restructuring based on the geometric separability of the selected measurements.

Recent research has explored the use of edge computing to address the limitations of centralized processing by enabling data analysis closer to the source [33,34]. Edge-based solutions reduce communication latency and improve system responsiveness, making them suitable for time-sensitive applications. In the context of power electronics, edge computing has been applied to monitoring [18] and control [35] tasks, as well as to the deployment of lightweight machine learning models [36]. Despite these advances, integrating edge-based intelligence into distributed monitoring architectures while ensuring scalability, reliability, and efficient resource utilization remains a significant challenge.

In summary, although significant progress has been made in condition monitoring, data-driven diagnostics, and distributed architectures, a critical gap persists between multi-dimensional data acquisition, geometric space optimization, and low-latency edge decision-making [23]. This work addresses these limitations through a fault diagnosis methodology that combines fault-space restructuring with supervised machine learning models trained offline and evaluated on resource-constrained embedded platforms.

3. Real-Time Simulation and Dataset

3.1. Real-Time Simulation Environment

Due to the absence of publicly available datasets containing comprehensive fault measurements for multiple power-converter operating conditions and fault sources, a real-time

simulation environment has been developed to generate the dataset used throughout this work. The proposed approach enables the controlled generation of electrical disturbances and fault conditions under reproducible operating scenarios while preserving the dynamic behavior of the converter system.

For this purpose, a detailed converter model was deployed in a Typhoon HIL HIL606 real-time simulator, which emulates the electrical behavior of the power stack and enables controlled fault injection under realistic operating conditions. The simulator continuously generates electrical measurements that are transmitted through UDP to a dedicated data acquisition platform for subsequent storage and analysis [6].

The simulated system corresponds to a 2 MW and 1 kHz 2L-VSI grid-tied inverter connected to a three-phase 400/690 V and 50 Hz grid and a 1500 V DC bus. In addition to the converter topology itself, the simulation environment incorporates multiple external fault sources associated with the electrical grid, the DC bus, and semiconductor operation. A basic scheme of the model is shown in Figure 1. The fault control function forces the state of different switches or changes some magnitudes in the electrical sources around the model, while some communication function blocks developed by Typhoon enable the sending of the measured data.

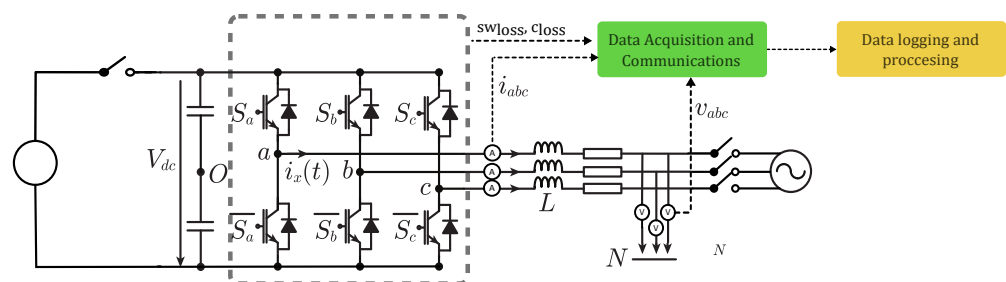


Figure 1. Schematic of the electrical fault model developed.

3.2. Fault Injection Strategy

The proposed fault injection strategy considers both physical faults and electrical disturbance conditions that could be present mostly in the grid, the DC connection (or the DC voltage source), or the connection of the power input and output of the power converter, excluding the faults of the components of the power stack: semiconductors, capacitors, and output filter. Physical faults include short-circuit and open-circuit conditions affecting different converter subsystems, whereas electrical disturbances correspond to abnormal operating conditions reflected through variations in electrical magnitudes and signal characteristics.

The simulated scenarios include the following:

- Short circuits between each grid phase and the ground, among two or even three phases, or a combination of them. These originate from the degradation of the wire insulation, animals making contact between multiple conductors, or even incidental connections between phases or with the ground by a tree branch.
- Short circuits between the DC-bus poles and the ground resulting from wire insulation degradation, as well as the presence of water from condensation or accidental spills.
- Grid-phase disconnections for various reasons, such as the destruction of the wires by a falling tree branch; the tripping of overload protections; blown fuses; improper connections; loose connections; and blackout conditions caused by the tripping of higher-level protections, transformer failures, or supply shutdowns.

- Grid impedance variations from other faults in the grid that do not affect the converter branch directly; reconfiguration of the grid because of maintenance or maneuvers; or the step-in/step-out of transformers, generators, and loads.
- Voltage amplitude variations, spikes, lags, frequency deviations, and phase-sequence alterations originating from switching operations; short circuits at certain points; lightning strikes; or imbalances between generation and consumption.
- Harmonic distortion and electrical noise perturbations resulting from the presence of other power converters, switching operations during maneuvers, lightning, electro-magnetic coupling between phases, or electric machines connected to the grid.
- Semiconductor blocking faults under open-state and closed-state conditions after the failure of the device driver due to a power supply fault or failures caused by high current or voltage, EMI, or overtemperature.

The simulation alternates between nominal operation and fault conditions in order to capture the variations introduced by each operating scenario over the monitored electrical magnitudes. This enables subsequent association between the acquired data and their corresponding operating conditions.

3.3. Data Acquisition and Dataset Generation

The acquired measurements include AC and DC voltages and currents, as well as conduction and switching losses associated with the IGBTs and diodes that compose the converter. The measurements are acquired using two complementary representations. First, RMS values are extracted to characterize the global operating behavior of the system. Second, instantaneous electrical measurements are sampled at the switching frequency, enabling the vectorized representation of complete grid cycles and preserving the transient behavior associated with the injected fault conditions.

To improve the generalization capability of the generated dataset and avoid phase-dependent bias, fault injection is performed at randomly selected temporal instants during converter operation. This strategy ensures that the generated disturbances affect different regions of the electrical waveforms and operating cycles, producing a more representative and structurally diverse dataset.

Finally, the generated measurements are encapsulated and transmitted through UDP to the proposed modular acquisition platform, where the information is processed and stored for subsequent analysis and supervised fault diagnosis tasks. The transmitted UDP packet contains the metadata required for measurement traceability together with the associated electrical values (see Figure 2). Specifically, each message includes the experiment identifier (`exp_id`), device identifier (`device_id`), measurement identifier (`mea_id`), fault category (`error_id`), and the corresponding measurement vector generated by the simulator (`measurement_values`).

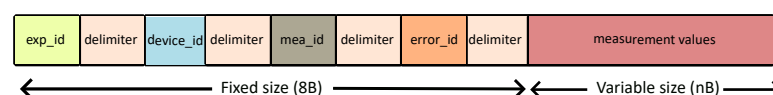


Figure 2. Typhoon format message.

After processing, the acquired measurements are stored in structured JSON format within the acquisition platform, following the representation shown in Listing 1. The stored structure preserves both the electrical measurements and the associated fault identifier (`id_error`), enabling traceability and monitoring of the converter operating conditions throughout the acquisition process, while also supporting the supervised training of the models employed for automatic fault identification.

Listing 1. Structured JSON representation of the electrical measurements stored by the acquisition platform.

```

{
  "_id": "...",                      /* Document identifier */
  "timestamp": "...",               /* Measurement timestamp */
  "device": "...",                 /* Device identifier */
  "id_mea": "...",                 /* Measurement identifier */
  "id_error": "...",               /* Fault identifier */
  "mea_value": [                   /* Measurement vector */
    value_1,
    value_2,
    value_3,
    ...,
    value_N
  ],
  "experiment_id": "...",          /* Experiment identifier */
  "store_timestamp": "..."       /* Storage timestamp */
}

```

3.4. Dataset Description

The generated dataset contains electrical measurements acquired under both nominal and faulty operating conditions of the analyzed power converter. Each acquired sample is associated with a specific fault category through the corresponding fault identifier (id_{error}), enabling supervised fault diagnosis and geometric fault-space analysis.

The dataset includes multiple measurement categories (id_{mea}) corresponding to different electrical and thermal variables generated by the simulator. Depending on the selected measurement type, the dimensionality of the generated measurement vectors (mea_{value}) varies significantly. While certain measurements are represented using low-dimensional RMS values, others correspond to high-frequency instantaneous electrical signals sampled at switching frequency, producing high-dimensional vector representations of complete grid cycles.

Table 1 summarizes the main characteristics of the generated dataset.

Table 1. Main characteristics of the generated dataset.

Dataset Characteristic	Value
Number of fault categories (id_{error})	48
Total number of labeled samples	673,207
Measurement categories (id_{mea})	10
Sampling frequency	Switching frequency
Storage format	Structured JSON

4. Proposed Methodology

The proposed methodology restructures the diagnostic label space according to the geometric separability of the selected measurement space before supervised machine learning is performed. First, the selected electrical measurements are projected into a lower-dimensional manifold using UMAP to characterize the geometric relationships among fault categories.

Based on the resulting manifold representation, pairwise inter-fault distances are computed to identify geometrically overlapping fault categories as well as faults exhibiting high similarity to nominal operating conditions. Subsequently, hierarchical clustering is applied to automatically group non-separable fault categories, generating a hybrid fault

space composed of clusters associated with nominal operation, multi-fault clusters, and isolated faults.

Finally, the transformed label space is used to train and evaluate several machine learning models, including Decision Tree and Random Forest, enabling the analysis of how geometric fault restructuring affects classification robustness, inter-fault confusion, and real-time deployment feasibility in edge-computing environments. To ensure reproducibility and facilitate the practical application of this methodology, the complete source code is publicly available in [37].

4.1. Geometric Fault-Space Analysis

To analyze the geometric separability of the fault space, the high-dimensional measurement samples are transformed into a lower-dimensional feature manifold using UMAP. The resulting manifold provides a compact representation of the geometric relationships between fault categories, facilitating the analysis of their relative separability while reducing the dimensional complexity of the original measurement space.

The objective of this stage is not to classify fault categories, but to determine which of them are reliably distinguishable from the selected measurements. The resulting low-dimensional representation is subsequently used to estimate inter-fault similarity, identify geometrically overlapping fault categories, and analyze the separability of the resulting fault space with respect to nominal operating conditions. In contrast to visualization-oriented applications of UMAP, the proposed methodology employs the obtained embedding space as the basis for quantitative distance analysis and automatic fault-space restructuring. This geometric analysis enables the identification of structurally ambiguous fault regions where different operating conditions exhibit similar measurement behavior, increasing the difficulty of reliable fault discrimination.

Figure 3 presents an illustrative synthetic example of fault-space overlapping in a UMAP projection. The example highlights the existence of geometrically ambiguous operating regions (faults 2-5 and 1-3-nominal), illustrating situations in which the selected measurements do not provide sufficient discriminative information to reliably distinguish between predefined fault categories. This motivates the evaluation of geometric separability before defining the diagnostic label space used for supervised learning.

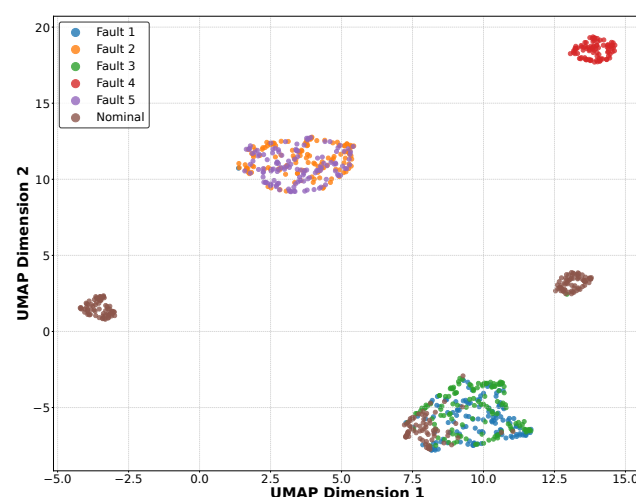


Figure 3. Geometric fault space example.

4.2. Inter-Fault Similarity Estimation

Once the low-dimensional embedding space is generated, pairwise distances between fault categories are computed in order to quantify their structural similarity within the transformed measurement space. For each pair of fault categories f_i and f_j , the proposed

methodology estimates a symmetric nearest-neighbor distance using the corresponding embedded samples.

More specifically, for each sample $x_k \in f_i$, the minimum Euclidean distance to any sample belonging to category f_j is computed within the embedding space:

$$d(x_k, f_j) = \min_{y \in f_j} \|x_k - y\|_2 \quad (1)$$

The obtained minimum distances are subsequently averaged across all samples from f_i in order to estimate a directed nearest-neighbor distance:

$$D(f_i \rightarrow f_j) = \frac{1}{|f_i|} \sum_{x_k \in f_i} d(x_k, f_j) \quad (2)$$

Finally, the symmetric inter-fault distance is obtained by averaging both directional distances:

$$D_{\text{sym}}(f_i, f_j) = \frac{D(f_i \rightarrow f_j) + D(f_j \rightarrow f_i)}{2} \quad (3)$$

Additionally, a directed distance analysis with respect to the nominal operating condition is performed in order to identify fault categories exhibiting strong geometric proximity to nominal operation.

$$D_{\text{nom}}(f_i) = D(f_i \rightarrow f_0) \quad (4)$$

where f_0 represents the nominal operating condition.

The computed distances are organized into an inter-fault distance matrix, which is subsequently used for geometric similarity analysis and hierarchical fault-space clustering. These distance analyses enable the identification of:

- Fault categories geometrically close to nominal operation
- Mutually overlapping faults
- Structurally similar fault categories
- Geometrically separable faults

Based on these geometric relationships, hierarchical clustering is applied to automatically group non-separable fault categories into shared fault clusters. Additionally, faults exhibiting strong proximity to nominal operation are incorporated into a shared nominal-operation cluster.

4.3. Hybrid Fault-Space Clustering

To adapt the diagnostic label space to the geometric separability of the selected measurement space, a two-stage clustering strategy is applied over the fault-space distance matrix. First, a specific analysis is performed with respect to the nominal operating condition. The distance between each fault category and the nominal operating regime is evaluated using the previously computed inter-fault distances. Faults categories whose distance to the nominal condition falls below a predefined threshold (τ_{normal}) are considered not reliably distinguishable from normal operation within the selected measurement space. These categories are grouped together with the nominal operating condition into a unified cluster. Based on the illustrative example shown in Figure 3, the nominal operating condition (brown points) appears distributed across multiple regions of the embedding space. Additionally, faults 1 and 3, blue and green points, respectively, exhibit strong geometric proximity to one of these nominal regions, indicating operating conditions that may become difficult to reliably separate from normal operation.

Subsequently, hierarchical clustering is applied to the remaining fault categories in order to identify groups of mutually overlapping faults. Clusters are generated according

to a distance-based threshold criterion ($\tau_{cluster}$), enabling the automatic discovery of fault groups that cannot be reliably separated within the embedding space. Only groups containing more than one fault category are considered valid fault clusters, while geometrically separable faults are preserved as independent classes. The clustering process is therefore not intended to perform unsupervised fault diagnosis. Instead, it is used to redefine the diagnostic label space according to the level of separability supported by the selected measurements before supervised learning is performed.

As a result, the proposed methodology generates a hybrid fault space composed of

- A nominal-operation cluster
- Multi-fault clusters
- Isolated fault categories

According to the geometric relationships previously illustrated in Figure 3, the proposed clustering strategy generates different operating groups within the hybrid fault space. Table 2 summarizes the resulting cluster organization obtained for the presented example.

Table 2. Illustrative example of the resulting hybrid fault-space organization obtained from the proposed clustering strategy.

Cluster Type	Associated Faults
Nominal-operation cluster	nominal, 1, 3
Multi-fault cluster	2, 5
Isolated fault category	4

4.4. Fault Classification and Edge-Oriented Evaluation

The hybrid fault space generated by the proposed methodology is subsequently used for supervised fault classification. In this stage, several machine learning models are trained and evaluated using the restructured fault labels in order to analyze how the proposed geometric fault-space organization affects classification performance under overlapping measurement conditions.

To evaluate the proposed methodology under different learning paradigms, the evaluation includes two representative learning paradigms:

- Decision Tree (DT), as an interpretable single-tree baseline.
- Random Forest (RF), as a bagging-based ensemble approach.

The evaluated models are trained offline using a resource-rich computing environment and subsequently deployed on resource-constrained edge devices for inference evaluation. This deployment-oriented strategy enables the analysis of the practical feasibility of the proposed methodology under realistic edge-computing conditions.

5. Experimental Validation of the Proposed Methodology

5.1. Experimental Setup

To evaluate the proposed methodology under different measurement dimensionalities, three representative electrical measurement categories were selected from the available measurement space, corresponding to $id_{mea} = 10$, $id_{mea} = 3$, and $id_{mea} = 2$. These measurement categories correspond to RMS three-phase grid voltages (3 components); instantaneous electrical measurements, including DC-link capacitor voltage and input current (40 components); and instantaneous three-phase grid currents (60 components), respectively. The measurements used in this study are publicly available in [38].

Prior to dimensionality reduction, all measurement vectors were normalized using standard score normalization. Subsequently, UMAP was applied using a two-dimensional embedding configuration ($n_{components} = 2$) with Euclidean distance as the similarity metric

and $n_{neighbors} = 30$. A fixed random seed ($random_{state} = 42$) was employed to ensure reproducibility of the generated embeddings.

Once the embedding space was obtained, inter-fault distances were estimated using the Euclidean distance between embedded samples. Hierarchical clustering was subsequently applied using average linkage clustering over the generated inter-fault distance matrix. Cluster generation was controlled through a distance threshold criterion ($\tau_{cluster} = 1.0$), while fault categories exhibiting strong proximity to nominal operation were identified using a nominal-distance threshold ($\tau_{normal} = 0.01$).

The evaluated machine learning models were configured using representative hyperparameter settings (see Table 3) in order to balance classification capability and computational complexity under embedded deployment conditions.

Table 3. Main hyperparameter configuration of the evaluated machine learning models.

Model	Parameter	Value
DT	Criterion	Gini
	Maximum depth	10
	Minimum samples split	2
	Minimum samples leaf	1
RF	Number of estimators	300
	Maximum features	$\sqrt{n_{features}}$

$n_{features}$ represents the dimensionality of the selected measurement.

To additionally evaluate the feasibility of the proposed methodology under embedded deployment conditions, inference experiments were conducted on a Raspberry Pi 4 Model B Rev 1.1 platform equipped with a quad-core ARM Cortex-A72 CPU running at 1.5 GHz and 4 GB RAM. The device operated under a Linux-based operating system. In order to obtain statistically representative inference measurements, each evaluated model was independently executed 10 times, and the reported results correspond to the mean and standard deviation obtained across all executions.

5.1.1. Classification Performance Metrics

To evaluate the fault diagnosis capability of the proposed methodology, several standard supervised classification metrics were considered, including accuracy (5), precision (6), recall (7), and F1-score (8).

Classification accuracy was computed as

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (5)$$

where TP , TN , FP , and FN represent true positives, true negatives, false positives, and false negatives, respectively.

Precision was estimated as

$$Precision = \frac{TP}{TP + FP} \quad (6)$$

which measures the proportion of correctly predicted positive samples among all predicted positive samples.

Recall was computed as

$$Recall = \frac{TP}{TP + FN} \quad (7)$$

representing the capability of the model to correctly identify positive fault samples.

Finally, the F1-score was calculated as the harmonic mean between precision and recall:

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad (8)$$

5.1.2. Edge-Oriented Inference Evaluation

Beyond classification performance, an edge-oriented evaluation was conducted to analyze the computational feasibility of the evaluated models under resource-constrained embedded deployment scenarios. In particular, average inference latency (9), throughput (10), and model size were measured in order to assess the suitability of the proposed methodology for real-time fault diagnosis applications.

Average inference latency was computed as

$$L_{avg} = \frac{T_{total}}{N} \quad (9)$$

where T_{total} represents the total inference time required to process N input samples.

Similarly, throughput was estimated as

$$TH = \frac{N}{T_{total}} \quad (10)$$

which represents the number of inferences processed per second.

5.2. Geometric Fault-Space Evaluation

5.2.1. Low-Dimensional Representation of Fault Distributions

Figure 4 shows the low-dimensional manifolds generated through UMAP for the three analyzed measurements. The obtained representations reveal notable differences in the geometric organization of the fault distributions, depending both on the dimensionality and on the physical nature of the associated measurement vectors.

For the measurement associated with $id_{mea} = 10$ (Figure 4a), corresponding to the lowest-dimensional measurement space, the nominal operating condition (blue points) exhibits a broad and highly dispersed distribution across the manifold. In contrast, several fault categories are concentrated around the central region, while others appear as sparse or localized groups. This structure indicates that, for this measurement, the nominal condition occupies a large portion of the transformed space, which increases the likelihood of overlap with certain fault categories and reduces their separability from normal operation.

In the case of the measurement associated with $id_{mea} = 3$ (Figure 4b), the generated manifold exhibits a substantially higher point density and a more spatially dispersed organization compared to the representation obtained for $id_{mea} = 10$. The nominal operating condition (blue points) forms multiple elongated and fragmented regions distributed throughout the embedding space, while several fault categories appear concentrated in localized areas with clearer geometric boundaries. Although partial overlap between certain faults and the nominal condition is still present, the obtained representation reveals a richer and more heterogeneous geometric structure, characterized by a larger number of embedded regions and a wider spatial distribution of the fault categories.

Finally, for the measurement associated with $id_{mea} = 2$ (Figure 4c), the generated manifold exhibits the highest geometric complexity among the analyzed measurements. The nominal operating condition (blue points) occupies a large portion of the embedding space through highly fragmented and spatially distributed regions, while multiple fault categories appear concentrated in localized areas near the center of the manifold. Although several faults form partially distinguishable groups, substantial overlap with the nomi-

nal distribution is still observed, indicating reduced geometric separability for multiple operating conditions.

Overall, the obtained manifolds indicate that the dimensionality of the measurement vectors affects the structural organization of the fault space. At the same time, the persistence of partially overlapping regions and fault categories located near nominal operating conditions motivates the subsequent inter-fault similarity estimation and fault-space restructuring stages, since these geometric ambiguities may limit the ability of conventional classifiers to reliably distinguish between certain fault conditions and normal operation and may also introduce confusion among geometrically similar fault categories.

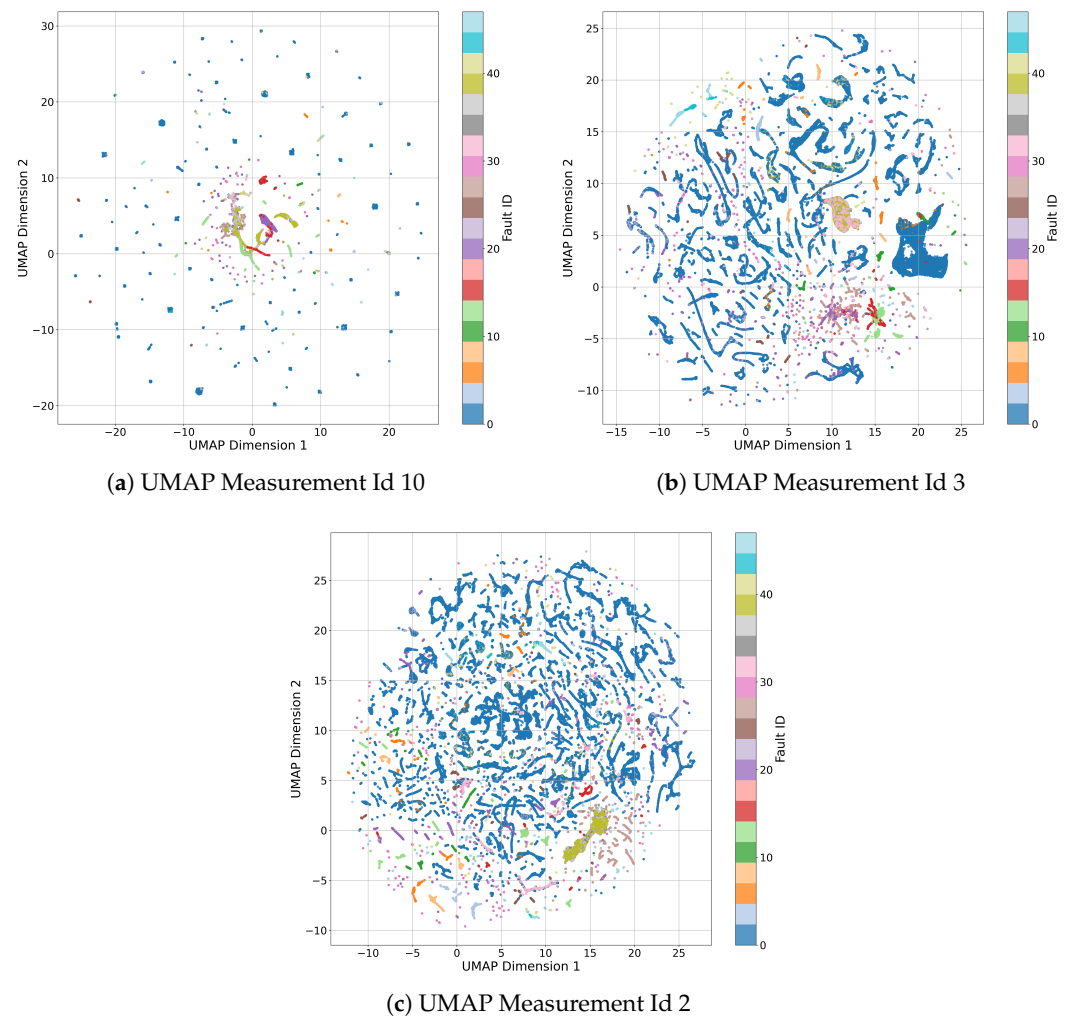


Figure 4. UMAP representations for different measurement configurations.

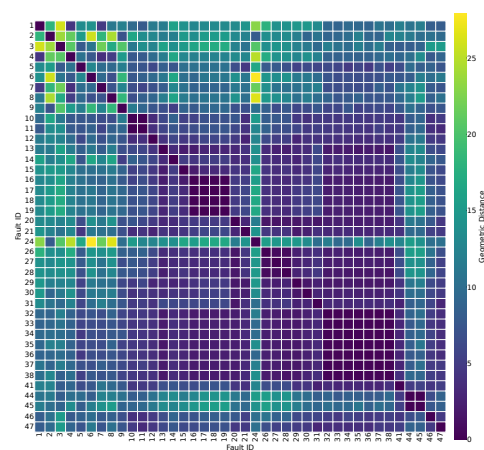
5.2.2. Inter-Fault Distance Analysis

To further quantify the structural relationships between fault categories, pairwise inter-fault distances were computed over the UMAP embedding space. Figure 5 represents the resulting geometric distance matrix using a color scale ranging from dark purple to yellow, where darker regions indicate low inter-fault distances and therefore strong structural similarity, while brighter regions correspond to larger geometric separations between fault distributions. This representation enables the identification of overlapping fault families, isolated operating conditions, and geometrically ambiguous fault regions that may reduce the separability of conventional supervised classifiers.

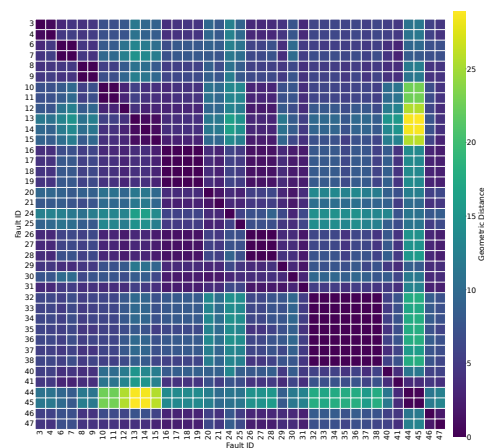
For the measurement associated with $id_{mea} = 10$ (three-phase RMS voltages), the inter-fault distance matrix shown in Figure 5a reveals the coexistence of both highly overlapping

and strongly separable fault categories within the analyzed measurement space. Several compact low-distance regions can be observed, particularly for fault pairs 10–11 and 44–45, as well as for the fault families spanning categories 17–19 and 32–38. These structures indicate the presence of fault distributions exhibiting strong geometric similarity and limited geometric separability. In contrast, fault 24 exhibits consistently large distances with respect to most remaining categories, suggesting a substantially more isolated geometric behavior.

For the measurement associated with $id_{mea} = 3$ (instantaneous DC voltage and current), the inter-fault distance matrix shown in Figure 5b reveals the coexistence of compact fault families and strongly isolated fault categories within the transformed feature space. Several low-distance regions can be clearly identified, particularly for fault groups 16–19, 26–28, and especially for categories 32–38, which form one of the most cohesive regions within the matrix. In addition, some neighboring fault categories exhibit intermediate-distance relationships, suggesting the presence of partially connected fault regions. In contrast, certain fault categories remain substantially separated from the remaining classes. This behavior is especially noticeable for fault 44, whose associated row and column contain consistently large distance values, indicating high intrinsic separability.

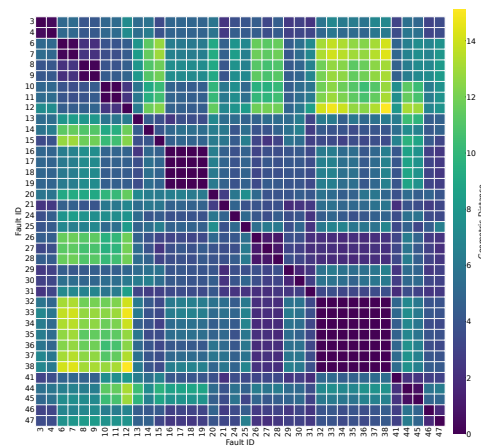


(a) Fault distance Id 10



(b) Fault distance Id 3

Figure 5. Cont.



(c) Fault distance Id 2

Figure 5. Fault distance heatmaps for the analyzed fault identifiers.

For the measurement associated with $id_{mea} = 2$ (three-phase currents), the inter-fault distance matrix shown in Figure 5c reveals a highly heterogeneous geometric organization characterized by the coexistence of compact fault groups and strongly separated operating conditions. Several large low-distance regions can be observed, particularly for the fault families spanning categories 15–18 and 33–38, indicating substantial geometric similarity and reduced separability among these operating conditions. Additional localized low-distance structures are also visible around fault categories 6–8 and 43–45. In contrast, multiple fault families exhibit consistently larger inter-fault distances with respect to the remaining categories, reflecting a more fragmented and spatially distributed fault space compared to the previous measurements.

Overall, the obtained inter-fault distance matrices reveal that the analyzed measurement spaces exhibit heterogeneous geometric organizations composed of both strongly overlapping fault families and highly separable operating conditions. While certain faults remain geometrically isolated and therefore intrinsically easier to distinguish, multiple fault groups consistently present reduced inter-fault distances, as illustrated by the top-10 most similar fault pairs reported in Tables 4–6. These low-distance regions indicate the presence of structurally ambiguous fault distributions whose geometric proximity may increase inter-fault confusion during supervised classification. Consequently, the obtained results support the need for the hybrid fault-space restructuring strategy proposed in this work, whose objective is to reduce geometric ambiguity and improve the robustness of the subsequent fault diagnosis stage.

Table 4. Top-10 similar fault pairs for $id_{mea} = 10$.

A	B	Dist.
10	11	0.007021
17	18	0.031530
18	19	0.033334
17	19	0.034433
44	45	0.055953
16	18	0.176792
16	19	0.180282
16	17	0.180604
35	38	0.197196
32	37	0.217271

Table 5. Top-10 similar fault pairs for $id_{mea} = 3$.

A	B	Dist.
18	19	0.006901
17	18	0.007232
17	19	0.009003
26	27	0.025023
32	36	0.027958
32	34	0.029142
34	36	0.029569
32	33	0.030367
33	36	0.030646
33	34	0.032180

Table 6. Top-10 similar fault pairs for $id_{mea} = 2$.

A	B	Dist.
17	18	0.007026
18	19	0.025484
36	37	0.026009
32	37	0.026885
37	38	0.027063
32	36	0.027978
36	38	0.028618
35	37	0.028789
32	33	0.028818
33	37	0.029097

5.2.3. Analysis of Faults Close to Nominal Operation

Following the hybrid fault-space clustering defined in the proposed methodology, this subsection investigates those fault categories whose embedded representations remain geometrically close to the nominal operating condition within the UMAP space. According to the proposed strategy, faults whose directed distance to the nominal manifold satisfies $d_{fault \rightarrow normal} < \tau_{normal}$ are automatically merged with the healthy operating condition. In the present study, the nominal merging threshold was fixed to $\tau_{normal} = 0.01$.

The obtained results show different geometric relationships with respect to nominal operation across the analyzed measurements. While certain fault categories remain relatively separated from the healthy operating manifold, several others satisfy the nominal merging criterion, indicating reduced geometric separability from healthy behavior. The fault categories associated with the nominal operating condition for each analyzed measurement are summarized in Table 7. Fault categories not satisfying the nominal merging criterion remain available for the subsequent fault-space clustering analysis, where they may either preserve isolated geometric behavior or be grouped according to their pairwise geometric distances within the transformed feature space.

Table 7. Fault categories geometrically associated with nominal operation based on the directed distance analysis.

id_{mea}	Faults Merged with Nominal Operation
10	22, 23, 25, 39, 40, 42, 43
3	1, 2, 5, 22, 23, 39, 42, 43
2	1, 2, 5, 22, 23, 39, 40, 42, 43

5.2.4. Automatic Discovery of Overlapping Fault Groups

As part of the proposed hybrid fault-space strategy, hierarchical clustering was applied over the inter-fault distance matrices in order to identify groups of fault categories exhibiting strong structural similarity within the transformed feature spaces. According to the proposed methodology, fault categories whose pairwise geometric distances satisfy $d_{i,j} < \tau_{cluster}$ are automatically grouped into the same hybrid fault cluster. In the present study, the clustering threshold was fixed to $\tau_{cluster} = 1.0$.

Table 8 summarizes the overlapping fault groups automatically identified for each analyzed measurement, together with the fault categories that remained geometrically isolated after the clustering stage. Interestingly, most of the detected fault groups consist of only two fault categories, suggesting that the proposed restructuring strategy performs localized ambiguity reduction while preserving the separability of the remaining fault space.

Table 8. Automatically discovered overlapping fault groups for each analyzed measurement configuration.

id_mea	Detected Fault Groups	Isolated Fault Categories
10	{10, 11}, {16, 17, 18, 19}, {26, 27, 28}, {32, 33, 34, 35, 36, 37, 38}, {44, 45}	{1, 2, 3, 4, 5, 6, 7, 8, 9, 12, 13, 14, 15, 20, 21, 24, 29, 30, 31, 41, 46, 47}
3	{3, 4}, {6, 7}, {8, 9}, {10, 11}, {13, 14, 15}, {16, 17, 18, 19}, {26, 27, 28}, {32, 33, 34, 35, 36, 37, 38, 44, 45}	{12, 20, 21, 24, 25, 29, 30, 31, 40, 41, 46, 47}
2	{3, 4}, {6, 7}, {8, 9}, {10, 11}, {16, 17, 18, 19}, {26, 27, 28}, {32, 33, 34, 35, 36, 37, 38}, {44, 45}	{12, 13, 14, 15, 20, 21, 24, 25, 29, 30, 31, 41, 46, 47}

5.3. Classification Performance Evaluation

This subsection evaluates the impact of the proposed fault-space restructuring methodology on the supervised fault diagnosis stage. For this purpose, two diagnostic settings are compared. The first one corresponds to the original fine-grained fault taxonomy, where classifiers are trained to discriminate all predefined fault categories without applying the proposed restructuring process. This setting is used as a baseline to quantify the difficulty of directly separating fault categories that may exhibit geometric overlap within the selected measurement space.

The second setting corresponds to the hybrid diagnostic label space generated by the proposed methodology. Thus, the results obtained after restructuring should be interpreted as macro-diagnostic performance over the redefined label space, rather than as evidence of improved fine-grained discrimination among the original fault categories merged into the same cluster.

Consequently, the comparison between both settings is intended to assess whether adapting the diagnostic label space to the separability supported by the selected measurements improves macro-diagnostic robustness under geometric ambiguity. This interpretation explicitly acknowledges the trade-off introduced by the proposed methodology: improved robustness and reduced inter-fault ambiguity at the cluster level, at the expense of lower fault-isolation granularity for the merged categories.

The improvement obtained after fault-space restructuring can be qualitatively explained by the relationship between class geometry and supervised decision boundaries. In the original fine-grained label space, classifiers are forced to discriminate between fault categories whose measurement distributions are geometrically close or partially overlapping. Under these conditions, the learned decision boundaries may become fragmented or unstable within the selected measurement space. By grouping these ambiguous categories into hybrid diagnostic labels, the proposed methodology reduces the need for unreliable sepa-

rations and produces target classes that are more consistent with the geometric structure observed in the data. Consequently, the observed performance gains are mainly attributed to a reduction in inter-fault ambiguity and to a better alignment between the supervised classification task and the separability supported by the available measurements.

For all experiments, the dataset was divided using an 80/20 train-test split. The classification stage was evaluated using the previously described Decision Tree and Random Forest models over the analyzed measurements. The employed target categories include the nominal clusters associated with healthy operating regions, the hybrid fault groups generated through the fault-space restructuring process, and the isolated fault categories that preserved sufficient geometric separability within the transformed measurement space, as summarized in Tables 7 and 8, respectively.

To facilitate comparative analysis, the obtained classification results are organized according to the analyzed measurements (id_{mea}). For each measurement, the performance of the evaluated models is reported both with and without the application of the proposed fault-space restructuring process, enabling direct evaluation of the impact of the generated nominal clusters and hybrid fault groups on the subsequent supervised fault diagnosis stage.

5.3.1. Results for Three-Phase RMS Voltages ($id_{mea} = 10$)

The obtained results reveal a substantial improvement in classification performance after applying the proposed fault-space restructuring process. As shown in Table 9, the use of hybrid fault representations considerably provides diagnostic target categories that are more consistent with the separability supported by the selected measurements, leading to substantial gains in precision, recall, and F1-score. In particular, accuracy increases from 0.88 to 0.98 for Decision Tree and from 0.90 to 0.99 for Random Forest when compared to the original fault-space representation (see Table 10), which exhibits significant inter-fault confusion and limited class separability, especially for the Decision Tree classifier. Similarly, the F1-score improves from 0.53 to 0.91 for Decision Tree and from 0.67 to 0.98 for Random Forest, confirming the effectiveness of the proposed methodology reducing classification ambiguity among geometrically overlapping fault categories. Detailed per-class classification results are additionally provided in Appendix A (Tables A1 and A2).

Table 9. Classification results for $id_{mea} = 10$ using hybrid fault-space clustering.

Model	Accuracy	Precision	Recall	F1-Score
DT	0.98	0.93	0.90	0.91
RF	0.99	0.99	0.98	0.98

Table 10. Classification results for $id_{mea} = 10$ without fault-space clustering.

Model	Accuracy	Precision	Recall	F1-Score
DT	0.88	0.58	0.54	0.53
RF	0.90	0.69	0.67	0.67

5.3.2. Results for Instantaneous DC Voltage and Current ($id_{mea} = 3$)

The obtained results again demonstrate the effectiveness of the proposed fault-space restructuring methodology for improving the separability of geometrically ambiguous operating conditions associated with $id_{mea} = 3$. As shown in Table 11, the use of hybrid fault-space clustering substantially improves the classification performance of both analyzed models. In particular, the F1-score increases from 0.41 to 0.65 for Decision Tree and from 0.62 to 0.95 for Random Forest when compared to the original fault-space representation (see Table 12), which exhibits considerable inter-fault confusion and limited class

separability. Similarly, precision improves from 0.52 to 0.79 for Decision Tree and from 0.66 to 0.97 for Random Forest, indicating a substantial reduction in inter-fault misclassification after applying the proposed restructuring process. These results indicate that the proposed restructuring process reduces the classification ambiguity caused by geometrically overlapping fault categories, thereby enabling more robust macro-diagnostic classification. Detailed per-class classification results are additionally provided in Appendix A (Tables A3 and A4).

Table 11. Classification results for $id_{mea} = 3$ using hybrid fault-space clustering.

Model	Accuracy	Precision	Recall	F1-Score
DT	0.93	0.79	0.61	0.65
RF	0.99	0.97	0.93	0.95

Table 12. Classification results for $id_{mea} = 3$ without fault-space clustering.

Model	Accuracy	Precision	Recall	F1-Score
DT	0.82	0.52	0.40	0.41
RF	0.88	0.66	0.60	0.62

5.3.3. Results for Three-Phase Currents ($id_{mea} = 2$)

The obtained results for $id_{mea} = 2$ further confirm the effectiveness of the proposed fault-space restructuring methodology for reducing inter-fault ambiguity during the supervised classification stage. Despite the relatively high dimensionality associated with this measurement configuration (60 dimensions), the use of hybrid fault-space clustering substantially improves the classification performance of both analyzed models, as shown in Table 13. In particular, the F1-score increases from 0.42 to 0.57 for Decision Tree and from 0.65 to 0.99 for Random Forest when compared to the original fault-space representation (Table 14), which still exhibits considerable geometric overlap and limited class separability despite the higher dimensionality of the analyzed measurement space. Similarly, precision also improves substantially, increasing from 0.51 to 0.82 for Decision Tree and from 0.68 to 1.00 for Random Forest, indicating a significant reduction in inter-fault misclassification after applying the proposed restructuring process. Detailed per-class classification results are additionally provided in Appendix A (Tables A5 and A6).

Table 13. Classification results for $id_{mea} = 2$ using hybrid fault-space clustering.

Model	Accuracy	Precision	Recall	F1-Score
DT	0.92	0.82	0.52	0.57
RF	1.0	1.0	0.98	0.99

Table 14. Classification results for $id_{mea} = 2$ without fault-space clustering.

Model	Accuracy	Precision	Recall	F1-Score
DT	0.83	0.51	0.42	0.42
RF	0.89	0.68	0.64	0.65

5.4. Practical Interpretation of Hybrid Fault Clusters

From a practical maintenance perspective, a hybrid fault cluster should be interpreted as a candidate fault group rather than as a single isolated fault. When the proposed methodology assigns a sample to a hybrid cluster, the selected measurements indicate that the faults contained in that cluster exhibit similar electrical behavior and cannot be

reliably distinguished at the current diagnostic level. This output is still actionable because it narrows the diagnostic search space and informs the operator about the subset of fault categories that require further inspection.

In such cases, maintenance decisions can be supported through a hierarchical diagnostic workflow. First, the proposed method provides a robust macro-diagnostic result by identifying the most plausible fault group. Then, fine-grained isolation may be performed using additional measurements, complementary electrical tests, expert inspection, or a secondary diagnostic model specifically trained to discriminate among the faults contained in the detected cluster. Therefore, the proposed method does not replace subsequent fine-grained analysis when several faults are merged, but provides a reliable first-stage diagnosis that guides maintenance actions and reduces unnecessary inspection of unrelated fault categories.

5.5. Edge Deployment Feasibility Analysis

In addition to classification performance, practical fault diagnosis systems must satisfy real-time execution and computational efficiency requirements, particularly in edge-oriented deployment scenarios involving resource-constrained devices.

To evaluate the practical feasibility of the proposed methodology, this subsection analyzes the computational characteristics of the evaluated classifiers using the generated hybrid fault-space representations. In particular, Decision Tree and Random Forest are evaluated in terms of inference latency, throughput, and model size across the analyzed measurement configurations. The models were trained offline and subsequently deployed on a Raspberry Pi platform in order to evaluate their execution behavior under realistic edge-oriented deployment conditions.

This analysis aims to assess whether the proposed methodology remains compatible with real-time edge-oriented deployment constraints despite the additional fault-space restructuring stage.

Inference Latency and Throughput Analysis

Table 15 summarizes the computational performance obtained for the evaluated classifiers under the analyzed measurements, including inference latency, throughput, and model size metrics. The obtained results reveal clear differences in the computational characteristics of the evaluated classifiers after applying the proposed fault-space restructuring process. Decision Tree consistently achieves the lowest inference latency and memory footprint across all analyzed measurement types, reaching inference times close to 0.3 ms and throughput values above 2600 inferences per second. In contrast, Random Forest introduces substantially higher computational costs due to the use of multiple estimators. In particular, it exhibits significantly larger model sizes, reaching up to 299 MB for certain measurements, together with considerably higher inference latencies compared to Decision Tree. Additionally, within the evaluated scope, variations in measurement dimensionality introduced only minor differences in the resulting computational performance.

Table 15. Edge-oriented computational performance obtained for the evaluated models under the analyzed measurement configurations.

id_mea	Model	L_{avg} (ms)	TH (inf/s)	Size (MB)
2	DT	0.373	2683.590	0.135
	RF	112.20	8.910	296.67
3	DT	0.368	2720.520	0.126
	RF	112.15	8.920	299.09
10	DT	0.319	3136.580	0.087
	RF	109.584	9.130	106.22

Despite these differences, both evaluated models remain compatible with edge-oriented monitoring scenarios requiring millisecond-level inference times. Overall, the obtained results indicate that the proposed restructuring methodology can improve fault diagnosis robustness while preserving computational characteristics compatible with practical edge-oriented deployment conditions.

6. Limitations and Research Directions

Several limitations of the proposed methodology should be considered.

First, the generated dataset was obtained under simulated operating conditions. Although the employed simulation environment enables controlled fault injection and high-frequency measurement acquisition, real industrial environments may introduce additional sources of variability, measurement noise, and hardware-related uncertainties not fully represented in the current study. Additional validation with experimental measurements obtained from real power electronic converters under practical operating conditions is therefore required.

Second, the proposed fault-space restructuring strategy depends on both the geometric properties of the generated embedding space and the selected clustering thresholds. Consequently, different UMAP configurations, threshold values, or measurement categories may affect the resulting fault organization and cluster generation. Additionally, the experimental validation was limited to three representative measurement categories selected from the available measurement space, which may not fully capture the complete variability of the monitored electrical and thermal measurements. In addition, data-driven or adaptive parameter selection strategies should be explored in order to reduce the dependence on manually fixed configuration values.

Third, the proposed hybrid clustering strategy improves diagnostic robustness by redefining the target label space when geometrically overlapping fault categories are detected. Consequently, the reported performance improvements should be interpreted within the resulting hybrid diagnostic label space, rather than as evidence of improved fine-grained discrimination among all original fault categories. When several faults are merged into the same hybrid cluster, additional measurements or a subsequent fine-grained analysis may be required to isolate the individual fault. A natural extension of this work is the development of hierarchical diagnostic schemes in which hybrid clusters provide a first-stage macro-diagnostic output, followed by cluster-specific models or complementary measurements for fine-grained fault isolation.

Furthermore, the experimental validation was limited to three representative measurement dimensionalities, obtained from three selected measurement types of the available dataset. Although these cases cover low-dimensional RMS measurements and higher-dimensional instantaneous waveform representations, they do not exhaust the complete range of monitored variables or possible measurement combinations. Therefore, the reported results should be interpreted within the scope of the analyzed measurement dimensionalities. A natural extension of this work is to evaluate the proposed methodology using additional measurement types, as well as intermediate and higher-dimensional representations, in order to assess how measurement dimensionality and signal selection affect geometric separability, diagnostic robustness, and computational cost.

Finally, the edge-oriented evaluation conducted in this work focuses on inference latency, throughput, and model size under embedded deployment conditions. Continuous real-time deployment aspects such as online adaptation, long-term execution stability, and energy consumption were not considered within the scope of the present study. Furthermore, the experimental validation was limited to Decision Tree and Random Forest classifiers, while alternative lightweight deep-learning architectures may exhibit different performance and computational trade-offs.

7. Conclusions

The proposed methodology was experimentally validated using multiple electrical measurements exhibiting substantially different dimensionalities and geometric organizations, ranging from low-dimensional RMS measurements to high-dimensional instantaneous waveform representations. The obtained results consistently demonstrated that the proposed fault-space restructuring strategy substantially improves supervised fault diagnosis performance across the three analyzed measurement spaces.

The conducted experiments revealed that the primary limitation of the analyzed fault diagnosis problem is not necessarily associated with the supervised learning models themselves, but rather with the intrinsic geometric organization of the fault space. In this context, the proposed restructuring methodology significantly reduced inter-fault ambiguity and improved class separability under geometrically overlapping operating conditions, enabling substantially more robust supervised classification boundaries. Ultimately, these performance gains validate the shift toward a robust macro-diagnostic paradigm rather than an advancement in fine-grained fault isolation. By consolidating overlapping categories, the methodology successfully mitigates ambiguous decision boundaries, resulting in highly dependable classifications within the hybrid label space.

Additionally, the obtained results demonstrated that increasing the dimensionality of the analyzed measurements does not necessarily eliminate geometric fault overlap. Despite evaluating measurement spaces containing up to 60 dimensions, several fault categories remained difficult to separate. This finding highlights the critical importance of prior geometric fault-space analysis during the diagnostic design stage, as the proposed methodology improved F1-scores by up to 72% compared with classification without fault-space restructuring in the evaluated scenarios.

The proposed methodology was compatible with the two evaluated supervised learning models. Although both Decision Tree and Random Forest classifiers benefited from the restructuring process, the results show that lightweight machine learning models can achieve strong fault diagnosis performance in the evaluated measurements when the underlying fault space is properly reorganized.

Finally, the edge-oriented experiments conducted on a Raspberry Pi 4 platform confirmed the practical feasibility of deploying the proposed methodology under constrained computational environments. In particular, the lightweight Decision Tree classifier achieved real-time inference performance with sub-millisecond latency (as low as 0.32 ms) and high inference throughput, while the Random Forest model remained executable on the embedded platform at the cost of higher latency and memory requirements. These results demonstrate the suitability of the proposed approach for embedded and edge-based fault diagnosis applications, especially when combined with lightweight classifiers.

Author Contributions: Conceptualization, J.M.G.-C., A.M.A., J.I.L. and A.L.-C.; data curation J.M.G.-C., A.M.A. and A.L.-C., methodology J.M.G.-C. and A.M.A.; software, J.M.G.-C. and A.L.-C.; formal analysis, J.M.G.-C., A.M.A. and J.I.L.; research, J.M.G.-C., A.M.A. and A.L.-C.; resources, J.M.G.-C. and A.M.A.; writing—original draft preparation, J.M.G.-C., A.M.A. and A.L.-C.; writing—review and editing, J.M.G.-C., A.M.A., J.I.L. and R.P.; visualization, J.M.G.-C. and A.M.A.; supervision, J.M.G.-C., A.M.A., J.I.L. and R.P.; funding acquisition A.M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been supported by Ministerio de Ciencia e Innovación under the project PID2023-152292OB-I00 as well as “UE-Ministerio de Hacienda y Función Publica-Fondos Europeos-Junta de Andalucía-Consejería de Universidad, Investigación e Innovación project 2024-31824-1828102303”, Shanghai Partner Research Program under Grant 25HB2705900.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Detailed Classification Reports

Appendix A.1. $id_{mea} = 10$

Table A1. Detailed per-class classification performance for $id_{mea} = 10$ under the proposed hybrid fault-space representation.

Class	DT			RF			Support
	Prec.	Rec.	F1	Prec.	Rec.	F1	
Cluster 0	0.99	1.00	1.00	1.00	1.00	1.00	8797
Cluster 1	1.00	1.00	1.00	1.00	1.00	1.00	128
Cluster 2	1.00	1.00	1.00	1.00	1.00	1.00	272
Cluster 3	0.99	0.99	0.99	0.99	0.99	0.99	191
Cluster 4	0.92	0.83	0.88	0.99	0.95	0.97	437
Cluster 5	0.98	0.97	0.97	0.99	1.00	1.00	128
Error 1	1.00	1.00	1.00	1.00	1.00	1.00	71
Error 12	0.80	1.00	0.89	1.00	1.00	1.00	84
Error 13	0.96	1.00	0.98	0.97	1.00	0.99	134
Error 14	0.99	0.96	0.98	1.00	1.00	1.00	126
Error 15	1.00	0.94	0.97	1.00	1.00	1.00	137
Error 2	1.00	1.00	1.00	1.00	1.00	1.00	66
Error 20	0.83	0.71	0.77	1.00	0.97	0.99	70
Error 21	0.44	0.88	0.59	1.00	1.00	1.00	64
Error 24	1.00	0.98	0.99	1.00	1.00	1.00	65
Error 29	0.90	0.45	0.60	0.95	1.00	0.98	62
Error 3	1.00	0.98	0.99	1.00	1.00	1.00	62
Error 30	0.98	0.65	0.78	1.00	1.00	1.00	65
Error 31	0.58	0.52	0.55	0.93	0.91	0.92	58
Error 4	1.00	1.00	1.00	1.00	1.00	1.00	69
Error 41	0.77	0.74	0.75	0.97	0.86	0.91	66
Error 46	0.86	0.76	0.81	1.00	0.76	0.86	66
Error 47	0.93	0.92	0.92	0.98	0.92	0.95	71
Error 5	1.00	1.00	1.00	1.00	1.00	1.00	73
Error 6	1.00	1.00	1.00	1.00	1.00	1.00	69
Error 7	1.00	1.00	1.00	1.00	1.00	1.00	68
Error 8	1.00	1.00	1.00	1.00	1.00	1.00	72
Error 9	1.00	1.00	1.00	1.00	1.00	1.00	63

Table A2. Detailed per-class classification performance for $id_{mea} = 10$ without hybrid fault-space restructuring.

Class	DT			RF			Support
	Prec.	Rec.	F1	Prec.	Rec.	F1	
Error 0	0.94	1.00	0.97	0.95	0.99	0.97	8330
Error 1	1.00	1.00	1.00	1.00	1.00	1.00	71
Error 10	0.82	0.82	0.82	0.84	0.79	0.82	68
Error 11	0.80	0.80	0.80	0.78	0.84	0.81	61
Error 12	0.88	0.27	0.42	0.99	1.00	0.99	84
Error 13	0.99	0.93	0.96	0.99	0.99	0.99	133
Error 14	0.98	0.98	0.98	0.98	1.00	0.99	126
Error 15	1.00	0.99	0.99	0.99	1.00	0.99	137

Table A2. Cont.

Class	DT			RF			Support
	Prec.	Rec.	F1	Prec.	Rec.	F1	
Error 16	0.61	0.23	0.33	0.34	0.32	0.33	74
Error 17	0.50	0.06	0.10	0.26	0.32	0.28	69
Error 18	0.27	0.93	0.42	0.17	0.18	0.18	68
Error 19	0.25	0.02	0.03	0.25	0.19	0.22	62
Error 2	1.00	1.00	1.00	1.00	1.00	1.00	66
Error 20	0.61	0.73	0.66	0.89	0.97	0.93	70
Error 21	0.75	0.23	0.36	1.00	0.98	0.99	64
Error 22	0.00	0.00	0.00	0.25	0.10	0.14	72
Error 23	0.00	0.00	0.00	0.25	0.07	0.11	60
Error 24	1.00	1.00	1.00	1.00	1.00	1.00	65
Error 25	0.00	0.00	0.00	1.00	0.97	0.99	71
Error 26	1.00	0.26	0.41	0.93	1.00	0.96	62
Error 27	0.71	0.61	0.66	1.00	0.92	0.96	66
Error 28	0.46	0.90	0.61	0.94	0.98	0.96	63
Error 29	0.68	0.42	0.52	0.95	1.00	0.98	62
Error 3	0.98	1.00	0.99	0.98	1.00	0.99	62
Error 30	0.97	0.48	0.64	0.97	1.00	0.98	65
Error 31	0.78	0.50	0.61	0.89	0.88	0.89	58
Error 32	0.00	0.00	0.00	0.13	0.13	0.13	62
Error 33	0.14	0.46	0.21	0.16	0.13	0.15	69
Error 34	0.12	0.03	0.05	0.19	0.19	0.19	63
Error 35	0.00	0.00	0.00	0.12	0.12	0.12	58
Error 36	0.20	0.02	0.03	0.20	0.18	0.19	65
Error 37	0.00	0.00	0.00	0.24	0.16	0.20	61
Error 38	0.11	0.20	0.14	0.07	0.07	0.07	60
Error 39	0.00	0.00	0.00	0.28	0.07	0.12	67
Error 4	1.00	1.00	1.00	1.00	1.00	1.00	69
Error 40	0.00	0.00	0.00	0.11	0.03	0.05	70
Error 41	0.81	0.77	0.79	0.91	0.79	0.85	66
Error 42	0.00	0.00	0.00	0.20	0.08	0.11	62
Error 43	0.00	0.00	0.00	0.28	0.08	0.12	64
Error 44	0.96	0.82	0.88	0.91	0.89	0.90	65
Error 45	0.82	0.94	0.87	0.89	0.90	0.90	62
Error 46	0.72	0.73	0.72	1.00	0.73	0.84	66
Error 47	0.98	0.86	0.92	0.97	0.94	0.96	71
Error 5	1.00	1.00	1.00	1.00	1.00	1.00	73
Error 6	0.91	1.00	0.95	1.00	1.00	1.00	69
Error 7	1.00	1.00	1.00	1.00	1.00	1.00	68
Error 8	0.28	1.00	0.43	1.00	1.00	1.00	72
Error 9	1.00	1.00	1.00	1.00	1.00	1.00	63

Appendix A.2. $id_{mea} = 3$ Table A3. Detailed per-class classification performance for $id_{mea} = 3$ under the proposed hybrid fault-space representation.

Class	DT			RF			Support
	Prec.	Rec.	F1	Prec.	Rec.	F1	
Cluster 0	0.93	0.99	0.96	0.99	1.00	0.99	34,801
Cluster 1	0.92	0.99	0.95	1.00	1.00	1.00	535
Cluster 2	0.99	0.71	0.83	1.00	1.00	1.00	533
Cluster 3	1.00	0.13	0.23	1.00	1.00	1.00	526

Table A3. Cont.

Class	DT			RF			Support
	Prec.	Rec.	F1	Prec.	Rec.	F1	
Cluster 4	1.00	0.75	0.86	1.00	0.99	0.99	508
Cluster 5	0.91	0.85	0.88	0.99	1.00	1.00	980
Cluster 6	0.92	0.93	0.93	1.00	0.99	1.00	1284
Cluster 7	0.90	0.87	0.88	1.00	1.00	1.00	954
Cluster 8	0.98	0.95	0.97	0.99	0.99	0.99	1799
Cluster 9	0.99	1.00	1.00	1.00	1.00	1.00	392
Error 12	0.70	0.95	0.80	1.00	1.00	1.00	280
Error 20	0.97	0.20	0.34	1.00	0.94	0.97	340
Error 21	0.00	0.00	0.00	1.00	0.95	0.97	332
Error 24	1.00	0.32	0.49	1.00	0.99	0.99	309
Error 25	0.00	0.00	0.00	0.99	0.99	0.99	330
Error 29	0.98	0.26	0.41	1.00	1.00	1.00	325
Error 30	0.46	0.47	0.47	1.00	1.00	1.00	327
Error 31	0.88	0.42	0.57	1.00	0.87	0.93	318
Error 40	0.00	0.00	0.00	0.36	0.11	0.17	202
Error 41	0.99	0.93	0.96	1.00	0.96	0.98	191
Error 46	0.97	0.81	0.88	1.00	0.81	0.90	196
Error 47	0.99	0.91	0.95	1.00	0.91	0.95	206

Table A4. Detailed per-class classification performance for $id_{mea} = 3$ without hybrid fault-space restructuring.

Class	DT			RF			Support
	Prec.	Rec.	F1	Prec.	Rec.	F1	
Error 0	0.89	0.99	0.94	0.93	0.99	0.96	32,753
Error 1	0.00	0.00	0.00	0.28	0.05	0.08	274
Error 10	0.95	0.68	0.79	0.93	0.92	0.92	263
Error 11	0.89	0.72	0.80	0.93	0.93	0.93	245
Error 12	0.95	0.99	0.97	1.00	1.00	1.00	280
Error 13	0.90	0.63	0.74	0.98	1.00	0.99	334
Error 14	0.94	0.69	0.80	0.99	0.99	0.99	315
Error 15	0.69	0.83	0.75	1.00	0.99	1.00	331
Error 16	0.24	0.25	0.25	0.31	0.28	0.29	311
Error 17	0.27	0.55	0.36	0.29	0.30	0.29	327
Error 18	0.30	0.08	0.13	0.29	0.27	0.28	329
Error 19	0.44	0.13	0.20	0.31	0.35	0.33	318
Error 2	0.00	0.00	0.00	0.14	0.03	0.05	260
Error 20	1.00	0.67	0.80	1.00	0.97	0.98	340
Error 21	0.94	0.23	0.37	1.00	0.95	0.98	332
Error 22	0.00	0.00	0.00	0.27	0.06	0.10	343
Error 23	0.00	0.00	0.00	0.21	0.05	0.08	307
Error 24	1.00	0.65	0.79	1.00	1.00	1.00	309
Error 25	1.00	0.06	0.11	0.99	0.98	0.99	330
Error 26	0.61	0.11	0.19	0.75	0.77	0.76	310
Error 27	0.29	0.60	0.39	0.74	0.71	0.72	328
Error 28	0.38	0.16	0.23	0.76	0.77	0.77	316
Error 29	0.89	0.28	0.43	0.99	1.00	1.00	325
Error 3	0.41	0.61	0.49	0.94	0.95	0.94	263
Error 30	0.98	0.76	0.85	1.00	1.00	1.00	327
Error 31	0.97	0.45	0.61	1.00	0.91	0.95	318
Error 32	0.16	0.77	0.27	0.16	0.20	0.18	292
Error 33	0.18	0.01	0.02	0.18	0.21	0.20	266
Error 34	0.13	0.04	0.06	0.16	0.16	0.16	257

Table A4. Cont.

Class	DT			RF			Support
	Prec.	Rec.	F1	Prec.	Rec.	F1	
Error 35	0.12	0.04	0.07	0.17	0.14	0.16	246
Error 36	0.17	0.00	0.01	0.17	0.17	0.17	264
Error 37	0.14	0.11	0.12	0.14	0.12	0.13	252
Error 38	0.12	0.01	0.02	0.07	0.05	0.06	222
Error 39	0.00	0.00	0.00	0.30	0.05	0.09	202
Error 4	0.79	0.15	0.25	0.95	0.94	0.94	272
Error 40	0.00	0.00	0.00	0.34	0.10	0.16	202
Error 41	0.99	0.90	0.94	1.00	0.94	0.97	191
Error 42	0.00	0.00	0.00	0.24	0.06	0.09	190
Error 43	0.00	0.00	0.00	0.39	0.07	0.12	189
Error 44	0.96	0.98	0.97	0.98	0.99	0.98	196
Error 45	0.98	0.95	0.97	0.99	0.98	0.98	196
Error 46	1.00	0.62	0.77	1.00	0.73	0.85	195
Error 47	0.48	0.79	0.60	1.00	0.95	0.98	206
Error 5	0.00	0.00	0.00	0.33	0.11	0.16	284
Error 6	0.48	0.71	0.57	0.97	0.98	0.98	275
Error 7	0.61	0.51	0.56	0.98	0.97	0.97	258
Error 8	0.90	0.58	0.71	0.93	0.97	0.95	271
Error 9	0.65	0.68	0.66	0.96	0.92	0.94	254

Appendix A.3. $id_{mea} = 2$ Table A5. Detailed per-class classification performance for $id_{mea} = 2$ under the proposed hybrid fault-space representation.

Class	DT			RF			Support
	Prec.	Rec.	F1	Prec.	Rec.	F1	
Cluster 0	0.95	1.00	0.97	1.00	1.00	1.00	59,420
Cluster 1	0.83	0.70	0.76	1.00	1.00	1.00	928
Cluster 2	1.00	0.71	0.83	1.00	1.00	1.00	929
Cluster 3	1.00	0.57	0.73	1.00	1.00	1.00	900
Cluster 4	1.00	0.56	0.72	1.00	1.00	1.00	902
Cluster 5	1.00	0.89	0.94	1.00	1.00	1.00	2059
Cluster 6	0.47	0.98	0.63	1.00	1.00	1.00	1540
Cluster 7	0.97	0.94	0.96	1.00	0.99	0.99	3152
Cluster 8	0.92	0.66	0.77	1.00	1.00	1.00	798
Error 12	0.28	0.70	0.40	1.00	1.00	1.00	489
Error 13	0.94	0.19	0.31	1.00	1.00	1.00	533
Error 14	0.33	0.05	0.09	1.00	1.00	1.00	502
Error 15	0.99	0.33	0.49	1.00	1.00	1.00	516
Error 20	0.98	0.60	0.74	0.99	0.97	0.98	541
Error 21	0.69	0.23	0.35	1.00	0.98	0.99	531
Error 24	0.99	0.52	0.68	1.00	1.00	1.00	499
Error 25	0.96	0.82	0.88	1.00	1.00	1.00	540
Error 29	0.85	0.43	0.57	1.00	0.99	1.00	479
Error 30	0.32	0.53	0.40	1.00	1.00	1.00	455
Error 31	0.83	0.15	0.25	0.99	0.93	0.96	448
Error 41	0.72	0.16	0.27	1.00	0.95	0.97	387
Error 46	0.83	0.15	0.26	0.98	0.77	0.86	390
Error 47	0.98	0.11	0.19	1.00	0.91	0.95	403

Table A6. Detailed per-class classification performance for $id_{mea} = 2$ without hybrid fault-space restructuring.

Class	DT			RF			Support
	Prec.	Rec.	F1	Prec.	Rec.	F1	
Error 0	0.90	1.00	0.95	0.93	0.99	0.96	55,408
Error 1	0.00	0.00	0.00	0.25	0.06	0.10	478
Error 10	0.55	0.96	0.70	0.98	0.97	0.97	459
Error 11	0.90	0.55	0.68	0.97	0.98	0.97	444
Error 12	0.99	0.71	0.83	1.00	1.00	1.00	489
Error 13	0.70	0.84	0.76	1.00	1.00	1.00	533
Error 14	0.79	0.84	0.82	1.00	1.00	1.00	502
Error 15	0.86	0.77	0.81	1.00	1.00	1.00	516
Error 16	0.49	0.06	0.10	0.30	0.27	0.29	510
Error 17	0.38	0.08	0.13	0.24	0.22	0.23	514
Error 18	0.25	0.03	0.06	0.27	0.29	0.28	526
Error 19	0.26	0.83	0.39	0.29	0.32	0.30	508
Error 2	0.00	0.00	0.00	0.29	0.06	0.10	453
Error 20	0.94	0.59	0.73	1.00	0.97	0.99	540
Error 21	0.85	0.34	0.49	1.00	0.96	0.98	531
Error 22	0.00	0.00	0.00	0.28	0.10	0.15	532
Error 23	0.00	0.00	0.00	0.26	0.08	0.12	500
Error 24	0.96	0.98	0.97	1.00	1.00	1.00	499
Error 25	0.98	0.99	0.98	1.00	1.00	1.00	540
Error 26	0.75	0.44	0.55	0.99	1.00	1.00	513
Error 27	0.42	0.81	0.56	1.00	0.99	0.99	518
Error 28	0.54	0.31	0.39	0.99	1.00	0.99	509
Error 29	0.85	0.53	0.65	0.99	1.00	1.00	479
Error 3	0.93	0.55	0.69	0.98	0.97	0.98	461
Error 30	0.87	0.48	0.62	1.00	1.00	1.00	455
Error 31	0.80	0.20	0.32	0.99	0.94	0.97	448
Error 32	0.12	0.13	0.13	0.18	0.18	0.18	443
Error 33	0.17	0.46	0.24	0.22	0.30	0.25	462
Error 34	0.21	0.03	0.06	0.20	0.20	0.20	456
Error 35	0.22	0.04	0.06	0.18	0.17	0.17	445
Error 36	0.18	0.06	0.09	0.17	0.17	0.17	465
Error 37	0.14	0.08	0.10	0.19	0.15	0.17	460
Error 38	0.22	0.08	0.12	0.21	0.18	0.19	421
Error 39	0.00	0.00	0.00	0.31	0.09	0.14	398
Error 4	0.57	0.67	0.62	0.97	0.99	0.98	467
Error 40	0.00	0.00	0.00	0.31	0.10	0.15	401
Error 41	0.79	0.49	0.60	1.00	0.95	0.97	387
Error 42	0.00	0.00	0.00	0.29	0.09	0.13	392
Error 43	0.00	0.00	0.00	0.26	0.08	0.12	374
Error 44	0.56	0.80	0.66	0.98	0.99	0.98	404
Error 45	1.00	0.31	0.48	0.99	0.97	0.98	394
Error 46	0.55	0.60	0.57	1.00	0.80	0.89	390
Error 47	0.44	0.83	0.58	0.99	0.89	0.94	403
Error 5	0.00	0.00	0.00	0.35	0.10	0.15	485
Error 6	0.37	0.97	0.54	0.98	0.98	0.98	475
Error 7	0.96	0.71	0.81	0.98	0.98	0.98	454
Error 8	0.94	0.60	0.73	0.98	1.00	0.99	460
Error 9	0.98	0.50	0.66	1.00	0.98	0.99	440

References

- Chen, X.; Jiang, Y.; Terzija, V.; Lu, C. Review on measurement-based frequency dynamics monitoring and analyzing in renewable energy dominated power systems. *Int. J. Electr. Power Energy Syst.* **2024**, *155*, 109520. <https://doi.org/10.1016/j.ijepes.2023.109520>.
- Hosseinabadi, F.; Chakraborty, S.; Bhoi, S.K.; Prochart, G.; Hrvanovic, D.; Hegazy, O. A Comprehensive Overview of Reliability Assessment Strategies and Testing of Power Electronics Converters. *IEEE Open J. Power Electron.* **2024**, *5*, 473–512. <https://doi.org/10.1109/OJPEL.2024.3379294>.
- Cheng, G.; Lin, Y.; Abur, A.; Gómez-Expósito, A.; Wu, W. A Survey of Power System State Estimation Using Multiple Data Sources: PMUs, SCADA, AML, and Beyond. *IEEE Trans. Smart Grid* **2024**, *15*, 1129–1151. <https://doi.org/10.1109/TSG.2023.3286401>.
- Sánchez, L.; Couso, I.; Costa, N. Self-supervised spectral learning with an application to condition monitoring of tunnel ventilation systems under sparse data conditions. *Comput. Ind. Eng.* **2026**, *211*, 111634. <https://doi.org/10.1016/j.cie.2025.111634>.
- Liserre, M.; Buticchi, G.; Leon, J.I.; Marquez Alcaide, A.; Raveendran, V.; Ko, Y.; Andresen, M.; Monopoli, V.G.; Franquelo, L. Power Routing: A New Paradigm for Maintenance Scheduling. *IEEE Ind. Electron. Mag.* **2020**, *14*, 33–45. <https://doi.org/10.1109/MIE.2020.2975049>.
- García-Campos, J.M.; Alcaide, A.M.; Letrado-Castellanos, A.; Portillo, R.; Leon, J.I. A Scalable Microservices Architecture for Condition Monitoring and State-of-Health Tracking in Power Conversion Systems. *Sensors* **2026**, *26*, 1282. <https://doi.org/10.3390/s26041282>.
- Atitallah, S.B.; Driss, M.; Ghzela, H.B. Microservices for Data Analytics in IoT Applications: Current Solutions, Open Challenges, and Future Research Directions. *Procedia Comput. Sci.* **2022**, *207*, 3938–3947. <https://doi.org/10.1016/j.procs.2022.09.456>.
- Dineva, A.; Mosavi, A.; Gyimesi, M.; Vajda, I. Multi-Label Classification for Fault Diagnosis of Rotating Electrical Machines. *arXiv* **2019**, arXiv:1908.01078. <https://doi.org/10.48550/arXiv.1908.01078>.
- Lundgren, A.; Jung, D. Data-driven fault diagnosis analysis and open-set classification of time-series data. *Control Eng. Pract.* **2022**, *121*, 105006.
- Aviña-Corral, V.; de Jesus Rangel-Magdaleno, J.; Barron-Zambrano, J.H.; Rosales-Núñez, S. Review of fault detection techniques in power converters: Fault analysis and diagnostic methodologies. *Measurement* **2024**, *234*, 114864. <https://doi.org/10.1016/j.measurement.2024.114864>.
- Awaisi, K.S.; Ye, Q.; Sampalli, S. A Survey of Industrial AIoT: Opportunities, Challenges, and Directions. *IEEE Access* **2024**, *12*, 96946–96996. <https://doi.org/10.1109/ACCESS.2024.3426279>.
- Savaglio, C.; Mazzei, P.; Fortino, G. Edge Intelligence for Industrial IoT: Opportunities and Limitations. *Procedia Comput. Sci.* **2024**, *232*, 397–405. <https://doi.org/10.1016/j.procs.2024.01.039>.
- Zhou, Z.; Chen, X.; Li, E.; Zeng, L.; Luo, K.; Zhang, J. Edge Intelligence: Paving the Last Mile of Artificial Intelligence with Edge Computing. *Proc. IEEE* **2019**, *107*, 1738–1762. <https://doi.org/10.1109/JPROC.2019.2918951>.
- Liu, F.; Tang, G.; Li, Y.; Cai, Z.; Zhang, X.; Zhou, T. A Survey on Edge Computing Systems and Tools. *Proc. IEEE* **2019**, *107*, 1537–1562. <https://doi.org/10.1109/JPROC.2019.2920341>.
- Yang, S.; Xiang, D.; Bryant, A.; Mawby, P.; Ran, L.; Tavner, P. Condition Monitoring for Device Reliability in Power Electronic Converters: A Review. *IEEE Trans. Power Electron.* **2010**, *25*, 2734–2752. <https://doi.org/10.1109/TPEL.2010.2049377>.
- Yu, X.; Kong, J.; Wang, N.; Zhang, K.; Blaabjerg, F.; Zhou, D. A Review on Junction Temperature and ON-state Voltage Condition Monitoring of Power Semiconductor Devices. *Chin. J. Electr. Eng.* **2025**, *11*, 17–37. <https://doi.org/10.23919/CJEE.2025.000144>.
- Khan, S.S.; Wen, H. A Comprehensive Review of Fault Diagnosis and Tolerant Control in DC-DC Converters for DC Microgrids. *IEEE Access* **2021**, *9*, 80100–80127. <https://doi.org/10.1109/ACCESS.2021.3083721>.
- Sumiti; Shrivastava, V. Condition Monitoring of Power Electronic Devices-A Review. In *Proceedings of the 2022 2nd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*; IEEE: Piscataway, NJ, USA, 2022; pp. 749–753. <https://doi.org/10.1109/ICACITE53722.2022.9823555>.
- Alcaide, A.M.; Guenter, S.; Leon, J.I.; Buticchi, G.; Kouro, S.; Franquelo, L.G. Common DC-Link Capacitor Lifetime Extension in Modular DC/DC Converters for Electric Vehicle Fast Chargers via Variable-Angle Interleaved Operation. *IEEE Trans. Ind. Electron.* **2023**, *70*, 10765–10774. <https://doi.org/10.1109/TIE.2022.3231251>.
- Alcaide, A.M.; Guenter, S.; Leon, J.I.; Portillo, R.; Lizana, R.; Buticchi, G.; Kouro, S.; Franquelo, L.G. Integral Hierarchical Degradation Management-Oriented Control of Parallel-Connected DC/DC Converters. *IEEE Trans. Ind. Electron.* **2026**, 1–10. <https://doi.org/10.1109/TIE.2026.3672741>.
- Gao, Z.; Cecati, C.; Ding, S.X. A Survey of Fault Diagnosis and Fault-Tolerant Techniques—Part I: Fault Diagnosis with Model-Based and Signal-Based Approaches. *IEEE Trans. Ind. Electron.* **2015**, *62*, 3757–3767.
- Han, P.; Liu, Z.; He, X.; Ding, S.X.; Zhou, D. Multi-Condition Fault Diagnosis of Dynamic Systems: A Survey, Insights, and Prospects. *IEEE Trans. Autom. Sci. Eng.* **2025**, *22*, 15728–15744. <https://doi.org/10.1109/TASE.2025.3571516>.
- Zhao, S.; Yang, H.; Kareck, T.L.; Khan, F.; Wang, Q. Data-driven fault detection and diagnosis in industrial process systems: A systematic review and perspective. *Reliab. Eng. Syst. Saf.* **2026**, *270*, 112159. <https://doi.org/10.1016/j.res.2025.112159>.

24. Pimentel, J.; McEwan, A.A.; Yu, H.Q. A Novel Real-Time Battery State Estimation Using Data-Driven Prognostics and Health Management. *Appl. Sci.* **2025**, *15*, 8538. <https://doi.org/10.3390/app15158538>.
25. Liu, Y.; Li, P.; Si, Y.; Ma, L. A Comprehensive Review of AI Integration for Fault Detection in Modern Power Systems: Data Processing, Modeling, and Optimization. *Energies* **2025**, *18*, 4983. <https://doi.org/10.3390/en18184983>.
26. Bharath Kurukuru, V.S.; Haque, A.; Kumar, R.; Khan, M.A.; Tripathy, A.K. Machine Learning based Fault Classification Approach for Power electronic converters. In *Proceedings of the 2020 IEEE International Conference on Power Electronics, Drives and Energy Systems (PEDES)*; IEEE: Piscataway, NJ, USA, 2020; pp. 1–6.
27. Bhat, D.; Muench, S.; Roellig, M. Estimation of remaining useful lifetime of power electronic components with machine learning based on mission profile data. *Power Electron. Devices Components* **2023**, *5*, 100040. <https://doi.org/10.1016/j.pedc.2023.100040>.
28. Bhoi, S.K.; Chakraborty, S.; Verbrugge, B.; Helsen, S.; Robyns, S.; El Baghdadi, M.; Hegazy, O. Intelligent data-driven condition monitoring of power electronics systems using smart edge–cloud framework. *Internet Things* **2024**, *26*, 101158. <https://doi.org/10.1016/j.iot.2024.101158>.
29. Wani, A.A. Comprehensive review of dimensionality reduction algorithms: challenges, limitations, and innovative solutions. *PeerJ Comput. Sci.* **2025**, *11*, e3025. <https://doi.org/10.7717/peerj-cs.3025>.
30. Siddiqui, H.; Khendek, F.; Toeroe, M. Microservices based architectures for IoT systems - State-of-the-art review. *Internet Things* **2023**, *23*, 100854. <https://doi.org/10.1016/j.iot.2023.100854>.
31. Lin, L.; Zhang, S.; Fu, S.; Liu, Y. FD-LLM: Large language model for fault diagnosis of complex equipment. *Adv. Eng. Inform.* **2025**, *65*, 103208. <https://doi.org/10.1016/j.aei.2025.103208>.
32. Xu, C.; Wang, Z.; Jin, Y.; Nong, W. An adaptive industrial large language model for mechanical fault diagnosis under variable operating conditions. *Adv. Eng. Inform.* **2026**, *74*, 104821. <https://doi.org/10.1016/j.aei.2026.104821>.
33. Kong, X.; Wu, Y.; Wang, H.; Xia, F. Edge Computing for Internet of Everything: A Survey. *IEEE Internet Things J.* **2022**, *9*, 23472–23485. <https://doi.org/10.1109/JIOT.2022.3200431>.
34. Deng, S.; Zhao, H.; Fang, W.; Yin, J.; Dustdar, S.; Zomaya, A.Y. Edge Intelligence: The Confluence of Edge Computing and Artificial Intelligence. *IEEE Internet Things J.* **2020**, *7*, 7457–7469. <https://doi.org/10.1109/JIOT.2020.2984887>.
35. Richelli, A.; Salem, M.; Colalongo, L. A Review of Fully Integrated and Embedded Power Converters for IoT. *Energies* **2021**, *14*, 5419. <https://doi.org/10.3390/en14175419>.
36. Terziyska, M.; Ilieva, I.; Terziyski, Z.; Komitov, N. A Review of the Transition from Industry 4.0 to Industry 5.0: Unlocking the Potential of TinyML in Industrial IoT Systems. *Sci* **2026**, *8*, 10. <https://doi.org/10.3390/sci8010010>.
37. García-Campos, J.M.; Alcaide, A.M.; Letrado-Castellanos. Hybrid Fault-Space Restructuring for Fault Diagnosis in Power Electronic Converters. GitHub Repository. 2026. Available online: <https://github.com/jmgarcam/ml-hybrid-fault-space-clustering> (accessed on 1 June 2026).
38. García-Campos, J.M.; Marquez-Alcaide, A.; Letrado-Castellanos, A.; Portillo, R.; Leon, J.I. Power Converter Fault Diagnosis Dataset. Zenodo, 2026. Available online: <https://doi.org/10.5281/zenodo.20484338> (accessed on 1 June 2026).

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.