

Article

Physics-Regularized Hybrid Learning Framework for Fault Location and Classification in Aging Underground Distribution Networks

Alexander Aguila Téllez ^{1,*} , Francisco Jurado ² , Manuel Jaramillo ¹  and Pengda Liu ³

¹ Department of Electrical Engineering, Universidad Politécnica Salesiana, Quito EC 170146, Ecuador

² Department of Electrical Engineering, University of Jaen, ES 23700 Linares, Spain

³ International Science and Technology Cooperation Base of Intelligent Manufacturing Service, Chongqing Technology and Business University, Chongqing 400072, China

* Correspondence: aaguila@ups.edu.ec; Tel.: +593-998643886

Abstract

Underground distribution networks increasingly rely on aging cable assets whose parameter drift modifies propagation velocity, attenuation, and fault-initiated transient signatures, thereby reducing the reliability of conventional traveling-wave (TW) fault location and purely data-driven diagnosis. This paper proposes a physics-regularized hybrid learning framework for joint fault-type classification, feeder-area identification, and continuous fault localization in aging underground distribution feeders. The methodology integrates (i) an aging-aware simulation pipeline driven by a normalized aging-stress index $\alpha \in [0, 1]$ that perturbs the per-unit-length cable matrices within a controlled domain; (ii) synchronized multi-sensor time–frequency representations of three-phase voltage and current transients; (iii) an area-aware multi-task architecture with fault-type and area-classification heads and area-specific local regression heads; and (iv) a propagation-consistency loss that depends explicitly on the model-predicted fault position and therefore contributes gradients during training. A branched underground feeder is evaluated using five synchronized sensing locations and a stratified dataset of $N_{\text{tot}} = 36,000$ simulated fault events covering 11 fault classes (SLG-A/B/C, LL-AB/BC/CA, DLG-ABG/BCG/CAG, LLL, and LLLG), six non-overlapping feeder areas, fault resistance $R_f \in [0.1, 50] \Omega$, measurement noise $\text{SNR} \in [20, 40] \text{ dB}$, and a 20 ms transient window sampled at 200 kHz. On a held-out test set of 7200 previously unseen event records drawn from the same simulation domain, the Hybrid model achieves a fault-type accuracy of 0.93, an area-identification accuracy of 0.96, a localization MAE of 0.011 p.u., and a 95th-percentile absolute error of 0.027 p.u. The proposed configuration outperforms the TW-TOA, purely data-driven Baseline, and physics-regularized PINN references across the reported diagnostic tasks within the prescribed simulator and parameter ranges. Time–frequency attribution is included only as a qualitative interpretability illustration and is not treated as quantitative evidence of explanation faithfulness. Accordingly, the results demonstrate comparative in-domain simulation performance rather than field or cross-simulator deployment readiness.



Academic Editor: Armando Pires

Received: 17 June 2026

Revised: 23 July 2026

Accepted: 27 July 2026

Published: 29 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](https://creativecommons.org/licenses/by/4.0/)

[Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: cable aging; explainable artificial intelligence; fault classification; fault location; multi-task learning; physics-regularized learning; robustness under noise; time–frequency analysis; traveling-wave transients; underground distribution networks

1. Introduction

Underground distribution networks offer improved public safety, reduced visual impact, and greater resilience to environmental disturbances. Nevertheless, fault restoration in underground feeders remains difficult because faulted sections must be identified under limited measurement availability, complex branching structures, and strongly attenuated transient signals. These challenges become more pronounced as cable insulation deteriorates under thermal, electrical, mechanical, and environmental stresses. Aging modifies dielectric behavior, conductivity, and effective electromagnetic parameters, thereby altering transient attenuation, dispersion, and propagation velocity [1–4].

Traveling-wave (TW) methods remain physically interpretable fault-location techniques because the arrival times and spectral content of fault-initiated high-frequency transients contain distance information. Single-ended, double-ended, time–frequency, full-waveform, and frequency-domain formulations have therefore been widely investigated [5–9]. Their practical accuracy, however, depends on the recorder bandwidth, sampling frequency, synchronization, sensor response, propagation-velocity estimation, and reliable wavefront detection. Underground and hybrid feeders introduce additional attenuation, dispersion, discontinuities, and branch-dependent reflections that can distort the first arrival and increase location ambiguity [10–15].

Learning-based methods can exploit multi-channel waveforms and time–frequency representations without relying exclusively on a single detected arrival. However, purely data-driven models may be sensitive to changes in cable parameters, fault resistance, operating point, feeder topology, and measurement conditions. Physics-informed and physics-regularized learning address this limitation by embedding governing relations or engineering constraints into the training process [16–20]. In protection-oriented applications, attribution methods can additionally support the examination of the signal regions associated with diagnostic decisions, although model explanations must be validated before they are interpreted as evidence of decision faithfulness [21].

Existing studies generally address cable aging, TW processing, learning-based diagnosis, physics-constrained training, and interpretability as separate problems. Their integration remains limited for branched underground feeders in which aging-induced parameter drift, fault resistance, waveform distortion, measurement noise, and nonuniform sensor observability simultaneously affect fault-type classification and continuous location estimation. To address this gap, this paper proposes a physics-regularized Hybrid framework for joint fault-type classification, feeder-area identification, and continuous fault localization. The framework combines controlled aging-aware simulation, synchronized multi-sensor time–frequency inputs, an output-dependent propagation-consistency loss, and area-specific local regression.

The principal contributions are as follows:

1. **Controlled aging-aware transient modeling:** A normalized aging-stress index $\alpha \in [0, 1]$ is used to impose bounded and physically interpretable variations on the per-unit-length cable matrices, enabling a comparative evaluation under structured parameter drift.
2. **Output-dependent physics regularization:** A propagation-consistency residual is formulated explicitly as a function of the model-predicted fault position, ensuring that the physics term contributes gradients to the trainable parameters.
3. **Area-aware multi-sensor localization:** A feeder-area classification head, fixed area-dependent sensor selectors, and six local regression heads are combined through a topology-dependent local-to-global coordinate mapping.

4. Unified multi-task evaluation: Fault-type classification, area identification, and continuous localization are evaluated using the same simulated events and signal representation for TW-TOA, Baseline, PINN, and Hybrid configurations.
5. Robustness and uncertainty assessment: This paper examines performance across normalized aging stress, fault resistance, additive noise, feeder areas, bootstrap resamples, calibration, and selective prediction. Time–frequency saliency is included only as a qualitative illustration rather than as quantitative evidence of explanation faithfulness.

The resulting formulation is evaluated as an in-domain comparative study using unseen events generated within the same controlled simulator and parameter ranges.

2. Related Work

Research relevant to the proposed framework spans underground-cable aging, traveling-wave fault location, learning-based diagnosis, and physics-informed modeling. Cable degradation is driven by interacting thermal, electrical, and environmental stresses that alter insulation properties, conductor losses, dielectric behavior, and expected service life [1,2,4]. Partial-discharge measurements provide complementary evidence of insulation deterioration and have been used to characterize degradation progression [3]. These studies motivate the inclusion of aging-related parameter variability, although they do not directly address joint post-fault classification and localization under transient propagation changes.

Traveling-wave methods estimate fault position from high-frequency transients generated at fault inception. Their theoretical foundations, measurement requirements, and principal implementation challenges are summarized in [5]. Time–frequency analysis and Park-transformation-based processing have been used to improve transient characterization and wavefront extraction [6,7]. Other formulations address unsynchronized measurements, series-compensated lines, simultaneous faults, and propagation-error compensation [22–25]. In distribution and underground networks, branching, wavefront distortion, attenuation, and frequency-dependent propagation remain important sources of uncertainty [9–11,13–15].

Data-driven diagnostic methods can use wavelet coefficients, time–frequency maps, envelopes, or full transient waveforms to capture information that is not represented by a single arrival-time estimate. Distributed transient-processing algorithms and learning-based classifiers have demonstrated the diagnostic value of representative waveform data [26–28]. Their performance, however, depends on the correspondence between the training and deployment domains, particularly when cable parameters, network topology, operating conditions, fault resistance, or measurement characteristics change. Physics-informed fault-diagnosis research therefore seeks to combine learned representations with physical constraints or prior knowledge [29].

Physics-informed learning has been applied to power-system dynamics, state estimation, high-impedance fault detection, and constrained optimization [16–20]. Explainable-learning methods have also been proposed for high-voltage equipment diagnostics to support the engineering interpretation of model decisions [21]. Nevertheless, the literature provides a limited integration of controlled cable-parameter drift, multi-sensor time–frequency learning, output-dependent physics regularization, and area-conditioned continuous localization within a common branched underground-feeder framework. This paper addresses this methodological intersection while treating the available saliency visualization as qualitative rather than as validated evidence of attribution faithfulness.

3. System Modeling and Cable Aging Formulation

This section defines the electrical model of the underground feeder, the cable parameterization under aging, and the mathematical objects used later to generate fault transients

and to impose physics-consistency constraints. The formulation is kept general (per-unit-length distributed parameters) and then specialized to a numerically tractable multi-section representation suitable for time-domain simulation.

3.1. Network and Measurement Model

Consider a radial or weakly meshed distribution network composed of buses (nodes) connected by underground cable segments. Let $\mathcal{N}$ be the set of buses and $\mathcal{E}$ the set of cable branches. Each branch $e \in \mathcal{E}$ connects buses (i, j) and has length ℓ_e .

Voltage and current measurements are assumed available at a subset of buses $\mathcal{M} \subseteq \mathcal{N}$, typically at the substation and optionally at selected monitoring points. For each monitored bus $m \in \mathcal{M}$, the measured three-phase voltage and current vectors, denoted by $\mathbf{v}_m(t)$ and $\mathbf{i}_m(t)$, respectively, are given by

$$\mathbf{v}_m(t) = \begin{bmatrix} v_{a,m}(t) \\ v_{b,m}(t) \\ v_{c,m}(t) \end{bmatrix}, \quad \mathbf{i}_m(t) = \begin{bmatrix} i_{a,m}(t) \\ i_{b,m}(t) \\ i_{c,m}(t) \end{bmatrix}. \quad (1)$$

These waveforms are used to detect and localize faults and to compute time–frequency representations (Section 4).

3.2. Fault Modeling

Faults are modeled as shunt connections at an unknown location $x_f \in (0, \ell)$ along a branch with fault resistance R_f and fault type $\mathcal{T} \in \{\text{SLG}, \text{LL}, \text{DLG}, \text{LLL}, \text{LLLG}\}$. At the fault point, the phase-domain relation can be expressed as

$$\mathbf{i}_f(t) = \mathbf{Y}_f(\mathcal{T}, R_f) \mathbf{v}_f(t), \quad (2)$$

where $\mathbf{v}_f(t)$ and $\mathbf{i}_f(t)$ are the three-phase voltage and injected fault current at the fault location, and $\mathbf{Y}_f$ is the fault admittance matrix encoding the fault connections. For example, for an a -phase-to-ground fault (SLG on phase a),

$$\mathbf{Y}_f^{\text{SLG-}a}(R_f) = \begin{bmatrix} 1/R_f & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad (3)$$

and analogous forms apply for LL, DLG, LLL, and LLLG faults using appropriate phase couplings. This admittance-based representation is convenient for embedding faults into nodal equations during simulation.

3.3. Traveling-Wave Signatures and Modeling Requirements

Traveling-wave-based fault location relies on capturing high-frequency transients initiated at fault inception. The ability of the model to represent attenuation, dispersion, and reflections is therefore critical. TW-based fault location methods and their practical challenges are reviewed in [5], and time–frequency traveling-wave characterization has been shown to be useful for enhancing location performance [6]. Distorted wavefronts in distribution networks can degrade classical estimators, motivating an explicit consideration of wavefront distortion in fault location algorithms [10]. The modeling approach adopted here (distributed parameters approximated by multi-section π) is designed to preserve these transient phenomena with controllable fidelity.

3.4. Implementation: Data Generation Pipeline (Pseudocode)

Algorithm 1 summarizes the aging-aware simulation procedure used to generate labeled fault data for subsequent learning. It explicitly samples the normalized aging-stress index α , fault type $\mathcal{T}$, fault resistance R_f , and fault location x_f .

Algorithm 1: Aging-aware fault transient data generation

Input: Network topology $(\mathcal{N}, \mathcal{E})$, nominal cable matrices $(\mathbf{R}_0, \mathbf{L}_0, \mathbf{C}_0, \mathbf{G}_0)$, discretization N_π

Input: Sampling ranges: $\alpha \sim \mathcal{D}_\alpha$, $\mathcal{T} \sim \mathcal{D}_\mathcal{T}$, $R_f \sim \mathcal{D}_R$, $x_f \sim \mathcal{D}_x$, operating point $\xi \sim \mathcal{D}_\xi$

Output: Labeled dataset $\mathcal{S} = \{(\{v_m(t), i_m(t)\}_{m \in \mathcal{M}}, \mathcal{T}, x_f, \alpha)\}$

Initialize empty set $\mathcal{S} \leftarrow \emptyset$;

for $n = 1$ **to** N_{samples} **do**

 Sample normalized aging-stress index $\alpha^{(n)} \sim \mathcal{D}_\alpha$;

 Compute aged matrices using (13)–(16);

 Discretize each cable branch into N_π π -sections with $(\mathbf{R}_k, \mathbf{L}_k, \mathbf{C}_k, \mathbf{G}_k)$;

 Sample operating point $\xi^{(n)}$ (loads, DG if applicable);

 Run pre-fault steady state under $\xi^{(n)}$;

 Sample fault type $\mathcal{T}^{(n)}$, location $x_f^{(n)}$, and resistance $R_f^{(n)}$;

 Apply fault admittance $\mathbf{Y}_f(\mathcal{T}^{(n)}, R_f^{(n)})$ at $x_f^{(n)}$ using (2);

 Simulate time-domain transient and record $\{v_m(t), i_m(t)\}_{m \in \mathcal{M}}$;

 Store $(\{v_m(t), i_m(t)\}_{m \in \mathcal{M}}, \mathcal{T}^{(n)}, x_f^{(n)}, \alpha^{(n)})$ in $\mathcal{S}$;

return $\mathcal{S}$.

3.5. Compact Diagram of the Modeling Workflow

Figure 1 provides a compact workflow diagram linking aging parameterization, cable discretization, fault insertion, and waveform recording. The diagram is designed to fit within a single column/page width.

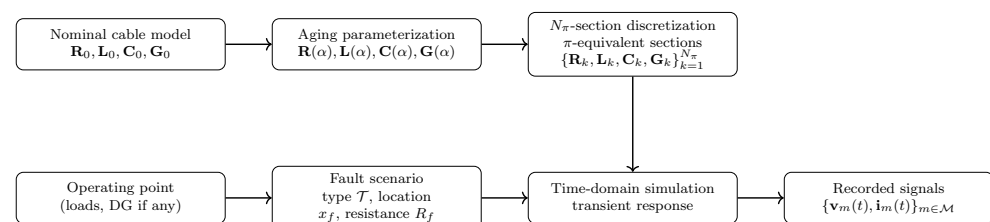


Figure 1. Aging-aware modeling workflow: nominal cable parameters are degraded via a normalized aging-stress index α , discretized into N_π cascaded π -sections, and combined with operating conditions and an explicit fault scenario as inputs to the time-domain simulation, which produces transient voltage and current measurements.

3.6. Underground Cable Electrical Model

This subsection formalizes the electrical representation of underground cables used throughout the proposed framework. We adopt a distributed-parameter multiconductor model to capture attenuation, dispersion, and coupling effects that critically shape fault-initiated high-frequency transients, which are central to traveling-wave-based fault analysis [5,6,10].

3.6.1. Multiconductor Telegrapher Equations (Frequency Domain)

Consider an underground cable segment of length ℓ with longitudinal coordinate $x \in [0, \ell]$. Let $\mathbf{V}(x, \omega) \in \mathbb{C}^p$ and $\mathbf{I}(x, \omega) \in \mathbb{C}^p$ denote the phasor voltage and current vectors of a p -conductor representation (e.g., $p = 3$ for three-phase phase-domain modeling). The multiconductor telegrapher equations are

$$\frac{d\mathbf{V}(x, \omega)}{dx} = -\mathbf{Z}(\omega) \mathbf{I}(x, \omega), \quad (4)$$

$$\frac{d\mathbf{I}(x, \omega)}{dx} = -\mathbf{Y}(\omega) \mathbf{V}(x, \omega), \quad (5)$$

where the per-unit-length impedance and admittance matrices are

$$\mathbf{Z}(\omega) = \mathbf{R} + j\omega \mathbf{L}, \quad (6)$$

$$\mathbf{Y}(\omega) = \mathbf{G} + j\omega \mathbf{C}. \quad (7)$$

Here, $\mathbf{R}, \mathbf{L}, \mathbf{G}, \mathbf{C} \in \mathbb{R}^{p \times p}$ are, respectively, the series resistance, series inductance, shunt conductance, and shunt capacitance matrices. Off-diagonal entries represent mutual coupling among conductors, and $\mathbf{G}$ accounts for dielectric loss.

Combining (4) and (5) yields second-order matrix wave equations:

$$\frac{d^2 \mathbf{V}(x, \omega)}{dx^2} = \mathbf{Z}(\omega) \mathbf{Y}(\omega) \mathbf{V}(x, \omega), \quad (8)$$

$$\frac{d^2 \mathbf{I}(x, \omega)}{dx^2} = \mathbf{Y}(\omega) \mathbf{Z}(\omega) \mathbf{I}(x, \omega). \quad (9)$$

The propagation matrix and characteristic impedance matrix are defined as

$$\mathbf{\Gamma}(\omega) = (\mathbf{Z}(\omega) \mathbf{Y}(\omega))^{1/2}, \quad \mathbf{Z}_c(\omega) = (\mathbf{Z}(\omega) \mathbf{Y}(\omega)^{-1})^{1/2}, \quad (10)$$

where $(\cdot)^{1/2}$ denotes the principal matrix square root. These quantities govern attenuation/phase shift and reflection behavior, which directly affects wavefront detection and time–frequency feature extraction in traveling-wave-based fault location [5,6].

3.6.2. Numerically Tractable Time-Domain Representation: Cascaded π -Sections

To simulate transients efficiently in the time domain while retaining distributed behavior, each cable segment is discretized into N_π cascaded π -sections of length $\Delta x = \ell / N_\pi$. For each section $k \in \{1, \dots, N_\pi\}$, the lumped matrices are

$$\mathbf{R}_k = \mathbf{R} \Delta x, \quad \mathbf{L}_k = \mathbf{L} \Delta x, \quad \mathbf{G}_k = \mathbf{G} \Delta x, \quad \mathbf{C}_k = \mathbf{C} \Delta x. \quad (11)$$

The corresponding series and shunt elements in the Laplace domain are

$$\mathbf{Z}_k(s) = \mathbf{R}_k + s \mathbf{L}_k, \quad \mathbf{Y}_k(s) = \mathbf{G}_k + s \mathbf{C}_k, \quad (12)$$

with s the Laplace variable. Increasing N_π improves the approximation of distributed propagation and waveform distortion effects, which is relevant when wavefronts are distorted by network conditions [10].

3.6.3. Monitoring Signals and Notation

The monitored three-phase voltage and current vectors are defined in (1). These signals constitute the measured boundary information used for transient recording, time–frequency feature construction, and physics-regularized inference.

3.6.4. Visual Summary (Electrical Model and Discretization)

Figure 2 summarizes the adopted modeling hierarchy: distributed parameters ($\mathbf{R}, \mathbf{L}, \mathbf{C}, \mathbf{G}$) per unit length are mapped to a cascaded π -section network, which is then used for time-domain transient simulation.

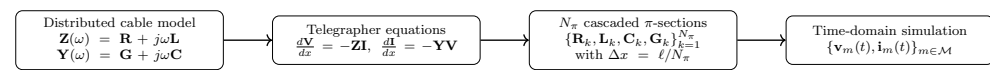


Figure 2. Electrical modeling hierarchy for underground cables: distributed parameters and telegrapher equations are approximated by cascaded π -sections for time-domain transient simulation.

3.7. Aging Parameterization

Underground cable aging is driven by combined thermal, electrical, and environmental stresses that progressively alter conductor losses and dielectric behavior [1,2,4]. Rather than assuming a single deterministic end-of-life, recent work emphasizes probabilistic lifetime assessment under thermal aging and the influence of uncertain operating conditions [4]. In parallel, diagnostic evidence such as partial discharge activity has been shown to correlate with insulation degradation progression [3]. Motivated by these findings, we introduce a normalized aging-stress index that enables the controlled stress testing of fault diagnosis and location under parameter drift.

3.7.1. Normalized Aging-Stress Index and Parameter Mapping

The scalar variable $\alpha \in [0, 1]$ is defined as a normalized degradation-stress coordinate. The value $\alpha = 0$ represents the nominal cable condition used to define the baseline per-unit-length matrices, whereas $\alpha = 1$ represents the upper bound of the controlled degradation domain considered in the simulations. This variable is not treated as a directly measured calendar age, lifetime fraction, or asset-health index. Instead, it provides a reproducible way to perturb cable parameters and to evaluate whether the diagnostic model remains robust under structured aging-induced parameter drift.

Let $\mathbf{R}_0$, $\mathbf{L}_0$, $\mathbf{C}_0$, and $\mathbf{G}_0$ denote the nominal per-unit-length resistance, inductance, capacitance, and conductance matrices, respectively. The aged matrices used in the simulation domain are parameterized as

$$\mathbf{R}(\alpha) = (1 + \kappa_R \alpha) \mathbf{R}_0, \quad (13)$$

$$\mathbf{L}(\alpha) = (1 + \kappa_L \alpha) \mathbf{L}_0, \quad (14)$$

$$\mathbf{C}(\alpha) = (1 - \kappa_C \alpha) \mathbf{C}_0, \quad (15)$$

$$\mathbf{G}(\alpha) = (1 + \kappa_G \alpha) \mathbf{G}_0, \quad (16)$$

where $\kappa_R, \kappa_L, \kappa_C, \kappa_G \geq 0$ are dimensionless sensitivity coefficients. Equations (13)–(16) preserve the nominal matrix coupling structure while imposing interpretable monotonic trends: increased effective series losses through $\mathbf{R}(\alpha)$, modified electromagnetic energy storage and propagation through $\mathbf{L}(\alpha)$ and $\mathbf{C}(\alpha)$, and increased dielectric leakage through $\mathbf{G}(\alpha)$. These trends are consistent with the qualitative effects of insulation degradation, thermal stress, and dielectric deterioration reported in cable-aging studies [1–4].

The coefficients $\kappa_R, \kappa_L, \kappa_C$, and κ_G should therefore be interpreted as controlled sensitivity parameters rather than field-calibrated aging constants. In the absence of asset-specific calibration data, their role is to define a bounded perturbation domain over which the learning models are compared under identical conditions. To avoid nonphysical parameter degeneration, the selected coefficients must satisfy

$$1 + \kappa_R \alpha > 0, \quad 1 + \kappa_L \alpha > 0, \quad 1 - \kappa_C \alpha > 0, \quad 1 + \kappa_G \alpha > 0, \quad \forall \alpha \in [0, 1]. \quad (17)$$

Thus, a sufficient condition is $\kappa_C < 1$, together with non-negative κ_R , κ_L , and κ_G . If the nominal matrices are positive definite or positive semidefinite according to their physical role, these scalar perturbations preserve the corresponding definiteness properties over the simulated aging interval.

The linear dependence on α in (13)–(16) can also be interpreted as a first-order approximation of a more general monotonic aging law. Specifically, one may write

$$\mathbf{R}(\alpha) = (1 + \kappa_R \psi_R(\alpha)) \mathbf{R}_0, \quad (18)$$

$$\mathbf{L}(\alpha) = (1 + \kappa_L \psi_L(\alpha)) \mathbf{L}_0, \quad (19)$$

$$\mathbf{C}(\alpha) = (1 - \kappa_C \psi_C(\alpha)) \mathbf{C}_0, \quad (20)$$

$$\mathbf{G}(\alpha) = (1 + \kappa_G \psi_G(\alpha)) \mathbf{G}_0, \quad (21)$$

where $\psi_q(0) = 0$, $\psi_q(1) = 1$, and $\psi_q(\alpha)$ is a monotonic shaping function for $q \in \{R, L, C, G\}$. The experiments in this paper use the first-order setting $\psi_q(\alpha) = \alpha$ to isolate the effect of controlled aging-driven drift without introducing additional calibration-dependent nonlinear parameters. Nonlinear choices, such as $\psi_q(\alpha) = \alpha^{\eta_q}$ with $\eta_q > 0$, can be used when laboratory, field, thermal-history, or partial-discharge data are available to identify asset-specific aging trajectories.

Real underground-cable aging may also be frequency dependent. A more general representation would replace the effective matrices in (13)–(16) by frequency-dependent quantities,

$$\begin{aligned} \mathbf{Z}(\omega, \alpha) &= \mathbf{R}(\omega, \alpha) + j\omega \mathbf{L}(\omega, \alpha), \\ \mathbf{Y}(\omega, \alpha) &= \mathbf{G}(\omega, \alpha) + j\omega \mathbf{C}(\omega, \alpha). \end{aligned} \quad (22)$$

This paper uses broadband effective parameters in the transient simulation to obtain a reproducible and computationally tractable validation domain. Therefore, the reported results quantify robustness to structured parameter drift—not the calibration accuracy of a specific aged cable asset.

The practical interpretation of the four sensitivity coefficients and the corresponding admissibility criteria are summarized in Table 1. The table distinguishes the physical trend represented by each coefficient from the criterion used to keep the simulated cable parameters within a bounded and physically admissible domain.

Table 1. Interpretation of the aging-sensitivity coefficients used in the controlled simulation domain.

Coeff.	Modeled Effect	Selection Criterion
κ_R	Increase of effective series losses.	Chosen to produce monotonic loss increase while preserving positive series resistance.
κ_L	Change in effective magnetic coupling and propagation characteristics.	Chosen as a bounded perturbation of the nominal inductive matrix without altering its coupling structure.
κ_C	Reduction of effective capacitive behavior associated with dielectric degradation.	Chosen with $\kappa_C < 1$ to ensure $\mathbf{C}(\alpha)$ remains nondegenerate for all $\alpha \in [0, 1]$.
κ_G	Increase of dielectric leakage and shunt losses.	Chosen to stress attenuation and damping effects while preserving a physically admissible shunt conductance.

Together, these coefficient definitions ensure that the aging parameterization is used as a controlled perturbation model for robustness assessment while avoiding the interpretation of κ_R , κ_L , κ_C , and κ_G as field-calibrated constants for a particular cable asset.

3.7.2. Aging-Aware π -Section Parameters

For a cable segment discretized into N_π sections, the aging-aware lumped parameters become

$$\begin{aligned}\mathbf{R}_k(\alpha) &= \mathbf{R}(\alpha) \Delta x, \\ \mathbf{L}_k(\alpha) &= \mathbf{L}(\alpha) \Delta x, \\ \mathbf{C}_k(\alpha) &= \mathbf{C}(\alpha) \Delta x, \\ \mathbf{G}_k(\alpha) &= \mathbf{G}(\alpha) \Delta x\end{aligned}\quad (23)$$

with $\Delta x = \ell / N_\pi$. This construction ensures that each simulated normalized aging-stress level modifies the time-domain cable network consistently with the per-unit-length parameterization. Since the same α is propagated from the continuous-domain matrices to every cascaded π -section, the resulting transient waveforms reflect a coherent aging-induced perturbation of propagation, attenuation, and reflection behavior. The construction also ensures that all methods are evaluated under the same aged network realization for each event so that performance differences are attributable to the diagnostic models rather than to inconsistent scenario generation.

3.7.3. Visual Summary (Aging Map)

Figure 3 illustrates the aging parameterization as a deterministic mapping from α to the per-unit-length matrices used in the electrical model.

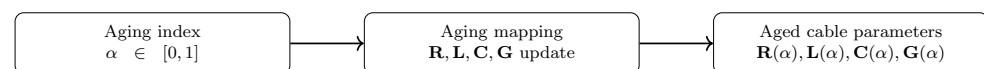


Figure 3. Aging parameterization: a normalized aging-stress index α is mapped to aged per-unit-length cable matrices used by the electrical model.

3.8. Role of the Aging Model in the Proposed Method

The above formulation serves two purposes in the proposed framework. First, it generates labeled transient waveforms under controlled parameter drift, allowing the diagnostic models to be evaluated across nominal and degraded cable conditions. Second, it provides an explicit physical structure that can be used in the physics-consistency regularization terms during learning. In particular, the dependence of $\mathbf{Z}(\omega, \alpha)$ and $\mathbf{Y}(\omega, \alpha)$ on α supplies a structured mechanism to create aging-induced domain shift and to test whether the proposed hybrid model maintains classification and localization performance when propagation and attenuation characteristics deviate from the nominal condition.

The relevance of aging for fault location can also be seen from the effective modal propagation velocity. For a weakly coupled modal approximation, the velocity can be interpreted as

$$v_{\text{eff}}(\alpha) \approx \frac{1}{\sqrt{L_{\text{eff}}(\alpha) C_{\text{eff}}(\alpha)}}, \quad (24)$$

where $L_{\text{eff}}(\alpha)$ and $C_{\text{eff}}(\alpha)$ denote effective modal inductive and capacitive parameters. A relative velocity perturbation produces an approximate distance bias

$$\Delta x \approx t_{\text{arr}} \Delta v_{\text{eff}}, \quad (25)$$

for an arrival time t_{arr} . Therefore, even bounded aging-induced changes in propagation and attenuation can bias TOA-based location estimates and alter the time–frequency features

used by learning-based methods. This motivates including α as a structured stress variable in the simulation and in the physics-consistency terms.

The aging model is not intended to replace asset-specific insulation diagnostics. Field deployment would require the calibration of α or replacement of α by measured condition indicators, such as thermal history, dielectric-loss measurements, partial-discharge features, or other cable-health diagnostics. Within the scope of this paper, however, the normalized aging-stress index provides a reproducible stress-testing variable for comparing traveling-wave, data-driven, physics-regularized, and hybrid diagnostic methods under identical simulated degradation conditions.

4. Dataset Generation and Fault Scenarios

This section formalizes the dataset generation process used to train and evaluate the proposed hybrid physics-informed and explainable framework within the controlled simulation domain. Each sample is produced by (i) selecting an operating condition, (ii) selecting a normalized cable degradation-stress level, (iii) inserting a fault event with controlled type, location, and resistance, and (iv) simulating the transient response to record three-phase voltage and current waveforms at monitoring locations. The design explicitly targets traveling-wave observables and waveform distortions that influence fault location and classification performance [5,6,10]. Moreover, the parameter drift induced by underground cable aging is incorporated to reflect controlled long-term degradation uncertainty rather than field-calibrated asset aging [1–4].

4.1. Fault Types and Locations

Let the feeder/cable path under study be parameterized by a one-dimensional coordinate $x \in [0, \ell]$, where ℓ is the total electrical length (or physical length, if a homogeneous propagation model is assumed). A fault is characterized by the tuple

$$\theta_f \triangleq (\mathcal{T}, x_f, R_f, t_f), \quad (26)$$

where $\mathcal{T}$ is the fault type, $x_f \in (0, \ell)$ is the fault location, $R_f > 0$ is the fault resistance, and t_f is the fault inception time.

4.1.1. Fault Type Set

We define a finite set of fault types

$$\mathcal{T} \in \mathcal{S}_{\mathcal{T}}, \quad (27)$$

where $\mathcal{S}_{\mathcal{T}}$ includes the relevant short-circuit classes for distribution systems. In practical dataset generation, $\mathcal{S}_{\mathcal{T}}$ is chosen to cover both ground-involved and phase-to-phase events, as traveling-wave signatures and time–frequency features can differ significantly across classes [5,6]. The dataset is built to enable multi-class fault classification jointly with continuous or discretized fault location estimation.

The fault set considered in this paper focuses on shunt short-circuit events, including single-line-to-ground, line-to-line, double-line-to-ground, three-phase, and three-phase-to-ground faults. These classes represent the primary short-circuit categories required for joint fault-type classification and location. Other events, such as open-conductor faults, intermittent arcing faults, evolving faults, simultaneous multi-location faults, or high-resistance insulation defects before complete breakdown, are not included in the present label space. These cases can occur in practical networks and are relevant for protection and asset-health monitoring; however, including them would require additional event models, labels, and validation data. This paper therefore uses the above short-circuit set as

a controlled diagnostic domain, while the proposed multi-task structure can be extended by enlarging $\mathcal{S}_T$ when additional event classes are available.

4.1.2. Location Parameterization

For reporting and sampling convenience, the fault location can be expressed as a normalized position

$$\lambda_f \triangleq \frac{x_f}{\ell} \in (0, 1), \quad (28)$$

or equivalently as a percentage along the feeder $100\lambda_f\%$. Dataset samples are generated by drawing λ_f from a prescribed distribution, for example uniform, over a constrained interval away from terminals to avoid degenerate reflections:

$$\lambda_f \sim \mathcal{U}(\lambda_{\min}, \lambda_{\max}), \quad 0 < \lambda_{\min} < \lambda_{\max} < 1. \quad (29)$$

This formulation ensures the coverage of interior locations where wavefront detection and distortion are informative [10].

4.1.3. Fault Resistance Model

The fault resistance R_f governs the magnitude and spectral content of the transient injection and thus strongly affects traveling-wave detection reliability and feature separability. We sample

$$R_f \sim \mathcal{U}(R_{\min}, R_{\max}), \quad 0 < R_{\min} < R_{\max}, \quad (30)$$

with bounds selected to include low-resistance, or strong, and higher-resistance, or weak, faults.

4.1.4. Fault Insertion as an Equivalent Boundary Condition

In time-domain simulation, the fault event is implemented as a shunt connection at node/location x_f with equivalent impedance $Z_f(t)$, which is typically modeled as

$$Z_f(t) = R_f \cdot u(t - t_f), \quad (31)$$

where $u(\cdot)$ is the unit step. This produces an abrupt change in the network boundary condition at $t = t_f$, initiating traveling waves whose arrival times and frequency-dependent dispersion contain the information exploited by fault location methods [5,6].

4.2. Simulation Scenarios

4.2.1. Scenario Parameter Vector

Each simulated sample is defined by a scenario vector

$$\boldsymbol{\theta} \triangleq (\boldsymbol{\theta}_f, \alpha, \boldsymbol{\theta}_{op}, \boldsymbol{\theta}_n), \quad (32)$$

where $\boldsymbol{\theta}_f$ is the fault descriptor in (26), $\alpha \in [0, 1]$ is the normalized aging-stress index, $\boldsymbol{\theta}_{op}$ encodes operating conditions, such as loading levels and, if applicable, distributed generation dispatch, and $\boldsymbol{\theta}_n$ encodes measurement noise settings.

4.2.2. Normalized Aging-Stress Sampling

To model degradation-driven parameter drift within the controlled simulation domain, each sample is assigned a normalized aging-stress level

$$\alpha \sim \mathcal{U}(0, 1), \quad (33)$$

and the aged per-unit-length parameters $\mathbf{R}(\alpha)$, $\mathbf{L}(\alpha)$, $\mathbf{C}(\alpha)$, and $\mathbf{G}(\alpha)$ are computed using Section 3.7. Uniform sampling is used to ensure balanced coverage of the full degradation-stress interval rather than to represent the empirical distribution of ages in a particular utility fleet. This enables systematic robustness assessment under aging-induced parameter drift and is consistent with treating cable degradation as uncertain and stress-dependent [1,2,4]. Partial-discharge-based degradation evidence further motivates testing a broad range of degradation conditions [3].

4.2.3. Operating Point Variability

Let θ_{op} denote the operating point parameters. At minimum, we represent operating variability through a load scaling factor γ_L :

$$\begin{aligned}\mathbf{p}_L &= \gamma_L \mathbf{p}_{L,0}, \\ \mathbf{q}_L &= \gamma_L \mathbf{q}_{L,0},\end{aligned}\tag{34}$$

where $\mathbf{p}_{L,0}$ and $\mathbf{q}_{L,0}$ are baseline nodal active and reactive demands, respectively, and γ_L is sampled from a prescribed interval $\gamma_L \in [\gamma_{L,\min}, \gamma_{L,\max}]$. This ensures that transients are generated under diverse prefault states, which can influence wavefront amplitude and detection thresholds.

4.2.4. Measurement Noise and SNR Definition

To emulate realistic measurement conditions, additive noise is injected into each recorded channel, including voltage and current. For a generic scalar signal $s(t)$, corresponding to one phase at one measurement location, we define the noisy observation

$$\tilde{s}(t) = s(t) + n(t),\tag{35}$$

where $n(t)$ is zero-mean noise with variance selected to meet a target signal-to-noise ratio (SNR). Using the energy-based definition,

$$\text{SNR}_{\text{dB}} = 10 \log_{10} \left(\frac{\sum_{t \in \mathcal{T}_w} s^2(t)}{\sum_{t \in \mathcal{T}_w} n^2(t)} \right),\tag{36}$$

over a time window $\mathcal{T}_w$ that contains the fault transient. For each sample, SNR_{dB} is drawn within a prescribed range and $n(t)$ is scaled accordingly.

4.2.5. Dataset Construction and Labeling

Let

$$\mathcal{D} = \left\{ \left(\mathbf{X}^{(i)}, \mathbf{y}^{(i)}, \zeta^{(i)} \right) \right\}_{i=1}^N\tag{37}$$

denote the final dataset, where $\mathbf{X}^{(i)}$ is the stacked time–frequency tensor constructed from the synchronized noisy voltage and current waveforms. The supervised target vector is

$$\mathbf{y}^{(i)} = \left(y_{\text{cls}}^{(i)}, y_{\mathcal{A}}^{(i)}, y_{\text{loc}}^{(i)} \right),\tag{38}$$

with

$$y_{\text{cls}}^{(i)} = \mathcal{T}^{(i)} \in \mathcal{S}_{\mathcal{T}},\tag{39}$$

$$y_{\mathcal{A}}^{(i)} = a^{(i)} \in \{1, \dots, A\},\tag{40}$$

$$y_{\text{loc}}^{(i)} = \lambda_f^{(i)} \in (0, 1).\tag{41}$$

The feeder-area label $a^{(i)}$ is assigned deterministically from the mutually exclusive area partition containing the simulated fault position. Thus, every event contains the three supervised targets required for fault-type classification, feeder-area identification, and continuous localization.

The auxiliary physics information associated with event i is

$$\zeta^{(i)} = (\alpha^{(i)}, t_0^{(i)}, \hat{\tau}^{(i)}), \quad (42)$$

where $\alpha^{(i)}$ is the normalized aging-stress index, $t_0^{(i)} = t_f^{(i)}$ is the simulator-defined fault-inception time, and

$$\hat{\tau}^{(i)} = \{\hat{\tau}_m^{(i)}\}_{m \in \mathcal{M}_{\text{valid}}^{(i)}} \quad (43)$$

contains the valid transient-arrival estimates obtained from the recorded waveforms. These auxiliary variables are used only to construct the propagation-consistency loss during training and are not additional prediction targets.

Accordingly, the learning-based models receive the same time–frequency input tensor and supervised labels, while the PINN and Hybrid configurations additionally use $\zeta^{(i)}$ during training. The raw voltage and current waveforms remain the source signals from which both $\mathbf{X}^{(i)}$ and the arrival-time estimates are obtained.

4.2.6. Algorithmic Procedure

Algorithm 2 summarizes the complete scenario-sampling, transient-simulation, pre-processing, and labeling procedure.

Algorithm 2: Dataset generation with aging-aware parameters, area labels, and physics metadata

Input: Network topology $(\mathcal{N}, \mathcal{E})$; nominal cable matrices $(\mathbf{R}_0, \mathbf{L}_0, \mathbf{C}_0, \mathbf{G}_0)$; discretization N_π ; monitoring set $\mathcal{M}$; feeder-area partition $\{\mathcal{A}_a\}_{a=1}^A$; numerical ranges specified in Section 7.3.

Output: Dataset $\mathcal{D} = \{(\mathbf{X}^{(i)}, \mathbf{y}^{(i)}, \zeta^{(i)})\}_{i=1}^N$.

for $i = 1$ **to** N **do**

Sample normalized aging-stress index $\alpha^{(i)} \sim \mathcal{U}(0, 1)$;
 Compute $\mathbf{R}(\alpha^{(i)})$, $\mathbf{L}(\alpha^{(i)})$, $\mathbf{C}(\alpha^{(i)})$, and $\mathbf{G}(\alpha^{(i)})$;
 Discretize the cable sections into N_π π -sections using (23);
 Sample operating point $\theta_{op}^{(i)}$, including the load factor $\gamma_L^{(i)}$;
 Sample fault type $\mathcal{T}^{(i)} \in \mathcal{S}_{\mathcal{T}}$;
 Sample fault location $\lambda_f^{(i)} \sim \mathcal{U}(\lambda_{\min}, \lambda_{\max})$ and determine the corresponding physical position $x_f^{(i)}$;
 Assign the feeder-area label $a^{(i)}$ from the unique partition satisfying $x_f^{(i)} \in \mathcal{A}_{a^{(i)}}$;
 Sample fault resistance $R_f^{(i)} \sim \mathcal{U}(R_{\min}, R_{\max})$;
 Set the fault-inception time $t_0^{(i)} = t_f^{(i)}$ and apply the fault using (31);
 Run the time-domain simulation and record $\{\mathbf{v}_m(t), \mathbf{i}_m(t)\}_{m \in \mathcal{M}}$;
 Sample $\text{SNR}_{\text{dB}}^{(i)}$ and inject additive noise using (35) and (36);
 Construct $\mathbf{X}^{(i)}$ using the STFT and channel-stacking procedure defined in Section 5.2;
 Detect $\hat{\tau}^{(i)} = \{\hat{\tau}_m^{(i)}\}_{m \in \mathcal{M}_{\text{valid}}^{(i)}}$ using (47);
 Set $\mathbf{y}^{(i)} \leftarrow (\mathcal{T}^{(i)}, a^{(i)}, \lambda_f^{(i)})$;
 Set $\zeta^{(i)} \leftarrow (\alpha^{(i)}, t_0^{(i)}, \hat{\tau}^{(i)})$;
 Add $(\mathbf{X}^{(i)}, \mathbf{y}^{(i)}, \zeta^{(i)})$ to $\mathcal{D}$.

4.2.7. Numerical Instantiation

The generic sampling procedure above is instantiated numerically in the case-study protocol of Section 7.3, where the feeder topology, monitored locations, dataset size, and scenario ranges are specified for the reported simulations.

5. Proposed Hybrid Physics-Informed XAI Framework

This section presents the proposed hybrid physics-informed and explainable framework for joint fault type classification and fault location in underground distribution networks under parameter drift induced by cable aging. The design targets two coupled challenges: (i) traveling-wave transients are strongly shaped by frequency-dependent propagation and waveform distortion [5,6,10], and (ii) underground cable aging alters the effective electrical parameters and thus the transient signatures [1–4]. Physics-informed learning is used to regularize the data-driven model so that predictions remain consistent with the governing cable dynamics (Section 3.6) and the aging parameterization (Section 3.7), while an explainability module provides feature-level attribution to support auditing and diagnostic interpretability [21].

To make the methodological separation among the evaluated configurations explicit, Table 2 summarizes the diagnostic principle, physics usage, and evaluation role of each method. This hierarchy is used to isolate the incremental contribution of supervised time–frequency learning, physics-consistency regularization, and the complete Hybrid formulation.

Table 2. Methodological hierarchy of the evaluated diagnostic configurations.

Configuration	Diagnostic Principle	Use of Physics	Role in the Evaluation
TW-TOA	Classical traveling-wave time-of-arrival estimation based on detected transient wavefronts.	Uses propagation-time interpretation and assumed wave velocity, but no learned representation.	Provides a physically interpretable reference for conventional TW-based localization.
Baseline	Supervised learning from the same transient time–frequency representation used by the learning-based models.	No explicit dynamic residual is imposed; equivalently, the physics-regularization weight is set to $\lambda_{\text{phy}} = 0$.	Isolates the performance of purely data-driven representation learning under the same dataset split and task definitions.
PINN	Physics-regularized learning using the same supervised task structure and an additional dynamic consistency penalty.	Uses the aging-aware cable matrices to construct $\mathcal{L}_{\text{phy}}$ during training.	Isolates the contribution of physics-consistency regularization relative to the purely data-driven baseline.
Hybrid	Complete proposed formulation combining transient time–frequency learning, aging-conditioned physics consistency, area-aware localization, multi-sensor information, and task-specific explanation.	Uses the aged cable model to regularize learning and to structure the diagnostic pipeline under parameter drift.	Evaluates the full integrated method proposed in this paper for simultaneous fault-type classification, area identification, and continuous fault localization.

Accordingly, the term Hybrid refers to the complete integration of data-driven transient representation learning with aging-aware physical structure and diagnostic explainability. It is therefore distinguished from the PINN configuration, which is used as an ablation baseline to evaluate the isolated effect of adding a physics residual to the supervised learning objective.

The evaluated configurations are intended to form an ablation-oriented benchmark hierarchy rather than an exhaustive survey of all possible machine-learning architectures. TW-TOA represents a conventional physically interpretable traveling-wave reference, Baseline isolates supervised time–frequency learning without explicit physics regularization, PINN isolates the effect of the dynamic residual, and Hybrid evaluates the complete proposed integration. Other data-driven baselines, such as wavelet-feature classifiers, support-vector machines, random forests, gradient-boosted trees, recurrent networks, or transformer-style architectures, are relevant alternatives for broader empirical benchmarking. However, including such models without matched inputs, task definitions, and training budgets can confound the specific contribution being tested here—namely, the incremental value of aging-aware physics regularization and area-aware multi-sensor fusion.

5.1. Traveling-Wave TOA Reference Implementation for Baseline Comparison

The TW-TOA configuration is implemented as a physically interpretable reference method based on detected wavefront arrival times. It is not a learning-based model. For each measured channel $s_m(t)$, a high-frequency residual is first extracted as

$$s_m^{\text{HF}}(t) = s_m(t) - \mathcal{L}_{\text{LP}}\{s_m(t)\}, \quad (44)$$

where $\mathcal{L}_{\text{LP}}\{\cdot\}$ is a low-pass smoothing operator used to remove the dominant fundamental component and slow variations. The transient envelope is then computed using the analytic signal:

$$e_m(t) = \left| s_m^{\text{HF}}(t) + j \mathcal{H}\{s_m^{\text{HF}}(t)\} \right|, \quad (45)$$

where $\mathcal{H}\{\cdot\}$ denotes the Hilbert transform. A pre-fault interval $\mathcal{T}_{\text{pre}}$ is used to estimate the envelope noise floor,

$$\mu_m^{\text{pre}} = \text{mean}_{t \in \mathcal{T}_{\text{pre}}} e_m(t), \quad \sigma_m^{\text{pre}} = \text{std}_{t \in \mathcal{T}_{\text{pre}}} e_m(t). \quad (46)$$

The first arrival time at sensor m is selected as

$$\hat{\tau}_m = \min \left\{ t \in \mathcal{T}_w : e_m(t) \geq \mu_m^{\text{pre}} + \rho \sigma_m^{\text{pre}} \right\}, \quad (47)$$

where $\rho > 0$ is a detection threshold multiplier. When the threshold crossing occurs between adjacent samples, a local quadratic interpolation of the envelope peak is used to obtain a sub-sample estimate of $\hat{\tau}_m$.

Given the detected arrivals, the TW-TOA reference jointly evaluates the feeder-area hypotheses and the candidate fault position within each area:

$$(\hat{a}_{\text{TOA}}, \hat{x}_{\text{TOA}}) = \arg \min_{\substack{a \in \{1, \dots, A\} \\ x \in \mathcal{A}_a}} \frac{\sum_{m \in \mathcal{M}_a^{\text{valid}}} w_m \left(\hat{\tau}_m - t_0 - \frac{d_m(x)}{\hat{v}(\alpha)} \right)^2}{\sum_{m \in \mathcal{M}_a^{\text{valid}}} w_m}, \quad (48)$$

where $\mathcal{M}_a^{\text{valid}} = \mathcal{M}_a \cap \mathcal{M}_{\text{valid}}$ is the subset of area-associated sensors for which a valid arrival time is detected, x is the ordered arc-length coordinate within area $\mathcal{A}_a$, $d_m(x)$ is the path distance from the candidate point to sensor m , $\hat{v}(\alpha)$ is the assumed aging-conditioned

propagation velocity, and $w_m > 0$ is a sensor-reliability weight. Normalization by the sum of the active sensor weights prevents areas with different numbers of available sensors from being favored solely because of the cardinality of their sensor subsets.

The resulting continuous estimate is mapped to the global normalized feeder coordinate as

$$\hat{\lambda}_{\text{TOA}} = \Phi_{\hat{a}_{\text{TOA}}} \left(\frac{\hat{x}_{\text{TOA}}}{\ell_{\hat{a}_{\text{TOA}}}} \right), \quad (49)$$

where $\ell_{\hat{a}_{\text{TOA}}}$ is the ordered arc length of the selected area and $\Phi_{\hat{a}_{\text{TOA}}}$ is the same local-to-global mapping used by the area-wise learning formulation. Therefore, TW-TOA does not use a separate learned area classifier: its predicted area is the area index associated with the minimum propagation-time residual in (48).

In the controlled simulation evaluation, t_0 and α are available from the generated scenario and are applied consistently to all TW-TOA test cases. In practical deployment, t_0 must be estimated from synchronized measured transients using a detector such as (47), while $\hat{v}(\alpha)$ must be obtained from a calibrated propagation model or an online condition-dependent estimate.

This baseline is intentionally sensitive to wavefront distortion, propagation-velocity uncertainty, measurement bandwidth, and sampling resolution. It therefore provides a conventional TW reference against which the data-driven, PINN, and Hybrid configurations can be compared under the same simulated scenarios.

5.2. Overall Architecture

Let $\{\mathbf{v}_m(t), \mathbf{i}_m(t)\}_{m \in \mathcal{M}}$ denote the recorded three-phase measurements (Section 3.6). The framework applies a time–frequency operator $\mathcal{T}\{\cdot\}$ to extract transient-localized representations; then, it performs joint inference of fault type and location with a physics-informed hybrid model $f_\theta(\cdot)$, and finally, it computes post hoc explanations $\mathcal{E}(\cdot)$ that quantify which time–frequency components drive the decision. The learning-based configurations in Table 2 use matched input representations, task definitions, and data splits; their differences arise from the inclusion or exclusion of physics-consistency regularization and from the additional area-aware and explainability components of the complete Hybrid formulation.

Formally, for each sensor $m \in \mathcal{M}$, modality $q \in \{v, i\}$, and phase $\phi \in \{a, b, c\}$, let

$$s_{m,q,\phi}[n], \quad n \in \{0, \dots, N_w - 1\}, \quad (50)$$

denote the corresponding discrete-time channel expressed in per-unit values. Each record contains $N_w = 4000$ samples obtained from the $T_w = 20$ ms observation window at $f_s = 200$ kHz.

The time–frequency operator $\mathcal{T}\{\cdot\}$ is the short-time Fourier transform (STFT). A periodic Hann window of length $L_w = 256$ samples is used:

$$w[n] = \frac{1}{2} - \frac{1}{2} \cos\left(\frac{2\pi n}{L_w}\right), \quad n = 0, \dots, L_w - 1. \quad (51)$$

Consecutive windows overlap by $L_{\text{ov}} = 192$ samples, corresponding to a 75% overlap and a hop size

$$H = L_w - L_{\text{ov}} = 64 \text{ samples} = 0.32 \text{ ms}. \quad (52)$$

Using an FFT length of $N_{\text{FFT}} = 256$, the STFT is defined as

$$\mathcal{T}\{s_{m,q,\phi}\}(k, r) = \sum_{n=0}^{L_w-1} s_{m,q,\phi}[n + rH] w[n] \exp\left(-j \frac{2\pi kn}{N_{\text{FFT}}}\right), \quad (53)$$

where

$$r = 0, \dots, N_\tau - 1, \quad N_\tau = 1 + \left\lfloor \frac{N_w - L_w}{H} \right\rfloor = 59. \quad (54)$$

The frequency spacing is

$$\Delta f = \frac{f_s}{N_{\text{FFT}}} = 781.25 \text{ Hz}. \quad (55)$$

Only the frequency interval from 0 to 50 kHz is retained. Therefore,

$$f_k = k\Delta f, \quad k = 0, \dots, 64, \quad N_f = 65. \quad (56)$$

The magnitude spectrum is logarithmically compressed as

$$X_{m,q,\phi}(f_k, \tau_r) = \log(\varepsilon + |\mathcal{T}\{s_{m,q,\phi}\}(k, r)|), \quad \varepsilon = 10^{-8}. \quad (57)$$

The logarithm is applied after the conversion of voltage and current signals to their corresponding per-unit bases. No sample-wise min-max scaling, z-score normalization, envelope extraction, frequency interpolation, or image resizing is subsequently applied to the learning inputs.

The complete tensor is constructed using the fixed channel order sensor–modality–phase with sensors ordered as $\{S_0, S_3, S_6, S_9, S_{12}\}$, modalities ordered as $\{v, i\}$, and phases ordered as $\{a, b, c\}$:

$$\mathbf{X} = \text{stack}\left(\{X_{m,q,\phi}\}_{m \in \mathcal{M}, q \in \{v,i\}, \phi \in \{a,b,c\}}\right) \in \mathbb{R}^{C_{\text{in}} \times N_f \times N_\tau}, \quad (58)$$

where

$$C_{\text{in}} = |\mathcal{M}| \times 2 \times 3 = 30. \quad (59)$$

Thus, every learning-based configuration receives an input tensor of dimensions

$$\mathbf{X} \in \mathbb{R}^{30 \times 65 \times 59}. \quad (60)$$

The term raw signal refers to the original per-unit time-domain voltage and current channels before the STFT. High-frequency residuals and HF-envelope signals are used only for transient visualization and onset interpretation in the results section; they are not substituted for the STFT tensors supplied to the learning models.

The proposed Hybrid model uses the complete time–frequency tensor for fault-type and feeder-area identification, while area-dependent channel selection is used for local fault-position regression. The shared representation and the two classification outputs are

$$\begin{aligned} \mathbf{z} &= e_\theta(\mathbf{X}), \\ \hat{\mathbf{p}} &= c_\theta(\mathbf{z}) \in [0, 1]^{|S_T|}, \\ \hat{\mathbf{p}}_{\mathcal{A}} &= g_\theta(\mathbf{z}) \in [0, 1]^A, \end{aligned} \quad (61)$$

where e_θ is the shared encoder, c_θ is the fault-type head, and g_θ is the feeder-area head.

For each feeder area $\mathcal{A}_a$, the fixed channel-selection operator $\mathcal{P}_a$ retains the corresponding voltage and current channels:

$$\mathbf{X}_a = \mathcal{P}_a(\mathbf{X}), \quad a \in \{1, \dots, A\}. \quad (62)$$

The selected tensor is processed by the shared-weight encoder and the corresponding local regression head:

$$\hat{u}_a = h_{a,\theta}(e_\theta(\mathbf{X}_a)) \in (0, 1), \quad (63)$$

where $\hat{u}_a$ is the normalized in-area coordinate. Each local output is transformed into a candidate global normalized location through the topology-dependent mapping

$$\hat{\lambda}_a = \Phi_a(\hat{u}_a), \quad a \in \{1, \dots, A\}. \quad (64)$$

The predicted feeder area is

$$\hat{a} = \arg \max_{a \in \{1, \dots, A\}} \hat{p}_{\mathcal{A}_a}, \quad (65)$$

and the final end-to-end location estimate is

$$\hat{\lambda} = \hat{\lambda}_{\hat{a}} = \Phi_{\hat{a}}(\hat{u}_{\hat{a}}). \quad (66)$$

During inference, no ground-truth area information or oracle correction is used. Consequently, an incorrect area prediction directly selects an incorrect local regression head and local-to-global mapping, and its effect is therefore included in the reported localization error.

The explanation module computes separate attribution maps for the predicted fault type and the final location output:

$$(\mathbf{A}_{\text{cls}}, \mathbf{A}_{\text{loc}}) = \mathcal{E}(f_\theta, \mathbf{X}), \quad (67)$$

enabling inspection of the sensor, phase, time, and frequency regions that contribute most strongly to the fault-type and location decisions.

The complete end-to-end architecture is presented in Figure 4. The diagram distinguishes the shared time–frequency encoder, the fault-type and feeder-area classification heads, the area-dependent sensor-selection operators, the six local regression heads, and the local-to-global location mapping. The corresponding training and inference procedures are formalized in Algorithm 3.

5.3. Physics-Informed Learning Strategy

The learning strategy combines (i) supervised multi-task objectives and (ii) physics-consistency regularization derived from the underground cable model and discretized dynamics. Physics-informed neural networks (PINNs) and related approaches motivate augmenting purely data-driven objectives with governing-equation residuals [16–18,20], while physics-informed fault diagnosis surveys emphasize model-based constraints to improve generalization under distribution shift [29]. Here, the constraints are instantiated using the discretized cable representation (Section 3.6) and the aging parameterization (Section 3.7).

5.3.1. Multi-Task Supervised Objective

Let $y_{\text{cls}} \in \mathcal{S}_{\mathcal{T}}$ denote the ground-truth fault type, $y_{\mathcal{A}} \in \{1, \dots, A\}$ the ground-truth feeder area, and $y_{\text{loc}} = \lambda_f \in (0, 1)$ the ground-truth global normalized fault location. The supervised objective jointly trains the fault-type head, the area-identification head, and the area-specific localization heads:

$$\mathcal{L}_{\text{sup}}(\theta) = \mathcal{L}_{\text{cls}} + \gamma \mathcal{L}_{\text{area}} + \beta \mathcal{L}_{\text{loc}}, \quad (68)$$

where $\gamma > 0$ and $\beta > 0$ weight the area-identification and localization terms, respectively.

The fault-type classification loss is

$$\mathcal{L}_{\text{cls}} = - \sum_{c \in \mathcal{S}_{\mathcal{T}}} \mathbb{I}[c = y_{\text{cls}}] \log \hat{p}_c, \quad (69)$$

and the feeder-area classification loss is

$$\mathcal{L}_{\text{area}} = - \sum_{a=1}^A \mathbb{I}[a = y_{\mathcal{A}}] \log \hat{p}_{\mathcal{A}_a}. \quad (70)$$

During training, the ground-truth area activates the corresponding local regression head. Its normalized in-area prediction is mapped to the global feeder coordinate as

$$\hat{\lambda}_{y_{\mathcal{A}}} = \Phi_{y_{\mathcal{A}}}(\hat{u}_{y_{\mathcal{A}}}), \quad (71)$$

and the localization loss is

$$\mathcal{L}_{\text{loc}} = \left(\hat{\lambda}_{y_{\mathcal{A}}} - y_{\text{loc}} \right)^2. \quad (72)$$

Accordingly, each local regression head is updated only by training events belonging to its associated feeder area. During inference, the ground-truth area is unavailable, and the final location is obtained exclusively through the area predicted by (65); therefore, area-identification errors are propagated to the reported end-to-end localization error.

Thus, each local regression head is trained only with events belonging to its corresponding feeder area, while the area head is trained independently through (70). During test-time inference, the ground-truth area is not used and routing is performed exclusively through the predicted area in (65).

5.3.2. Physics-Consistency Regularization Through Propagation Residuals

To ensure that the physics loss contributes gradients to the trainable parameters, the residual is defined explicitly as a function of the location predicted by the learning model. For each training event, the transient arrival time $\hat{\tau}_m$ at sensor m is extracted from the measured waveform using the onset-detection procedure in (47). Let $\hat{x}_f(\theta)$ denote the physical feeder position associated with the normalized model output $\hat{\lambda}(\theta)$. The propagation-consistency residual at sensor m is defined as

$$r_m(\theta; \alpha) = \hat{\tau}_m - t_0 - \frac{d_m(\hat{x}_f(\theta))}{\hat{v}(\alpha)}, \quad m \in \mathcal{M}_{\text{valid}}, \quad (73)$$

where t_0 is the fault-inception time, $d_m(\hat{x}_f)$ is the path distance between the predicted fault position and sensor m , $\hat{v}(\alpha)$ is the aging-conditioned effective propagation velocity, and $\mathcal{M}_{\text{valid}} \subseteq \mathcal{M}$ contains the sensors for which a valid transient onset is detected.

The corresponding physics-consistency loss is

$$\mathcal{L}_{\text{phy}}(\theta) = \frac{1}{\sum_{m \in \mathcal{M}_{\text{valid}}} w_m} \sum_{m \in \mathcal{M}_{\text{valid}}} w_m [r_m(\theta; \alpha)]^2, \quad (74)$$

where $w_m > 0$ is a sensor-reliability weight. Unlike a residual computed exclusively from simulator states, (74) depends directly on the trainable location output. Its gradient is

$$\nabla_{\theta} \mathcal{L}_{\text{phy}} = - \frac{2}{\hat{v}(\alpha) \sum_{m \in \mathcal{M}_{\text{valid}}} w_m} \sum_{m \in \mathcal{M}_{\text{valid}}} w_m r_m \frac{\partial d_m(x)}{\partial x} \Big|_{x=\hat{x}_f} \nabla_{\theta} \hat{x}_f(\theta). \quad (75)$$

Therefore, the physics term produces a nonzero training gradient whenever the propagation residual and the sensitivity of the predicted location are nonzero. The path-distance function is evaluated on the feeder branch associated with the predicted position and is piecewise differentiable along each cable section.

During simulation-based training, t_0 and α are obtained from the generated scenario, whereas $\hat{\tau}_m$ is extracted from the recorded voltage or current transient. Internal simulator state vectors are not used in (73). During online inference, the trained diagnostic model requires only the measurement tensor $\mathbf{X}$; the physics loss is not evaluated. For field implementation, t_0 must be estimated from synchronized measured transients, and $\hat{v}(\alpha)$ must be obtained from a calibrated cable model or condition-dependent propagation estimate. This output-dependent formulation follows the general physics-informed principle of constraining trainable predictions through governing physical relations [16,18,20].

5.3.3. Complete Training Objective

The overall objective combines supervised and physics-informed terms:

$$\min_{\theta} \mathcal{L}(\theta) = \mathcal{L}_{\text{sup}}(\theta) + \lambda_{\text{phy}} \mathcal{L}_{\text{phy}}(\theta) + \lambda_{\text{reg}} \|\theta\|_2^2, \quad (76)$$

where $\lambda_{\text{phy}} \geq 0$ controls the strength of physics regularization and $\lambda_{\text{reg}} \geq 0$ is a standard weight decay term. This formulation follows the general physics-informed paradigm of augmenting data losses with equation residuals [16,29].

5.4. Neural Network Architecture and Training Details

The learning-based configurations use a convolutional encoder composed of three consecutive blocks. Each block contains a 3×3 convolution, batch normalization, a ReLU activation, and 2×2 max pooling. The resulting feature maps are flattened and passed to a shared fully connected layer with 256 neurons.

The full $C_{\text{in}} = 30$ -channel tensor is used by two common classification heads. The fault-type head contains $|\mathcal{S}_{\mathcal{T}}| = 11$ softmax outputs, whereas the area-identification head contains $A = 6$ softmax outputs.

For the Hybrid configuration, area-dependent localization is implemented using fixed channel-selection operators $\mathcal{P}_a$. Each operator retains the voltage and current channels associated with the sensor subset $\mathcal{M}_a$ defined for each feeder area and masks the remaining sensor channels:

$$\mathbf{X}_a = \mathcal{P}_a(\mathbf{X}), \quad a \in \{1, \dots, A\}. \quad (77)$$

The selected tensor $\mathbf{X}_a$ is processed by the shared-weight convolutional encoder and the corresponding area-specific regression head. Each local head uses a sigmoid output to produce $\hat{u}_a \in (0, 1)$. The mappings Φ_a subsequently convert these local outputs into the global normalized coordinate used for evaluation.

The evaluated learning configurations are distinguished as follows. The Baseline uses the common fault-type and area heads together with one global location-regression head and sets $\lambda_{\text{phy}} = 0$. The PINN retains the same global supervised heads and adds the propagation-consistency loss. The proposed Hybrid retains the fault-type and area heads, replaces the single global location head with six area-specific local heads, applies the sensor-selection operators in (77), and includes the propagation-consistency loss. Consequently, the architecture and routing differences among the evaluated configurations are stated explicitly rather than being treated as identical output structures.

Figure 4 summarizes the complete Hybrid architecture, including the fault-type head, area-identification head, area-dependent sensor selection, and local regression heads.

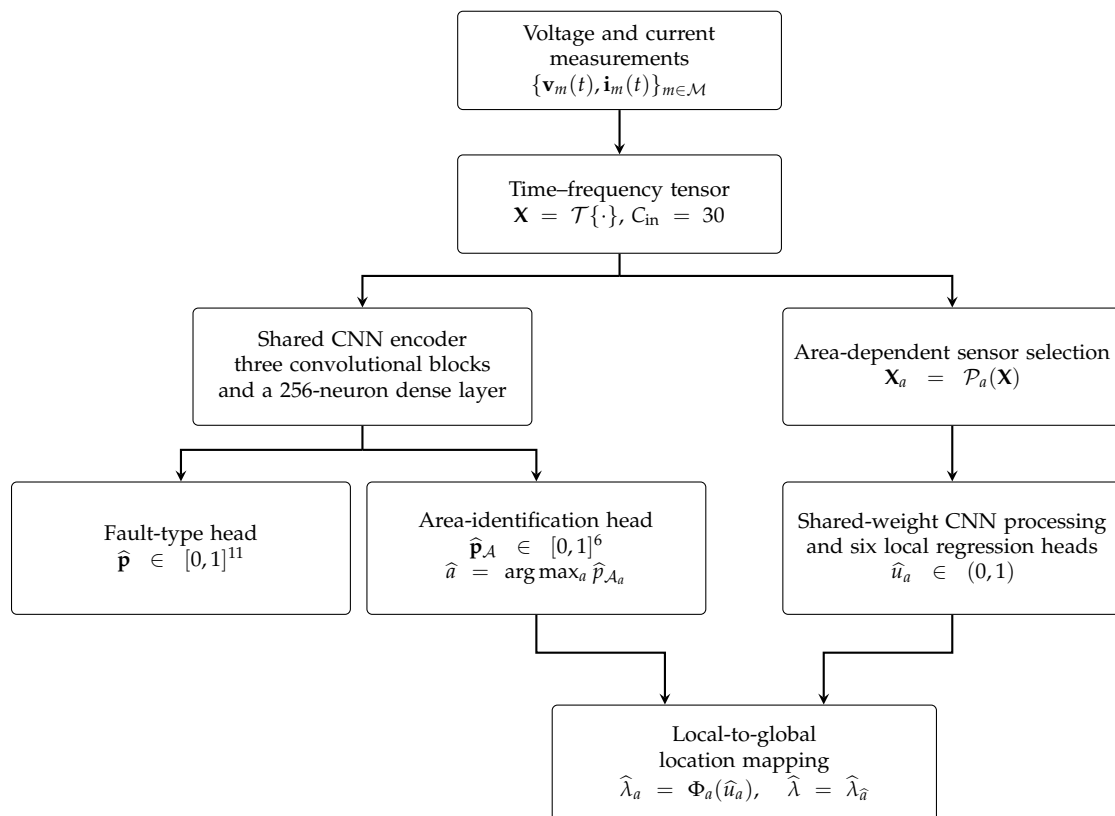


Figure 4. Unified architecture of the proposed Hybrid model. The complete time–frequency tensor is processed by the shared encoder for fault-type and feeder-area identification. In parallel, fixed area-dependent channel selectors retain the sensor subsets associated with each feeder area. The selected tensors are processed using shared-weight convolutional operations and six local regression heads. During inference, the predicted area selects the corresponding local estimate, which is subsequently transformed into the global normalized fault location.

The model is trained in MATLAB R2025b using the Adam optimizer with a learning rate of 10^{-3} , a batch size of 64, and early stopping based on validation loss with a patience of 15 epochs. Training is conducted for a maximum of 150 epochs. The physics-regularization weight λ_{phy} is selected through a validation-grid search over $\{0, 10^{-4}, 10^{-3}, 10^{-2}\}$. The selected value is then kept fixed for all test-set evaluations to avoid test-set leakage.

5.5. Explainable AI Module

Explainability is integrated to provide transparent evidence for both fault type classification and location inference. Since the inputs are time–frequency tensors, explanations should identify (i) the specific time intervals around wavefront arrivals and (ii) the frequency bands that contribute most to decisions. Recent work demonstrates that explainable AI can guide and audit fault diagnostics, including unsupervised fault detection contexts [21].

5.5.1. Attribution Targets

Let $\mathbf{X} \in \mathbb{R}^{H \times W \times C}$ denote the time–frequency tensor (height H frequency bins, width W time bins, channels C corresponding to phases and measurement locations). We compute explanations for

$$\text{Classification score: } s_{\text{cls}}(\mathbf{X}) \triangleq \hat{p}_{y^*}, \quad (78)$$

$$\text{Location output: } s_{\text{loc}}(\mathbf{X}) \triangleq \hat{\lambda}, \quad (79)$$

where $y^* = \arg \max_c \hat{p}_c$.

5.5.2. Gradient-Based Explanations (Grad-CAM-Style)

For convolutional backbones, gradient-based class activation mapping produces spatial attributions over feature maps. Let $\mathbf{F}^l \in \mathbb{R}^{H_l \times W_l \times K_l}$ be the feature maps at layer l . The Grad-CAM weights for score $s(\cdot)$ are

$$\alpha_k^{(l)} = \frac{1}{H_l W_l} \sum_{i=1}^{H_l} \sum_{j=1}^{W_l} \frac{\partial s(\mathbf{X})}{\partial F_{ij,k}^l}, \quad (80)$$

and the corresponding attribution map is

$$\mathbf{A}^{(l)} = \text{ReLU} \left(\sum_{k=1}^{K_l} \alpha_k^{(l)} \mathbf{F}_{:, :, k}^l \right), \quad (81)$$

which is then upsampled to the input resolution to align with $\mathbf{X}$. Computing $\mathbf{A}^{(l)}$ for both s_{cls} and s_{loc} yields distinct explanations for classification and location tasks.

5.5.3. Feature-Importance Explanations (SHAP-Style)

Complementarily, SHAP-based explanations quantify the contributions of input features to the output by comparing model outputs under feature perturbations. Let ϕ_q denote the SHAP value for feature group q (e.g., a time–frequency patch or channel group). The explanation is the vector $\boldsymbol{\phi} = (\phi_1, \dots, \phi_Q)$ satisfying an additive decomposition

$$s(\mathbf{X}) \approx \phi_0 + \sum_{q=1}^Q \phi_q, \quad (82)$$

where ϕ_0 is a baseline output and Q is the number of feature groups. In time–frequency settings, grouping features into structured patches preserves physical interpretability (time localization and frequency selectivity), enabling domain experts to validate that the model relies on plausible wavefront signatures rather than noise artifacts.

5.5.4. Interpretation Protocol

For each predicted event, the explainability module outputs the following:

- A time–frequency attribution map $\mathbf{A}_{\text{cls}}$ supporting the predicted fault type;
- A time–frequency attribution map $\mathbf{A}_{\text{loc}}$ supporting the predicted location estimate;
- Channel-wise importances indicating which measurement locations $m \in \mathcal{M}$ and which phases contribute most.

These outputs provide a structured audit trail that can be compared against expected traveling-wave behavior and known dispersion/distortion effects [5,10], supporting trust and diagnosis in deployment [21].

5.5.5. Attribution-Validation Metrics for Model-Native Explanations

To avoid ambiguity between illustrative saliency overlays and model-native explanations, the following diagnostics define a quantitative validation protocol for attribution maps computed directly from the trained model outputs. Let $\mathbf{A}_{n,q} \in \mathbb{R}_+^{N_f \times N_\tau}$ denote a non-negative model-native attribution map for sample n and task $q \in \{\text{cls}, \text{loc}\}$, where N_f is the number of frequency bins and N_τ is the number of time frames. The normalized attribution mass is defined as

$$\tilde{A}_{n,q}(f_k, \tau_r) = \frac{A_{n,q}(f_k, \tau_r)}{\sum_{k=1}^{N_f} \sum_{r=1}^{N_\tau} A_{n,q}(f_k, \tau_r) + \epsilon}, \quad (83)$$

where $\epsilon > 0$ avoids numerical division by zero.

The first diagnostic is the transient-attribution ratio, which measures whether attribution mass is concentrated around the known simulated fault-inception instant $t_{f,n}$:

$$\text{TAR}_{n,q} = \sum_{k=1}^{N_f} \sum_{\tau_r \in \mathcal{W}(t_{f,n})} \tilde{A}_{n,q}(f_k, \tau_r), \quad (84)$$

where $\mathcal{W}(t_{f,n}) = \{\tau_r : |\tau_r - t_{f,n}| \leq \Delta t_{\text{xai}}\}$ is a short transient window centered at fault inception. High values of $\text{TAR}_{n,q}$ indicate that the explanation is concentrated around the fault-initiated transient rather than over stationary background regions.

The second diagnostic is the high-frequency attribution ratio,

$$\text{HFAR}_{n,q} = \sum_{f_k \in \mathcal{B}_{\text{HF}}} \sum_{r=1}^{N_\tau} \tilde{A}_{n,q}(f_k, \tau_r), \quad (85)$$

where $\mathcal{B}_{\text{HF}}$ denotes the frequency band used for transient analysis. This metric evaluates whether the explanation emphasizes the high-frequency content associated with traveling-wave-like excitation.

The third diagnostic is attribution compactness. It is computed from the normalized entropy of the attribution map,

$$H_{n,q} = -\frac{1}{\log(N_f N_\tau)} \sum_{k=1}^{N_f} \sum_{r=1}^{N_\tau} \tilde{A}_{n,q}(f_k, \tau_r) \log(\tilde{A}_{n,q}(f_k, \tau_r) + \epsilon), \quad (86)$$

and reported as

$$\text{Comp}_{n,q} = 1 - H_{n,q}. \quad (87)$$

Higher compactness indicates that the explanation is localized in a smaller set of time–frequency regions, which is expected for fault-initiated transient evidence.

The fourth diagnostic is attribution stability. If $\mathbf{A}_{n,q}$ and $\mathbf{A}'_{n,q}$ denote attribution maps obtained from the original sample and from a small admissible perturbation of the same sample, respectively, their stability is measured by cosine similarity:

$$\text{Stab}_{n,q} = \frac{\langle \text{vec}(\mathbf{A}_{n,q}), \text{vec}(\mathbf{A}'_{n,q}) \rangle}{\|\text{vec}(\mathbf{A}_{n,q})\|_2 \|\text{vec}(\mathbf{A}'_{n,q})\|_2 + \epsilon}. \quad (88)$$

Finally, occlusion-based faithfulness is evaluated for the classification task. Let $s_{\text{cls}}(\mathbf{X}_n)$ be the predicted-class score for the original input, $s_{\text{cls}}(\mathbf{X}_n^{\text{top}})$ the score after occluding the top- η most attributed time–frequency cells, and $s_{\text{cls}}(\mathbf{X}_n^{\text{rand}})$ the score after occluding the same number of randomly selected cells. The attribution is considered faithful when

$$\Delta s_n^{\text{top}} = s_{\text{cls}}(\mathbf{X}_n) - s_{\text{cls}}(\mathbf{X}_n^{\text{top}}) \quad (89)$$

is consistently larger than

$$\Delta s_n^{\text{rand}} = s_{\text{cls}}(\mathbf{X}_n) - s_{\text{cls}}(\mathbf{X}_n^{\text{rand}}). \quad (90)$$

This verifies that removing the most attributed regions has a stronger effect on the diagnostic score than removing randomly selected regions of equal size. These metrics are meaningful only when $\mathbf{A}_{n,q}$ is computed from the trained diagnostic model f_θ and when the occluded inputs are re-evaluated by the same model. Therefore, proxy or schematic attribution overlays should not be used to report numerical faithfulness, stability, or concentration scores.

Algorithm 3: Hybrid Physics-Regularized Area-Aware Fault Diagnosis

Input: Training set $\mathcal{D} = \{(\mathbf{X}^{(i)}, y_{\text{cls}}^{(i)}, y_{\mathcal{A}}^{(i)}, y_{\text{loc}}^{(i)}, \alpha^{(i)}, t_0^{(i)}, \hat{\tau}^{(i)})\}_{i=1}^N$; area selectors $\{\mathcal{P}_a\}_{a=1}^A$; local-to-global mappings $\{\Phi_a\}_{a=1}^A$; loss weights $\beta, \gamma, \lambda_{\text{phy}}$, and λ_{reg} .
Output: Trained parameters θ ; fault type $\hat{y}_{\text{cls}}$; feeder area $\hat{a}$; global normalized location $\hat{\lambda}$; and attribution maps $\mathbf{A}_{\text{cls}}$ and $\mathbf{A}_{\text{loc}}$.

Offline training;

Initialize the shared encoder e_θ , the fault-type head c_θ , the area head g_θ , and the local regression heads $\{h_{a,\theta}\}_{a=1}^A$;

foreach mini-batch $\mathcal{B} \subset \mathcal{D}$ **do**

foreach sample $i \in \mathcal{B}$ **do**

$\mathbf{z}^{(i)} \leftarrow e_\theta(\mathbf{X}^{(i)})$;

$\hat{\mathbf{p}}^{(i)} \leftarrow c_\theta(\mathbf{z}^{(i)})$;

$\hat{\mathbf{p}}_{\mathcal{A}}^{(i)} \leftarrow g_\theta(\mathbf{z}^{(i)})$;

for $a = 1$ **to** A **do**

$\mathbf{X}_a^{(i)} \leftarrow \mathcal{P}_a(\mathbf{X}^{(i)})$;

$\mathbf{z}_a^{(i)} \leftarrow e_\theta(\mathbf{X}_a^{(i)})$;

$\hat{u}_a^{(i)} \leftarrow h_{a,\theta}(\mathbf{z}_a^{(i)})$;

$\hat{\lambda}_a^{(i)} \leftarrow \Phi_a(\hat{u}_a^{(i)})$;

$\mathcal{L}_{\text{cls}}^{(i)} \leftarrow -\log \hat{p}_{y_{\text{cls}}^{(i)}}^{(i)}$;

$\mathcal{L}_{\text{area}}^{(i)} \leftarrow -\log \hat{p}_{y_{\mathcal{A}}^{(i)}}^{(i)}$;

$\mathcal{L}_{\text{loc}}^{(i)} \leftarrow \left(\hat{\lambda}_{y_{\mathcal{A}}^{(i)}}^{(i)} - y_{\text{loc}}^{(i)} \right)^2$;

 Map $\hat{\lambda}_{y_{\mathcal{A}}^{(i)}}^{(i)}$ to the corresponding physical feeder position $\hat{x}_f^{(i)}$;

$r_m^{(i)} \leftarrow \hat{\tau}_m^{(i)} - t_0^{(i)} - \frac{d_m(\hat{x}_f^{(i)})}{\hat{v}(\alpha^{(i)})}, \quad m \in \mathcal{M}_{\text{valid}}^{(i)}$;

$\mathcal{L}_{\text{phy}}^{(i)} \leftarrow \frac{\sum_{m \in \mathcal{M}_{\text{valid}}^{(i)}} w_m (r_m^{(i)})^2}{\sum_{m \in \mathcal{M}_{\text{valid}}^{(i)}} w_m}$;

$\mathcal{L}^{(i)} \leftarrow \mathcal{L}_{\text{cls}}^{(i)} + \gamma \mathcal{L}_{\text{area}}^{(i)} + \beta \mathcal{L}_{\text{loc}}^{(i)} + \lambda_{\text{phy}} \mathcal{L}_{\text{phy}}^{(i)} + \lambda_{\text{reg}} \|\theta\|_2^2$;

 Update θ by minimizing $|\mathcal{B}|^{-1} \sum_{i \in \mathcal{B}} \mathcal{L}^{(i)}$;

Online inference;

Construct $\mathbf{X}$ from synchronized voltage and current measurements;

$\mathbf{z} \leftarrow e_\theta(\mathbf{X})$;

$\hat{\mathbf{p}} \leftarrow c_\theta(\mathbf{z})$;

$\hat{\mathbf{p}}_{\mathcal{A}} \leftarrow g_\theta(\mathbf{z})$;

$\hat{y}_{\text{cls}} \leftarrow \arg \max_c \hat{p}_c$;

$\hat{a} \leftarrow \arg \max_a \hat{p}_{\mathcal{A}_a}$;

$\mathbf{X}_{\hat{a}} \leftarrow \mathcal{P}_{\hat{a}}(\mathbf{X})$;

$\hat{u}_{\hat{a}} \leftarrow h_{\hat{a},\theta}(e_\theta(\mathbf{X}_{\hat{a}}))$;

$\hat{\lambda} \leftarrow \Phi_{\hat{a}}(\hat{u}_{\hat{a}})$;

Use the predicted area $\hat{a}$ directly, without oracle routing;

Compute model-derived attribution maps $\mathbf{A}_{\text{cls}}$ and $\mathbf{A}_{\text{loc}}$ for $\hat{y}_{\text{cls}}$ and $\hat{\lambda}$, respectively;

return $\hat{y}_{\text{cls}}, \hat{a}, \hat{\lambda}, \mathbf{A}_{\text{cls}}, \mathbf{A}_{\text{loc}}$.

6. Algorithm Description

This section provides a rigorous end-to-end description of the proposed hybrid physics-informed and explainable fault diagnosis algorithm. The procedure operates on multi-location

three-phase measurements, extracts transient-localized time–frequency representations, trains a multi-task model under explicit physics-consistency regularization informed by the aged underground cable model, and produces both predictions and explanatory attribution maps.

7. Case Study: Large-Scale Aging-Aware Underground Distribution Feeder

This section defines a large-scale aging-aware underground distribution feeder case study used to evaluate the proposed hybrid physics-regularized and explainable framework under network branching, traveling-wave-like transient propagation, and aging-driven parameter drift of underground cables. The case study is synthetic but explicitly specified in terms of topology, area partitioning, sensor placement, sampling ranges, dataset size, and split protocol. Therefore, the results should be interpreted as simulation-based validation rather than field-data validation.

7.1. Study System Topology and Area Partitioning

The feeder is represented as a connected graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ denotes electrical nodes, including junctions and terminations, and $\mathcal{E}$ denotes underground cable sections. Each cable section $e \in \mathcal{E}$ has length ℓ_e and per-unit-length parameter matrices $\mathbf{R}_e(\alpha)$, $\mathbf{L}_e(\alpha)$, $\mathbf{C}_e(\alpha)$, and $\mathbf{G}_e(\alpha)$ computed from the normalized aging-stress index $\alpha \in [0, 1]$ using the aging parameterization in Section 3.7. Each section is discretized into $N_{\pi,e}$ cascaded π -sections according to the underground cable model in Section 3.6.

The feeder is partitioned into $A = 6$ non-overlapping areas to create a coarse-to-fine localization structure aligned with the sensing geometry and branching layout. The number of areas is selected so that each partition contains a physically contiguous feeder region with distinguishable multi-sensor transient observability while avoiding excessively small areas that would produce unstable labels near junctions. Boundaries are placed at or near electrically meaningful points, such as sensor locations, feeder junctions, or changes in branching structure. Thus, each area groups cable sections with similar dominant propagation paths to the available sensors. This partitioning supports area-balanced sampling, reduces ambiguity in the global regression task, and enables local in-area localization after the coarse area has been identified.

Figure 5 shows the case-study layout. The colored regions indicate mutually exclusive and collectively exhaustive areas, while black-filled nodes denote synchronized voltage/current sensors.

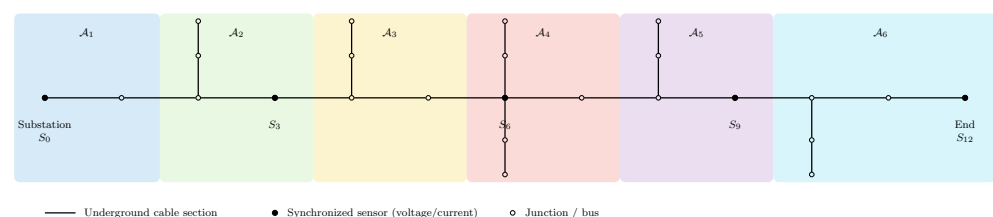


Figure 5. The large-scale underground feeder is partitioned into $A = 6$ mutually exclusive and collectively exhaustive areas to enable area-wise coarse-to-fine fault localization in a branched network. Each lateral is assigned uniquely to the area of its root junction to avoid ambiguity in area labeling.

The five monitoring locations represent synchronized transient voltage/current recorders installed at electrically meaningful points, such as the substation, switching nodes, or feeder junctions. They should not be interpreted as sensors installed at every cable section. This placement defines a multi-ended observability configuration for evaluating the proposed method under branching and area-dependent sensing. Practical de-

ploysments may use fewer devices or different locations depending on cost, communication infrastructure, and available switchgear; such changes would require a sensor-placement sensitivity study.

The assumed measurements correspond to transient-capable voltage and current channels with sufficient analog bandwidth to preserve the frequency range used by the time–frequency representation. In a field implementation, the effective signal available to the algorithm would be shaped by CT/VT or low-power sensor transfer functions, anti-aliasing filters, cabling, recorder resolution, synchronization error, and possible saturation. These effects are summarized by the measurement-chain model in Section 7.5. Therefore, the simulated sensor configuration should be interpreted as an ideal synchronized reference configuration rather than as a complete specification of field instrumentation.

For the reported simulations, the sampling frequency is fixed at $f_s = 200$ kHz, corresponding to the sampling interval

$$\Delta t_s = \frac{1}{f_s} = 5 \mu\text{s}. \quad (91)$$

This sampling rate provides a Nyquist frequency of

$$f_{\text{Nyq}} = \frac{f_s}{2} = 100 \text{ kHz}, \quad (92)$$

which exceeds the upper frequency of 50 kHz retained in the time–frequency representation defined in Section 5.2. Thus, the complete frequency band supplied to the learning models is represented without aliasing under the assumed ideal anti-aliasing condition.

The numerical results reported in this paper correspond exclusively to $f_s = 200$ kHz. A lower sampling rate would modify both the temporal resolution and the available spectral bandwidth. In particular, sampling rates of 10, 20, and 50 kHz would limit the Nyquist frequencies to 5, 10, and 25 kHz, respectively, thereby removing part of the 0–50 kHz band used by the present model. Consequently, direct decimation of the recorded signals would not preserve the input representation on which the reported networks were trained.

Evaluation at lower sampling rates would require an anti-aliasing filter, recomputation of the STFT parameters and retained frequency bins, and retraining of all learning-based configurations with matched lower-rate data. Such a sampling-rate sensitivity experiment is not included in the present evaluation. Therefore, the reported performance should not be extrapolated to recorders operating at 10–50 kHz. Although the Hybrid model uses the complete multi-sensor time–frequency tensor rather than a single arrival-time estimate, its localization performance remains dependent on the temporal resolution, measurement bandwidth, synchronization accuracy, and spectral content preserved by the acquisition system.

Topology Summary

Table 3 summarizes the case-study structure and sensing layout used for simulation and evaluation. The case study is formulated as a normalized simulation benchmark rather than as a utility-specific feeder model; therefore, voltages, currents, and loads are interpreted in per-unit quantities, while the cable behavior is governed by the nominal matrices and the aging-dependent perturbations defined in Section 3.7.

7.2. Area-Wise Localization Formulation

Let $\mathbf{X}$ denote the time–frequency tensor constructed from the synchronized voltage and current measurements. The proposed localization strategy follows a coarse-to-fine structure, as illustrated in Figure 6. This formulation is particularly useful in branched

feeders, where reflections from lateral junctions and terminations can produce ambiguous traveling-wave arrivals.

Table 3. Case-study topology and simulation configuration.

Item	Value/Description
Network type	Synthetic branched underground distribution feeder
Electrical representation	Cascaded π -section underground-cable model
Numerical tool	MATLAB 2025b transient simulation
Number of trunk nodes	13, from n_0 to n_{12}
Number of laterals	6, including one double-lateral junction at n_6
Feeder partitions	$A = 6$ non-overlapping areas
Synchronized sensors $\mathcal{M}$	$\{S_0, S_3, S_6, S_9, S_{12}\}$
Sensing configuration	Multi-ended, area-dependent transient monitoring
Signal channels	Three-phase voltages and currents at each sensor
Sampling rate	$f_s = 200$ kHz
Transient window	$T_w = 20$ ms
Operating-point variability	Load scaling $\gamma_L \in [0.80, 1.20]$
Cable parameters	Nominal $\mathbf{R}_0, \mathbf{L}_0, \mathbf{C}_0, \mathbf{G}_0$ perturbed by α as in Section 3.7

It is noted that areas A2 and A5 include three sensors due to their proximity to branching points, which increases waveform distortion and reflection complexity. The additional sensor is therefore used to improve observability in topologically complex regions. In contrast, two-sensor configurations are sufficient for linearly connected areas with reduced reflection ambiguity. The first area (A1) represents a boundary region where only downstream wave propagation is observable. This limitation is consistent with practical deployment scenarios and is intentionally preserved to evaluate model robustness under reduced observability conditions. Intermediate areas such as A3 may experience increased wavefront distortion due to reflections from adjacent branching regions. The overlapping sensor configuration is designed to mitigate this effect by capturing bidirectional transient information.

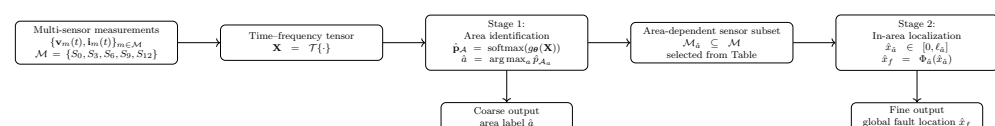


Figure 6. Coarse-to-fine area-wise localization strategy. The model first identifies the most probable feeder area and then estimates the in-area coordinate using the corresponding area-dependent sensor subset.

Stage 1: Area identification.

The complete multi-sensor time–frequency tensor is processed by the shared encoder and the area-identification head to obtain

$$\begin{aligned} \hat{\mathbf{p}}_A &= \text{softmax}(g_\theta(e_\theta(\mathbf{X}))) \in [0, 1]^A, \\ \sum_{a=1}^A \hat{p}_{A_a} &= 1. \end{aligned} \quad (93)$$

The predicted feeder area is

$$\hat{a} = \arg \max_{a \in \{1, \dots, A\}} \hat{p}_{A_a}. \quad (94)$$

Stage 2: Area-dependent sensor selection and local regression.

For each feeder area $\mathcal{A}_a$, the fixed channel-selection operator $\mathcal{P}_a$ retains the voltage and current channels associated with the sensor subset $\mathcal{M}_a$:

$$\mathbf{X}_a = \mathcal{P}_a(\mathbf{X}), \quad a \in \{1, \dots, A\}. \quad (95)$$

The selected tensor is processed by the shared-weight encoder and the corresponding local regression head:

$$\hat{u}_a = h_{a,\theta}(e_\theta(\mathbf{X}_a)) \in (0, 1). \quad (96)$$

The normalized output $\hat{u}_a$ defines the physical in-area coordinate

$$\hat{x}_a = \ell_a \hat{u}_a, \quad \hat{x}_a \in (0, \ell_a), \quad (97)$$

where ℓ_a is the ordered arc length assigned to area $\mathcal{A}_a$. The corresponding candidate global normalized location is

$$\hat{\lambda}_a = \Phi_a(\hat{u}_a), \quad \Phi_a : [0, 1] \rightarrow [0, 1], \quad (98)$$

where Φ_a incorporates the area boundary, cable-section ordering, and cumulative feeder arc length used to define the global location label.

During training, the ground-truth area activates the corresponding local regression head. During inference, the ground-truth area is unavailable and the final location is obtained exclusively from the predicted route:

$$\hat{\lambda} = \hat{\lambda}_{\hat{a}} = \Phi_{\hat{a}}(\hat{u}_{\hat{a}}). \quad (99)$$

No oracle correction is applied when the predicted area is incorrect. Therefore, an area-identification error selects an incorrect local regression head and local-to-global mapping, and its effect is included directly in the reported end-to-end localization error.

Sensor subsets per area.

Define $\mathcal{M}_a \subseteq \mathcal{M}$ as the sensor subset used for area $\mathcal{A}_a$. The area-wise sensing geometry is summarized in Table 4.

Table 4. Sensor subsets used for each area in the area-wise localization strategy.

Area	Sensor Subset $\mathcal{M}_a$
$\mathcal{A}_1$	$\{S_0, S_3\}$
$\mathcal{A}_2$	$\{S_0, S_3, S_6\}$
$\mathcal{A}_3$	$\{S_3, S_6\}$
$\mathcal{A}_4$	$\{S_6, S_9\}$
$\mathcal{A}_5$	$\{S_6, S_9, S_{12}\}$
$\mathcal{A}_6$	$\{S_9, S_{12}\}$

7.3. Simulation Scenarios and Dataset Protocol

Each simulated event is defined by the scenario vector

$$\boldsymbol{\theta} = (\mathcal{T}, e_f, x_f, R_f, t_f, \alpha, \gamma_L, \text{SNR}_{\text{dB}}), \quad (100)$$

where $\mathcal{T} \in \mathcal{S}_{\mathcal{T}}$ is the fault type, $e_f \in \mathcal{E}$ is the cable section where the fault is applied, $x_f \in (0, \ell_{e_f})$ is the local fault coordinate along that section, $R_f > 0$ is the fault resistance, t_f is the fault inception time, $\alpha \in [0, 1]$ is the normalized aging-stress index, γ_L is the load scaling factor, and SNR_{dB} sets the measurement-noise level.

7.3.1. Area-Balanced Scenario Sampling

To avoid bias toward any feeder area and to obtain statistically balanced coverage, sampling is performed in a stratified manner:

$$a \sim \mathcal{U}\{1, \dots, A\}, \quad (101)$$

$$e_f \sim \mathcal{U}(\mathcal{E}(\mathcal{A}_a)), \quad (102)$$

$$x_f \sim \mathcal{U}(\epsilon, \ell_{e_f} - \epsilon), \quad (103)$$

where $\mathcal{E}(\mathcal{A}_a)$ denotes the set of cable sections belonging to area $\mathcal{A}_a$, and $\epsilon > 0$ prevents degenerate placements exactly at junctions or terminations.

7.3.2. Fault Resistance, Aging, Operating Point, and Noise

Fault resistance is sampled as

$$R_f \sim \mathcal{U}(R_{\min}, R_{\max}), \quad (104)$$

with $(R_{\min}, R_{\max})$ given in Table 5. Aging is sampled as $\alpha \sim \mathcal{U}(0, 1)$, and the cable parameters are updated using the aging mapping in Section 3.7. Operating conditions are varied using $\gamma_L \sim \mathcal{U}(\gamma_{L,\min}, \gamma_{L,\max})$. Measurement noise is injected to achieve $\text{SNR}_{\text{dB}} \sim \mathcal{U}(\text{SNR}_{\min}, \text{SNR}_{\max})$ using the SNR definition in (36).

Table 5. Simulation scenario parameter ranges used for dataset generation.

Parameter	Minimum	Maximum	Unit
Fault resistance R_f	0.1	50	Ω
Fault location in section x_f / ℓ_{e_f}	0.05	0.95	–
Normalized aging-stress index α	0	1	–
Load scaling γ_L	0.80	1.20	–
SNR	20	40	dB

7.3.3. Scenario Ranges

The scenario ranges used in the case study are listed in Table 5. These values define the controlled simulation domain explored in the experiments. The fault-resistance interval includes low-resistance faults with strong transient excitation and higher-resistance faults with weaker signatures. The fault-location interval excludes placements exactly at section boundaries, junctions, and terminations to avoid degenerate labels and ambiguous boundary reflections. The SNR interval represents moderate-to-noisy measurement conditions for robustness assessment under additive channel noise. The load-scaling interval introduces prefault operating variability without changing the feeder topology. Finally, the normalized aging-stress interval $\alpha \in [0, 1]$ provides balanced coverage from nominal to maximally perturbed simulated cable parameters without implying that the endpoints correspond to a calibrated field lifetime.

7.3.4. Dataset Size, Stratification, and Splits

A large dataset is constructed to support evaluation across feeder areas, fault types, normalized aging-stress levels, fault resistances, operating points, and noise levels. Let N_{tot} denote the total number of simulated events and N_A the number of events per area:

$$N_A = \frac{N_{\text{tot}}}{A}. \quad (105)$$

In this case study, $N_{\text{tot}} = 36,000$ fault events are generated and stratified uniformly across $A = 6$ areas, yielding $N_A = 6000$ events per area. Within each area, scenarios are further stratified across the fault types in $\mathcal{S}_{\mathcal{T}}$ and across backbone and lateral placements using a 50%/50% allocation so that lateral events are not underrepresented.

The event records are assigned once to mutually disjoint training, validation, and test subsets:

$$\begin{aligned}\mathcal{D} &= \mathcal{D}_{\text{tr}} \cup \mathcal{D}_{\text{val}} \cup \mathcal{D}_{\text{te}}, \\ \mathcal{D}_{\text{tr}} \cap \mathcal{D}_{\text{val}} &= \mathcal{D}_{\text{tr}} \cap \mathcal{D}_{\text{te}} = \mathcal{D}_{\text{val}} \cap \mathcal{D}_{\text{te}} = \emptyset, \\ |\mathcal{D}_{\text{tr}}| : |\mathcal{D}_{\text{val}}| : |\mathcal{D}_{\text{te}}| &= 70 : 10 : 20.\end{aligned}\quad (106)$$

This record-level partition prevents the same simulated event from appearing in more than one subset. Stratification is applied jointly by feeder area and fault type so that the class and area proportions remain comparable across the three subsets.

All subsets are nevertheless generated using the same simulator, feeder topology, cable-parameter formulation, and numerical ranges for α , R_f , γ_L , SNR, and fault location. No tolerance-based grouping was applied to force neighboring points in the continuous scenario space into the same subset. Therefore, although exact event duplication across subsets is excluded, scenarios with similar continuous parameters may occur in different subsets.

Accordingly, the reported test results quantify interpolation performance on unseen event records drawn from the same controlled simulation domain. They do not constitute a grouped-scenario, out-of-range, independent-simulator, or field-domain generalization test. A stricter extrapolation assessment would require constructing groups from neighborhoods of the scenario vector

$$\boldsymbol{\theta} = (a, \mathcal{T}, e_f, x_f, R_f, \alpha, \gamma_L, \text{SNR}) \quad (107)$$

and assigning complete groups, parameter intervals, feeder sections, or simulator domains exclusively to the test set before model training. Such a grouped or out-of-domain evaluation is not reported in this paper and remains necessary before making claims about performance beyond the specified simulation domain.

The resulting dataset composition is summarized in Table 6.

Table 6. Dataset composition and interpretation of the large-case-study split.

Subset	Samples	Stratification	Role and Domain
Training $\mathcal{D}_{\text{tr}}$	25,200	Uniform by area and fault type	Model fitting using the complete prescribed ranges of aging, fault resistance, loading, noise, and fault location.
Validation $\mathcal{D}_{\text{val}}$	3600	Uniform by area and fault type	Hyperparameter selection and training monitoring within the same simulation domain.
Test $\mathcal{D}_{\text{te}}$	7200	Uniform by area and fault type	Held-out event-level evaluation within the same simulator and parameter ranges; not a grouped or out-of-domain test.

7.4. Simulation Workflow and Recorded Outputs

For each scenario $\boldsymbol{\theta}$, the aged cable parameters are computed, the feeder is discretized into cascaded π -sections, and a time-domain transient simulation is executed with a fault applied at t_f . The simulator records three-phase voltages and currents at each synchronized sensor $m \in \mathcal{M}$:

$$\{\mathbf{v}_m(t), \mathbf{i}_m(t)\}_{m \in \mathcal{M}}, \quad t \in \mathcal{T}_w, \quad (108)$$

where $\mathcal{T}_w = [t_{\text{start}}, t_{\text{start}} + T_w]$ is the transient window containing the fault inception and subsequent wave reflections. In the case study, $T_w = 20$ ms and the sampling rate is $f_s = 200$ kHz, yielding $N_w = T_w f_s = 4000$ samples per recorded channel. The fault inception time t_f is placed inside the window so that each record contains a pre-fault portion for baseline and noise-floor estimation, which is followed by a post-inception portion that captures the initial broadband transient, early reflections from branches and terminations, and the beginning of the fault-driven waveform evolution. Therefore, the window length is selected to support transient-onset processing and time–frequency feature extraction—not to represent a steady-state post-fault interval.

Each sample is stored with the labels

$$(y_{\text{cls}}, y_{\text{loc}}, y_A) = (\mathcal{T}, \lambda_f, a), \quad (109)$$

where y_{cls} is the fault-type label, $y_{\text{loc}} = \lambda_f$ is the global normalized location, and $y_A = a$ is the feeder-area label. Time–frequency tensors $\mathbf{X}$ are constructed from the recorded signals and passed to the hybrid physics-regularized model and the explainability module described in Section 5.

All transient simulations were implemented in MATLAB 2025b using the cascaded π -section representation described in Section 3.6. The preserved methodological specification includes the feeder topology, sensing locations, scenario ranges in Table 5, the 20 ms record length, the 200 kHz sampling rate, the time–frequency parameters defined in Section 5.2, the dataset partition, and the training protocol.

The archived simulation record does not retain the numerical entries of the nominal matrices $\mathbf{R}_0$, $\mathbf{L}_0$, $\mathbf{C}_0$, and $\mathbf{G}_0$, the realized values of κ_R , κ_L , κ_C , and κ_G , the section-specific values of $N_{\pi,e}$, the nominal nodal load vector, the numerical random seed, or the final validation-selected values of λ_{phy} and λ_{reg} . Consequently, the available specification supports the interpretation and internal comparison of the reported configurations but does not guarantee exact numerical regeneration of the original transient dataset or trained models. Reproduction on an independently implemented benchmark requires assigning and reporting these quantities explicitly before scenario generation and model training.

The complete scenario generation and evaluation pipeline is illustrated in Figure 7. The diagram summarizes the interaction between stratified sampling, aging-aware parameterization, transient simulation, dataset labeling, and physics-regularized explainable inference.

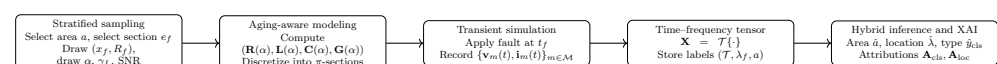


Figure 7. Large-case study workflow for aging-aware scenario generation, transient simulation, dataset labeling, and hybrid physics-regularized explainable inference.

7.5. Scope, Deployment Assumptions, and Domain-Shift Boundary

The case study is a controlled simulation-based validation study. It does not use field records, laboratory cable-fault measurements, or waveforms generated by an independent electromagnetic-transient simulation environment. Consequently, the numerical results reported in Section 8 should be interpreted as evidence of internal consistency, comparative behavior, and robustness within the explicitly defined simulation domain rather than as direct evidence of field deployment readiness.

This distinction is important because practical underground-cable fault diagnosis is affected not only by cable propagation physics but also by measurement-chain effects, sensor bandwidth, synchronization errors, transducer saturation, anti-aliasing filters, recorder resolution, and uncertainty in the actual cable parameters. To make this validation boundary explicit, let $s_m(t)$ denote an ideal simulated voltage or current channel at sensor location m ,

and let $\tilde{s}_m(t)$ denote the corresponding measured channel available to a practical diagnostic system. A generic measurement-chain representation can be written as

$$\tilde{s}_m(t) = (h_m * s_m)(t - \tau_m) + d_m(t) + n_m(t), \quad (110)$$

where $h_m(t)$ represents the bandwidth-limited impulse response of the sensor, transducer, cabling, anti-aliasing filter, and recorder; τ_m is the effective timing or synchronization error; $d_m(t)$ represents slow drift, offset, clipping, or saturation effects; $n_m(t)$ is additive measurement noise; and $*$ denotes convolution. In the baseline simulation study reported in this paper, the measurement chain is idealized as $h_m(t) = \delta(t)$, $\tau_m = 0$, and $d_m(t) = 0$, while $n_m(t)$ is scaled according to the prescribed SNR range. Therefore, the present results isolate the effects of fault type, fault location, feeder topology, fault resistance, load variability, measurement noise, and aging-induced cable-parameter drift, but they do not fully reproduce all imperfections of a field measurement system.

The same consideration applies to cable aging. The variable $\alpha \in [0, 1]$ is used as a normalized aging-stress index that induces controlled parameter drift in the per-unit-length matrices $\mathbf{R}(\alpha)$, $\mathbf{L}(\alpha)$, $\mathbf{C}(\alpha)$, and $\mathbf{G}(\alpha)$. It is not claimed to be a directly measured lifetime indicator or a field-calibrated health index. Thus, the aging model is suitable for robustness assessment under structured parameter drift, but additional calibration would be required before assigning a specific value of α to a real cable asset.

Table 7 summarizes the main validation assumptions and the corresponding requirements for transition toward practical deployment.

Table 7. Validation scope and practical deployment assumptions of this paper.

Aspect	Assumption in This Study	Requirement Before Field Deployment
Data source	Controlled simulation using the cascaded π -section underground-cable model.	Validation with independent electromagnetic-transient simulations and, where available, laboratory or field transient records.
Measurement chain	Ideal synchronized voltage/current channels with additive noise set by the prescribed SNR range.	Sensitivity analysis including sensor bandwidth, timing skew, filtering, saturation, recorder resolution, and channel-dependent transfer functions.
Cable aging	Normalized aging-stress index $\alpha \in [0, 1]$ used to induce monotonic drift in cable parameters.	Calibration of α or replacement by measured condition indicators, such as thermal history, dielectric loss, partial-discharge indicators, or asset-specific diagnostic data.
Cable parameters	Nominal matrices perturbed according to the aging parameterization.	Assessment under uncertain nominal parameters, spatially heterogeneous cable sections, and frequency-dependent dielectric behavior.
Performance claims	Comparative simulation-based performance under identical sampled scenarios for all methods.	Cross-domain validation to quantify degradation when the training and testing domains differ in simulator, sensor model, topology, operating conditions, or aging mechanism.

Accordingly, the objective of the present case study is to evaluate whether the proposed physics-regularized hybrid model provides a consistent advantage over the selected baselines under controlled, aging-aware, and topology-aware scenarios. The study is not intended to replace independent electromagnetic-transient validation or experimental verification. Instead, it establishes a reproducible simulation benchmark and identifies the additional validation layers required for deployment-oriented assessment.

8. Results and Discussion

This section evaluates the proposed framework using the stratified simulation dataset described in Section 4 and the large-scale underground feeder case study defined in Section 7. The analysis is organized around five objectives: (i) verifying that the simulated voltage and current signals contain physically meaningful transient signatures; (ii) illustrating whether the attribution module is aligned with physically meaningful transient time–frequency regions; (iii) quantifying multi-task performance in fault-type classification, feeder-area identification, and continuous fault localization; (iv) evaluating robustness under measurement noise, aging-induced parameter drift, fault-resistance variation, and feeder-topology effects; and (v) assessing confidence reliability and statistical uncertainty.

All quantitative results are obtained from simulation-based data and must be interpreted under the validation boundary defined in Section 7.5. The purpose of the evaluation is to compare diagnostic models under identical controlled scenarios—not to claim direct field deployment readiness. The compared methods follow the hierarchy defined in Table 2: TW-TOA provides a classical traveling-wave reference, Baseline isolates purely supervised time–frequency learning, PINN isolates the effect of physics-consistency regularization, and Hybrid evaluates the complete proposed formulation. Since all methods are evaluated on the same held-out scenarios, the reported metrics quantify relative performance within the specified synthetic domain. Cross-domain performance under independent simulators, non-ideal measurement chains, and field-calibrated aging conditions remains an important validation step.

8.1. Evaluation Protocol and Metrics

All quantitative metrics are computed on a held-out test set stratified by feeder area and fault type. Three prediction tasks are evaluated: fault-type classification over 11 classes, area identification over 6 feeder partitions, and continuous fault localization expressed by the normalized per-unit distance $\lambda \in [0, 1]$.

For fault-type classification and area identification, accuracy is defined as

$$\text{Acc} = \frac{1}{N} \sum_{n=1}^N \mathbb{I}\{\hat{y}_n = y_n\}, \quad (111)$$

where y_n and $\hat{y}_n$ are the true and predicted labels, respectively, and $\mathbb{I}\{\cdot\}$ denotes the indicator function.

In addition to accuracy, the class-level precision, recall, and F1-score are computed for the fault-type classification task. For class $c \in \mathcal{S}_{\mathcal{T}}$, let TP_c , FP_c , and FN_c denote the true-positive, false-positive, and false-negative counts. The corresponding metrics are

$$\text{Precision}_c = \frac{\text{TP}_c}{\text{TP}_c + \text{FP}_c + \epsilon'}, \quad (112)$$

$$\text{Recall}_c = \frac{\text{TP}_c}{\text{TP}_c + \text{FN}_c + \epsilon'}, \quad (113)$$

$$\text{F1}_c = \frac{2 \text{Precision}_c \text{Recall}_c}{\text{Precision}_c + \text{Recall}_c + \epsilon'}, \quad (114)$$

where $\epsilon > 0$ prevents division by zero. The macro-averaged F1-score is defined as

$$\text{F1}_{\text{macro}} = \frac{1}{|\mathcal{S}_{\mathcal{T}}|} \sum_{c \in \mathcal{S}_{\mathcal{T}}} \text{F1}_c. \quad (115)$$

These metrics complement aggregate accuracy by verifying whether the classifier maintains balanced behavior across the eleven fault categories.

For fault localization, the mean absolute error is defined as

$$\text{MAE} = \frac{1}{N} \sum_{n=1}^N |\hat{\lambda}_n - \lambda_n|, \quad (116)$$

where λ_n and $\hat{\lambda}_n$ are the true and estimated normalized fault locations, respectively. In addition, the 95th percentile of $|\hat{\lambda} - \lambda|$ is reported to characterize tail error, which is relevant for protection and restoration decisions.

Confidence quality is evaluated through reliability-style calibration curves and expected calibration error (ECE), while selective prediction curves quantify how classification risk and localization error change when low-confidence samples are rejected.

Binning rules and sample support.

All binned analyses are performed on the fixed test set of $N = 7200$ events without resampling, post hoc balancing, or replacement. Each event is assigned to exactly one interval for each analyzed variable, and the metric associated with a bin is calculated using all events contained in that interval.

For classification calibration, the Baseline, PINN, and Hybrid confidence scores are partitioned into $B_{\text{cal}} = 10$ equal-width intervals over $[0, 1]$:

$$\mathcal{I}_b^{\text{cal}} = \left\{ n : c_n \in \left[\frac{b-1}{10}, \frac{b}{10} \right) \right\}, \quad b = 1, \dots, 9, \quad (117)$$

and

$$\mathcal{I}_{10}^{\text{cal}} = \{ n : c_n \in [0.9, 1] \}. \quad (118)$$

The sample count of bin b is

$$N_b^{\text{cal}} = |\mathcal{I}_b^{\text{cal}}|, \quad \sum_{b=1}^{10} N_b^{\text{cal}} = 7200. \quad (119)$$

All ten confidence intervals contain test events. Empirical accuracy, mean confidence, and the ECE contribution are computed from the corresponding N_b^{cal} samples. The deterministic TW-TOA method is excluded from this probabilistic binning because it does not produce native class probabilities.

The one-dimensional robustness analyses use five intervals for each continuous stress variable. The SNR intervals are

$$[20, 24), \quad [24, 28), \quad [28, 32), \quad [32, 36), \quad [36, 40] \text{ dB}, \quad (120)$$

and the normalized aging-stress intervals are

$$[0, 0.2), \quad [0.2, 0.4), \quad [0.4, 0.6), \quad [0.6, 0.8), \quad [0.8, 1]. \quad (121)$$

Fault resistance is grouped uniformly in logarithmic space. Defining

$$r_f = \log_{10}(R_f), \quad r_f \in [-1, \log_{10}(50)], \quad (122)$$

the six bin edges are

$$r_j = -1 + \frac{j}{5} [\log_{10}(50) + 1], \quad j = 0, \dots, 5. \quad (123)$$

Thus, fault-resistance bin j contains the events satisfying

$$r_f \in [r_{j-1}, r_j), \quad j = 1, \dots, 4, \quad (124)$$

whereas the fifth interval includes its upper boundary.

For a generic robustness variable z , the number of samples assigned to interval b is

$$N_b^{(z)} = \left| \left\{ n : z_n \in \mathcal{I}_b^{(z)} \right\} \right|, \quad \sum_{b=1}^5 N_b^{(z)} = 7200. \quad (125)$$

Fault-type accuracy, area-identification accuracy, and localization MAE are calculated independently within each bin while marginalizing over all remaining scenario variables.

The two-dimensional sensitivity maps use the Cartesian products of the same five aging intervals with the five SNR or logarithmic fault-resistance intervals. Therefore, each map contains $5 \times 5 = 25$ cells. The exact number of test events in each cell is displayed inside the corresponding cell of the two-dimensional sensitivity maps, and the displayed counts sum to 7200 for each two-dimensional partition.

Table 8 summarizes the evaluation protocol used in the Results section. The numerical scenario ranges are not repeated here; they are defined in Table 5 to avoid duplicating the case-study specification.

Table 8. Evaluation protocol used for the reported test-set analysis.

Item	Value/Description
Test-set size	$N = 7200$ held-out simulated events
Stratification	By feeder area and fault type
Evaluated configurations	TW-TOA, Baseline, PINN, and Hybrid
Discrete tasks	Fault-type classification and feeder-area identification
Continuous task	Normalized fault-location estimation $\lambda \in [0, 1]$
Classification metrics	Accuracy, precision, recall, and F1-score
Localization metrics	MAE, P95, empirical coverage, and bootstrap uncertainty
Robustness variables	SNR, normalized aging-stress index α , fault resistance R_f , and topology subset
Interpretability output	Time–frequency attribution illustration and model-native attribution-validation protocol

The area labels are defined by the mutually exclusive feeder partition and sensor layout shown in Figure 5. For the Baseline, PINN, and Hybrid configurations, the predicted area is obtained from the corresponding area-probability vector:

$$\hat{a}_n = \arg \max_{a \in \{1, \dots, A\}} \hat{p}_{\mathcal{A}_a}^{(n)}. \quad (126)$$

For TW-TOA, no separate area classifier is used; the predicted label is $\hat{a}_{\text{TOA},n}$, which is obtained jointly with the continuous location estimate from (48). Area-identification accuracy is then calculated for every method using (111) with $y_n = a_n$ and $\hat{y}_n = \hat{a}_n$.

The area partition follows the same deterministic assignment rule used during dataset generation, so every feeder section and ordered arc-length coordinate belongs to exactly one area. Faults are sampled away from junctions, terminations, and area boundaries; consequently, the ground-truth area labels are unambiguous. Any numerical estimate falling exactly on a shared boundary is assigned using the same fixed boundary convention used to construct the dataset labels.

To verify that the reported metrics are not dominated by sampling imbalance, Figure 8 summarizes the generated dataset marginals, including fault resistance, normalized aging-stress index, SNR, fault location, area distribution, and trunk/lateral topology.

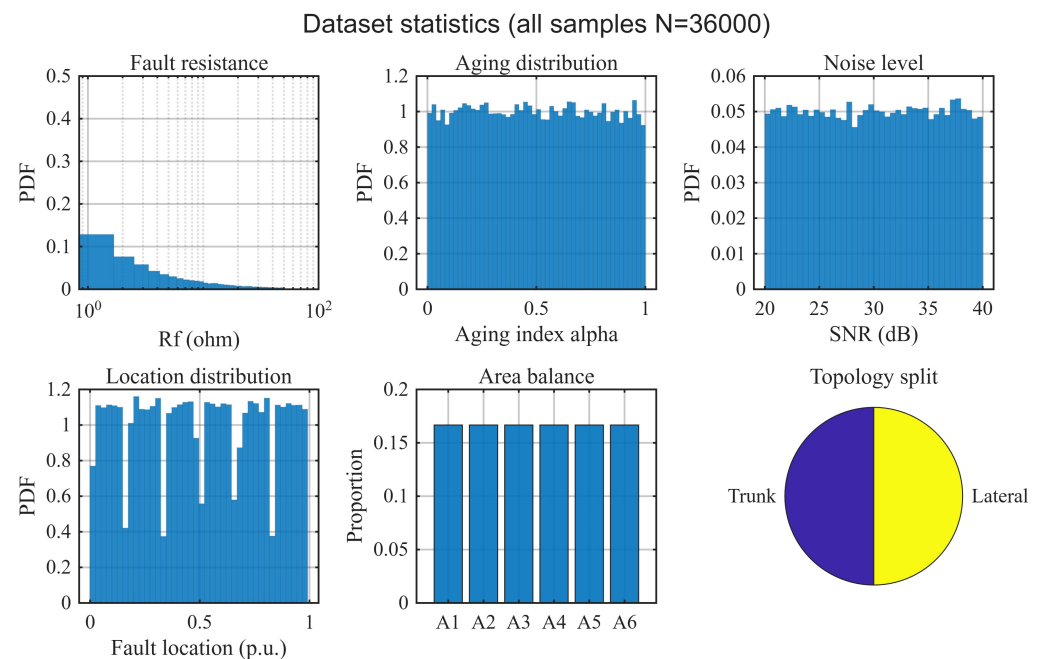


Figure 8. Dataset statistics for all generated samples: distributions of R_f , normalized aging-stress index α , SNR, fault location λ , area balance, and trunk/lateral topology split. These distributions verify broad coverage over the scenario variables and reduce the risk of biased aggregate metrics.

8.2. Signal-Level Validation and Interpretability of Synthetic Fault Signatures

A necessary condition for scientifically meaningful simulation-based evaluation is that the generated signals exhibit structured and physically consistent transient behavior. In underground cables, fault events produce fast electromagnetic disturbances whose informative content is concentrated in the high-frequency (HF) range. Therefore, the analysis focuses on HF-envelope representations and time–frequency maps to validate that the simulated data preserve traveling-wave-like characteristics relevant for fault diagnosis and localization.

8.2.1. Voltage HF-Envelope Across Fault Types

Figure 9 presents the HF-envelope of three-phase voltages measured at a fixed synchronized sensor for all 11 fault types under a representative operating scenario. Each row corresponds to a distinct fault class, enabling a direct comparison of transient amplitude and phase involvement. The figure shows that faults involving specific phases systematically produce stronger HF activity in those phases, indicating that the envelope representation preserves the discriminative information required for multi-class classification.

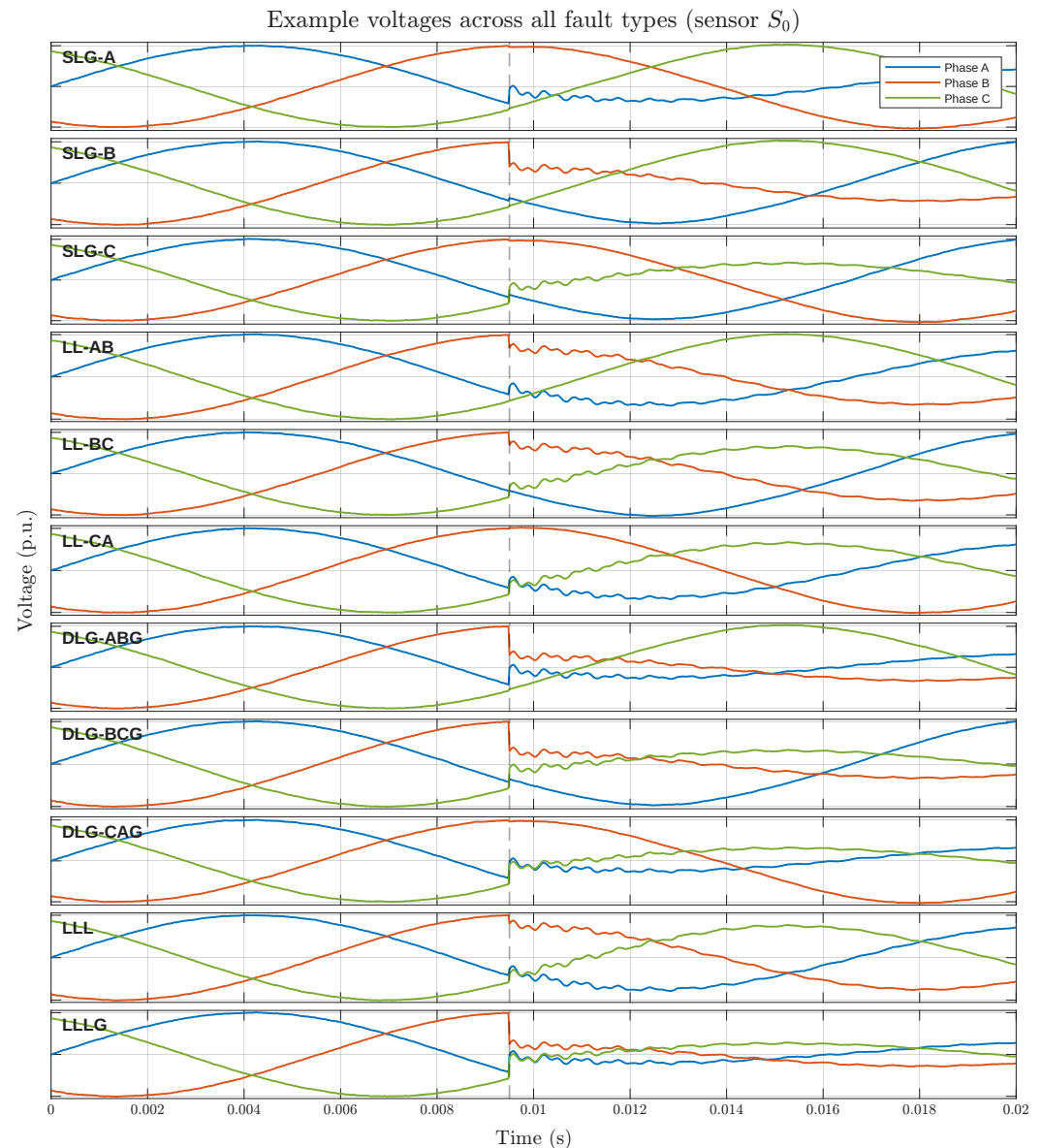


Figure 9. HF-envelope voltage waveforms across all fault types for a representative scenario at a fixed sensor. The row-wise layout highlights phase-dependent transient energy around fault inception, which is consistent with traveling-wave-based interpretation.

8.2.2. Current HF-Envelope Across Fault Types

Figure 10 shows the corresponding HF-envelope representation for current signals. Compared to voltages, currents exhibit sharper discontinuities and stronger sensitivity to ground-return paths, which is particularly evident in ground-involved faults. This complementary behavior supports the use of multi-channel voltage and current measurements to enhance both classification and localization performance.

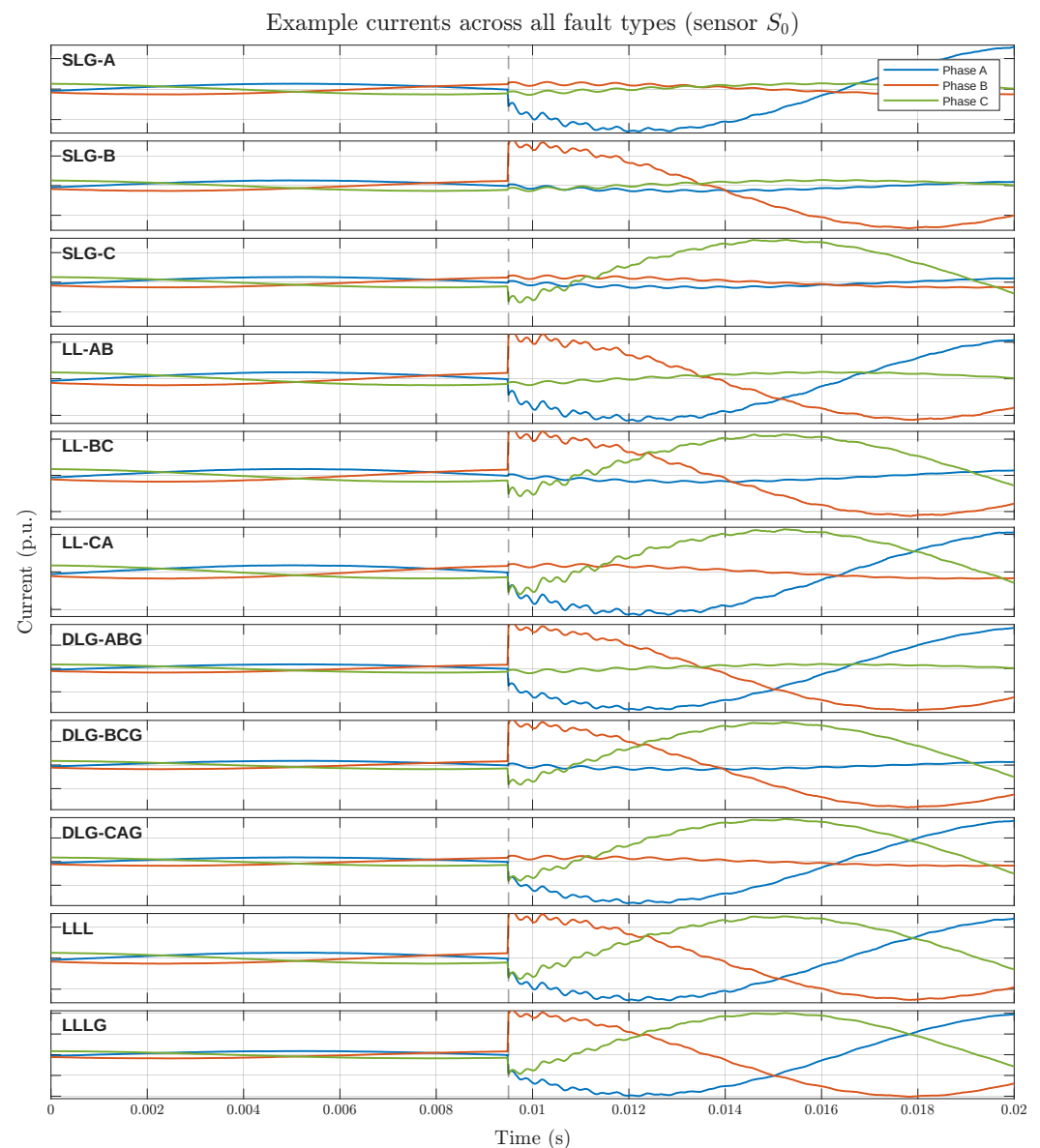


Figure 10. HF-envelope current waveforms across all fault types for the same scenario as Figure 9. Current signals provide complementary transient information that enhances discrimination among fault categories and supports localization.

8.2.3. Raw Signals Versus HF-Envelope

Although HF-envelope representations improve interpretability, the model ultimately operates on realistic time-domain signals or derived features. Figure 11 compares raw three-phase voltages and currents with their corresponding HF-envelope views. The raw signals exhibit relatively subtle disturbances under moderate SNR, whereas the envelope representation clearly highlights the transient onset and phase involvement. This comparison justifies the use of HF-enhanced representations for both visualization and feature extraction.

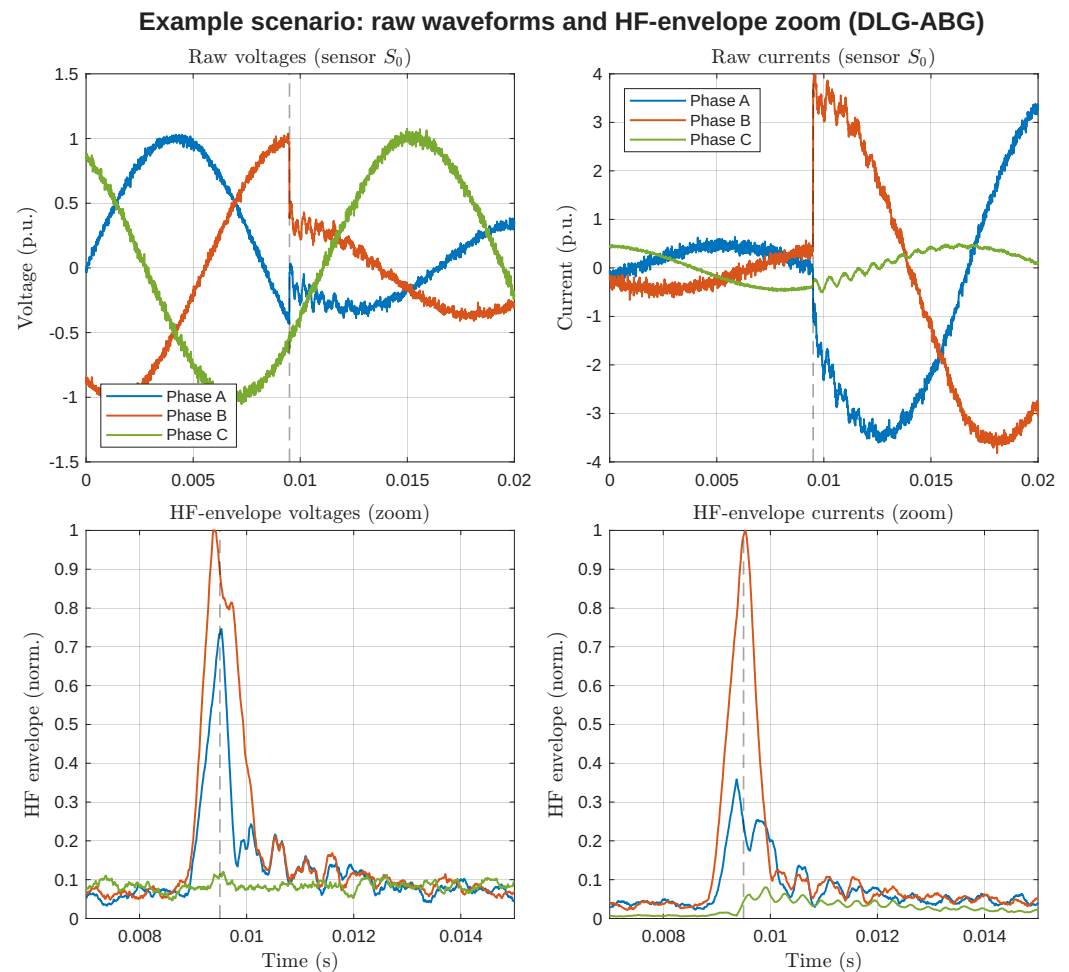


Figure 11. Comparison between raw three-phase signals and HF-envelope representations for an illustrative fault scenario. The envelope emphasizes transient energy and improves the visibility of fault inception and phase involvement. The dashed vertical line indicates the simulator-defined fault-inception instant.

8.2.4. Time–Frequency Evidence of Transient Behavior

To further validate the physical consistency of the simulated signals, Figure 12 presents time–frequency maps computed from a high-frequency residual of the measured voltage at sensor S_0 . The dashed vertical line indicates the known fault inception instant used by the simulator to generate the scenario; it is shown only as a visual reference and is not provided to the learning model as an input. In practical records, this instant would need to be estimated from the measured transient onset—for example, using the TW-TOA detector defined in Section 5.1. Across all fault types, the onset is characterized by a localized broadband energy burst spanning several kilohertz, which is followed by a decay pattern that varies with frequency.

This behavior is consistent with dispersive and lossy propagation in underground cables, where higher-frequency components attenuate more rapidly. Importantly, the energy is temporally localized rather than stationary, confirming that the signals reflect transient events rather than steady-state spectral artifacts.

Differences among fault categories are also observable. Ground-involved faults, such as SLG and DLG, exhibit stronger broadband responses due to zero-sequence excitation and ground-return currents, whereas line-to-line and three-phase faults produce more symmetric spectral patterns. The LLLG condition combines strong broadband excitation with ground coupling. These observations confirm that the simulated dataset preserves physically meaningful transient signatures required for reliable learning and inference.

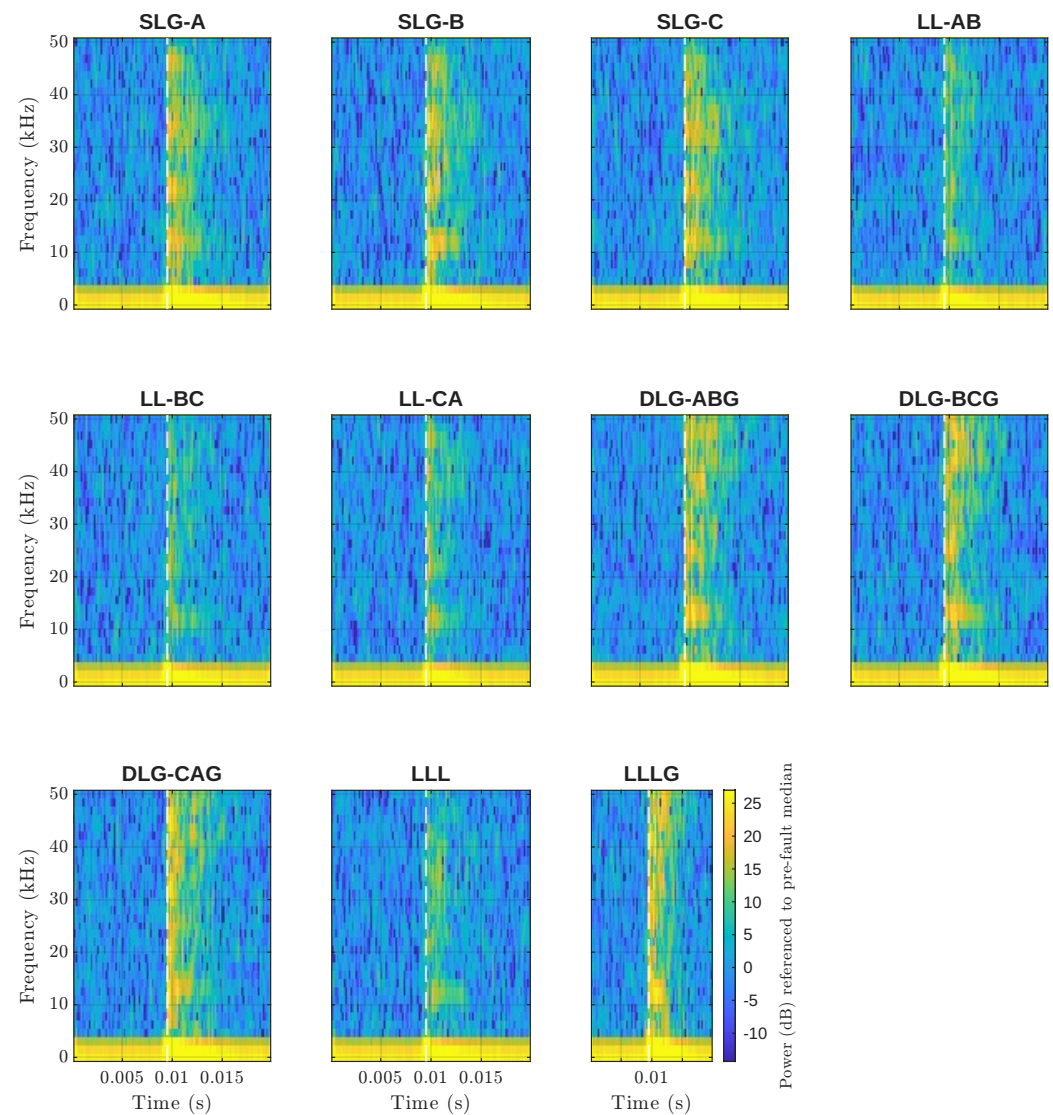


Figure 12. Time–frequency maps for all fault types using a high-frequency voltage residual at sensor S_0 . Each subplot shows a localized broadband energy increase around fault inception, indicated by the dashed line, confirming transient behavior consistent with traveling-wave propagation rather than stationary noise.

8.2.5. Fault-Inception and Wavefront-Arrival Timing

The dashed vertical reference used in Figures 9–12 denotes the simulator-defined fault-inception instant t_f , which is equivalent to t_0 in the propagation-residual formulation. This source event time must be distinguished from the first wavefront-arrival time at sensor m , because the latter includes the propagation delay between the fault position and the measurement location.

For event n , the reference first-arrival time at sensor m is expressed as

$$\tau_{m,n}^{\text{ref}} = t_{f,n} + \frac{d_m(x_{f,n})}{\hat{v}(\alpha_n)}, \quad (127)$$

where $x_{f,n}$ is the simulated fault position, $d_m(x_{f,n})$ is the corresponding feeder-path distance, and $\hat{v}(\alpha_n)$ is the aging-conditioned propagation velocity. The onset detector in (47) produces the measured estimate $\hat{\tau}_{m,n}$.

Once the TW-TOA position $\hat{x}_{\text{TOA},n}$ has been obtained, an operational estimate of the fault-inception time can be reconstructed from the detected sensor arrivals as

$$\hat{t}_{0,n}^{\text{TOA}} = \frac{\sum_{m \in \mathcal{M}_{\text{valid}}^{(n)}} w_m \left[\hat{\tau}_{m,n} - \frac{d_m(\hat{x}_{\text{TOA},n})}{\hat{v}(\alpha_n)} \right]}{\sum_{m \in \mathcal{M}_{\text{valid}}^{(n)}} w_m}. \quad (128)$$

The corresponding inception-time error is

$$e_{t_{0,n}}^{\text{TOA}} = \hat{t}_{0,n}^{\text{TOA}} - t_{f,n}. \quad (129)$$

Wavefront-detection accuracy is more directly characterized at the sensor level through

$$e_{\tau,m,n} = \hat{\tau}_{m,n} - \tau_{m,n}^{\text{ref}}, \quad (130)$$

with aggregate measures

$$\text{MAE}_{\tau} = \frac{1}{N_{\tau}} \sum_n \sum_{m \in \mathcal{M}_{\text{valid}}^{(n)}} |e_{\tau,m,n}|, \quad \text{P95}_{\tau} = Q_{0.95}(\{|e_{\tau,m,n}|\}), \quad (131)$$

where N_{τ} is the total number of valid sensor–event arrival estimates.

The Hybrid inference architecture predicts the fault type, feeder area, and fault location; it does not contain a fault-inception-time or wavefront-arrival-time output head. Therefore, $\hat{t}_0$ is not a Hybrid model prediction. The detected arrivals are used in the TW-TOA reference and in the propagation-consistency term during training, whereas online Hybrid inference operates directly on the time–frequency tensor.

The retained evaluation outputs contain the diagnostic predictions and aggregate performance metrics but not the complete set of per-event arrival-time estimates $\hat{\tau}_{m,n}$ or reconstructed inception times $\hat{t}_{0,n}^{\text{TOA}}$. Consequently, Figures 9–12 provide a signal-level visualization of the known fault inception and the associated transient response rather than a quantitative evaluation of wavefront-detection timing.

8.2.6. Qualitative Time–Frequency Saliency Illustration

Figure 13 compares the time–frequency representation of a representative DLG-ABG fault at sensor S_0 with a proxy saliency overlay constructed on the same time–frequency grid. The input map is computed from the high-frequency residual of the measured signal and referenced to the pre-fault condition. The proxy overlay emphasizes the transient broadband region surrounding the known simulated fault-inception instant.

The concentration of the overlay around fault inception is physically compatible with the expected high-frequency excitation of a fault-initiated traveling wave. However, the displayed saliency pattern is not computed from the gradients, activations, or feature contributions of the trained diagnostic network. It therefore provides only a qualitative visualization of the signal region that a physically plausible model explanation would be expected to emphasize.

Model-specific interpretability requires attribution maps computed directly from the trained fault-type and localization outputs, such as the Grad-CAM- and SHAP-style formulations defined in Section 5.5. The quantitative diagnostics introduced in Section 5.5.5 are consequently not applied to Figure 13, because transient concentration, attribution stability, compactness, and occlusion-based faithfulness cannot be established from a proxy overlay.

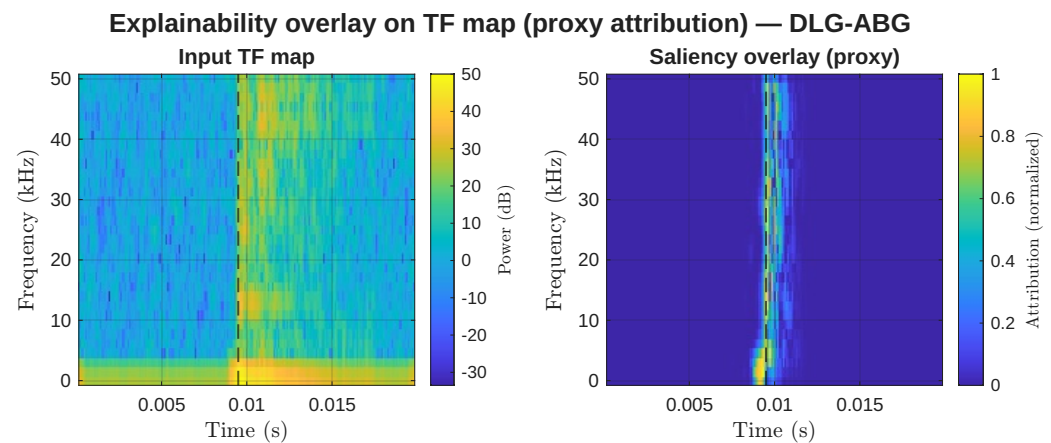


Figure 13. Qualitative time–frequency saliency illustration for a representative DLG-ABG fault at sensor S_0 . The left panel shows the input time–frequency representation, whereas the right panel shows a proxy saliency overlay constructed to emphasize the broadband transient region around the known simulated fault-inception instant. The proxy overlay is not a model-native Grad-CAM or SHAP attribution and is not used as quantitative evidence of explanation faithfulness.

8.3. Main Quantitative Performance on the Stratified Test Set

We next quantify the multi-task performance of the four predictors under identical test-set conditions. The central objective is to determine whether the proposed physics-regularized and hybrid learning strategies provide systematic and consistent improvements across the three coupled tasks considered in this work: (i) discrete fault-type classification, (ii) discrete feeder-area identification, and (iii) continuous fault localization. Particular attention is given to the localization tail error, since large deviations are especially relevant in protection, maintenance, and service-restoration applications.

8.3.1. Primary Quantitative Summary

Table 9 reports the principal test-set metrics for all models. These values provide the numerical basis for the comparative analysis that follows. For the localization task, both the mean absolute error (MAE) and the 95th percentile (P95) of the absolute error are reported. While MAE describes average estimation accuracy, P95 characterizes the upper tail of the error distribution and therefore captures high-error cases that may dominate operational risk.

Table 9. Test-set performance summary across fault-type classification, area identification, and localization. Localization errors are expressed in per-unit distance.

Model	Fault Acc.	Area Acc.	MAE	P95
TW-TOA	0.76	0.86	0.027	0.069
Baseline	0.83	0.90	0.021	0.054
PINN	0.89	0.93	0.014	0.036
Hybrid	0.93	0.96	0.011	0.027

8.3.2. Computational Scope and Inference Latency

Table 9 reports diagnostic performance only. The retained experimental outputs do not include execution-time traces, processor or accelerator specifications, memory usage, model-parameter counts, or floating-point operation counts. Consequently, a numerical inference time cannot be reported without performing a new controlled benchmark, and the present results do not establish real-time or near-real-time execution capability.

The online Hybrid inference path comprises per-unit signal scaling, construction of the $30 \times 65 \times 59$ time–frequency tensor, evaluation of the shared encoder and the fault-type and area-identification heads, selection of the area-dependent sensor channels, and evaluation of the local regression head associated with the predicted area. Cable-transient simulation, model training, backpropagation of the physics-consistency loss, bootstrap analysis, and attribution-map generation are offline operations and are not required for each online diagnostic prediction.

A reproducible latency assessment should separately measure signal preprocessing and neural-network execution using a batch size of one after an initial warm-up period. For N_{run} repeated evaluations, the per-event end-to-end latency can be defined as

$$t_{\text{inf}}^{(r)} = t_{\text{output}}^{(r)} - t_{\text{input}}^{(r)}, \quad r = 1, \dots, N_{\text{run}}, \quad (132)$$

and summarized through the median and upper-tail latency:

$$\tilde{t}_{\text{inf}} = Q_{0.50} \left(\left\{ t_{\text{inf}}^{(r)} \right\}_{r=1}^{N_{\text{run}}} \right), \quad t_{\text{inf,P95}} = Q_{0.95} \left(\left\{ t_{\text{inf}}^{(r)} \right\}_{r=1}^{N_{\text{run}}} \right). \quad (133)$$

Such measurements must be reported together with the processor or accelerator model, available memory, numerical precision, software environment, thread configuration, and the inclusion or exclusion of signal acquisition and STFT computation. Without these implementation-dependent quantities, deployment feasibility cannot be inferred from the diagnostic accuracy metrics alone.

8.3.3. Physical Scale of the Localization Error

The global fault coordinate is expressed as a normalized arc-length position with respect to a reference feeder length ℓ_{ref} . Accordingly, a normalized absolute error e_{pu} corresponds to the physical distance

$$e_{\text{phys}} = e_{\text{pu}} \ell_{\text{ref}}. \quad (134)$$

The localization results obtained by the Hybrid model can therefore be written as

$$\text{MAE}_{\text{phys}} = 0.011 \ell_{\text{ref}}, \quad e_{\text{P95,phys}} = 0.027 \ell_{\text{ref}}. \quad (135)$$

When ℓ_{ref} is expressed in meters or kilometers, the corresponding localization errors are obtained directly in the same unit. In the normalized feeder adopted in this paper, the cable sections are represented through relative arc lengths; therefore, the numerical values in meters or kilometers depend on the physical section lengths assigned to a particular network implementation.

The sampling frequency of 200 kHz corresponds to a sampling interval of

$$\Delta t_s = \frac{1}{f_s} = 5 \mu\text{s}. \quad (136)$$

For an effective propagation velocity $\hat{v}(\alpha)$, the associated one-way propagation distance over one sampling interval is

$$\Delta d_s = \hat{v}(\alpha) \Delta t_s, \quad \Delta \lambda_s = \frac{\hat{v}(\alpha) \Delta t_s}{\ell_{\text{ref}}}. \quad (137)$$

For a differential-arrival formulation, the corresponding first-order spatial increment is approximately

$$\Delta d_{\text{diff}} \approx \frac{\hat{v}(\alpha) \Delta t_s}{2}. \quad (138)$$

These quantities provide a sampling-based reference for interpreting the reported errors. The TW-TOA implementation additionally uses sub-sample interpolation and multiple sensor arrivals, whereas the learning-based models infer location from the complete time–frequency tensor. Their effective resolution is therefore influenced jointly by sampling frequency, measurement bandwidth, synchronization, propagation-velocity uncertainty, and the information contained in the multi-channel transient representation.

Table 9 shows progressively improved performance across the four evaluated configurations. The Hybrid model achieves the highest accuracy in both discrete tasks, reaching 0.93 for fault-type classification and 0.96 for area identification. For localization, MAE decreases from 0.027 for TW-TOA to 0.011 for the Hybrid configuration, while P95 decreases from 0.069 to 0.027. The reduction in P95 indicates that the complete Hybrid formulation improves both the average localization accuracy and the upper tail of the absolute-error distribution.

The Baseline–PINN comparison isolates the effect of adding the propagation-consistency loss to the matched global supervised architecture. Relative to the Baseline, the PINN increases fault-type accuracy from 0.83 to 0.89, increases area-identification accuracy from 0.90 to 0.93, reduces MAE from 0.021 to 0.014, and reduces P95 from 0.054 to 0.036. Within the controlled test domain, these differences are consistent with a beneficial effect of the output-dependent physics regularization under otherwise matched task definitions and input representations.

8.3.4. Relative Gains of the Hybrid Model

Table 10 reports the relative gains of the Hybrid model with respect to each reference configuration. Accuracy improvements are expressed in percentage points, whereas localization improvements are expressed as relative reductions in MAE and P95.

Table 10. Relative gains of the Hybrid model with respect to each reference configuration. “pp” denotes percentage points.

Hybrid vs. Reference	Δ Fault (pp)	Δ Area (pp)	MAE Red. (%)	P95 Red. (%)
TW-TOA	+17.0	+10.0	59.3	60.9
Baseline	+10.0	+6.0	47.6	50.0
PINN	+4.0	+3.0	21.4	25.0

Relative to TW-TOA, the Hybrid configuration increases fault-type accuracy by 17.0 percentage points and area-identification accuracy by 10.0 percentage points while reducing MAE by 59.3% and P95 by 60.9%. Relative to the purely data-driven Baseline, it increases fault-type and area-identification accuracies by 10.0 and 6.0 percentage points, respectively, and reduces MAE and P95 by 47.6% and 50.0%.

The additional differences between PINN and Hybrid are 4.0 percentage points in fault-type accuracy, 3.0 percentage points in area-identification accuracy, a 21.4% reduction in MAE, and a 25.0% reduction in P95. These differences quantify the advantage of the complete Hybrid configuration over the global PINN configuration. Because the Hybrid simultaneously introduces area-conditioned routing, area-dependent sensor selection, and local regression heads, the PINN–Hybrid comparison does not identify the marginal contribution of any individual component.

8.3.5. Graphical Summary of Overall Performance

For a compact visual comparison, Figure 14 summarizes the main quantitative metrics across all models, including fault-type accuracy, area-identification accuracy, localization MAE, and localization P95.

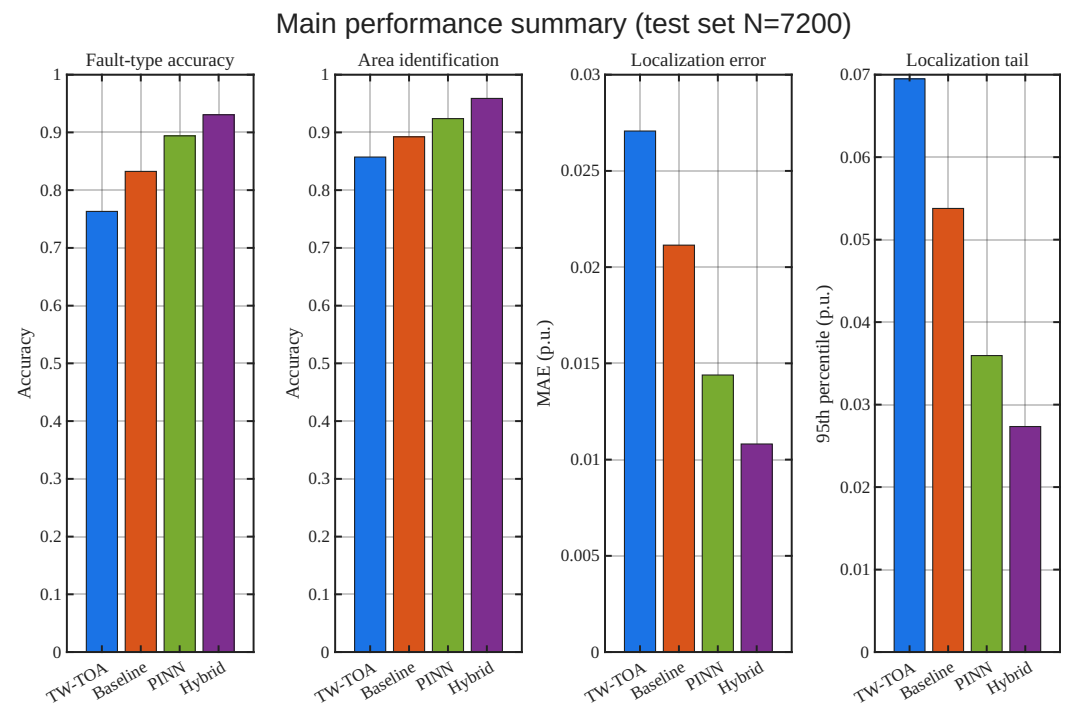


Figure 14. Main performance summary on the test set: fault-type accuracy, area-identification accuracy, localization MAE, and localization P95 across TW-TOA, Baseline, PINN, and Hybrid models. The Hybrid model consistently achieves the best performance across all metrics, including reduced tail error.

As shown in Figure 14, a consistent ordering is observed across the four evaluation criteria. TW-TOA provides a physically interpretable reference, while the Baseline and PINN models progressively improve performance. The Hybrid model achieves the highest classification and area-identification accuracies while simultaneously minimizing both MAE and P95. This confirms that the improvement is not restricted to a single metric; instead, it reflects a balanced enhancement across discrete and continuous tasks. Therefore, the Hybrid configuration provides a superior overall operating point by improving classification reliability, area identification, and localization precision at the same time.

8.4. Confusion Matrices and Fine-Grained Diagnostic Analysis

Aggregate metrics summarize global performance but do not reveal whether the remaining errors are physically plausible or randomly distributed across unrelated classes. Therefore, a class-level and area-level diagnostic analysis is required. For fault-type classification, the relevant question is whether misclassifications occur mainly between fault categories with similar phase involvement or grounding conditions. For area identification, errors are expected to occur primarily between adjacent feeder partitions, because faults close to area boundaries may generate similar multi-sensor transient patterns.

Figure 15 presents the diagnostic suite of the Hybrid model on the test set with $N = 7200$ samples. The figure includes row-normalized confusion matrices for fault-type classification and area identification together with per-class precision, recall, F1-score, and area-wise accuracy. The class-level metrics are computed according to (112)–(114), so the diagnostic evaluation is not limited to aggregate accuracy.

The fault-type confusion matrix in Figure 15 exhibits strong diagonal dominance, indicating that correct classifications dominate across the eleven fault classes. The remaining off-diagonal entries are sparse and mainly concentrated among fault types with related electrical behavior, such as faults sharing a similar phase involvement or ground-return components. This pattern suggests that the errors are consistent with intrinsic waveform similarity rather than arbitrary class confusion.

The area-identification confusion matrix also shows diagonal dominance. When area misclassifications occur, they are mostly located near the main diagonal, meaning that incorrect predictions tend to involve neighboring areas. This behavior is physically reasonable because faults located close to area boundaries can produce similar arrival patterns at adjacent sensor subsets. The area-wise accuracy bars further indicate that no feeder partition dominates the error distribution, supporting balanced performance across the six areas.

Overall, Figure 15 shows that the Hybrid model errors are structured, spatially coherent, and physically interpretable. This strengthens the validity of the aggregate results reported in Tables 9 and 10, because high global accuracy is accompanied by plausible error patterns rather than by uncontrolled misclassification behavior.

Hybrid model: confusion matrices and detailed metrics (test set N=7200)

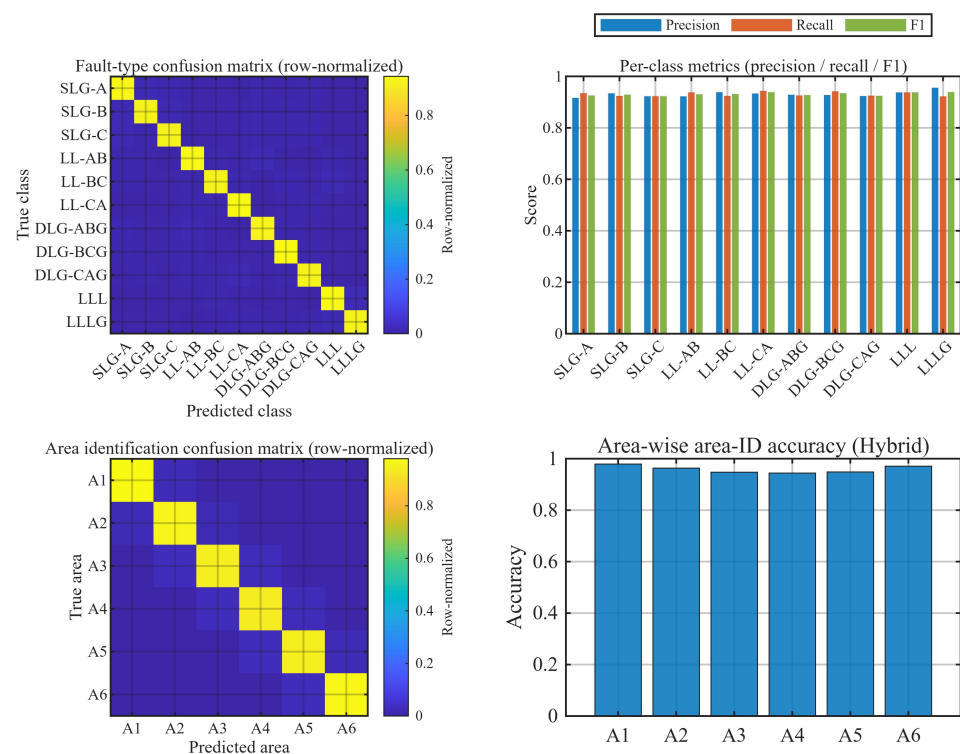


Figure 15. Hybrid diagnostic suite on the test set ($N = 7200$): row-normalized fault-type confusion matrix with per-class precision, recall, and F1-score metrics, and row-normalized area-identification confusion matrix with area-wise accuracy. The dominant diagonal structure indicates accurate predictions, while the sparse off-diagonal entries are mainly associated with physically related fault types or neighboring feeder areas.

8.5. Localization Regression Behavior: Bias, Dispersion, and Tail Risk

A rigorous evaluation of fault-location performance requires more than scalar indicators such as MAE and P95. Although these metrics summarize average and tail errors, they do not reveal whether the estimator exhibits systematic bias, position-dependent deviations, asymmetric errors, or aging-dependent dispersion. Therefore, Figure 16 presents

a detailed regression diagnostic suite for the Hybrid model on the stratified test set with $N = 7200$ samples.

The predicted-versus-true location plot in Figure 16 evaluates global bias and calibration over the complete normalized feeder span. A concentration of samples around the identity line indicates that the estimator preserves proportionality between the true and predicted locations and does not introduce strong position-dependent distortions. The signed-error histogram evaluates whether localization errors are centered around zero, while the empirical cumulative distribution function (CDF) of the absolute error quantifies the fraction of events localized within a given tolerance. Finally, the absolute-error-versus-aging plot evaluates whether increasing degradation severity produces systematic error growth or heteroscedastic behavior.

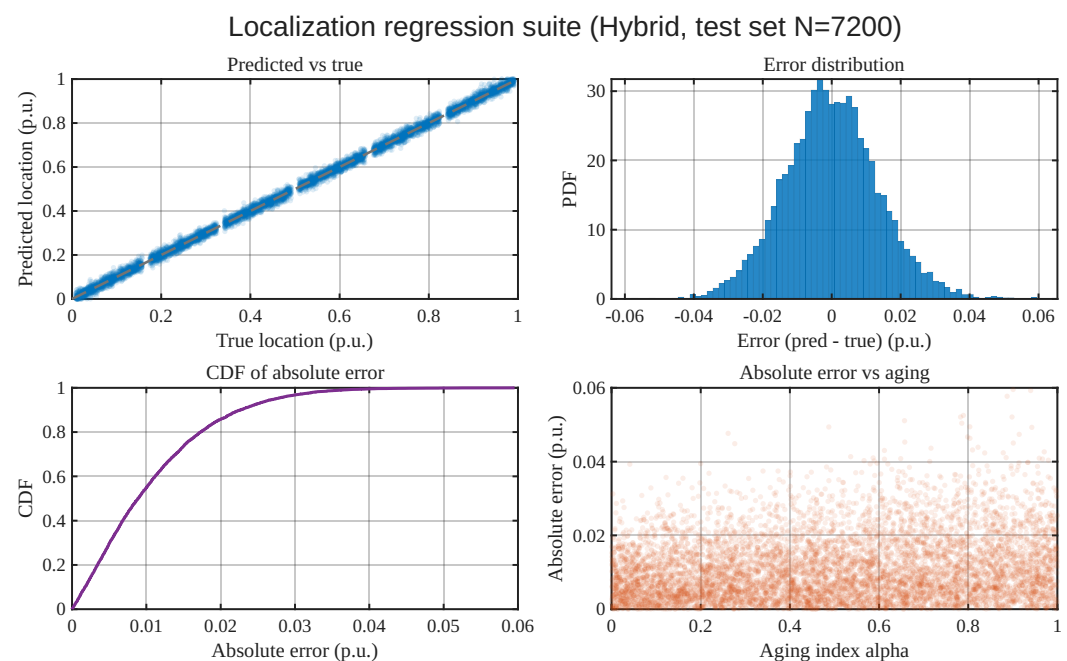


Figure 16. Hybrid localization regression diagnostics on the test set ($N = 7200$): predicted versus true normalized location λ , signed error distribution, empirical CDF of absolute localization error, and absolute error versus normalized aging-stress index α . These diagnostics evaluate bias, dispersion, tail behavior, and sensitivity to aging-induced parameter drift.

As shown in Figure 16, the predicted locations are concentrated close to the identity line, indicating low systematic bias across the feeder. The signed-error distribution is approximately centered around zero, which supports the absence of a dominant positive or negative localization bias. The empirical CDF confirms that most events fall below small absolute-error thresholds, which is consistent with the MAE and P95 values reported in Table 9. The absolute-error-versus-aging panel shows that localization errors remain bounded over the full aging range, although moderate dispersion is expected at higher values of α because aging modifies attenuation, dispersion, and apparent propagation velocity. Thus, Figure 16 provides distributional evidence that complements the aggregate localization metrics and supports the robustness analysis developed in the following subsection.

8.6. Robustness Studies: Noise (SNR), Aging, and Fault Resistance

For practical deployment in underground distribution networks, a fault diagnosis framework must maintain stable performance under variations in measurement noise, insulation degradation, and fault resistance. Robustness is therefore evaluated through

controlled binned analyses in which performance metrics are computed as functions of a single perturbation variable while marginalizing over all remaining scenario parameters. The analyzed metrics include fault-type classification accuracy, area-identification accuracy, and localization mean absolute error (MAE). This methodology isolates the sensitivity of each model to individual stress factors while preserving the statistical representativeness of the test set.

8.6.1. Robustness Within the Prescribed Additive-Noise Range

Figure 17 evaluates performance as a function of SNR over the interval 20–40 dB used for dataset generation. Within this interval, decreasing SNR progressively reduces the fault-type and area-identification accuracies and increases the localization MAE for all evaluated configurations, reflecting the reduced visibility of the high-frequency transient components used for diagnosis and location.

The Hybrid configuration maintains the highest classification accuracies and the lowest localization MAE throughout the evaluated SNR range. The persistence of this ordering across the SNR bins indicates that its comparative advantage is not restricted to the least noisy events in the controlled dataset.

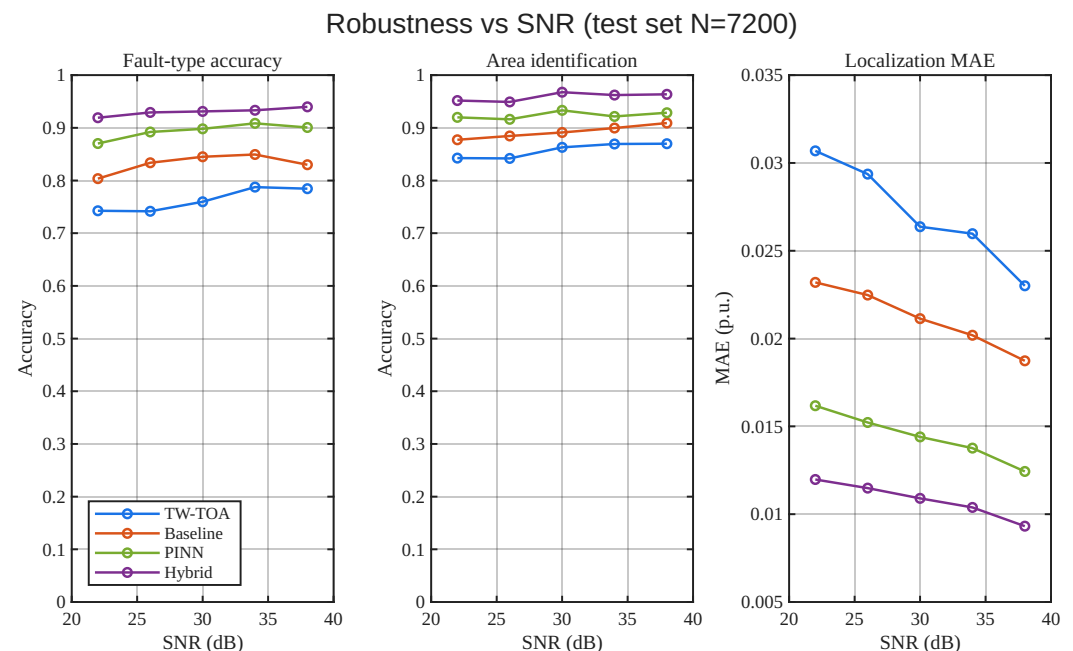


Figure 17. Binned test-set performance as a function of SNR over the simulated range of 20–40 dB: fault-type accuracy, area-identification accuracy, and localization MAE. Decreasing SNR degrades all configurations, while the Hybrid model retains the highest accuracies and the lowest MAE within the prescribed additive-noise domain.

The SNR analysis represents additive channel-noise variation under the ideal synchronized measurement chain defined in Section 7.5. It does not reproduce non-additive field disturbances such as current-transformer saturation, clipping, channel-dependent filtering, timing skew, synchronization loss, impulsive interference, or sensor transfer-function distortion. These effects cannot generally be represented by an equivalent SNR value because they modify waveform shape and arrival-time information rather than only the noise energy. Accordingly, the trends in Figure 17 characterize robustness to additive noise within the simulated interval and do not establish robustness under severe measurement-chain distortion or SNR values below 20 dB.

The normalized aging-stress index modifies the distributed electrical parameters of underground cables and, consequently, affects attenuation, dispersion, and effective propagation velocity within the controlled simulation domain. Figure 18 presents performance as a function of α , where each point represents the average metric within an α -bin. This analysis should be interpreted as sensitivity to structured aging-induced parameter drift—not as validation against field-calibrated cable age.

A consistent trend is observed: increasing α reduces classification and area-identification accuracy while increasing the localization MAE. This behavior is expected because the progressive perturbation of $\mathbf{R}(\alpha)$, $\mathbf{L}(\alpha)$, $\mathbf{C}(\alpha)$, and $\mathbf{G}(\alpha)$ changes the transient attenuation, apparent propagation velocity, and wavefront morphology, thereby increasing the difficulty of both classification and localization. This reflects the growing difficulty of the problem as the waveform distortion increases. Importantly, the Hybrid model exhibits the smallest degradation slope across all metrics, indicating improved robustness to aging-induced variations in signal propagation.

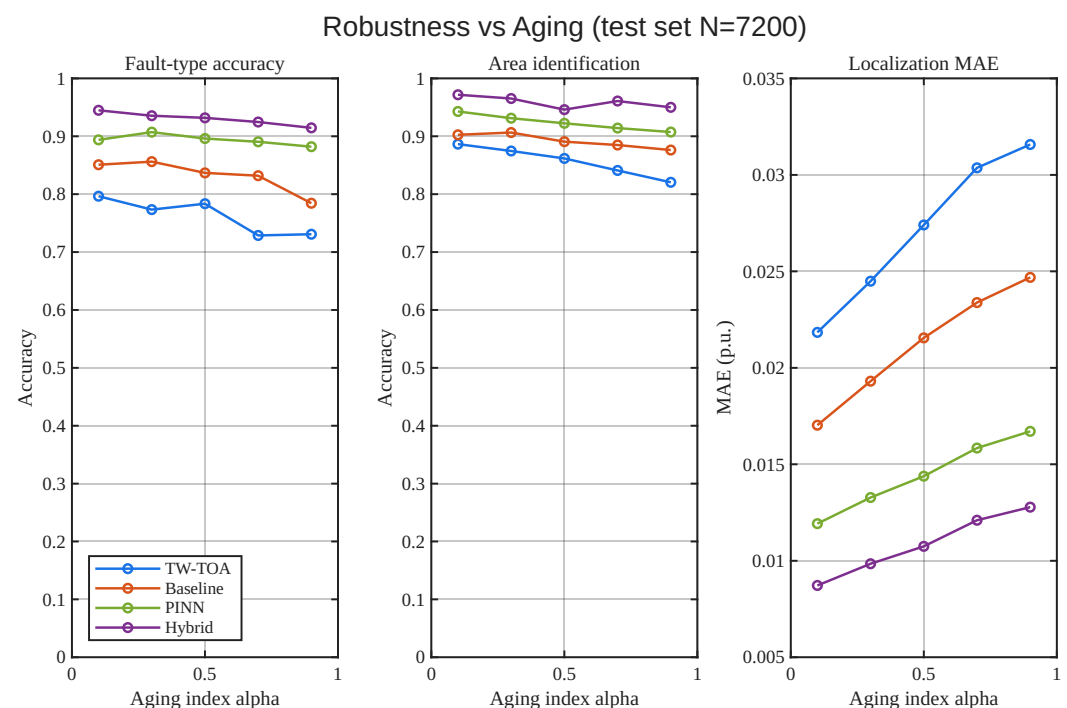


Figure 18. Binned robustness versus normalized aging-stress index α on the test set: fault-type accuracy, area-identification accuracy, and localization MAE. Increasing aging degrades all models, whereas the Hybrid model shows the smallest performance degradation, indicating improved robustness to aging-driven propagation effects.

8.6.2. Robustness Versus Fault Resistance R_f

Fault resistance determines the magnitude of fault currents and the strength of the resulting transient signals. Figure 19 reports performance as a function of R_f , showing that increasing resistance weakens the transient signature and reduces discriminability.

All models exhibit decreased classification accuracy and increased localization MAE as R_f increases. Nevertheless, the Hybrid model consistently maintains the best performance across the full resistance range, achieving the highest discrete-task accuracies and the lowest localization error. This indicates improved stability under high-impedance fault conditions, which are particularly challenging in practical systems.

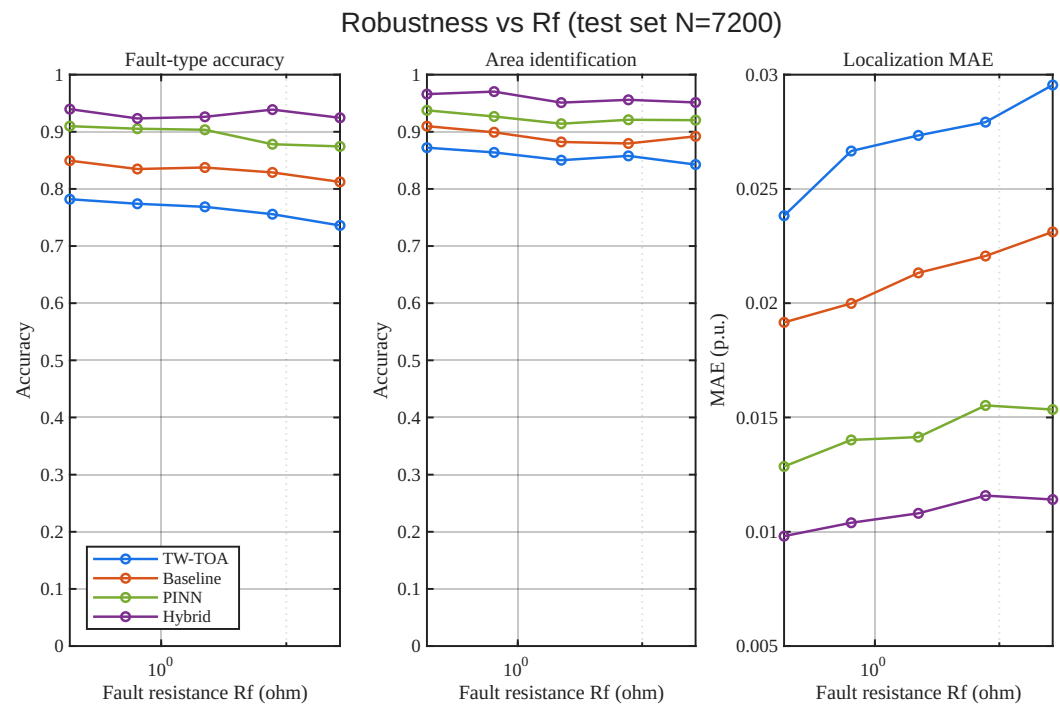


Figure 19. Binned robustness versus fault resistance R_f on the test set: fault-type accuracy, area-identification accuracy, and localization MAE. Increasing R_f weakens transient signatures and degrades performance for all models; the Hybrid model remains the most stable across the evaluated range.

8.7. Topology Sensitivity and Area-Wise Heterogeneity

Underground distribution feeders are not topologically uniform. Lateral branches introduce impedance discontinuities, additional reflection points, and modified propagation paths, which can alter transient signatures and complicate arrival-time interpretation. Therefore, performance is evaluated not only globally but also under two topology-aware perspectives: trunk versus lateral faults and area-wise feeder partitions.

8.7.1. Trunk Versus Lateral Comparison

Figure 20 compares the model performance for faults located on the main trunk and on lateral branches. This figure evaluates whether the reported gains are preserved when the propagation path includes additional branching effects.

Across all models, lateral faults produce slightly lower classification accuracy and higher localization MAE than trunk faults. This behavior is physically consistent with the additional reflections and local impedance mismatches introduced by lateral branches. However, the relative model ordering remains stable in both cases. The Hybrid model achieves the highest fault-type and area-identification accuracies and the lowest localization MAE for both trunk and lateral events, indicating that its improvement is not restricted to the simpler main-trunk topology.

8.7.2. Area-Wise Performance and Junction-Related Ambiguity

Figure 21 reports the fault-type accuracy, area-identification accuracy, and localization MAE separately for the six feeder areas. This disaggregation captures variations in sensor observability, branching structure, and dominant propagation paths that are concealed by the aggregate test-set metrics.

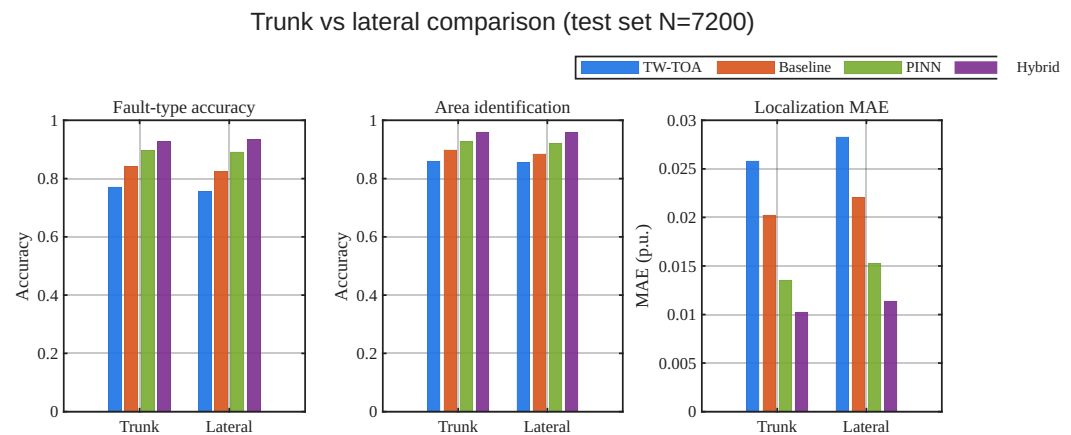


Figure 20. Topology-conditioned performance on the test set ($N = 7200$): fault-type accuracy, area-identification accuracy, and localization MAE for trunk and lateral faults. Lateral events are more challenging due to additional branching effects, but the Hybrid model maintains superior performance in both regimes.

Moderate area-to-area variability is observed particularly in localization MAE. Areas containing or adjoining branching points are exposed to additional reflected and refracted wave components, whereas terminal and boundary areas may provide fewer independent propagation paths. These effects can reduce the separability of first-arrival signatures and increase the possibility that similar transient patterns are associated with neighboring locations.

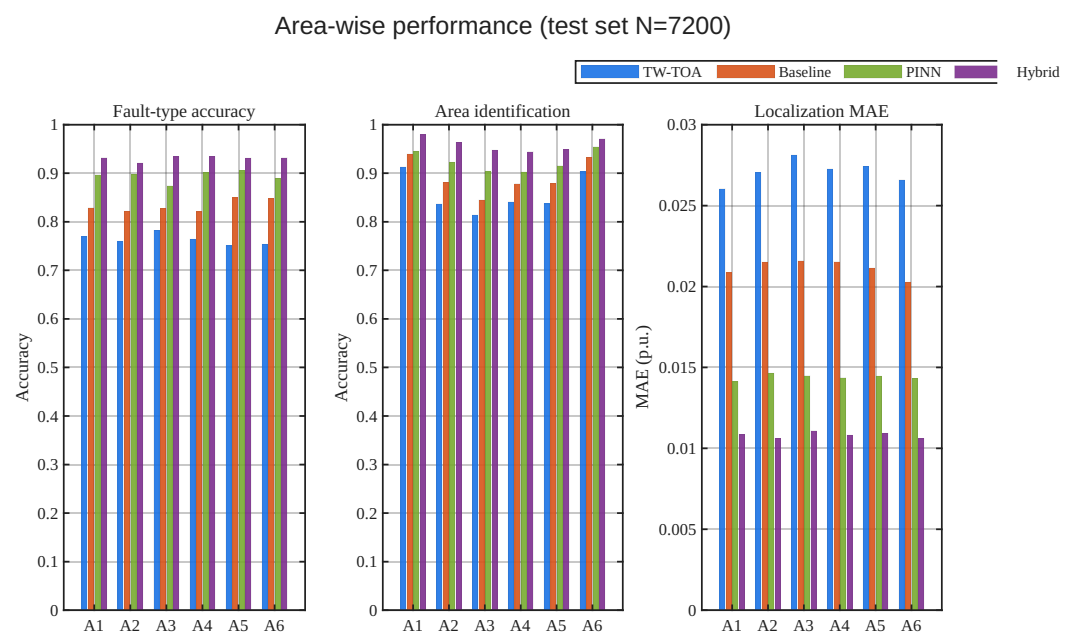


Figure 21. Area-wise test-set performance for feeder partitions $\mathcal{A}_1$ – $\mathcal{A}_6$: fault-type accuracy, area-identification accuracy, and localization MAE. The results show the influence of the area-dependent sensing geometry and branching structure on diagnostic performance.

The Hybrid configuration maintains the most favorable performance among the evaluated methods in each area. Nevertheless, the area-wise averages do not isolate the error distribution as a continuous function of distance from a junction or area boundary. Fault locations are sampled within

$$x_f/\ell_{ef} \in [0.05, 0.95],$$

so exact section endpoints and junction coordinates are excluded, but events remain distributed across regions with different levels of proximity to impedance discontinuities.

A boundary-conditioned assessment can be defined using

$$\delta_{J,n} = \min_{j \in \mathcal{J}} d_{\text{path}}(x_{f,n}, x_j), \quad (139)$$

where $\mathcal{J}$ is the set of feeder junctions and area boundaries, and $d_{\text{path}}(\cdot, \cdot)$ is the shortest feeder-path distance. Localization errors may then be grouped according to $\delta_{J,n}$ to distinguish near-junction events from faults occurring in the interior of cable sections.

The present area-wise results demonstrate consistency across the six predefined partitions, but they should not be interpreted as a dedicated validation of localization immediately adjacent to junctions or area boundaries. Such locations remain especially sensitive to reflection overlap, area-routing errors, propagation-path ambiguity, and uncertainty in the local cable parameters.

8.8. Localization Error Sensitivity Maps: Joint Effects of Aging, SNR, and R_f

The single-stressor analyses in Section 8.6 quantify the marginal effect of SNR, aging, and fault resistance separately. However, the localization performance may also be affected by the interactions among stressors. To evaluate these interactions, Figure 22 presents two-dimensional sensitivity maps for the Hybrid model by binning the test-set samples over pairs of stress variables and aggregating localization-error statistics within each bin.

Let $\lambda \in [0, 1]$ denote the true normalized fault location and $\hat{\lambda}$ its estimate. The signed localization error is defined as

$$e = \hat{\lambda} - \lambda, \quad (140)$$

and the analysis focuses on the absolute error $|e|$. For each bin, Figure 22 reports two statistics: the mean absolute error $\mathbb{E}[|e|]$, which describes the typical localization performance, and the 95th-percentile absolute error $\text{P95}(|e|)$, which characterizes tail deviations. The maps are constructed over (α, SNR) and $(\alpha, \log_{10}(R_f))$, where α is the normalized aging-stress index, SNR is expressed in dB, and R_f is the fault resistance.

The (α, SNR) maps show that the localization error increases when aging becomes more severe and measurement quality decreases. This trend is physically consistent with transient-based localization in underground cables: aging modifies attenuation, dispersion, and effective propagation velocity, whereas low SNR reduces the observability of high-frequency fault-initiated components. Their combined effect increases uncertainty in transient detection and location regression.

The $(\alpha, \log_{10}(R_f))$ maps further indicate that larger errors occur when high normalized aging-stress levels coincide with moderate-to-high fault resistance. As R_f increases, the injected transient becomes weaker and phase-dependent features become less pronounced. Consequently, the high-frequency content used for localization becomes less observable, increasing both the central error and tail risk. The P95 maps exhibit stronger contrast than the mean-error maps, indicating that compounded stressors affect extreme localization deviations more strongly than average performance.

Each bin in Figure 22 is annotated with the corresponding number of test samples. These annotations verify that the observed high-error regions are supported by sufficient samples and are not artifacts of severe bin imbalance. Therefore, Figure 22 provides a joint sensitivity analysis showing how aging, noise, and fault resistance interact to influence Hybrid localization accuracy and tail behavior.

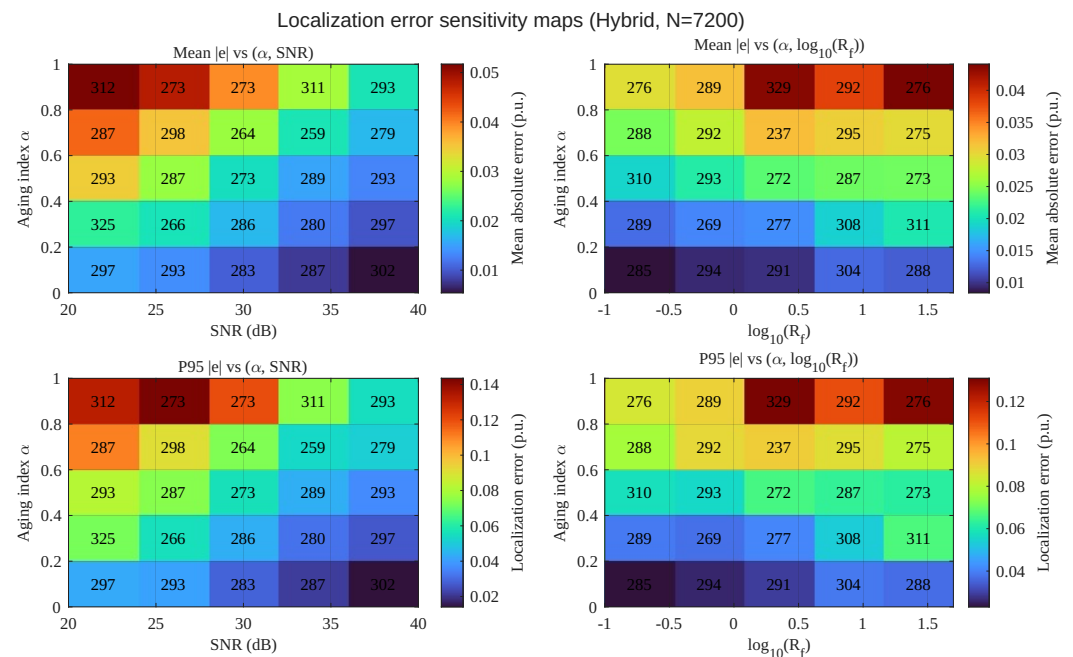


Figure 22. Hybrid localization error sensitivity maps on the test set: binned mean absolute error (top row) and binned 95th-percentile absolute error (bottom row) as functions of (α, SNR) and $(\alpha, \log_{10}(R_f))$. Numbers inside cells denote per-bin sample counts, allowing verification that high-error regions are supported by adequate sample coverage.

8.9. Reliability and Calibration of Localization and Classification Confidence

Point metrics such as MAE and P95 summarize central tendency and tail behavior, but they do not directly quantify operational reliability. For deployment in protection and monitoring systems, it is essential to evaluate the probability that localization errors remain within admissible tolerances and to verify whether confidence estimates reflect actual predictive correctness. This subsection therefore analyzes reliability through empirical coverage functions and evaluates confidence calibration for classification outputs.

8.9.1. Reliability Curves for Localization

Let $\lambda \in [0, 1]$ denote the true normalized fault location and $\hat{\lambda}$ its estimate. For a tolerance $\epsilon \geq 0$, the empirical coverage function is defined as

$$C(\epsilon) = \mathbb{P}(|\hat{\lambda} - \lambda| \leq \epsilon), \quad (141)$$

which corresponds to the cumulative distribution function of the absolute localization error evaluated at ϵ . The function $C(\epsilon)$ is monotonically non-decreasing with $C(0) = 0$ and $\lim_{\epsilon \rightarrow \infty} C(\epsilon) = 1$.

Figure 23 reports $C(\epsilon)$ for all models on the test set. Faster-rising curves indicate a higher probability of achieving a small localization error. For any admissible tolerance ϵ^* , the value $C(\epsilon^*)$ directly represents the probability that the localization error remains within that bound.

As shown in Figure 23, the Hybrid model dominates across the full tolerance range. This implies first-order stochastic improvement in localization accuracy: for any threshold ϵ , the Hybrid model achieves a higher probability of satisfying $|\hat{\lambda} - \lambda| \leq \epsilon$. The steep initial slope of the Hybrid curve indicates a strong concentration of probability mass near zero error, which is consistent with the reduced MAE and P95 reported in Table 9.

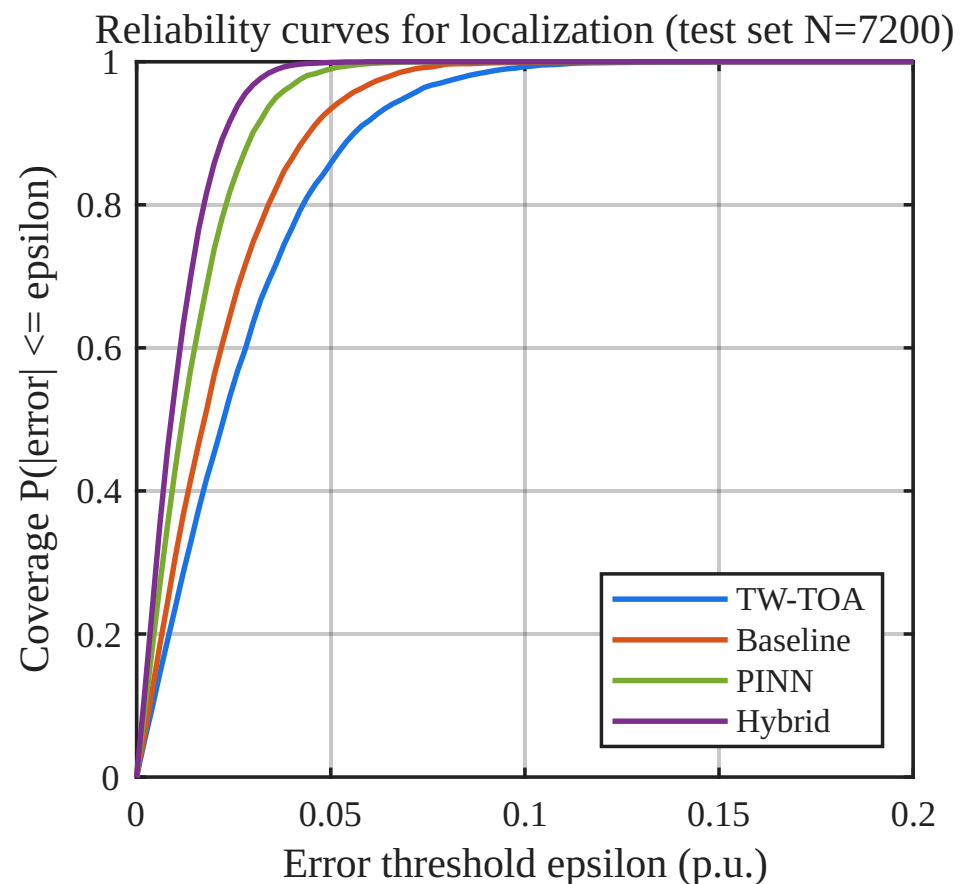


Figure 23. Localization reliability curves on the test set: empirical coverage $C(\epsilon) = \mathbb{P}(|\hat{\lambda} - \lambda| \leq \epsilon)$ as a function of tolerance ϵ . Faster-rising curves indicate a higher probability of satisfying strict localization bounds. The Hybrid model achieves uniformly higher coverage across tolerances, indicating improved reliability.

8.9.2. Calibration and Confidence Analysis

Accurate classification does not necessarily imply reliable probabilistic confidence. For the Baseline, PINN, and Hybrid models, the classification confidence assigned to sample n is defined as the maximum predicted fault-type probability:

$$c_n = \max_{c \in \mathcal{S}_T} \hat{p}_{n,c}. \quad (142)$$

The test samples are partitioned into equal-width confidence intervals, and the empirical accuracy and mean confidence are computed within each nonempty bin. The expected calibration error is

$$\text{ECE} = \sum_{b=1}^{B_{\text{cal}}} \frac{|\mathcal{I}_b|}{N} |\text{acc}(\mathcal{I}_b) - \text{conf}(\mathcal{I}_b)|, \quad (143)$$

where $\mathcal{I}_b$ is the set of samples assigned to confidence bin b , $|\mathcal{I}_b|$ is its sample count, and B_{cal} is the number of calibration bins.

Accordingly, the quantitative calibration interpretation is restricted to the Baseline, PINN, and Hybrid outputs in the empirical-accuracy, reliability, and ECE panels. The bottom density panel illustrates possible differences in confidence concentration but does not constitute an empirical distribution of trained-model outputs. Likewise, the TW-TOA entries are not interpreted as probabilistic calibration results because TW-TOA is a deterministic reference method without a native softmax or posterior-probability output. These calibration results are summarized in Figure 24.

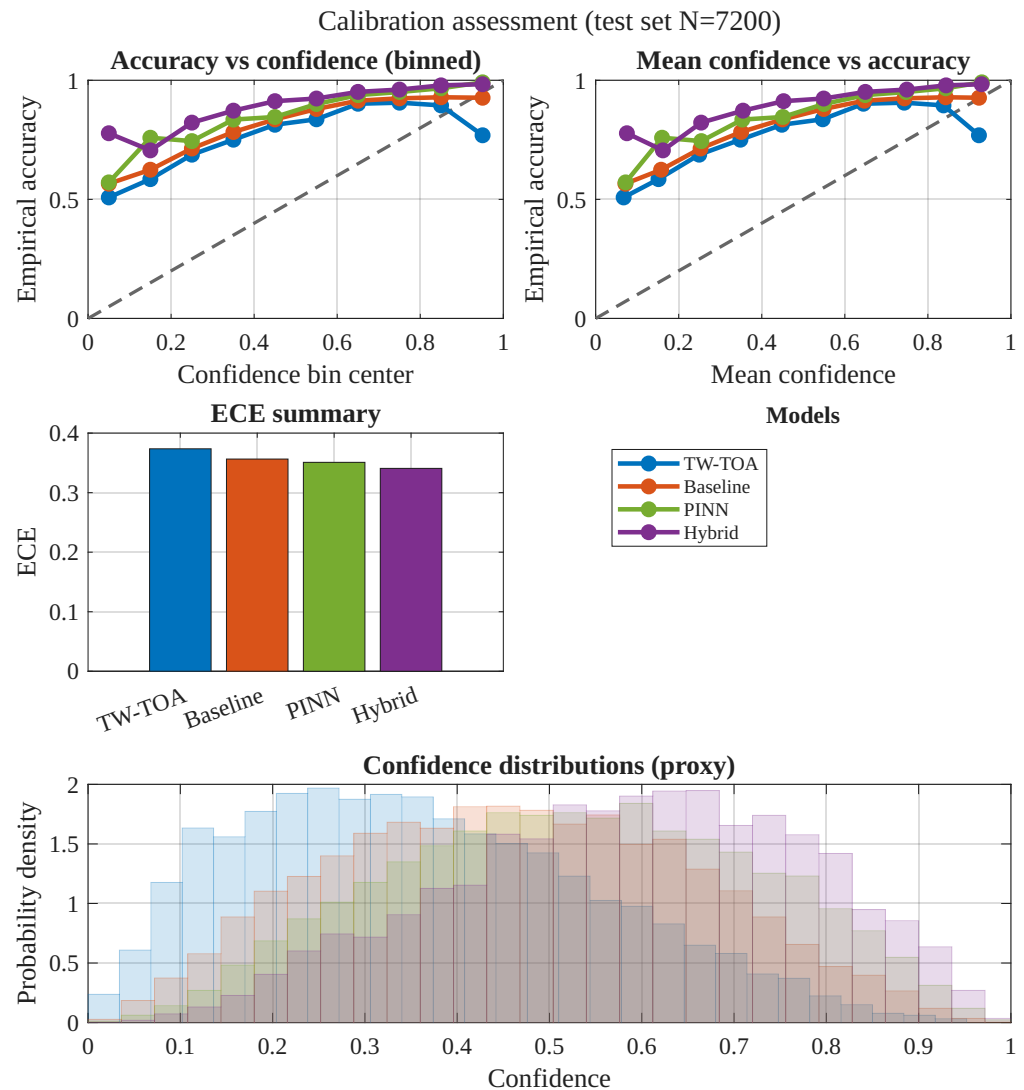


Figure 24. Calibration assessment on the test set. The top-left and top-right panels show binned empirical accuracy and reliability results, respectively, and the middle-left panel reports expected calibration error. For the learning-based models, these three panels are calculated from the fault-type probabilities produced by the trained networks. The TW-TOA traces and the bottom confidence-density panel are proxy visualizations because the deterministic TW-TOA implementation does not produce native class probabilities and complete per-sample confidence distributions were not retained. Proxy elements are included only for qualitative context and are excluded from the quantitative calibration conclusions.

Within the learning-based comparison, the Hybrid model exhibits the closest alignment between confidence and empirical correctness among the evaluated neural configurations. This result supports only a relative calibration comparison within the controlled test domain; it does not establish field-calibrated uncertainty or justify interpreting the proxy panels as measured predictive-confidence distributions.

8.10. Selective Prediction: Coverage–Risk and Coverage–Error Trade-Offs

Selective prediction evaluates diagnostic performance after retaining only the samples with the highest confidence scores. For the learning-based configurations, the fault-type classification confidence of sample n is defined as

$$c_{\text{cls},n} = \max_{c \in S_T} \hat{p}_{n,c} \quad (144)$$

where $\hat{p}_{n,c}$ is the predicted probability of fault class c .

Because the localization heads produce point estimates rather than predictive distributions, the model does not provide a native coordinate-level uncertainty estimate. The confidence used to rank localization predictions is therefore the feeder-area routing confidence

$$c_{\text{loc},n} = \max_{a \in \{1, \dots, A\}} \hat{p}_{\mathcal{A}_a, n}. \quad (145)$$

This quantity measures confidence in selecting the area-specific regression path; it should not be interpreted as a calibrated confidence interval for the continuous location estimate.

For each task, the test samples are sorted in descending order of the corresponding confidence score. At coverage level q , the retained index set is

$$\mathcal{K}(q) = \{\pi(1), \dots, \pi(N_q)\}, \quad N_q = \lfloor qN \rfloor, \quad (146)$$

where π is the confidence-based ordering and

$$q \in \{0.1, 0.2, \dots, 1.0\}. \quad (147)$$

For the test set with $N = 7200$, the retained subsets therefore contain between 720 and 7200 events.

Classification selective risk is defined as

$$R_{\text{cls}}(q) = 1 - \frac{1}{N_q} \sum_{n \in \mathcal{K}_{\text{cls}}(q)} \mathbb{I}[\hat{y}_{\text{cls},n} = y_{\text{cls},n}], \quad (148)$$

whereas localization error at coverage q is

$$E_{\text{loc}}(q) = \frac{1}{N_q} \sum_{n \in \mathcal{K}_{\text{loc}}(q)} |\hat{\lambda}_n - \lambda_n|. \quad (149)$$

Figure 25 reports these coverage–performance relationships for the learning-based configurations. The TW-TOA trace is included only as a qualitative reference because the deterministic TW-TOA method does not produce native fault-class or area probabilities comparable with (144) and (145).

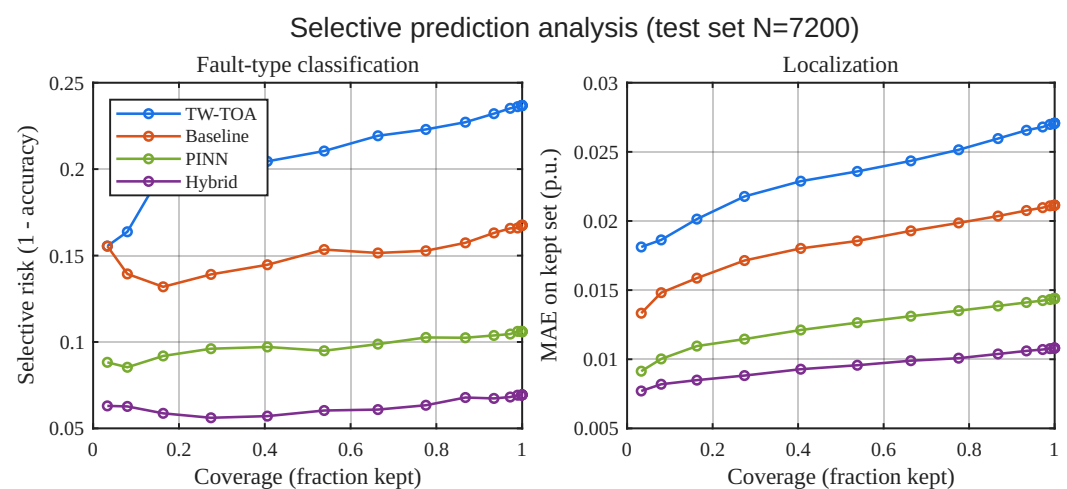


Figure 25. Selective prediction analysis on the test set. The left panel reports classification selective risk, and the right panel reports localization MAE over the retained subsets. For the learning-based configurations, samples are ranked using the maximum fault-type probability for classification and the maximum feeder-area probability for localization. The TW-TOA trace is shown only as a qualitative reference because TW-TOA does not provide native probabilistic confidence outputs.

For the learning-based configurations, both the classification risk and localization MAE decrease as coverage is reduced. This behavior indicates that the probability outputs provide useful rankings of prediction difficulty within the controlled test domain. The Hybrid configuration maintains the lowest selective risk and localization MAE across the evaluated coverage levels. Nevertheless, the localization confidence in (145) quantifies routing certainty rather than uncertainty in the continuous coordinate itself; coordinate-level uncertainty would require a probabilistic regression head, an ensemble, or another dedicated uncertainty-estimation mechanism.

8.11. Uncertainty Quantification via Bootstrap Distributions

Point estimates such as the classification accuracy and localization MAE do not capture the sampling variability associated with a finite test set. Statistical uncertainty is therefore quantified using a paired non-parametric bootstrap procedure applied to the fixed test-set predictions.

Given the test set of size $N = 7200$, a total of $B = 1400$ bootstrap resamples are generated. For each bootstrap repetition $b \in \{1, \dots, B\}$, an index vector

$$\mathcal{I}^{(b)} = \{i_1^{(b)}, \dots, i_N^{(b)}\}, \quad i_k^{(b)} \sim \mathcal{U}\{1, \dots, N\}, \quad (150)$$

is sampled with replacement from the original test-set indices. The same resampled index vector $\mathcal{I}^{(b)}$ is applied to every evaluated method within repetition b , thereby preserving the paired structure of the common test set. The models are not retrained during bootstrapping; the classification accuracy and localization MAE are recomputed from the previously obtained test predictions for each resample.

For a performance metric M with bootstrap values $\{M^{(b)}\}_{b=1}^B$, the reported central estimate is the bootstrap median, and the empirical two-sided 95% percentile confidence interval is

$$\text{CI}_{95\%}(M) = \left[Q_{0.025}(\{M^{(b)}\}_{b=1}^B), Q_{0.975}(\{M^{(b)}\}_{b=1}^B) \right], \quad (151)$$

where $Q_p(\cdot)$ denotes the empirical quantile of order p . This procedure quantifies finite-sample variability without assuming Gaussian metric distributions and ensures that comparisons among methods are based on identical bootstrap realizations.

8.11.1. Bootstrap Distributions

Figure 26 presents the empirical distributions of classification accuracy and localization MAE across bootstrap resamples. The central marker denotes the median of each distribution, while the intervals represent empirical 95% percentile ranges.

As shown in Figure 26, the distributions for different models remain clearly separated across resamples. This indicates that the observed ranking of methods is stable and not driven by a particular realization of the test set, thereby strengthening the reliability of the comparative results.

8.11.2. Bootstrap Confidence Interval Summary

To provide numerical values consistent with Figure 26, Table 11 reports the bootstrap medians and corresponding empirical 95% percentile confidence intervals for the classification accuracy and localization MAE.

Table 11 confirms that the Hybrid model not only achieves the best median performance but also exhibits relatively narrow confidence intervals. This indicates that its superiority is robust under sampling variability and not sensitive to the particular composition of the test set.

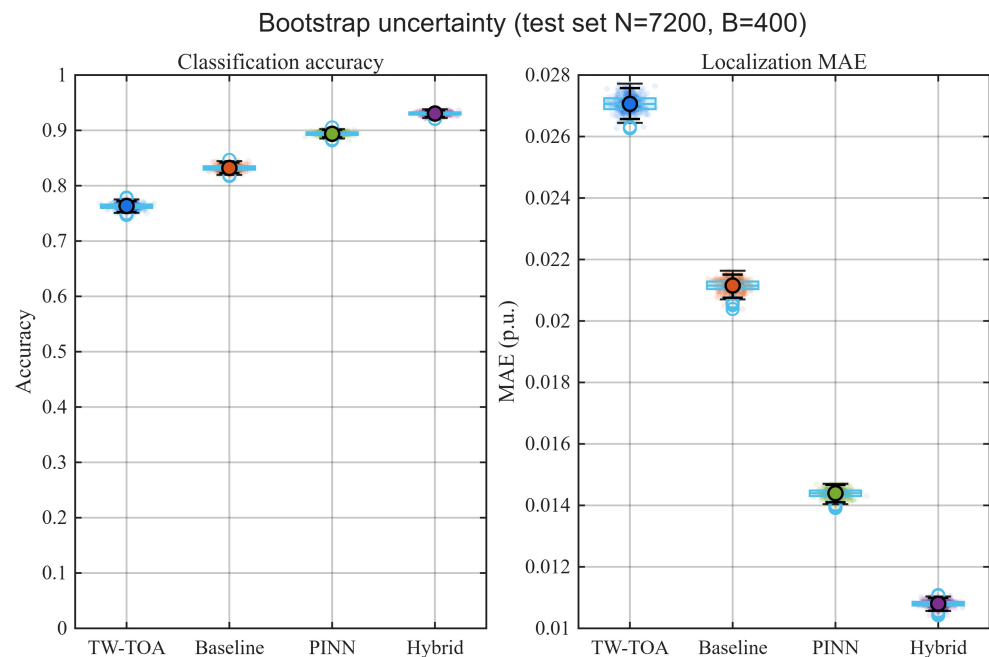


Figure 26. Bootstrap distributions on the test set: classification accuracy (**left**) and localization MAE (**right**) across bootstrap resamples. Markers indicate medians and intervals represent empirical 95% percentile ranges.

Table 11. Bootstrap summary on the test set: median and empirical 95% percentile confidence intervals for classification accuracy and localization MAE.

Model	Accuracy (Median [95% CI])	MAE (p.u.) (Median [95% CI])
TW-TOA	0.76 [0.75, 0.77]	0.027 [0.0265, 0.0275]
Baseline	0.83 [0.82, 0.84]	0.021 [0.0208, 0.0216]
PINN	0.89 [0.88, 0.90]	0.014 [0.0140, 0.0146]
Hybrid	0.93 [0.92, 0.94]	0.011 [0.0105, 0.0111]

8.12. Global Trade-Offs: Pareto Frontier and Radar Comparison

Because the proposed framework jointly addresses fault-type classification, area identification, and continuous localization, performance must be interpreted in a multi-objective setting. A model is operationally preferable only if improvements in one objective do not incur degradation in others. Therefore, joint performance is analyzed using Pareto and normalized multi-metric visualizations.

8.12.1. Pareto Trade-Off Between Localization Error and Fault-Type Accuracy

Figure 27 represents each model in a two-dimensional objective space defined by fault-type accuracy (to be maximized) and localization MAE (to be minimized). In this plane, a model Pareto-dominates another if it achieves no worse performance in both objectives and strictly better performance in at least one.

As observed in Figure 27, the Hybrid model occupies the most favorable position in the Pareto plane, achieving both higher classification accuracy and lower localization error than the alternatives. This demonstrates that the improvement is not obtained through a trade-off but reflects a genuinely superior joint operating point.

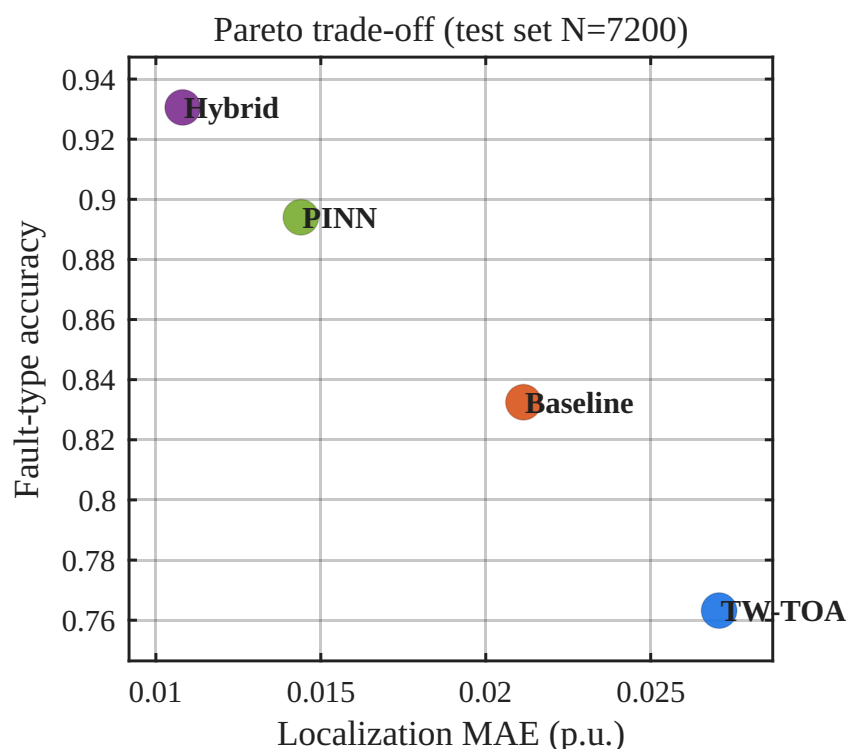


Figure 27. Pareto representation of fault-type accuracy versus localization MAE on the test set. Each point corresponds to one model. Points closer to the upper-left region indicate higher classification accuracy and lower localization error.

8.12.2. Radar Comparison Across Normalized Objectives

While Figure 27 focuses on two primary objectives, Figure 28 extends the comparison to multiple metrics, including fault-type accuracy, area-identification accuracy, and localization performance expressed through normalized inverse MAE and inverse P95. All axes are oriented so that larger values correspond to better performance.

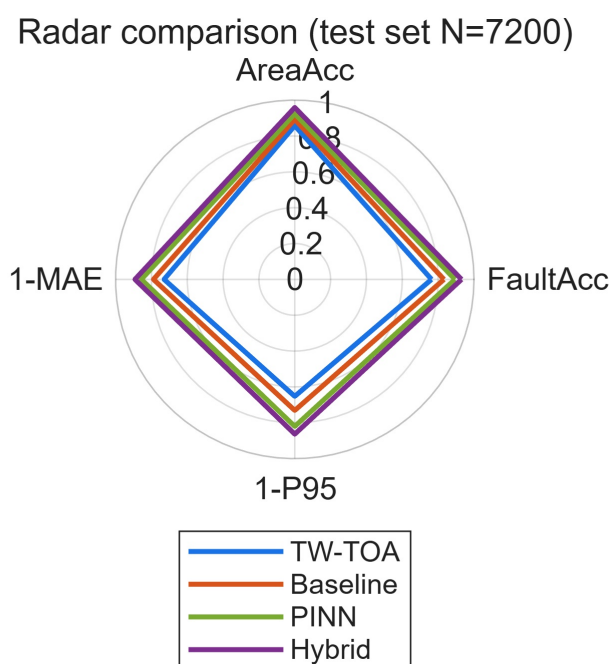


Figure 28. Radar comparison across normalized objectives with larger values indicating better performance on all axes: fault-type accuracy, area-identification accuracy, inverse localization MAE, and inverse localization P95. Normalization ensures consistent interpretation across metrics.

Figure 28 shows that the Hybrid model expands outward across all axes, indicating balanced improvements across classification, area identification, and both average and tail localization errors. This confirms that the proposed framework achieves a consistent multi-objective advantage rather than isolated gains in individual metrics.

8.13. Integrated-Configuration Summary and Attribution Limits

Figure 29 provides a compact synthesis of the principal task metrics for the four evaluated diagnostic configurations. It consolidates fault-type accuracy, area-identification accuracy, localization MAE, and localization P95, and it reproduces the numerical trends reported in Tables 9 and 10.

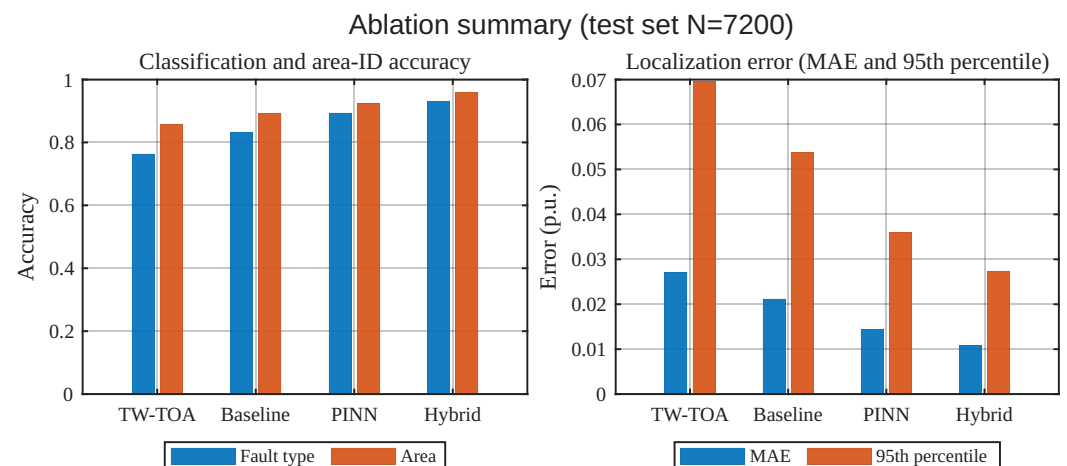


Figure 29. Integrated-configuration performance summary on the test set. The panels compare fault-type and area-identification accuracies, localization MAE, and localization P95 for TW-TOA, Baseline, PINN, and Hybrid. The figure compares complete diagnostic configurations and should not be interpreted as a component-wise ablation of the Hybrid architecture.

The Baseline–PINN comparison provides a controlled assessment of adding the propagation-consistency term while retaining the same global supervised architecture. By contrast, the difference between PINN and Hybrid reflects the combined effect of several simultaneous modifications: area-conditioned routing, fixed sensor-subset selection, area-specific local regression heads, and physics regularization within the complete diagnostic pipeline. Consequently, the additional improvement of the Hybrid configuration cannot be assigned to any one of these components individually.

Within this scope, Figure 29 shows that the complete Hybrid configuration achieves the most favorable joint operating point among the evaluated alternatives. The comparison supports the effectiveness of the integrated formulation as a whole, while a factorial ablation in which routing, sensor selection, local regression, and physics regularization are independently enabled or disabled would be required to quantify their separate marginal contributions.

Comparison with Learning-Based Fault-Location Approaches

The learning-based fault diagnosis and location methods reported in the literature include support-vector-machine classifiers, fuzzy diagnostic systems, wavelet-based transient-feature models, and data-driven fault-location algorithms [26–28]. These approaches demonstrate that transient or transformed electrical signals can support accurate diagnostic decisions when the training data represent the operating domain of interest.

A direct numerical ranking against this paper’s results is not appropriate because the reported studies differ in feeder topology, voltage level, fault-label space, sensor configuration, sampling frequency, signal representation, physical distance scale, noise model, and training–test partition. In particular, the classification accuracy obtained for a fault-type-

only problem is not equivalent to the joint fault-type, feeder-area, and continuous-location task considered here, and a location error expressed in meters or as a percentage of a specific line cannot be converted uniquely to the normalized branched-feeder coordinate used in this paper.

The performance improvement reported in Tables 9 and 10 is therefore meaningful as a controlled within-study comparison: TW-TOA, Baseline, PINN, and Hybrid are evaluated on the same event records, input channels, label definitions, and parameter ranges. The results establish that the complete Hybrid configuration outperforms these matched references within the prescribed simulation domain.

Wavelet-feature SVMs, tree ensembles, recurrent networks, and transformer-based models were not trained on the present dataset. Consequently, the reported results do not establish superiority over those architecture families. A direct architecture-level comparison would require identical training, validation, and test records; the same fault and area labels; matched preprocessing and sensor information; comparable hyperparameter-selection budgets; and common classification and localization metrics. Under these conditions, differences could be attributed to the learning architecture rather than to incompatible datasets or evaluation protocols.

Taken together, the results support four main observations. First, the waveform, HF-envelope, and time–frequency analyses in Figures 9–12 confirm that the simulated dataset contains event-localized, phase-dependent transient signatures consistent with traveling-wave-like propagation. This indicates that the learning problem is grounded in physically meaningful signal content rather than artificial class separability.

Second, the quantitative results in Tables 9 and 10, together with Figure 14, show that the Hybrid model provides coherent multi-task improvement across fault-type classification, area identification, and localization. The reduction in both MAE and P95 indicates that the model improves average accuracy and reduces tail-error risk.

Third, the robustness analyses in Figures 17–22 demonstrate that performance trends remain stable under variations in SNR, normalized aging-stress index, fault resistance, and compounded stress conditions. The topology-conditioned and area-wise analyses in Figures 20 and 21 further show that the gains are not restricted to a single feeder region or topology subset.

Fourth, the reliability, calibration, selective prediction, and bootstrap analyses in Figures 23–26 indicate that the reported performance is stable under test-set resampling and that confidence scores carry useful information for ranking prediction difficulty. This supports the use of confidence-aware decision policies, such as abstaining from low-confidence cases or prioritizing them for operator review.

Overall, the empirical evidence indicates that the Hybrid approach provides consistent improvements across complementary objectives while maintaining stability under structured perturbations and finite-sample uncertainty. Nevertheless, because the evaluation is based on controlled simulation data, the results should be interpreted as simulation-based validation. Future work should incorporate laboratory or field cable-fault measurements, heterogeneous aging scenarios, and independently validated frequency-dependent cable models to further assess deployment readiness.

9. Conclusions

This paper developed a physics-regularized Hybrid framework for joint fault-type classification, feeder-area identification, and continuous fault localization in aging underground distribution networks. The formulation combines synchronized multi-sensor time–frequency inputs, an output-dependent propagation-consistency loss, area-conditioned routing, and area-specific local regression heads. Cable degradation is represented through

a normalized aging-stress index that imposes controlled parameter drift rather than a field-calibrated asset-health model.

The framework was evaluated on a synthetic branched feeder using 36,000 simulated events spanning 11 fault classes, 6 feeder areas, trunk and lateral locations, $R_f \in [0.1, 50] \Omega$, $\text{SNR} \in [20, 40] \text{ dB}$, and $\alpha \in [0, 1]$. On the held-out test set of 7200 event records, the Hybrid configuration achieved a fault-type accuracy of 0.93, an area-identification accuracy of 0.96, a localization MAE of 0.011 p.u., and a 95th-percentile absolute error of 0.027 p.u. It provided the most favorable joint performance among TW-TOA, Baseline, PINN, and Hybrid under the common simulation protocol.

The Baseline–PINN comparison supports the contribution of the propagation-consistency term under matched global supervised architectures. The additional difference between PINN and Hybrid represents the combined effect of area-conditioned routing, sensor-subset selection, local regression, and physics regularization; therefore, it cannot be assigned to any one component individually. The robustness and bootstrap analyses further show that the ordering of the evaluated configurations remains stable within the prescribed ranges of additive noise, normalized aging stress, fault resistance, and feeder topology.

The time–frequency saliency overlay included in this paper is a qualitative proxy and does not constitute model-native evidence of explanation faithfulness. Quantitative interpretability claims require Grad-CAM-, SHAP-, or equivalent attribution maps computed directly from the trained classification and localization outputs and evaluated using the proposed concentration, stability, compactness, and occlusion criteria.

The reported results characterize the interpolation performance for unseen events generated by the same simulator and parameter domain. They do not establish cross-simulator generalization, field accuracy, real-time execution, performance at lower sampling rates, or robustness to non-additive measurement-chain disturbances. Practical implementation would require synchronized transient-capable voltage and current sensors, sufficient analog bandwidth, calibrated sensor transfer functions, accurate feeder topology and cable parameters, condition-dependent propagation-velocity estimation, and calibration of the aging representation using laboratory or field indicators.

Further validation should use independently generated electromagnetic-transient waveforms, laboratory cable-fault experiments, and field recordings. It should also examine heterogeneous and frequency-dependent aging, imperfect synchronization, sensor saturation and filtering, grouped or out-of-range test scenarios, lower sampling frequencies, component-wise ablations, matched wavelet/SVM/transformer baselines, and measured end-to-end inference latency. These steps are required to determine whether the comparative advantages observed in the controlled simulation domain are retained under operational conditions.

Author Contributions: Conceptualization, A.A.T. and F.J.; methodology, A.A.T.; software, A.A.T.; validation, A.A.T., F.J., M.J. and P.L.; formal analysis, A.A.T.; investigation, A.A.T., F.J., M.J. and P.L.; resources, A.A.T. and M.J.; data curation, A.A.T.; writing—original draft preparation, A.A.T. and F.J.; writing—review and editing, A.A.T., F.J., M.J. and P.L.; visualization, A.A.T.; supervision, F.J. and M.J.; project administration, A.A.T.; funding acquisition, A.A.T. and M.J. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding. The APC was funded by Universidad Politécnica Salesiana.

Data Availability Statement: The data supporting the findings of this study were generated through the controlled numerical simulation framework described in the manuscript. The electrical modeling formulation, algorithms, pseudocode, scenario ranges, signal-processing procedures, model architecture, and evaluation protocol are presented in this paper to support methodological transparency and

reproducibility. Additional information and materials related to the simulation and implementation workflow may be obtained from the corresponding author upon reasonable request. The data are not publicly available because the simulation datasets have not been deposited in a public repository.

Acknowledgments: The authors acknowledge the support of Universidad Politécnica Salesiana for the development of this research. During the preparation of this manuscript, the authors used OpenAI's ChatGPT (GPT-5.6 Thinking) solely to improve the wording and language quality of the manuscript. It was not used for data generation, simulation, methodology development, data analysis, result interpretation, or the creation of scientific content. The authors reviewed and edited the final text and take full responsibility for the content of this publication.

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

References

- Choudhary, M.; Shafiq, M.; Kiitam, I.; Hussain, A.; Palu, I.; Taklaja, P. A review of aging models for electrical insulation in power cables. *Energies* **2022**, *15*, 3408. <https://doi.org/10.3390/en15093408>.
- Stojanović, J.; Klimenta, D.; Panić, P.; Mitrović, D.; Prokić, M.; Jevtić, M.; Petrović, M. Thermal aging management of underground power cables using FEM-based Arrhenius analysis. *Electr. Eng.* **2023**, *105*, 647–662. <https://doi.org/10.1007/s00202-022-01689-z>.
- Hassan, W.; Shafiq, M.; Hussain, G.A.; Choudhary, M.; Palu, I. Investigating the progression of insulation degradation in power cable based on partial discharge measurements. *Electr. Power Syst. Res.* **2023**, *220*, 109452. <https://doi.org/10.1016/j.epsr.2023.109452>.
- Li, G.; Wang, Z.; Lan, R.; Wei, Y.; Nie, Y.; Li, S.; Lei, Q. The lifetime prediction and insulation failure mechanism of XLPE for high-voltage cable. *IEEE Trans. Dielectr. Electr. Insul.* **2023**, *30*, 761–768. <https://doi.org/10.1109/TDEI.2022.3215605>.
- Karimi, M.; Galijasevic, Z.; Lehtonen, M.; Lahtinen, M. Review of fault location methods for transmission lines based on traveling waves. *IET Gener. Transm. Distrib.* **2016**, *10*, 1125–1135. <https://doi.org/10.1049/iet-gtd.2015.0899>.
- Lin, S.; He, Z.; Li, X. Travelling wave time-frequency characteristic-based fault location method for transmission lines. *IET Gener. Transm. Distrib.* **2012**, *6*, 764–772. <https://doi.org/10.1049/iet-gtd.2011.0703>.
- Lopes, F.V.; Fernandes, D.; Neves, W.L.A. A traveling-wave detection method based on Park's transformation for fault locators. *IEEE Trans. Power Deliv.* **2013**, *28*, 1626–1634. <https://doi.org/10.1109/TPWRD.2013.2260182>.
- Deng, F.; He, Z.; Zhang, Y.; Hu, W. A single-ended fault location method based on full waveform feature extraction of traveling waves. *IEEE Trans. Power Deliv.* **2023**, *38*, 2585–2595. <https://doi.org/10.1109/TPWRD.2023.3248102>.
- Pathirage, T.; Wijayapala, W.H.M.S.; Ekanayake, J. Fault location in power distribution networks using frequency analysis of traveling waves. *Electr. Power Syst. Res.* **2025**, *236*, 111971. <https://doi.org/10.1016/j.epsr.2025.111971>.
- Wang, Y.; Xie, L.; Liu, F.; Yu, K.; Zeng, X.; Bi, L.; Tang, X. Fault location method for distribution network considering distortion of traveling wavefronts. *Int. J. Electr. Power Energy Syst.* **2024**, *159*, 110065. <https://doi.org/10.1016/j.ijepes.2024.110065>.
- Sessa, S.D.; Sanniti, F.; Greco, A.; Talomo, S.; Pajussin, M.; Benato, R. An online single-ended traveling waves fault detection algorithm for high-voltage multi-branch overhead lines. *IEEE Access* **2024**, *12*, 89691–89706. <https://doi.org/10.1109/ACCESS.2024.3419234>.
- Cheng, L.; Wang, T.; Wang, Y. A novel fault location method for distribution networks with distributed generations based on the time matrix of traveling-waves. *Prot. Control Mod. Power Syst.* **2022**, *7*, 46. <https://doi.org/10.1186/s41601-022-00265-8>.
- Tariq, R.; Alhamrouni, I.; Rehman, A.U.; Eldin, E.T.; Shafiq, M.; Ghamry, N.A.; Hamam, H. An optimized solution for fault detection and location in underground cables based on traveling waves. *Energies* **2022**, *15*, 6468. <https://doi.org/10.3390/en15176468>.
- Wang, Z.; Zhang, Y.; Li, M.; Chen, H. A micro-PMU-based fault location method for distribution networks with multiple branches. *Int. J. Electr. Power Energy Syst.* **2025**, *166*, 111042. <https://doi.org/10.1016/j.ijepes.2025.111042>.
- Biswal, C.; Sahu, B.K.; Rout, P.K.; Mishra, M. A critical review on traveling wave-based fault assessment and enhanced protection of distribution networks in smart grid scenario. *Unconv. Resour.* **2025**, *8*, 100242. <https://doi.org/10.1016/j.uncres.2025.100242>.
- Huang, B.; Wang, J. Applications of physics-informed neural networks in power systems: A review. *IEEE Trans. Power Syst.* **2023**, *38*, 572–588. <https://doi.org/10.1109/TPWRS.2022.3162473>.
- Stiasny, J.; Misyris, G.S.; Chatzivasileiadis, S. Physics-informed neural networks for nonlinear system identification for power system dynamics. In Proceedings of the 2021 IEEE Madrid PowerTech, Madrid, Spain, 28 June–2 July 2021; pp. 1–6. <https://doi.org/10.1109/PowerTech46648.2021.9495063>.
- Tran, M.-Q.; Zamzam, A.S.; Nguyen, P.H. Physics-informed graphical neural network for power system state estimation. *IEEE Trans. Smart Grid* **2021**, *12*, 3326–3336. <https://doi.org/10.1109/TSG.2021.3052998>.
- Li, W.; Deka, D. Physics-informed learning for high-impedance fault detection. In Proceedings of the 2021 IEEE Madrid PowerTech, Madrid, Spain, 28 June–2 July 2021; pp. 1–6. <https://doi.org/10.1109/PowerTech46648.2021.9494807>.

20. Nellikkath, R.; Chatzivasileiadis, S. Physics-informed neural networks for AC optimal power flow. *Electr. Power Syst. Res.* **2022**, *212*, 108412. <https://doi.org/10.1016/j.epsr.2022.108412>.
21. Hsu, C.-C.; Frusque, G.; Forest, F.; Macedo, F.; Franck, C.M.; Fink, O. Explainable AI guided unsupervised fault diagnostics for high-voltage circuit breakers. *Reliab. Eng. Syst. Saf.* **2025**, *244*, 111199. <https://doi.org/10.1016/j.ress.2025.111199>.
22. Naidu, O.P.; Pradhan, A.K. A traveling wave-based fault location method using unsynchronized current measurements. *IEEE Trans. Power Deliv.* **2019**, *34*, 505–513. <https://doi.org/10.1109/TPWRD.2018.2875598>.
23. Naidu, O.P.; Pradhan, A.K. Model-free traveling wave based fault location method for series compensated transmission line. *IEEE Access* **2020**, *8*, 193128–193137. <https://doi.org/10.1109/ACCESS.2020.3032458>.
24. El-Ghany, H.A.A.; Azmy, A.M.; Abeid, A.M. A general travelling-wave-based scheme for locating simultaneous faults in transmission lines. *IEEE Trans. Power Deliv.* **2020**, *35*, 130–139. <https://doi.org/10.1109/TPWRD.2019.2931178>.
25. Mu, D.; Lin, S.; He, P.; Li, X. An improved method of traveling wave protection for DC lines based on the compensation of line-mode fault voltage. *IEEE Trans. Power Deliv.* **2023**, *38*, 1720–1730. <https://doi.org/10.1109/TPWRD.2022.3222375>.
26. Li, M.; Zhang, H.; Chen, Z.; Wang, Q. Design of fault location algorithm based on online distributed travelling wave for HV power cable. *PLoS ONE* **2023**, *18*, e0296513. <https://doi.org/10.1371/journal.pone.0296513>.
27. Catucuamba, J.; Aguila Téllez, A. Design of a generic fault diagnosis model for electrical distribution networks using a support vector machine (SVM) algorithm. *IEEE Access* **2025**, *13*, 160175–160192. <https://doi.org/10.1109/ACCESS.2025.3603450>.
28. Perez, R.; Inga, E.; Aguila, A.; Vásquez, C.; Lima, L.; Vilorio, A.; Henry, M. Fault diagnosis on electrical distribution systems based on fuzzy logic. In *Advances in Swarm Intelligence*; Tan, Y., Shi, Y., Tang, Q., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 174–185. https://doi.org/10.1007/978-3-319-93818-9_17.
29. Xu, L.; Wang, X.; Zhao, Z. Physics-informed machine learning for fault diagnosis: A review. *Adv. Eng. Inform.* **2024**, *60*, 102806. <https://doi.org/10.1016/j.aei.2024.102806>.

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.