Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Objective-Driven Efficient Sparse Metering for Machine Learning-Based Fault Diagnosis and Self-Healing Restoration in Active Distribution Networks

Omar Gamal Abdelmohsen

Omar.Gamal56@sh-eng.menofia.edu.eg

Menoufia University

Tamer Amin Kawady

Menoufia University

Tamer Fatouh Atia

Menoufia University

Eman Salah Ali

Menoufia University

Article

Keywords:

Posted Date: November 2nd, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11245800/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Objective-Driven Efficient Sparse Metering for Machine Learning-Based Fault Diagnosis and Self-Healing Restoration in Active Distribution Networks

Omar Gamal Abdelmohsen[1,*], Tamer Amin Kawady[1], Tamer Fatouh Atia[1], and Eman Salah Ali[1]

[1]Department of Electrical Power and Machines, Faculty of Engineering, Menoufia University, Shebin El-Kom 32511, Egypt

[*]Corresponding author: Omar.Gamal56@sh-eng.menofia.edu.eg

Machine learning is a leading research approach to the diagnostic stage of fault location, isolation and service restoration (FLISR), but most schemes assume dense or synchronized metering and few link diagnostic accuracy to reliability. We present an objective-driven framework in which a few low-cost, unsynchronized feeder meters support a learning pipeline for fault classification, section localization, isolation, and self-healing restoration on the IEEE 33-bus system in OpenDSS/MATLAB. The two diagnostic tasks differ markedly in their sensing requirements: classification saturates with three to four meters, whereas localization keeps improving up to sixteen. The difference extends to model choice: the learner most accurate on clean data is not the one that minimizes interruption duration. A greedy wrapper run separately per task, under an event-held-out protocol, beats coverage, filter and random placement at all three budgets tested. Minimizing a lifetime cost that prices meters and unserved energy selects seven meters; there faults are localized to within one hop with 97.3% accuracy and FLISR cuts SAIDI, the system average interruption duration index, by 82.2% under IEEE Std 1366, with a nine-month payback. A data-driven isolability measure explains why localization errors fall on adjacent sections, and a closed-form relation maps accuracy to SAIDI.

Introduction

Distribution feeders account for the majority of customer interruption minutes. Therefore, quickly locating, isolating, and restoring a faulted section is essential for improving the reliability indices reported by utilities[1, 2, 3]. Fault location, isolation, and service restoration (FLISR) aims to automate this process, and machine learning has become a leading research direction for the diagnostic part of FLISR, although deployed systems still rely mainly on fault indicators and SCADA logic.

Recent research has substantially improved machine-learning algorithms for fault diagnosis. Mirshekali *et al.* achieved accurate fault location using a capsule-based convolutional network applied to feeder waveforms[4]. Jiang developed a data-driven fault locator that remains effective even when distributed generation (DG) changes fault-current contributions[5]. Mousavi *et al.* used neural networks to locate faults in DG-integrated distribution networks while explicitly considering fault impedance[6]. Bhagwat *et al.* developed an artificial neural network for distribution fault detection[7]. Although these

studies improve diagnostic performance, they all assume that the measurement infrastructure is predefined. The number and locations of instrumented buses are fixed, and none examines how performance changes when fewer meters are installed or when their locations are modified. In addition, these studies evaluate only diagnostic accuracy without assessing its impact on system reliability.

Several studies have also investigated optimal meter placement. Krysander and Frisk formulated sensor placement based on fault detectability and isolability within a model-based residual framework[8]. Other studies used observability- or coverage-based methods to identify the minimum number of sensors required to monitor the network[9]. Closest to the present work, two recent studies place devices so as to maximize the performance of a learning-based locator rather than to satisfy a coverage rule. Marín-Quintero *et al.* allocate fault indicators to maximize the accuracy of an artificial-neural-network fault locator[10]. Orozco-Henao *et al.* jointly optimize the number and position of intelligent electronic devices (IEDs) and fault indicators, using a fitness function that combines the prediction error of a classification and regression tree (CART) model with the standard deviation of unserved power. Their method reaches 98.3% FLISR accuracy on the IEEE 123-node feeder[11]. That figure and the accuracies reported here are not comparable: the feeders, the zone definitions, and the accuracy conventions all differ, and a like-for-like comparison would require re-implementing their method on the present benchmark, which is left to future work. These works establish that device placement can be driven by learner performance and that unserved energy belongs in the objective, and the present paper claims neither idea as new. What they share with the earlier literature is a *single* placement applied to all FLISR functions. They do not examine whether different FLISR tasks require different levels of spatial information. As a result, it remains unclear whether a meter placement that performs well for one task is sufficient, excessive, or inadequate for another.

Another group of studies focuses on service restoration after the fault has been identified. Zidan *et al.* reviewed the development and future trends of FLISR systems[3]. Wang *et al.* proposed a multistage restoration strategy for medium-voltage distribution networks with distributed generation[12], while Wang and Wang developed a self-healing strategy based on network sectionalization into microgrids[13]. Although these studies improve restoration strategies, they generally assume that the faulted section has already been identified, and their economic analyses are limited. Classical fault-indicator placement already weighs device cost against interruption cost and restoration time[14, 15], but it does not involve a learned diagnosis whose accuracy must itself be measured. To our knowledge, no study has yet carried the measured accuracy of a learned fault diagnosis through to IEEE Std 1366 reliability indices and an investment decision.

Three important limitations can be identified in the existing literature. **First**, most FLISR methods assume dense or synchronized metering, although such infrastructures are often impractical because of their high installation and maintenance costs. **Second**, meter placement is generally optimized for a single objective, without considering that different FLISR tasks may require different levels of sensing information. **Third**, diagnostic performance is usually reported without evaluating its impact on system reliability or the economic benefit of the proposed solution.

To address these limitations, this paper makes the following contributions.

1. **Objective-driven meter placement** (addresses the second limitation). A greedy wrapper is applied independently to fault classification and fault localization. Unlike the shared placement assumed in earlier sensor-siting work[8, 9], paired, event-aware tests show that the two tasks require different sens-

ing levels. Classification reaches stable performance with three to four meters, whereas localization keeps benefiting from additional meters up to the seven-meter economic budget and beyond.

2. **Low-cost unsynchronized metering** (addresses the first limitation). A metering framework is developed that uses only per-phase magnitudes and locally referenced sequence components, eliminating the need for a GPS time base[16]. Single-ended unsynchronized measurement is classical for impedance-based location[17]; the contribution here is to quantify what such inexpensive sensing delivers inside a sparse learning pipeline.

3. **A complete diagnosis-to-reliability framework** (addresses the third limitation). Localization accuracy is linked to isolation, restoration, IEEE Std 1366 indices[2] and the value of lost load (VOLL)[18], providing both technical and economic evaluation.

4. **Metric-aware model selection.** Four learners are evaluated on both tasks under paired tests, and in both the nominal leader is reported but not deployed. Selection considers not only prediction accuracy but also robustness to realistic instrument error and the resulting reduction in customer interruption duration.

5. **Leakage-aware evaluation.** Complete fault events, and in the strictest case entire buses, are held out instead of individual samples. The protocol quantifies the impact of leakage rather than asserting it, and shows that leakage affects localization far more than classification.

The scope is a single benchmark feeder: the study is offered as a methodology demonstration, and the transferable claims are the procedure and the cross-task difference rather than the specific meter count. Table 1 positions the work against the studies discussed above.

Results

Evaluation protocol and the cost of leakage

All results refer to the IEEE 33-bus feeder of Fig. 1, where k denotes the number of installed meters. The samples are grouped into events, each defined by the faulted bus, the fault resistance and the loading. An event holds on average 12.2 samples: the ten shunt fault variants simulated at that operating point and, in the lowest-resistance groups, the open-conductor events of the same bus and loading (three open phases with four measurement-noise replicas each). Because every sample carries its own measurement noise, no two are identical, but a random split places different fault types from the same operating point on both sides of the split. The size of this leakage is measured here instead of being assumed (Supplementary Table S3). For classification it is small. For the deployed random forest (RF) at the classification placement, a random split over samples and a 10-fold cross-validation return a macro-F_1 higher than the event-held-out protocol by only 0.27 and 0.33 percentage points (pp), within the seed-to-seed spread of ± 0.41 pp.

Localization is affected much more. At the seven-meter operating point, a random split raises the exact section accuracy from 80.0% to 90.7% ($+10.7$ pp, paired $p = 0.0002$) and the ± 1-hop accuracy from 97.3% to 98.8%, and it lowers the distance mean absolute error (MAE) from 0.32 to 0.26 hop. The fault types simulated at one operating point share its location signature but not its class label. When the other fault types of the same point fall in the training set, they reveal the location of a test sample but not its fault type, which is why leakage inflates localization far more than classification. Holding out whole buses

Table 1. Positioning against representative prior work. Dashes denote aspects the study does not address.

Study (year)	Measurement	Placement objective	Tasks covered	Evaluation	Reliab. + econ.
Mirshekali *et al.* (2023)[4]	Feeder waveforms	Not addressed	Location	Conventional split	No
Jiang (2020)[5]	Feeder meas.	Not addressed	Location	Conventional split	No
Mousavi *et al.* (2024)[6]	Bus meas.	Not addressed	Location	Conventional split	No
Bhagwat *et al.* (2024)[7]	Bus meas.	Not addressed	Detection	Conventional split	No
Krysander & Frisk (2008)[8]	Sensors	Detectability, isolability	Placement only	Analytical	No
Marín-Quintero *et al.* (2024)[10]	Fault indicators	Maximize ANN locator accuracy	Location	Conventional split	Unserved power only
Orozco-Henao *et al.* (2026)[11]	IEDs + fault indicators	ML error + s.d. of unserved power	Location (zones)	Conventional split	Unserved power only
Shahsavari *et al.* (2014)[14]; Farajollahi *et al.* (2019)[15]	Fault indicators (+ switches)	Interruption cost vs. device cost	Location (indicators)	Analytical optimization	Yes
Wang & Wang (2015)[13]; Wang *et al.* (2017)[12]	—	Not addressed	Isolation + restoration	Scenario	Partial
This work	**Unsync. feeder meters**	**Objective-driven, vs. cover/filter/random**	**Class. + loc. + isol. + restor.**	**Event- and location-held-out, multi-seed**	**IEEE 1366 + VOLL**

is a stricter test, of spatial extrapolation and not of leakage, and it raises the distance MAE to 0.62 hop. Event-held-out evaluation is used throughout.

Placement and the classification/localization difference

The central result is that the two FLISR tasks need different numbers of meters (Fig. 2). Classification approaches its ceiling almost immediately. In the meter-count ablation, the RF reaches 98.76% macro-F_1 at the three-meter knee, within 0.3 pp of its five-meter peak; Table 2 gives $98.71 \pm 0.41\%$ for a separate five-seed run. A paired test still separates the two ($p = 0.036$), but from four meters onward the difference from the peak is no longer significant ($p = 0.265$). Localization behaves differently. Its exact section accuracy rises from 64.0% to 92.9% between 2 and 16 meters. Under an event-aware, multi-seed paired test, objective-driven placement brings a large gain for localization ($+13.3$ pp macro-F_1, $p = 1 \times 10^{-5}$) and a negligible one for classification ($+0.06$ pp, $p = 0.53$).

The greedy wrapper of equation (3) also beats the three placement baselines at all three budgets tested, four, seven and eight meters (Supplementary Fig. S1). At the seven-meter operating point it reaches 80.1% exact localization accuracy, against 75.1% for random placement ($+5.0$ pp, $p = 0.006$), 74.0% for coverage ($+6.1$ pp, $p = 0.001$) and 72.2% for the ReliefF filter ($+8.0$ pp, $p = 3 \times 10^{-5}$). The margin is widest at four meters, where the filter reaches only 56.5% ($+17.5$ pp), and all nine comparisons remain significant after a Holm correction. Placement therefore matters most when the budget is tight. The margins are computed before rounding.

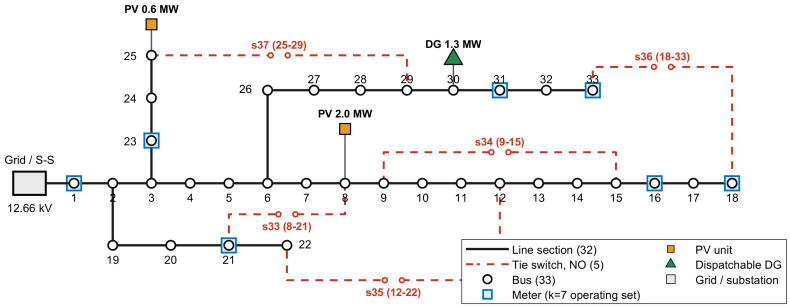


Figure 1. IEEE 33-bus test feeder used throughout. Solid lines are the 32 line sections; dashed red lines are the five normally open tie switches available for restoration. The DG portfolio comprises a 2.0 MW photovoltaic unit at bus 8, a 1.3 MW dispatchable unit at bus 30, and a 0.6 MW photovoltaic unit at bus 25. Blue squares mark the seven meters of the cost-optimal placement at buses 1, 16, 18, 21, 23, 31 and 33.

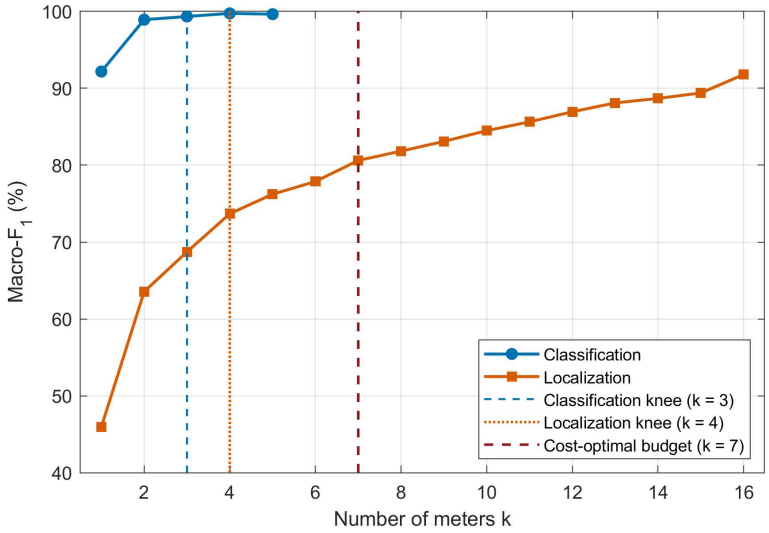


Figure 2. Objective-driven placement: event-split macro-F_1 against meter count for the two diagnostic tasks. The blue dashed line marks the knee of the classification curve at three meters, and from four meters onward classification is statistically at its peak ($p = 0.265$). The orange dotted line marks the knee of the localization curve at four meters, where its gains begin to diminish, although it keeps improving with density up to sixteen meters. The red dashed line marks the cost-optimal budget of seven meters (Fig. 3): past the knee each added meter brings less accuracy, but up to seven meters it still avoids more outage cost than it adds.

Classification: the clean-versus-robust trade-off

Table 2 reports the event-held-out classification results. On clean data, the gradient-boosting model (GBM) leads with $99.86 \pm 0.10\%$ macro-F_1, ahead of the RF at $98.71 \pm 0.41\%$ (McNemar $\chi^2 = 80.1$, $p = 4 \times 10^{-19}$). The support-vector machine (SVM) reaches $87.99 \pm 0.80\%$ and the artificial neural network (ANN) $85.24 \pm 1.02\%$. In the RF confusion matrix (Supplementary Fig. S2), errors occur only for ground faults. About 3% of single line-to-ground (SLG) faults are read as open-conductor (OC) events and 1% of double line-to-ground (DLG) faults as SLG or OC, while every other class is recognized without error. As a result, the worst single-class F_1 of both tree ensembles stays above 0.95. The strength of

Table 2. Event-split classification: clean accuracy, robustness, and the effect of noise-augmented training. Bold marks the best value in each column. The asterisk denotes the deployed model, in the noise-augmented mode, for both stages. The RF leads only the clean-trained 40 dB column here; the GBM leads the clean metrics and the augmented 40, 30 and 20 dB columns, and the SVM leads the clean-trained low-SNR columns and the augmented 10 dB column. The deployment argument rests not on this table but on Supplementary Table S4, where the augmented RF leads under every instrument error while giving up almost no clean accuracy to augmentation. κ and Min F_1 (worst single class) are five-seed event-split means; the GBM augmented row is averaged over ten noise realizations.

				Clean-trained accuracy (%)				**Noise-augmented accuracy (%)**			
Model	**Macro-F_1**	κ	**Min F_1**	40 dB	30 dB	20 dB	10 dB	40 dB	30 dB	20 dB	10 dB
ANN	85.24 ± 1.02	0.875	0.647	89.2	87.2	80.6	61.7	86.3	86.0	83.4	66.1
RF*	98.71 ± 0.41	0.985	0.953	**91.9**	82.9	62.8	45.2	93.0	88.5	82.7	61.0
SVM	87.99 ± 0.80	0.891	0.726	90.7	**88.5**	**81.1**	**62.3**	86.8	86.6	84.3	**66.2**
GBM	$\mathbf{99.86 \pm 0.10}$	**0.999**	**0.996**	87.4	70.7	53.6	40.2	**93.3**	**90.5**	**84.7**	65.2

tree ensembles on tabular electrical features agrees with broader evidence[19].

Noise reverses the ranking. At a signal-to-noise ratio (SNR) of 10 dB, the SVM retains 62.3% and the ANN 61.7%, while the RF falls to 45.2%. The GBM, which leads on clean data, collapses to 40.2%, a fragility to input noise that resembles the known sensitivity of boosting to noisy training labels[20].

This weakness can be corrected. Training on the clean set plus one additive white Gaussian noise (AWGN) replica acts as a regularizer, related to Tikhonov regularization for smooth models[21], and raises the 10 dB accuracy of the RF from 45.2% to 61.0% and that of the GBM from 40.2% to 65.2%. With augmentation the GBM leads the RF at every SNR (Table 2), which would favor it if white noise were the only concern. Instrument errors give a different picture. Under these errors the augmented RF is at least as accurate as the augmented GBM in every condition measured, clearly so with a fixed $\pm 1\%$ ratio error added to the base error (96.3% against 94.4%, 2.2σ) and within seed variation elsewhere (0.6–0.9 pp). The decisive factor is the cost of augmentation. It lowers the clean accuracy of the RF by 0.4 pp ($99.5 \rightarrow 99.1\%$) but that of the GBM by about 3 pp ($99.9 \rightarrow 96.7\%$, the clean reference taken from the classification study). **The RF is therefore deployed for classification**, with augmentation applied. This is an engineering preference, not a claim of statistical superiority. The RF is never worse under any measured instrument error, keeps its clean accuracy under augmentation and is easier to interpret. Under white noise the ranking is reversed, and Table 2 gives the figures needed to make the other choice.

Additive white noise is the mildest error a field device suffers. It has zero mean and is independent from sample to sample, so an ensemble averages much of it away. Three errors do not average away: a fixed ratio error in the instrument transformers, a fixed offset, and the finite word length of the analog-to-digital converter. Supplementary Table S4 reports the accuracy under each error, added to the base measurement noise and drawn per channel and averaged over ten draws. The classifier tolerates these errors far better than low-SNR noise. It retains 96.1% with a fixed $\pm 1\%$ ratio error, the limit of a class-1 instrument transformer, added to the random base error, and 99.0% when each channel is rounded to 12 bits over a fixed full scale. Accuracy drops noticeably only at a 5% ratio error or with an 8-bit converter, which correspond to aging or poorly specified equipment. With all errors combined, accuracy falls to 80.8% for clean training and recovers to 90.8% with the same augmentation. In any case, a ratio error is a calibration fault, and calibration corrects it.

Fault location, isolation, and the metering optimum

In the placement comparison, at the cost-optimal placement of $k = 7$ meters, the section classifier reaches $80.0 \pm 1.8\%$ exact accuracy, $97.3 \pm 0.4\%$ ± 1-hop accuracy (the predicted section is the faulted one or shares a bus with it) and a distance MAE of 0.32 hop. At the three-meter classification placement they fall to 60.6%, 75.2% and 0.99 hop. Holding out whole buses raises the location-held-out MAE to 0.62 hop. Three independent five-seed runs of this model at $k = 7$ give ± 1-hop values of 97.3, 97.4 and 97.6% (Supplementary Table S5), a spread within seed variation; the exact accuracy of the three runs agrees to within 0.4 pp (Supplementary Table S5). The confusion matrix (Supplementary Fig. S3) shows a strong diagonal, and most of the remaining errors fall on adjacent sections.

Accuracy depends strongly on the fault resistance (Supplementary Table S7). The shunt faults are simulated at twelve resistances spaced logarithmically from 0.001 to $100\ \Omega$, so seven of the twelve lie at or below $0.53\ \Omega$. For these near-bolted faults the exact accuracy is 97.4% and the ± 1-hop accuracy 99.98%. Between 1.5 and $12.3\ \Omega$ they fall to 73.7% and 97.6%, and at 35 and $100\ \Omega$ to 54.7% and 93.5%. Open-conductor events, which have no fault resistance, are the hardest case (34.1% exact, 87.3% ± 1-hop), followed by three-phase faults (74.2% and 91.9%), which produce no negative- or zero-sequence components. The pooled accuracy therefore depends on the resistance distribution of the sweep. If every shunt fault had a resistance of 35–$100\ \Omega$, with the open-conductor events unchanged, the ± 1-hop accuracy that drives the reliability model would fall to 93.0% and the SAIDI reduction to 78.5%.

The learner for this stage is chosen by measurement, not carried over from the classification result, and the two tasks do not give the same answer. Table 3 compares the four learners at the cost-optimal placement, together with a non-learning nearest-centroid baseline that assigns each test signature to the closest section mean in the same feature space. Two observations follow. First, the accuracy comes from learning. The centroid rule reaches only 39.7% exact and 64.8% ± 1-hop accuracy, so the features alone do not explain the results. Second, the learners rank differently depending on the accuracy measure. The ANN names the exact section more often (86.2% against 80.4%), but the RF keeps its errors closer: 88% of RF errors fall on an adjacent section, against 62% for the ANN, so its ± 1-hop accuracy is higher (97.6% against 94.8%). The axis-aligned splits of the RF follow the monotone impedance ordering along the feeder, whereas the smoother decision surface of the network spreads its mistakes further along the graph.

This difference decides the deployment. Only ± 1-hop success, with probability α, enters equation (6). Substituting each learner's α into equation (9) therefore converts accuracy into the system average interruption duration index (SAIDI), the quantity by which the system is judged. The RF gives a SAIDI of 0.510 h against 0.578 h for the ANN, a difference of 0.069 h per customer-year (82.4% against 80.1% reduction), worth about \$2.3k/yr of avoided outage cost at the assumed VOLL. The RF is therefore deployed although it is not the most precise locator, and it also trains about six times faster than the ANN. If the objective were precise crew routing instead of fast restoration, the ANN would be the right choice.

Because isolation needs the fault zone only to within one hop, dense metering is not required. Combining meter cost with VOLL-weighted outage cost through equation (8) gives a total-cost curve whose minimum lies inside the search range, at seven meters (Fig. 3). Every budget is evaluated on the same five seeds and splits, so the costs can be compared in pairs. With fourteen comparisons and five seeds, these tests need a correction for multiplicity. After a Holm correction, seven meters remains separated from every budget except eight, which costs \$133/yr more ($p = 0.084$); the stricter Bonferroni correction also

Table 3. Localization at the cost-optimal placement ($k = 7$, five seeds). The asterisk denotes the deployed model. SAIDI is obtained by substituting each learner's measured α into equation (9), so the RF entry (0.510 h) is marginally below the 0.515 h reported below, where the chain is driven by the $\alpha = 97.4\%$ of the meter-count sweep rather than by this re-run (97.6%); the comparison between rows is unaffected. The ANN is the more precise locator, whereas the RF keeps more of its errors adjacent, which is what the reliability model rewards.

Method	Exact (%)	±1-hop (%)	SAIDI (h)	Train (s)
RF*	80.4 ± 1.2	$\mathbf{97.6 \pm 0.6}$	**0.510**	**28**
ANN	$\mathbf{86.2 \pm 0.5}$	94.8 ± 0.3	0.578	171
GBM	76.4 ± 1.2	92.9 ± 1.5	0.625	394
SVM	62.1 ± 1.4	83.1 ± 0.8	0.865	59
Nearest centroid	39.7 ± 0.8	64.8 ± 1.7	1.314	—

leaves four, six, nine and ten meters unresolved. The choice therefore rests on the one-standard-error rule, which selects the fewest meters whose mean cost lies within one standard error ($176/yr) of the minimum and returns seven. In money terms the differences are small. Every budget from four to sixteen meters costs at most 5.1% more than the minimum (at fourteen meters), whereas three meters cost 9.7% more and two meters 41.5% more. **The operating point is therefore seven meters** (21% of buses), with eight as an equivalent alternative, and it is used throughout. Supplementary Table S6 shows the trade-off. Only the ±1-hop criterion of equation (4) enters the reliability model, so the cost follows the ±1-hop curve and not the exact-accuracy curve. Raising the budget from seven to sixteen meters adds 12.8 exact points but only 1.9 ±1-hop points, while adding $2,400/yr in metering.

The choice depends only moderately on the economic assumptions (Supplementary Fig. S4). The minimum stays at seven meters for VOLL between $4.2 and $13.6/kWh, a range that includes the assumed $9/kWh. Below this range the minimum falls to four meters, and to three below $2.0/kWh. Above it, the minimum rises to sixteen. A meter price between $2,000 and $8,000 moves it between seven and sixteen meters. These shifts are expected, because the budgets they reorder differ in cost by only a few percent.

The end-to-end economics follow. Seven meters cost $28,000, and with an indicative $30,000 for automated sectionalizing and tie control, the investment is $58,000. The avoided outage cost is $80.0k/yr, from the reduction in energy not supplied (ENS) valued at $9/kWh. This gives a simple payback of 0.73 years and a net present value of about $0.63 M at an 8% discount rate over 15 years. The undiscounted 15-year benefit is $1.20 M, about 21 times the investment. Doubling the meter price keeps the payback near one year.

Self-healing and reliability value

Of the load interrupted by a fault, 86.2% can be restored before repair. Of the restored load, 84% returns through reconfiguration and 16% through DG islanding, and the rest waits for the repair crew. The rest is essentially the faulted section itself. Its share of the interrupted load is $(1/32)/(\mathrm{SAIFI}/\sum_s \lambda_s) = 13.8\%$, where SAIFI is the system average interruption frequency index, and this agrees with the 13.7% implied by the measured indices. Combining this restorable fraction β with the ±1-hop isolation success $\alpha = 0.974$ gives the effective fast-restoration fraction $f = \alpha\beta = 0.840$ used in equation (6). For this reason the reduction settles near 82% and not at the 95% that full restorability would give.

In IEEE Std 1366 terms (Table 4), SAIDI falls from 2.903 to 0.515 h per customer per year, ENS from $10{,}808$ to $1{,}919$ kWh/yr, and the customer average interruption duration index (CAIDI) from 4.00 to 0.710 h. Each is a reduction of 82.2%, and the average service availability index rises from 0.99967

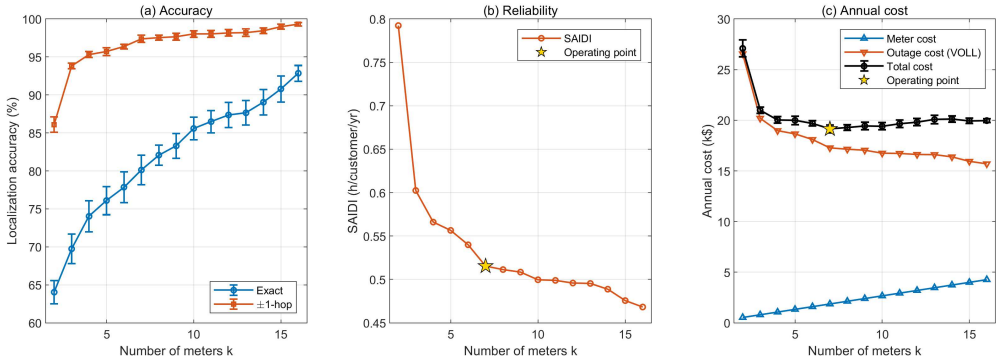


Figure 3. Meter-count sensitivity and cost-benefit. (**a**) Localization accuracy against meter count; (**b**) the resulting SAIDI; (**c**) total annual cost, comprising amortized meter cost plus VOLL-weighted outage cost, with seed-to-seed error bars. After a Holm correction only eight meters cannot be separated from the minimum at seven; every budget from four to sixteen meters costs at most 5.1% more, whereas two and three meters cost 41.5% and 9.7% more.

Table 4. Reliability indices before and after FLISR, computed under IEEE Std 1366 at the seven-meter placement. SAIFI is unchanged because customers restored by switching are still counted as interrupted (Methods); the three duration-based indices therefore move together by construction.

Index	Before	After	Improvement
SAIFI (int./yr)	0.726	0.726	0.0%
SAIDI (h/cust/yr)	2.903	0.515	82.2%
CAIDI (h/int.)	4.000	0.710	82.2%
ASAI	0.99967	0.99994	+0.027 pp
ENS (kWh/yr)	10,808	1,919	82.2%

to 0.99994. SAIFI is unchanged at 0.726, because customers restored by switching are still counted as interrupted under the convention described in Methods. The identical 82.2% is an algebraic identity: with SAIFI fixed, CAIDI scales with SAIDI, and ENS is proportional to duration under the per-section load model. The reliability chain uses the $\alpha = 97.4\%$ of the meter-count sweep at the operating point. Under $N-1$ tie contingencies, DG adds margin. In the worst case, the 18–33 tie, the share of interrupted load that is restored rises from 81% with reconfiguration alone to 84% with DG. Without any reconfiguration, DG islanding alone would supply on average 45% of the interrupted load.

Mechanism: data-driven isolability and a closed-form design relation

Two analyses explain the localization results. First, each fault section is treated as a class. The distinguishability of two sections is measured as the Mahalanobis distance D between their signature centroids at the seven-meter set, with a single covariance pooled across sections. D rises with the graph-hop distance between the sections (Fig. 4b). Its mean increases at every one of the 19 hop distances, from 3.59 for adjacent sections to 11.7 for the most distant pairs, although the largest distances rest on few pairs. The correlation over all 496 section pairs is 0.62. Adjacent sections are therefore the hardest to tell apart with sparse, unsynchronized meters, which is why most localization errors fall on an adjacent section. This is the data-driven counterpart of fault isolability in model-based diagnosis[8]. Evaluated against meter count (Fig. 4a), the median section pair keeps separating as meters are added, from 1.7 at two meters to 9.3 at sixteen, in step with the localization accuracy. The class pairs grow much less, and along the three classification meters the worst pair reaches only $D = 0.74$. The classifiers are nevertheless close to 99% accurate at three meters, because they exploit nonlinear differences between fault types that a centroid

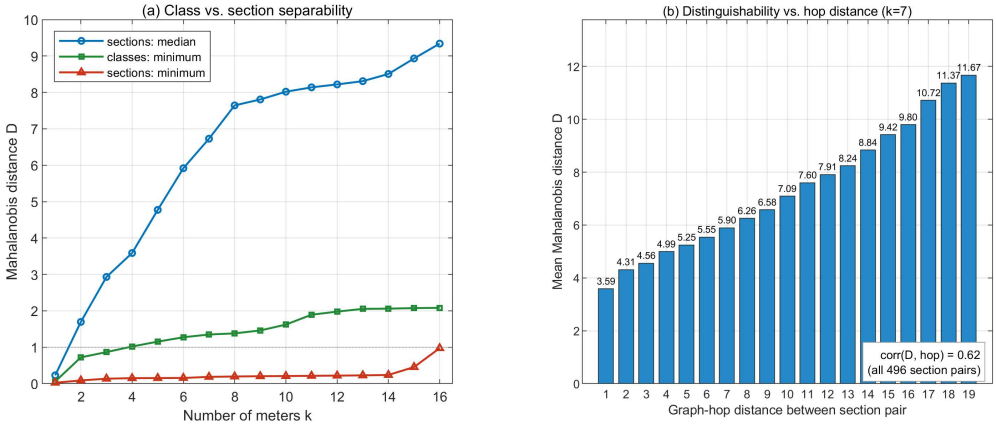


Figure 4. Data-driven isolability, measured as the Mahalanobis distance D between group centroids under a covariance pooled within groups. (**a**) D against meter count along the localization selection path: the worst class pair saturates near $D \approx 2$, the median section pair keeps separating with density, and the worst section pairs stay low until $k \geq 15$; the dotted line marks $D = 1$, a centroid gap of about one pooled standard deviation. (**b**) At $k = 7$, the mean D between section centroids rises at every hop distance; hop distances 17–19 rest on 7, 4 and 2 section pairs respectively. The most confusable pairs are therefore adjacent, which is why residual errors remain within one hop.

distance does not capture. D therefore explains why localization keeps benefiting from density, while the saturation of classification is shown by its accuracy curve.

Second, the reliability model has a closed form. Substituting equation (6) into equation (7) makes the simulated SAIDI an affine function of the ± 1-hop accuracy $\alpha(k)$, given in equation (9). The relation is derived, not fitted, so a goodness-of-fit statistic has no meaning here. The expected restoration time is linear in α by construction, and equation (9) only states the resulting slope and intercept, which the simulation reproduces to machine precision. Its value is practical. Once the $\alpha(k)$ curve has been measured, the SAIDI of any budget follows without further simulation, and the metering optimum follows from the balance against meter cost.

Discussion

This study evaluated an end-to-end machine-learning framework for fault classification, localization, isolation and self-healing with a few low-cost, unsynchronized meters. Its main finding is that the two diagnostic tasks place different demands on the sensing layer. Classification reaches its statistical peak at three to four meters and depends little on where the meters are placed, whereas localization keeps improving with density and its economic budget is seven meters. Running the placement search separately for each task exploits this difference and outperforms coverage-, filter- and random-based placement at all three budgets tested, with the widest margin at the tightest budget.

Model selection follows the same task-dependent logic. Gradient boosting is the most accurate learner on clean data and under white noise. Nevertheless, the random forest is deployed for both stages, because it keeps its clean accuracy under noise augmentation and keeps more of its localization errors within one hop. The second property is the one that matters in operation. A locator is judged by the interruption time its errors cause, not by how often it names the exact section, and by that measure the less precise learner is better by 0.069 h per customer-year. This is a general caution for applied work in this area, where nominal accuracy is the usual selection criterion.

Linking the diagnosis to a standards-based outcome turns these observations into a design rule. The lifetime cost is lowest at seven meters (21% bus coverage), and after a Holm correction only eight meters cannot be separated from it. At this point FLISR cuts SAIDI by 82.2%, with a payback of 0.73 years. Every budget from four to sixteen meters stays within 5.1% of the minimum cost, so the exact choice in this range depends on the assumed prices and on the measurement-error model (Methods), whereas sparser metering is clearly more expensive. The isolability measure shows why localization keeps improving with density and why most of its errors fall on adjacent sections, and the closed-form relation gives the reliability of any metering budget without new simulations.

Several limitations bound these claims. The study is a methodology demonstration on a single benchmark feeder in simulation. The transferable results are the cross-task difference, the placement procedure and the closed-form relation; the meter count, costs and reliability gains are specific to this feeder, DG portfolio and VOLL. Localization errors beyond ± 1 hop and variable switching delay are simplified. The five-minute switching time assumes prompt communication and control, and real latency would lengthen restoration. Diagnosis is also evaluated only on the nominal radial configuration. This matches the operating sequence, because a fault is diagnosed in the configuration that exists when it occurs and reconfiguration comes after the diagnosis. However, a feeder left in a post-restoration or loss-minimizing configuration would present the learner with a topology it was not trained on. Topology-adaptive localization is a separate problem that this paper does not address.

Three modeling choices should also be read as limits of scope. The failure rate is applied uniformly as $\lambda = 0.1$ faults/yr to every section regardless of its length, whereas utilities usually express it per kilometer. Because the 33-bus sections differ in length, a length-weighted λ would change which sections dominate SAIDI and could change the buses the wrapper selects. In addition, the fault-resistance sweep stops at $100\ \Omega$. High-impedance faults, which reach hundreds to thousands of ohms, produce the weakest signatures and are the hardest case for any magnitude-based scheme. They are outside the scope of this study, and the accuracies reported here do not cover them. Within the sweep, accuracy already declines as the resistance grows (Supplementary Table S7), so it would be lower still for high-impedance faults. The base measurement-error model also leaves the per-unit voltage channels nearly uninformative (Methods); a per-channel quantizer would likely raise accuracy at every budget and could move the cost-optimal count. Finally, graph-based and deep localizers may score higher and are left to future work.

Future work will extend the framework to larger and meshed feeders, evaluate graph-based and deep localizers under the same leakage-aware protocol, and replace the fixed switching delay with a model of communication latency. This would also prepare the way for hardware-in-the-loop validation.

Methods

Benchmark feeder and DG portfolio

The test system is the 12.66 kV IEEE 33-bus radial distribution feeder[22], which consists of 33 buses, 32 line sections, and five normally open tie switches, with a total load of $3{,}715$ kW $+ j2{,}300$ kvar. It is modeled in OpenDSS (version 11.0) and interfaced with MATLAB (R2019b), and the substation is located at bus 1 (Fig. 1). The locations of the distributed generation units are determined using a loss-sensitivity analysis followed by power-flow verification[23, 24]. The final configuration includes a 2.0 MW photovoltaic (PV) unit at bus 8, a 1.3 MW dispatchable DG unit at bus 30, and a 0.6 MW PV unit at bus 25.

This configuration maintains all bus voltages within $\pm 5\%$ while reducing the total active power loss by 72.6% ($202.7 \rightarrow 55.5$ kW). In addition, the dispatchable unit provides grid-forming capability to support islanded operation during restoration[25].

Unsynchronized measurement model

Faults affect the balance among the three phases. Therefore, the measured phase quantities are resolved into symmetrical components using the Fortescue transformation, which separates that imbalance into interpretable parts. Each meter records the per-phase voltage and current magnitudes together with the sequence components obtained from locally referenced phasors,

$$\begin{bmatrix} V_0 \\ V_1 \\ V_2 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & a & a^2 \\ 1 & a^2 & a \end{bmatrix} \begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix}, \qquad a = e^{j2\pi/3}, \tag{1}$$

where (V_a, V_b, V_c) in equation (1) are measured against a *local* reference; no GPS time base or wide-area phasor telemetry[16] is required. For a meter set S the feature vector $\mathbf{x}(S)$ concatenates these magnitudes at the metered buses. Single-ended unsynchronized location is classical for impedance-based methods[17]; the contribution here is to use such measurements inside a sparse learning pipeline that separates the FLISR sub-tasks and prices the result. Every sample carries the steady-state measurement-error model applied at generation. The raw features are per-unit voltage magnitudes, line currents in amperes and sequence-voltage magnitudes in volts. Each channel receives a random gain error with a standard deviation of 0.5%, an RMS inflation for 2.5% harmonic distortion, additive Gaussian noise with a standard deviation of 0.01 in raw units, and a uniform error whose width is $1/4096$ of the sample's raw feature range. That range is never smaller than the positive-sequence voltage of about 7.3 kV, so the last term has a half-width of at least 0.89 and a standard deviation of at least 0.52 in raw units. This is negligible for the voltage channels in volts and moderate for small currents, but comparable to the per-unit magnitudes themselves (about 1.0). These channels therefore carry little information, and the voltage information is carried mainly by the sequence voltages. A field meter would quantize each channel over its own full scale. The reported accuracies are therefore conservative, whereas the cost-optimal meter count is conditional on this error model as well as on the prices. Results labeled clean refer to this base condition. The robustness tests add white noise at a given signal-to-noise ratio (SNR), or the instrument errors of Supplementary Table S4, on top of it; the quantization test there rounds each channel over its range across the dataset, as a converter with a fixed full scale would. Series open-conductor events are included as a sixth class.

Dataset and leakage-aware protocol

Six classes are simulated: normal operation and single line-to-ground (SLG), line-to-line (LL), double line-to-ground (DLG), three-phase (3LG) and open-conductor (OC) faults. Shunt faults are applied to all 32 sections at twelve fault resistances spaced logarithmically from 0.001 to $100\ \Omega$ (ratio 2.85) and five loading levels (0.70–1.30 p.u. in steps of 0.15), with one sample for each combination and phase variant. Open-conductor events carry no fault resistance and are recorded with $R_f = 0$. This process generates $22{,}110$ samples (Supplementary Table S1) described by 294 features. Each bus contributes three phase-voltage magnitudes and three sequence magnitudes, plus three current magnitudes for every incident line,

so a bus of degree d carries $6 + 3d$ features. This gives nine features at the substation and the four feeder ends, twelve at ordinary buses and fifteen at the three lateral branch points. Because the current of a line is shared by the two buses it joins, the per-bus assignments reduce to $6 \times 33 + 3 \times 32 = 294$ distinct columns, so the total is not a multiple of the bus count.

Because the fault types simulated at one operating point share its location signature, a random sample-level split can cause data leakage. Therefore, an *event* is defined as a (bus, fault-resistance, loading) group. The $2{,}080$ swept fault operating points form $1{,}728$ groups and, with the 33 normal-operation events, the splits hold out $1{,}761$ groups (Supplementary Table S1 details the merging). Merging is never looser than splitting, since it can only place more samples on the same side. Whole event groups are then assigned to disjoint training and test sets,

$$e = (b, R_f, \ell), \quad \mathcal{D}_{\text{test}} = \{\mathbf{x} : \text{event}(\mathbf{x}) \in \mathcal{E}_{\text{test}}\}, \quad \mathcal{E}_{\text{train}} \cap \mathcal{E}_{\text{test}} = \varnothing. \tag{2}$$

Alongside equation (2), a *location-held-out* test with whole buses unseen measures spatial extrapolation[26, 27]. All runs use five seeds (reported mean $\pm$ standard deviation). Model pairs are compared with McNemar's test on the pooled held-out predictions[28] and, across seeds, with a paired t-test. All tests are two-sided at a significance level of $\alpha = 0.05$. Paired t-tests are used across the five seeds because a Wilcoxon signed-rank test cannot return a p-value below 0.0625 with five pairs, and a Holm correction is applied wherever several budgets or methods are compared. Rank-based procedures for comparing classifiers over multiple data sets[29] are not applicable here, since the replicates are seeds on a single data set rather than independent data sets. Feature standardization is fitted on the *training partition* of each split; this is immaterial for the deployed tree ensembles, which are invariant to monotone feature scaling, and negligible for the distance-based learners at this sample size. The synthetic minority over-sampling technique (SMOTE)[30] is applied to the training partition only[31].

Objective-driven sparse placement

Meter placement is formulated as a feature-selection problem, because each instrumented bus provides a corresponding block of features to the model. A greedy forward wrapper[32] selects the meter locations. At each step it adds the bus that gives the largest improvement in the validation macro-F_1 of the corresponding model, and the Kneedle criterion[33] then locates the knee of the resulting curve,

$$S_t^{(m+1)} = S_t^{(m)} \cup \underset{b \notin S_t^{(m)}}{\arg\max}\, F1_{\text{val}}\big(h_t(S_t^{(m)} \cup \{b\})\big). \tag{3}$$

The wrapper is run separately for the two tasks $t \in \{\text{cls}, \text{loc}\}$, in both cases with a random forest as the inner model h_t. The inner model is also the deployed one for both stages, so the placement is selected and used under the same learner and no transfer assumption is required. Baselines are greedy line set-cover (coverage), the ReliefF filter[34, 35] and random placement; the substation is mandatory. The knee sets the classification budget. For localization, whose accuracy keeps rising with density, the budget is set instead by the cost minimization of equation (8). The classification path is $\{1, 24, 22\}$ ($k = 3$); the location path begins $\{1, 33, 18, 21, 31, 16, 23\}$ at the deployed budget ($k = 7$), extending to 16 buses.

Learning models and training setup

Four machine-learning models are evaluated on identical splits: a bagged tree ensemble, a boosted tree ensemble, a kernel-based model, and a feed-forward neural network. The random forest (RF)[36] samples $\lfloor\sqrt{p}\rfloor$ predictors per split, where p is the number of features at the placement in use (five at the three-meter classification set). The gradient-boosting model (GBM) is AdaBoost.M2[37] over depth-limited trees. The support-vector machine (SVM)[38, 39] uses a radial-basis-function kernel in an error-correcting output-code arrangement, and the neural network (ANN)[40] is a multilayer perceptron. The hyperparameters listed in Supplementary Table S2 are selected using random search followed by Bayesian optimization[41, 42], both performed on the training partition only. Tree learners absorb class imbalance natively, whereas SVM and ANN receive SMOTE[30, 31, 43] applied to the training partition alone. The 32 localization sections are balanced overall (660 samples each) but become mildly uneven once whole events are held out, so the localization training partition is SMOTE-balanced too. Reported metrics are macro-averaged and per-class F_1, Cohen's κ[44, 45] and the worst single-class F_1. Permutation importance from the RF[36] supports interpretation and independently corroborates the placement.

Fault location

The section classifier identifies which of the 32 sections is faulted. It reuses the RF tuned above without further tuning (100 trees, minimum leaf size 3), with $\lfloor\sqrt{p}\rfloor$ predictors sampled per split for the placement under test. Reusing the same hyperparameters ensures that the meter-count sweep is not confounded by simultaneous changes in placement and model configuration. The choice of learner for this stage is settled empirically in the Results rather than assumed from the classification result. Its training partition is SMOTE-balanced as described above.

Two accuracy notions are needed because the operational requirement is weaker than the nominal one. An *exact* hit is the correct section index, whereas a ±1-*hop* hit counts the true section or a graph-adjacent neighbor. The latter is the operationally relevant tolerance, since isolation opens the sectionalizers bounding the suspected zone[3], and it is defined as

$$\mathrm{Acc}_{\pm1} = \frac{1}{n}\sum_{i} \mathbb{1}\big[\hat{s}_i \in \{s_i\} \cup \mathcal{N}(s_i)\big], \tag{4}$$

where n is the number of test samples, s_i and $\hat{s}_i$ are the true and predicted sections of sample i, and $\mathcal{N}(s_i)$ are the sections sharing a bus with s_i. A regressor estimates fault distance in hops, with a location-held-out MAE quantifying spatial extrapolation[26, 27].

Isolation and restoration

Isolation starts at the predicted section. Its bounding sectionalizers are opened and the feeder is re-energized. If the prediction is exact, the fault is isolated. If the prediction is a neighbor of the faulted section, re-energization shows on which side the fault lies: a fault upstream trips the source breaker again, and a fault downstream trips when that part of the feeder is restored through a tie. The zone is then extended by the adjacent section on that side. At a branch point more than one adjacent section can lie on the same side, and a second attempt may be needed. A ±1-hop hit is therefore resolved with one or two extra switching steps, whereas a prediction two or more sections away is treated as a failed isolation that costs

the full repair time. Load is restored by radiality-preserving tie reconfiguration under voltage and thermal limits[3, 46, 47, 48], with DG islanding[12, 25] for the remainder. Two islanding modes are evaluated, grid-forming and dispatchable-only, and $N-1$ tie contingencies value DG as a resilience resource.

Reliability and economics

The diagnostic result is next expressed in the language a utility actually reports. Since the benchmark carries no customer data, the customer count at load point i is set proportional to its connected load, $N_i = \max(1, \mathrm{round}(P_i/10\,\mathrm{kW}))$, which gives 372 customers on the 32 load buses. The rule also assigns one customer to the substation bus, which carries no load and is never interrupted, so it enters only the denominators of the indices and lowers them by $1/373$ (0.3%); the total is 373. The indices are therefore load-weighted and are not drawn from a metered census. The relevant indices of IEEE Std 1366[1, 2] are the system average interruption frequency and duration indices (SAIFI, SAIDI), the customer average interruption duration index (CAIDI) and the energy not supplied (ENS). Between them they capture how often supply is lost, for how long, and how much energy goes unserved. With a section failure rate $\lambda_s = 0.1$ faults/yr for every line, a manual repair time $t_{\mathrm{rep}} = 4$ h and a switching time $t_{\mathrm{sw}} = 5$ min, these are

$$\mathrm{SAIFI} = \frac{\sum_i \lambda_i N_i}{\sum_i N_i}, \quad \mathrm{SAIDI} = \frac{\sum_i \lambda_i r_i N_i}{\sum_i N_i}, \quad \mathrm{CAIDI} = \frac{\mathrm{SAIDI}}{\mathrm{SAIFI}}, \quad \mathrm{ENS} = \sum_i \lambda_i r_i P_i. \tag{5}$$

where N_i and P_i are the customers and average load at load point i, λ_i is its interruption rate, that is the sum of the failure rates λ_s of the sections whose faults interrupt it, and r_i is its average outage duration. IEEE Std 1366 counts an interruption as sustained only when it lasts more than five minutes. With $t_{\mathrm{sw}} = 5$ min, customers restored by switching sit at this boundary. They are counted here as sustained interruptions of duration t_{sw}, a deliberately conservative convention that leaves SAIFI unchanged; counting them as momentary would lower SAIFI and SAIDI further. The re-energization used to resolve a neighbor hit can also trip the source breaker a second time, adding momentary events ($\mathrm{MAIFI_E}$) that these indices do not include. In equation (5) FLISR shortens the duration only where the fault is correctly isolated and the load is restorable. For load point i under a fault on section s, with α the ±1-hop isolation-success probability and $\beta_i(s) \in \{0, 1\}$ a restorability indicator (reconfiguration or DG island),

$$\mathbb{E}[r_i(s)] = \alpha\,\beta_i(s)\,t_{\mathrm{sw}} + \big[1 - \alpha\,\beta_i(s)\big]\,t_{\mathrm{rep}}. \tag{6}$$

Equating α with the classifier's ±1-hop accuracy assumes that a correct zone decision is executed as a successful isolation. Switch and telemetry reliability are folded into α and are not modeled separately, which makes the reported improvement an upper bound with respect to device failure. A second simplification concerns the isolation sequence described above. At $k = 7$, 17.3% of faults are located to a neighbor. For these faults the sequence isolates one healthy section with the faulted one and needs at least one extra switching step, and equation (6) charges neither. Both effects are bounded here, assuming customers are distributed uniformly over sections. In the optimistic case the over-isolated section is back-fed, so its customers see t_{sw}, which adds 0.0014 h to SAIDI, and one extra switching step delays the other restorable customers of these faults by t_{sw}, which adds 0.009 h; together they lower the improvement from 82.2% to 81.9%. In the pessimistic case the over-isolated customers wait for the repair and every neighbor hit needs two extra steps, which adds 0.087 h and lowers the improvement to 79.2%. The headline reduc-

tion is also insensitive to the assumed repair time. Writing $\mathrm{SAIDI} = \mathrm{SAIFI}\,[f\,t_{\mathrm{sw}} + (1-f)\,t_{\mathrm{rep}}]$ with $f = 0.840$ the effective fraction restored by switching, both the before and after indices scale with t_{rep}, so their ratio barely moves: the reduction is 80.5% at $t_{\mathrm{rep}} = 2$ h, 82.2% at 4 h and 82.8% at 6 h. The reported 82.2% therefore leaves out isolation effects worth at most 3.0 points, and the conclusion is unaffected in direction.

Aggregating this expected duration over every section that can fail, weighted by the customers each outage affects, gives the feeder-level index

$$\mathrm{SAIDI} = \frac{\sum_s \lambda_s \sum_i N_i\,\mathbb{E}[r_i(s)]}{\sum_i N_i}. \tag{7}$$

where λ_s is the failure rate of section s. Thus SAIFI is unchanged while SAIDI, CAIDI and ENS improve, and the improvement is bounded below 100% because non-restorable load ($\beta_i = 0$) and mislocated faults still incur the full repair time.

The metering investment trades capital against expected outage cost,

$$C(k) = k\,\frac{C_{\mathrm{meter}}}{L} + \mathrm{VOLL}\cdot\mathrm{ENS}(k), \qquad k^\star = \arg\min_k C(k), \tag{8}$$

with $C_{\mathrm{meter}} = \$4{,}000$, $L = 15$ yr and $\mathrm{VOLL} = \$9$/kWh[18]. Substituting equation (6) into equation (7) makes SAIDI affine in $\alpha(k)$,

$$\mathrm{SAIDI}(k) = -2.45\,\alpha(k) + 2.90, \tag{9}$$

exact by construction, so the metering optimum can be read off analytically instead of by repeated simulation.

Data-driven isolability

Treating each fault class or section as a group, the distinguishability of two groups is the Mahalanobis distance D between their signature centroids under a single covariance pooled within groups, with a ridge of $0.1\,\mathbf{I}$ added for numerical stability. This linear-discriminant convention ranks all pairs on one scale. It is evaluated for the 32 sections and the six classes at $k = 1, \ldots, 16$ meters along the localization selection path, and for the classes also along the classification selection path. It is also evaluated for all 496 section pairs at the seven-meter set, where each pair is assigned its graph-hop distance (sections sharing a bus are one hop apart). This provides a data-driven counterpart to fault isolability in model-based diagnosis[8].

Use of large language models

A large language model (Claude, Anthropic) was used for writing and language-editing assistance on the manuscript and as a programming assistant during development of the analysis scripts. All code was reviewed and executed by the authors; every reported result was produced by, or derived analytically from, the authors' own simulation runs; and the authors verified all content and take full responsibility for it. The model is not an author of this work.

Data availability

The IEEE 33-bus test-system data used in this study are publicly available from the original publication[22]. The datasets generated and analyzed during the current study, comprising the simulated fault records, the extracted feature matrices and the model outputs, are available from the corresponding author on reasonable request. The OpenDSS/MATLAB simulation pipeline and the analysis scripts that reproduce every figure and table in this article and in its Supplementary Information are likewise available from the corresponding author on reasonable request, and will be provided to the editors and reviewers during peer review.

Code availability

The OpenDSS and MATLAB code used to generate the dataset, train the models and compute every result in this article is available from the corresponding author on reasonable request.

References

1. Billinton, R. & Allan, R. N. *Reliability Evaluation of Power Systems* 2nd edn (Plenum, 1996).
2. *IEEE Guide for Electric Power Distribution Reliability Indices*. IEEE Std 1366-2012 (IEEE, 2012).
3. Zidan, A. et al. Fault detection, isolation, and service restoration in distribution systems: state-of-the-art and future trends. *IEEE Trans. Smart Grid* **8**, 2170–2185 (2017).
4. Mirshekali, H., Keshavarz, A., Dashti, R., Hafezi, S. & Shaker, H. R. Deep learning-based fault location framework in power distribution grids employing convolutional neural network based on capsule network. *Electr. Power Syst. Res.* **223**, 109529 (2023).
5. Jiang, Y. Data-driven fault location of electric power distribution systems with distributed generation. *IEEE Trans. Smart Grid* **11**, 129–137 (2020).
6. Mousavi, A. et al. Artificial neural networks-based fault localization in distributed generation integrated networks considering fault impedance. *IEEE Access* **12**, 82880–82896 (2024).
7. Bhagwat, A., Dutta, S., Jadoun, V. K., Veerendra, A. S. & Sahu, S. K. A customised artificial neural network for power distribution system fault detection. *IET Gener. Transm. Distrib.* **18**, 2105–2118 (2024).
8. Krysander, M. & Frisk, E. Sensor placement for fault diagnosis. *IEEE Trans. Syst. Man Cybern. A Syst. Humans* **38**, 1398–1410 (2008).
9. Abur, A. & Expósito, A. G. *Power System State Estimation: Theory and Implementation* (Marcel Dekker, 2004).
10. Marín-Quintero, J., Orozco-Henao, C. & Herrera-Orozco, A. Fault indicators allocation to maximize the performance of a fault locator based on artificial intelligence. *Electr. Power Syst. Res.* **234**, 110701 (2024).

11. Orozco-Henao, C., Marín-Quintero, J. & Herrera-Orozco, A. Optimal placement of intelligent electronic devices and fault indicators for enhancing FLISR performance in active distribution networks. *Electr. Power Syst. Res.* **254**, 112595 (2026).

12. Wang, F. et al. A multi-stage restoration method for medium-voltage distribution system with DGs. *IEEE Trans. Smart Grid* **8**, 2627–2636 (2017).

13. Wang, Z. & Wang, J. Self-healing resilient distribution systems based on sectionalization into microgrids. *IEEE Trans. Power Syst.* **30**, 3139–3149 (2015).

14. Shahsavari, A., Mazhari, S. M., Fereidunian, A. & Lesani, H. Fault indicator deployment in distribution systems considering available control and protection devices: a multi-objective formulation approach. *IEEE Trans. Power Syst.* **29**, 2359–2369 (2014).

15. Farajollahi, M., Fotuhi-Firuzabad, M. & Safdarian, A. Simultaneous placement of fault indicator and sectionalizing switch in distribution networks. *IEEE Trans. Smart Grid* **10**, 2278–2287 (2019).

16. *IEEE Standard for Synchrophasor Measurements for Power Systems*. IEEE Std C37.118.1-2011 (IEEE, 2011).

17. Anderson, P. M. *Analysis of Faulted Power Systems* (IEEE Press, 1995).

18. Sullivan, M. J., Schellenberg, J. & Blundell, M. *Updated Value of Service Reliability Estimates for Electric Utility Customers in the United States*. Report LBNL-6941E (Lawrence Berkeley National Laboratory, 2015).

19. Grinsztajn, L., Oyallon, E. & Varoquaux, G. Why do tree-based models still outperform deep learning on typical tabular data? In *Advances in Neural Information Processing Systems 35, Datasets and Benchmarks Track* (2022).

20. Dietterich, T. G. An experimental comparison of three methods for constructing ensembles of decision trees: bagging, boosting, and randomization. *Mach. Learn.* **40**, 139–157 (2000).

21. Bishop, C. M. Training with noise is equivalent to Tikhonov regularization. *Neural Comput.* **7**, 108–116 (1995).

22. Baran, M. E. & Wu, F. F. Network reconfiguration in distribution systems for loss reduction and load balancing. *IEEE Trans. Power Deliv.* **4**, 1401–1407 (1989).

23. Acharya, N., Mahat, P. & Mithulananthan, N. An analytical approach for DG allocation in primary distribution network. *Int. J. Electr. Power Energy Syst.* **28**, 669–678 (2006).

24. Kersting, W. H. *Distribution System Modeling and Analysis* 4th edn (CRC Press, 2017).

25. *IEEE Guide for Design, Operation, and Integration of Distributed Resource Island Systems with Electric Power Systems*. IEEE Std 1547.4-2011 (IEEE, 2011).

26. Roberts, D. R. et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40**, 913–929 (2017).

27. Forman, G. & Scholz, M. Apples-to-apples in cross-validation studies: pitfalls in classifier performance measurement. *ACM SIGKDD Explor. Newsl.* **12**, 49–57 (2010).

28. Dietterich, T. G. Approximate statistical tests for comparing supervised classification learning algorithms. *Neural Comput.* **10**, 1895–1923 (1998).

29. Demšar, J. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **7**, 1–30 (2006).

30. Chawla, N. V., Bowyer, K. W., Hall, L. O. & Kegelmeyer, W. P. SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002).

31. Blagus, R. & Lusa, L. SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics* **14**, 106 (2013).

32. Kohavi, R. & John, G. H. Wrappers for feature subset selection. *Artif. Intell.* **97**, 273–324 (1997).

33. Satopää, V., Albrecht, J., Irwin, D. & Raghavan, B. Finding a 'kneedle' in a haystack: detecting knee points in system behavior. In *Proc. 31st International Conference on Distributed Computing Systems Workshops* 166–171 (IEEE, 2011).

34. Kononenko, I. Estimating attributes: analysis and extensions of RELIEF. In *Machine Learning: ECML-94* 171–182 (Springer, 1994).

35. Robnik-Šikonja, M. & Kononenko, I. Theoretical and empirical analysis of ReliefF and RReliefF. *Mach. Learn.* **53**, 23–69 (2003).

36. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).

37. Freund, Y. & Schapire, R. E. A decision-theoretic generalization of on-line learning and an application to boosting. *J. Comput. Syst. Sci.* **55**, 119–139 (1997).

38. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20**, 273–297 (1995).

39. Schölkopf, B. & Smola, A. J. *Learning with Kernels* (MIT Press, 2002).

40. Goodfellow, I., Bengio, Y. & Courville, A. *Deep Learning* (MIT Press, 2016).

41. Bergstra, J. & Bengio, Y. Random search for hyper-parameter optimization. *J. Mach. Learn. Res.* **13**, 281–305 (2012).

42. Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian optimization of machine learning algorithms. In *Advances in Neural Information Processing Systems 25* 2951–2959 (2012).

43. He, H. & Garcia, E. A. Learning from imbalanced data. *IEEE Trans. Knowl. Data Eng.* **21**, 1263–1284 (2009).

44. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **20**, 37–46 (1960).

45. McHugh, M. L. Interrater reliability: the kappa statistic. *Biochem. Med.* **22**, 276–282 (2012).

46. Civanlar, S., Grainger, J. J., Yin, H. & Lee, S. S. H. Distribution feeder reconfiguration for loss reduction. *IEEE Trans. Power Deliv.* **3**, 1217–1223 (1988).

47. Baran, M. E. & Wu, F. F. Optimal capacitor placement on radial distribution systems. *IEEE Trans. Power Deliv.* **4**, 725–734 (1989).

48. *IEEE Standard for Calculating the Current-Temperature Relationship of Bare Overhead Conductors*. IEEE Std 738-2012 (IEEE, 2012).

Author contributions

O.G.A. conceived and designed the study, developed the simulation and learning pipeline, performed the analysis and wrote the manuscript. T.A.K., T.F.A. and E.S.A. reviewed and critically revised the manuscript. All authors read and approved the final manuscript.

Funding

The authors received no specific funding for this work.

Additional information

Competing interests: The authors declare no competing interests.

Supplementary Information accompanies this paper. Correspondence and requests for materials should be addressed to O.G.A. (`Omar.Gamal56@sh-eng.menofia.edu.eg`). E.S.A. can be reached at `eman_salah@sh-eng.menofia.edu.eg`.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SRsupplementaryinformation.pdf