International Journal of Computers and Informatics (IJCI)
Vol. (5), No. (11)
IJCI
November 2026
المجلة الدولية للحاسبات والمعلوماتية
الإصدار (5)، العدد (11)

Lightweight Neural Fault Diagnosis of Analog Electronic Circuits: The Value of a Second Measurement Node and the Limits of Transfer under Severity and Operating-Condition Shift

Roaa Ali Hussein
Assistant lecturer, Electrical Engineering, Department of Electronic and Communications Technologies Engineering, Polytechnic College Mosul, Northern Technical University, Iraq
Roaa.eng8@ntu.edu.iq

Shaymaa A. Ahmed
Assistant lecturer, Electrical Engineering, Department of Sensing and Nanophotonics, Laser and Photonics Center, University of Al-Hamdaniya, Mosul, Iraq
shymaalhabbar@uohamdaniya.edu.iq

Abstract

A classifier can identify large component deviations in a simulated analog circuit and still fail when the deviation becomes smaller or the measurement conditions change. This study evaluates a compact neural diagnostic method with that distinction made explicit, and reports the negative results as carefully as the positive ones. Complex transfer estimates at eight excitation frequencies are normalized by the nominal circuit response and supplied to a multilayer perceptron with 32 and 16 hidden units. Two observation nodes are used, and an output-only variant isolates the value of the additional measurement. The numerical benchmark covers a Sallen–Key low-pass filter and a cascade of two active first-order filters. Circuit realizations are generated independently for training, validation, calibration and six test domains, and the test domains enter none of the model-selection decisions. Across five training seeds, the two-node network reaches 100.00 % matched-domain accuracy on both circuits and 98.25 % and 99.96 % accuracy at a waveform signal-to-noise ratio of 20 dB. Accuracy falls to 79.85 % and 83.56 % for unseen 12–24 % component deviations, and to 56.77 % and 76.41 % under combined stress. A support vector machine performs better under combined stress, by 4.41 and 4.33 percentage points, whereas the neural model improves on the output-only network by 7.84 and 11.49 points. The network has 1,737 trainable parameters. A local sensitivity analysis explains both the benefit and the limit of the second observation node, and an exact parameter-scaling invariance of the Sallen–Key topology is shown to bound what any classifier can recover from these measurements. Confidence analysis shows that nominal-domain temperature scaling does not ensure reliability after a severity shift: expected calibration error increases after calibration in every shifted domain, and a fixed rejection threshold accepts fewer than 10 % of mild-fault cases. The results support a small diagnostic model inside a specified operating envelope, identify the measurements and experiments required before deployment, and should not be read as a claim of robustness outside that envelope.

Keywords: Analog Circuit Diagnosis, Parametric Faults, Compact Neural Network, Component Tolerances, Measurement Noise, Distribution Shift, Confidence Calibration, Selective Classification.

1. Introduction

Component drift changes an analog circuit continuously. A resistor that has moved beyond its specified tolerance may alter a filter response without stopping the circuit, and the resulting waveform may resemble the response of a healthy circuit with different component values. Fault diagnosis therefore depends on the circuit topology, the chosen observations, and the allowed parameter spread. Testability-aware neural diagnosis has already combined multi-frequency information with single-fault identification (Aizenberg et al., 2021). More recent work on nonlinear circuits emphasizes that an inverse diagnosis can admit more than one parameter explanation (Hałgas, 2025; Hałgas, 2026). The high classifying score does not mean the lack of the physical ambiguity.

Recent work on analog diagnosis based on convolutional features, spatial and temporal attention, and the selection of informative measurement nodes has been done (Yuan et al., 2025: Zhang et al., 2025; Chang et al., 2025; Yang et al., 2026). With the mentioned progress made, it would be reasonable to formulate a more specific question before further improvement of the model: How much diagnostic information can be extracted from a properly chosen combination of two measurements by a small network? The second question concerns the test distribution. Now domain adaptation for analog circuits considers fault severity and the operation point (Zhang et al., 2026). The mentioned factors become important since faults at the tolerance edge can be substantially harder to discriminate than faults with a 25-50% deviation.

The paper is concerned with three research questions through experiments, not hypotheses. What level of diagnostic accuracy can a second, intermediate observation node brings to the output in terms of width, optimization, and training data, RQ1? Is there an explanation for it through the circuit equations? RQ2. Will the accuracy in the matched domain be able to predict the behavior under new conditions of fault severity, component variance, measurement noise, or operating point variation in amplifiers if it is beyond the training domain, and which one will be more influential? RQ3. Does nominal-domain confidence calibration support a reliable rejection policy after such a shift?

This paper employs a controlled numerical benchmark to address the above questions. No long waveform or two-dimensional time-frequency map is taken as the input diagnosis. There is a short vector of complex-valued residual responses. The classifier is made of two very small hidden layers that are unseen. A radial-basis SVM and a random forest are chosen as non-neural references, and ablations are carried out to assess the measurement node selection and training with noisy measurements. This is not a comparison between these models versus those published networks on different circuits; this is a comparison between models built on the same dataset.

The protocol first separates the independent realizations of the circuit before generating the repeated training measurements. This avoids the scenario where noisy copies of the same realization of the circuit parameters end up in both training and testing splits. This type of leakage may overestimate a

model's generalizability, and the problem extends to all machine learning based sciences (Kapoor & Narayanan, 2023). In this paper, validation decides the epoch, another set decides the confidence intervals, and the stress domains are kept away from both.

This research provides an audit trail based on diagnosing benchmarks using node and noise ablation, with a small classifier and an evaluation of what is left unknown in changed deployment conditions. This research produces both poor and good results. It neither proposes any new neural network structure nor suggests that neural networks always perform better than classical methods in classification. It is intended for setting up a defendable numeric base for further SPICE and lab verification.

1.1 Organization:

Section 2 positions the work. Section 3 defines the circuits, the fault model and the identifiability bound. Section 4 describes the measurement features, the classifier and the confidence machinery. Section 5 specifies the evaluation protocol and the threats to validity. Section 6 reports results against RQ1–RQ3. Sections 7 and 8 discuss implications, limitations and the validation sequence required before any deployment claim.

2. Related work and study scope

(Yuan et al..2025) uses efficient channel attention to provide different processing paths for time-domain and frequency-domain information. Zhang et al. (2025) consider a spatial-temporal feature attention network. Both report high accuracy on their own benchmarks and these methods address representation learning. A further experimental path is explored in the present study: a compact description of the complex response is achieved by using eight known excitation tones and varying the number of observed nodes without increasing the network depth. Our goal is not to compete with the accuracy of deeper models, but rather to quantify the percentage of achievable accuracy that can be attributed to the measurement configuration, as opposed to the learned representation- a decomposition that cannot be obtained from an end-to-end comparison.

Measurement location is an established part of analog fault diagnosis. (Chang et al.,2025) combine node selection with diagnostic learning, while (Yang et al., 2026) formulate test-point selection as a sequential decision problem under a cost constraint. The two nodes used here are fixed from the circuit equations rather than searched. This choice makes the comparison easy to reproduce, but it does not establish an optimal node set for a larger circuit. The extra node also requires an extra measurement channel and cannot be treated as a free source of information; Section 7 returns to that cost.

Data quantity and the type of fault change the problem being solved. Jia et al. (2023) combine generative models with a Transformer for diagnosis from limited samples. Shi et al. (2025) fuse entropy and amplitude information to distinguish intermittent faults from noise. Nedjah et al. (2023) study parameter diagnosis through swarm-based optimization. The present benchmark concerns

steady single-component parametric deviations. It does not claim to detect transient interruptions, to estimate the exact faulty value, or to reproduce the inference procedure of those studies.

DiffDA-Net (Zhang et al., 2026) evaluates analog diagnosis under data imbalance and variable operating conditions, including a change in fault severity. That work motivates treating severity transfer as a separate test instead of mixing all severities into one random split. The present study uses balanced classes and performs no adaptation to the target domains, so the mild-fault and joint-stress scores describe an unadapted source-domain model and are a lower bound on what an adapted model could reach.

Two gaps follow from this literature. First, the reported accuracy of a diagnostic model is rarely decomposed into the part contributed by the measurement configuration and the part contributed by the classifier, even though the first is a hardware decision with a recurring cost and the second is not. Second, severity transfer is usually either absorbed into a random split or addressed only after adaptation, so the unadapted failure envelope — the quantity an engineer needs in order to decide whether adaptation is necessary at all — is seldom published. This study addresses both, at the price of a deliberately simple circuit model. The scope comparison in Table (1) records research directions rather than unsupported claims that earlier studies omitted particular safeguards.

Table (1) Related research directions and the scope of the numerical benchmark.

Research direction	Representative work	Relationship to this benchmark
Feature fusion and attention	Yuan et al. (2025)	Compact response residuals are used instead of a deep feature extractor.
Spatial and temporal representation	Zhang et al. (2025)	The data here are steady multifrequency measurements.
Measurement node selection	Chang et al. (2025); Yang et al. (2026)	Two fixed nodes are compared with the final output alone.
Limited sample learning	Jia et al. (2023)	Training-set size is fixed; no generated minority-class augmentation is used.
Intermittent faults and noise	Shi et al. (2025)	Only persistent parametric deviations are included.
Severity and operating domain transfer	Zhang et al. (2026)	Disjoint mild-fault severity and joint stress are tested without target adaptation.
Inverse diagnosis and ambiguity	Aizenberg et al. (2021); Hałgas (2025, 2026)	A sensitivity calculation and an exact scaling invariance bound interpretation.

3. Circuit models and fault definitions

3.1 Active filter topologies:

The first circuit is the Sallen–Key low-pass filter in Fig. (1a). Its topology follows the standard active-filter analysis (Karki, 2002), with capacitor names fixed by the figure rather than assumed from another drawing. R_1 connects the input to V_a; R_2 connects V_a to V_b. C_1 connects V_a to the amplifier output, and C_2 connects V_b to ground. The amplifier senses V_b through its non-inverting input. A feedback divider supplies a nominal β of 2/3, corresponding to a low-frequency gain close to 1.5. V_a and V_o are the diagnostic observation nodes.

The operational amplifier is represented by a one-pole open-loop response. Its nominal DC gain A_0 is 10^5 and its nominal unity-gain frequency f_t is 1 MHz. The response and the resulting non-inverting closed-loop gain are

$$A(s) = \frac{A_0}{1 + sA_0/(2\pi f_t)}, \qquad K(s) = \frac{A(s)}{1 + \beta A(s)} \tag{1}$$

Applying Kirchhoff's current law at V_a and V_b gives the denominator and the two observable transfer functions

$$D(s) = 1 + s[C_2(R_1 + R_2) + R_1C_1\{1 - K(s)\}] + s^2R_1R_2C_1C_2 \tag{2}$$

$$H_o(s) = \frac{K(s)}{D(s)}, \qquad H_a(s) = \frac{1 + sR_2C_2}{D(s)} \tag{3}$$

The nominal resistance and capacitance values are 10 kΩ and 10 nF. The undamped natural frequency is 1,591.55 Hz; it should not be read as the exact −3 dB frequency. With the ideal constant-gain approximation, the nominal quality factor is 2/3. All data use the finite-gain response in the preceding equations.

The second circuit, shown in Fig. (1b) is a cascade of two inverting first-order active low-pass stages. Each stage has a fixed 10 kΩ input resistor and a feedback resistor in parallel with a feedback capacitor. The stage output V_1 and the final output V_o are observed. If $Z_{f,i}$ denotes the feedback impedance, the stage response is

$$Z_{f,i}(s) = \frac{R_{f,i}}{1 + sR_{f,i}C_{f,i}}, \qquad G_i(s) = \frac{-Z_{f,i}(s)/R_{\mathrm{in},i}}{1 + \{1 + Z_{f,i}(s)/R_{\mathrm{in},i}\}/A(s)} \tag{4}$$

Thus $V_1/V_{in} = G_1$ and $V_o/V_{in} = G_1G_2$. R_{f1} and R_{f2} are nominally 10 kΩ, while C_{f1} is 10 nF and C_{f2} is 4.7 nF. The ideal individual pole frequencies are 1,591.55 and 3,386.28 Hz. Both amplifiers share the same sampled bandwidth within a circuit realisation; independent amplifier mismatch is outside this benchmark.

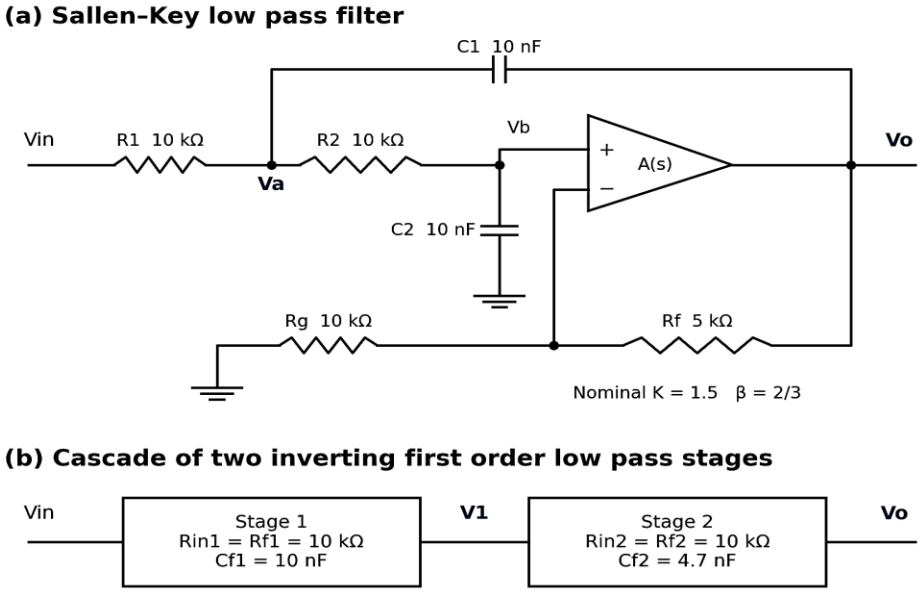


Figure (1) Circuit definitions. (a) Sallen–Key schematic with C1 connected to the output and C2 to ground. (b) Cascade stage diagram; each stage is an inverting amplifier with Rf in parallel with Cf. The input reference is measured in addition to the two indicated observation nodes.

Table (2) Circuit values and model scope.

Quantity	Sallen–Key	Cascade
Fault-target components	R1, R2, C1, C2	Rf1, Cf1, Rf2, Cf2
Nominal target values	10 kΩ, 10 kΩ, 10 nF, 10 nF	10 kΩ, 10 nF, 10 kΩ, 4.7 nF
Fixed resistors	Rg = 10 kΩ; Rf = 5 kΩ	Rin1 = Rin2 = 10 kΩ
Observation nodes	Va and Vo	V1 and Vo
Nominal amplifier parameters	$A_0 = 10^5$; f_t = 1 MHz; β = 2/3	$A_0 = 10^5$; f_t = 1 MHz
Training healthy tolerance	Resistors ±1 %; capacitors ±5 %	Resistors ±1 %; capacitors ±5 %
Unmodelled effects	Input and output impedance, saturation, slew rate, parasitic ESR, loading	The same effects; independent amplifier mismatch

3.2 Healthy spread and single component faults:

A healthy component is sampled independently around its nominal value. Uniform spread is a controlled modelling choice, not an assertion about a manufacturing distribution. For component j,

$$p_j = p_{j,0}(1 + \varepsilon_j), \qquad \varepsilon_j \sim \mathcal{U}(-t_j, t_j) \tag{5}$$

The nominal resistor tolerance t_j is 0.01 and the nominal capacitor tolerance is 0.05. A faulty target component replaces that healthy draw with a nominal-relative deviation,

$$p_k = p_{k,0}(1 + d\delta), \qquad d \in \{-1, +1\}, \qquad \delta \sim \mathcal{U}(\delta_{\min}, \delta_{\max}) \tag{6}$$

The other target components retain their healthy draws. Training faults use δ from 0.25 to 0.50, while the mild-fault test uses δ from 0.12 to 0.24. The intervals are disjoint. Even the smallest mild deviation lies outside the doubled 10 % capacitor tolerance used under stress. The healthy class together with the decreased and increased value of each of four components yields nine classes per circuit. Gain-divider faults and input-resistor faults are not included.

These labels identify a component and a deviation direction under a single-fault assumption. They do not identify every arbitrary combination of parameter changes. The simultaneous double-fault set is used only to probe confidence rejection and is excluded from classifier training.

3.3 Algebraic verification and parameter ambiguity:

The closed-form circuit responses were checked against independent complex nodal matrix solves at 120 randomly selected component and frequency conditions per topology. The maximum absolute discrepancies were 4.19×10^{-15} for Sallen–Key and 5.80×10^{-16} for the cascade. This verifies the response implementation. It does not validate the one-pole amplifier model against a physical device or a transistor-level simulator.

For Sallen–Key, multiplying both resistors by a positive factor q and dividing both capacitors by the same factor leaves all RC products and both transfer functions unchanged:

$$(R_1, R_2, C_1, C_2) \;\mapsto\; (qR_1,\, qR_2,\, C_1/q,\, C_2/q), \qquad q > 0 \qquad (7)$$

This complete invariance perseveres even when the limited response of the amplifier is taken into account. At a value of $q = 1.2$, the numerical deviation from the nominal transfer function is 2.22×10^{-16}. Consequently, it is impossible—even with two noise-free observation points—to uniquely determine the values of the four unconstrained components. Nevertheless, the classification of individual faults remains feasible, as the set of admissible parameters in this scenario is more narrowly defined. This distinction is crucial when interpreting the confusion matrix; it is an inherent property of the measurement configuration itself, rather than the classifier, and no increase in model capacity can overcome this limitation

4 Measurement features and compact diagnosis

4.1 Coherent multi-frequency excitation:

The excitation contains eight sinusoidal tones. The sample rate is 51.2 kHz and the record length is 512 samples, giving a 10 ms observation record and a frequency resolution of 100 Hz. The frequency-bin indices are 2, 5, 8, 13, 17, 24, 35 and 51, corresponding to 200, 500, 800, 1,300, 1,700, 2,400, 3,500 and 5,100 Hz. The tones are placed to bracket the nominal pole frequencies of both topologies on a quasi-logarithmic grid; they are not the solution of an optimal input-design problem, and Section 8 lists that optimization as an open item (Pintelon & Schoukens, 2012). The input is

$$u[n] = a \sum_{m=1}^{8} \cos\left(2\pi k_m n/N + \phi_m\right), \qquad n = 0, \ldots, N-1 \tag{8}$$

The nominal per-tone amplitude a is 0.1 V; a common amplitude factor is sampled from 0.95 to 1.05 for each measurement. The fixed phase vector is [0, 0.8, 1.7, 2.5, 3.4, 4.2, 5.1, 5.8] rad. The analysis assumes that circuit transients have settled before acquisition. An input-reference channel permits a response estimate for each observation node q:

$$\widehat{U}_m = \frac{2}{N} \sum_{n=0}^{N-1} u\,[n] e^{-\mathrm{j}2\pi k_m n/N}, \qquad \widehat{H}_{q,m} = \frac{\widehat{V}_{q,m}}{\widehat{U}_m} \tag{9}$$

Three voltage channels are therefore required for the two-node method: the excitation reference and the two circuit observations. Synchronized coherent sampling is part of the method. A system with an unknown excitation, a clock mismatch or significant channel loading would require a different measurement model.

4.2 Noise model and response residuals:

Independent additive white Gaussian measurement noise is assigned to each voltage channel. Waveform SNR is defined relative to the noiseless RMS value of that channel. Rather than synthesizing every complete time record, the dataset generator samples the exact Gaussian distribution of its coherent Fourier projections. For an N-sample real white-noise record, the real and imaginary projected errors at the selected positive-frequency bins have variance $2\sigma_q^2/N$:

$$\sigma_q = v_{q,\mathrm{rms}} 10^{-\mathrm{SNR}/20}, \qquad \eta_{q,m} \sim \mathcal{N}\left(0,\, 2\sigma_q^2/N\right) + \mathrm{j}\,\mathcal{N}\left(0,\, 2\sigma_q^2/N\right) \tag{10}$$

The noisy input denominator is shared by both response estimates, which preserves the induced correlation between their measurement errors. Training measurements use waveform SNR sampled uniformly from 30 to 50 dB, so the 20 dB test lies outside that range. Fourier averaging lowers the noise in the estimated tone amplitudes; the reported waveform SNR should therefore not be interpreted as the SNR of each extracted feature. Coloured noise and interference aligned with an excitation tone are not represented by this model.

Each complex response is divided by the nominal response at the same frequency, and its residual from unity is separated into real and imaginary parts:

$$r_{q,m} = \frac{\widehat{H}_{q,m}}{H_{q,m}^{(0)}} - 1, \qquad \mathbf{x} = [\Re\mathbf{r}_1,\ \Im\mathbf{r}_1,\ \Re\mathbf{r}_2,\ \Im\mathbf{r}_2] \tag{11}$$

This produces 32 real features for two nodes and 16 for the output-only ablation. It retains gain and phase information while removing the dominant nominal response scale. The feature mean and standard deviation are fitted on training records only. Normalization by a small nominal high-

frequency response can amplify measurement noise; the highest tone is limited to 5.1 kHz, and the stress tests retain this sensitivity.

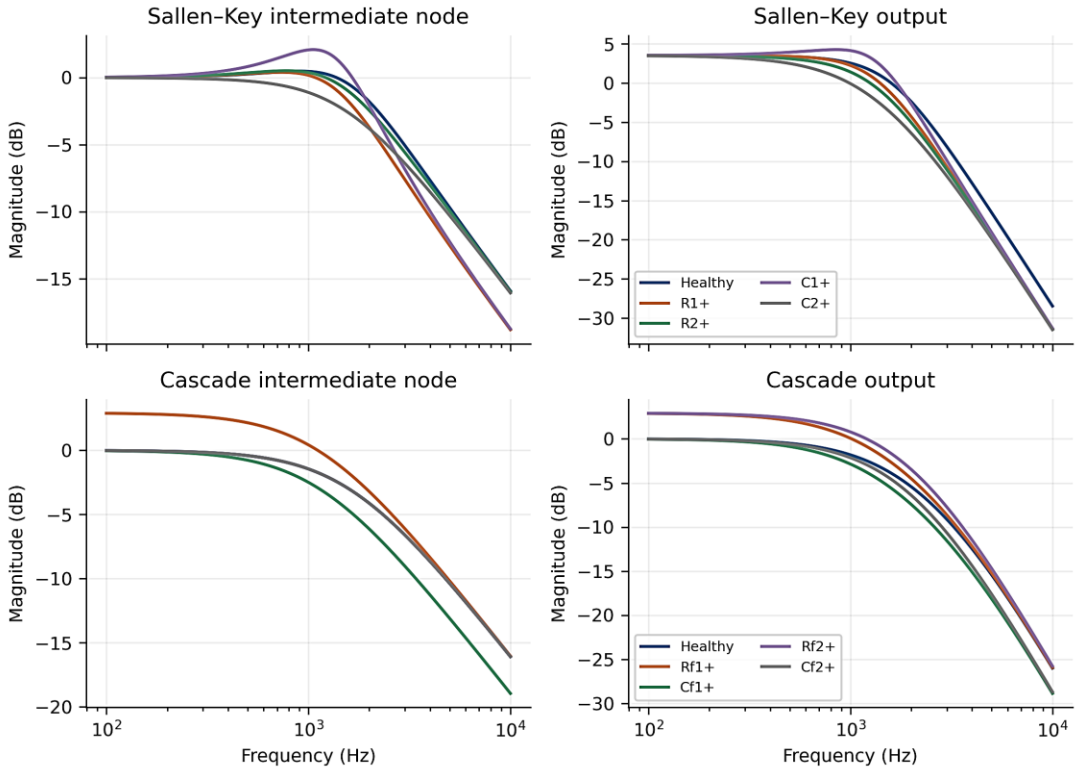


Figure (2) Computed frequency responses at nominal amplifier parameters. Each coloured fault curve increases one target component by 40 % and leaves the other components nominal. The response differences at the intermediate node help interpret the output-only ablation.

4.3 Classifier and reference models:

The main classifier is a multilayer perceptron with 32 and 16 repaired linear hidden units and nine class logits. The standardize of input feature are calculated before forward process.

Cross-entropy training uses Adam (Kingma & Ba, 2015), a learning rate of 0.002, batches of 256, and the scikit-learn L2 regularization parameter $\alpha = 0.001$. The logit vector is

$$\mathbf{z} = W_3\,\mathrm{ReLU}(W_2\,\mathrm{ReLU}(W_1\tilde{\mathbf{x}} + \mathbf{b}_1) + \mathbf{b}_2) + \mathbf{b}_3 \qquad (12)$$

Training runs for at most 150 epochs. The model snapshot with the lowest independent validation log loss is retained, and training stops after 18 epochs without an improvement of at least 10^{-4}. Repeated noisy training records are never used as a validation split. Separate models are fitted for the two circuit topologies; cross-topology transfer is not claimed.

The support vector machine uses a radial-basis kernel and the standard training-derived gamma scale (Cortes & Vapnik, 1995). C is selected from 1, 10 and 100 by validation macro F_1; $C = 1$ is selected for both circuits. The random forest contains 150 trees, square-root feature sampling and at least two samples per leaf (Breiman, 2001). Both references receive the same noisy training data and two-node features as the main model. Support-vector-machine accuracy is computed from its class decision rather than from the argmax of a fitted probability calibration. The support vector machine is deterministic given fixed data and hyper parameters, which is why its standard deviation across the five fits is exactly zero in Tables (6) and (7); the random forest retains estimator randomness and the neural models retain both initialization and batch-order randomness.

Table (3) Implemented models and ablation definitions.

Model	Measurements	Training measurements	Configuration
Two-node MLP	32 features	SNR 30–50 dB	Hidden units 32 and 16; 9 outputs
Output-only MLP	16 features	SNR 30–50 dB	Same hidden units and optimisation
Clean-training MLP	32 features	SNR 100 dB	Same circuit realisations and optimisation
SVM	32 features	SNR 30–50 dB	RBF; C = 1; gamma from training scale
Random forest	32 features	SNR 30–50 dB	150 trees; minimum leaf size 2

4.4 Calibration and selective decisions:

A separate calibration set is used for scalar temperature scaling (Guo et al., 2017). The temperature is selected by minimizing calibration log loss within the fixed interval 0.25 to 5:

$$p_c(\mathbf{x};T) = \frac{\exp(z_c/T)}{\sum_{j=1}^{9}\exp\left(z_j/T\right)}, \qquad T^* = \arg\min_{0.25\le T\le 5}\left[-\sum_i \log p_{y_i}(\mathbf{x}_i;T)\right] \tag{13}$$

A calibrated probability is an estimate for the calibration distribution. It is not a guarantee for the unseen operating condition, as uncertainty quantification studies under dataset shift have pointed out (Ovadia et al., 2019). In this scenario, each main model fit chooses a temperature near the lower limit of 0.25. The calibration data are easy and almost perfectly classified, and thus their optimization serves to sharpen instead of soften probabilities. This effect is explicitly examined in section 6.4 when stressed instead of being taken for granted.

The reject option uses selective-classification principle (Geifman & El-Yaniv, 2017). Acceptance is calculated using stable winning-class log odds, since confident softmax probability may round to one:

$$g(\mathbf{x}) = \frac{z_{c^*}}{T} - \log\sum_{j\neq c^*}\exp\left(z_j/T\right), \qquad c^* = \operatorname{argmax}_j z_j \tag{14}$$

The threshold τ is the tenth percentile of this score on the calibration set, which targets about 90 % calibration coverage without selecting a threshold on the stress domains. A lower score produces an abstention that requests further inspection. Coverage and error among accepted cases are reported together:

$$\text{coverage} = \frac{1}{n}\sum_i \mathbf{1}\,[g_i \geq \tau], \qquad \text{risk} = \frac{\sum_i \mathbf{1}\,[g_i \geq \tau]\,\mathbf{1}[\hat{y}_i \neq y_i]}{\sum_i \mathbf{1}\,[g_i \geq \tau]} \tag{15}$$

Equation (14) is the maximum-logit family of confidence scores, of which the softmax-probability baseline is a special case (Hendrycks & Gimpel, 2017). Feature-space and energy-based alternatives exist (Lee et al., 2018; Liu et al., 2020) and are not evaluated here; Section 7 argues that they are the natural next comparison given the present results.

For confidence discrimination of double faults, the negative log-odds score is evaluated as an unknown-condition score. The area under its receiver-operating characteristic curve is calculated against matched single-fault and healthy test cases. This auxiliary analysis assesses rejection behaviour; it does not add a tenth diagnostic class.

5 Evaluation protocol

5.1 Independent circuit partitions:

Each topology has 1,980 training realizations, comprising 220 examples per class. Three independently noisy measurements are generated for each training realization, producing 5,940 training records. The validation partition has 630 independent realizations and the calibration partition has 540. Each test domain contains 1,620 new realizations, with 180 per class. The same circuit seed is deliberately used for the clean and noisy training ablations so that their parameter realizations match; it is not reused for validation, calibration or testing.

The base circuit seed is 20,261,005 for Sallen–Key and 20,361,005 for the cascade. Validation and calibration use the respective base plus 1 and plus 2; the six stress sets use the base plus 10 through plus 15. Double faults use the base plus 99. Training seeds are 11, 22, 33, 44 and 55. These seeds vary initialization or estimator randomness on one fixed generated dataset. The five fits are therefore five fits of one dataset, not five independent manufacturing datasets, and every interval reported below is conditional on that dataset.

Table (4) Dataset sizes per circuit topology.

Partition	Circuits per class	Distinct realizations	Records used
Training	220	1,980	5,940 with three noise draws each
Validation	70	630	630
Calibration	60	540	540
Each of six test domains	180	1,620	1,620
Auxiliary double-fault set	Not a nine-class partition	1,620	1,620

5.2 Stress domains:

Table (5) defines the conditions before comparison. Normal amplifier bandwidth is sampled uniformly from 0.85 to 1.15 MHz. Amplifier-shift domains employ 0.60 and 1.40 MHz, respectively, while in Sallen-Key, ß is scaled by a uniform factor, ranging between 0.98 and 1.02. For cascade,

there is no sampled divider, and thus its amplifier-shift domains are limited to bandwidth. The individual test domains have independently generated circuits, and comparisons in a model within the test domains are made using the same instances.

The six test domains are arranged such that one factor is isolated by each of them while the last domain involves the combination of three factors. This ensures that the dominant factor can be determined instead of being inferred since if the model fails after successfully completing the other domains, it must be due to severity coverage and not measurement accuracy.

Table (5) Held-out test-domain definitions.

Test domain	Waveform SNR	R and C healthy spread	Fault deviation	Amplifier condition
Matched	40 dB	±1 % and ±5 %	25–50 %	Training distribution
Noise	20 dB	±1 % and ±5 %	25–50 %	Training distribution
Tolerance	40 dB	±2 % and ±10 %	25–50 %	Training distribution
Mild fault	40 dB	±1 % and ±5 %	12–24 %	Training distribution
Amplifier	40 dB	±1 % and ±5 %	25–50 %	Bandwidth shift; β spread in SK
Joint	20 dB	±2 % and ±10 %	12–24 %	Bandwidth shift; β spread in SK

5.3 Metrics and comparison uncertainty:

Accuracy and macro F_1 are the metrics for the nine-class classification problem. False alarm rate for healthy circuits is the ratio of correctly predicted healthy circuits among the falsely classified circuits. Missed fault rate is the ratio of circuits predicted to be healthy despite being faulty. Confusion between two faulty circuits is not considered to be a missed fault. For the confidence assessment, the log-loss, Brier Score and Expected Calibration Error are calculated using 15 equally wide bins:

$$\mathrm{ECE} = \sum_{b=1}^{15} \frac{|B_b|}{n} |\mathrm{acc}(B_b) - \mathrm{conf}(B_b)| \qquad (16)$$

Performance tables give the mean and sample standard deviation across the five fitted seeds. The five fits, that have standard deviation is a major instant which reported as an expressive range as the foundation of any insinuation. The model accuracy differences are evaluated with 1,200 paired, class-stratified bootstrap resamples of the held-out circuit recognitions (Efron, 1979). The correctness has been averaged over five fits before resampling operation, with each realization.

The perccential intervals pointwise are calculated to be (95%) on the fixed training dataset, with fitted models. Usually, the resampling of new training population is not included for uncertainty form. It is representing an acceptable intervals description without many comparisons correction, these parameters (intervals) act as exclude zero with small margin must not treated as confirm results.

Computation uses Python with NumPy 2.3.5, SciPy 1.17.0 and scikit-learn 1.8.0 (Pedregosa et al., 2011; Virtanen et al., 2020). The companion material preserves the configuration, generated feature arrays, response checks, predictions, representative neural weights and summary metrics. Its hashes

identify each dataset used in the tables. No target-domain labels enter training, early stopping, temperature fitting or the rejection threshold.

5.4 Threats to validity:

Construct validity. can be defined as an Accuracy over nine classes that used as the primary amounts, while the operational charge of a healthy-to-fault error changes from that elements confusion. Macro F_1, the false-alarm rate and the missed-fault rate are therefore reported alongside it, and the confusion matrices are retained.

Internal validity. Represents a significant internal threat that causes leakage between noisy copies of one parameter realization to another. It is organized by generating realizations before amounts to keeps similar registers inside a single partition. A second threat is that early stopping, temperature fitting and the rejection threshold could each leak target information; all three use only the validation or calibration partitions.

External validity. The benchmark is a closed-form frequency-domain model of two low-order topologies with uniform, independent component spread. Transfer to a transistor-level simulation, to correlated production tolerances, or to a physical instrument is not established by these results and is the subject of Section 8.

Statistical conclusion validity. All variability reported here originates from model fitting on one generated dataset. Between-dataset variability is unmeasured, so the reported intervals understate the uncertainty that a replication with new circuit draws would exhibit.

6 Numerical results

6.1 Accuracy within and outside the training envelope:

Tables (6) and (7) report all implemented models. Every two-node model reaches 100.00 % matched-domain accuracy on both topologies. That result is consistent with the broad 25–50 % fault deviations, the controlled excitation and the simplified steady circuit model. It should not be read as a general probability of correct diagnosis in a laboratory or a field system; it indicates instead that the matched domain is saturated and cannot discriminate between candidate models. The informative differences appear only when the test condition changes, which is the reason the stress domains, rather than the matched domain, carry the argument of this paper.

Table (6) Sallen–Key accuracy in percent as mean ± standard deviation over five fits.

Domain	SVM	Random forest	Output MLP	Clean MLP	Two-node MLP
Matched	100.00 ± 0.00	100.00 ± 0.00	97.73 ± 0.27	100.00 ± 0.00	100.00 ± 0.00
Noise	99.38 ± 0.00	100.00 ± 0.00	94.16 ± 0.99	92.48 ± 2.94	98.25 ± 0.39
Tolerance	99.14 ± 0.00	96.80 ± 0.17	83.67 ± 0.50	99.19 ± 0.53	99.41 ± 0.44
Mild fault	74.07 ± 0.00	66.69 ± 0.69	69.78 ± 2.27	79.35 ± 1.82	79.85 ± 2.91
Amplifier	99.94 ± 0.00	100.00 ± 0.00	92.49 ± 0.96	98.69 ± 0.87	99.75 ± 0.11
Joint	61.17 ± 0.00	56.28 ± 0.72	48.93 ± 1.73	49.10 ± 2.87	56.77 ± 2.70

Table (7) Cascade accuracy in percent as mean ± standard deviation over five fits.

Domain	SVM	Random forest	Output MLP	Clean MLP	Two-node MLP
Matched	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
Noise	100.00 ± 0.00	99.95 ± 0.03	98.53 ± 0.31	99.96 ± 0.03	99.96 ± 0.03
Tolerance	100.00 ± 0.00	98.53 ± 0.07	98.99 ± 0.19	99.98 ± 0.03	99.99 ± 0.03
Mild fault	84.14 ± 0.00	82.65 ± 0.83	83.44 ± 1.98	84.15 ± 1.45	83.56 ± 1.98
Amplifier	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00	100.00 ± 0.00
Joint	80.74 ± 0.00	78.63 ± 0.49	64.91 ± 2.09	76.83 ± 1.16	76.41 ± 1.40

On Sallen–Key, the two-node MLP retains 98.25 % accuracy at 20 dB waveform SNR and 99.41 % under doubled component tolerances. The corresponding cascade results are 99.96 % and 99.99 %. The support vector machine can hold its own in such a setting just as well. Under isolated amplifier change, the primary network achieves 99.75 % in Sallen-Key and 100.00 % In the cascade. This holds true for the given bandwidth and divider disturbance range, but not any other replacement of an operational amplifier.

In the context of mild faults, however, greater deviation occurs. Accuracy will be at 79.85 % for Sallen-Key and 83.56 % for the cascade despite having a signal-to-noise ratio of 40 dB. The healthy to faulty separation boundary derived from 25-50 % deviations, therefore, does not apply to all 12-24% deviations. When under joint stress, accuracy falls to 56.77 % and 76.41 %. In turn, the support vector machine obtains 61.17 % and 80.74 % in that case, bettering the small network by 4.41 and 4.33 percentage points, respectively

This is a direct answer to research question 2: out of the four isolated shifts, the severity shift prevails. The measurement noise shift affects the two-node network by 1.75 points on Sallen-Key and 0.04 points on cascade; the doubled tolerance by 0.59 points on Sallen-Key and 0.01 points on cascade; and the amplifier by 0.25 and 0.00 points. The three mild factors are not additive with it either: the joint domain is 23.08 and 7.15 points below the mild-fault domain, so interaction between severity and spread is itself a substantial term.

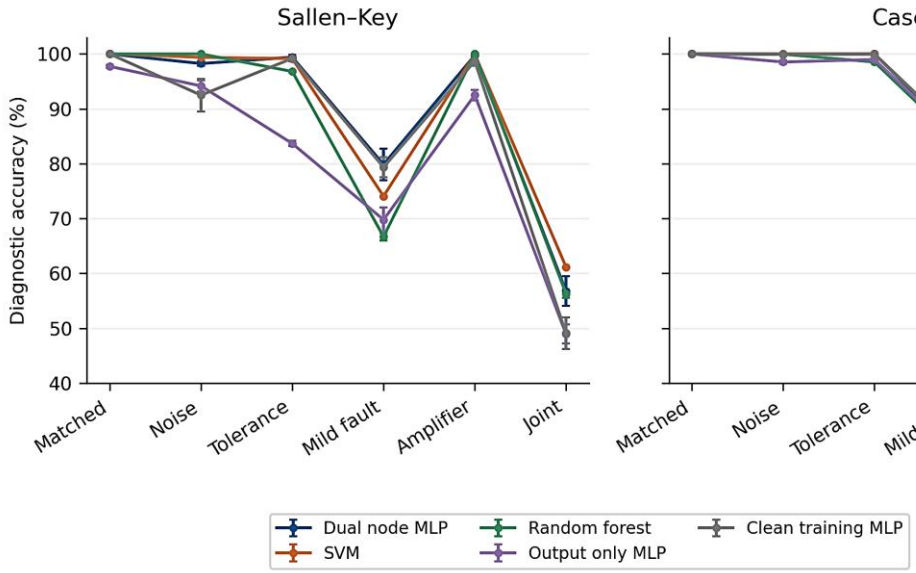


Figure (3) Mean accuracy in the six held-out domains. Error bars show standard deviation across five fitted seeds on the same fixed data partitions. Domain labels correspond to Table 5.

6.2 What the node and noise ablations establish:

For the Sallen–Key circuit, the output-only network achieves an accuracy of 97.73% in the matched regime, 94.16% in the presence of noise, and 48.93% under common-mode stress conditions. Including the variable V_a increases accuracy under common-mode stress by 7.84 percentage points, with a 95% confidence interval ranging from 5.79 to 9.68 points. For the cascaded version, the output-only approach yields perfect results in the matched regime, but performance drops to 64.91% under common-mode stress. Incorporating the variable V_1 boosts this result by 11.49 percentage points, with a confidence interval of 9.88 to 13.15 points. These gains reflect the impact of additional observation, while the width of the hidden layers and the optimization process remain constant. Since the two-node input layer also introduces 512 additional weights, the ablation study affects both the measurement data and the number of parameters in the input layer. Consequently, the Jacobian analysis—presented in Section 6.3—is employed to disentangle these two factors, demonstrating that the observed improvement is attributable to "observability" rather than "model capacity."

Noise exposure has a topology-dependent effect. On Sallen–Key, noisy training improves the two-node network over clean training by 5.77 points in the noise domain and 7.67 points under joint stress. On the cascade, the joint-stress change is −0.42 points, with an interval of −0.79 to −0.06 points. The tested augmentation therefore improves the noise-sensitive Sallen–Key case but does not improve every transfer setting. Fault severity and circuit response geometry remain relevant after measurement noise is included. Taken together with the node ablation, this answers RQ1: a measurement change

produced a larger and more consistent gain across both topologies than a data-augmentation change, which is a hardware conclusion rather than a modelling one.

Table (8) Selected paired accuracy differences for the two-node noisy-training MLP.

Circuit and domain	Comparator	Main minus comparator in pp	Paired 95 % interval in pp
SK noise	Clean MLP	5.77	4.96 to 6.49
SK mild	SVM	5.78	4.69 to 6.90
SK joint	Output MLP	7.84	5.79 to 9.68
SK joint	SVM	−4.41	−6.09 to −2.54
Cascade joint	Output MLP	11.49	9.88 to 13.15
Cascade joint	Clean MLP	−0.42	−0.79 to −0.06
Cascade joint	SVM	−4.33	−5.69 to −2.90

6.3 Diagnostic errors and local sensitivity:

The joint-stress confusion matrices in Fig. (4) show that failures include both wrong component labels and faults assigned to the healthy class. The main network's healthy false-alarm and missed-fault rates are 22.11 % and 27.92 % on Sallen–Key, and 4.00 % and 24.40 % on the cascade. Its joint-stress macro F_1 values are 59.19 % and 79.82 %. Those values explain why a single accuracy number cannot characterise the diagnostic consequence of the shift: on Sallen–Key, more than one healthy circuit in five would be sent for unnecessary inspection, while more than one faulty circuit in four would be released as healthy.

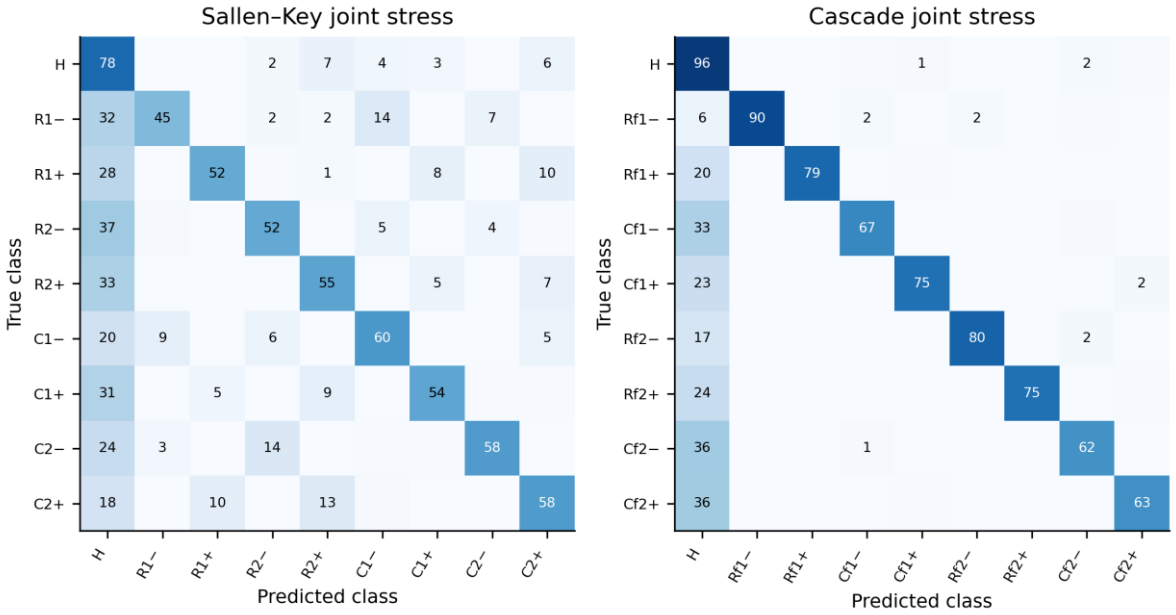


Figure (4) Row-normalized joint-stress confusion matrices for the main network, averaged across five fits. Entries are percentages and values below 1 % are omitted. H denotes healthy; minus and plus denote decreased and increased nominal-relative component values.

To relate these results to the circuit measurements, a local Jacobian is calculated at the nominal parameter vector. The columns use logarithmic parameter perturbations and the rows use the unstandardized normalized-response features. Central differences with a logarithmic step of 10^{-5} give

$$J_{\ell j} = \frac{\partial x_\ell}{\partial \log p_j}\Big|_{\mathbf{p}=\mathbf{p}_0}, \qquad J = U\Sigma V^{\mathrm{T}} \tag{17}$$

For Sallen–Key, the third output-only singular value relative to the largest is 3.38×10^{-6}; with V_a included it becomes 0.180, an improvement of more than four orders of magnitude. A fourth independent parameter direction remains absent because of the exact RC scaling invariance of Eq. (7). For the cascade case, the minimum relative singular value increases from 9.79×10^{-7} for the output at the end of the cascade to 0.273 with V_1 taken into account. The finite amplification factor gives a weak distinguishing power for an otherwise nearly unobservable output-only direction, whereas the middle observation provides much stronger relative sensitivity.

This is the quantitative connection between Sections 6.2 and 3.3. The gains in measurement accuracy by 7.84 and 11.49 points occur with directions having a relative sensitivity increase from below 10^{-5} to above 0.18, which is what is expected when an unobservable parameter direction becomes observable. The 512 additional weights from the input cannot provide a similar interpretation, since they affect variables carrying little parameter information if the node was not observed. Figure 5 shows this relative sensitivity without making the local derivative measurement become a statement about global identifiability: the single fault labels bound the inverse problem, while any multiple faults may overlap.

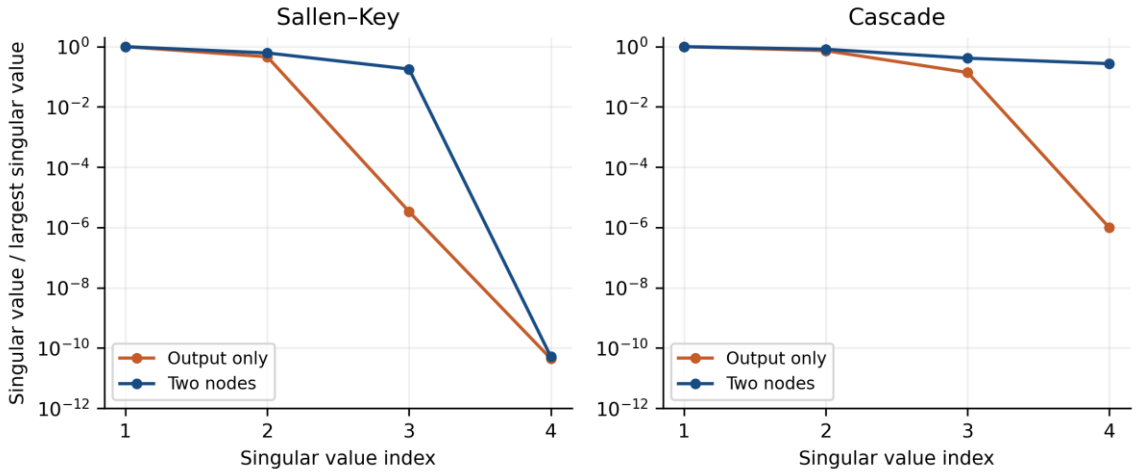


Figure (5) Singular values of the nominal normalised-response Jacobian, each divided by the largest value for that observation configuration. The very small fourth Sallen–Key value is numerical residue of an exact parameter-scaling invariance. The cascade output-only direction is weak, rather than exactly absent, with the finite amplifier model.

6.4 Confidence reliability and rejection:

The calibration set leads every main-model temperature to the lower allowed bound. Matched-domain ECE is effectively zero, but calibration worsens ECE after a mild-fault or joint shift. For Sallen–Key, mild-fault ECE increases from 9.90 % before calibration to 17.79 % afterwards, and joint-stress ECE increases from 29.93 % to 40.09 %. For the cascade, the corresponding changes are 6.48 % to 13.83 % and 12.16 % to 21.14 %. This is proof that calibration for the easy domain should not be transferred blindly into the modified diagnostic domain.

The process is not a mystery but can be described. Temperature scaling is designed to minimize log loss over a set on which the model performs almost ideally; the minimizer of such an objective is the minimum permissible temperature, which makes all probabilities more certain. The latter operation is beneficial when the model is correct and dangerous when the model is incorrect, thus making the procedure optimal on the calibration domain and detrimental on the diagnostic domain with accuracy inferior to the calibration domain. The lower limit of 0.25 in Eq. (13) is not a random choice in the code but a requirement, without which the fitted temperature is unbounded from below.

Table (9) Calibration error and the fixed-threshold rejection policy for the main network.

Circuit	Domain	Raw ECE %	Cal ECE %	Coverage %	Accepted error %
SK	Matched	0.02	0.00	89.19	0.00
SK	Noise	0.78	1.53	70.21	0.00
SK	Mild	9.90	17.79	10.06	0.84
SK	Joint	29.93	40.09	8.74	28.27
Cascade	Matched	0.02	0.00	88.89	0.00
Cascade	Noise	0.18	0.04	79.81	0.00
Cascade	Mild	6.48	13.83	8.42	0.00
Cascade	Joint	12.16	21.14	3.90	18.72

Fixed threshold enables 89.19% and 88.89% matching test samples close to the calibration coverage target. As for the scenario of the mild fault type, the acceptance rate will be reduced to 10.06% and 8.42%.

In this case, the rate of the acceptance error will be minimal, however, the majority of the diagnoses will be deferred. Thus, the acceptance strategy cannot become the safe low-coverage mode, but rather turns into the universal deferral strategy. Under the two stress factors, the acceptance rate is 8.74% and 3.90%, while the rate of errors of accepted samples is 28.27% and 18.72%, respectively. Thus, the deferral strategy does not ensure the safe operating mode under the extreme joint shift, answering the RQ3 negatively. For the additional dataset of simultaneous double faults, the mean confidence-score AUROC equals to 0.799 for Sallen-Key and 0.865 for the cascade, correspondingly. At the same time, the rejection rate is 67.84% and 74.83%, respectively. It means partial discrimination, but not recognition of all possible untrained fault combinations.

For the auxiliary dataset of simultaneous double faults, the mean confidence-score AUROC is 0.799 for Sallen-Key and 0.865 for the cascade. At that, the rejection rate is 67.84% and 74.83%,

respectively. This means partial discrimination, but not identification of all possible untrained fault combinations. Unknown conditions may look like known single faults with a high confidence score.

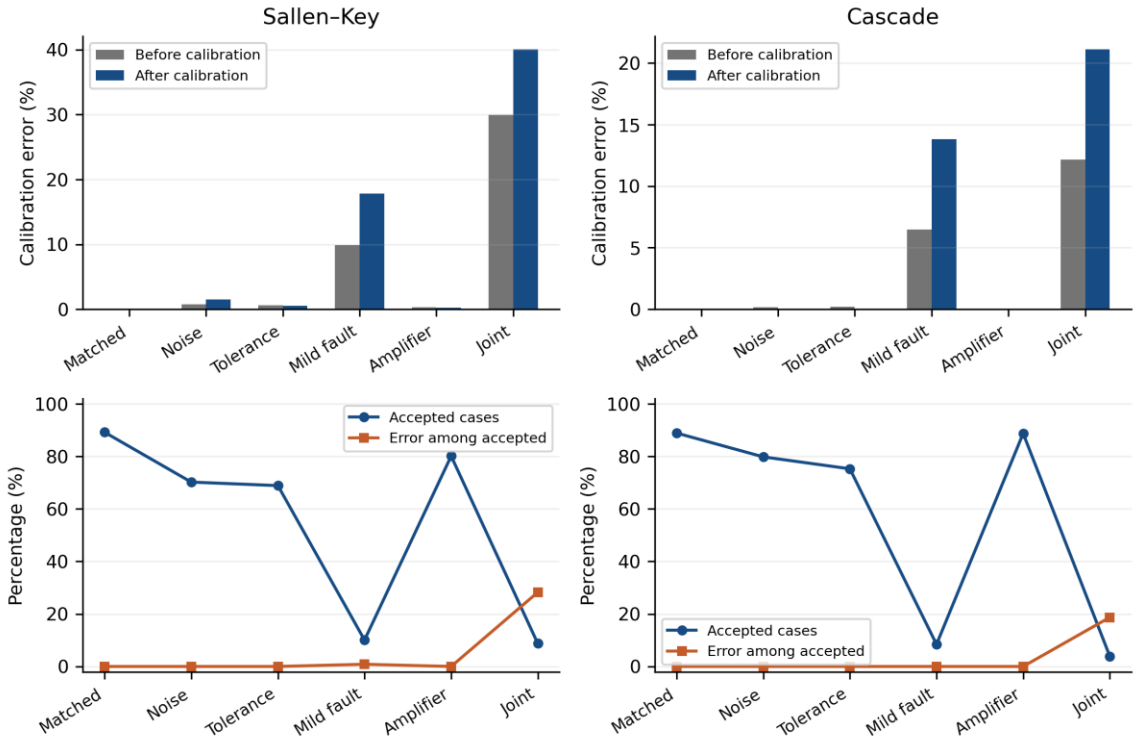


Figure (6) Expected calibration error before and after nominal-domain temperature fitting, together with coverage and error among accepted cases. The rejection threshold is fixed using calibration data; it is not chosen separately for each stress domain.

6.5 Computational size:

The 32–32–16–9 network has 1,737 trainable weights and biases and requires 1,680 matrices multiply–accumulate operations per forward logit calculation. They would take up 6.79 KiB in storage as 32-bit floating-point numbers. The output-only neural network uses 1,225 parameters and performs 1,168 operations. The above do not include feature extraction, scaling coefficients, activation functions, softmax, code and acquisition buffer sizes. They define the size of the classifier but not the memory requirements of a full-fledged embedded diagnostic tool, and they are not in any way comparable with those of the support vector machine, which have a variable storage depending on the number of kept support vectors.

The indicative host processor logit timing per one record after scaling is 9.35 µs for the Sallen-Key fits and 8.97 µs for the cascade fits, averaged over five different models after 50 warming calls and 2,000 timed calls per model. The above NumPy timing is performed on the host computer and uses

float64 arithmetic; it is not a measurement done on the microcontroller or FPGA. The acquisition record alone takes 10 ms, so the classifier does not form the latency bottleneck of a full diagnosis, and there are additional settling and preprocessing.

7. Discussion

Most importantly, the separation between fit on easy-domain and transfer of diagnosticity to difficult domain has been shown. The small network successfully detects the large controlled deviations in the benchmark. In addition, the network survives the noise and increased tolerance without increasing fault severity out of the range of the training domain. Small deviations alter the task since they shift the faulty behavior closer to the healthy one. Combining this alteration with increased dispersion and noise leads to the formation of the failure envelope which is not seen in the matched-domain performance. Since the matched-domain is saturated for all five models, an analysis of it would have mistakenly stated that all the models are equal.

A related corollary has to do with the interpretation of the severity of the failure. There is no evidence here that the compact architecture cannot recognize deviations at 12-24% levels; there is only evidence that an architecture trained on 25-50% deviations cannot extend its recognition to the 12-24% deviations. The two assertions carry very different implications, and a control experiment where the training severity range is broadened to cover the test range, with everything else remaining the same, would distinguish between the two. Such a test is outlined in Section 8 as the number one validation task, and until such an experiment is done, the current results should be considered coverage statistics rather than capability limits.

The new observation node provides a more definite benefit compared to noisy training in both cases. This can be explained by the Jacobian matrix: the method amplifies the parameter-sensitive directions, which are not very well captured by the end-of-the-line output. There is an obvious drawback associated with the benefit above. It is possible that the intermediate node is hidden in the packaged circuit and could be loaded by a probing circuitry or ADC front-end.

The findings in relation to confidence align with the common problem of how uncertainty estimates tend to be compromised by shifts in the datasets (Ovadia et al., 2019). The almost zero-error calibration set contributes towards overly confident probabilities. It may be possible to correct this through the inclusion of near-boundary faults in a fresh training and calibration process, but the test sets used in the current study should be preserved for posterity rather than used for calibration.

8. Limitations, recommendations and validation priorities:

The results are a numerical study of two low-order active RC topologies. They use closed-form frequency-domain models verified by nodal equations and not transistor-level SPICE models or physical measurements. The operational-amplifier model omits finite output impedance, input loading, offsets, slew-rate limits, saturation, parasitic components and coloured noise. No measured

temperature dependence or ageing trajectory is used. The component deviations represent fault scenarios, not a validated lifetime model.

The baseline is balanced and employs uniform, independent component distributions. In contrast, a production batch may exhibit correlated tolerances, non-uniform defect distributions, and component values that cluster together. The primary classifier also assumes the presence of only one faulty target component at a time. A dual-fault study serves solely to assess confidence levels, and the robustness analysis of the Sallen-Key topology demonstrates that it is not possible to reconstruct unconstrained parameters from the classifier's results.

A stable, coherent multi-tone stimulus is applied throughout the study, with the eight tone frequencies selected via empirical inspection rather than through optimal input design (Pentelion and Shokins, 2012). Timing errors, frequency mismatches, transient windowing, and narrowband interference can alter the extracted response residuals. Such effects must be incorporated through comprehensive time-domain simulations and a model of the measurement front-end before any practical claims can be made; furthermore, probe loading must be explicitly modeled when selecting an intermediate node.

All reported variability comes from fitting one generated dataset with five seeds. Between-dataset variability is therefore unmeasured, and the reported intervals are narrower than a full replication would produce.

The validation sequence that would convert this baseline into a deployment claim is, in order of priority:

1. **Severity-coverage control.** Repeat the entire protocol with training deviations spanning 12–50 % and all other settings fixed, reserving a new mild-fault test set. This separates a coverage limitation from a capability limitation and is the single experiment most likely to change the interpretation of Section 6.1.
2. **Dataset replication.** Regenerate the benchmark with at least three independent circuit seeds and report between-dataset as well as between-seed variability, so that the intervals in Table 8 become unconditional.
3. **Confidence redesign.** Evaluate at least one calibration-free confidence score (Hendrycks & Gimpel, 2017; Lee et al., 2018; Liu et al., 2020) and a risk-targeted threshold against the current log-odds policy on the same shifted domains.
4. **Model-size parity.** Report the support-vector-machine storage and inference cost under the same accounting used in Section 6.5, so that the compactness claim is comparative rather than absolute.
5. **Macro-model cross-check.** Repeat the same component draws with manufacturer operational-amplifier macro-models in a circuit simulator, preserving a new held-out test set.
6. **Laboratory prototype.** Build a low-voltage prototype, compare nominal and independently measured components, synchronize all three voltage channels, and quantify loading and quantization.

7. **Embedded implementation.** Implement the compact model on a microcontroller and measure full memory use, latency and power.

Items 5 to 7 would support a deployment claim; they are not represented as completed in this manuscript.

9 Conclusion

A compact two-node neural classifier provides a reproducible baseline for single-component parametric diagnosis in the two modelled active filters. Its 1,737 parameters support perfect matched-domain classification and high accuracy under the tested isolated noise, tolerance and amplifier changes. The second observation node improves joint-stress accuracy over the output-only network by 7.84 and 11.49 percentage points on the two circuits, and a local sensitivity analysis attributes that gain to parameter directions that the final output measures weakly rather than to the additional input weights. Severity transfer and combined stress remain substantial limitations: accuracy falls to 79.85 % and 83.56 % for unseen mild deviations and to 56.77 % and 76.41 % under combined stress, where a support vector machine performs better. Nominal-domain confidence calibration not only fails to establish a dependable rejection policy after that shift but degrades the calibration error in every shifted domain, for a reason that is traced to the near-perfect calibration population. These results show the numerical range of the method and the set of organized data that are needed to test it beyond simulation.

Nomenclature

Symbol	Meaning
A(s), A_0, f_t	Open-loop amplifier response, DC gain, unity-gain frequency
K(s), β	Non-inverting closed-loop gain and feedback-divider ratio
D(s)	Common denominator of the Sallen–Key transfer functions
H_o(s), H_a(s)	Transfer functions at the output node and at node Va
G_i(s), Z_f,i	Gain and feedback impedance of cascade stage i
p_j, p_j,0, t_j	Component value, nominal value and tolerance of component j
ε_j, δ, d	Healthy relative spread, fault deviation magnitude, fault direction
q	Scaling factor of the exact RC parameter invariance
N, k_m, φ_m	Record length, excitation bin index, excitation phase
$\hat{U}$_m, $\hat{V}$_q,m, $\hat{H}$_q,m	Estimated input, node and transfer spectra at tone m
σ_q, η_q,m, SNR	Channel noise level, projected spectral error, waveform SNR
r_q,m, x	Normalized response residual and the assembled feature vector
z, T, p_c	Logit vector, calibration temperature, calibrated class probability
g(x), τ	Selective-classification score and its acceptance threshold
J, Σ	Normalized-response Jacobian and its singular-value matrix
ECE	Expected calibration error with 15 equal-width bins

References

- Aizenberg, I., Belardi, R., Bindi, M., Grasso, F., Manetti, S., Luchetta, A., & Piccirilli, M. C. (2021). A neural network classifier with multi-valued neurons for analog circuit fault diagnosis. Electronics, 10(3), 349. https://doi.org/10.3390/electronics10030349.
- Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5–32.

  https://doi.org/10.1023/A:1010933404324.
- Chang, Y., Xu, X., Deng, K., Xu, Y., Tu, B., Gao, X., & Wei, Q. (2025). An analog circuit fault diagnosis method incorporating multi-objective selection of measurement nodes. Electronics, 14(8), 1528. https://doi.org/10.3390/electronics14081528.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297. https://doi.org/10.1007/BF00994018.
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. The Annals of Statistics, 7(1), 1–26. https://doi.org/10.1214/aos/1176344552.
- Geifman, Y., & El-Yaniv, R. (2017). Selective classification for deep neural networks. In Advances in Neural Information Processing Systems (Vol. 30). https://arxiv.org/abs/1705.08500.
- Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. In Proceedings of the 34th International Conference on Machine Learning (PMLR Vol. 70, pp. 1321–1330). https://proceedings.mlr.press/v70/guo17a.html.
- Hałgas, S. (2025). Challenges in fault diagnosis of nonlinear circuits. Electronics, 14(22), 4427. https://doi.org/10.3390/electronics14224427.
- Hałgas, S. (2026). A method for identifying single faults in nonlinear circuits. Electronics, 15(10), 2198. https://doi.org/10.3390/electronics15102198.
- Hendrycks, D., & Gimpel, K. (2017). A baseline for detecting misclassified and out-of-distribution examples in neural networks. In Proceedings of the International Conference on Learning Representations. https://arxiv.org/abs/1610.02136.
- Jia, Z., Yang, Q., Li, Y., Wang, S., Xu, P., & Liu, Z. (2023). A fault diagnosis strategy for analog circuits with limited samples based on the combination of the Transformer and generative models. Sensors, 23(22), 9125. https://doi.org/10.3390/s23229125.
- Kapoor, S., & Narayanan, A. (2023). Leakage and the reproducibility crisis in machine-learning-based science. Patterns, 4(9), 100804. https://doi.org/10.1016/j.patter.2023.100804.
- Karki, J. (2002). Analysis of the Sallen-Key architecture (Application Report SLOA024B). Texas Instruments. https://www.ti.com/lit/pdf/sloa024.

- Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. In Proceedings of the 3rd International Conference on Learning Representations. https://arxiv.org/abs/1412.6980.
- Lee, K., Lee, K., Lee, H., & Shin, J. (2018). A simple unified framework for detecting out-of-distribution samples and adversarial attacks. In Advances in Neural Information Processing Systems (Vol. 31). https://arxiv.org/abs/1807.03888.
- Liu, W., Wang, X., Owens, J. D., & Li, Y. (2020). Energy-based out-of-distribution detection. In Advances in Neural Information Processing Systems (Vol. 33). https://arxiv.org/abs/2010.03759
- Nedjah, N., Luna Galindo, J. D., de Macedo Mourelle, L., & Reis de Oliveira, F. D. V. (2023). Fault diagnosis in analog circuits using swarm intelligence. Biomimetic, 8(5), 388. https://doi.org/10.3390/biomimetics8050388.
- Ovadia, Y., Fertig, E., Ren, J., Nado, Z., Sculley, D., Nowozin, S., Dillon, J. V., Lakshminarayanan, B., & Snoek, J. (2019). Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift. In Advances in Neural Information Processing Systems (Vol. 32).
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.
- Pintelon, R., & Schoukens, J. (2012). System identification: A frequency domain approach (2nd ed.). Wiley–IEEE Press.
- Shi, J., Hou, Y., Wang, Z., Yang, Z., & Lv, Z. (2025). A diagnosis method for noise and intermittent faults in analog circuits based on the fusion of multiscale fuzzy entropy features and amplitude features. Sensors, 25(4), 1090. https://doi.org/10.3390/s25041090.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. Nature Methods, 17(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2.
- Yang, H., Liu, X., Gao, T., & Yang, J. (2026). Analog circuit test point selection method for fault diagnosis based on deep reinforcement learning. Engineering Applications of Artificial Intelligence, 164(B), 113294. https://doi.org/10.1016/j.engappai.2025.113294.
- Yuan, H., Shi, Y., Li, L., Ling, G., Zeng, J., & Wang, Z. (2025). Fault diagnosis in analog circuits using a multi-input convolutional neural network with feature attention. Computation, 13(4), 94. https://doi.org/10.3390/computation13040094.

- Zhang, A., Yang, Q., & Han, L. (2026). DiffDA-Net: Diffusion-augmented domain adaptive network for analog circuit fault diagnosis under imbalanced and variable operating conditions. Scientific Reports, 16, 30784. https://doi.org/10.1038/s41598-026-60313-3.
- Zhang, C., Li, M., Li, W., Dong, Z., He, S., & Zhou, Z. (2025). Fault diagnosis of analog circuit based on spatial–temporal feature attention network. Electronics, 14(7), 1341. https://doi.org/10.3390/electronics14071341.