Journal of Computational Electronics (2026) 25:166
https://doi.org/10.1007/s10825-026-02611-7

From microstructure to instability: a multimodal data-driven workflow for leakage-aware screening of memristive devices

Saba Zamankhani[1] · Hauke Honig[2] · Michaela Blum[2] · Eric Troebs[1] · Peter Schaaf[2] · Kai-Uwe Sattler[1]

Received: 23 March 2026 / Accepted: 15 July 2026 / Published online: 4 August 2026
© The Author(s) 2026

Abstract

Stress-induced microstructural and interfacial changes are widely implicated in the electrical instability of memristive devices, yet they are rarely quantified within a single leakage-aware predictive framework. Here we present a proof-of-concept small-data study that integrates registered pre- and post-stress SEM differencing, DC current–voltage hysteresis descriptors, and illumination-sensitive electrochemical impedance spectroscopy features to predict a device-level electrical instability proxy. All analyses are performed using strict device- and batch-aware grouping to prevent cross-device leakage and to evaluate cross-batch generalization. Multimodal fusion improves predictive performance relative to single-modality baselines under device-grouped validation, whereas leave-one-batch-out analysis reveals heterogeneous transfer across fabrication batches. Grouped permutation testing and bootstrap uncertainty estimates support a modest average cross-batch benefit while highlighting substantial fabrication-dependent domain shift. Feature attribution identifies complementary predictive signals across modalities: SEM texture and Fourier descriptors summarize stress-associated image changes, DC metrics capture switching asymmetry, and impedance phase features reflect illumination-sensitive interfacial response. These attributions are interpreted as physically motivated, hypothesis-generating indicators rather than direct evidence of unique causal mechanisms. Together, these results establish a reproducible workflow for instability-oriented screening and batch-level process assessment in memristive materials, while defining a small-data benchmark for future validation against direct reliability endpoints.

Keywords Memristive devices · Multimodal fusion · Scanning electron microscopy (SEM) · Current–voltage (*I-V*) · Electrochemical impedance spectroscopy (EIS)

1 Introduction

Memristive devices are promising building blocks for next-generation non-volatile memory and neuromorphic computing architectures owing to their scalability, analog tunability, and low-energy operation [1]. However, large-scale deployment remains limited by reproducibility challenges and stress-induced electrical instability, which manifest as device-to-device variability, cycle-to-cycle drift, and threshold-voltage scatter [2]. These behaviors arise from field-driven ionic motion, interfacial redox reactions, and stochastic filament formation within the oxide layer,

Saba Zamankhani, Hauke Honig and Michaela Blum are contributed equally to this work.

✉ Saba Zamankhani
saba.zamankhani@tu-ilmenau.de

Hauke Honig
hauke-lars.honig@tu-ilmenau.de

Michaela Blum
michaela.blum@tu-ilmenau.de

Eric Troebs
eric.troebs@tu-ilmenau.de

Peter Schaaf
peter.schaaf@tu-ilmenau.de

Kai-Uwe Sattler
kus@tu-ilmenau.de

[1] Department of Computer Science and Automation, TU Ilmenau, Ilmenau, Germany

[2] Chair Materials for Electrical Engineering and Electronics, MEEE, Institute of Micro and Nanotechnologies IMN and Institute of Materials Science and Engineering, Department of Electrical Engineering and Information Technology, TU Ilmenau, Ilmenau, Germany

processes that couple structural and electronic degrees of freedom and evolve during device operation [3–6]. Understanding how such microstructural evolution relates to electrical instability is therefore central to predictive reliability assessment and fabrication control.

Experimental studies have shown that degradation and variability in memristive devices are often accompanied by morphological and interfacial changes in the active oxide layer. Electron microscopy can reveal features such as filament formation, densification, void evolution, and microcracking during device operation [7–9]. Complementary electrical characterization techniques probe the transport mechanisms associated with these structural changes. Direct-current (I–V) measurements capture nonlinear transport and hysteresis behavior, while electrochemical impedance spectroscopy (EIS) provides frequency-dependent insight into interfacial charge-transfer dynamics and trap-mediated processes [4, 10, 11]. Despite these advances, structural and electrical measurements are typically analyzed independently, making it difficult to quantitatively relate observed microstructural transformations to device-level instability.

Recent progress in materials informatics suggests that combining microstructural descriptors with electrical measurements can enable data-driven discovery of structure–property relationships in complex materials systems. However, multimodal analyses of memristive devices remain scarce, and the small size of typical experimental datasets poses challenges for robust statistical validation and information-leakage control.

Here we test the hypothesis that microstructural signatures captured in registered pre- and post-stress SEM images can serve as quantitative predictors of electrical instability metrics derived from I–V and EIS measurements. Representative DC characteristics illustrating switching-like and non-switching-like behavior are shown in Fig. 1. To evaluate this hypothesis, we introduce a multimodal workflow that integrates normalized SEM difference maps, DC hysteresis descriptors, and illumination-sensitive impedance features within a device-grouped, leakage-controlled analysis framework. As illustrated in Fig. 2, SEM image pairs are aligned using feature-based homography estimation and converted into difference maps from which morphological, textural, and spectral descriptors are extracted. These structural features are then combined with electrical indicators such as nonlinearity, ON/OFF separation, and impedance phase slopes.

Using a curated dataset of 32 metal/oxide/metal memristor devices, we show that multimodal feature fusion

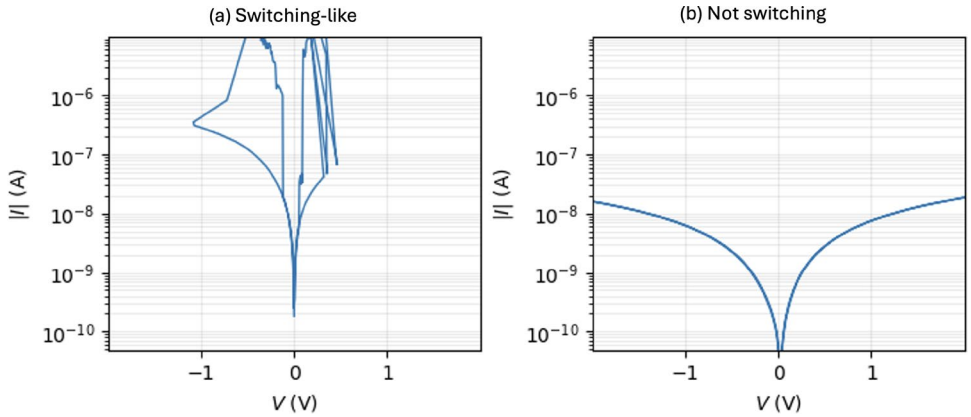


Fig. 1 Representative DC current–voltage characteristics of **a** a switching-like device and **b** a not-switching device, illustrating stress-induced variability in memristive transport behavior

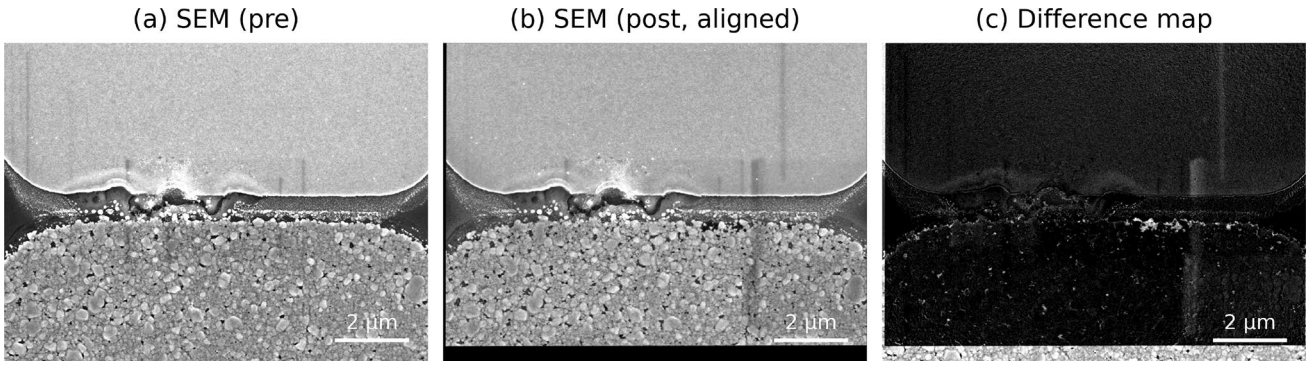


Fig. 2 Representative SEM pair. **a** Pre-stress SEM image, **b** post-stress SEM after feature-based alignment, and **c** Normalized difference map. Device-level instability metric $\bar{s}_1 = 1.000 \pm 0.002$. Alignment quality metrics: ORB matches = 295, inlier ratio = 0.5, ECC correlation = 0.653

improves predictive performance relative to single-modality models under device-grouped cross-validation. Feature attribution analysis identifies complementary predictive contributions across modalities: SEM texture and Fourier-domain descriptors are consistent with stress-induced morphological change, DC hysteresis metrics summarize transport asymmetry, and illumination-dependent EIS phase descriptors capture sensitivity to interfacial electrical response. We interpret these descriptors cautiously as physically motivated indicators rather than as direct mechanistic proof, since attribution in a correlated small-data feature space cannot by itself establish causal roles. Together, these results demonstrate that multimodal descriptors capture complementary signatures of memristor instability and provide a reproducible framework for reliability screening and batch-level process assessment in memristive materials.

2 Related work

Early studies on resistive switching established the memristor concept and linked its operation to field-driven ion migration, filamentary conduction, and interfacial redox reactions [6, 12–14]. Subsequent research revealed a strong dependence of switching behavior on fabrication conditions and operational stress, underscoring the importance of reproducible reliability methodologies [3–5].

Electron microscopy has provided key insights into resistive switching by visualizing filament growth, void evolution, and interface roughening during cycling [8, 15–17]. However, many of these studies rely on qualitative before-and-after comparisons or manual inspection, which limits throughput and reproducibility. Quantitative image analysis methods–such as gray-level co-occurrence matrices for texture, local binary patterns for fine structure, and Fourier-domain metrics for characteristic length scales–enable consistent comparison when supported by robust registration and normalization [17–19]. Automated alignment using feature-based descriptors (e.g., ORB) and model fitting (e.g., RANSAC) has become standard for creating reproducible difference maps, allowing observed structural evolution to be translated into measurable, uncertainty-aware descriptors [20–22].

Electrical characterization complements these imaging approaches by probing transport dynamics and interfacial phenomena. DC sweeps yield metrics such as nonlinearity, hysteresis energy, and ON/OFF current ratio, while EIS captures frequency-dependent impedance behavior [3, 4]. Environmental perturbations such as ultraviolet illumination further reveal trap dynamics and interface states [10, 11, 23, 24]. Although modern EIS studies emphasize confidence interval reporting and model identifiability [25, 26], the quantitative use of illumination-induced impedance shifts as indicators of device stability remains limited. Moreover, few studies integrate microscopy, DC, and EIS data within a unified predictive framework.

Recent developments in materials informatics have shown that combining engineered features with machine-learned representations can effectively link microstructure and macroscopic properties [27–30]. In the context of memristive systems, data-driven models trained on electrical or fabrication metadata have been used to predict variability and switching statistics [31, 32]; yet, these approaches often overlook dependencies introduced by repeated measurements and rarely incorporate imaging data. Group-aware validation and permutation-based significance testing are essential to prevent data leakage and ensure robust inference [33, 34], but publicly available multimodal datasets remain scarce, and uncertainty quantification is often incomplete.

The present work addresses these limitations by introducing a unified, measurement-driven workflow that integrates SEM differencing, DC metrics, and illumination-sensitive EIS descriptors into a single quantitative framework. Through robust registration, standardized feature extraction, and group-aware validation, this approach bridges the structural and electrical domains to yield physically motivated, process-relevant descriptors for predictive modeling and reliability screening. The resulting framework improves predictive accuracy and provides attribution-supported hypotheses about morphology, hysteresis behavior, and interfacial response, while avoiding the stronger claim that individual descriptors uniquely identify specific physical mechanisms.

3 Methods

This work establishes a multimodal workflow that links microstructural evolution to electrical stability by integrating registered scanning electron microscopy (SEM) differencing, direct-current (I-V) analysis, and illumination-aware electrochemical impedance spectroscopy (EIS) summaries (Fig. 3). Metadata and raw signals were structured into hierarchical `HDF5` containers organized by wafer, batch, and device. In this study, a batch denotes one die containing four devices that were fabricated and characterized together. Thus, devices within a batch share the same local process and measurement history, while devices from different batches may reflect die-level fabrication variability and local microstructural differences.

3.1 Target construction and validation

The prediction target is a device-level quarter-sweep symmetry score, $\bar{s}_1$, derived from even-indexed DC sweeps; odd sweeps are reserved for feature extraction (Sect. 3.2).

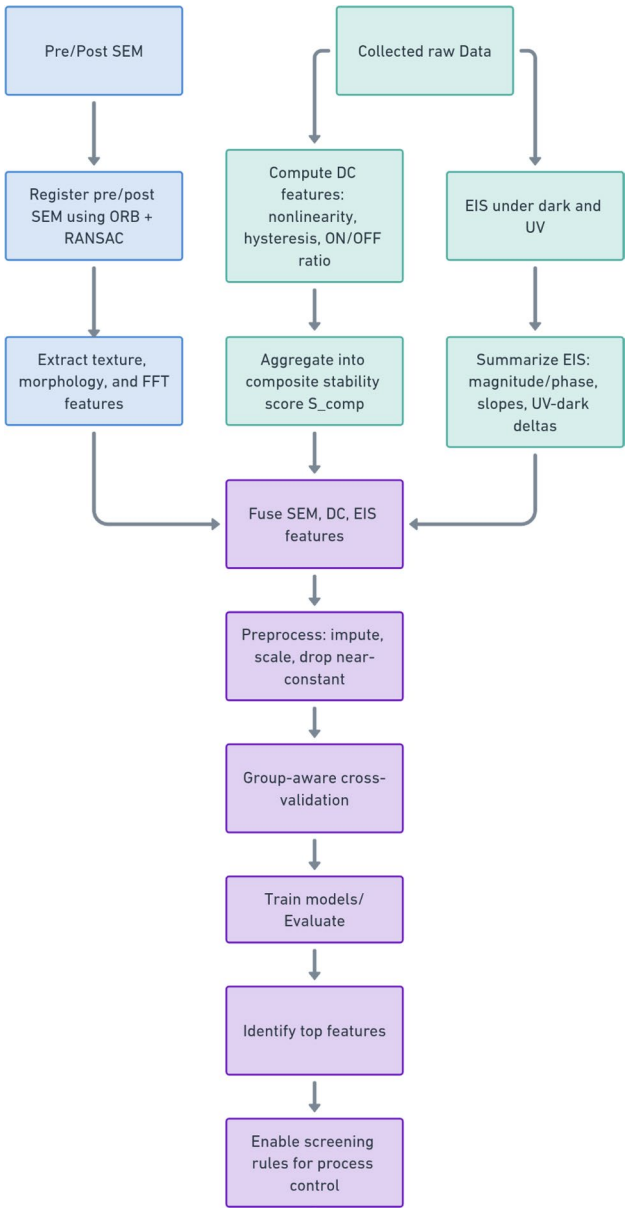


Fig. 3 Schematic of our method integrating SEM differencing, DC $I-V$ descriptors, and illumination-sensitive EIS features into a leakage-controlled, group-aware predictive pipeline for memristor instability

For each sweep with samples $\{(U_k, I_k)\}_{k=0}^{N-1}$, where U_k is the applied voltage (V) and I_k the measured current (A), the local resistance is computed as

$$R_k = \frac{U_k}{I_k} \quad (\Omega),$$

following constraints to suppress numerical artifacts:

$$I \in [-10^{-6}, 10^{-6}]\ \text{A}, \quad |I| \geq 10^{-9}\ \text{A},$$
$$R \text{ winsorized to } [P_2, P_{98}], \quad |R| \leq 10^6\ \Omega.$$

With trimming parameter $\ell = 0.2$ and midpoint $m = (N-1)/2$, the quarter-symmetry statistic is

$$s_1 = \gamma \sum_{k \in \mathcal{K}} \left(\frac{R_k}{R_{m-k}} + \frac{R_{N-1-k}}{R_{m+k}} \right), \quad \gamma = \frac{2}{(1-\ell)(N-1)}, \quad (1)$$

where $\mathcal{K}$ indexes valid symmetric resistance pairs.

The device-level label $\bar{s}_1$ is the average of valid sweep scores:

$$\bar{s}_1 = \frac{1}{T} \sum_{t=1}^{T} s_1^{(t)},$$

computed only when at least five valid symmetric pairs exist.

This dimensionless metric summarizes reciprocal behavior in the I–V loops and penalizes deviations from quarter-sweep symmetry, while excluding the high-noise region near $U \approx 0$ V. We therefore interpret $\bar{s}_1$ as an electrical instability proxy rather than as a direct lifetime or endurance endpoint.

Because s_1 is computed using index-matched resistance pairs, current-compliance events can shift the voltage regions contributing to individual resistance ratios. Future versions of the score could instead compare resistance values at matched voltages; however, voltage-matched interpolation is not always straightforward in the presence of volatile resistance changes and non-monotonic sweep segments.

Figure 4a, b illustrate the distribution of s_1 before and after filtering, respectively.

Robustness of $\bar{s}_1$ was verified using Bland–Altman bias plots, 95% limits of agreement, and the intraclass correlation coefficient ICC(2, 1), as shown in Fig. 4. Sensitivity to guard thresholds $(U_{\text{skip}}, I_{\text{min}})$ was analyzed through heatmaps of the median absolute deviation $|\Delta s_1|$ and the Spearman correlation coefficient ρ (Fig. 4d, e), confirming rank stability for permissive thresholds ($\rho \gtrsim 0.85$).

3.2 DC descriptors and composite quality index

DC features were extracted from trajectories disjoint from those defining the target. For each valid sample with $|U| \geq U_{\text{skip}}$ and $|I| \geq I_{\text{min}}$, the positive and negative nonlinearities were computed as the absolute slopes of $\log |I|$ versus U:

$$\text{nonlin}_+ = |\ \text{slope}\ (\log |I| \text{ vs. } U | U \geq U_{\text{skip}})|,$$
$$\text{nonlin}_ = |\ \text{slope}\ (\log |I| \text{ vs. } U | U \leq -U_{\text{skip}})|.$$

The hysteresis area,

$$A_h = \left| \int_{\circlearrowleft} I\, \mathrm{d}U - \int_{\circlearrowleft} I\, \mathrm{d}U \right|,$$

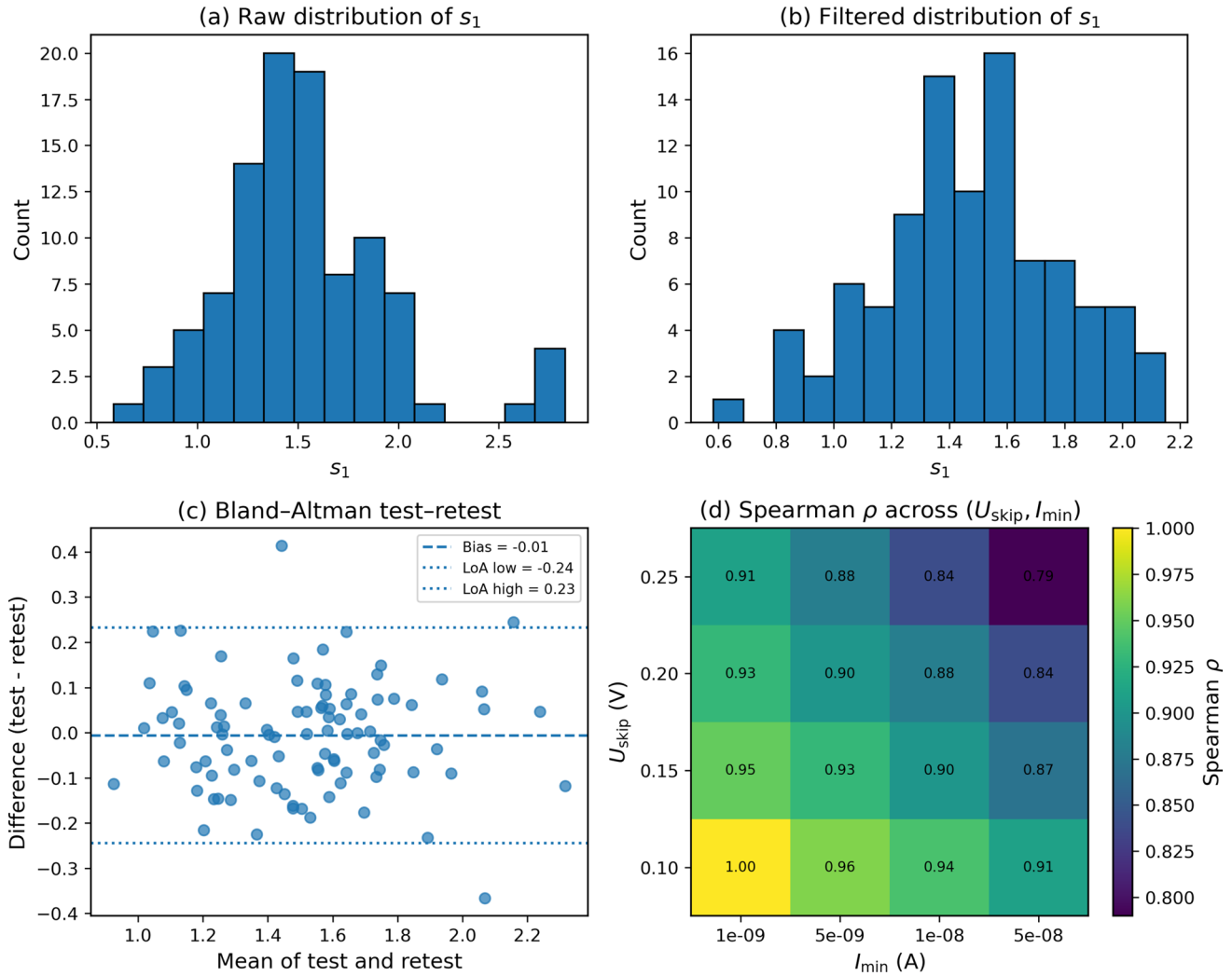


Fig. 4 **a** Raw distribution of the s_1 stability metric for measurements. **b** Filtered s_1 distribution after removal of extreme values. **c** Bland-Altman analysis of test-retest agreement, showing minimal bias and tight limits of agreement. **d** Spearman stability map illustrating the rank consistency of s_1 across (U_{skip}, I_{min}) parameter variations

was evaluated over forward and reverse sweeps, excluding the skip region and currents below I_{min}. The ON/OFF separation was derived from median resistances in protocol-defined read windows:

$$\mathrm{ON/OFF} = \max\left(\frac{\tilde{R}_{off}}{\tilde{R}_{on}}, \frac{\tilde{R}_{on}}{\tilde{R}_{off}}\right) \geq 1,$$

with limits $\mathrm{ON/OFF} \leq \rho_{max}$ and $A_h \leq A_{h,max}$ to prevent outlier distortion. These parameters were summarized in a composite DC index:

$$Q_{DC} = \sum_{q \in \{\mathrm{nonlin}_+, \mathrm{nonlin}_-, A_h, \mathrm{ON/OFF}-1\}} \log(1+q),$$

where the logarithmic transform stabilizes scale and enforces non-negativity. The device median of Q_{DC} was used for modeling.

3.3 SEM registration and differencing

Pre- and post-stress SEM images were cropped to a shared field of view and aligned using ORB keypoints (1000 descriptors per image), brute-force Hamming matching, and RANSAC homography estimation [20–22]. The transformation matrix **H** was estimated using the best $K = \min(200, n_{match})$ pairs with a reprojection threshold of 2.5 px. Registrations were accepted if the inlier ratio exceeded 0.5 and mean reprojection error was below 3.0 px. Quality metrics (inlier ratio, reprojection error, and match count) were retained as covariates.

A normalized absolute difference map,

$$(I_{\text{diff}} = \left|(\frac{I_{\text{pre}} - \mu}{\sigma}) - (\frac{I_{\text{post}} - \mu}{\sigma})\right|,$$

was computed using global normalization parameters μ and σ across both frames and rescaled to 8-bit. From I_{diff}, we extracted intensity, texture, morphological, and frequency-domain features. Intensity statistics included mean, standard deviation, skewness, kurtosis, and entropy. Texture was characterized using gray-level co-occurrence matrices across multiple offsets and orientations, and uniform local binary patterns with radii $R \in \{1, 2, 3\}$. Morphological descriptors were obtained via adaptive thresholding and connected-component analysis, reporting particle count, area, perimeter, solidity, eccentricity, and aspect ratio. Frequency-domain content was quantified from the 2D FFT magnitude using the fine-to-coarse ratio

$$\text{FR} = \frac{E_{\text{high}}}{E_{\text{low}} + 10^{-12}},$$

representing the relative contribution of high-frequency components.

3.4 EIS features and illumination deltas

Impedance spectra under dark and ultraviolet illumination were summarized without circuit fitting. For each condition, we computed the mean, extrema, and slopes of $|Z|$ and $\angle Z$ versus frequency by least-squares regression:

$$\text{slope}_{\log|Z|} = \text{slope}\,(\log_{10}|Z| \text{ vs. } \log_{10} f), \quad \text{slope}_{\angle Z} = \text{slope}\,(\angle Z_{\text{rad}} \text{ vs. } \log_{10} f).$$

Illumination sensitivity was captured through differences between UV and dark conditions, for instance,

$$\Delta\text{phase} = \overline{\angle Z}_{\text{UV}} - \overline{\angle Z}_{\text{dark}}, \quad \Delta\,\text{slope}_{\log|Z|} = \text{slope}_{\log|Z|,\text{UV}} - \text{slope}_{\log|Z|,\text{dark}}.$$

3.5 Data assembly and preprocessing

Each device-magnification pair was represented as a single row containing SEM, DC, and EIS descriptors along with registration quality metrics and metadata. Because the prediction target is defined at device level, all rows belonging to the same physical device were kept within the same validation fold. Model errors were computed after aggregating predictions to device level by averaging the out-of-fold predictions across magnifications for each device. This prevents devices with multiple SEM magnifications from receiving disproportionate weight in the final RMSE calculation. Infinite values were set to missing; variables with excessive missingness were discarded, and remaining gaps were median-imputed within cross-validation folds. Scaling and dimensionality reduction (via mean-variance or robust interquartile range normalization) were fitted on the training data only [35, 36].

3.6 Models and training

We evaluated a compact set of tabular regression models chosen to span complementary inductive biases while remaining appropriate for the available sample size. These models are used as regularized statistical probes for comparing the predictive contributions of different data modalities, rather than as deployable predictors.

The model set comprised six regressors. As an ensemble baseline, we used a random forest regressor (RF), selected for its robustness to feature scaling and its ability to capture nonlinear interactions in heterogeneous tabular data. To examine whether simple batch-level centering could improve transfer across fabrication groups, we also evaluated a batch-offset variant of the random forest, in which target normalization was performed using training-fold statistics only. As a linear reference, we trained a ridge-regularized regression model on principal components explaining 95% of the variance, providing a low-variance baseline for assessing the incremental benefit of nonlinear models. We further included Kernel Ridge Regression (KRR) with a radial-basis-function kernel as a smooth nonlinear alternative with convex optimization, Gaussian Process Regression (GPR) as a probabilistic kernel method for small-data regression, and histogram gradient boosting (HGB) as a second tree-based nonlinear model with a boosting inductive bias distinct from bagged ensembles.

Model hyperparameters were tuned within grouped inner cross-validation loops using only the training portion of each outer split. For RF and its batch-offset variant, the search included tree depth and minimum leaf size; analogous grouped tuning was applied to the regularization and kernel parameters of ridge, KRR, and GPR, and to the boosting parameters of HGB. This design preserves strict separation between training and evaluation data under device- and batch-aware grouping.

Although we also explored convolutional and hybrid neural-network architectures for SEM image analysis and multimodal learning, the dataset was too small to support stable training. These models consistently overfit and underperformed the tabular baselines, even with strong regularization

and augmentation, and are therefore excluded from the main analysis.

Predictive performance was quantified using the root-mean-square error (RMSE),

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2},$$

computed out-of-fold for each validation split. To account for batch-dependent target scaling, we additionally report the normalized RMSE (NRMSE),

$$\mathrm{NRMSE} = \frac{\mathrm{RMSE}}{y_{\max} - y_{\min}},$$

where $y_{\max}$ and $y_{\min}$ are computed from the corresponding training fold. For leave-one-batch-out evaluation, we also report a weighted mean RMSE across held-out batches, with each batch weighted by its number of devices:

$$\mathrm{RMSE}_w = \frac{\sum_{b=1}^{B} n_b \,\mathrm{RMSE}_b}{\sum_{b=1}^{B} n_b}.$$

3.7 Group-aware validation and permutation testing

Validation respected device-level grouping, ensuring that all magnifications of a device were assigned to the same fold. Outer validation used grouped K-folds ($K \leq 5$) and inner grouped loops for hyperparameter optimization. Leave-one-batch-out (LOBO) evaluation tested cross-batch transfer, with preprocessing and tuning confined to the $N-1$ training batches.

Statistical significance of observed performance gains was quantified through grouped permutation testing. Keeping the validation structure fixed, the observed error E_{obs} (e.g., RMSE) was compared against a null distribution obtained by permuting device-level targets across devices for $P = 1000$ replicates, refitting the full pipeline for each permutation. The corrected one-sided p-value was computed as

$$p = \frac{1 + \#\{\, b : E_b \leq E_{\mathrm{obs}} \}}{1 + P},$$

preserving intra-device structure during reshuffling [37, 38].

4 Results and discussion

4.1 Target distribution, and robustness

The device-level instability target $\bar{s}_1$ was computed for each device as the mean of valid even-index DC sweeps after trace-quality filtering. A total of 32 devices satisfied the minimum requirements for stable score evaluation.

To reduce sensitivity to rare noise-driven extremes while preventing information leakage, we applied a robust trimming procedure within each training fold based on the median absolute deviation (MAD) together with a high-percentile cap. This approach removes a small number of extreme values while preserving the dynamic range required for regression modeling. Figure 4 compares the raw and trimmed target distributions of $\bar{s}_1$. The trimmed distribution remains approximately unimodal with a moderate spread, indicating that the instability metric retains sufficient variance for predictive modeling.

The stability of $\bar{s}_1$ was evaluated using test-retest comparisons, Bland–Altman analysis, and intraclass correlation (ICC). The resulting agreement metrics demonstrate low systematic bias and narrow limits of agreement (Fig. 4c), indicating that the score provides a robust summary of device-level electrical symmetry.

Sensitivity to threshold parameters ($U_{\mathrm{skip}}, I_{\mathrm{min}}$) was further examined through grid-based perturbation analysis. Across a broad range of parameter choices, the rank ordering of devices remains stable with Spearman correlations $\rho \gtrsim 0.85$ (Fig. 4d). This result confirms that the instability metric is not strongly dependent on specific guard-threshold settings.

4.2 Predictive value of multimodal features

We evaluated predictive performance using device-grouped five-fold cross-validation, ensuring that all rows from the same physical device were assigned to the same fold. This analysis provides the primary test of whether multimodal fusion adds information beyond individual measurement modalities. Table 1 compares SEM-only, DC-only, EIS-only, and fused SEM + DC + EIS predictors under the same validation protocol.

Among the single-modality predictors, EIS-derived descriptors provided the strongest signal, with an RMSE

Table 1 Device-grouped fivefold cross-validation comparing unimodal and fused predictors in the same-batch setting

Model (features)	RMSE	
SEM only (random forest)	0.369	0.098
DC only (random forest)	0.339	0.071
EIS only (random forest)	0.247	0.097
Fusion: SEM + DC + EIS (RF)	**0.152**	**0.037**

Fusion of SEM, DC, and EIS descriptors yields the lowest mean RMSE across folds

It indicates the lowest mean RMSE and therefore the best predictive performance among the models compared

of 0.247 ± 0.097. SEM-only and DC-only models showed weaker performance, with RMSE values of 0.369 ± 0.098 and 0.339 ± 0.071, respectively. The fused SEM + DC + EIS random-forest model achieved the lowest error, with an RMSE of 0.152 ± 0.037, corresponding to an approximately 38% reduction relative to the strongest single-modality baseline. This comparison, rather than the train-median baseline, is the main evidence that the modalities provide complementary predictive information.

Although the instability target is derived from DC measurements, the DC-only model is not a direct reconstruction of the target. The target s_1 is computed from even-indexed DC sweeps, whereas the DC predictors are extracted from disjoint odd-indexed sweeps. Moreover, s_1 quantifies reciprocal quarter-sweep symmetry, while the DC feature set summarizes nonlinearity, hysteresis area, ON/OFF separation, and the composite Q_{DC}. Thus, the DC descriptors are informative but incomplete proxies for the target.

4.3 Cross-batch generalization

To evaluate robustness under fabrication-dependent domain shift, we performed leave-one-batch-out (LOBO) validation. In each fold, one complete fabrication batch was held out for testing, while preprocessing, imputation, scaling, dimensionality reduction, hyperparameter tuning, and model fitting were restricted to the remaining batches. This design prevents information from the held-out batch from influencing the fitted pipeline.

Table 3 summarizes the leave-one-batch-out comparison of single-modality and fused feature sets. Consistent with the device-grouped validation results, the fused SEM + DC + EIS representation achieves the lowest LOBO error, indicating that multimodal descriptors retain complementary predictive value even when an entire fabrication batch is withheld.

Pooled across held-out device-level predictions, the best-performing model achieved an RMSE of 0.152 with a 95% confidence interval of [0.060, 0.244] (Table 2). Alternative models, including Gaussian-process regression and kernel ridge regression, yielded overlapping confidence intervals, suggesting that performance is constrained primarily by dataset size and batch-to-batch variability rather than by the choice of regressor alone.

Table 2 Pooled RMSE computed over all held-out predictions across LOBO folds

Model	RMSE	95% CI	NRMSE
RF (batch-offset)	0.152	[0.060, 0.244]	0.238
GPR	0.170	[0.089, 0.252]	0.187
RF	0.182	[0.086, 0.269]	0.200
KRR (tuned)	0.237	[0.096, 0.385]	0.261
HGB/GB	0.295	[0.195, 0.388]	0.324

In the LOBO setting, we additionally compared the multimodal model with a train-median baseline. This baseline is not intended to demonstrate the value of multimodal fusion over individual modalities. Instead, it provides a conservative leakage-controlled reference for asking whether a model trained on the remaining batches transfers useful signal to a completely held-out batch beyond the central tendency of the training targets. The main evidence for multimodal benefit remains the comparison against SEM-only, DC-only, and EIS-only predictors in the device-grouped validation setting shown in Table 1.

Batch-wise RMSE values for the evaluated models are shown in Fig. 5. Performance relative to the train-median baseline is shown as per-batch ΔRMSE in Fig. 6 and numerically in Table 4. Improvement is defined as

$$\Delta\mathrm{RMSE} = \mathrm{RMSE}_{\mathrm{baseline}} - \mathrm{RMSE}_{\mathrm{model}},$$

so positive values indicate lower error for the multimodal model. Under this convention, the multimodal model improves performance in six of eight held-out batches, while two batches show degradation. The magnitude of improvement varies considerably across batches, indicating that the learned relationship between microstructural, impedance, and DC descriptors transfers unevenly across fabrication conditions.

Figures 6, 7, 8 and 9 summarize the LOBO results using complementary views: per-batch improvement relative to the train-median baseline, a forest-style paired comparison of model and baseline RMSE, grouped permutation testing, and bootstrap uncertainty estimates. These analyses support a modest average cross-batch benefit of multimodal fusion while emphasizing substantial uncertainty caused by the limited number of fabrication batches.

4.4 Statistical significance of multimodal gains

To determine whether the observed improvements could arise by chance, we conducted grouped permutation testing while preserving the device-level validation structure.

The observed weighted mean improvement in RMSE lies in the lower tail of the permutation distribution (Fig. 7), indicating that the performance gain of the multimodal model is unlikely under the null hypothesis of no relationship between descriptors and the instability metric.

Bootstrap resampling across batches (Fig. 9) further quantifies the uncertainty associated with cross-batch performance. While the mean improvement remains positive, the confidence interval reflects the limited number of fabrication batches available in the present dataset.

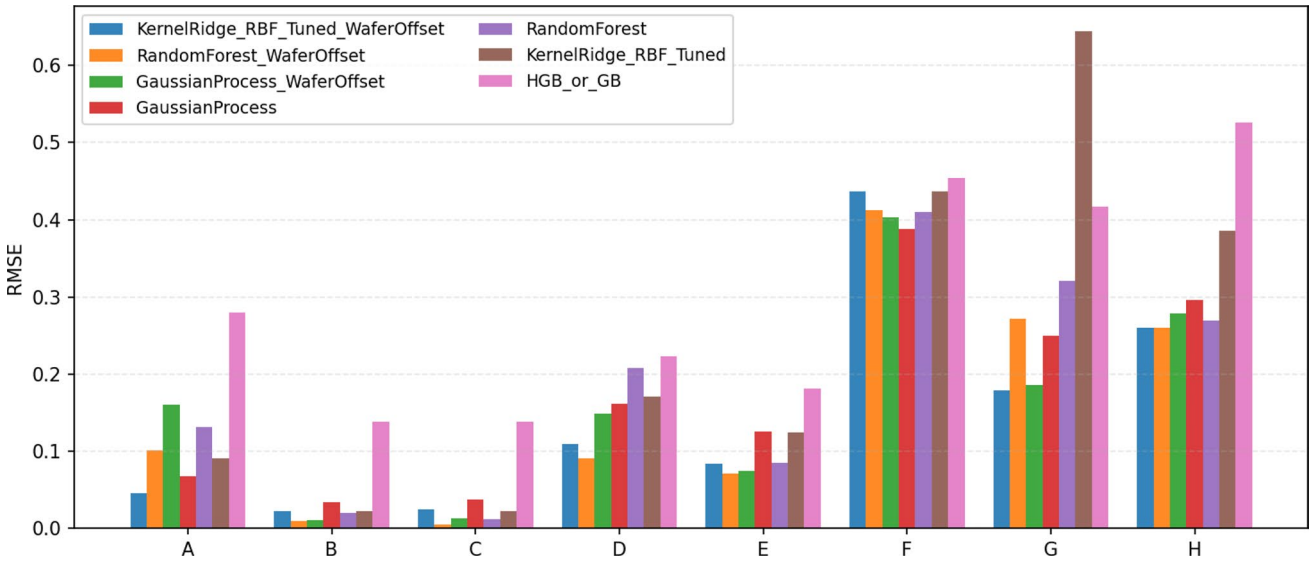


Fig. 5 per-batch RMSE for all models. Bars are grouped by batches, illustrating substantial batch-to-batch variability in prediction difficulty and relative model performance

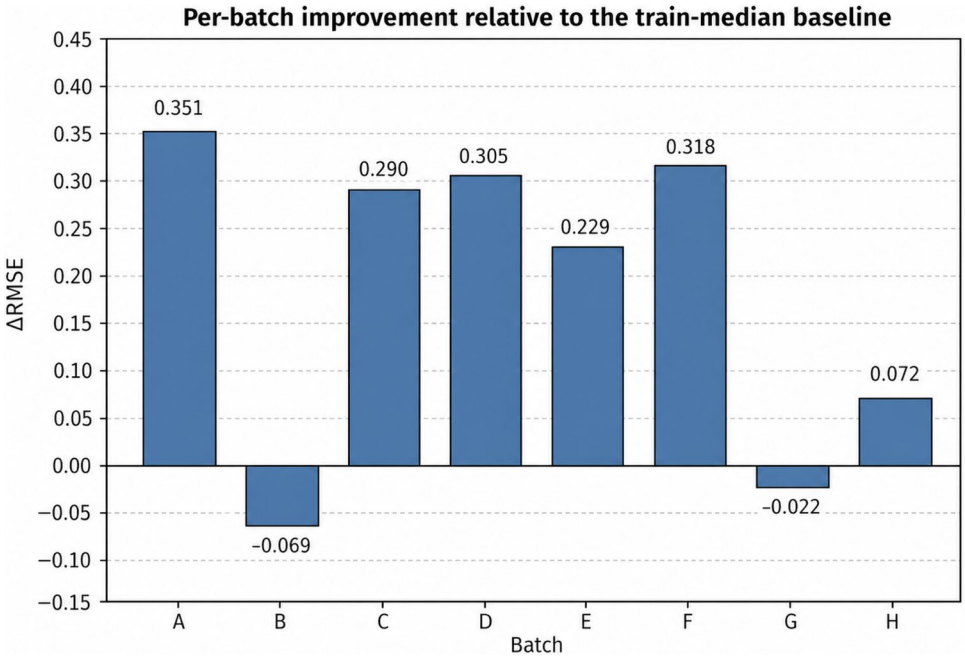


Fig. 6 Relative to the train-median baseline, the multimodal model improves performance in six of eight held-out batches, but the gains are heterogeneous and two batches show degradation. This indicates that the learned structure–property relationship transfers unevenly across fabrication batches

Table 3 Leave-one-batch-out comparison of single-modality and fused predictors

Feature set	RMSE	95% CI	NRMSE
Train-median baseline	0.322	[0.18, 0.50]	0.461
SEM only	0.281	[0.15, 0.45]	0.405
DC only	0.228	[0.12, 0.35]	0.307
EIS only	0.184	[0.09, 0.30]	0.259
SEM + DC + EIS	0.146	[0.06, 0.23]	0.181

Together, these results indicate that multimodal fusion provides a modest average predictive advantage under realistic cross-batch transfer conditions, while the batch-wise results show that this benefit is not uniform across all held-out batches.

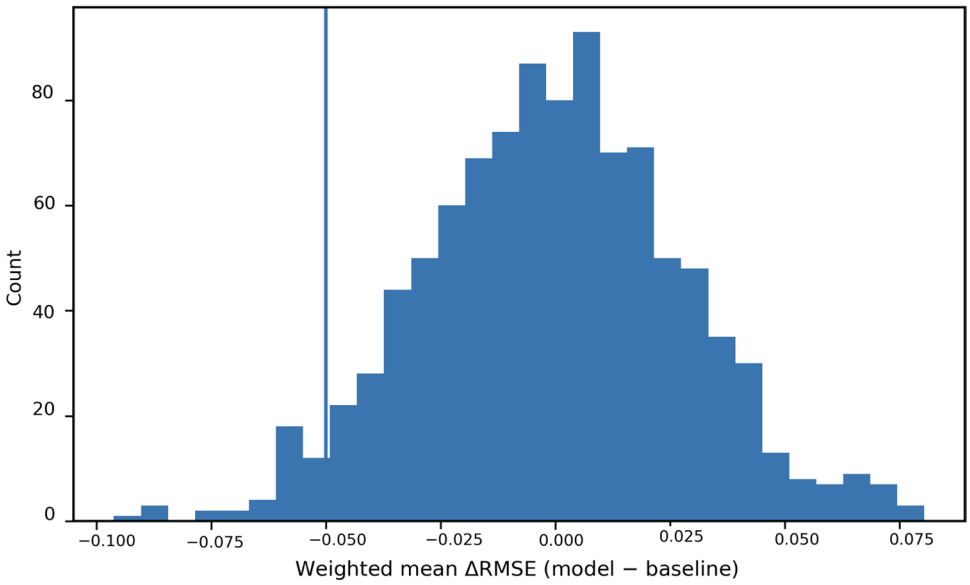


Fig. 7 Permutation test on the weighted mean ΔRMSE in the LOBO setting. The observed statistic lies in the lower tail of the null distribution, confirming real but modest cross-batch gains

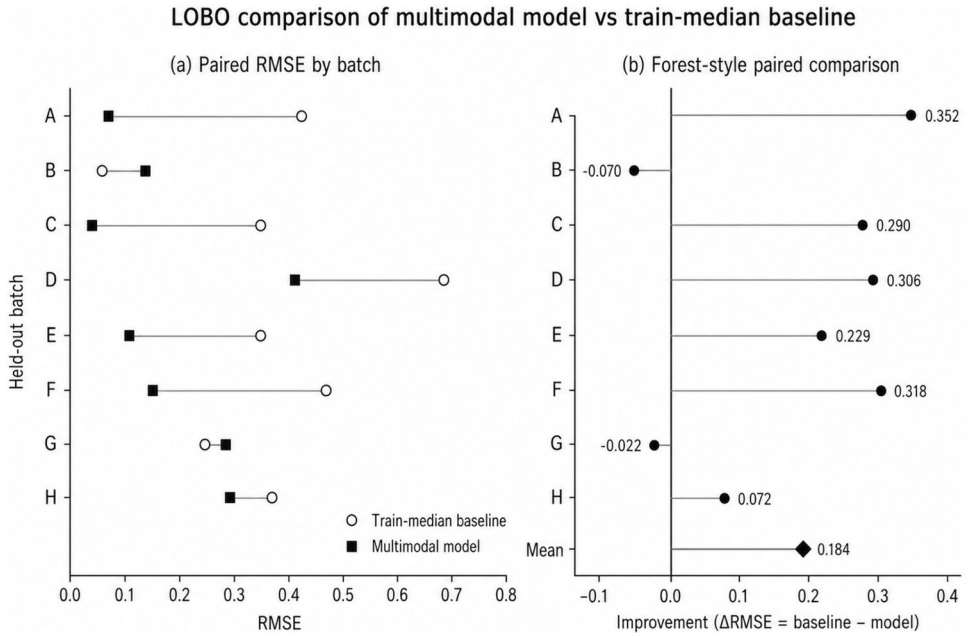


Fig. 8 Forest-style paired comparison of train-median baseline RMSE and multimodal-model RMSE across held-out batches in LOBO validation. **a** shows paired RMSE values for each held-out batch. **b** shows the corresponding improvement, defined as $\Delta\text{RMSE} = \text{RMSE}_{\text{baseline}} - \text{RMSE}_{\text{model}}$; positive values indicate lower RMSE for the multimodal model. The diamond denotes the mean improvement across batches

Table 4 batch-wise RMSE compared with a train-median baseline (original scale)

Batch	RMSE (model)	RMSE (baseline)	Improvement
A	0.061	0.413	0.351
B	0.124	0.054	−0.069
C	0.035	0.325	0.290
D	0.395	0.701	0.305
E	0.097	0.326	0.229
F	0.140	0.458	0.318
G	0.258	0.236	−0.022
H	0.277	0.349	0.072

Improvement = baseline − model

4.5 Drivers of performance

We analyzed the fitted random-forest model using grouped permutation importance to identify descriptor groups that contributed to predictive performance. This analysis should be interpreted as model-level attribution, not as a physical decomposition of the instability mechanism. In particular, several descriptors are correlated within and across modalities, the feature space is high-dimensional relative to the number of devices, and feature interactions cannot be fully resolved at this data scale. Therefore, the attribution results are used to identify physically plausible

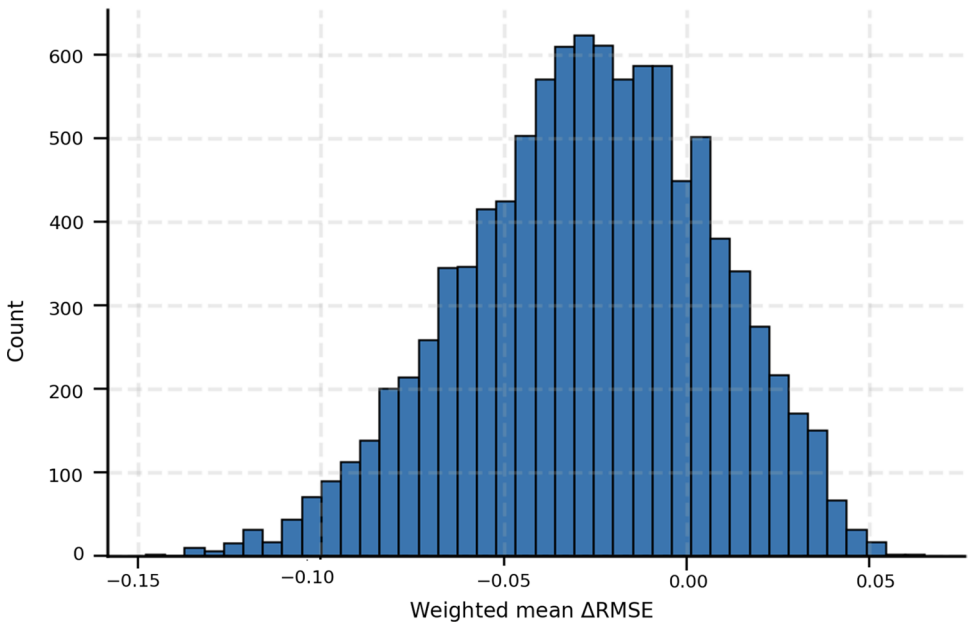


Fig. 9 Bootstrap distribution of the weighted mean ΔRMSE across batches, estimated from 10,000 bootstrap resamples, showing uncertainty in cross-batch improvements

and hypothesis-generating descriptor groups, not to assign unique causal roles to individual features.

The interpretability of the workflow arises from the use of predefined, physically motivated descriptors rather than from the permutation-importance ranking alone. The DC descriptors summarize measurable aspects of the current–voltage response, including nonlinear transport, hysteresis area, and ON/OFF separation. These quantities are directly related to switching asymmetry and therefore provide an interpretable electrical summary of device behavior. The EIS descriptors summarize frequency-dependent impedance magnitude and phase, including changes under UV illumination, and are therefore interpreted as indicators of illumination-sensitive interfacial or trap-mediated response. The SEM descriptors summarize registered pre-/post-stress image changes through intensity, texture, morphology, and Fourier-domain statistics. These descriptors quantify spatial contrast changes and characteristic length-scale content in the SEM difference maps, but they should not be interpreted as direct observations of a unique microscopic degradation pathway.

Permutation importance (Fig. 10) identifies the composite DC descriptor Q_{DC} as the dominant predictor. This is physically plausible because Q_{DC} summarizes nonlinear transport, hysteresis area, and ON/OFF separation, all of which are directly related to switching asymmetry and instability in the measured current–voltage response. The importance of DC nonlinearities is consistent with the same interpretation:

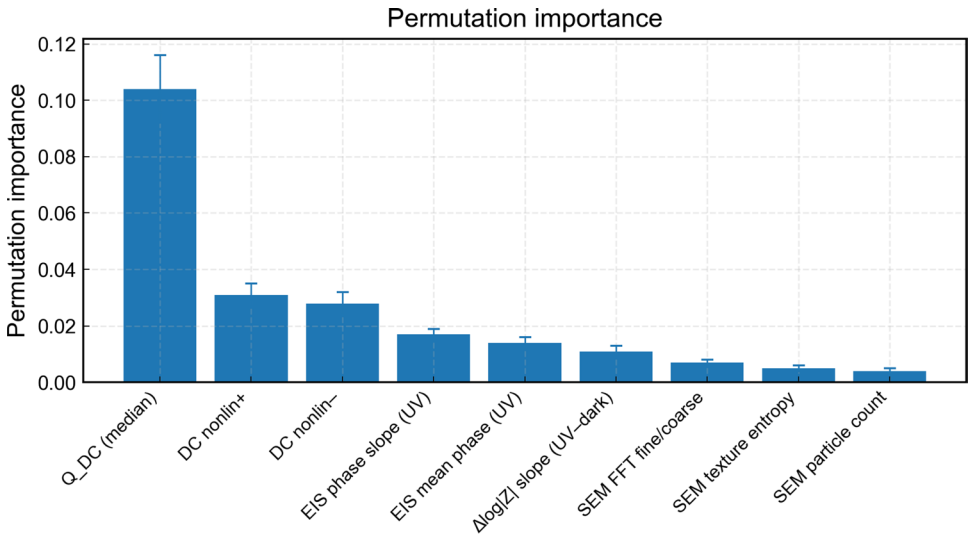


Fig. 10 Permutation importance for the random-forest model across device-grouped folds. DC, EIS, and SEM features all contribute, with Q_{DC} emerging as the most influential

devices with stronger bias-dependent transport asymmetry tend to carry more information about the quarter-sweep symmetry score $\bar{s}_1$.

EIS descriptors provide additional predictive signal, particularly UV-induced phase slopes and magnitude-difference metrics. These features quantify frequency-dependent impedance response and its change under illumination. We therefore interpret them as indicators of interfacial or trap-mediated electrical response, rather than as direct proof of a specific redox or defect mechanism. Their contribution supports the view that the instability proxy is not captured by DC hysteresis alone, but also reflects illumination-sensitive interfacial dynamics.

SEM-derived descriptors, including FFT fine-to-coarse ratios, texture entropy, and morphology counts, contribute smaller but non-negligible information. These descriptors summarize changes in spatial contrast, characteristic length-scale content, and thresholded morphological heterogeneity in registered pre-/post-stress SEM difference maps. Their attribution is consistent with stress-induced microstructural evolution, but it does not uniquely identify the underlying structural process. For example, a change in the FFT fine-to-coarse ratio may reflect altered fine-scale contrast or roughness, while particle-count-like morphology descriptors depend on the thresholded difference-map representation and should therefore be interpreted as quantitative image-change summaries rather than literal particle counts.

Representative bivariate relationships with $\bar{s}_1$ are shown in Fig. 11, illustrating that no single descriptor fully explains the instability metric. Instead, the fused model benefits from combining partially complementary structural, electrical, and impedance-derived information. Image-only convolutional models underperform at this data scale, indicating that the observed fusion benefit arises from domain-relevant engineered descriptors rather than from unconstrained raw-image learning.

5 Conclusion and future work

This work presents a multimodal framework for analyzing the relationship between microstructural evolution and electrical instability in oxide memristive devices. By combining SEM difference maps with descriptors derived from DC current–voltage hysteresis and illumination-sensitive electrochemical impedance spectroscopy, the proposed workflow enables joint analysis of structural and electrical signatures of device behavior. Across device-grouped and cross-batch validation settings, multimodal feature fusion improves predictive accuracy relative to single-modality models while retaining descriptor-level transparency through predefined, physically motivated features.

Feature attribution analysis indicates that electrical hysteresis metrics provide the strongest predictive signal, while EIS-derived phase slopes and SEM-derived texture descriptors contribute complementary, physically motivated information consistent with illumination-sensitive interfacial responses and stress-associated image changes. These results suggest that electrical instability in oxide memristors reflects coupled structural and electrochemical processes that can be captured through combined multimodal characterization.

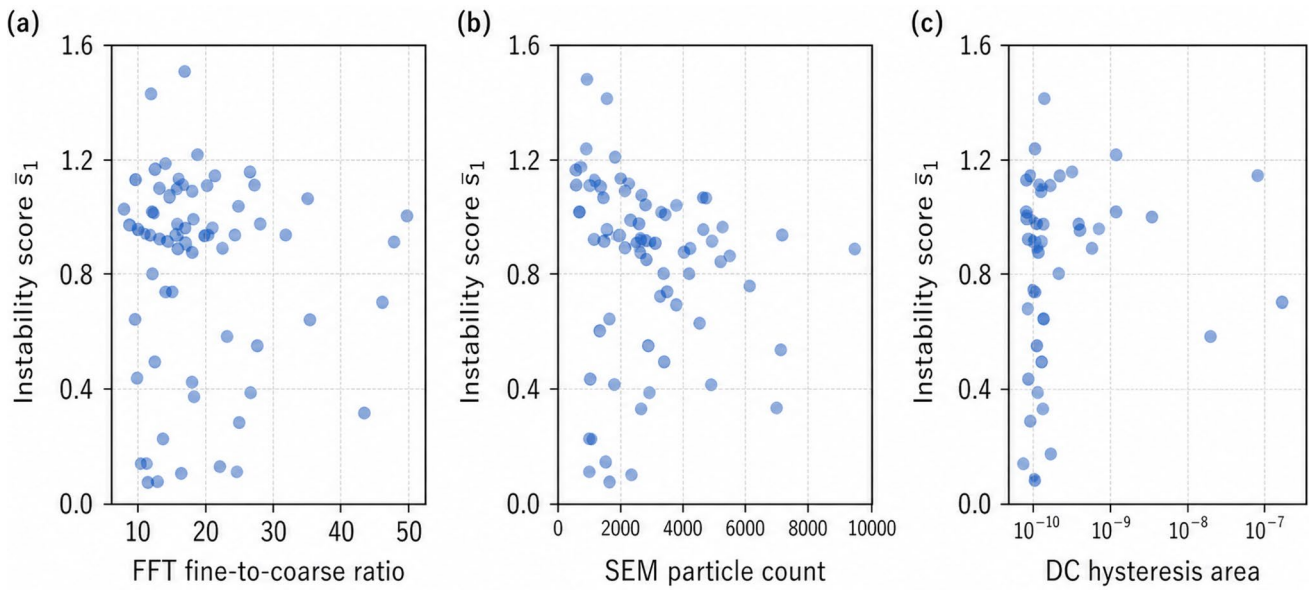


Fig. 11 Representative bivariate associations between the device-level instability score $\bar{s}_1$ and selected descriptors from the fused feature set: **a** SEM FFT fine-to-coarse ratio, **b** SEM particle count, and **c** DC hysteresis area

The present study is limited by the modest size of the available dataset and by fabrication-dependent variability across batches, which constrain the generalizability of predictive models. Consequently, the models should be interpreted primarily as statistical probes for identifying multimodal signatures of instability rather than as deployable predictors.

Future work will focus on expanding the dataset with additional fabrication batches and incorporating controlled variations in stress protocols and device architectures. Such datasets will enable more detailed modeling of degradation pathways and facilitate the development of physics-informed predictive tools for reliability screening. More broadly, the presented framework illustrates how multimodal characterization combined with leakage-aware statistical analysis can support interpretable reliability assessment in memristive and other oxide-based electronic materials.

Appendix A: Notation (for reference)

Symbol	Meaning
U [V]	Applied voltage
I [A]	Measured current
$R = U/I$ [Ω]	Local resistance
f [Hz]	Frequency (EIS)
$\angle Z$ [rad]	Impedance phase
U_{skip} [V]	Bias window excluded near 0 V
$I_{\min}$ [A]	Current floor for stability
A_h [A V]	Hysteresis area
ON/OFF [–]	Median-resistance ratio
Q_{DC} [–]	Composite DC index
s_1, $\bar{s}_1$ [–]	Quarter-sweep symmetry (sweep / device)
ℓ [–]	Trimming fraction
$\mathcal{K}$	Valid symmetric-pair index set

Acknowledgements This work was supported by the Carl-Zeiss Foundation via Project Memwerk. Additional support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project-ID 434434223 SFB 1461 is acknowledged. Additional funding was provided by the BMBF project ForLab NSME (Grant No. 16ES0939). Support from the Center of Micro- and Nanotechnologies (ZMN), a DFG-funded core facility of TU Ilmenau, is gratefully acknowledged.

Author Contributions S.Z. performed the analysis, prepared the figures, and wrote the manuscript. M.B. and H.H. provided the data and reviewed the manuscript. E.T. assisted with programming. P.S. and K.-U.S. supervised the work throughout all stages. All authors reviewed and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. the Carl-Zeiss Foundation via Project Memwerk. Additional support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project-ID 434434223 SFB 1461 is acknowledged. Additional funding was provided by the BMBF project ForLab NSME (Grant No. 16ES0939). Support from the Center of Micro- and Nanotechnologies (ZMN), a DFG-funded core facility of TU Ilmenau, is gratefully acknowledged.

Data Availability A single representative example of the multimodal dataset, including registered pre-and post-stress SEM images and the corresponding extracted DC and EIS features, is publicly available for transparency and methodological illustration. The full datasets generated and analyzed during this study are not publicly available due to institutional and project-level restrictions but can be discussed with the corresponding author upon reasonable request.

Code Availability All analysis scripts and workflows for multimodal memristor modeling–including guarded $\bar{s}_1$ computation, EIS/SEM feature extraction, and leave-one-batch-out validation–are available at: https://github.com/memwerk/memristor-multimodal-modeling

Declarations

Conflict of interest The authors declare no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Nirmal, K.A., Kumbhar, D.D., Kesavan, A.V., Dongale, T.D., Kim, T.G.: Advancements in 2d layered material memristors: unleashing their potential beyond memory. NPJ 2D Mater. Appl. **8**(1), 83 (2024). https://doi.org/10.1038/s41699-024-00522-4
2. Jaafar, A.H., Al Habsi, S.K.S., Braben, T., Venables, C., Francesconi, M.G., Stasiuk, G.J., Kemp, N.T.: Unique coexistence of two resistive switching modes in a memristor device enables multifunctional neuromorphic computing properties. ACS Appl. Mater. Interfaces. **16**(33), 43816–43826 (2024). https://doi.org/10.1021/acsami.4c07820
3. Wong, H.-S.P., Lee, H.-Y., Yu, S., Chen, Y.-S., Wu, Y., Chen, P.-S., Lee, B., Chen, F.T., Tsai, M.-J.: Metal-oxide RRAM. Proc. IEEE **100**(6), 1951–1970 (2012). https://doi.org/10.1109/JPROC.2012.2190369
4. Ielmini, D.: Resistive switching memories based on metal oxides: mechanisms, reliability and scaling. Semicond. Sci. Technol. **31**(6), 063002 (2016). https://doi.org/10.1088/0268-1242/31/6/063002
5. Yang, J.J., Strukov, D.B., Stewart, D.R.: Memristive devices for computing. Nat. Nanotechnol. **8**, 13–24 (2013). https://doi.org/10.1038/nnano.2012.240
6. Waser, R., Aono, M.: Nanoionics-based resistive switching memories. Nat. Mater. **6**(11), 833–840 (2007). https://doi.org/10.1038/nmat2023

7. Yang, J.J., Pickett, M.D., Li, X., Ohlberg, D.A.A., Stewart, D.R., Williams, R.S.: Memristive switching mechanism for metal/oxide/metal nanodevices. Nat. Nanotechnol. **3**(7), 429–433 (2008). https://doi.org/10.1038/nnano.2008.160
8. Celano, U., Goux, L., Belmonte, A., Opsomer, K., Franquet, A., Schulze, A., Detavernier, C., Richard, O., Bender, H., Jurczak, M., Vandervorst, W.: Three-dimensional observation of the conductive filament in nanoscaled resistive memory devices. Nano Lett. **14**(5), 2401–2406 (2014). https://doi.org/10.1021/nl500049g
9. Sun, W., Gao, B., Chi, M., Xia, Q., Yang, J.J., Qian, H., Wu, H.: Understanding memristive switching via in situ characterization and device modeling. Nat. Commun. **10**(1), 3453 (2019). https://doi.org/10.1038/s41467-019-11411-6
10. Marquardt, R., Zahari, F., Carstensen, J., Popkirov, G., Gronenberg, O., Kolhatkar, G., Kohlstedt, H., Ziegler, M.: Impedance spectroscopy on hafnium oxide-based memristive devices. Adv. Electron. Mater. **9**(6), 2201227 (2023). https://doi.org/10.1002/aelm.202201227
11. Barsukov, Y., Macdonald, J.: Electrochemical impedance spectroscopy. In: Kaufmann. E.N. (ed.) Characterization of Materials. Wiley (2012). https://doi.org/10.1002/0471266965
12. Chua, L.O.: Memristor-the missing circuit element. IEEE Trans. Circuit Theory **18**(5), 507–519 (1971). https://doi.org/10.1109/TCT.1971.1083337
13. Strukov, D.B., Snider, G.S., Stewart, D.R., Williams, R.S.: The missing memristor found. Nature **453**(7191), 80–83 (2008). https://doi.org/10.1038/nature06932
14. Waser, R., Dittmann, R., Staikov, G., Szot, K.: Redox–based resistive switching memories—nanoionic mechanisms, prospects, and challenges. Adv. Mater. **21**(25–26), 2632–2663 (2009). https://doi.org/10.1002/adma.200900375
15. Qian, K., Cai, G., Nguyen, V., Chen, T., Lee, P.S.: Direct observation of conducting filaments in tungsten oxide based transparent resistive switching memory. ACS Appl. Mater. Interfaces. **8**, 27885–27891 (2016). https://doi.org/10.1021/acsami.6b08154
16. Valov, I.: Redox-based resistive switching memories (RERAMs): electrochemical systems at the atomic scale. ChemElectroChem **1**(1), 26–36 (2014). https://doi.org/10.1002/celc.201300165
17. Li, Y., Le Pape, Y., Rodriguez, E.T., Torrence, C., Mena, J.A., Rosseel, T., Sircar, M.: Microstructural characterization and assessment of mechanical properties of concrete based on combined elemental analysis techniques and fast-Fourier transform–based simulations. Constr. Build. Mater. **257**, 119500 (2020). https://doi.org/10.1016/j.conbuildmat.2020.119500
18. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Trans. Syst., Man, Cybernet. SMC **3**(6), 610–621 (1973). https://doi.org/10.1109/TSMC.1973.4309314
19. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002). https://doi.org/10.1109/TPAMI.2002.1017623
20. Rublee, E., Rabaud, V., Konolige, K., Bradski, G.: Orb: An efficient alternative to sift or surf. In: 2011 International Conference on Computer Vision, pp. 2564–2571 (2011). https://doi.org/10.1109/ICCV.2011.6126544
21. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. In: Fischler, M.A., Firschein, O. (eds.) Readings in Computer Vision, pp. 726–740. Morgan Kaufmann, San Francisco (1987). https://doi.org/10.1016/B978-0-08-051581-6.50070-2
22. Lowe, D.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**, 91 (2004). https://doi.org/10.1023/B:VISI.0000029664.99615.94
23. Xiao, Y., Jiang, B., Zhang, Z., Ke, S., Jin, Y., Wen, X., Ye, C.: A review of memristor: material and structure design, device performance, applications and prospects. Sci. Technol. Adv. Mater. **24**(1), 2162323 (2023). https://doi.org/10.1080/14686996.2022.2162323
24. Barnes, B.K., Das, K.S.: Resistance switching and memristive hysteresis in visible-light-activated adsorbed ZNO thin films. Sci. Rep. **8**(1), 2184 (2018). https://doi.org/10.1038/s41598-018-20598-5
25. Macdonald, J.: Impedance spectroscopy. Ann. Biomed. Eng. **20**, 289–305 (1992). https://doi.org/10.1007/BF02368532
26. Barsoukov, E., Macdonald, J.R. (eds.): Impedance Spectroscopy: Theory, Experiment, and Applications, 3rd edn. Wiley (2018). https://doi.org/10.1002/9781119381860
27. DeCost, B.L., Francis, T., Holm, E.A.: Exploring the microstructure manifold: image texture representations applied to ultrahigh carbon steel microstructures. Acta Mater. **133**, 30–40 (2017). https://doi.org/10.1016/j.actamat.2017.05.014
28. Zhang, Y., Ling, C.: A strategy to apply machine learning to small datasets in materials science. NPJ Comput. Mater. **4**(1), 25 (2018). https://doi.org/10.1038/s41524-018-0081-z
29. Ramprasad, R., Batra, R., Pilania, G., Mannodi-Kanakkithodi, A., Kim, C.: Machine learning in materials informatics: recent applications and prospects. NPJ Comput. Mater. **3**(1), 54 (2017). https://doi.org/10.1038/s41524-017-0056-5
30. Wang, S., Chen, X., Zhao, C., Kong, Y., Lin, B., Wu, Y., Bi, Z., Xuan, Z., Li, T., Li, Y., Zhang, W., Ma, E., Wang, Z., Ma, W.: An organic electrochemical transistor for multi–modal sensing, memory and processing. Nat. Electron. **6**(4), 281–291 (2023). https://doi.org/10.1038/s41928-023-00950-y
31. Lyu, Y., Chen, W., Zheng, M., Yin, B., Li, J., Cai, L.: Machine learning-assisted device modeling with process variations for advanced technology. IEEE J. Electron Dev. Soc. **11**, 303–310 (2023). https://doi.org/10.1109/JEDS.2023.3277548
32. Dongale, T.D., Sutar, S.S., Dange, Y.D., Khot, A.C., Kundale, S.S., Patil, S.R., Patil, S.V., Patil, A.A., Khot, S.S., Patil, P.J., et al.: Machine learning-assisted design guidelines and performance prediction of CMOS-compatible metal oxide-based resistive switching memory devices. Appl. Mater. Today **29**, 101650 (2022). https://doi.org/10.1016/j.apmt.2022.101650
33. Roberts, D., Bahn, V., Ciuti, S., Boyce, M., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J., Schröder, B., Thuiller, W., Warton, D., Wintle, B., Hartig, F., Dormann, C.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography (2016). https://doi.org/10.1111/ecog.02881
34. Winkler, A.M., Webster, M.A., Vidaurre, D., Nichols, T.E., Smith, S.M.: Multi-level block permutation. Neuroimage **123**, 253–268 (2015). https://doi.org/10.1016/j.neuroimage.2015.05.092
35. Huber, P.J., Ronchetti, E.M.: Robust Statistics, 2nd edn. Wiley, Hoboken (2009)
36. Wilcox, R.R.: Introduction to Robust Estimation and Hypothesis Testing, 5th edn. Academic Press, San Diego (2021)
37. Good, P.I.: Permutation, Parametric, and Bootstrap Tests of Hypotheses, 3rd edn. Springer, New York (2005). https://doi.org/10.1007/b138696
38. Efron, B., Tibshirani, R.J.: An Introduction to the Bootstrap, Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, New York (1993)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.