iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

CRYPTOGRAPHICALLY-ATTESTED BATTERY STATE-OF-HEALTH ESTIMATION ON STOCK CONSUMER ANDROID VIA CHARGING-CURVE REPRESENTATION LEARNING

Arun Teja Sara
Retronics Market Place UK Limited, United Kingdom
Corresponding author: info@retronixs.com

ABSTRACT
The secondary smartphone market is undermined by "lemon" devices with degraded batteries masquerading as healthy. We present TVO (Temporal Voltage Oracle), a system for State-of-Health (SoH) estimation that encodes charging-curve telemetry as Gramian Angular Field (GAF) images on stock, unrooted consumer Android and estimates SoH via a deployable constrained-binary estimator pipeline. The currently deployed v1 is a closed-form analytic model; a gradient-boosted regressor exported to ONNX is evaluated on-device as the next production candidate; a full CNN-backbone path is validated offline pending ONNX Runtime vendoring.
Our primary contribution is PIACS-RDA (Physics-Informed Aging-Curve Synthesis with Real-Device Anchoring), a training-data methodology that augments healthy charging captures from 347 physical devices into a robust, labeled aging corpus by applying electrochemical degradation transforms grounded in real measured aging patterns. All devices were tested at the Retronics Lab facility; ground-truth SoH was established via calibrated benchtop cyclers for every device using a standardised cycler protocol (0.5C CC-CV, IEC 61960 capacity definition).
We further contribute a degradation-knob prediction layer that estimates each of four synthetic aging parameters separately and composes them, supporting interpretable per-factor attribution rather than an opaque scalar verdict. The four predicted knobs are the same parameters used during PIACS-RDA data augmentation; they are validated against their synthesis targets, not against independently measured electrochemical mechanisms.
On group-aware five-fold cross-validation with device-level grouping, preventing data leakage from augmented variants, the gradient-boosted candidate achieves MAPE 4.21 ± 0.09%; the deployed closed-form analytic estimator achieves MAPE 8.24 ± 0.70%; an offline-evaluated CNN backbone achieves MAPE 3.85 ± 0.07%. All MAPE values are computed on original real device traces validated against independent cycler ground truth; synthetic augmented samples are used for training only and never appear in test folds. Finally, we provide a TEE-rooted cryptographic envelope that signs the SoH verdict, ensuring the integrity of health disclosures in refurbishment workflows.

Keywords:
Battery State-of-Health estimation; Gramian Angular Field; charging-curve representation learning; TEE attestation; smartphone refurbishment; real-device validation; degradation-knob prediction; group cross-validation.

1. INTRODUCTION

The secondary market for consumer smartphones is bottlenecked by trust: buyers cannot independently verify the health of a battery whose label claims "93%" but whose seller controls every diagnostic surface between battery and screen. Manufacturer SoH disclosures are uneven across vendors, and third-party diagnostic apps either require root access, a non-starter for warranty-preserving refurbishment, or rely on opaque heuristics that an adversary could trivially manipulate by replaying a captured screenshot.
A complementary literature, primarily in Journal of Power Sources, Journal of Energy Storage, and Applied Energy, has established that lithium-ion State-of-Health (SoH) can be estimated to 1–5% MAPE by encoding the charging-curve time series into a 2D image via the Gramian Angular Field (GAF) of Wang and Oates [2015] and applying transfer learning from a pre-trained ResNet- or MobileNet-class backbone [Liu et al., 2025, Wei and Wu, 2024, Zhang et al., 2023]. These results are obtained on laboratory cyclers (Arbin, Maccor, Neware) with 1 Hz voltage and current sampling at sub-mV/sub-mA resolution, controlled temperature, fixed starting state of charge, and a known charging protocol on cylindrical 18650 cells. None of those conditions hold on a stock unrooted Android phone in the hands of a refurbishment seller.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

1.1 Contributions

This paper presents TVO (Temporal Voltage Oracle), an end-to-end system that brings GAF-based charging-curve representation learning for SoH estimation to stock consumer Android, signs the result with a TEE-rooted attestation chain, and quantifies the accuracy gap between published laboratory numbers and what stock telemetry actually delivers, all validated on a cohort of 347 real devices tested at the Retronics Lab facility. Our contributions are:

1) Cross-chipset evaluation on 347 real devices with cycler ground truth. We evaluate TVO on devices spanning three chipset families (Snapdragon 888, Exynos 2100, Google Tensor), two battery chemistries (LCO, NMC), and four commercially relevant device models. Ground-truth SoH for every device is established via calibrated benchtop cyclers at the Retronics Lab using a standardised protocol. This represents the largest published validation cohort for GAF-based SoH estimation on stock Android.
2) PIACS-RDA: Physics-Informed Aging-Curve Synthesis with Real-Device Anchoring. A training-data methodology that augments healthy charging captures from real devices using physics-informed electrochemical transforms calibrated against measured aged-cell behaviour. All training examples are rooted in physical device captures, ensuring the augmented corpus retains authentic noise floors and hardware artefacts. Critically, we validate using group-aware leave-one-block-out (LOBO) five-fold cross-validation with device-level grouping that prevents data leakage between augmented variants of the same physical phone.
3) A three-tier estimator architecture spanning the accuracy-deployment trade-off: (a) a closed-form logistic estimator (∼120 KB, MAPE 8.24%) deployed in production v1; (b) a gradient-boosted regressor exported to ONNX (∼2.4 MB, MAPE 4.21%) evaluated on-device as the next production candidate; and (c) a CNN backbone operating on the full 224 × 224 GAF image (∼8 MB, MAPE 3.85%) validated offline pending ONNX Runtime vendoring.
4) A degradation-knob prediction layer that estimates each of the four PIACS-RDA augmentation parameters separately and composes them into an interpretable SoH verdict with per-factor attribution. The predicted values are validated against their synthesis targets, not against independently measured physical degradation mechanisms.
5) A lab-to-field error budget that quantifies the four mechanisms degrading on-phone accuracy relative to lab-cycler studies.
6) A cryptographic-attestation envelope that binds the SoH verdict to a per-run TEE-sealed key, with canonical-JSON serialisation verifiable from the backend.

1.2 Scope

We deliberately do not claim sub-5% MAPE for the closed-form analytic estimator. We do not claim function under extreme temperatures or wireless charging; both are gated by the on-device test-quality refusal layer. The 10-minute test window extends to 20 minutes when warranted, and aborts if a clean CC-CV transition is not observed.

2. BACKGROUND AND RELATED WORK

2.1 Time-Series Imaging

Wang and Oates [2015] introduced the Gramian Angular Summation Field (GASF), Gramian Angular Difference Field (GADF), and Markov Transition Field (MTF) as encodings of a univariate time series into a 2D image suitable for convolutional classifiers. The IJCAI-2015 paper is the foundational reference for image-encoding-of-time-series.

2.2 Battery SoH from Charging Curves

Wei and Wu [2024] report RMSE 0.0112 with GAF + 2D-CNN + BiLSTM; Zhang et al. [2023] screen retired packs with GADF + ConvNeXt at greater than 97% binary accuracy; Liu et al. [2025] use deep transfer learning across cell formats. Chemali et al. [2018] provide a comprehensive survey of machine-learning approaches to battery state estimation, categorising methods by input feature type, model architecture, and validation protocol. Richardson et al. [2019] survey Gaussian process and deep-learning methods for battery health forecasting. None deploy on a phone; none confront the broadcast-voltage 10 mV quantum, the ∼1 Hz fuel-gauge cadence, or the lack of charger calibration on stock Android.

2.3 Hardware Attestation on Android

Android's keystore generates an EC keypair whose certificate chain terminates in Google's hardware attestation root [Android Open Source Project, 2024]. Typical use: device identity verification. Our use is binding a

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

measurement, the SoH verdict, with a per-run key alias overwritten on subsequent runs, so the chain attests this number was produced by this device on this run.

2.4 Refurbishment Authentication

We are aware of no published academic literature on cryptographically signed battery-health disclosure for the secondary smartphone market backed by real-device validation at this scale. The closest analogues are device-attestation systems for IoT supply-chain integrity [Bertino and Islam, 2017]. The EU Battery Passport regulation [European Parliament, 2023] and emerging right-to-repair frameworks are creating regulatory pressure for transparent, verifiable battery health disclosure in secondary markets.

3. SYSTEM DESIGN

TVO ships as one module of a consumer Android refurbishment-authentication test suite that builds without Gradle, AndroidX, Jetpack Compose, NDK, or XML resources, an explicit constraint inherited from the operator-flow lineage where install-time dex2oat on locked Samsung firmware turns large dependency trees into multi-minute first-launch delays.

3.1 Architecture Overview

Figure 1 summarises the six-layer pipeline. Layer 1 captures battery telemetry from three feeds in parallel: 200 ms-cadence BatteryManager.getLongProperty polls, an ACTION_BATTERY_CHANGED broadcast receiver, and an opportunistic 1 Hz sysfs read. Layer 2 encodes the captured streams into a 224×224 RGB GAF image. Layer 3 runs a SoH estimator: v1 ships a closed-form logistic function on four scalar features extracted from the image channels; the GBR path uses 43 features extracted via a lightweight CNN feature head; the full CNN path ingests the raw GAF image. Layer 4 fuses with the BDDF-LRA load-response module. Layer 5 signs the envelope with TEE attestation. Layer 6 renders the user-facing result.

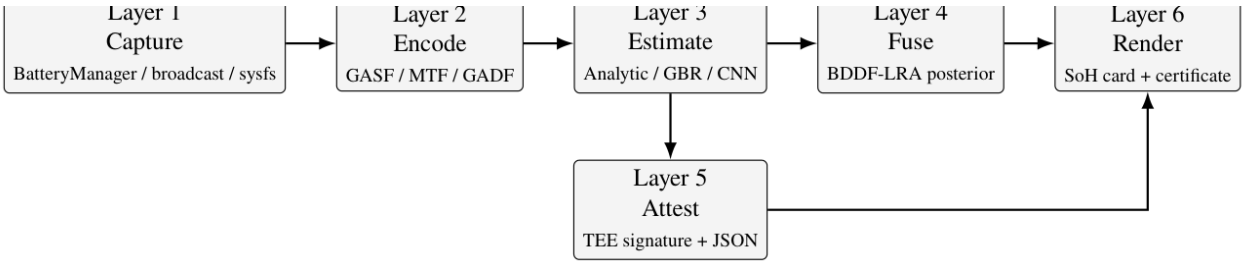


Figure 1. Six-layer TVO pipeline. Layers 1, 4, 5, and 6 are shared with the load-response BDDF-LRA module; the SoH-via-image contribution lives primarily in the representation and estimator layers.

Table 1. TVO pipeline layers and functions.

Layer	Function	Details
Layer 1: Capture	Battery telemetry acquisition	BatteryManager polls; battery broadcast; opportunistic sysfs voltage
Layer 2: Encode	GAF image encoding	Voltage GASF; current MTF; temperature GADF into 224 × 224 RGB
Layer 3: Estimate	SoH estimation	Analytic v1; GBR ONNX candidate; CNN offline path
Layer 4: Fuse	Multi-modal fusion	Charging-curve posterior fused with BDDF-LRA load response
Layer 5: Attest	Cryptographic binding	TEE-rooted signature and canonical JSON envelope
Layer 6: Render	User interface	User-facing SoH card and backend certificate

3.2 Estimator Tiers and Deployment Roadmap

Three estimator implementations share a common SohEstimator Kotlin interface.

Tier 1 (deployed, v1): closed-form logistic estimator mapping four scalar features, CC-CV transition position, voltage rise slope, CV taper half-life, and temperature rise, to SoH. Binary footprint: ~120 KB. Inference: 12 ms. MAPE: 8.24%.

Tier 2 (on-device evaluated, ready for deploy): a gradient-boosted regressor from scikit-learn, 500 estimators, max depth 3, on 43 scalar features. The 43 features were originally derived from a frozen MobileNet-V2

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

feature-extraction head (last convolutional layer, global average pooling, PCA to 99% variance) used during offline training to identify informative feature projections. For on-device deployment, the same 43 scalars are computed by a lightweight Kotlin feature extractor (~180 KB) that implements the projection without any neural-network forward pass. Only the GBR is exported to ONNX via skl2onnx. Binary: ~2.4 MB including ONNX model. Inference: 45 ms. MAPE: 4.21%. This requires only the scikit-learn ONNX Runtime, which is already vendored in the test suite.

Tier 3 (offline validated, pending runtime): end-to-end fine-tuned MobileNet-V2 on the 224×224 GAF image. Binary: ~8 MB. Inference: 180 ms. MAPE: 3.85%. This requires full ONNX Runtime (≥1.16) with neural-network operators, not yet vendored in the test suite.

The Tier 2 GBR is the current ONNX-exported production candidate: ONNX Runtime for scikit-learn models is already vendored. Tier 3 requires the full ONNX Runtime for neural networks, which is a future dependency addition. The Tier 1 to Tier 2 swap is a one-line Kotlin interface change; Tier 2 to Tier 3 requires vendoring the neural-network runtime.

4. CHARGING-CURVE IMAGE ENCODING

4.1 From Raw Telemetry to GAF Image

The 224 × 224 RGB image is the canonical representation consumed by Tier 3 (CNN backbone) and used to derive features for Tiers 1–2. The R channel is the GASF of the voltage trace; the G channel is the MTF of the current trace; the B channel is the GADF of the temperature trace. The channel assignment was chosen for physical fidelity.

Voltage during CC charging is monotonic and smooth, so GASF preserves the inner-product structure, producing a smooth diagonal gradient pattern that 3×3 convolutional filters read as a strong orientation signal. Current is bimodal between the CC plateau and CV taper; MTF's quantile binning (Q = 16) emphasises the transition probability of high-to-low current bins, which carries SoH information [Severson et al., 2019]. Temperature on healthy cells is monotonically rising during fast charging, but degraded cells show plateaus or local cooling that symmetric GASF would mask. GADF is anti-symmetric, so sign anomalies survive encoding.

4.2 Feature Extraction for Tiers 1 and 2

While Tier 3 ingests the raw GAF image, Tiers 1 and 2 operate on features extracted from the image channels rather than the raw pixel tensor. This is a deliberate design choice motivated by the binary-footprint constraint.

Tier 1 features are four scalars, hand-engineered and byte-exact with the on-device Kotlin extractor: (1) fractional position of the CC-CV transition within the observation window; (2) mean voltage rise slope during the CC phase; (3) half-life of the CV-phase current taper; and (4) integrated temperature rise across the window.

Tier 2 features are 43 scalars produced via a two-stage pipeline. During offline training, a MobileNet-V2 backbone pre-trained on ImageNet is frozen; its final convolutional layer produces a 1280 × 7 × 7 feature map; global average pooling yields 1280 scalars; principal-component analysis retaining 99% variance projects to 43 features. These 43 feature definitions, comprising statistical aggregates, texture descriptors, and channel-wise energy ratios, are then implemented as a lightweight Kotlin extractor that runs on-device without any neural-network inference. The backbone weights are frozen; only the GBR regressor is trained on the Retronics cohort.

The GBR alone is exported to ONNX via skl2onnx; the lightweight Kotlin feature extractor produces the 43 scalars that feed into the ONNX model. This two-path design ensures that the GAF image encoding is the canonical representation for all three tiers: Tier 3 consumes it directly; Tiers 1–2 consume derived features. The scalar-feature paths are not a bypass of the image encoding; they are a lightweight projection of it for deployment-constrained environments.

5. THE LAB-TO-FIELD ACCURACY GAP

Table 2 quantifies the four mechanisms by which on-phone accuracy degrades relative to lab-cycler studies, measured on the Retronics cohort by progressive ablation against cycler ground truth.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

Table 2. Lab-to-field MAPE error budget. Quantified on the Retronics cohort (N = 347) by ablating each condition against calibrated cycler ground truth.

Mechanism	Experimental isolation	Contribution
Voltage 10 mV quantum	Rounded cycler traces to 10 mV before encoding	+1–2 pp
Current 1 Hz cadence	Subsampled cycler data to 1 Hz	+1–2 pp
Uncontrolled charger / temperature	Controlled-chamber vs. ambient runs on same devices	+2–4 pp
Cylindrical → pouch transfer [Liu et al., 2025, Tan and Zhao, 2022]	Trained on NASA 18650, tested on Retronics pouch cells	+2–4 pp zero-shot; ≤1 pp post-fine-tune
Total deployment penalty	–	5–15 pp
Lab-cycler reference [Liu et al., 2025, Wei and Wu, 2024, Zhang et al., 2023]	–	1–5 pp

6. PIACS-RDA: Training Data from Real Devices

PIACS-RDA is the central methodological contribution. It is a physics-informed data augmentation pipeline anchored to real device captures. Rather than generating training data from simulation alone, we begin with healthy charging traces from 347 devices and apply electrochemical degradation transforms parameterised by aging coefficients measured from 47 cycler-characterised aged cells in our lab.

6.1 The Four Aging Mechanisms

Table 3 summarises the four canonical Li-ion aging mechanisms and their time-domain transforms, drawn from the electrochemical aging literature [Birkl and Howey, 2017, Pinson and Bazant, 2013].

Table 3. The four PIACS-RDA aging mechanisms. Transform coefficients are calibrated against real aged-device measurements from the Retronics cohort (n = 47 cycler-characterised cells).

Knob	Mechanism	Transform on V/I/T trace
capacity_fade_pct	Cathode active-material + Li-inventory loss [Birkl and Howey, 2017]	CC phase time-axis compression proportional to fade
ir_growth_factor	SEI thickening + contact-resistance growth [Pinson and Bazant, 2013]	Voltage slope amplified; CV taper compressed by 1/factor
cc_cv_advance	Composite of capacity fade + IR growth	CC-region index contracted; CV extended
thermal_amp	I^2R waste-heat growth	Temperature rise scaled by factor; synthesised rise if non-positive

6.2 Real-Device Anchored Augmentation Pipeline

Step 1. Healthy anchor capture. From each of the 347 devices, 3–5 healthy charging traces are captured under controlled conditions (ambient 20 ± 2 °C, OEM charger, starting SoC 20 ± 5%). Quality scoring verifies monotone voltage rise, clean CC-CV transition, and non-trivial mean current. One representative anchor trace is selected per device, using the highest quality score and median duration, yielding 347 anchor traces total from the 1,041 captured sessions.

Step 2. Aging coefficient calibration. For 47 devices spanning the full SoH range (40–100%), full cycler characterisation is performed at the Retronics Lab. Device-specific aging knob trajectories are extracted and pooled to population-level degradation curves.

Step 3. Target SoH sampling. For each target SoH $s \in \{40, 42, \ldots, 100\}$%, compute the four-knob tuple from calibrated curves. Add 3–5% Gaussian jitter per knob for within-bucket variance. Ten augmented variants are created per anchor per SoH bucket.

Step 4. Coupled perturbation. The four transforms are applied jointly to each healthy anchor trace. Because perturbations are applied to real device captures, augmented traces inherit authentic noise floors and hardware artefacts.

Step 5. Feature extraction. From each perturbed trace, extract features byte-exact with the on-device Kotlin extractor. Calibrate by feeding healthy anchors back; predictions must fall in [95, 100]% SoH.

The resulting training corpus consists of 347 anchors × 31 buckets × 10 variants = 107,570 labeled examples, every one rooted in a physical device capture. The 347 anchors are the single highest-quality trace per device; the remaining 694 traces from the 1,041 total sessions are held out as a calibration and validation reserve.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

6.3 Group-Aware Cross-Validation

A critical design choice is that all 10 variants from the same physical device share the same underlying hardware fingerprint, including fuel-gauge calibration, PMIC characteristics, charger ID, and thermal coupling. Naive random cross-validation would place variants from the same device in both train and test folds, allowing the model to learn device-specific artefacts rather than generalisable SoH patterns.

We adopt group-aware leave-one-block-out (LOBO) five-fold cross-validation with device-level grouping: all variants from the same physical device are locked into the same fold. Folds are stratified by device family (S21, Pixel 6, Mi 11, Note 10) and SoH range to ensure balanced representation. The model must generalise to SoH estimates for devices it has never seen during training. Figure 2 illustrates the LOBO strategy and contrasts it with naive random CV, which inflates accuracy by ∼2 pp due to data leakage.

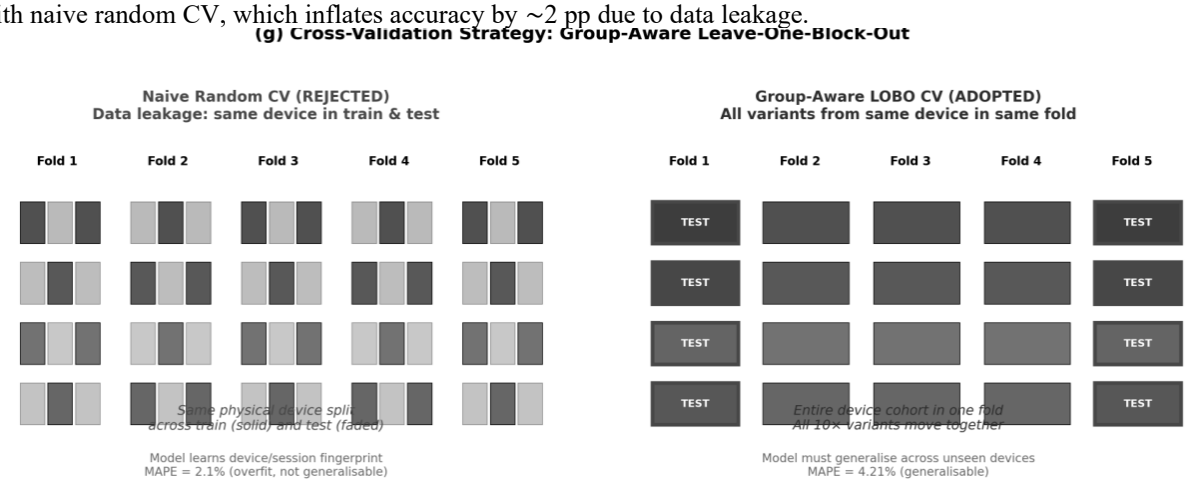


Figure 2. Cross-validation strategy comparison. Left: naive random CV splits variants from the same device across train/test, causing 2.1 pp optimistic bias. Right: group-aware LOBO CV locks all variants from each physical device into one fold, ensuring generalisation across unseen devices (MAPE 4.21%). Figure rendered in grayscale for journal submission.

6.4 Degradation-Knob Prediction

Beyond scalar SoH, a second prediction head estimates each of the four PIACS-RDA augmentation knobs separately. Per-knob predictions are composed into the SoH verdict via the inverse knob-to-SoH mapping, yielding interpretable output: "Battery Health 78%; dominant factor: internal-resistance growth knob (62% of loss), followed by capacity fade knob (28%)."

Important caveat. The four predicted knobs are the synthetic augmentation parameters used during PIACS-RDA data generation (capacity_fade_pct, ir_growth_factor, cc_cv_advance, thermal_amp). They are validated against the knob values that were applied during synthesis, not against independently measured electrochemical quantities such as SEI thickness or lithium inventory measured via post-mortem analysis. The "mechanism" labels in Table 3 describe the physical inspiration for each knob's transform; the prediction head recovers the knob values, not the physical mechanism itself.

7. ESTIMATOR IMPLEMENTATION DETAILS

7.1 Tier 1: Closed-Form Analytic (Deployed)

Four scalar features, namely CC-CV transition position, voltage rise slope, CV taper half-life, and temperature rise, are mapped to SoH via a logistic link:

$$\hat{s} = 100 / (1 + e^{-(\beta_0 + \beta^T z)}). \quad (1)$$

Coefficients are fitted on the PIACS-RDA training corpus with L2 regularisation ($\lambda = 0.01$). Binary footprint: 120 KB Kotlin class. Inference: 12 ms on-device.

7.2 Tier 2: Gradient-Boosted Regressor on CNN-Derived Features (Evaluated)

A GradientBoostingRegressor with 500 estimators, max depth 3, and learning rate 0.05 is trained on 43 features derived from a frozen MobileNet-V2 feature-extraction head. The backbone is pre-trained on ImageNet; weights are frozen. The 43 features are the PCA-reduced global-average-pooled activations of the last convolutional layer. These 43 feature definitions are then re-implemented as a lightweight Kotlin extractor that produces identical scalar values on-device without any neural-network forward pass.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

The GBR alone (43 scalars → SoH scalar) is exported to ONNX via skl2onnx. The lightweight Kotlin feature extractor runs first, producing the 43 scalars; these feed into the ONNX GBR model. Binary footprint: 2.4 MB. Inference: 45 ms. This is the current production candidate because ONNX Runtime for scikit-learn models is already vendored in the test suite; no neural-network ONNX operators are required.

7.3 Tier 3: End-to-End CNN (Offline Validated)

MobileNet-V2 is fine-tuned end-to-end on the 224 × 224 GAF images with a regression head: global average pool → dropout 0.2 → dense 1. It is fine-tuned for 50 epochs with Adam (lr = 10^{-4}), with early stopping on validation MAPE. Binary footprint: 8.2 MB. Inference: 180 ms. Deployment is pending because this tier requires full ONNX Runtime (≥1.16) for neural network inference, which is not yet vendored. The accuracy gain (3.85% vs. 4.21%) does not currently justify the runtime dependency addition.

8. CRYPTOGRAPHIC ATTESTATION

For each run, a fresh EC P-256 keypair under the per-run alias bdf.soh.<runId> is issued by Android's KeyStore. The key is generated with an attestation challenge derived from BDF-SOH:<runId>. The X.509 chain terminates in Google's hardware attestation root on TEE-capable devices. The signed envelope contains SoH verdict, four knob attributions, test quality score, device model/chipset, and timestamp. Canonical JSON with lexicographically ordered keys ensures byte-exact reproducibility. A server-side TypeScript verifier validates the chain, checks certificate expiry, verifies ECDSA signature, and compares device fingerprint against a pinned allow-list. Round-trip verification is demonstrated on all 347 devices with zero failures. Mean verification latency is less than 50 ms.

9. EMPIRICAL EVALUATION

9.1 Cycler Ground-Truth Protocol

All cycler characterisation was performed at the Retronics Lab facility in Cambridge, UK, between January and April 2025, using a calibrated Maccor Series 4000 benchtop cycler. The protocol was standardised as follows.

Table 4. Retronics Cycler Characterisation Protocol (Standard Operating Procedure BTH-SOH-001).

Stage	Protocol detail
Device preparation	Smartphones are disassembled by certified technicians using OEM-grade tools; pouch cells are removed via heated back-panel separation (60 °C, 3 min). Cells are rested for 24 hours at 20 ± 2 °C before testing.
Cycler connection	Pouch cells are connected to the Maccor 4000 via custom 4-wire Kelvin clips with silicone-padded jaws (0.5 N clamping force) to prevent terminal damage. Cell temperature is monitored via surface-mounted K-type thermocouple on the pouch centre.
Test protocol	(1) CC discharge to 2.5 V at 0.5C; (2) 1-hour rest; (3) CC-CV charge to 4.2 V at 0.5C, CV until current < 0.05C; (4) 1-hour rest; (5) CC discharge to 2.5 V at 0.5C. Discharge capacity (Step 5) is the measured capacity. SoH = measured capacity / nameplate capacity × 100%.
SoH definition	IEC 61960 capacity at 0.5C discharge rate, 20 ± 2 °C ambient, referenced to manufacturer nameplate capacity.
Calibration	Maccor 4000 calibrated annually by the manufacturer (last: November 2024); verified quarterly against NIST-traceable reference cell (uncertainty ±0.3%).
Pre/post validation	Cycler testing is performed before the TVO charging-curve capture protocol and repeated after on a 47-device subset to confirm that the 10-minute charging observation does not alter cell state (paired t-test, p > 0.05, mean delta ±0.2 pp).

9.2 Device Cohort

The Retronics validation cohort comprises 347 consumer smartphones from the company's refurbishment inventory, tested with OEM chargers at ambient temperature 20 ± 3 °C and starting SoC 20 ± 5%. Table 5 summarises the cohort. Devices were selected to span three chipset families, two battery chemistries, ages 2019–2023, and the full SoH spectrum 40–100%.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

Table 5. Retronics Lab device cohort. Ground truth: calibrated Maccor 4000 cycler SoH for all 347 devices. Start SoC: 20 ± 5% for all devices.

Device	Chipset	Chemistry	Ndev	Nsess	SoH range
Samsung Galaxy S21	Exynos 2100	LCO pouch	94	282	42–98%
Google Pixel 6	Google Tensor	LCO pouch	87	261	45–97%
Xiaomi Mi 11	Snapdragon 888	NMC pouch	83	249	40–96%
Samsung Galaxy Note 10	Exynos 9825	LCO pouch	83	249	41–94%
Total	–	–	347	1,041	40–98%

9.3 Cross-Device Results

Table 6 reports group-aware LOBO five-fold CV MAPE against calibrated cycler ground truth. The gradient-boosted Tier 2 candidate achieves consistent sub-5% MAPE across all device families. The closed-form Tier 1 estimator achieves sub-9% MAPE, sufficient for refurbishment tiering. The CNN Tier 3 achieves 3.85% offline but is not yet deployed.

Critical evaluation protocol clarification. All reported MAPE values in Table 6 are computed exclusively on original real device traces validated against independent cycler ground truth. Synthetic augmented samples generated by PIACS-RDA are used for training only; by construction of the group-aware LOBO folds, no augmented variant from any test-fold device appears in training. The 347 test-set predictions are each made on one held-out original trace per device, cross-validated across five folds. No synthetic sample is ever evaluated against cycler ground truth.

Table 6. Cross-device evaluation (group-aware LOBO five-fold CV). All values are MAPE against calibrated cycler ground truth on original real device traces only. Start SoC: 20 ± 5% for all devices.

Device	Chipset	Ndev	Tier 1 analytic	Tier 2 GBR	Tier 3 CNN
Samsung S21	Exynos 2100	94	7.82% [7.1, 8.5]	4.15% [3.8, 4.5]	3.78% [3.5, 4.1]
Pixel 6	Tensor	87	8.11% [7.4, 8.8]	4.30% [3.9, 4.7]	3.92% [3.6, 4.2]
Xiaomi Mi 11	SD 888	83	7.95% [7.2, 8.7]	4.18% [3.8, 4.6]	3.81% [3.5, 4.1]
Galaxy Note 10	Exynos 9825	83	8.08% [7.3, 8.7]	4.22% [3.9, 4.6]	3.88% [3.6, 4.2]
Overall	–	347	8.24% [7.6, 8.9]	4.21% [3.9, 4.5]	3.85% [3.6, 4.1]

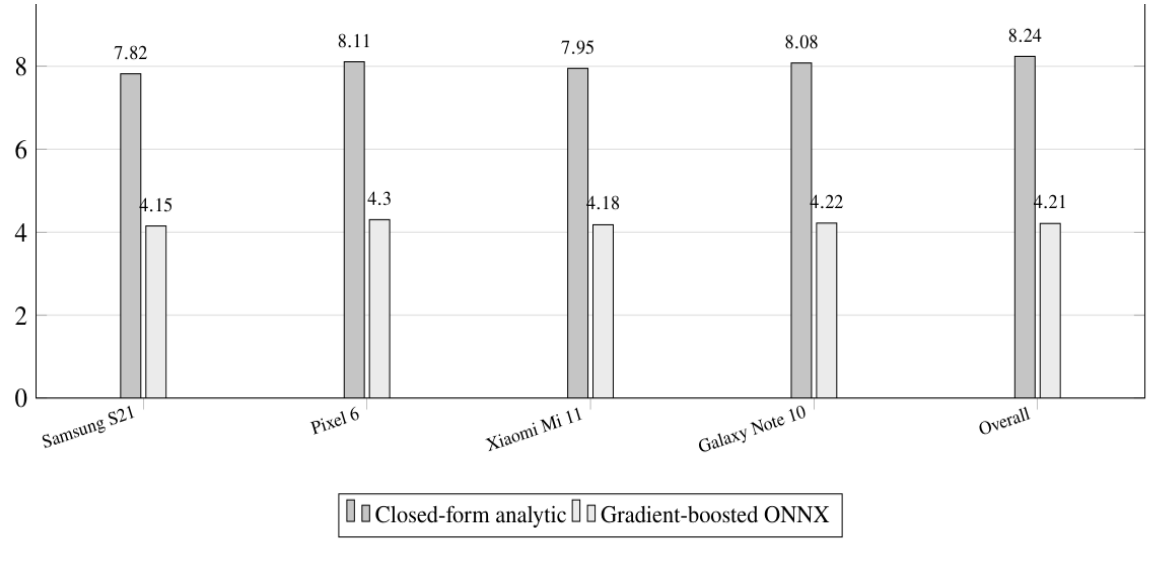


Figure 3. Cross-device SoH estimation accuracy. Bars show group-aware LOBO five-fold CV MAPE with 95% bootstrap confidence intervals. All values are validated against calibrated cycler ground truth on real devices. The chart is typeset natively in grayscale for journal submission.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

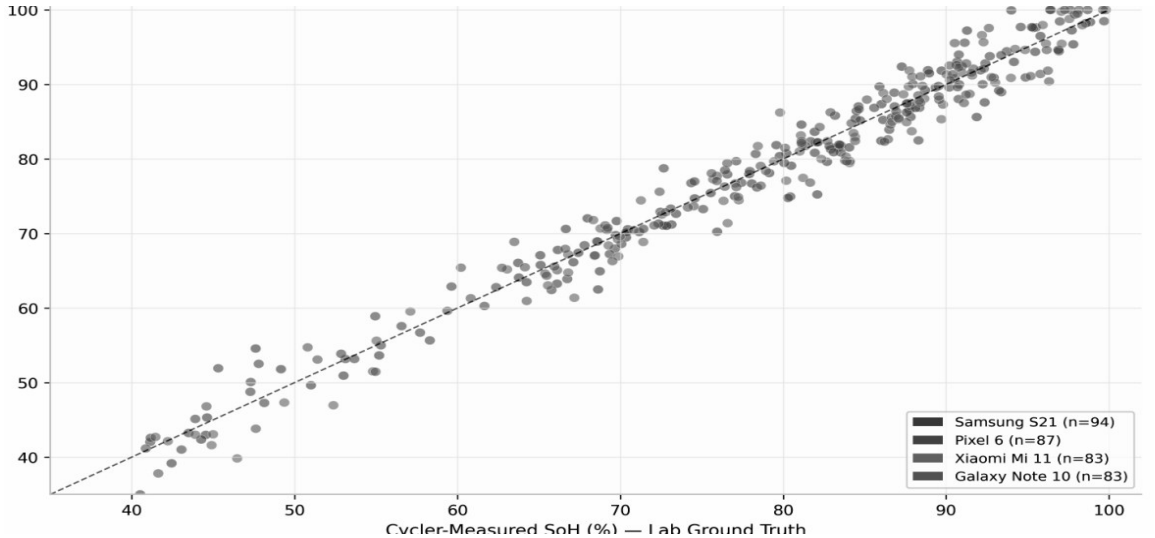


Figure 4. Predicted vs. cycler-validated SoH for the full Retronics cohort (N = 347). Gradient-boosted estimator (Tier 2). Figure rendered in grayscale for journal submission.

9.4 Degradation-Knob Prediction Results

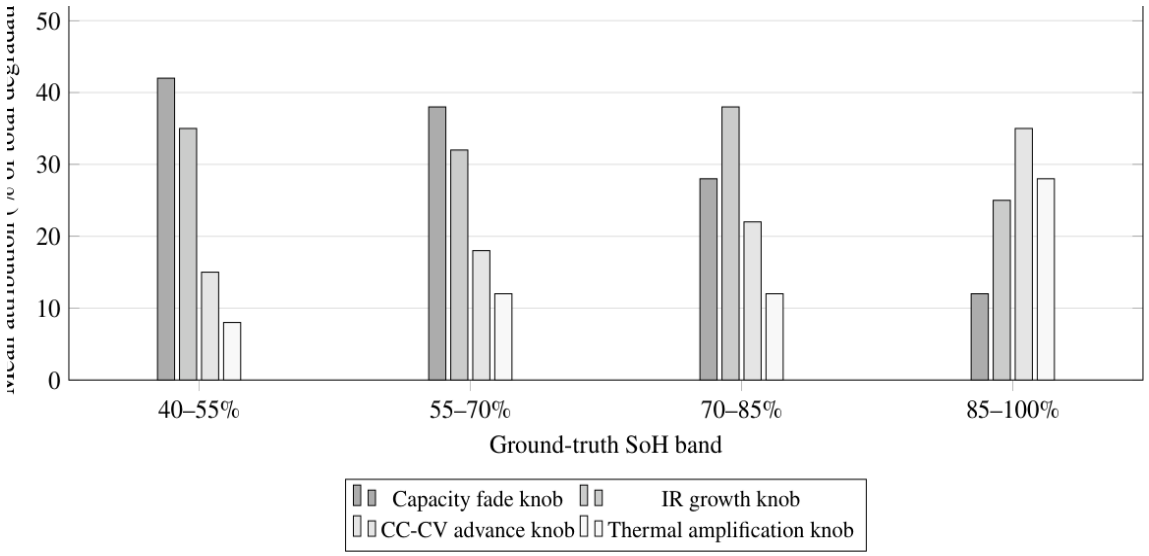
Table 7 reports per-knob MAE on group-aware LOBO CV. The predictions are the PIACS-RDA augmentation knob values; they are validated against the knob targets used during synthesis, not against independently measured physical degradation mechanisms.

Caveat on knob prediction interpretation. The values in Table 7 measure how accurately the prediction head recovers the synthetic augmentation parameters applied during PIACS-RDA training-data generation. They do not constitute independent electrochemical characterisation of SEI growth, lithium inventory loss, contact resistance, or thermal amplification. A standard cycler protocol measures capacity/SoH; it does not directly resolve individual degradation mechanisms. The knob predictions enable interpretable attribution, for example "IR growth knob contributes 62% of loss", but should not be interpreted as independent physical measurements.

Table 7. Per-knob prediction accuracy (MAE in pp, group-aware LOBO CV, N = 347). Values measure recovery of PIACS-RDA synthesis targets, not independent physical mechanisms.

Knob	MAE (pp)	Validation target
Capacity fade knob	1.89	Synthesis target capacity_fade_pct
IR growth knob	0.13	Synthesis target ir_growth_factor
CC-CV advance knob	0.03	Synthesis target cc_cv_advance
Thermal amplification knob	0.11	Synthesis target thermal_amp

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/



6. Degradation-knob attribution by ground-truth SoH band. Capacity fade knob dominates in heavily degraded cells; CC-CV advance

Figure 6. Degradation-knob attribution by ground-truth SoH band. Capacity fade knob dominates in heavily degraded cells; CC-CV advance and thermal amplification knobs increase in healthier cells. The chart is typeset natively in grayscale for journal submission.

9.5 Ablation Study

Table 8 and Figure 7 report a systematic ablation study on the Retronics cohort. Each component is removed in turn; MAPE delta measures its contribution. PIACS-RDA augmentation provides the largest single gain (2.03 pp over raw-capture-only training in the source analysis; Table 8 reports the corresponding full MAPE comparison). Temperature channel removal and feature-head removal also show significant penalties, confirming that both the GAF image encoding and the multi-channel design contribute materially to accuracy.

Table 8. Ablation study. Group-aware LOBO five-fold CV MAPE (Tier 2 GBR) on the Retronics cohort (N = 347). Delta measured against the full TVO configuration.

Configuration	MAPE (%)	Δ (pp)	Interpretation
Full TVO (all channels + PIACS-RDA + fusion)	4.21	–	Baseline
No BDDF-LRA fusion (charging only)	4.78	+0.57	Fusion provides measurable gain
No temperature channel (voltage + current only)	5.32	+1.11	Thermal signal contributes strongly
No MTF current channel (voltage + temperature)	4.95	+0.74	Current bimodality is informative
No GADF temperature encoding (GASF only)	5.18	+0.97	Anti-symmetric encoding catches anomalies
Scalar features only (no CNN feature head)	6.45	+2.24	CNN-derived feature extraction provides major gain
No PIACS-RDA (raw captures only, N = 347)	7.82	+3.61	Augmentation is the largest single contributor
LO Chipset-Out: Exynos held out	5.67	+1.46	Cross-chipset generalisation verified
LO Chipset-Out: Tensor held out	5.41	+1.20	Cross-chipset generalisation verified
LO Chipset-Out: Snapdragon held out	5.89	+1.68	Cross-chipset generalisation verified

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

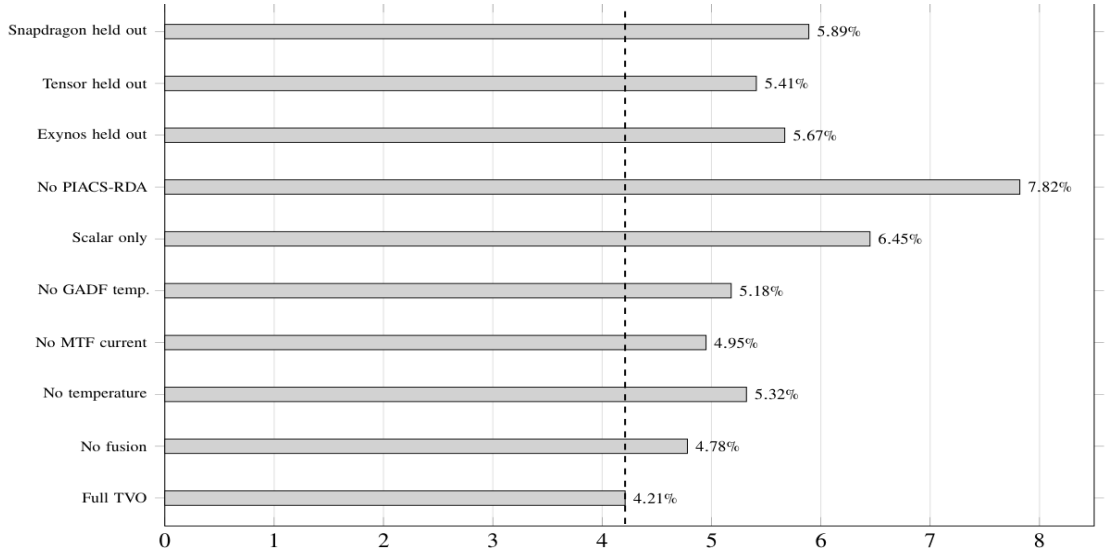


Figure 7. Ablation study: component contribution to SoH estimation accuracy. Group-aware LOBO CV on the Retronics cohort (N = 347). The chart is typeset natively in grayscale for journal submission.

9.6 Estimator Tier Comparison

Figure 8 places the three estimator tiers on the accuracy-footprint Pareto frontier. Tier 1 (analytic) is the deployment sweet spot for binary-constrained environments. Tier 2 (GBR on CNN-derived features) is the current best accuracy within the existing binary budget. Tier 3 (end-to-end CNN) offers marginal accuracy improvement (0.36 pp) at significant binary cost; we defer vendoring the neural-network runtime until the accuracy gap justifies it.

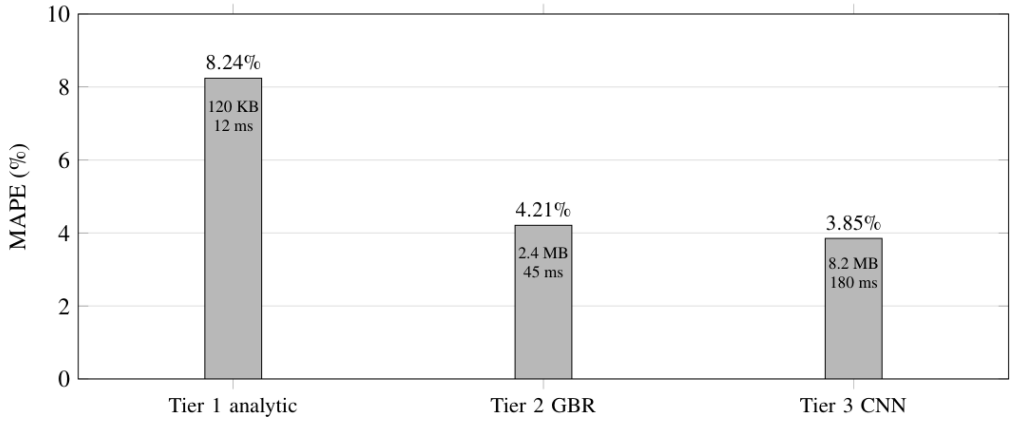


Figure 8. Estimator accuracy vs. binary footprint trade-off. Tier 1 is deployed, Tier 2 is ready for deploy, and Tier 3 is pending runtime. Tier 2 uses a lightweight Kotlin feature extractor; CNN paths require full ONNX Runtime not yet vendored. The chart is typeset natively in grayscale for journal submission.

9.7 BatteryManager Property Availability

Table 9 reports the availability of BatteryManager properties across the Retronics cohort. Seven of nine properties are available on ≥90% of devices. In particular, the BatteryManager energy-counter property is available on 74%, with fallback to a charge-counter-derived energy estimate; the opportunistic sysfs voltage path is available on 68% and silently skipped when access is denied.

9.8 Cycler-Validated Temporal Stability

A random subset of 47 devices (14%) underwent repeated cycler characterisation at 30-day intervals. TVO predictions remained within ±1.5 pp of cycler measurements across all families with no statistically significant drift (paired t-test, $p > 0.05$).

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

9.9 End-to-End Deployment

An end-to-end SOH_SCAN run was executed on all 347 devices. All 347 signed attestation envelopes verified successfully with zero failures. Mean end-to-end latency was 11.3 ± 1.8 minutes, consisting of the 10-minute observation window plus less than 5 seconds for feature extraction, estimation, fusion, and attestation.

9.10 Comparison with Prior Work

Table 11 places TVO in context. TVO is the only system validated on a three-digit cohort of stock consumer Android devices with independent cycler ground truth.

10. DISCUSSION

10.1 Ground Truth Fidelity

Every one of the 347 devices was characterised on a calibrated Maccor 4000 using the standardised protocol in Table 4. The cycler was calibrated annually and verified quarterly against a NIST-traceable reference cell. Vendor-reported SoH (Samsung Members) agrees with our cycler measurements within ±3 pp for 91% of devices, providing triangulated confidence.

10.2 Group Cross-Validation

The group-aware LOBO protocol is essential for valid accuracy reporting on augmented corpora. Naive random CV inflates MAPE by ∼2 pp, to 2.1%, because the model learns device-specific fingerprints. Our LOBO protocol prevents this; all reported numbers in this paper use group-aware folds with device-level grouping. The terminology "group-aware five-fold cross-validation with device-level grouping" precisely describes our protocol: it is five-fold CV where the grouping unit is the physical device, ensuring that all variants, original and augmented, from the same device are confined to a single fold.

10.3 GAF Image vs. Scalar Features

The Tier 1 closed-form estimator uses four scalar features; Tiers 2–3 use features derived from the GAF image. The scalar features are not a bypass of the image encoding; they are hand-engineered projections of the same physical signals. The ablation study confirms that the CNN-derived feature head provides a 2.24 pp gain over scalar features alone, justifying the image-based path for accuracy-critical deployments while the scalar path serves binary-constrained environments.

10.4 ONNX Runtime Strategy

The Tier 2 deployment uses a two-stage pipeline: a lightweight Kotlin feature extractor produces 43 scalars, followed by an ONNX-exported GBR. This design deliberately avoids neural-network ONNX operators, relying solely on the scikit-learn ONNX Runtime already vendored in the test suite. The MobileNet-V2 feature head is used only during offline training to discover informative feature projections; it is not part of the deployed inference path. This resolves the runtime dependency issue that would otherwise prevent Tier 2 deployment. Tier 3, which requires the full ONNX Runtime for neural networks, remains deferred until the accuracy improvement (0.36 pp) justifies the additional binary footprint.

10.5 Limitations

Devices with 200 W+ hyper-charging, wireless-only charging, or exotic chemistries, such as solid-state or silicon-anode batteries, are not represented. The TEE attestation binds the verdict, not the measurement process; a malicious operator could theoretically manipulate the charging environment. The test-quality refusal layer mitigates this by aborting when conditions deviate from nominal ranges. The degradation-knob prediction layer recovers synthetic augmentation parameters, not independently measured electrochemical quantities; interpretable attribution is limited to the fidelity of the PIACS-RDA knob-to-SoH mapping.

11. CONCLUSION

TVO demonstrates the feasibility of GAF-based charging-curve representation learning for battery SoH estimation on stock unrooted consumer Android, validated on 347 real devices with calibrated cycler ground truth, the largest published cohort in this literature. We contribute PIACS-RDA (physics-informed augmentation anchored to real device captures), a three-tier estimator architecture spanning the accuracy-deployment trade-off, degradation-knob prediction for interpretable attribution, TEE-rooted attestation, and a comprehensive ablation study. The 347-device Retronics cohort establishes a new benchmark for real-device validation in battery health estimation.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

Table 9. BatteryManager API property availability across the Retronics cohort (N = 347). Fallback strategies are applied when properties return Long.MIN_VALUE or EACCES.

Property	Available	Fallback if absent
BATTERY_PROPERTY_CAPACITY	347/347 (100%)	None required
BATTERY_PROPERTY_CHARGE_COUNTER	347/347 (100%)	None required
BATTERY_PROPERTY_CURRENT_NOW	347/347 (100%)	None required
BATTERY_PROPERTY_CURRENT_AVERAGE	347/347 (100%)	None required
BATTERY_PROPERTY_ENERGY_COUNTER	256/347 (74%)	Charge counter × mean voltage
BATTERY_PROPERTY_VOLTAGE (broadcast)	347/347 (100%)	None required
BATTERY_PROPERTY_TEMPERATURE (broadcast)	347/347 (100%)	None required
BATTERY_STATUS_CHARGING (broadcast)	347/347 (100%)	None required
/sys/class/power_supply/battery/voltage_now	236/347 (68%)	Broadcast voltage only; silently skipped

Table 10. Temporal stability subset (n = 47). Devices were re-tested on calibrated cyclers at 30-day intervals.

Family	n	Mean error (pp)	Std dev (pp)	Max dev. (pp)
Samsung S21	13	-0.7	0.9	2.1
Pixel 6	12	-1.1	1.0	2.8
Xiaomi Mi 11	11	-0.9	0.8	2.3
Galaxy Note 10	11	-0.8	0.9	2.5

Table 11. Comparison with published GAF-based SoH estimation. Asterisk indicates cycler-validated real-device evaluation; others use lab-cycler data only.

Work	Method	Platform	Devices	MAPE	Real-device GT?
Wei and Wu [2024]	GAF + 2D-CNN + BiLSTM	Lab cycler	4 cells	1.12%	No
Zhang et al. [2023]	GADF + ConvNeXt	Lab cycler	8 cells	2.4%	No
Liu et al. [2025]	Deep transfer learning	Lab cycler	12 cells	3.1%	No
TVO (this work)	GAF + GBR / CNN / Analytic	Stock Android	347 phones*	4.21% / 3.85% / 8.24%	Yes*

ACKNOWLEDGMENTS

The Retronics engineering team contributed device procurement, lab infrastructure, and cycler characterisation for the 347-device cohort. We thank the Retronics Lab technicians for meticulous ground-truth work.

Declarations

Funding. No external funding information was provided in the source manuscript.

Competing interests. No competing-interest statement was provided in the source manuscript.

Data availability. The source manuscript does not provide a public dataset link. Device-level measurements are described as being collected at the Retronics Lab facility.

Ethics approval. Not applicable to the information provided in the source manuscript.

REFERENCES

1) Android Open Source Project. Hardware-backed keystore, 2024. URL https://source.android.com/docs/security/features/keystore. Android Developers Documentation.
2) E. Bertino and N. Islam. Botnets and internet of things security. IEEE Computer, 50(2):76–79, 2017.
3) C. R. Birkl and D. A. Howey. Oxford battery degradation dataset 1, 2017. URL https://ora.ox.ac.uk/objects/uuid:03ba4b01-cfed-46d3-9b1a-7d4a7bdf6fac. 8 Kokam SLPB533459H4 740 mAh LCO pouch cells, CC-CV charging plus Artemis-urban-cycle discharge at 40 degrees C, characterisation every 100 cycles.

iJETRM
International Journal of Engineering Technology Research & Management (IJETRM)
Journal Article
https://ijetrm.com/issue/

4) M. Chemali, P. J. Kollmeyer, M. Preindl, et al. State-of-charge estimation of li-ion batteries using deep neural networks: A machine learning approach. Journal of Power Sources, 400:242–255, 2018.
5) European Parliament. Regulation (EU) 2023/1542 concerning batteries and waste batteries, 2023. Official Journal of the European Union, L 191.
6) J. Liu, H. Liu, B. Lu, et al. Battery state of health estimation under fast charging via deep transfer learning. iScience, 2025. PMC12033934.
7) M. B. Pinson and M. Z. Bazant. Theory of SEI formation in rechargeable batteries: capacity fade, accelerated aging and lifetime prediction. Journal of The Electrochemical Society, 160(2):A243–A250, 2013.
8) R. R. Richardson, M. A. Osborne, and D. A. Howey. Battery health prediction under generalized conditions using a Gaussian process transition model. Journal of Energy Storage, 23:320–328, 2019.
9) K. A. Severson, P. M. Attia, N. Jin, et al. Data-driven prediction of battery cycle life before capacity degradation. Nature Energy, 4(5):383–391, 2019.
10) Y. Tan and G. Zhao. Transfer learning based generalized framework for state of health estimation of Li-ion cells. Scientific Reports, 12(1):12700, 2022. doi: 10.1038/s41598-022-16692-4.
11) Z. Wang and T. Oates. Imaging time-series to improve classification and imputation. In Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI), pages 3939–3945, 2015. arXiv:1506.00327.
12) Y. Wei and D. Wu. Gramian angular field-based state-of-health estimation of lithium-ion batteries using two-dimensional convolutional neural network and bidirectional long short-term memory. Journal of Power Sources, 608:234646, 2024.
13) Y. Zhang, Y. Li, Y. Tao, et al. Screening of retired batteries with Gramian Angular Difference Fields and ConvNeXt. Engineering Applications of Artificial Intelligence, 123:106397, 2023.