

## **CROSS-BAND ANTENNA HEALTH INFERENCE FOR SMARTPHONE REFURBISHMENT: A MULTI-DEVICE, CRYPTOGRAPHICALLY-ATTESTED 90-SECOND RECEPTION TEST**

**Arun Teja Sara**

Retronics Market Place UK Limited, Middlesbrough, United Kingdom

[info@retronixs.com](mailto:info@retronixs.com)

### **ABSTRACT**

When purchasing a pre-owned smartphone, physical defects like cracked glass are obvious, but damaged internal antennas remain invisible to the consumer. Reception failures following handover are significant drivers of resale disputes; however, current on-device “RF tests” conflate hardware quality with local environmental attenuation. We present Cross-Band Antenna Health Inference (CBAHI), a 90-second on-device test that differentiates hardware deficits from environmental factors on stock Android. CBAHI captures concurrent reception data across WiFi (2.4/5/6 GHz), Cellular (LTE, 5G NR), GNSS, and Bluetooth-LE. Our core contribution is a leave-one-out studentised residual test with an exact t-distribution null, allowing for calibrated p-values. We evaluated CBAHI on 420 physical devices tested in the Retronics refurbishment laboratory. The corpus spans 120 healthy calibration units and 300 labelled test units across Samsung, Google, and Xiaomi OEMs, incorporating documented water damage, post-screen-replacement, post-back-glass-replacement, and drop-history cohorts. Zero false-positive hardware-concern flags were observed in the cross-environment stress test (60 runs across environment shifts), while the full labelled-cohort evaluation at  $\eta = 0.01$  yielded one false positive out of 120 healthy devices (specificity 99.2%). Each result is signed by a hardware-attested AndroidKeyStore key bound to a buyer-supplied nonce plus the device’s IMEI hash, mitigating replay, twin-device substitution, pre-caching, and band-uniform selective-environment manipulation. The system achieves a weighted sensitivity of 87.1% for documented water and mechanical damage (C1–C3), supported by a differentially-private federated protocol with cohort-rolling keys. We further characterise the system through a per-band Bayesian posterior baseline, a software-perturbation grid that exercises the graceful-degradation behaviour of the leave-one-out test under partial radio loss, and a cross-environment validation yielding zero false-positive hardware-concern flags across a  $> 13$  dB cellular shift, a  $> 7$  dB-Hz GPS L1 shift, and an opposite-direction Bluetooth shift. We are explicit about what we do not claim: per-antenna isolation (closed at the OEM combiner firmware), graded NFC sensitivity (binary functional check only), and reliability in pathological RF environments. A flagged band is a hint to inspect, not a verdict.

### **Keywords:**

Smartphone refurbishment; antenna health; RF inference; hardware attestation; cross-band analysis; studentised residuals; differential privacy; federated learning.

## **1. INTRODUCTION**

A drop event breaks one antenna; a basement weakens every antenna. To our knowledge, CBAHI is the first stock-Android cross-band hardware-health inference system designed to distinguish between these two latent variables. While storage-attestation addressed the trust gap for data wiping, the question of whether a phone actually receives signals as intended has remained a domain of ad-hoc commercial apps [1, 2].

### **1.1 Central Observation**

Damaged-antenna phones and attenuating environments differ in the cross-band distribution of reception loss. Mechanical damage typically impacts a specific feed line or solder joint, whereas environmental factors attenuate all bands shared across the device’s architecture. CBAHI leverages this asymmetry, providing a calibrated size-controlled test that remains robust even across significant environmental shifts.

### **1.2 Real-World Validation at Retronics**

This paper reports results from an extensive experimental programme conducted at the Retronics refurbishment laboratory in Middlesbrough, UK. Over a six-month period (November 2024–April 2025), we tested 420 physical smartphone devices drawn from live refurbishment intake stock. These were not simulated or emulated units; every device was a physical handset that passed through the standard Retronics intake

pipeline, with ground-truth labels assigned by certified technicians using documented damage criteria. The test corpus comprises:

- 120 healthy devices (C0) serving as the calibration baseline, spanning Samsung Galaxy S21/S22/S23/S24 series, Google Pixel 6/7/8 series, and Xiaomi 13/14 series;
- 80 water-damaged devices (C1) with triggered chemical indicators at intake;
- 100 post-screen-replacement devices (C2) from third-party repair shops;
- 60 post-back-glass-replacement devices (C3) with known antenna-compromising repairs;
- 60 drop-history devices (C4) with self-reported drops but no visible damage.

All 420 devices were physically present in the Retronics lab, tested under controlled RF conditions across three matched environments (outdoor sky-view, typical indoor home, basement/dense-indoor), with each device-environment pair repeated  $r = 5$  times to estimate within-run variance.

### 1.3 Contributions

This paper makes seven contributions:

1. The first multi-device stock-Android cross-band hardware-health inference system, operating with consumer-permission APIs and no vendor primitives.
2. An application of the externally-studentised deletion residual to cross-band antenna health, with closed-form null and calibrated p-values.
3. A conjugate Normal-Inverse-Gamma baseline-update protocol with Student-t posterior predictive.
4. A differentially-private, attestation-anchored federated aggregation scheme for refreshing the corpus without exposing per-device readings.
5. Mitigation of replay, twin-device substitution, pre-caching, and band-uniform selective-environment manipulation via an attack-model-aware attestation challenge.
6. A completed large-scale empirical evaluation on 420 physical devices with TP/FP/TN/FN metrics, Wilson confidence intervals, and per-OEM stratification.
7. An empirical characterisation of the construction's cross-environment robustness on real-device captures across  $> 13$  dB cellular shifts.

### 1.4 What This Paper Does Not Claim

We are explicit about the boundaries of CBAHI:

- Per-antenna isolation: CBAHI operates at band resolution; physical antenna localisation is closed at the OEM combiner firmware and below our resolution.
- Graded NFC sensitivity: NFC is a binary functional check only.
- Pathological RF environments: Faraday cages, intentional jamming, and active near-field interferers break the model.

A flagged band is a hint to inspect, not a verdict. CBAHI is positioned as a pre-screen, not a substitute for calibrated lab measurement.

## 2. BACKGROUND AND RELATED WORK

### 2.1 Smartphone Reception Measurement on Android

The Android GnsMeasurement API (Android 7.0+) exposes per-satellite carrier-to-noise density  $C/N_0$ , constellation, and carrier frequency. Prior work on smartphone GNSS quality assessment [3, 4] demonstrates that smartphone-grade  $C/N_0$  medians are stable for a given device but vary significantly across devices and environments. WiFi ScanResult.level is a per-AP RSSI value with a vendor-defined unit (IEEE 802.11-2020, Clause 11) [5], comparable within a chipset but not across chipsets. Cellular RSRP and RSRQ are defined by 3GPP TS 36.214 [6] and TS 38.215 [7] with absolute dBm units and are directly comparable across chipsets.

### 2.2 Hardware Fingerprinting from RF Signals

There is a substantial WiFi-CSI fingerprinting literature on inferring hardware identity from PHY-layer reception patterns. Brik et al. [8] identified individual 802.11 NICs via PHY-layer radiometric signatures; Liu et al. [9] extended the per-chain hardware-imperfection idea using channel state information; and Yang et al. [10] treated the RSSI distribution across multiple APs as a device-distinguishing feature. These works share with CBAHI the goal of extracting hardware-level information from software-visible RF reception, but their target is identification rather than health, and they require CSI access (Linux firmware patches such as Atheros CSI Tool or Intel 5300) that is not available on stock Android. CBAHI works the other side of the problem: we accept that we cannot read CSI on stock Android, but we show that the population statistics accessible through public APIs are sufficient for hardware-versus-environment discrimination. Our experiments on 420 physical devices confirm this claim empirically.

### 2.3 GNSS Multipath and Hardware-vs-Environment Classification

The GNSS community has long studied separating hardware from environment within a single band. Suzuki and Kubo [11] and Hsu [12] use elevation, azimuth, C/N0 patterns, and 3-D map information to classify multipath conditions, and modern smartphone GNSS quality work [3, 4, 13] treats per-satellite C/N0 as a physically grounded multi-path indicator. These works are within-band and require either external maps or high-rate/raw-pseudorange access. CBAHI is across-band and uses only the median per-satellite C/N0 above a band-adaptive elevation cutoff. We re-use the multipath-aware filter from Suzuki and Kubo [11] as a sample-selection step, not as a classification feature. A related line of work uses RF data to classify the environment (urban canyon, indoor, foliage) without isolating the device. Nasiri et al. [14], for example, use Wi-Fi and cellular signal statistics for environmental awareness. CBAHI inverts the target: rather than label the environment, we con-

dition on it via the cross-band shared-attenuation model so that the residual variation is attributable to hardware.

#### **2.4 Pre-Screen Applications for the Resale Market**

A handful of commercial Android apps target the same use case as CBAHI: Phone Check (and Test) [1], TestM [2], and Phone Doctor Plus [15] each include a “Wi-Fi test”, “cellular test”, or “GPS test” in their refurbishment-pre-screen flow. All three read raw single-band metrics and present them with rule-of-thumb thresholds. None compute a cross-band statistic, none publish a per-model baseline, and none sign their reports cryptographically. The labelled-cohort validation design includes head-to-head comparison against these baselines on the same labelled corpus. Carrier trade-in flows collect a similar set of raw metrics but do not expose them to the buyer side of the transaction. OEM service menus display vendor-internal metrics including per-chain RSSI on some devices, but they are operator-only, not exposed to a third-party app, and do not produce a verifiable report [16]. CBAHI’s contribution is not that it measures more than these tools; it measures less, since it cannot read per-chain RSSI. Its contribution is that it produces a calibrated, attested, cross-band assessment a buyer can rely on, backed by a 420device empirical validation.

#### **2.5 Hardware-Fault Inference from Software-Visible Signals**

The general framing—use software-visible analog properties of the device to infer hardware state—is established by work on power and thermal side channels [17]. We borrow the framing to infer health, not secrets.

#### **2.6 HCI Work on Resale Trust**

The trust gap in peer-to-peer phone resale has been studied from the human-factors side [18, 19]. CBAHI is the engineering operationalisation of one of the information asymmetries those papers identify: the buyer’s inability to inspect functional internals, specifically for antenna and reception health. Sibling Retronics modules address adjacent gaps (storage wipe, OLED degradation, touch-controller behaviour), and CBAHI uses the same shared AttestationSigner primitive so the buyer-side verifier handles all modules with one trust path.

#### **2.7 Refurbishment-Grade Attestation**

CBAHI plugs into Android Keystore-backed hardware attestation [21] the same way the existing wipe-flow, OLED, and touch modules do. The trust root is Google’s hardware-attestation X.509 root; the chain-walk implementation is shared across modules.

### **3. THEORETICAL FRAMING**

We now formalise why a cross-band statistic separates hardware faults from environmental attenuation, why the natural choice of statistic is a leave-one-out studentised residual, and where the formal guarantee breaks down.

#### **3.1 Model**

Let  $b \in B = \{2.4, 5, 6, \text{LTE}, \text{NR}, \text{L1}, \text{L5}, \text{BT}\}$  index the bands the analyser considers (NFC is not included). Let  $(\mu_b, \sigma_b)$  be the per-model healthy-population mean and standard deviation of the chosen reception metric for band  $b$  (median RSSI in dBm, RSRP in dBm, C/N0 in dB-Hz, etc.). All metrics are higher-is-better after the sign convention of the feature layer. We model the observed median for a single run as

$$X_b = \mu_b - \alpha \epsilon_b - H_b + \sigma_b \epsilon_b \quad (1)$$

where  $\alpha \geq 0$  is a scalar environmental attenuation amount,  $\epsilon_b \geq 0$  is a band-specific environmental coefficient,  $H_b \geq 0$  is the per-band hardware deficit (zero on healthy devices), and  $\epsilon_b \sim N(0, 1)$  are i.i.d. measurement noise terms after within-run aggregation. Define the standardised concern

$$C_b = -(X_b - \mu_b)/\sigma_b = (\alpha \epsilon_b + H_b)/\sigma_b - \epsilon_b \quad (2)$$

so that positive  $C_b$  corresponds to worse-than-baseline reception.

#### **3.2 Leave-One-Out Studentised Residual**

The natural cross-band statistic for our setting is the externally studentised deletion residual, also known as the generalised extreme-studentised-deviate (ESD) statistic in the classical outlier-detection literature [22, 23, 24]. Earlier prototypes used a more naive max-minus-mean asymmetry index  $A = \max_b C_b - 1/m \sum_b C_b$ , which is a reasonable point estimate but has three drawbacks: its null distribution depends on  $m = |B'|$ , different runs with different available bands are not directly comparable, and the maximum appears in both terms. The ESD construction fixes all three; the application novelty is its use as a cross-band hardware-vs-environment discriminator on stock-Android reception medians. For a candidate band  $b^*$ , let

$$\bar{C}_{-b^*} = (1/(m-1)) \sum_{b \neq b^*} C_b \quad (3)$$

$$s_{-b^*}^2 = (1/(m-2)) \sum_{b \neq b^*} (C_b - \bar{C}_{-b^*})^2 \quad (4)$$

$$T_{b^*} = (C_{b^*} - \bar{C}_{-b^*}) / (s_{-b^*} \sqrt{1 + 1/(m-1)}) \quad (5)$$

The CBAHI test statistic is  $T = \max_{b^*} T_{b^*}$ , with hardware concern declared when  $T > t_{1-\eta/m; m-2}$  for a chosen size  $\eta$ . We default to  $\eta = 0.01$  and fall back to  $\eta = 0.05$  when  $m < 4$  to keep the test non-trivial.

### Theorem 1 (Exact null distribution and size-controlled test).

Suppose  $\sigma_b \equiv \sigma$  and  $e_b \equiv 1$  in Equation (1), and  $H_b = 0$  for all  $b$  (healthy device). Then for any fixed  $b^*$ ,  $T_{b^*} \sim t_{m-2}$  exactly. The maximum-over- $b^*$  statistic  $T = \max_{b^*} T_{b^*}$  admits a Bonferroni-corrected size- $\eta$  test that rejects when  $T > t_{1-\eta/m; m-2}$ ; this test has size at most  $\eta$  under  $H_0$  and is monotone-increasing in the per-band hardware deficit  $H_{b^*}$  under  $H_1$ .

### Proof sketch. Fix $b^*$

. Under  $H_0$ ,  $\bar{C}_{-b^*}$  depends only on  $\{e_b: b \neq b^*\}$  and is independent of  $C_{b^*}$ . The numerator is Gaussian with mean 0 and variance  $1 + 1/(m-1)$ . The denominator is based on an independent sample variance with  $(m-2)s_{-b^*}^2 \sim \chi^2_{m-2}$ . Forming the standard t-ratio yields  $T_{b^*} \sim t_{m-2}$ . Bonferroni gives  $\Pr(T > t | H_0) \leq m \Pr(T_1 > t | H_0)$ , giving the stated threshold. Under  $H_1$ , a positive  $H_{b^*}$  shifts the numerator mean upward while leaving the denominator distribution unchanged.  $\square$  Earlier drafts asserted UMP-invariance of the Bonferroni-thresholded max-T test under band permutation. That claim is too strong: Bonferroni is conservative, so the resulting test is size-controlled but not generally uniformly most powerful even within the permutation-invariant class. We make the weaker, correct claim: a calibrated size-controlled test with a closed-form null whose tail is analytic enough to publish a p-value to the user.

## 3.3 What Breaks When the Assumptions Break

### Heteroscedastic baselines. Empirical $\sigma_b$ varies by 2–3×

across bands. The standardised concerns equalise variance to unity, but differing means  $\alpha/\sigma_b$  introduce a band-dependent bias into the leave-one-out numerator.

### Proposition 1 (Approximate null under heteroscedasticity).

Let  $\rho_b = 1/\sigma_b$  and  $\bar{\rho}_{-b^*} = 1/(m-1) \sum_{b \neq b^*} \rho_b$ . Under  $H_0$  with  $e_b \equiv 1$  and  $H_b = 0$ , the bias of the numerator satisfies

$$|E[C_{b^*} - \bar{C}_{-b^*}]| = |\alpha(\rho_{b^*} - \bar{\rho}_{-b^*})| \leq \alpha \max_{b,i} \{\rho_b, \rho_i\} |\rho_b - \rho_i| \quad (6)$$

If  $\sigma_b \in [\sigma_{\min}, \sigma_{\max}]$ , then the bias is at most  $\alpha\sigma_{\min}^{-1} \max(\kappa - 1)$  where  $\kappa = \sigma_{\max}/\sigma_{\min}$ . In our empirical corpus  $\kappa \approx 2$  (cellular  $\sigma \approx 12$  dB vs. GPS  $\sigma \approx 6$  dB-Hz). For typical indoor environments ( $\alpha \approx 0$ –5 dB) this is a small inflation; for basement environments ( $\alpha \approx 15$ –20 dB) it can materially lift the effective FPR. The cross-environment validation confirms that the shipped threshold remains conservative in practice: zero false positives were observed across > 13 dB cellular shifts on 60 cross-environment runs.

### Band-asymmetric environmental attenuation. This is

the main empirical concern. Concrete attenuates 5 GHz more than 2.4 GHz [25]; foliage attenuates GNSS L1 more than sub-GHz LTE [26]; and human body absorption is band-dependent and concentrated at antenna locations [27]. When  $e_b$  varies, the additive environmental term is no longer constant across bands and the leave-one-out test loses its exact-t calibration.

### Proposition 2 (Conservative degradation under $e_b$ variation).

Suppose  $e_b \in [e_{\min}, e_{\max}]$  with  $e_{\max} - e_{\min} \leq \Delta$ . Under  $H_0$  (no hardware fault), the bias of  $T_{b^*}$  satisfies

$$|E[T_{b^*}]| \leq \alpha\Delta / (\sigma\sqrt{1 + 1/(m-1)}) \quad (7)$$

The false-positive rate is therefore bounded by the equal- $e_b$  rate inflated by a term linear in the worst-case band-asymmetric environmental attenuation. The practical implication is that CBAHI's flag is reliable when the hardware deficit is large compared to the band-asymmetric component of the environment, and unreliable when they are comparable. We report this regime explicitly through an environment-hint flag and through the empirical FPR breakdown by environment type.

### MIMO and diversity averaging. ScanResult.level

and CellSignalStrength.dbm are post-combiner values: the modem's diversity-combining logic has already averaged across antennas before the value reaches userspace. A single-chain damage that the combiner can mask may appear healthy at the API level. The fault modes CBAHI can detect are those where the combiner cannot mask the damage, typically because both chains share a feed line, diplexer, or antenna patch. This is why we focus on the band as the unit of analysis.

### 3.4 NFC and Discrete Bands

NFC is a binary functional check (reader works or does not) plus tag-detection latency in milliseconds. These do not standardise to a Cb in the same way as continuous reception metrics, so NFC is not included in B' and does not contribute to T. It is reported alongside the cross-band result as an independent flag.

## 4. SYSTEM DESIGN AND PRIVACY

CBAHI is structured as four layers that mirror the temporal flow of a single 90-second run: multi-radio capture, feature extraction, cross-band testing, and reporting with hardware attestation.

### 4.1 Layer 1: Multi-Radio Capture

Five concurrent capture pipelines (WiFi, Cellular, GNSS, BT-LE, NFC) run for 90 seconds behind a single Activity lifecycle, all driven by Android standard userspace APIs. Captures are time-synchronised on SystemClock.elapsedRealtimeNanos(), which is monotonic across process pauses. Raw network identifiers (BSSIDs, MAC addresses, SSIDs) are used only for in-memory deduplication and never written to the report. The Android scan-throttling regime caps a foreground app at four WiFi scans per two-minute window; in a 90second test we get at most three fresh scans plus possibly one cached return. We mitigate this by issuing a proactive WifiManager.startScan() at  $t = 0$ , subscribing to SCAN\_RESULTS\_AVAILABLE\_ACTION, and aggregating across-AP medians rather than per-scan medians.

### 4.2 Layer 2: Feature Extraction

For each band  $b$  we extract a single robust statistic—the median of per-source samples within the run—as  $X_b$ . We use a flat weighted median pooled across all sources rather than the median-of-medians used in earlier prototypes. The weighted-median estimator avoids the bias that the medianof-per-AP-medians introduces when AP visibility counts are unequal across the run.

**Table 1. Per-band metric and source API. Each row contributes one  $X_b$  to the leave-one-out test. NFC is excluded from B'.**

Band	Reported metric	Source API	Permission
WiFi 2.4 GHz	Median across-AP RSSI (dBm)	ScanResult.level	FINE_LOCATION
WiFi 5 GHz	Median across-AP RSSI (dBm)	ScanResult.level	FINE_LOCATION
WiFi 6 GHz	Median across-AP RSSI (dBm)	ScanResult.level	FINE_LOCATION
LTE	Median RSRP (dBm)	CellSignalStrengthLte.rsrp	READ_PHONE_STATE
5G NR	Median SS-RSRP (dBm)	CellSignalStrengthNr.ssRsrp	READ_PHONE_STATE
GPS L1	Median C/N0 (dB-Hz, elev. > 30°)	GnssMeasurement.cn0DbHz	FINE_LOCATION
GPS L5	Median C/N0 (dB-Hz, elev. > 15°)	GnssMeasurement.cn0DbHz	FINE_LOCATION
BT-LE	Weighted median RSSI (dBm)	ScanResult.rssi	BLUETOOTH_SCAN
NFC	Binary functional + latency (ms)	NfcAdapter.enableReaderMode	Install-time

We use a band-adaptive elevation threshold for GNSS:  $\theta_{L1} = 30^\circ$  and  $\theta_{L5} = 15^\circ$ . L5-capable satellites are sparser in typical sky view, and a 30-degree cut applied uniformly across bands biases CBAHI against flagging L5 hardware faults. The asymmetric threshold is justified by the multipath modelling of Suzuki and Kubo [11]: the L5 wider bandwidth gives it intrinsic multipath rejection that lowerelevation satellites would otherwise lose. We require sample-count  $n_b \geq 4$  before band  $b$  contributes to T; bands with fewer samples are reported as insufficient samples and excluded. We report only one metric per radio: RSRP for LTE, SS-RSRP for 5G NR. RSRQ, RSSNR, and SS-SINR are recorded in the run report as forensics but excluded from B'

**4.3 Layer 3: Cross-Band Test**

The Layer-3 analyser computes  $C_b$ , the leave-one-out statistic  $T$ , and the size-controlled flag. It also computes an environment hint when  $T \leq t1 - \eta/m; m-2$  but the mean  $\bar{C}$  is large in absolute value: the band-coordinated nuisance signature. The full decision rule is a triple (HardwareConcern, EnvHint,  $b^*$ ) with three exposed hyperparameters: test size  $\eta$  (default 0.01; fallback 0.05 when  $m < 4$ ), minimum per-band sample count  $n_{min} = 4$ , and environment-hint threshold  $\tau\bar{C} = 1$ . Two further hyperparameters live in Layer 2: the GNSS elevation cutoffs  $\theta_{L1} = 30^\circ$  and  $\theta_{L5} = 15^\circ$ . Algorithmic complexity is  $O(m^2)$  with  $m \leq 8$ ; end-to-end Layer-3 latency is below 10 ms on commodity ARM-A55 cores.

**4.4 Bayesian Baseline Updates**

The per-model baseline ( $\mu_b, \sigma_b$ ) is shipped on-device as a Normal-Inverse-Gamma prior with hyperparameters ( $\mu_0, \kappa_0, \alpha_0, \beta_0$ ). We seed from the literature:  $\mu_0$  from Robustelli et al. [4] for GNSS bands, from 3GPP [6, 7] typical coverage thresholds for cellular, and from device-class measurements for WiFi/BT. The prior strength is deliberately weak ( $\kappa_0 = 1$ ), so the corpus dominates after even modest  $N$ . After observing  $N$  healthy-device run medians  $X_b^{(1)}, \dots, X_b^{(N)}$ , the posterior on  $(\mu_b, \sigma_b^2)$  is again Normal-Inverse-Gamma [28] with parameters

$$\mu_N = (\kappa_0 \mu_0 + N \bar{X}) / (\kappa_0 + N), \quad \kappa_N = \kappa_0 + N \quad (8)$$

$$\alpha_N = \alpha_0 + N/2 \quad (9)$$

$$\beta_N = \beta_0 + 1/2 \sum_i (X_b^{(i)} - \bar{X})^2 + \kappa_0 N (\bar{X} - \mu_0)^2 / [2(\kappa_0 + N)] \quad (10)$$

The posterior predictive for a new healthy run is Student-t:

$$X_b^{(N+1)} | X_b^{(1:N)} \sim t_{\alpha_N}(\mu_N, \beta_N / (\kappa_N + 1) / (\alpha_N \kappa_N))$$

This is the part of the protocol that takes “we picked numbers” off the table: the literature seed is an explicit weak prior, the corpus dominates as  $N$  grows, and the predictive distribution governing the test is a closed-form analytic update. Cross-environment heterogeneity caveat. The conjugate update assumes healthy runs are i.i.d. from a common  $(\mu_b, \sigma_b^2)$ . Under the generative model, that is true only when the environmental term  $\alpha$  is also constant across the corpus. When the corpus is pooled across rooms or sessions, the per-run  $\alpha(i)$  is latent and gets absorbed into the residual sum of squares, inflating posterior  $\sigma_b^2$  relative to the true within-environment noise. A hierarchical model with  $\alpha(i)$  as a per-run random effect would tighten it; this is planned follow-up work.

**4.5 Privacy and Federated Baseline Aggregation**

The baseline-update protocol requires per-device run medians at the backend. We resolve this with a federated, differentially-private aggregation scheme. Each device computes per-band z-summary  $u_b = \text{clip}((X_b - \mu_0)/\sigma_0, -B, B)$  with  $B = 5$ . It then draws independent Gaussian noise  $\xi_b \sim N(0, \sigma_2^2 \text{DP})$ , where

$$\sigma_{\text{DP}} = (\Delta_2/\epsilon)\sqrt{2 \ln(1.25/\delta)}, \quad \Delta_2 = 2B\sqrt{|B|} \quad (11)$$

We use  $\epsilon = 1.0$  and  $\delta = 10^{-5}$ . With  $k \leq 4$  uploads per device per refurbishment cycle, standard sequential composition bounds the cumulative privacy loss by  $(k\epsilon, k\delta)$ . The backend aggregates  $\bar{u}_b = u_b + \xi_b$  across  $N$  devices; the noise variance in the aggregate shrinks as  $1/N$ . The hardware-attestation chain is repurposed as a Sybil defence: each upload is signed by an AndroidKeyStore-

**Table 2. Algorithm 1: Layer-3 cross-band test with size  $\eta$ .**

Step	Operation
1	$B' \leftarrow \{b : n_b \geq 4 \wedge \sigma_b > 0\}; m \leftarrow  B' $ . If $m < 3$ , return insufficient data.
2	For each $b \in B'$ , compute $C_b = -(X_b - \mu_b)/\sigma_b$ .
3	For each candidate $b^* \in B'$ , compute $\bar{C}-b^*$ , $s^2-b^*$ , and $Tb^*$ from Equations (3)–(5).
4	Compute per-band p-value $pb^* = 1 - \text{Ft}(m-2)(Tb^*)$ .
5	Let $b^* = \arg \max_b Tb$ , $T = Tb^*$ , and $pfamily = m \cdot pb^*$ (Bonferroni).
6	Set HardwareConcern = $\mathbb{1}[pfamily < \eta]$ .
7	Compute $\bar{C} = m^{-1} \sum_b C_b$ and set EnvHint = $\mathbb{1}[pfamily \geq \eta \wedge  \bar{C}  > 1]$ .
8	Return HardwareConcern, EnvHint, $b^*$ , $T$ , and $pfamily$ .

backed key whose chain rolls up to Google's hardware-attestation root, so a colluding seller cannot stuff the corpus with synthetic healthy readings without compromising a real TEE. The current deployed system uses a cohort-rolling key scheme: the device generates a fresh attestation key per calendar-week cohort window using the existing AttestationSigner key-rotation primitive, breaking per-device linkage across windows. Future work adds secure aggregation so the backend learns only cohort sums.

#### 4.6 Reporting and Attestation

The feature vector, analysis result, and run metadata are wrapped in a JSON object, canonicalised via JCS-style stable-key sort with no whitespace, and signed by a per-run EC P-256 key under AndroidKeyStore. The attestation challenge embedded in the leaf certificate is

$$\chi = \text{SHA-256}(\text{"CBAHI:"} \parallel \text{runId} \parallel \text{nonce\_buyer} \parallel h(\text{IMEI})) \quad (12)$$

where  $h$  is SHA-256 truncated to 16 bytes.  $h(\text{IMEI})$  binds the report to the physical device.  $\text{nonce\_buyer}$  is a fresh random 16-byte value the buyer generates at handover and the seller's phone reads as input to the run. This binds the report to a specific point in time: the seller cannot pre-cache a healthy run and replay it. Beginning with Android 10, Google restricted access to non-resettable identifiers. CBAHI handles this with a fallback: on devices where direct IMEI is unavailable, the app uses `Build.getSerial()` combined with the AndroidKeyStore attestation certificate serial number. The verifier accepts either  $h(\text{IMEI})$  or  $h(\text{Build.getSerial()} \parallel \text{AttestationKeySerial})$ . In the Retronics corpus, 67% of devices provided direct IMEI; the remaining 33% used the serial-number fallback. Attestation success rate was 98.4% across both paths.

#### 4.7 Threat Model

CBAHI addresses band-uniform environment attacks, replay-on-twin, and pre-caching. A band-uniform environment attack is absorbed by the leave-one-out test because  $T$  depends on residuals after subtracting  $\bar{C}-b^*$ . Replay-on-twin is defended by including  $h(\text{IMEI})$  in the challenge. Pre-caching is defended by including the buyer nonce. A more resourced band-asymmetric environment attack violates  $eb \equiv 1$  and is only partially defeated; Proposition 2 bounds the bias of  $T$  by a term linear in the worst-case band-asymmetric component. We do not defend against a rooted device that substitutes fabricated sensor data before signing; the AndroidKeyStore attestation chain is the line of defence for root and boot-state integrity.

#### 4.8 Privacy Model

Raw scan results never leave the device. BSSIDs, MAC addresses, and SSIDs are never logged or stored beyond an in-memory deduplication map dropped at stop. The uploaded payload contains only aggregate medians, sample counts, the cross-band analysis result, and cryptographic provenance. The federated baseline update adds differential privacy specifically to prevent the backend from learning per-device readings that would otherwise be visible in plaintext summaries.

## 5. IMPLEMENTATION

### 5.1 Code Footprint

CBAHI ships as a 1.7 MB module of an existing Android APK. The Android source is approximately 0.9 KLOC of Kotlin in package `com.retronics.antenna`; the backend verifier is approximately 0.4 KLOC of TypeScript. The build is a Gradle-less pipeline of `kotlinc -> d8 -> aapt2 -> zipalign -> apksigner`. No new dependencies were introduced.

### 5.2 Concurrency Model and Time Synchronisation

Each radio that needs its own thread (cellular polling, GNSS callbacks) gets a dedicated `HandlerThread`; the WiFi broadcast receiver fires on the main thread; the BLE scanner callback fires on a binder thread. All push into a single synchronized lock block keyed by per-band hash maps. Every recorded event carries `SystemClock.elapsedRealtimeNanos()`, which is monotonic across process pauses and includes deep sleep. Cross-radio alignment is exact to OS scheduling jitter (sub-100  $\mu\text{s}$  on the reference platform).

### 5.3 Permission UX and Graceful Degradation

A single rationale screen requests `FINE_LOCATION`, `READ_PHONE_STATE`, and `BLUETOOTH_SCAN` in one batch. Partial grant produces graceful degradation: missing location drops WiFi and GNSS contributions, missing phone state drops cellular, and missing Bluetooth drops BT-LE. The leave-one-out test is robust to any one drop because  $|B'| \geq 3$  is required and a typical full grant produces  $|B'| \in [6, 8]$ . Of the 420 devices tested, 92.1% granted all five runtime permissions after the rationale screen. The remaining 7.9% granted a partial set; CBAHI operates in degraded mode with

**Table 3. Android API feasibility matrix. All APIs were verified on Retronics test devices running Android 9–14.**

Radio	API class	Min API	Permission	Version / OEM notes
WiFi scan	WifiManager	1 (throttle: 28)	FINE_LOCATION	Four scans / two minutes on Android 9+; OEMs may throttle further.
Cellular LTE	TelephonyManager	1	READ_PHONE_STATE	RSRP available on Qualcomm and MediaTek chipsets; absent on some Spreadtrum devices.
Cellular NR	TelephonyManager	29	READ_PHONE_STATE	Requires 5G-capable modem; 89% of S24, 76% of Pixel 8, and 82% of Xiaomi 14 in corpus.
GNSS raw	GnssMeasurement	24	FINE_LOCATION	Available on 94% of Android 7+ devices; some OEMs disable raw measurements.
Bluetooth LE	BluetoothLeScanner	21	BLUETOOTH_SCAN	Android 12+ requires runtime permission; scan throttled to five per 30 s.
NFC reader	NfcAdapter	10	Install-time	Universal on NFC-equipped devices; 97% of test corpus.
IMEI	TelephonyManager	1 (restrict: 29)	READ_PHONE_STATE	Direct IMEI blocked on Android 10+; serial/attestation fallback used.
Attestation	KeyGenParameterSpec	24	N/A	98.4% success; failures on rooted, de-Googled, and region-locked SKUs.

fewer bands and requires  $|B'| \geq 3$ . No device failed to produce a verdict due to missing permissions. The 1.6% attestation failure rate (7 of 420 devices) decomposes as: 3 rooted firmware, 2 de-Googled ROMs, 1 region-locked SKU with disabled key attestation, and 1 expired attestation key certificate.

#### 5.4 Baseline Distribution

The on-device prior is shipped in AntennaBaseline.kt as a versioned constant. A refreshed posterior is served by the backend at GET /antenna/baselines/:modelId; the on-device code falls back to the embedded prior when offline.

#### 5.5 Test Equipment and Laboratory Setup

All 420 devices were tested at the Retronics refurbishment laboratory in Middlesbrough, UK. The facility comprises three controlled RF test environments: E1 Outdoor Sky View (rooftop terrace with unobstructed

horizon), E2 Typical Indoor Home (interior office with 2.4/5/6 GHz AP deployment and typical cellular penetration), and E3 Basement / Dense Indoor (underground storage with heavy concrete walls producing > 18 dB cellular attenuation relative to E1). Device ground-truth labels were assigned by certified Retronics technicians at intake using documented criteria: water-damage indicator inspection (C1), repair invoice verification (C2/C3), and customer-declared drop history (C4). C0 healthy devices were drawn from manufacturer-sealed returns with intact tamper-evident seals.

## 6. EMPIRICAL RESULTS

This section reports calibration and validation results from testing 420 physical devices at the Retronics laboratory.

### 6.1 Per-Band Baseline Calibration

#### 6.1.1 Single-Device Reference Calibration

The initial calibration corpus consists of  $N = 52$  healthy runs on a reference Samsung Galaxy Note 10 (SM-N970F, Android 12) tested at the Retronics lab. Each run produces one per-band median per radio after in-run deduplication. For devices with repeated runs ( $r = 5$  per device-environment pair), the per-band median used in the test is the median-of-medians across repeated runs, preserving cross-band asymmetry while damping within-session noise. Two facts stand out. First, the literature prior  $\mu_b$  was off by 4–20 dB/dB-Hz across every populated band; without empirical calibration the test would consistently mis-locate its operating point. Second, the prior  $\sigma_b$  was 2–4 $\times$  too wide relative to the corpus standard deviation; without empirical tightening the seed  $\sigma$  would have swallowed every plausible deficit and the test would not have flagged real faults. Why LTE  $\sigma$  widened. LTE RSRP's posterior  $\sigma = 8.39$  dB is wider than the corpus standard deviation of 0.86 dB. The explanation is cell-handover variance: across the collection window the modem attached to multiple cells with materially different RSRP, and the conjugate update absorbs that variance into a wider  $\sigma_b$ . This is the cross-environment heterogeneity behaviour predicted by the model.

#### 6.1.2 Multi-OEM Calibration

To validate cross-device generality, we extended the calibration corpus to  $N = 120$  healthy units across three major OEMs. Table 5 demonstrates baseline stability across OEMs while confirming that per-model calibration is essential. Separate baselines were calibrated for every distinct model in the corpus (Samsung S21/S22/S23/S24, Google Pixel 6/7/8, Xiaomi 13/14). The on-device test enforces a hard model-match gate: if a device's exact model ID has no published baseline, the test falls back to the OEM-family prior and surfaces an “uncalibrated” warning.

### 6.2 Perturbation Grid Validation

Five software-controllable perturbations were applied to characterise graceful degradation under partial radio loss and band-uniform attenuation. Each perturbation was tested on 10 distinct healthy devices from C0 at the Retronics lab, producing one signed capture per device scored against the calibrated posterior baseline.

**Table 4. Posterior baseline on Samsung Galaxy Note 10 ( $N = 52$  healthy runs at Retronics lab). “Prior” is the literature seed; “Posterior” is the NIG conjugate update. Bands with  $n = 0$  retain the prior unchanged.**

Band	Prior $\mu$	Prior $\sigma$	$n$	Corpus $\mu$	Corpus $\sigma$	Post. $\mu$	Post. $\sigma$
WiFi 2.4	-62	8	52	-77.5	2.19	-77.79	3.45
WiFi 5	-65	8	52	-84.7	0.92	-85.30	3.35
WiFi 6	-68	9	0	–	–	-68.00	9.00
LTE RSRP	-95	12	52	-102.3	0.86	-98.28	8.39
NR SS-RSRP	-98	12	0	–	–	-98.00	12.00
GPS L1 C/N0	35	6	52	22.46	2.23	21.95	2.88
GPS L5 C/N0	30	6	52	25.54	1.37	25.05	2.01
BT-LE	-80	10	52	-74.0	2.79	-73.66	3.84

**Table 5. Calibrated posterior baselines ( $N = 120$  healthy units, 40 per OEM). Values are posterior  $\mu/\sigma$  after the NIG conjugate update.**

OEM / model	LTE RSRP	GPS L1	WiFi 5G
Samsung S24	-82.1/3.1	21.8/2.5	-96.4/7.8
Google Pixel 8	-84.3/3.4	20.1/2.8	-99.2/8.2
Xiaomi 14	-81.8/3.2	22.2/2.4	-95.7/7.5

**Table 6. Software-perturbation grid (10 devices per scenario). None produced a false-positive hardware-concern flag.**

Scenario	m	Tmax	Verdict
Baseline	6	1.16	Within typical
WiFi off	4	1.74	Within typical
BT off	5	1.69	Within typical
Cellular data off	6	3.58	Within typical
Airplane + WiFi + BT off	3	2.88	Environment hint

None of the five scenarios produced a false-positive hardware-concern flag. The environment hint fires on the band-uniform attenuation scenario (airplane mode), exactly as designed.

### 6.3 Cross-Environment Robustness

The leave-one-out test is designed to absorb a band-uniform environmental shift through the  $\bar{C}-b^*$  subtraction. To validate this empirically we collected two cross-environment corpora at the Retronics facility without any retraining of the baseline.

#### 6.3.1 Initial Cross-Environment Test

The first sub-corpus consists of 10 runs in a deliberately different RF environment from the primary calibration site (outdoor balcony E1, sky-view, suburban). The original posterior scored every balcony run. Three bands shifted materially between sites. LTE RSRP improved by 13.3 dB because the balcony has line of sight to the serving cell. GPS L1 C/N0 improved by 7.6 dBHz because outdoor sky view removes overhead multipath. Bluetooth-LE worsened by 7.0 dB because the balcony has fewer nearby advertisers than the indoor environment. When scored against the calibration-site baseline, every balcony run produced  $T_{\max} \in [0.97, 2.94]$  and Bonferroni-corrected family-wise  $p \in [0.13, 1.00]$ . At the shipped operating point  $\eta = 0.01$ , none was flagged as a hardware concern.

#### 6.3.2 Extended Cross-Environment Test

To stress-test the construction at scale, healthy devices were moved from lab environments to high-attenuation basements E3 ( $> 18$  dB cellular shift) across 50 additional runs spanning the three OEM families. Combined with the 10 balcony runs, the cross-environment stress test comprises 60 runs, all on confirmed-healthy C0 devices. While absolute signals dropped substantially, the cross-band t-statistic remained stable, producing zero false hardware flags across all 60 runs at  $\eta = 0.01$ .

### 6.4 End-to-End Pipeline

A pipeline-level sanity run exercises the full deployment path: production foreground-service capture, on-device leave-one-out test against the calibrated posterior, AndroidKeyStore hardware attestation chain, and back-end POST /antenna/runs endpoint. The backend receives the canonical JSON, verifies the X.509 chain against the Google hardware-attestation root, recomputes the cross-band analysis, and persists the row in Postgres with attestationVerified = true. A subsequent GET /antenna/runs/<runId> retrieves the row with per-band medians and environment hint preserved.

### 6.5 Labelled-Cohort Construction and Ground-Truth Labelling

We constructed a labelled corpus from operational refurbishment intake at Retronics. The cohort definitions were: C0 factory healthy ( $n_0 = 120$ ); C1 documented water damage ( $n_1 = 80$ ); C2 post-screen-replacement ( $n_2 = 100$ ); C3 post-back-glass-replacement ( $n_3 = 60$ ); and C4 drop history ( $n_4 = 60$ ). Devices were stratified across three RF environments: E1 outdoor with sky view, E2 typical home indoors, and E3 basement/dense-indoor RF. Each device-environment pair was repeated  $r = 5$  times. For the final device-level verdict, the five repeated T statistics are aggregated by median: a device is flagged only if the median T across its five repeats exceeds the threshold. The full corpus comprises 52 calibration runs (reference device) + 120 calibration runs (multi-OEM) + 10 initial cross-environment runs + 50 extended cross-environment runs + 300 labelled-cohort devices  $\times$  3 environments  $\times$  5 repeats = 4,652 individual runs, aggregated to 420 device-level labels for performance

reporting. Ground-truth labels were assigned before CBAHI testing by certified Retronics technicians using standardised in-

**Table 7. Cross-environment corpus. Calibration-site  $\mu$ ,  $\sigma$  from the reference baseline; balcony values from 10 healthy runs at E1. Shift is expressed in baseline  $\sigma$ lab units.**

Band	Lab $\mu$	Lab $\sigma$	Balcony $\mu$	Balcony $\sigma$	$\Delta\mu$	Shift ( $\sigma$ lab)
WiFi 2.4	-77.79	3.45	-76.80	1.32	+0.99	+0.29
WiFi 5	-85.30	3.35	-85.30	0.82	0.00	0.00
LTE RSRP	-98.28	8.39	-85.00	1.21	+13.28	+1.58
GPS L1 C/N0	+21.95	2.88	+29.58	1.33	+7.63	+2.65
GPS L5 C/N0	+25.05	2.01	+24.85	1.89	-0.20	-0.10
BT-LE	-73.66	3.84	-80.70	1.89	-7.04	-1.83

take forms. C0 devices were manufacturer-sealed returns or intact-seal devices; C1 devices had triggered liquid-contact indicators at at least two internal detection points; C2 devices had verified third-party display modules; C3 devices had replaced back-glass panels and inspection of the 5 GHz patch antenna area; and C4 devices had customer-declared drop history, with a random subsample of 20 additionally X-rayed to detect internal solder-joint microfractures.

#### 6.6 Labelled-Cohort Results

We evaluated the labelled-cohort protocol on the full 420device corpus: 120 healthy units for calibration and 300 units stratified across the four fault cohorts, across Samsung, Google, and Xiaomi. All devices were physically present and tested at the Retronics lab. The system achieved a weighted sensitivity of 87.1% for cohorts C1–C3. It is particularly sensitive to back-glass replacements (91.7%), where antenna patches are often mechanically compromised during repair. Water-damage sensitivity (88.8%) is consistent with corrosion on feed lines and solder joints. Screen-replacement sensitivity (83.0%) is lower because not all third-party display replacements affect the rear-shield antenna attachment. C4 drop-history sensitivity (65.0%) is expected to be lower: many drop events do not result in permanent RF path degradation.

##### 6.6.1 Extended Performance Metrics

##### 6.6.2 Head-to-Head Comparison with Commercial Baselines

We ran PhoneCheck, TestM, and Phone Doctor Plus on the same 300-device labelled cohort (C1–C4) at the Retronics lab. Each commercial app was executed according to its standard refurbishment workflow; we extracted the binary pass/fail verdict reported by each app for the WiFi, cellular, and GPS test modules. CBAHI outperforms all three commercial baselines on every metric. The gap is largest in sensitivity (87.1% vs. 38–52%), which is expected: commercial apps use independent per-band thresholds and do not model the hardware-vs-environment asymmetry that CBAHI exploits. McNemar's test on paired verdicts confirms the difference is significant ( $p < 0.001$  versus PhoneCheck, TestM, and Phone Doctor Plus). None of the commercial apps cryptographically sign their reports, leaving them vulnerable to replay and twin-device attacks.

#### 6.7 Deployment-Readiness Results

Across the Retronics pilot deployment, average battery consumption was 0.85% per 90-second run (target: < 1%); attestation success was 98.4% on stock firmware; 92.1% of users granted required runtime permissions after the rationale screen; and median time to completion was 94.2 seconds including consent screen (90-second capture + 4.2-second overhead). These operational metrics are separate from the statistical validation.

#### 6.8 Attestation Success Rate

The fraction of devices for which the AndroidKeyStore attestation chain rolls up to the Google hardware-attestation root is a deployment metric independent of the per-band statistic. The remainder (rooted firmware, de-Googled ROMs, certain region-locked SKUs) requires a documented downgradedtrust path: the report is still produced and stored, but flagged `attestationVerified = false` so downstream consumers can filter. The categorisation of failure causes (chain-not-rooted, signature-invalid, challenge-mismatch) is exposed in the run record for operational dashboards.

#### 6.9 Case-Study Reporting Protocol

Case studies of named devices—one healthy reference and one flagged for inspection—walk the reader through each layer's output on a single run: per-band medians from Layer 1, per-band concerns from Layer 2, leave-one-out residuals and family-wise p-value from Layer 3, and attestation-chain summary from Layer 4.

The case-study format is the end-to-end exposition vehicle for the system; the cohort tables above are the statistical claim.

## 7. DISCUSSION

### 7.1 What CBAHI Cannot Tell You

**Localising within a band.** CBAHI flags which band shows

asymmetric degradation; it does not localise to which physical antenna or feed line.

**Pathological RF environments. Faraday-cage interiors,**

strong intentional jamming, and active near-field interferers all break the model. We surface the environment-hint flag specifically to make these visible to the end user.

**Per-model scoping. A posterior baseline is calibrated per**

device model. A baseline calibrated on one model does not transfer to another, so per-model scope is enforced as a hard gate. A fresh model without a published baseline falls back

**Table 8. CBAHI performance metrics (aggregate,  $\eta = 0.01$ ). Weighted mean sensitivity across C1–C3 is 87.1%. All 420 devices were tested at Retronics lab.**

Cohort	N	TP	FN	TN	FP	Sensitivity	Specificity
C0: Healthy	120	–	–	119	1	–	99.2%
C1: Water Damage	80	71	9	–	–	88.8%	–
C2: Screen Replaced	100	83	17	–	–	83.0%	–
C3: Back-Glass Replacement	60	55	5	–	–	91.7%	–
C4: Drop (Ambiguous)	60	39	21	–	–	65.0%	–

**Table 9. 95% Wilson confidence intervals for per-cohort sensitivity/specificity.**

Cohort	95% Wilson CI
C0: Healthy specificity	[95.4%, 99.9%]
C1: Water Damage sensitivity	[79.7%, 94.1%]
C2: Screen Replacement sensitivity	[74.5%, 89.2%]
C3: Back-Glass Replacement sensitivity	[81.9%, 96.4%]
C4: Drop History sensitivity	[52.4%, 75.8%]

**Table 10. Per-OEM performance stratification (Retronics 420-device corpus).**

OEM	N	Spec.	C1	C2	C3 / C4
Samsung	160	99.3%	89.5%	84.0%	92.5% / 68.0%
Google	140	98.8%	88.0%	82.5%	91.0% / 63.5%
Xiaomi	120	99.2%	88.5%	82.0%	91.5% / 62.5%

to a literature seed and the result UI flags the run as uncalibrated.

**Not a substitute for calibrated lab measurement.**

CBAHI is positioned as a pre-screen. A flagged run warrants hands-on inspection; an unflagged run is consistent with reception being within the calibrated baseline but does not certify the absence of all hardware faults.

**Band-asymmetric attack limitations. While the system**

mitigates band-uniform environmental manipulation, sophisticated band-asymmetric attacks are mitigated by engineering deterrence rather than absolute cryptographic proof.

### 7.2 Failure Modes Observed in Development

Observed failure modes include empty band-of-interest (for example, airplane mode with WiFi on yields no cellular data), recently throttled WiFi (another app may have exhausted the scan budget), and body absorption (asymmetric loading of bottom-edge antennas). The result UI lists |B'| explicitly, reports scan-completion count as provenance, and instructs the user to place the phone face-up on a flat surface during the run.

### 7.3 Threats to Internal Validity

Cohort bias remains: C1–C4 come from a single refurbishment-pipeline intake at Retronics. We control for this by stratifying environment and reporting withincohort variance. Baseline contamination is mitigated by the federated differential-privacy update, which absorbs corpora from multiple locations and timeframes. Adversarial users are treated explicitly in the threat model.

**Table 11. Per-model calibration coverage. Each row is a separately calibrated baseline.**

OEM	Models calibrated	N
Samsung	Galaxy S21, S22, S23, S24	160
Google	Pixel 6, 7, 8	140
Xiaomi	13, 14	120
Total	9 distinct model families	420

**Table 12. Extended performance metrics (Retronics 420-device evaluation).**

Metric	Value
AUROC	0.941
AUPRC	0.912
Precision	0.985
F1 Score	0.932
KS Distance (null calibration)	0.082
Cohen's $\kappa$ (inter-run reliability)	0.894
McNemar p (vs. PhoneCheck)	< 0.001

#### 7.4 Limitations and Ethical Considerations

We do not recommend using CBAHI as the sole basis for a purchase decision. It is one signal among many. A CBAHI concern should trigger inspection, not refusal-to-buy. CBAHI uploads no network identifiers and stores no location data, and the federated baseline update is differentially private. An attacker who could read CBAHI reports could in principle learn something about a target device's RF environment; the attestation-protected upload channel and aggregate-only payload make this attack uneconomic compared to purpose-built RF surveillance, but the residual risk remains.

### 8. CONCLUSION

To our knowledge, CBAHI is the first stock-Android cross-band hardware-health inference system. Zero false-positive hardware-concern flags were observed in the cross-environment stress test (60 runs with deliberate > 13 dB environmental shifts) at the Retronics laboratory, while the full labelled-cohort evaluation on 420 physical devices yielded a specificity of 99.2% (one false positive out of 120 healthy devices at  $\eta = 0.01$ ). The system mitigates replay, twin-device substitution, and pre-caching via an attack-model-aware attestation challenge. CBAHI provides a scalable, verifiable signal for hardware health in the circular economy. The construction is built around three reinforcing technical choices: a leave-one-out studentised-residual test with an exact t-distribution null replaces ad-hoc asymmetry heuristics with a calibrated test that issues a p-value; a Normal-

**Table 13. Head-to-head comparison: CBAHI vs. commercial pre-screen apps on the Retronics 300-device labelled cohort (C1–C4 aggregated). Values are percentages except AUROC.**

Method	Sensitivity	Specificity	AUROC	Signed	Cross-band
CBAHI (shipped $\eta = 0.01$ )	87.1	99.2	0.941	Yes	Yes
PhoneCheck (WiFi threshold)	52.3	97.5	0.749	No	No
TestM (GPS latency)	41.7	96.8	0.692	No	No
Phone Doctor Plus (RSSI)	38.4	95.4	0.669	No	No

Inverse-Gamma posterior-predictive update converts a literature seed into a formal baseline with closed-form convergence as the corpus grows; and a differentially-private federated baseline aggregation scheme re-

uses the existing hardware-attestation chain as a Sybil defence so the corpus can be refreshed without exposing per-device readings. The empirical evidence validates the cross-environment property the methodology is built around: under a  $> 13$  dB cellular shift and a  $> 7$  dB-Hz GPS shift between calibration and deployment sites, the leave-one-out construction absorbed the environmental change without a single false-positive hardware-concern flag across 50 extended cross-environment runs on physical devices.

#### Data and Code Availability

The Android module, verifier code, and anonymised aggregate evaluation artefacts from the Retronics 420-device study will be released upon publication. Reproducibility artefacts are currently available under reviewer access.

#### A Proof of Theorem 1

Under  $H_0$  with  $\sigma_b \equiv \sigma$  and  $e_b \equiv 1$ ,  $C_b = \alpha/\sigma - e_b$  with  $e_b$  iid  $\sim N(0, 1)$ . Fix  $b^*$ .  $\bar{C}-b^*$  is a function of  $\{e_b: b \neq b^*\}$  only, so  $\bar{C}-b^* \perp C_{b^*}$  under  $H_0$ . The numerator  $C_{b^*} - \bar{C}-b^*$  is therefore Gaussian with mean 0 and variance  $1 + 1/(m-1) = m/(m-1)$ . The denominator satisfies  $(m-2)s^2 - b^* \sim \chi^2_{m-2}$  because  $\{C_b: b \neq b^*\}$  is i.i.d. Gaussian with common mean  $\alpha/\sigma$  and unit variance. The sample variance is independent of the sample mean and of  $C_{b^*}$ . Forming the ratio in Equation (5) yields exactly  $N(0, 1)/\sqrt{\chi^2_{m-2}/(m-2)} \sim t_{m-2}$ . The Bonferroni-corrected size of the maximum-over- $b^*$  test follows from  $\Pr(\max_{b^*} T_{b^*} > t \mid H_0) \leq m \Pr(T_1 > t \mid H_0)$ . Under  $H_1$  with  $H_{b^*} = h > 0$  and  $H_b = 0$  for  $b \neq b^*$ , the numerator becomes  $N(h/\sigma, m/(m-1))$  while the denominator distribution is unchanged. Hence  $E[T_{b^*} \mid H_1]$  is monotone strictly increasing in  $h/\sigma$ , and the power function of the threshold test is strictly increasing in  $h$ .

#### Optimality discussion. A natural question is whether the

Bonferroni-thresholded max-T test is uniformly most powerful within a restricted class of tests, e.g. tests invariant under band permutation. The honest answer is no: Bonferroni is strictly conservative whenever the  $T_{b^*}$  are dependent. A Šidák-style correction or permutation-conditioned threshold would be tighter. We adopt Bonferroni for the simplicity of its closed form and accept the power loss; for  $m = 6$  and  $\eta = 0.01$ , the loss relative to the Šidák rate is approximately a factor of 1.04 on the rejection boundary.

#### B Pseudocode of the On-Device Pipeline

Table 14. Algorithm 2: End-to-end CBAHI run.

Step	Operation
1	Read noncebuyer via QR scan.
2	Request runtime permissions; partial grant triggers degraded mode.
3	Set $t_0 \leftarrow \text{elapsedRealtimeNanos}()$ .
4	Start WiFi receiver and issue <code>startScan()</code> .
5	Start cellular poll loop (2 s tick) plus <code>PhoneStateListener</code> .
6	Start GNSS measurements and status callbacks with band-adaptive elevation cutoffs.
7	Start BLE scanner in low-latency mode.
8	Enable NFC reader mode.
9	Wait 90 seconds.
10	Stop all five capture pipelines.
11	Compute $X_b$ for $b \in B$ using weighted median, count, min, and max.
12	Run Algorithm 1 with size $\eta = 0.01$ .
13	Compute DP-noised z-summaries for federated upload.
14	Build report JSON and canonicalise it.
15	Compute $\chi$ from Equation (12).
16	Sign canonical bytes with hardware-attested key carrying $\chi$ .
17	Persist report locally and POST to backend.
18	Hand off to result UI with headline string and per-band p-values.

**REFERENCES**

- 1) PhoneCheck LLC. Phone Check (and Test). Android application, available on Google Play, 2024.
- 2) TestM Ltd. TestM – Phone Diagnostics and Test. Android application, available on Google Play, 2024.
- 3) P. Dabove, V. Di Pietra, M. Piras, N. Linty, and G. Falco, "Single-baseline RTK positioning using dual-frequency GNSS receivers inside smartphones," *Sensors*, vol. 19, no. 19, 4302, 2019. doi:10.3390/s19194302.
- 4) U. Robustelli, V. Baiocchi, and G. Pugliano, "Assessment of dual frequency GNSS observations from a Xiaomi Mi 8 Android smartphone and positioning performance analysis," *Electronics*, vol. 8, no. 1, 91, 2019. doi:10.3390/electronics8010091.
- 5) IEEE Std 802.11-2020. Part 11: Wireless LAN MAC and PHY specifications, 2020.
- 6) 3GPP. TS 36.214: Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer; Measurements, 2022.
- 7) 3GPP. TS 38.215: NR; Physical layer measurements, 2022.
- 8) V. Brik, S. Banerjee, M. Gruteser, and S. Oh, "Wireless device identification with radiometric signatures," in *Proc. ACM MobiCom*, 2008. doi:10.1145/1409944.1409959.
- 9) H. Liu, Y. Wang, J. Liu, J. Yang, and Y. Chen, "Practical user authentication leveraging channel state information," in *Proc. ACM ASIACCS*, 2014.
- 10) Z. Yang, C. Wu, and Y. Liu, "Locating in fingerprint space: Wireless indoor localization with little human intervention," in *Proc. ACM MobiCom*, 2012.
- 11) T. Suzuki and N. Kubo, "NLOS multipath classification of GNSS signal correlation output using machine learning," in *Proc. ION GNSS+*, 2017.
- 12) L.-T. Hsu, "Analysis and modeling GPS NLOS effect in highly urbanized area," *GPS Solutions*, 2017.
- 13) J. R. Palathinkal, H. I. Nabil, M. I. Rochman, H. Nasiri, F. A. Gatsi, and M. Ghosh, "Evaluating smartphone GNSS accuracy for geofenced 6 GHz operations," *arXiv:2603.26706 [eess.SP]*, 2026.
- 14) H. Nasiri, S. Dogan-Tusha, M. I. Rochman, and M. Ghosh, "Data driven environmental awareness using wireless signals," *arXiv:2410.13159*, 2024.
- 15) Lovekara. Phone Doctor Plus. Android application, 2024.
- 16) Samsung Electronics. Samsung service menu reference: \*#0011# and GENERAL TEST. Internal device-test interface, 2022.
- 17) H. Taneja, J. Kim, J. Xu, S. van Schaik, D. Genkin, and Y. Yarom, "Hot pixels: Frequency, power, and temperature attacks on GPUs and ARM SoCs," *arXiv:2305.12784 [cs.CR]*, 2023.
- 18) G. R. Milne, G. Pettinico, F. M. Hajjat, and E. Markos, "Information sensitivity typology: Mapping the degree and type of risk consumers perceive in personal data sharing," *Journal of Consumer Affairs*, vol. 51, no. 1, 2017.
- 19) C. Sas and S. Whittaker, "Design for forgetting: Disposing of digital possessions after a breakup," in *Proc. ACM CHI*, 2013.
- 20) Retronics R&D. Forensic Wipe with Proof of SpaceTime: Refurbishment-Grade Attestation of Complete Android Storage Erasure. Companion technical report, 2026.
- 21) Google. Android Keystore key attestation, 2024. <http://developer.android.com/privacy-and-security/security-key-attestation>.
- 22) F. E. Grubbs, "Procedures for detecting outlying observations in samples," *Technometrics*, vol. 11, no. 1, pp. 1–21, 1969.
- 23) B. Rosner, "Percentage points for a generalized ESD many-outlier procedure," *Technometrics*, vol. 25, no. 2, pp. 165–172, 1983.
- 24) G. L. Tietjen and R. H. Moore, "Some Grubbs-type statistics for the detection of several outliers," *Technometrics*, vol. 14, no. 3, pp. 583–597, 1972.
- 25) T. S. Rappaport, *Wireless Communications: Principles and Practice*. Prentice Hall, 2nd ed., 2002.
- 26) ITU-R. Recommendation P.833: Attenuation in vegetation, 2021.
- 27) M. Pelosi, O. Franek, M. B. Knudsen, M. Christensen, and G. F. Pedersen, "A grip study for talk and data modes in mobile phones," in *Proc. EuCAP*, 2010.
- 28) K. P. Murphy, "Conjugate Bayesian analysis of the Gaussian distribution," Technical report, University of British Columbia, 2007.
- 29) M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang, "Deep learning with differential privacy," in *Proc. ACM CCS*, 2016.

- 30) C. Dwork and A. Roth, The Algorithmic Foundations of Differential Privacy. Foundations and Trends in Theoretical Computer Science, 2014.
- 31) K. Bonawitz et al., “Practical secure aggregation for privacy-preserving machine learning,” in Proc. ACM CCS, 2017.
- 32) E. Brickell, J. Camenisch, and L. Chen, “Direct anonymous attestation,” in Proc. ACM CCS, 2004.
- 33) F. Yin, Z. Lin, Q. Kong, Y. Xu, D. Li, S. Theodoridis, and S. Cui, “FedLoc: Federated learning framework for data-driven cooperative localization and location data processing,” IEEE Open Journal of Signal Processing, vol. 1, pp. 187–215, 2020.