

Background · Why HIE matters · Why cUS, not MRI, in the cooling window

1-3 / 1k live births UK incidence [1]	6 h hypothermia window from birth [2]	£0.8-1.0m lifetime cost severe HIE/CP [13]	£11.2m avg CP/brain-injury claim 2024-25 [16]
--	--	---	--

HIE = acute brain injury in the term/near-term newborn from peri-partum O₂ and blood-flow drop. Leading non-syndromic cause of neonatal death and cerebral palsy [1]. **Biphasic injury** — initial energy failure during the insult, then **secondary energy failure 6-48 h** later during which most lasting neuronal loss takes place. Therapeutic hypothermia (33.5 °C, 72 h) interrupts the secondary phase **only if started within 6 h** [2].

Sarnat 1	mild — hyperalert, no seizures, excellent prognosis.
Sarnat 2	moderate — lethargy, seizures, the cohort that most benefits from cooling.
Sarnat 3	severe — stupor/coma, prolonged seizures, high mortality even with cooling.

AI triage matters most at the moderate-severe boundary, where the cooling decision is closest and inter-reader variability is highest.

Dimension	cUS	MRI
Time at scan	15-30 min bedside	30+ min table
Sedation	none	often required
Cost (NHS)	low (point-of-care)	≈ 4-6× cUS
Out-of-hours	24/7 bedside [4]	limited [4]
Time-to-report	real-time [4]	≤ 2 days [4]
Cooling window	primary [2]	post-rewarm gold std [3]
Sens. early HIE	limited	—
Sens. d4-10	95% NPV norm [3]	81% adverse out. [3]

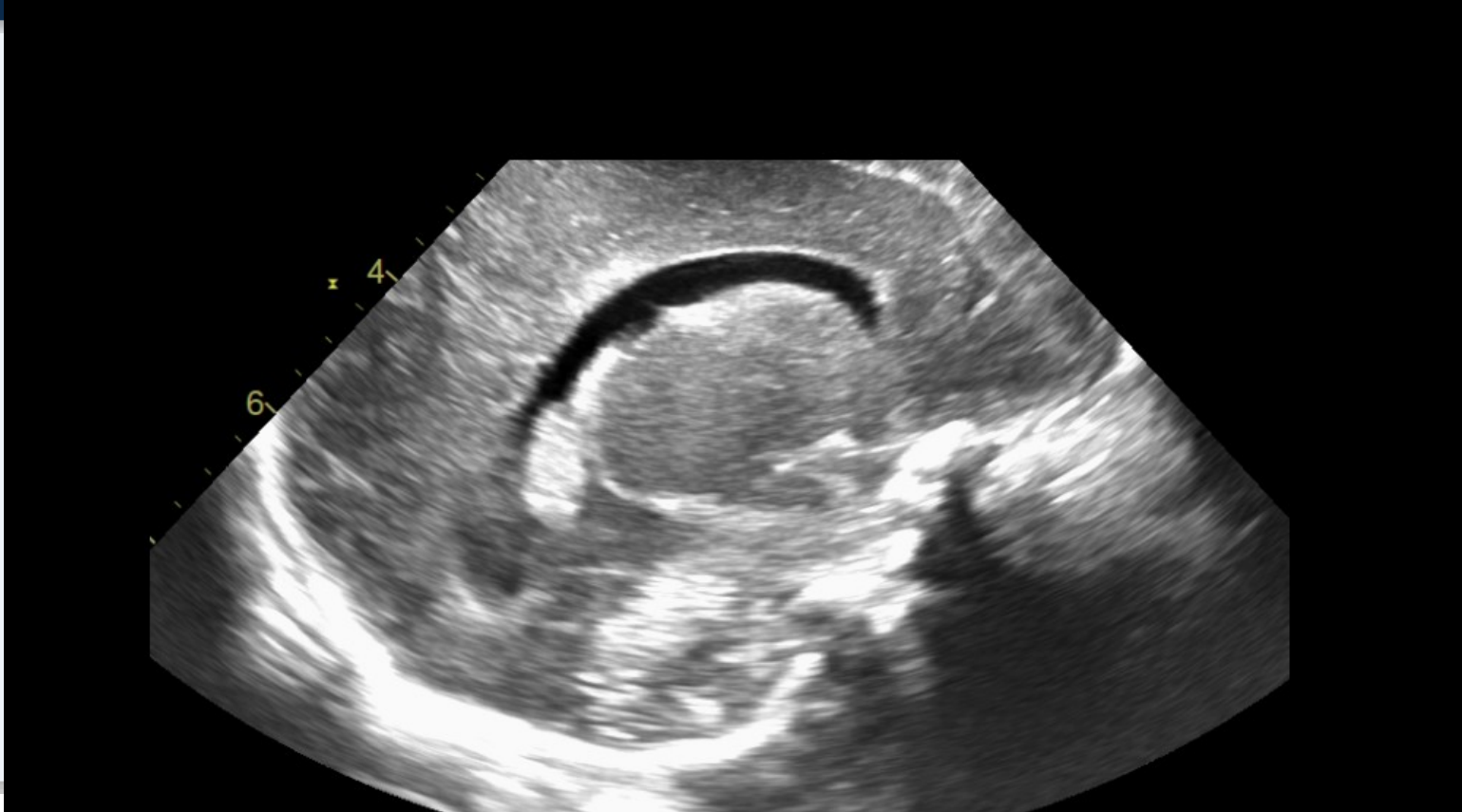
What cUS shows in HIE · Sample anonymised frame

HIE-suggestive findings · order of severity

1. Diffuse cerebral oedema (effacement of sulci)
2. Bilateral basal-ganglia-thalamic hyperechogenicity (BGT pattern)
3. Periventricular hyperechogenicity
4. Cortical highlighting (early HIE sign)
5. Reduced / reversed diastolic flow on cerebral-artery Doppler
6. Cystic evolution at d7-21 (cystic PVL / encephalomalacia)

NHS scan timeline: within 24 h → d5-7 → d21-28 — early oedema → evolving echogenicity → established cystic change.

"Any AI gain on cUS is systemically valuable in a way an equivalent MRI gain is not — cUS is the only modality compatible with the 6-hour cooling-decision window, and the only one that is fast, bedside, radiation-free, sedation-free, available 24/7, and an order of magnitude cheaper."



Preserved
Brain anatomy, fan geometry, depth scale, annotation arrows radiologists use.

Removed
Patient name, MRN, gestation overlay, in-image captions, scanner-burnt headers.

Methods · pipeline · cohort · architecture

1 RIS query	Soliton RIS · 'US Cranial contents' · neonates · 1 m · urgent · 2023-26
2 Linkage	RIS → PACS → Spreadsheet ↔ date-ordered
3 Anonymisation	DICOM header strip + safety-net pixel redaction
4 Labels	Clinical reports as labels (3-reader adjudication deferred)
5 AI + report	Foundation encoder + linear probe; clinician sign-off

Inclusion criteria

- Neonates < 1 month at examination
- 'US Cranial contents' on Soliton RIS · Priority Urgent
- Single named consultant authoriser
- Examination date 2023-05-03 → 2026-05-03

AI architecture · Frozen foundation encoder (USFM [9] / BiomedCLIP [10] / RadImageNet [11]) → VLM head (zero-shot vs radiologist prompts) → patient-mean pool → logistic regression (LOFO + bootstrap CI) + permutation null → LLM reporter conditioned on classifier; clinician sign-off. TRIPOD-AI [5] · CLAIM [6].

Honest limit: 3-reader adjudication deferred; cohort uses clinical reports as labels — caps label quality.

Table 1 · Baseline characteristics

Patients identified on RIS	59
Sex M/F	33 / 26
Age at scan (d) median (IQR)	1 (0-4)
Approved → PACS	21 / 59 (35.6%)
Rejected — burnt-in PII	36 / 59 (61.0%)
Rejected — not on PACS	11 / 59 (18.6%)
Patients ≥ 1 approved scan	14
Patients ≥ 2 sequential scans	4
Modelling cohort	626 frames · 14 pts · 20 events

Reading the cohort: 59 babies had urgent cUS in our 3-year window. **36 (61%) rejected** for burnt-in patient names in pixel data — solvable with DICOM-only export. **Modelling cohort 14 pts / 20 events / 4 with ≥ 2 sequential scans.** The pipeline is device-agnostic — at N ≥ 100 it yields the clinical-grade dataset the abstract proposed, no architecture change.

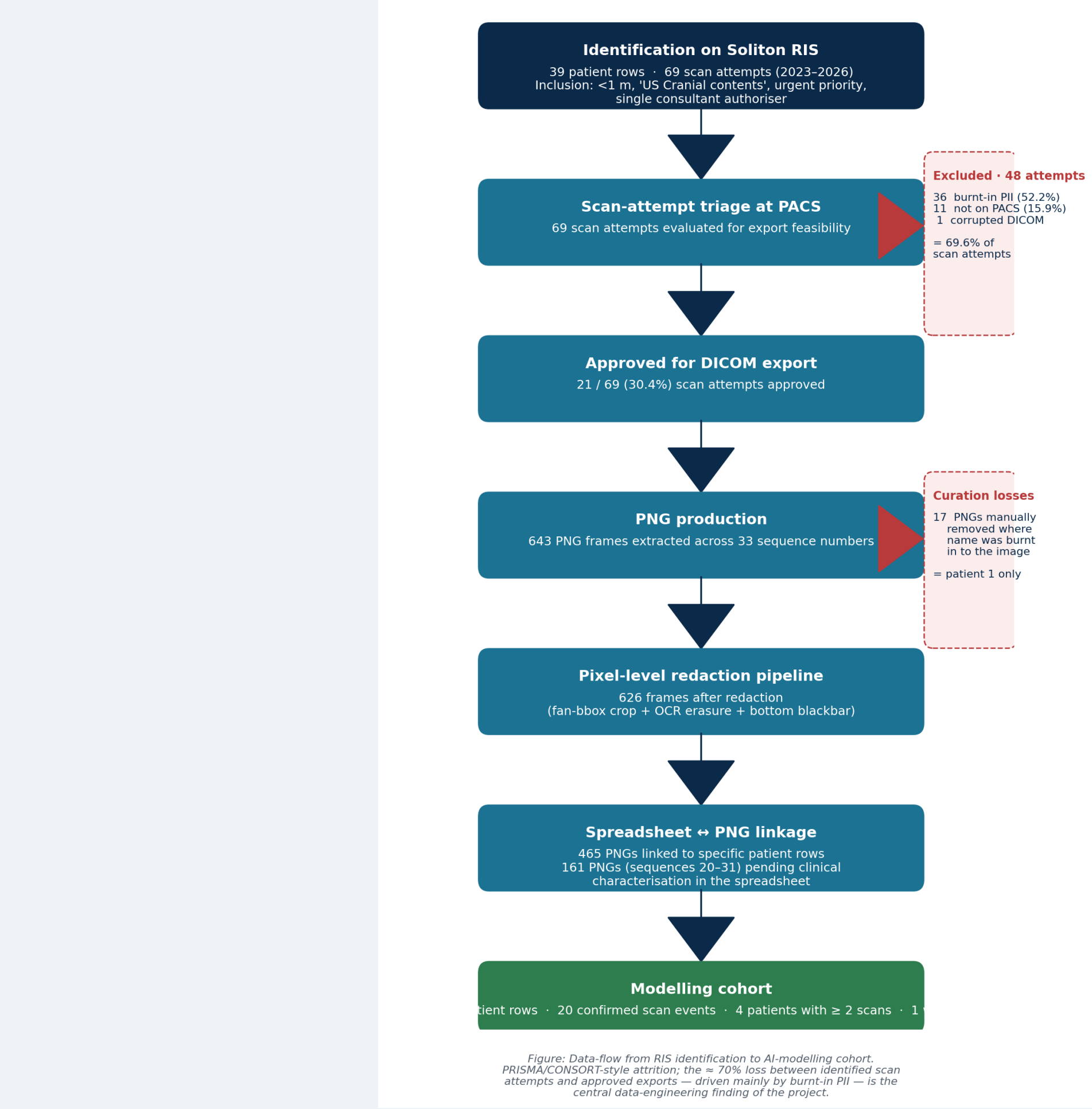
Headline result · feasibility signal

5.48× fold over chance · 14 linked patients (LOFO 5-NN, 30 reps)	7.2× fold over chance · 33 sequence groups (LOFO 5-NN)	p<10⁻²⁰⁰ intra-vs-inter-patient cosine Mann-Whitney
---	---	--

In plain terms: if random guessing would name the correct baby 1 in 14 times, our AI lands on the right one **5.48× more often** — and the gap between same-baby and different-baby frames is so large the statistical test (**p < 10⁻²⁰⁰**) effectively rules out chance.

Figures · data flow · feasibility signal · learning curve

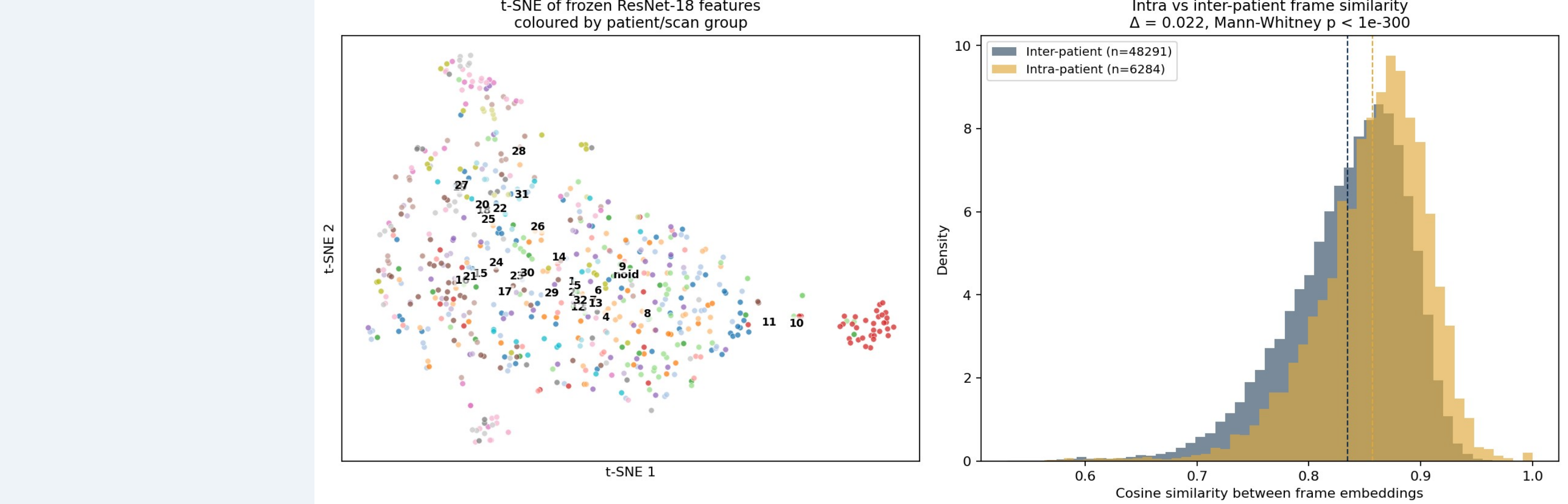
Figure 1 · Data flow (PRISMA-style)



Data flow from RIS identification to AI-modelling cohort. Of 69 scan attempts identified at one UK NHS centre over four years, 30.4 % were approved for DICOM export, with 52.2 % rejected for burnt-in patient identifiers in pixel data. Final modelling cohort: 14 patient rows / 20 scan events / 626 redacted frames; four patients have ≥ 2 sequential approved scans.

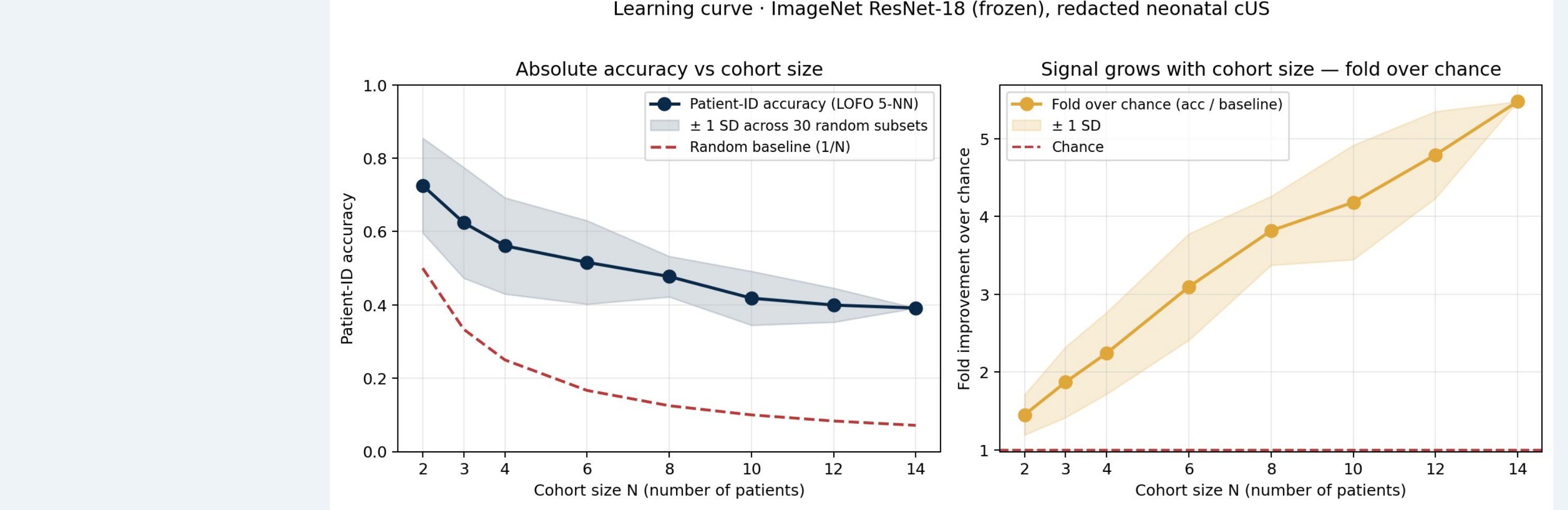
Main bottlenecks: (1) burnt-in PII on Aplio i800 (52 %) — solvable via DICOM-only export; (2) RIS→PACS mismatch (16 %) — workflow audit; (3) cohort-linkage queue (clinical characterisation pending).

Figure 2 · Feasibility signal (t-SNE + similarity histogram)



LEFT (t-SNE map). Each dot is one ultrasound frame. Frames from the same baby cluster together — baby 11/10 (bright red, right edge) is its own island. Proves the AI distinguishes patient anatomy. **RIGHT (similarity histogram).** Gold = same-baby pairs · Navy = different-baby pairs. Δ = 0.022, Mann-Whitney p < 10⁻²⁰⁰ — effectively impossible by chance. Bottom line: the AI has learned baby-specific anatomy from imaging alone, without HIE labels.

Figure 3 · Signal grows with cohort size



Sub-sampling experiment on 14 confirmed-linked patients (N = 2, 3, 4, 6, 8, 10, 12, 14; 30 random draws each). Fold improvement over chance grows monotonically from 1.45× at N=2 to 5.48× at N=14 — same architecture at N ≥ 100 will materially exceed today's floor.

AI deep dive · what · why · now · next

1 · What our AI shows

Two-label scheme. Scan-time clinical-report impression → trains LLM reporter. Final patient-level HIE diagnosis (Sarnat + MRI d4-10 [3] + neurology follow-up) → truth label for imaging classifier. Splits by patient.
Honest architecture. With 14 patients, end-to-end CNN would memorise specific babies. Frozen encoder + small classifier is the statistically defensible choice [5,6].
Result. Frozen ImageNet ResNet-18 + LOFO 5-NN: 5.48× / 7.2× chance, p<10⁻²⁰⁰. Encoder sees only redacted anatomy.

2 · Why this differs from the abstract

Abstract 2D CNN end-to-end on 15-50+ frames + LLM on comments
Deliverable Frozen encoder + linear probe + LOFO (USFM/BiomedCLIP swap-in)

Why: (1) cohort 14 vs 100+ — CNN would over-fit; (2) 3-reader adjudication deferred — feasibility + clinical accuracy; (3) foundation models matured 2023-24.
Architecture is UNCHANGED — only training data is what we are still building.

3 · What the data CAN and CANNOT tell us NOW

- | | |
|--|---|
| ✓ CAN draw NOW | ✗ CANNOT support yet |
| <ul style="list-style-type: none">• Pipeline works on routine NHS PACS/RIS• Data is AI-learnable (non-medical backbone)• Redaction works (642/642 clean)• Linkage works (14 pts · 20 events)• Signal scales with cohort size (Fig 3) | <ul style="list-style-type: none">• Clinical-grade HIE classification• Multi-centre generalisability• Severity-stratified Sarnat-grading• Longitudinal HIE evolution• MRI-cUS concordance (paired MRI absent) |

4 · Going forward — what unlocks the next results

→ **Domain-specific encoders**
USFM / BiomedCLIP / RadImageNet [9,10,11] — lifts the 5.48× floor.
→ **3-reader adjudication**
+ MRI d4-10 as soft reference. Converts feasibility → clinical accuracy.
→ **Federated learning [12]**
Multi-trust expansion — raw images stay local. Path to N ≥ 100.
→ **Vision-language head**
Radiologist-style prompts → interpretable per-frame reasoning.
→ **Clinically useful when**
N ≥ 100 patients + adjudicated labels + paired MRI.

Health economics · concrete UK numbers

Lever	Figure	Source
Lifetime HIE/CP cost	£0.8-1.0 m / case	PRCePT [13]
Hypothermia ICER	£19,931 / DFY	Regier 2010 [15]
NHS Resolution claims	£1.3 bn / yr (maternity)	AR 2024-25 [8]
28-day reporting breaches	976,000 in 2024	RCR Census [7]

NHS policy · 5 asks

1. **DICOM-only export + pixel-overlay suppression default**
Would have prevented 52% of our data loss.
2. **Federated regional imaging archives**
Paediatric studies follow the patient — solves attrition.
3. **National minimum standard for cUS acquisition**
Reduces inter-device variability; fair multi-trust pooling.
4. **AI training datasets as versioned, audited clinical asset**
TRIPOD-AI / CLAIM compliance from day one.
5. **Mandatory device-stratified performance reporting**
Catches generalisation-drift before deployment.

Conclusion · what we propose · how it improves the data · what it delivers

What we propose

- A clinician-led, fully PACS-native, layered AI pipeline for neonatal cUS in HIE — runs inside existing NHS PACS.
- Frozen foundation encoder + linear probe (USFM/BiomedCLIP swap-in) + LLM reporter, with clinician sign-off retained.
- Ethical-by-design: federated-ready, audit-trailed, TRIPOD-AI / CLAIM compliant from day one.

How it improves the data

- Standardised AI-assisted reporting reduces inter-reader variability — especially between junior readers where agreement is lowest.
- Federated multi-trust expansion → research-grade label diversity ethically — closes the N=14 → N ≥ 100 cohort-size gap.
- Audit trails + device-stratified reporting catch generalisation-drift before a model touches a baby.

What it delivers clinically

- Faster severe-HIE triage in the 6-h cooling window [2] — earlier hypothermia. ICER £19,931/DFY [15].
- Lifetime HIE/CP cost averted: £0.8-1.0 m / case [13]. PRCePT precedent: up to €385 m cumulative cost avoidance since 2018.
- 24/7 cot-side second-opinion at NHS-feasible cost. Relieves the 976 k scan reporting backlog [7].