Blood-based staging of Alzheimer's disease — and the label circularity that inflates it

Machine learning multiclass classification of cognitive stages using plasma biomarkers, clinical assessments & genetic features · a repeated, nested cross-validation study in ADNI with external evaluation in CNTN

Jiayuan Xu[1], **Andrew J. Doig**[2], **Sofia Michopoulou**[3], **Petroula Proitsi**[4], **Fumie Costen**[1,*]

[1]Department of Electrical & Electronic Engineering, University of Manchester · [2]Division of Neuroscience, School of Biological Sciences, University of Manchester · [3]Medical Physics, University Hospital Southampton NHS Foundation Trust, University of Southampton · [4]Centre for Preventive Neurology, Wolfson Institute of Population Health, Queen Mary University of London · ***Correspondence**

BOTTOM LINE Remove the circular clinical scales and blood-based CN/MCI/AD staging is **≈ 0.74** — not **0.96**. **pT217** carries the signal; **MCI** is the bottleneck.

1 Background & question

AD progresses along a continuum — **CN → MCI → AD**. Ultrasensitive assays now detect AD proteins (pT217, Aβ42/40, NfL, GFAP) in blood, promoted as scalable, low-cost staging tools. Yet head-to-head comparisons against the clinical scales that *define* diagnosis remain scarce, and reported ML "fusion" gains may reflect circularity rather than biology.

2 The circularity problem

CENTRAL CAVEAT

In ADNI, **MMSE, CDR-SB & FAQ** are the very instruments used to assign CN / MCI / AD labels (e.g. CDR-global cut-offs). A model trained on them partly **re-derives the label** — so its high accuracy is **definitional, not predictive**. Clinical-only and biomarker-only AUCs therefore measure different quantities and can't be compared directly.

3 Objectives

1. Can plasma biomarkers + demographics + genetics classify CN / MCI / AD?
2. How do they compare with validated clinical scales alone?
3. Does fusing all feature types add *genuine* discriminative gain?

SO WHAT · CLINICAL IMPLICATION

A triage tool, not a replacement

A minimally-invasive plasma panel could **triage** who proceeds to confirmatory amyloid-PET or CSF — but on the current feature set it **cannot yet replace** clinical staging. The limiting case is early **MCI**, where plasma overlaps with normal aging.

HONEST LIMITATIONS

- Cross-sectional; **ADNI** is highly-educated, non-Hispanic-White — limited generalizability
- External **CNTN** uses pT181 (not pT217); AD n = 17 → estimates realistic but preliminary

4 Data & methods

ADNI · INTERNAL

655 participants

CN 296 · MCI 168 · AD 191

CNTN · EXTERNAL

130 participants

CN 57 · MCI 56 · AD 17

- **3 feature sets:** clinical-only (3) · biomarker + demo + genetic (8) · fusion (11)
- **4 classifiers:** Logistic Regression · SVM · Random Forest · XGBoost
- **Repeated nested CV:** 5×3 outer (15 folds) + 5-fold inner; balanced weights; 95% CIs

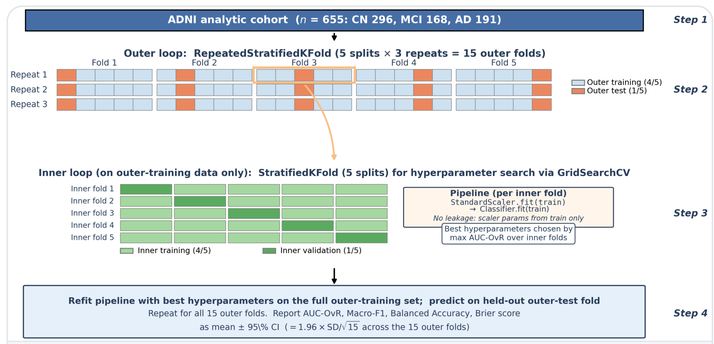

Fig 1. Repeated, nested CV: tuning is strictly separated from evaluation to avoid optimistic bias.

FINDING 1 · THE FUSION MIRAGE

Fusion adds nothing over clinical scales

Clinical-only ≈ full fusion — **statistically indistinguishable** (Wilcoxon p = 0.33). The apparent gain was never real.

Clinical-only inflated	0.954
Fusion · 11 feat. inflated	0.956
Plasma + demo + genetic de-confounded	0.746

FINDING 2 · CIRCULARITY, QUANTIFIED

Clinical scales own 78% of the signal

78% · 17%

Clinical scales **78.3%** · All plasma **17.1%** · Demo/genetic **4.6%**

Fusion-model SHAP: CDR-SB alone = 41.7%. Direct evidence the ≈0.95 is definitional.

FINDING 3 · PT217 CARRIES THE BIOLOGY

One marker dominates

pT217	0.106
NfL	0.059
GFAP	0.024
Aβ42/40 · APOE4 · age…	≤.019

Mean |SHAP|, biomarker model. Best subset **pT217 + NfL = 0.753**; Aβ42/40 adds nothing.

FINDING 4 · WHERE IT BREAKS & TRANSFERS

MCI is the bottleneck

CN·AD	MCI·AD	CN·MCI
1.00	0.97	0.93

Pairwise fusion AUC. Plasma-only **CN vs MCI falls to 0.70** — the hardest contrast; MCI recall 73%, errors toward CN.

0.702 External **CNTN** transfer (95% CI 0.635–0.764), reduced panel without pT217 — same high-CN / low-MCI pattern.

FUTURE DIRECTIONS

- External validation of a **pT217-inclusive** panel — the top priority
- Longitudinal **MCI → AD** conversion prediction
- Next-generation markers (**pT231**, synaptic proteins)

CONCLUSIONS

- → Apparent fusion gains are **largely label circularity**, not new biological signal.
- → Realistic blood-based staging ≈ **0.74 internal / 0.70 external**; **pT217** carries the plasma signal and is the top validation priority.
- → **MCI detection is the principal bottleneck** — the target for richer panels and pathology-anchored labels.

SELECTED REFERENCES & DATA

Xu J, Costen F. Diagnostics 2026;16:1755.

Jack CR et al. NIA-AA Framework. Alzheimers Dement 2018.

Ashton NJ et al. Plasma pT217. 2024.

Data: ADNI (adni.loni.usc.edu) · CNTN (nevadacntn.org).

CONTACT

Jiayuan Xu
jiayuan.xu@manchester.ac.uk
University of Manchester

Full paper · DOI · LinkedIn