AI-Driven Recruitment for Alzheimer's Disease Clinical Trials: A Pilot Analysis on the A4 Study Dataset

Luis R. Peraza[1], Richard Joules[1], Robin Wolz[1,2]
1 IXICO, London UK | 2 Imperial College London UK.

Background

Efficient recruitment is a challenge in Alzheimer's disease (AD) clinical trials, and deployment of sensitive and specific biomarkers early in the screening funnel can help screen out non-AD cases. Current approaches focus on the patient's clinical history, blood plasma biomarkers, and neuroimaging to assess participant suitability.

We previously trained an MRI-based ML model for classification of common dementia types: AD, FTD as well as normal controls (NC) and other dementias (OTH) [1]. Here we investigate the model's performance when applied as a pre-screening tool in a clinical trial. To answer this question, we used data from the A4 Study [2].

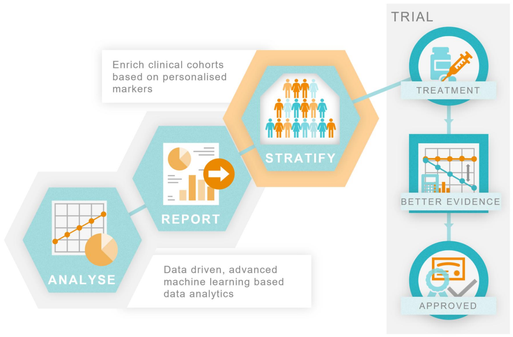

Figure 1: Neuroimaging based pre-screening in clinical trials. In this setup, structural/functional brain imaging are analysed to further inform researchers about participant suitability and enrich clinical cohorts.

Methods

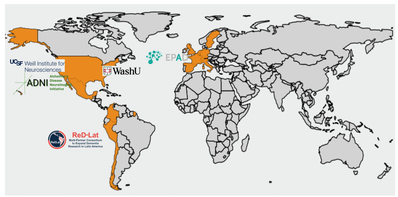

Figure 2: Geographic locations of the investigated databases.

Neuroimaging metrics, age, sex, and AI embeddings from a pretrained U-Net model were used as input variables to train a CatBoost classifier with AD, FTD, NC, and OTH target classes. For AD/NC groups within the ADNI and OASIS-3 datasets, only participants that longitudinally remained in their class were included. For EPAD we used a SUVR > 1.13 to assign AD. For the FTD group, NIFD and BrainLat, clinical diagnosis was used to assign participants to this group.

Model's performance on the evaluation set classifying each of group is shown in Figure 4.

y ~ Structural Metrics + Deep-Learning Metrics + Age + Sex

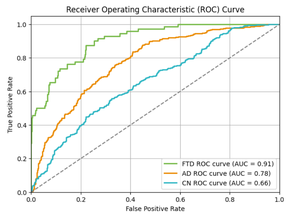

Figure 4: Model performance for 1-vs-rest classification.

The ML model was trained using T1w brain images from the ADNI, EPAD, NIFD, BrainLat, and OASIS-3 databases. Representing a diverse cohort spanning North America (USA and Mexico), pacific South America and Europe.

A4 Study's T1w images were processed with FastSurfer for normalised brain regional volumes and cortical thickness metrics [3].

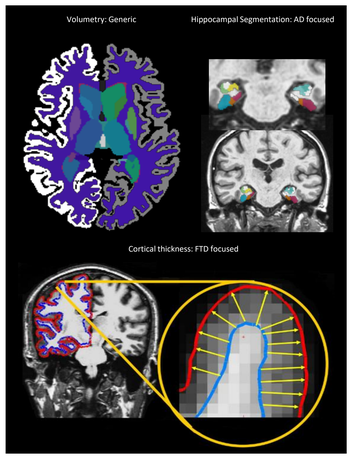

Figure 3: The model was trained with MRI-based metrics only: Volumetrics (including hippocampal subfields) and cortical thickness.

Results

The A4 study sample comprised 1304 participants, and we included all those who had tau imaging acquired. Participants were at early stages of AD with a MMSE score of 25 to 30, CDR=0, and evidence of brain amyloid pathology as seen on PET. SUVR values were referenced to the whole cerebellum with the temporal cortex as target region. Participants that had tau PET imaging acquired had significant higher levels of amyloid burden as indicated by the AB PET SUVR values (Figure 5). This is in line with current recommendations for using tau imaging as a confirmatory AD marker, Table II.

Table I: A4Study sample demographics and clinical variables

A4	MMSE	CDRSB	Age	Sex (F/M)	amyloid SUVR
N=1304	28.8 (1.28)	0.061 (0.17)	71.45 (4.67)	770/534	1.23 (0.21)

Table II: A4 sample demographics divided by tau imaging

tau imaging	MMSE	CDRSB	Age	Sex (F/M)	amyloid SUVR	tau SUVR
Yes: N=400	28.56 (1.38)	0.069 (0.19)	72.2 (4.8)	230/170	**1.31 (0.19)**	1.18 (0.11)
No: N=904	28.85 (1.22)	0.057 (0.16)	71.12 (4.6)	540/364	**1.19 (0.21)**	NA

Our ML model identified a dementia type landscape with most of the participants classified as NC, followed by FTD, AD, and other dementias (OTH). Participants classified as NC and FTD showed lower levels of amyloid burden than those classified as AD and OTH (Kruskal-Wallis, p-value < 0.005). However, the different classified groups didn't show a significant difference in their levels of tau burden, Figure 5.

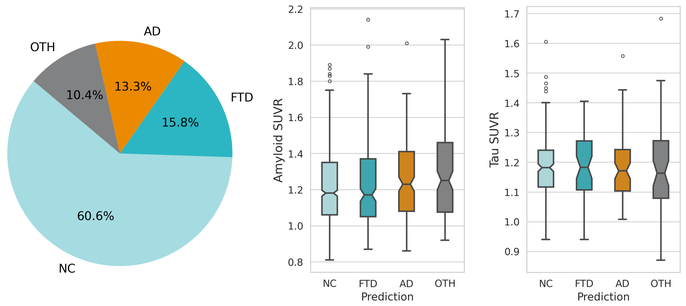

Figure 5: Disease landscape identified by our trained model, 60.6% of the sample classified as NC. Participants classified as NC and FTD showed lower levels of amyloid burden compared with those classified as AD and OTH. tau levels amongst the classified groups were not different.

Inferred probabilities of AD class in the analysed dataset showed a strong negative correlation with hippocampal volume, the inverse was true for the probabilities of NC class. amyloid burden also showed a strong correlation with hippocampal volume but the same was not the true for tau burden, see Figure 6.

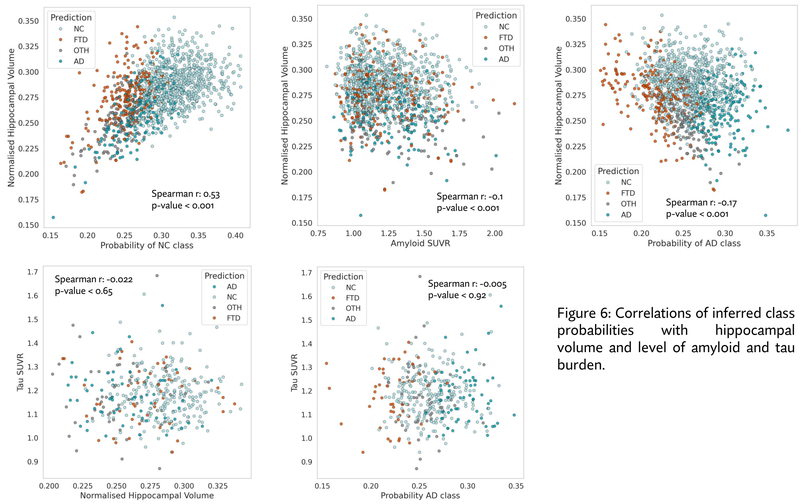

Figure 6: Correlations of inferred class probabilities with hippocampal volume and level of amyloid and tau burden.

Performance of our ML model as a pre-screening tool is shown in Table III. We selected pre-screening thresholds of 25 and 75 percentiles for hippocampal volumes and model class probabilities, and for MMSE a score <= 29. Suitable candidates were defined as those with an amyloid SUVR >= 1.1. Screen failure rate (SFR), sensitivity of suitable candidates and % savings were estimated.

From the eight pre-screening methods investigated (Table III), the AD class probability outputs showed significant lower SFR than without pre-screening, with a sensitivity of 78%, see Figure 8.

The interaction between AD probability class and MMSE when applied as a pre-screening threshold also showed a significantly lower SFR with a small decrease in sensitivity, 75%.

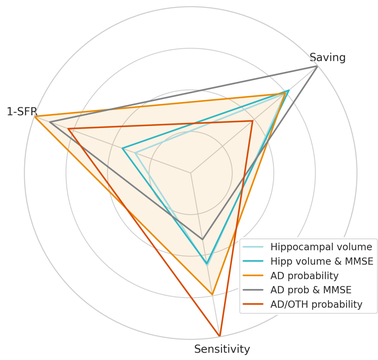

Figure 7: Spider web plot showing the performance of the pre-screening methods. AD probability class showed the most balanced performance.

Table III: Pre-screening methods' performance. * p-value < 0.05.

Pre-screened method	MRI (pre)	SUVR (suitable)	Saving	SFR	Sensitivity
None	1304	871	0%	0.332	100%
MMSE	831	565	36%	0.320	65%
Hippocampal vol	978	663	25%	0.322	76%
Hipp vol & MMSE	974	662	25%	0.320	76%
AD probability	**978**	**677**	**25%**	***0.308**	**78%**
AD/OTH probability	1014	697	22%	0.313	80%
FTD probability	978	664	25%	0.321	76%
AD prob & MMSE	**942**	**650**	**27%**	***0.310**	**75%**

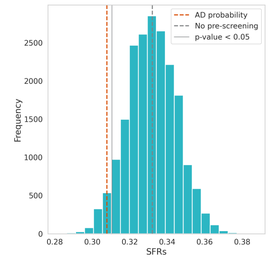

Figure 8: Significance was assessed with a bootstrap test of SFR estimates with replacement.

Conclusions

In this investigation we proved that an MRI-based ML model for disease classification can be successfully applied as a pre-screening tool in clinical trials. Although in this investigation we used a recruitment criteria of SUVR >= 1.1, future research will be done testing a wider range of selection criteria and values.

Our model consistently predicted higher levels of amyloid SUVR in participants classified as AD and OTH compared to those classified as NC and FTD. In our previous work [1], the OTH group comprised misdiagnoses from the training databases, many of these were PDD (Parkinson's disease dementia) and DLB (dementia with Lewy bodies) participants. This may explain the higher amyloid SUVR predictions.

Our model outputs did not correlate with tau PET burden for the patients that had tau imaging acquired, probably because the A4 Study cohort is at the very early stages of the disease. However, previous research has shown that tau correlates better with disease symptoms and previous investigations in patients at later disease stages have found significant correlations between tau burden and brain volumetrics [4].

Overall, using the AD prediction probabilities from our MRI-based ML model showed to improve participant selection in a clinical trial with a beneficial trade-off between sensitivity, SFR and savings, Figure 7.

References

[1] Wolz et al. 2025. AI-Driven Classification of Alzheimer's Disease and Frontotemporal Dementia from Magnetic Resonance Imaging. ADPD 2025.
[2] www.a4studydata.org
[3] Henschel L, et al. 2020. A fast and accurate deep learning-based neuroimaging pipeline, NeuroImage.
[4] Matthews DC. et al. 2024. Relationships between plasma biomarkers, tau PET, FDG PET, and volumetric MRI in mild to moderate Alzheimer's disease patients. Alzheimer's & Dementia TR&CI.