

Multi-Task Learning for Alzheimer's Disease Progression

Biomarker Assessment and Temporal Modelling

Xulong Wang

Supervisors: Prof Po Yang • Prof Vitaveska Lanfranchi
School of Computer Science • University of Sheffield

PhD Viva • Viva date: 23 June 2026

Outline



Background



Motivation



Aim & Objective



Research Questions



Knowledge Contribution

The scale of the problem

60–70%

of all dementia

55M → 139M

cases by 2050

every 3 s

one new case

\$1.3T / yr

global cost → \$2.8T by 2030

The Delayed Nature of Current Clinical Diagnosis

Brain damage / neuronal loss

Clinical symptom threshold

Diagnosed here

*damage accumulating —
clinically silent*

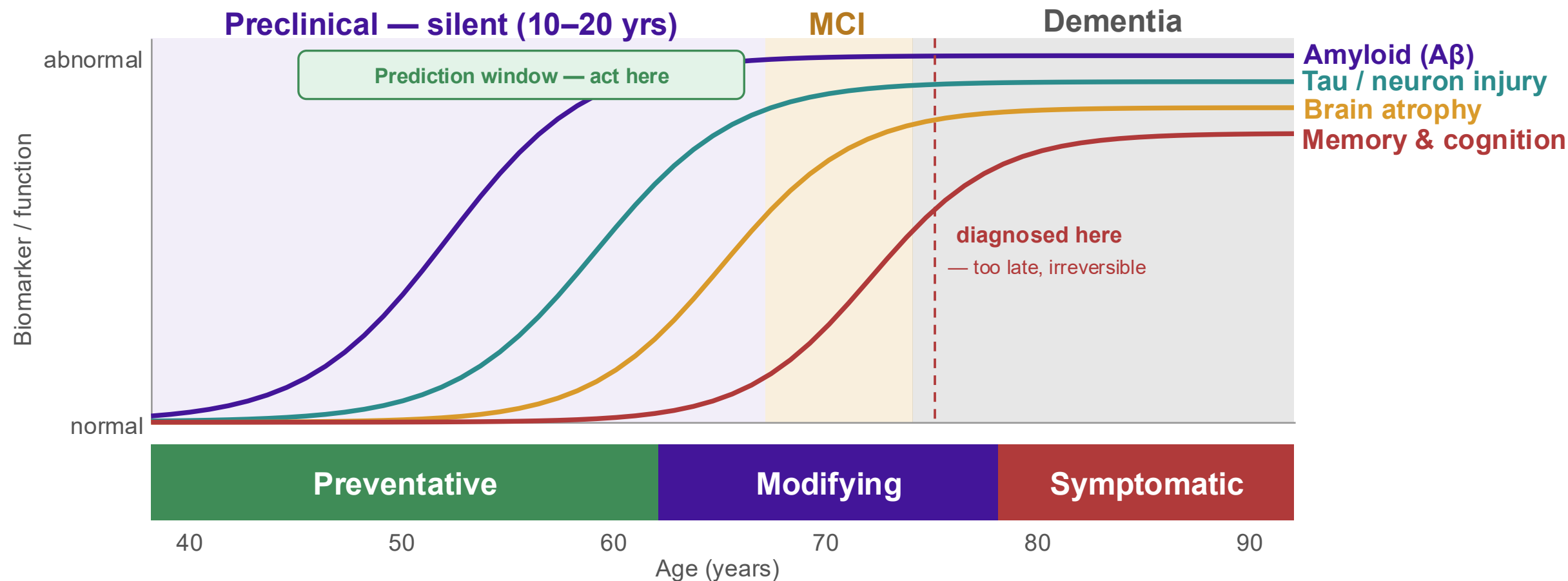
Time / age →

WHY IT MATTERS — WE CATCH IT TOO LATE

- Clinical diagnosis is established only after extensive and silent brain damage.
- We diagnose when the disease is already irreversible.

When does the disease start?

Biomarkers turn abnormal 10–20 years before symptoms; cognition is the last domain to decline.



Treatment (Lecanemab and Donanemab) benefit is restricted to early disease stages. Predictive modelling within this early window defines the core clinical value.

Sources: WHO; Alzheimer’s Disease International, 2024; cascade after Jack et al. (2010).

Why structural MRI and its challenges?

	Amyloid (A β)	Tau / injury	MRI
Non-invasive & repeatable	X invasive (PET/CSF)	X invasive (PET/CSF)	✓ non-invasive, repeatable
Routine & low-cost	X costly, specialist	X costly, specialist	✓ routine, low-cost

Structural MRI offers a **non-invasive and routine clinical tool**, making it the most viable biomarker for scalable early prognosis and long-term tracking.

Yet three modelling challenges remain



High dimensionality & noise

thousands of noisy voxels per scan



Inter-subject variability

anatomy differs across individuals



Limited sample size

small longitudinal AD cohorts

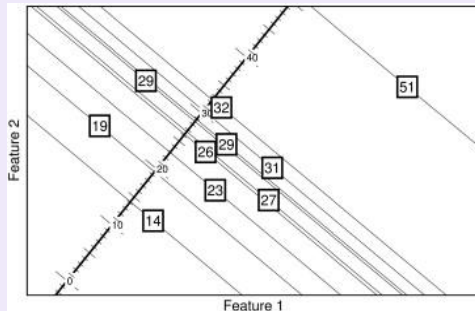
Previous work: from isolated predictions to joint temporal modelling

Single-task approaches

Predict one score at a single time point in isolation.

Limitation

- Lacks temporal constraints
- Over-fits high-dim MRI



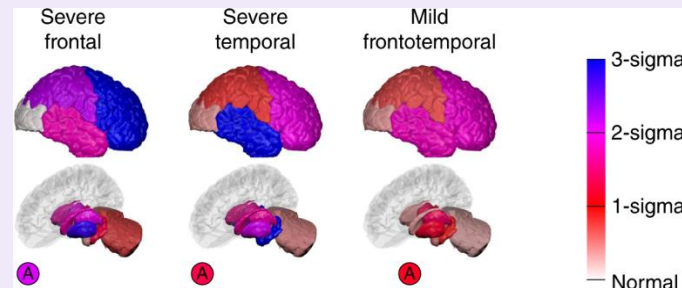
Stonnington et al., NeuroImage 2010

Event-Based Models

Recover the data-driven order of biomarker abnormality.

Limitation

- Discrete staging only
- No continuous forecast
- Group-level, not individual



Young Brain 2014 · Nature Communications 2018

OUR FOCUS

Multi-task learning (MTL)

Jointly learn predictions across time points.

Advantage

- Robust to small samples
- Predicts the clinical outcome
- Cross-task feature sharing
- Interpretable biomarker selection
- Flexible temporal structure

Zhou 2011/2013 (MTL) · Wang 2019 (MTL-MRI)

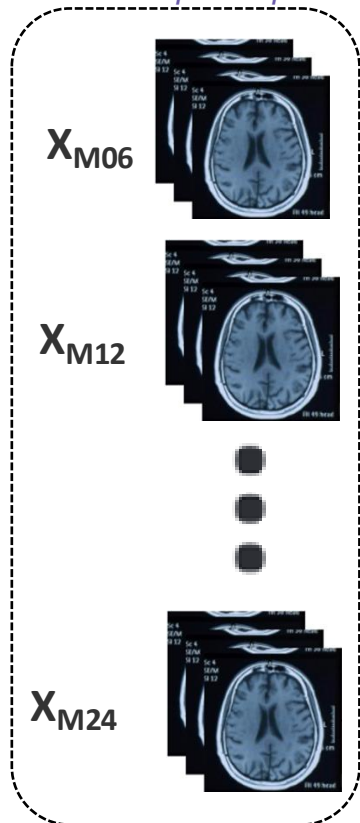


MTL emerges as the one of viable approaches for joint temporal modelling with limited data

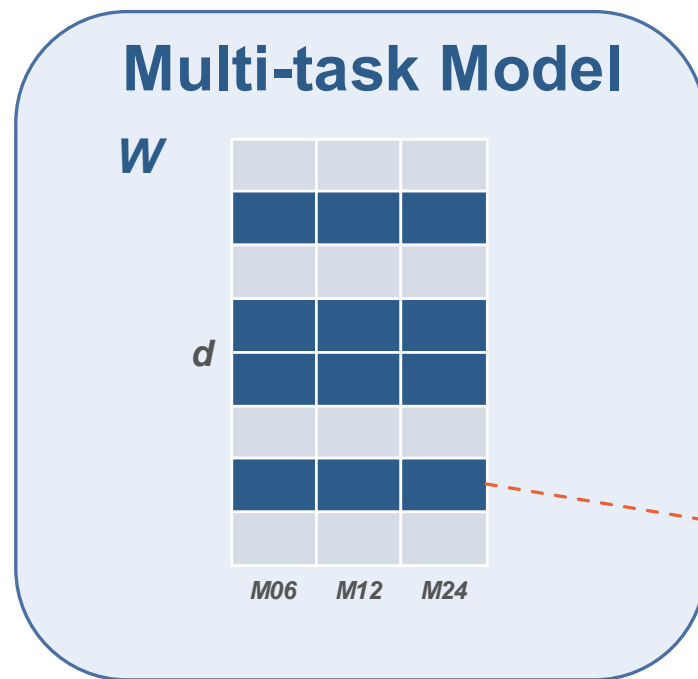
Aim & Objectives

Use regularised multi-task learning to **predict** individual AD trajectories and **interpret** biomarker importance from longitudinal neuroimaging.

Tasks = follow-up time points (M06 / M12 / M24).



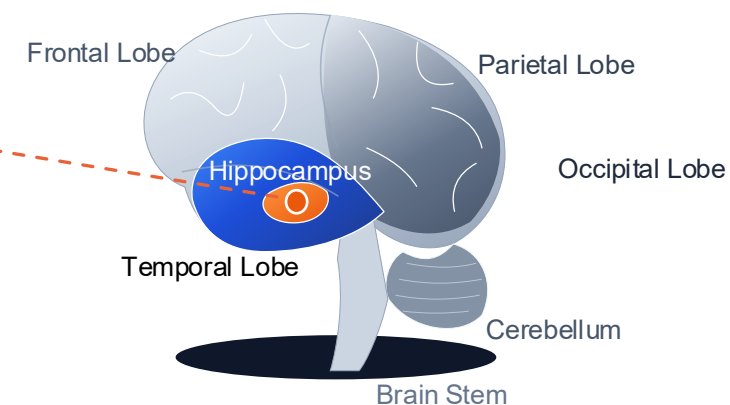
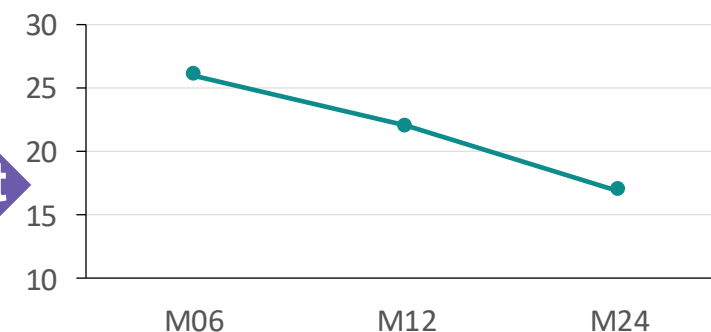
Train



Predict

Output

Predicted MMSE



O1 Review & research gaps

O2 Regularised MTL framework

O3 Robust feature learning

O4 Stable biomarker

O5 Non-linear temporal dynamics

O6 Real-world cohort evaluation

Research Questions & Framework

RQ1

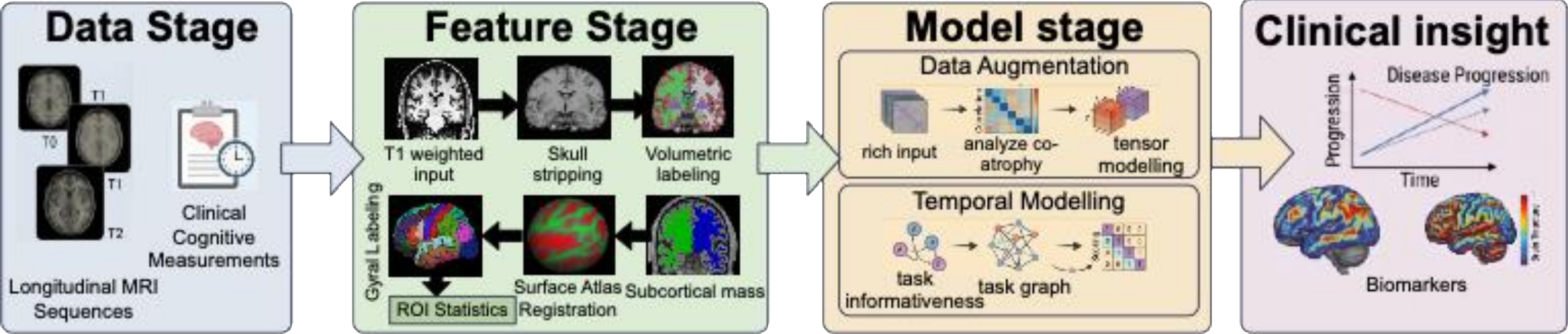
Formulate & evaluate a regularised MTL framework for AD progression.

RQ2

Assess biomarker importance via inter-regional interaction for interpretable insight.

RQ3

Model non-linear temporal dynamics across prediction tasks.



Stabilising AD Biomarker Selection

Correlated MRI features destabilise ℓ_1 selection $\rightarrow$ RMFL yields a stable, reproducible biomarker set.

CHALLENGE · mechanistic motivation

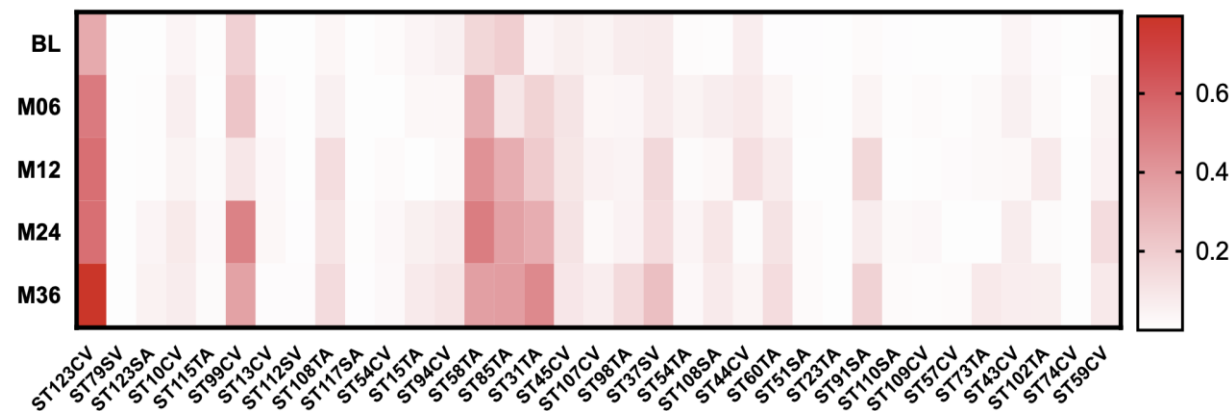
- 327 MRI features, $n = 429 \rightarrow$ high-dim, strongly correlated (low-rank)
- ℓ_1 picks one feature per correlated group $\rightarrow$ random across splits

APPROACH · RMFL

- MTEN: $\ell_{2,1}$ + Frobenius $\rightarrow$ grouping effect (correlated regions together)
- Two-stage stability selection: sub-sample $\lfloor n/2 \rfloor$; retain features with $\pi \geq 0.96$

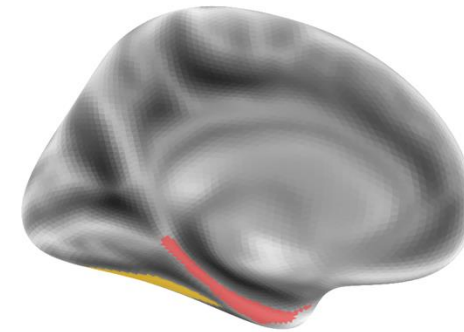
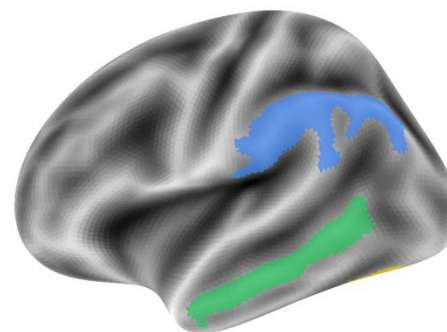
RESULTS

- **Reproducible set: 327 $\rightarrow$ 35 core stable features**
- MMSE nMSE 0.745 vs Lasso 0.945 vs Ridge 2.088



Left hemisphere — lateral view

Left hemisphere — medial view



■ Inferior Parietal ■ Hippocampus ■ Middle Temporal ■ Fusiform

Reproducible 35-feature core; four regions consistent with Braak / AD atrophy direction.

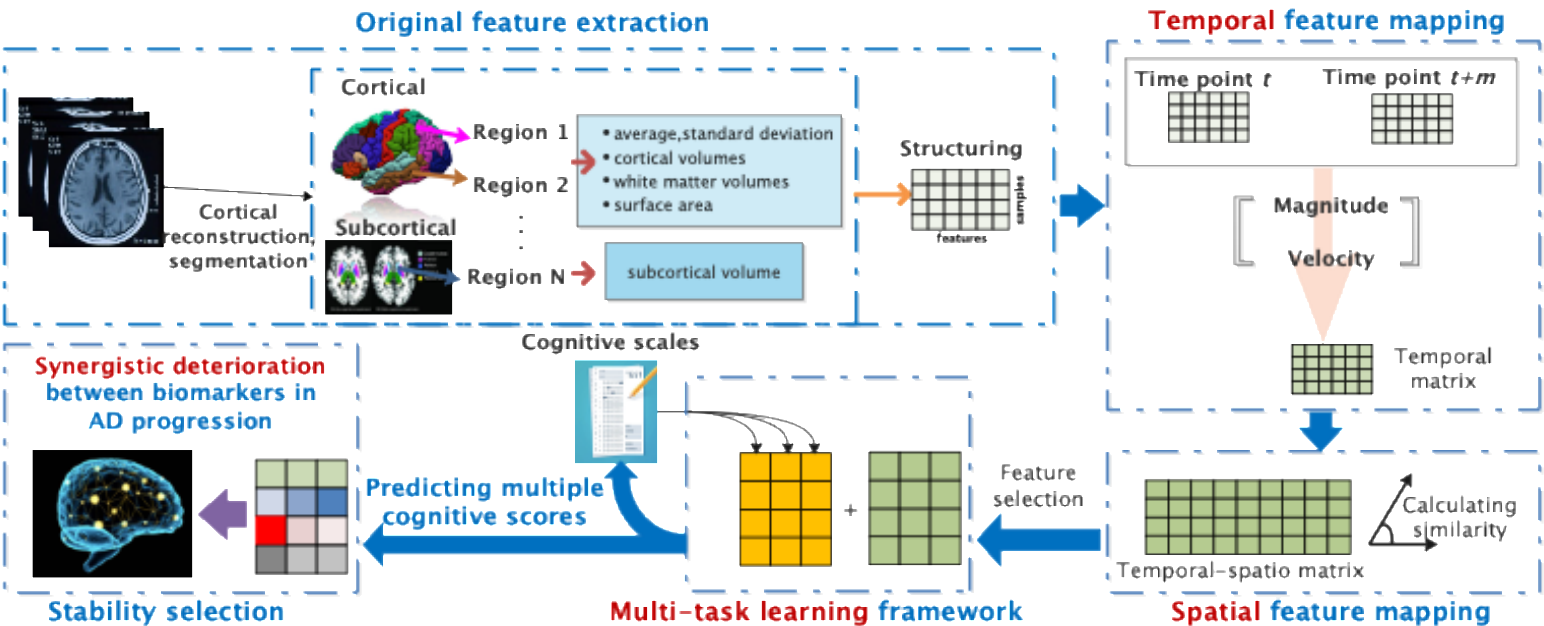
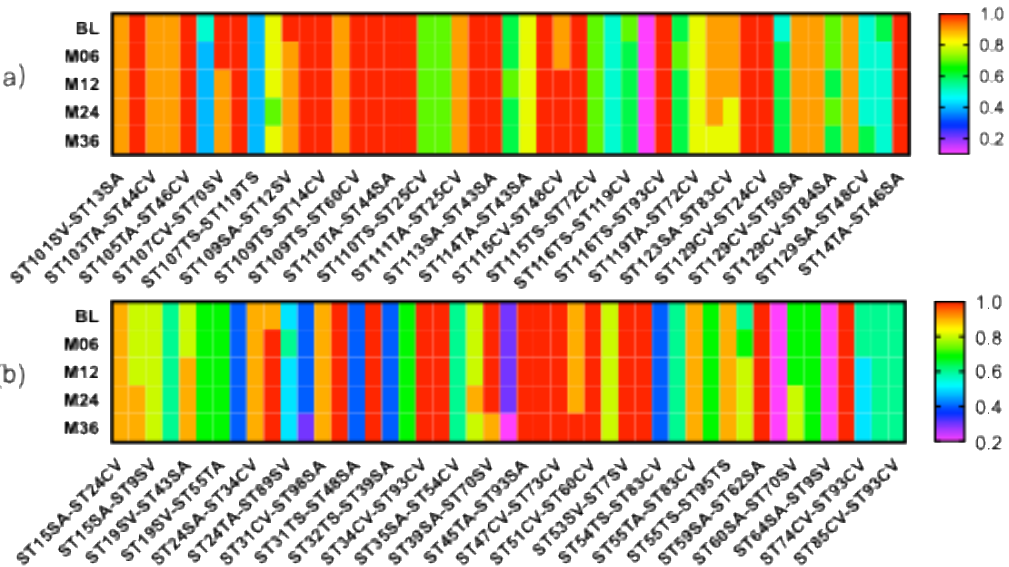
Co-atrophy signals between brain regions

Challenge

- Single-region scores miss co-atrophy synergy

Approach

- A cosine similarity to measure the region pair trend
- Weight = predictive region-pair, not one region



Result

- 94 / 87 stable pairs (MMSE / ADAS)
- Cluster into 2 known AD networks
 - Medial temporal memory (entorhinal, parahippocampal)
 - Posterior default-mode (post. cingulate, retrosplenial)

Ref: JAMA Neurology • Acta Neuropathologica • Annals of Neurology

Pairwise interactions via tensor modelling

Challenge

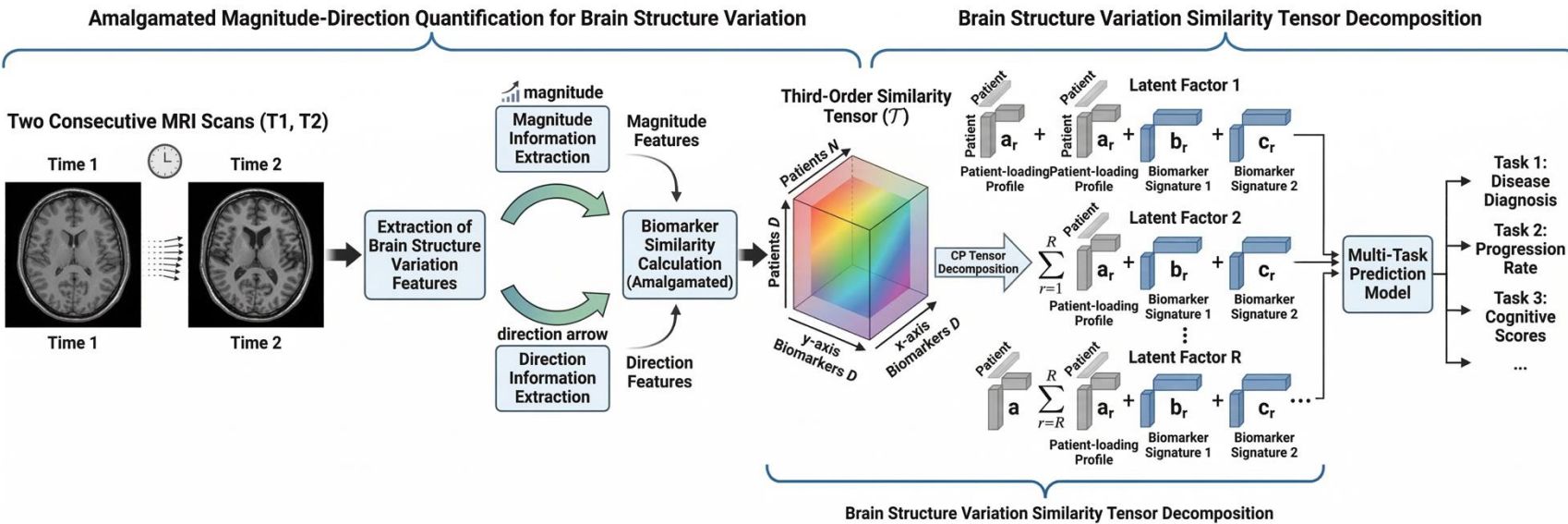
- capture magnitude + direction of change

Approach — AMDQ-TMTL

- Mahalanobis + 5 direction types $\rightarrow [-1, +1]$
- $D \times D \times N$ tensor $\rightarrow$ CP decomp. $\rightarrow$ latent factors

Method	nMSE ↓	wR ↑	M12 ↓	M24 ↓	M36 ↓	M48 ↓	M60 ↓
Target: MMSE							
Ridge	2.487 ± 4.853	0.252 ± 0.096	4.489 ± 1.969	5.676 ± 1.274	6.098 ± 2.078	7.520 ± 1.721	7.299 ± 1.894
Lasso	0.895 ± 0.091	0.431 ± 0.081	1.971 ± 0.219	2.311 ± 0.286	2.858 ± 0.639	3.547 ± 0.633	4.503 ± 1.027
TGL	0.402 ± 0.052	0.818 ± 0.030	1.383 ± 0.212	1.426 ± 0.161	1.778 ± 0.347	2.165 ± 0.556	2.763 ± 0.330
cFSGL	0.327 ± 0.024	0.835 ± 0.025	1.353 ± 0.128	1.462 ± 0.214	1.661 ± 0.269	2.010 ± 0.372	2.640 ± 0.685
nCFGL1	0.356 ± 0.042	0.841 ± 0.014	1.336 ± 0.242	1.422 ± 0.146	1.746 ± 0.255	2.126 ± 0.367	2.713 ± 0.532
FL-SGL	0.499 ± 0.051	0.748 ± 0.031	1.691 ± 0.182	1.738 ± 0.137	2.141 ± 0.521	2.662 ± 0.459	3.078 ± 0.544
NC-CMTL	0.497 ± 0.060	0.767 ± 0.043	1.573 ± 0.165	1.599 ± 0.172	1.881 ± 0.120	2.620 ± 0.421	4.145 ± 1.221
AMDQ-TMTL	0.277 ± 0.010	0.874 ± 0.006	1.315 ± 0.031	1.364 ± 0.142	1.332 ± 0.181	1.706 ± 0.314	2.305 ± 0.374
Target: ADAS-Cog							
Ridge	2.310 ± 0.317	0.251 ± 0.104	11.111 ± 1.529	11.144 ± 1.466	13.585 ± 2.073	14.958 ± 2.728	18.921 ± 2.297
Lasso	0.886 ± 0.125	0.407 ± 0.095	6.105 ± 1.257	7.013 ± 1.258	8.183 ± 1.571	11.048 ± 1.836	12.854 ± 2.568
TGL	0.318 ± 0.067	0.857 ± 0.031	3.892 ± 0.751	3.517 ± 0.438	3.950 ± 0.614	4.352 ± 1.039	8.580 ± 1.346
cFSGL	0.321 ± 0.072	0.858 ± 0.028	3.979 ± 0.861	3.534 ± 0.435	4.021 ± 0.263	4.639 ± 0.982	8.003 ± 1.274
nCFGL1	0.285 ± 0.072	0.875 ± 0.040	4.161 ± 0.639	3.625 ± 0.772	3.649 ± 0.649	5.821 ± 2.486	6.688 ± 2.082
FL-SGL	0.482 ± 0.093	0.784 ± 0.008	4.761 ± 1.091	4.363 ± 0.496	4.377 ± 0.352	6.058 ± 1.009	9.791 ± 1.666
NC-CMTL	0.239 ± 0.032	0.883 ± 0.022	3.998 ± 0.689	3.643 ± 0.780	3.801 ± 0.505	4.748 ± 1.987	7.211 ± 1.610
AMDQ-TMTL	0.159 ± 0.047	0.910 ± 0.034	1.892 ± 0.456	1.975 ± 0.150	1.943 ± 0.212	2.494 ± 0.743	3.191 ± 0.428

Integrated Pipeline for Brain Structure Variation Analysis via Amalgamated Quantification and Tensor Decomposition



Result

- Beats 8 baselines: MMSE -15.4%, ADAS -33.7% nMSE
- Variance 5.2× smaller $\rightarrow$ stable



Task graphs learned from data

Challenge

- Fixed temporal smoothness (TGL) assumes adjacent tasks alike
- Too rigid for heterogeneous AD progression

Approach — AutoTG (in MAGPP)

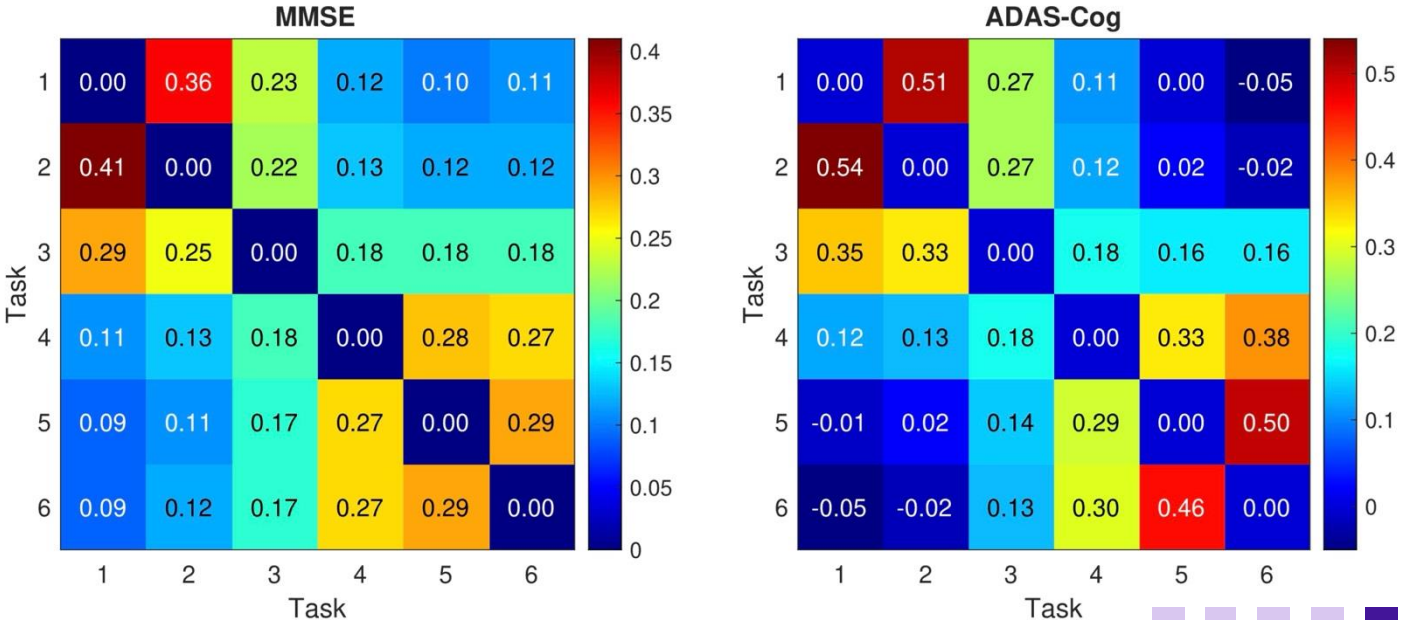
$$\min_{W,R} \frac{1}{2} \sum_{i=1}^m \|X_i w_i - y_i\|_2^2 + \lambda_1 \|W - WR\|_F^2 + \lambda_2 \|R \odot S\|_{1,1} + \lambda_3 \|W^T\|_{2,1} + \lambda_4 \|W^T\|_{1,1}$$

- Learn task relation graph R (W and R both learnable)
- Allows asymmetric, non-adjacent, repulsive relation

Method	nMSE ↓			wR ↑		
	$\beta = 0.4$	$\beta = 0.6$	$\beta = 0.8$	$\beta = 0.4$	$\beta = 0.6$	$\beta = 0.8$
TGL	0.620	0.621	0.602	0.616	0.618	0.631
cFSG	0.631	0.597	0.583	0.610	0.632	0.650
FLSGL	0.651	0.639	0.626	0.594	0.607	0.619
VSTG	0.649	0.659	0.641	0.588	0.593	0.608
NCCMTL	0.626	0.632	0.627	0.611	0.609	0.600
LSA	0.630	0.601	0.579	0.610	0.619	0.650
GAMTL	0.635	0.627	0.591	0.608	0.611	0.637
LSTM	0.850	0.860	0.756	0.438	0.444	0.562
SGLasso	0.671	0.666	0.653	0.590	0.599	0.612
AutoTG	0.642	0.638	0.617	0.601	0.607	0.629
MAGPP	0.624	0.585	0.567	0.619	0.636	0.663

Result

- Matches/edges best fixed baseline (ADAS nMSE 0.453 vs cFSG 0.459)
- Learned R: asymmetric + repulsive
- Long-range task dependencies confirm non-linear Alzheimer's progression



Quantifying pairwise task relevance

Challenge

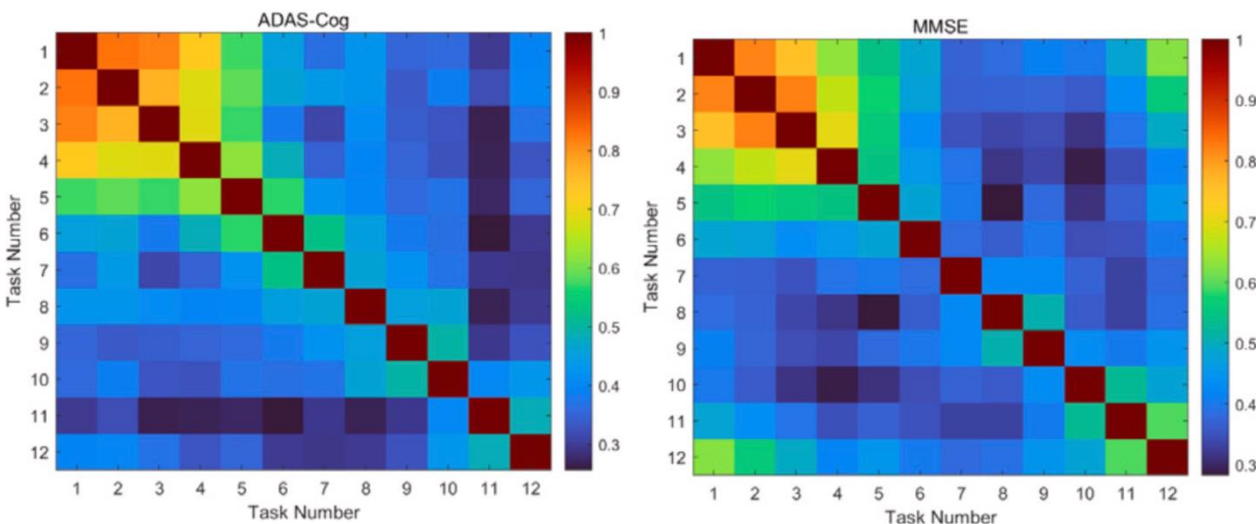
- MTL prescribes how to share, not whether it helps
- Informativeness varies by disease stage

Approach — MTIRL

$$\min_{H,P} \mathcal{L}(W) + \lambda_1 \text{tr}(H\Omega^{-1}H^T) + \lambda_2 \|P\|_{1,2}, \quad W = H + P,$$

- Learned task covariance $\Omega \rightarrow$ quantifies **sharing** informativeness
- Exclusive Lasso on $P \rightarrow$ **task-specific** feature competition

Method	$\beta = 0.6$	$\beta = 0.7$	$\beta = 0.8$	$\beta = 0.9$
eLasso	11.414 ± 0.550	8.814 ± 0.205	8.567 ± 0.220	8.807 ± 0.537
RMTL	8.230 ± 0.342	8.084 ± 0.212	7.643 ± 0.239	7.112 ± 0.265
RMTFL	9.713 ± 0.408	8.936 ± 0.205	8.785 ± 0.301	8.428 ± 0.404
SPATS	9.028 ± 0.191	8.330 ± 0.232	7.863 ± 0.368	7.584 ± 0.119
MTERL	9.128 ± 0.200	8.603 ± 0.230	8.301 ± 0.453	8.061 ± 0.582
RoTS	6.373 ± 0.293	6.594 ± 0.175	6.376 ± 0.206	6.366 ± 0.250
MTIRL	6.413 ± 0.261	6.182 ± 0.122	6.440 ± 0.144	6.212 ± 0.265
MTSIRL	6.275 ± 0.341	6.580 ± 0.152	6.343 ± 0.214	6.088 ± 0.295



Result

- ADAS rMSE 11.410 vs SPATS 12.112 (−5.8%)
- Learned $\Omega \rightarrow$ three-phase temporal pattern
- Mid-stage (M60–M96) most complex, least uniform

Conclusions: stable biomarkers, flexible modelling

01

Stable feature learning

RMFL + stability selection

Reproducible
biomarker sets (Ch 4)

02

Interaction assessment

Pairwise similarity + AMDQ-TMTL

Richer space capturing
co-atrophy (Ch 5)

03

Temporal modelling

AutoTG + MTIRL

Learned task
structure replaces
rigid priors (Ch 6)

LIMITATIONS & FUTURE WORK

- Cohort limited to ADNI-1 → external validation on ADNI-2 / 3 needed
- Future: graph neural networks for dynamic task-graph learning

**Thank you —
questions welcome**