Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

AI enabled Tau–Neurodegeneration mismatch in Alzheimer's disease

Xueying Lyu
University of Pennsylvania
Mengjin Dong
University of Pennsylvania
Pulkit Khandelwal
Massachusetts General Hospital
Christopher Brown
University of Pennsylvania
Yue Li
University of Pennsylvania
Michael Tran Duong
University of Pennsylvania
Long Xie
University of Pennsylvania
Amanda Denning
University of Pennsylvania
Ilya M Nasrallah
University of Pennsylvania
Paul A. Yushkevich
University of Pennsylvania
David A. Wolk
University of Pennsylvania
Sandhitsu R. Das
sudas@seas.upenn.edu

University of Pennsylvania

Article

Keywords:

Posted Date: November 18th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11613998/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

AI enabled Tau–Neurodegeneration mismatch in Alzheimer's disease

Xueying Lyu[1], Mengjin Dong[1], Pulkit Khandelwal[5], Christopher Brown[2], Yue Li[3], Michael Tran Duong[3], Long Xie[3], Amanda Denning[4], Ilya M Nasrallah[3], Paul A. Yushkevich[3], David A. Wolk[2†], Sandhitsu R. Das[2†] for the Alzheimer's Disease Neuroimaging Initiative*

[†]co-corresponding authors

Author affiliations:

Departments of Bioengineering[1], Neurology[2], Radiology[3], Neuroscience Graduate Group Medicine[4], 3451 Walnut Street, University of Pennsylvania, Philadelphia, PA, 19104, USA

[5]Laboratories for Computational Neuroimaging, Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston MA, 02129, USA

*Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. A complete listing of ADNI investigators can be found at: http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Correspondence to: Sandhitsu R. Das and David A. Wolk

Full address: Penn Memory Center, Ralston House #212A, 3615 chestnut street, Philadelphia, PA19104

E-mail: sudas@seas.upenn.edu ; David.Wolk@pennmedicine.upenn.edu

Abstract

The heterogeneity of Alzheimer's disease (AD) and lack of validated markers for non-AD co-pathologies limits targeted therapies. Our prior work showed that deviations ("mismatch") from a normative Tau (T)–Neurodegeneration (N) relationship can identify co-pathologies and resilience in AD. However, T–N relationships are nonlinear and spatially complex. In this study, we implemented a T–N mismatch framework using inter-modality image translation to synthesize "expected" cortical thickness maps from tau PET, and comparing these to observed thickness as a mismatch metric. T–N mismatch was quantified as region-wise “reconstruction error” (AI-predicted versus observed thickness) and subsequently used for data-driven clustering. The AI-predicted N aligned with the observed N while deviations still commonly existed. The AI-enabled T-N mismatch identified distinct data-driven T-N phenotypes associated with specific co-pathological signatures. Our findings support that AI-based T-N mismatch for identifying non-AD modulators in AD may have advantages over a statistical linear-regression approach.

Introduction

The pathophysiology of Alzheimer’s disease (AD) is characterized by accumulation of β-amyloid (Aβ) plaques and tau neurofibrillary tangles (T)[1,2]. While both Amyloid (A) and T are the key components of the biological definition of AD based on the ATN framework[1,2], T is thought to be more tightly linked to downstream neurodegeneration (N) and cognitive impairment and, thus, stage of disease[3–5]. With the availability of multi-modal imaging biomarkers, a number of studies[3,6,7] have demonstrated relatively strong T-N associations *in vivo* for AD patients. Despite these overall associations, there is moderate variance in these T-N relationships across individuals. While some of this variability may reflect measurement

noise or physiologic variation, a significant component is also linked to various sources of AD heterogeneity that alter the T-N relationship, such as the presence of co-pathologies[8–11] (e.g. TAR DNA-binding protein 43 (TDP-43)[12,13] deposits, vascular disease[8]), which are commonly present in AD patients. These non-AD factors influence N outside of its direct relationship with tau pathology.

Several imaging and fluid biomarkers have been studied as potential surrogates for specific co-pathologies in AD, such as white matter hyperintensity (WMH)[14,15] volume for vascular disease, entorhinal (ERC)/parahippocampal (PHC) thickness ratio[16], lower hippocampal volume[17–19], or a specific metabolic signature on ^{18}F-FDG-PET[20] for limbic-predominant age-related TDP-43 encephalopathy (LATE)[19], and fluid biomarkers[21,22] of TDP-43/LATE and α-synuclein[23]. However, given the existence of multiple co-pathologies in AD with suboptimal specific biomarkers for many, there is increasing need for approaches that can identify broader non-AD modulators of neurodegeneration.

In our prior work[24–26], we exploited the non-specific nature of neurodegeneration by investigating the magnitude and spatial pattern of deviation from the level of neurodegeneration (N) expected for a given level of tau (T). We modeled T-N relationships using linear regression in individual brain regions with measures of cortical thickness as a measure of N burden and ^{18}F-Flortaucipir SUVR as the measure of T burden. Using a data-driven clustering approach based on regional T-N residuals, we found clusters, hereafter also referred to as phenotypes, that had greater N than expected for T ("vulnerable") and or less N than expected for T ("resilient") in specific regional distribution patterns, likely reflecting the influence from distinct non-AD modulators. The vulnerable phenotypes showed evidence of enrichment for specific co-pathologies, such as TDP-43 and cerebrovascular disease, and exhibited more rapid cognitive decline, whereas the resilient ones demonstrated slower cognitive progression. In general, our previous work supports the hypothesis that a measure of

deviation from a normative T-N relationship across brain regions can reveal vulnerability and resilience within the AD continuum and that regional patterns of deviation may provide prediction of likely co-pathologies driving this vulnerability.

However, even in the case of "pure AD" pathology without other factors, where tau accumulation is thought to drive atrophy, the relationship between T and N may be complex and potentially non-linear. Moreover, tau pathology propagation and neurodegeneration unfold along brain networks[27,28], whereby tau pathology in one region may influence degeneration in anatomically connected or functionally coupled areas [7,29–31]. Therefore, region-by-region T-N regression modeling may underestimate nonlocal dependencies and complex pathology-structure coupling. Given these potential limitations of the linear T-N modeling, a more sophisticated approach that can capture non-linear and high-dimensional T-N relationships may provide more accurate estimates of N not explained by T and thereby enhance the prediction capability of the T-N mismatch approach.

Deep learning–based image translation provides a potential framework to capture these complex relationships[32,33]. In clinical imaging, deep-learning translation models have mostly been used to simulate missing scans to reduce cost, radiation exposure, or patient burden. As such, the primary goal of image translation has been to optimize visual realism and maximize image reconstruction accuracy[34].

Here, we extend the concept of T–N mismatch by utilizing inter-modality image translation from Tau PET to MRI as an artificial intelligence (AI)–based approach for uncovering heterogeneity within AD, rather than merely an image synthesis task. Specifically, instead of judging model performance solely based on prediction accuracy of image synthesis, we interpret image reconstruction errors as a meaningful signal of deviation from the typical T–N relationship. We posit that the reconstruction error (predicted versus actual N) may serve as a

proxy for T–N mismatch, likely reflecting patterns of N linked to co-pathologies or resilience that a deep learning model is unable to infer from tau PET alone. Methodologically, the use of error signal in this way shares similarities with an anomaly detection framework[33,35], but instead of detecting abnormalities relative to healthy individuals, our approach aims to identify deviations from typical T-N relationships within the AD continuum.

Specifically, we trained an AI model to perform direct voxel-wise image translation from tau PET SUVR maps to cortical thickness maps. We hypothesize that the AI model will learn nonlinear, spatially distributed relationships that extend beyond conventional region-constrained, linear regression–based T–N modeling. Importantly, the AI model was designed to learn to generate the component of N that is primarily explained by tau pathology, while minimizing the influence of non-tau information. Besides using only T for generating N, we also explored the model to include both amyloid PET (A) and tau PET (T) as inputs to study whether beta-amyloid pathology patterns contribute additional explanatory value for N. In both settings, regional reconstruction error was used as the AI-based measure of T–N mismatch. Overall, this framework reframes image translation as a tool for identifying biologically informative heterogeneity within the AD continuum, with potential applications in phenotype discovery and patient stratification.

Methods

Study Overview

The study includes the following experiments, with a flowchart summarizing them along with corresponding hypotheses in Figure S1. (1) We first trained AI models to perform voxel-wise image translation from tau PET SUVR maps alone ($T \rightarrow \hat{N}_T$) or from amyloid and tau PET SUVR maps together ($T+A \rightarrow \hat{N}_{T+A}$) to cortical thickness maps derived from structural MRI. For reducing inter-individual variability in cortical geometry, we mapped the thickness and tau

SUVR maps to a population template space. This spatial normalization limits the reliance of the model on participant-specific anatomical features, thereby encouraging the learning of intensity-based associations between tau pathology and cortical thickness. (2) We then evaluated model performance in test sets by assessing regional image reconstruction accuracy ($\hat{N}$ vs N) and evaluating AI-predicted T-N associations. (3) Regional reconstruction error (AI-predicted versus observed cortical thickness, i.e., $\hat{N}$-N) in different test sets was utilized as a measure of T–N mismatch for data-driven clustering to identify T-N phenotypes. (4) Finally, we compared AI-based and linear regression-based T–N mismatch approaches in terms of their consistency in classifications, discordance, and sensitivity to non-AD features.

Participants

ADNI

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD).

We included 184 cognitively impaired (CI) ADNI participants diagnosed with MCI or dementia, all of whom were amyloid-positive (A+). All participants had both a tau PET and a T1-weighted MRI scan. For each individual, the MRI acquired closest in time to the tau PET was selected for analysis, with an average inter-scan interval of 14 ± 10 months (mean ± SD). Most of the included participants overlapped with those in our prior work. Clinical and demographic characteristics of the ADNI cohort are summarized in Table S1.

Avid

We included 70 A+ symptomatic participants (36 females, 34 males) from the Avid Radiopharmaceuticals A05 study as an external testing cohort, each with a paired Tau PET and T1-weighted MRI scan acquired during the same visit. Inclusion criteria were age ≥ 50 years and MMSE over ten.[36] All participants provided written informed consent, and both the study protocol and consent procedures were approved by the relevant Institutional Review Boards. The cohort comprised 29 individuals with dementia and 41 with MCI. The mean age was 73.4 ± 9.9 years, mean ADAS11 score was 15.8 ± 8.3 and mean MMSE was 24.7 ± 4.3.

Image acquisition and processing

Image acquisition

Both ADNI and Avid datasets followed similar image acquisition procedures, with similar ^{18}F-flortaucipir tracer dosing, uptake times, and scan duration. Tau PET images were standardized to an 8 mm full width at half maximum (FWHM). For ADNI, T1-weighted MRI scans were acquired with a spatial resolution of either 1.0 × 1.0 × 1.0 mm³ (73%) or 1.0 × 1.0 × 1.2 mm³ (27%). T2 FLAIR MRI was collected during the same session as the T1-weighted MRI. For the Avid dataset, MRI scans were obtained at a 1.0 × 1.0 × 1.2 mm³ resolution.

T1-MRI processing

T1-weighted MRI and ^{18}F-flortaucipir tau PET scans were processed to extract cortical thickness and tau SUVR measures, respectively, across 100 bilateral gray matter regions of interest (ROIs).

The image-processing procedures followed our previously published methods. In brief, T1-weighted MRI scans were processed using the ANTs cortical thickness pipeline[37]. MRI images were parcellated into cerebellar, cortical, and subcortical ROIs using a multi-atlas segmentation approach[38]. Cortical thickness maps were then computed with the DiReCT algorithm[38], which estimates voxel-wise thickness across the cortical mantle. The gray matter thickness map for

each participant, which is a scalar map defined at each MRI voxel classified as gray matter, was then mapped to the population template space, which is then used as the input N for deep learning-based T-N image translation. For T-N mismatch modeling, ROI-level cortical thickness measurements were determined as the mean of template-space thickness within gray-matter voxels of each region. The same image processing pipeline was applied to both ADNI and Avid cohorts. A composite medial temporal lobe (MTL)[39] thickness measure was computed as the average thickness across the bilateral parahippocampal gyrus, entorhinal cortex, amygdala, inferior temporal gyrus, and fusiform gyrus.

T1-weighted MRI scans were also processed with the Automated Segmentation of Hippocampal Subfields in T1-MRI (T1-ASHS)[40–42] pipeline to further segment MTL subregions in ADNI patients. The MTL gray matter surface was modeled and vertex-wise thickness computed using the Cortical Reconstruction for ASHS (CRASHS)[43] pipeline. Mean subregional MTL thickness values for each patient was also obtained. Intracranial volume (ICV) was derived using ASHS as well[40–42]. We segmented WMH using a nn-UNet[44] model described elsewhere[45].

<u>^{18}F-flortaucipir PET processing</u>

For the ADNI cohort, post-processed tau PET images were obtained from the ADNI database. These images were generated by averaging co-registered individual 5-minute acquisition frames. PET images were coregistered to each participant's T1-weighted structural MRI using Advanced Normalization Tools (ANTs)[46]. For the Avid dataset, ^{18}F-flortaucipir PET images were processed following a similar workflow. Attenuation-corrected dynamic image frames were first motion-corrected using MCFLIRT[47] with six degrees of freedom, and the corrected frames were averaged to create static post-processed images.

^{18}F-flortaucipir tau PET-derived standardized uptake value ratio (SUVR) maps were generated and utilized for three purposes: (1) to generate 3D voxel-wise tau maps paired with cortical thickness for deep-learning-based T–N image translation, (2) to extract regional SUVR values for the region-wise linear T–N mismatch modeling and for analyses of composite tau burden in the meta-MTL ROI, (3) to compute a scalar global tau burden measure (TauMaX[48]) reflecting both the magnitude and spatial extent of tau pathology across the brain. The details were listed in sMethods (^{18}F-flortaucipir PET processing).

Amyloid-β

Amyloid-β status for ADNI participants was determined based on amyloid-β PET imaging, using tracer- and reference-specific SUVR thresholds, by UC Berkeley. Centiloid values were obtained from ADNI.

T(A)-N inter-modality image translation

The schematic overview of the workflow is presented in Figure 1. The deep-learning network was trained using both $T \rightarrow \hat{N}_T$ and $T + A \rightarrow \hat{N}_{T+A}$ image translation paradigms to learn the typical T–N relationship. At inference, cortical thickness was predicted from T or T + A inputs, and the reconstruction error (actual – predicted thickness) was utilized as the AI-derived T–N mismatch.

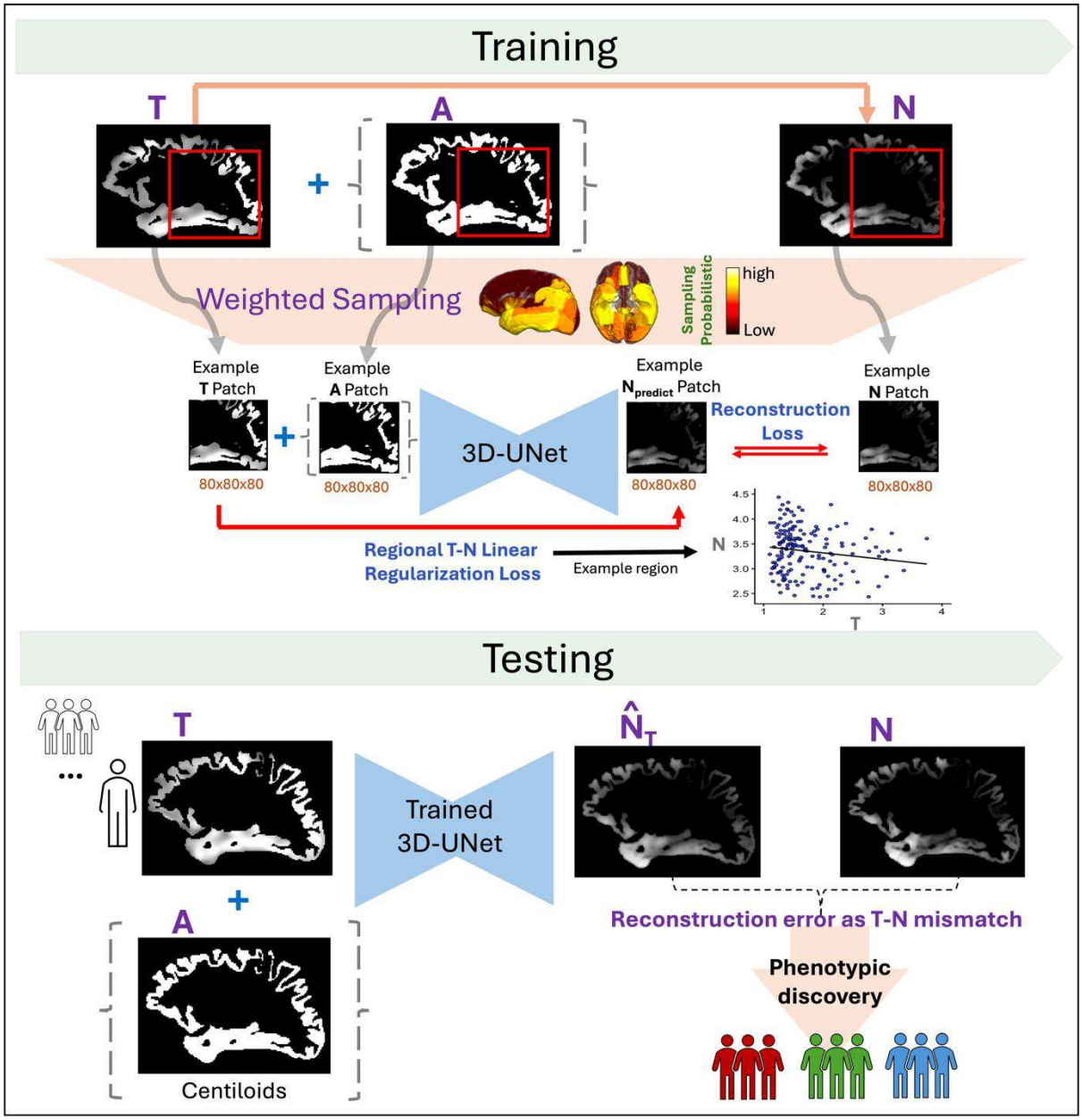


Figure 1. Schematic of the AI-based T–N mismatch framework. The network learns the mapping from T(A) to N during training. During testing, the trained model predicts N based on T(A), and the reconstruction error between the predicted and actual N serves as the T–N mismatch, which is used for phenotypic discovery.

Pre-processing Inputs for T-N image translation

Processed cortical thickness and tau SUVR maps in the template space, masked by the cortical gray matter, were used as paired neurodegeneration (N) and tau (T) inputs for inter-modality T–N image translation. The tau SUVR map served as the input, and the corresponding cortical

thickness map served as the target. The paired thickness and tau maps were further cropped to remove background padding, retaining a two-voxel margin around the brain boundary. For the $T + A \rightarrow \hat{N}_{T+A}$ image translation, we used the Centiloid[49] value to represent global amyloid burden. The Centiloid values were Winsorized at 150 and then normalized to a range of 0 to 2. For each participant, a 3D Centiloid image was generated by assigning the participant's normalized Centiloid value to every voxel within the cortical gray matter mask. This image was used together with the tau maps to reconstruct the corresponding 3D cortical thickness map.

Network implementation

Several approaches have been used for image translation, including U-Nets[50], Generative Adversarial Networks (GANs)[51], and diffusion models[52]. Here, we employed 3D U-Net[50] from the MONAI[53] framework in a simple supervised regression design to provide deterministic voxel-wise mapping from PET inputs to cortical thickness (sMethods Network implementation). The model accepted a single input channel for $T \rightarrow \hat{N}_T$ translation, and two input channels (tau and amyloid) for $T + A \rightarrow \hat{N}_{T+A}$ translation.

Training Loss

To reconstruct thickness images from the paired tau images, the network was trained using a composite loss function. Specifically, the total loss included three terms that capture voxel-wise similarity: mean squared error (MSE)[54] term, a 3D structural similarity (SSIM)[55] loss to preserve local anatomical structure of the thickness map, and a symmetric Kullback–Leibler (KL)[56] divergence term to align the global intensity distributions of predicted and true thickness images (sMethods Training loss section).

A key methodological challenge in training the $T \rightarrow \hat{N}_T$ translation model is avoiding overfitting to non-tau influences on thickness within the training cohort, as this may obscure biologically meaningful deviations later used to define T–N mismatch. For example, partial volume effects unrelated to tau pathology may carry information about cortical thinning even in regions with minimal tau burden, allowing the model to predict thickness from these non-tau cues rather than from the biological relationship between T and N. To address this, we further introduced a linear regularization loss term that simultaneously constrains the AI-predicted thickness toward that predicted by simple regional T–N linear regression, thereby encouraging the model to learn robust and generalizable tau-related neurodegeneration patterns. Specifically, this loss minimizes the discrepancy at the ROI level between the 3D U-Net–predicted thickness and the thickness estimated by simple region-wise T-N linear regression, with detailed description in sMethods. Such regional constraint helps prevent the model from overfitting to specific T-N associations in the training cohort, given the heterogeneity of the cohort where neurodegeneration (N) may not be fully explained by Tau (T).

The final loss function was defined as below:

$$\text{Loss} = 3 \cdot \text{MSE}(Y, \hat{Y}) + 5 \cdot \text{SSIM}(Y, \hat{Y}) + 5 \cdot \text{KL}(H(\hat{Y}) \parallel H(Y)) + 5 \cdot \text{KL}(H(Y) \parallel H(\hat{Y})) + 3 \cdot \sum_{r=1}^{44} w_r [\, \bar{z}_r - \bar{z}_r^{\text{lin}} \,]^2$$

where Y and $\hat{Y}$ denote the ground-truth and predicted thickness images, respectively. H(Y) and $H(\hat{Y})$ represent their normalized intensity histograms. The weights for the reconstruction terms (MSE, SSIM, KL) were empirically set to balance their differing native magnitudes based on pilot experiments, after which the weight for the global linear regularization term was optimized within the range of 0–1000, with a value of 3 yielding the strongest negative T–N association in the medial temporal lobe (Table S3). The total composite loss function allowed the model to reconstruct accurate cortical thickness maps based on tau (and amyloid) images while learning a robust and generalizable T–N relationship.

Training and testing

For the 184 unique A+ CI ADNI patients, we conducted five-fold cross-validation with 70% of participants used for training, 10% for validation, and 20% for testing. Within each fold, there was no overlap among training, validation, and testing sets. Across folds, each participant was included in the testing set exactly once and was used for training or validation in the rest of the folds. The demographics for patients in each fold are described in Table S4. Given that global tau burden was skewed toward higher values, global Tau-MaX[48] values were first Winsorized at 25, and uniform resampling was applied within each training fold to obtain a more balanced representation across the full tau range from 0 to 25. Details were described in sMethods (Training and testing section).

To assess external generalizability, the five $T \rightarrow \hat{N}_T$ image translation models trained via 5-fold cross-validation in ADNI were each applied to the independent Avid dataset. For each Avid participant, the resulting thickness predictions from all five ADNI-trained models were averaged to produce a single prediction, used for later analyses.

T-N mismatch modeling

AI-based T-N phenotypic clustering

Among the 100 ROIs across the whole brain, each regional AI-based T–N mismatch was represented by the mean reconstruction error, computed as the difference between the actual and AI predicted mean regional cortical thickness (actual thickness – AI predicted thickness). The predicted thickness for each participant was obtained from the corresponding held-out test fold in which that participant was not included in the training or validation sets across the 5-fold cross-validation. We then input this 100-element vector of AI derived T-N mismatch into hierarchical clustering[57] based on Euclidean distance for data-driven T-N phenotypic discovery.

The number of clusters was empirically determined based on dendrogram (Figure 4A) and inspired by our prior work[24–26].

LATE-related T-N mismatch modeling

Given our previous work[24–26] which suggested that the T-N mismatch approach can identify AD patients enriched with TDP-43/LATE as co-pathology and neuropathological evidence that LATE can be present in up to 50% of patients with AD[58], we specifically evaluated the potential of the AI-derived T-N mismatch to identify LATE-related co-pathology in AD.

We first selected eight bilateral ROIs (16 total ROIs) that are linked to LATE, specifically the amygdala, hippocampus, anterior cingulate gyrus, entorhinal cortex, inferior temporal gyrus, temporal pole, anterior orbitofrontal gyrus and middle frontal gyrus[19]. The T-N mismatch from those regions were used for LATE-related T-N mismatch modeling, with more details listed in sMethods.

We assessed the consistency of group assignments between the AI and linear regression approaches. We further evaluated differences in LATE-related signatures between individuals with concordant versus discordant group assignments.

Statistical Analyses

General statistical modeling

Statistical analyses were performed in R (v4.5). All statistical tests were two-sided.

Reconstruction accuracy

The reconstruction accuracy of the T–N image translation model was evaluated on the testing set across five folds, as well as in the Avid cohort. Performance was quantified using voxel-wise mean absolute error (MAE) within the cortical thickness mask and Spearman correlation

between predicted and observed cortical thickness values. Differences in reconstruction accuracy across folds were assessed using the Kruskal–Wallis test.

Regional T–N associations

Both local and remote regional T–N associations were assessed using partial Spearman correlations, adjusting for age. Resulting p-values were corrected for multiple comparisons using the false discovery rate (FDR). Differences in correlation strength were assessed using Hittner's z test.

Between-group comparisons

Between-group comparisons of continuous variables, including age, global tau burden (TauMaX), imaging-derived measures, and cognitive scores, were performed using general linear models, with group as the independent variable and relevant covariates included for adjustment. The comparison of TauMaX was adjusted for age and sex. The ERC/PHC thickness ratio was adjusted for age, sex, and global tau represented by TauMaX. Hippocampal volume and white matter hyperintensity volume were adjusted for age, sex, global tau, and ICV. The comparison of MMSE was adjusted for age, sex, and global tau. Bonferroni correction was applied to account for multiple comparisons.

Associations with T–N mismatch metrics

Associations between T–N mismatch metrics and global tau burden, imaging features, and cognitive performance were assessed using partial Spearman correlations with relevant covariates. Specifically, associations with global tau burden were adjusted for age and sex. Associations with ERC/PHC thickness ratio were adjusted for age, sex, and global tau. Associations with hippocampal volume were adjusted for age, sex, global tau, and ICV. Differences in correlation strength were assessed using Hittner's z test.

Longitudinal cognitive modeling

Longitudinal cognitive trajectories of AI-derived T–N groups were aligned along the time dimension to match the estimated time of tau onset (T+) and analyzed using linear mixed-effects models. The time of tau onset (referred to below as T+ age) was estimated using the Sampled Iterative Local Approximation (SILA)[59] model based on longitudinal TauMaX[48] measurements[60]. Cognitive score was modeled as the dependent variable, with fixed effects including time from T onset (T+), group, and their interaction (T+ × group), along with covariates for age and sex. A random intercept was included to account for within-participant correlations across repeated measures. Further details were described in sMethods (Longitudinal cognitive modeling)

Between-group voxel-wise atrophy comparisons

Vertex-wise comparisons of MTL thickness across T–N groups were performed using a general linear model implemented in meshglm (https://github.com/pyushkevich/cmrep). Patient-level MTL thickness maps were smoothed using surface-based diffusion smoothing with a Gaussian kernel (σ = 3 mm), with details described in sMethods.

Results

T(A)-N image reconstruction performance

We evaluated the performance of the T–N image translation model on the testing sets across five folds. The translation results between T and N were assessed both qualitatively and quantitatively. In Figure 2B, representative examples of the predicted and actual cortical thickness maps are displayed for several participants. Overall, the AI-predicted thickness maps derived from T ($T \rightarrow \hat{N}_T$) or T + A ($T+A \rightarrow \hat{N}_{T+A}$) inputs demonstrated strong visual agreement with the corresponding actual thickness maps. However, in some cases, the model showed less

accurate predictions, either underestimating (thinner than expected) or overestimating (thicker than predicted) cortical thickness, especially in the MTL region.

Quantitatively, the voxel-wise MAE within gray matter cortical thickness mask for the AI-predicted cortical thickness was 0.366 ± 0.048 mm for $T \rightarrow \hat{N}_T$ and 0.365 ± 0.051 mm for $T+A \rightarrow \hat{N}_{T+A}$. The MAE across the five cross-validation folds is summarized in Figure 2A, showing consistent performance with no significant difference between folds (*p=0.16* for $T \rightarrow \hat{N}_T$; *p=0.49* for $T+A \rightarrow \hat{N}_{T+A}$). Similarly, the voxel-wise Spearman correlation (ρ) between predicted and actual cortical thickness maps demonstrated strong agreement, with an average ρ of 0.911 ± 0.022 for $T \rightarrow \hat{N}_T$ and 0.912 ± 0.022 for $T+A \rightarrow \hat{N}_{T+A}$, with no significant variation across folds (p = 0.45 for $T \rightarrow \hat{N}_T$; p=0.45 for $T+A \rightarrow \hat{N}_{T+A}$). We then examined the regional MAE for both the linear T–N model and the AI-based T–N models (Figure 2C). The linear model exhibited generally higher regional errors, whereas the deep learning models showed lower prediction errors across regions, with the largest prediction error observed in temporal lobe regions.

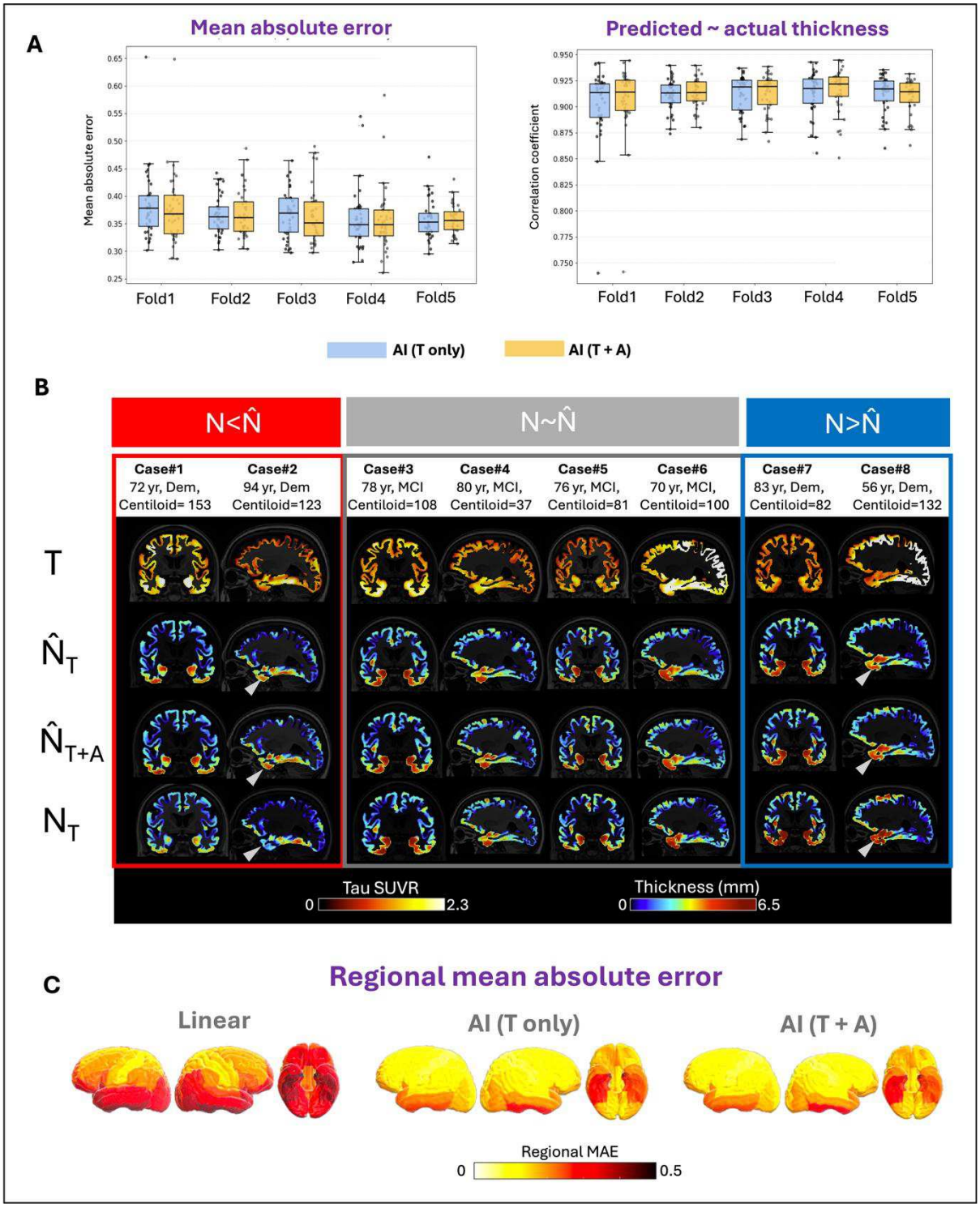


Figure 2. Evaluation of T(A)–N image translation performance. (A) Voxel-wise mean absolute error (MAE) and Spearman correlation between predicted and actual cortical thickness within the gray matter mask across 5-fold ADNI test sets for both AI models. (B) Example participants showing the Tau SUVR map, predicted cortical thickness maps derived from T alone ($\hat{N}_T$) and from T + A ($\hat{N}_{T+A}$), and the corresponding actual cortical thickness (N) map. The arrows point to the area with mismatch between predicted and actual thickness. (C) Regional MAE patterns between predicted and actual thickness for linear regression–based T–N mismatch, AI(T) mismatch, and AI(T + A) mismatch models.

Image translation effectively captured T-N relationships

Next, we evaluated whether the AI model effectively learned T-N relationships. We hypothesized that the predicted T-N relationship would be comparable to or stronger than the actual T-N relationship since it would not be able to account for other factors that modulate it. Across the whole brain (Figure 3A), the regional T–N association patterns derived from both the $T \rightarrow \hat{N}_T$ and $T+A \rightarrow \hat{N}_{T+A}$ models closely mirrored the spatial patterns of the actual T–N relationships. Furthermore, while the reconstruction error was relatively higher in the MTL than in other regions, the T–N association within the meta-MTL region was more pronounced in the AI predictions (Figure 3B). The $T \rightarrow \hat{N}_T$ model ($\rho = -0.471$, $p < 0.001$) and the $T + A \rightarrow \hat{N}_{T+A}$ model ($\rho = -0.401$, $p < 0.001$) both showed significant T-$\hat{N}$ correlations, with numerically greater coefficient and tighter associations than the observed T–N relationship ($\rho = -0.371$, $p < 0.001$). The predicted T-N association under $T \rightarrow \hat{N}_T$ model was significantly stronger than the observed T-N relationship ($z = -2.20$, $p = 0.028$). However, adding A as an additional input did not significantly alter the strength of the T–N association, suggesting that tau alone provided most of the power to predict neurodegeneration in this context.

We further hypothesized that inter-modality T–N image translation would capture non-local effects of tau pathology on neurodegeneration. To test this, we examined remote T–N relationships beyond local effects. Figure 3C quantified the number of effective remote T-N associations, defined as T-N connections with $\rho < -0.3$ and $p < 0.05$, for both the actual and AI predicted T N relationships. The $T \rightarrow \hat{N}_T$ and $T + A \rightarrow \hat{N}_{T+A}$ models displayed a greater number of effective cross-regional connections, with 88 (46 within-lobe and 42 between-lobe) and 67 associations (29 within-lobe and 38 between-lobe), respectively, compared with 27 connections (13 within-lobe and 14 between-lobe) in the observed T~N data. These findings suggest that the AI models capture both local and remote T–N relationships, with predicted

thickness showing greater reliance on tau SUVR, as expected given that tau serves as the primary model input. Again, adding A as an additional input here did not further strengthen the predicted T–N associations.

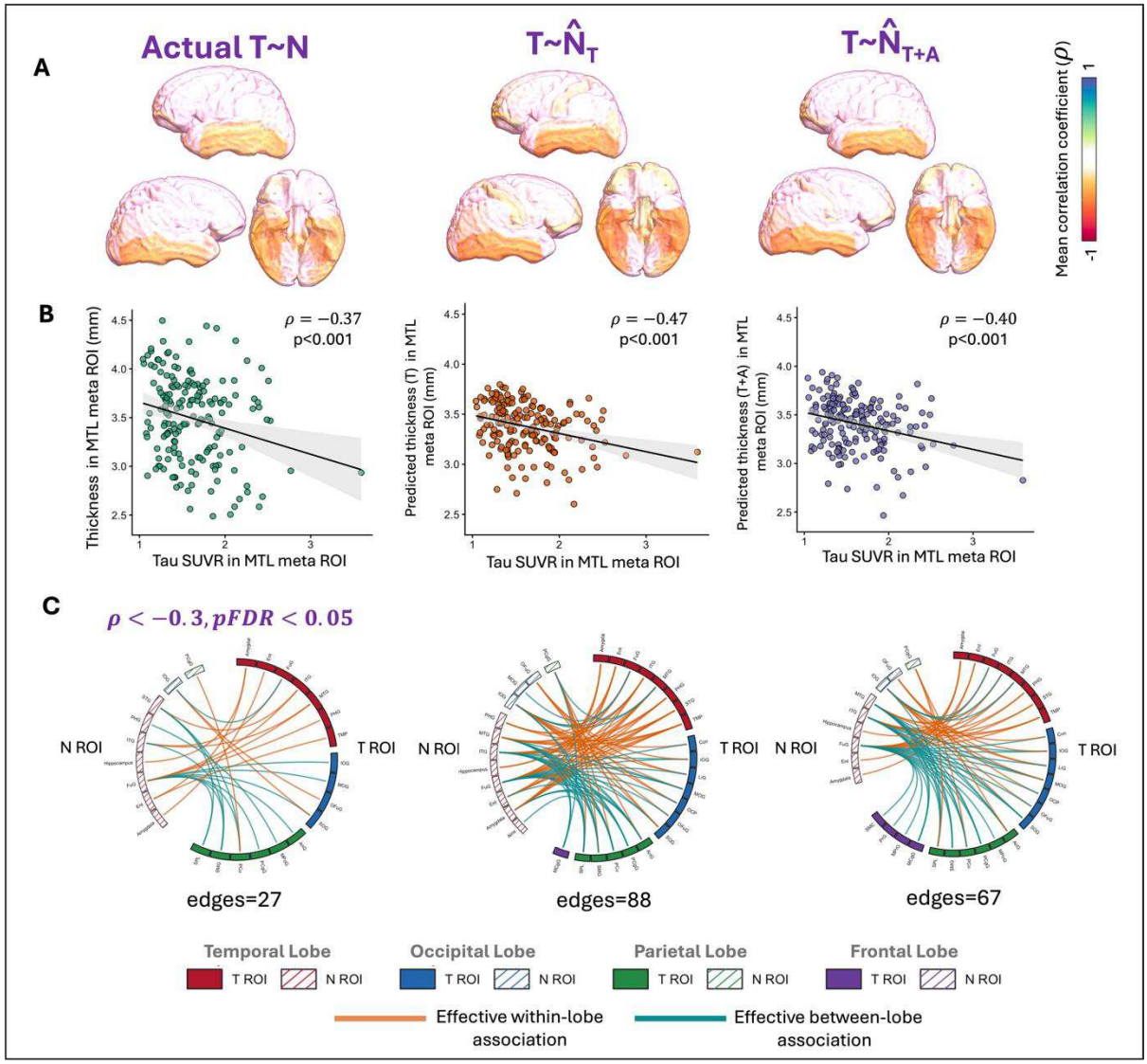


Figure 3. AI-predicted T–N associations. (A) Regional Spearman correlations between tau and neurodegeneration (T–N) for the actual data, AI(T)-predicted T→$\hat{N}_T$, and AI(T+A)-predicted T→$\hat{N}_{T+A}$, shown from left to right. (B) T–N Spearman associations within the meta-MTL region of interest (ROI) for the same three models (actual, AI(T), and AI(T+A)). (C) Circos plots illustrating significant remote T–N associations ($\rho < -0.3$, $p_{FDR} < 0.05$) for actual, AI(T), and AI(T+A) models, respectively from left to right. All the T-N associations were covaried with age. Each sector represents an individual brain ROI.

"Reconstruction error" as the T-N mismatch for AD patient stratification

We explored the reconstruction error (N-$\hat{N}_T$) in T→$\hat{N}_T$ as a measure of T-N mismatch. Using regional reconstruction error patterns, we clustered ADNI participants into five data-driven groups. Among them, the group with the largest number of patients was labeled as "canonical", and this group was characterized by overall low error (N~$\hat{N}_T$) in the whole brain but slightly overestimating thickness (N>$\hat{N}_T$) in the temporal lobe (Figure 4B). Three clusters exhibited lower thickness than predicted (N<$\hat{N}_T$) and were thus labeled as "vulnerable." One of these showed a predominantly anterior pattern, with negative residual (N<$\hat{N}_T$) in the orbitofrontal and frontal regions (anterior vulnerable; Figure 4B). The other two clusters displayed highly similar temporal–limbic vulnerability patterns and belonged to the same branch in the dendrogram (Figure 4A). Given their similarity (Figure S3), we merged these two clusters into a single "limbic vulnerable" group, resulting in four AI-derived T–N groups (Figure 4B). Additionally, one cluster exhibited a global pattern of observed N being greater than predicted thickness (N>$\hat{N}_T$), which we designated as "resilient".

The limbic vulnerable group was significantly older ($p<0.01$, mean age = 79.2 ± 6.7, Figure 4I) and showed higher global tau burden ($p<0.01$, Figure 4E) compared with the canonical group. Nevertheless, after adjusting for global tau and age, this group showed significantly lower ERC/PHC thickness ratio (Figure 4F), reduced hippocampal volume (Figure 4G), and worse global cognitive impairment (Figure 4J), especially in memory (Table S2), relative to all other groups -- features suggestive of comorbid LATE in AD. Conversely, the anterior vulnerable group had significantly greater white matter hyperintensity volume than the canonical group (Figure 4H), suggesting an association with vascular disease.

We then evaluated the structural differences between AI-derived T-N groups both within the MTL (Figure 4C) and across the whole brain (Figure 4D). Within the MTL, the limbic vulnerable group exhibited greater anterior atrophy compared with the canonical group, involving the entorhinal cortex, BA35, and anterior hippocampal regions. At the whole-brain

level, this group showed more extensive atrophy extending into the anterior temporal–limbic regions. Consistent with our prior findings[24,25,61], these atrophy patterns closely resembled LATE-like atrophy[62]. While the anterior vulnerable group did not show significant atrophy within the MTL compared to the canonical group, there was broader cortical involvement with greater neurodegeneration across the orbital–frontal, frontal, and posterior cortical regions. Conversely, the resilient group displayed less atrophy both within and beyond the MTL compared with the canonical group, consistent with widespread structural preservation.

The AI-derived T-N groups also exhibited distinct cognitive profiles longitudinally anchored to the estimated T onset. The limbic vulnerable group demonstrated a significantly faster decline in memory ($p < 0.05$) and language ($p < 0.001$) compared with the canonical group, whereas the resilient group showed a slower decline in memory ($p < 0.001$) and language ($p < 0.05$). No significant group differences were observed in executive function.

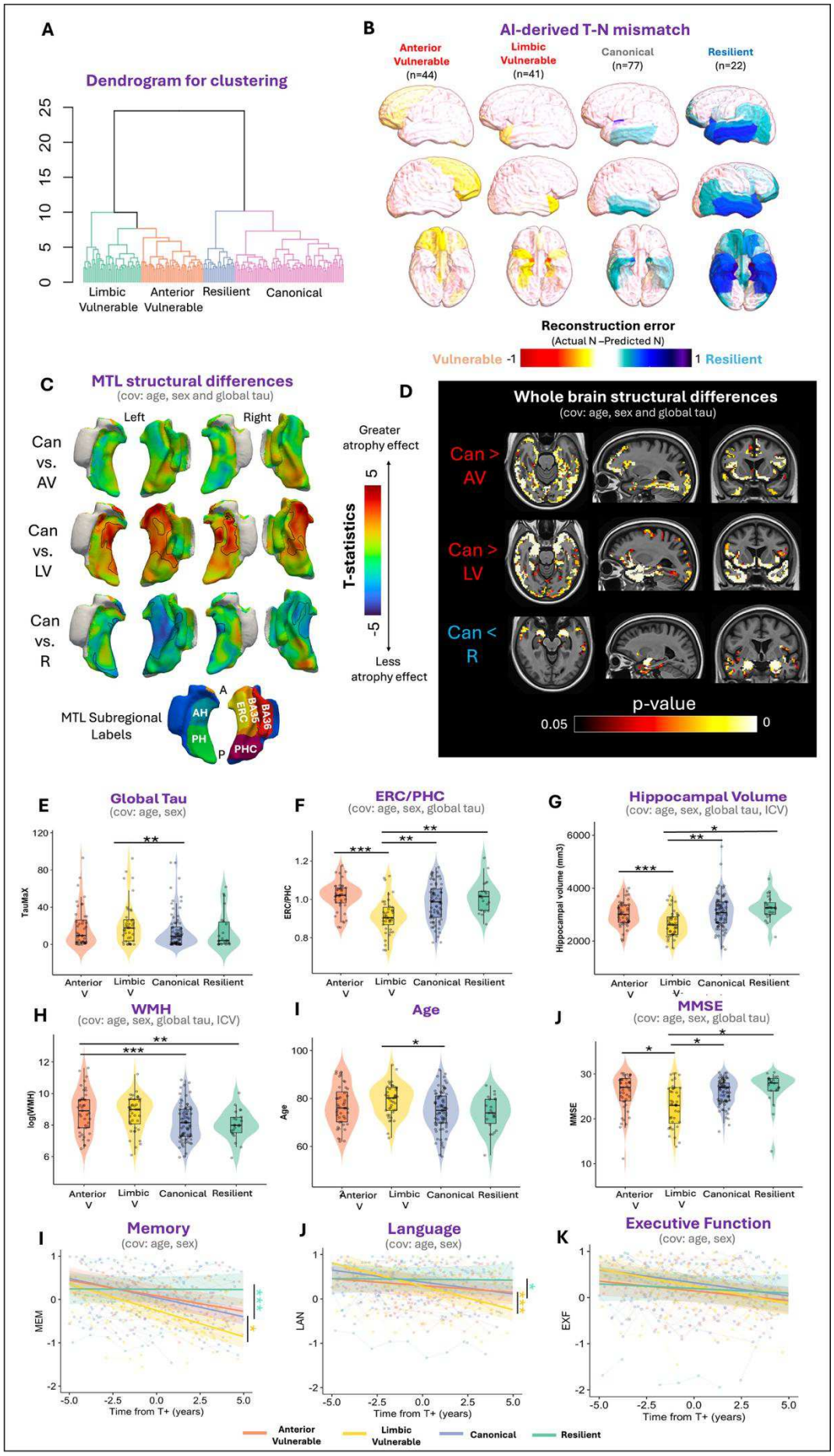

A
Dendrogram for clustering
Limbic Vulnerable
Anterior Vulnerable
Resilient
Canonical
B
AI-derived T-N mismatch
Anterior Vulnerable (n=44)
Limbic Vulnerable (n=41)
Canonical (n=77)
Resilient (n=22)
Reconstruction error
(Actual N –Predicted N)
Vulnerable -1
1 Resilient
C
MTL structural differences
(cov: age, sex and global tau)
Left
Right
Can vs. AV
Can vs. LV
Can vs. R
MTL Subregional Labels
AH
PH
ERC
BA35
BA36
PHC
D
Whole brain structural differences
(cov: age, sex and global tau)
Greater atrophy effect
T-statistics
Less atrophy effect
Can > AV
Can > LV
Can < R
p-value
0.05
0
E
Global Tau
(cov: age, sex)
TauMaX
F
ERC/PHC
(cov: age, sex, global tau)
ERC/PHC
G
Hippocampal Volume
(cov: age, sex, global tau, ICV)
Hippocampal volume (mm3)
H
WMH
(cov: age, sex, global tau, ICV)
log(WMH)
I
Age
Age
J
MMSE
(cov: age, sex, global tau)
MMSE
Anterior
Limbic
Canonical
Resilient
I
Memory
(cov: age, sex)
MEM
J
Language
(cov: age, sex)
LAN
K
Executive Function
(cov: age, sex)
EXF
Time from T+ (years)
Anterior Vulnerable
Limbic Vulnerable
Canonical
Resilient

Figure 4. "Reconstruction error" ($T \rightarrow \hat{N}_T$) as the T–N mismatch. (A) Dendrogram for AI-derived clustering of T–N mismatch based on regional reconstruction errors. (B) Spatial patterns of reconstruction errors for each AI-derived T–N mismatch group. (C) MTL vertex-wise cortical thickness comparisons between AI-derived T–N mismatch groups and the canonical group, controlling for age, sex, and global tau. Black edges delineate regions with significant differences ($p_{FWER} < 0.05$). (D) Whole-brain voxel-wise cortical thickness comparisons between the vulnerable and canonical groups (top), and between the resilient and canonical groups (bottom), adjusted for age, sex, and global tau. Colored area indicates significant differences ($p_{FWER} < 0.05$). (E–J) Group comparisons across additional imaging and biomarker features, adjusted for relevant covariates as indicated. (I–K) Longitudinal cognitive trajectories anchored to estimated T onset, adjusted for age and sex. Significance levels corrected for Bonferroni multiple comparisons are denoted as *p < 0.05, **p < 0.01, ***p < 0.001. Only significant differences from the canonical group are shown for I-K.

Sensitivity of linear regression vs. AI-derived T–N mismatch for identifying LATE-like features

Next, we compared the sensitivity of linear regression and AI-based T–N mismatch approaches for identifying a LATE-like pattern of atrophy beyond effects of T in AD patients. The LATE-related mismatch metrics (as described in Methods: LATE-Related T–N Mismatch Modeling) were computed using regional mismatch values derived from LATE-relevant regions, such that higher values indicate greater vulnerability and lower values indicate greater resilience, for both the linear and AI-derived models. As shown in Figure 5A, the mismatch metric defined by linear regression showed no significant association with global tau, whereas both AI-derived mismatch metrics demonstrated modest positive associations with global tau burden ($\rho = 0.129$, $p < 0.05$ for $T \rightarrow \hat{N}_T$; $\rho = 0.151$, $p < 0.05$ for $T + A \rightarrow \hat{N}_{T+A}$). Nevertheless, after adjusting for global tau and age, AI-derived mismatch metrics showed stronger associations with LATE-related imaging features than the linear model. Specifically, both AI-derived metrics were significantly associated with lower ERC/PHC thickness ratio ($\rho = -0.236$, $p < 0.001$ for $T \rightarrow \hat{N}T$; $\rho = -0.226$, $p < 0.001$ for $T + A \rightarrow \hat{N}T + A$) and reduced hippocampal volume ($\rho = -0.371$, $p < 0.001$ for $T \rightarrow \hat{N}_T$; $\rho = -0.405$, $p < 0.001$ for $T + A \rightarrow \hat{N}_{T+A}$). Although the linear model also

showed significant associations with ERC/PHC thickness ratio ($\rho = -0.157$, $p < 0.001$) and hippocampal volume ($\rho = -0.258$, $p < 0.001$), direct comparison demonstrated that the association between mismatch and ERC/PHC thickness ratio was significantly stronger for the AI-derived $T \rightarrow \hat{N}_T$ model than for the linear model ($z = 1.98$, $p = 0.048$), and the $T + A \rightarrow \hat{N}_{T+A}$ model showed a numerically stronger but non-significant difference from the linear model ($z = 1.51$, $p = 0.13$). For hippocampal volume, both AI-derived mismatch metrics exhibited significantly stronger associations than the linear model ($z = 2.89$, $p = 0.0039$ for $T \rightarrow \hat{N}_T$; $z = 3.33$, $p < 0.001$ for $T + A \rightarrow \hat{N}_{T+A}$). No significant differences were observed between the two AI-derived models in their associations with either ERC/PHC thickness ratio ($z = -0.34$, $p = 0.73$) or hippocampal volume ($z = 1.31$, $p = 0.19$), suggesting that incorporating amyloid as an additional input did not further enhance the relationship between mismatch and these LATE-related signatures.

Beyond evaluating mismatch metrics alone, we partitioned participants into data-driven T–N groups based on mismatch patterns restricted to LATE-relevant ROIs (see Methods: LATE-Related T–N Mismatch Modeling) and compared how each approach captured LATE-consistent patterns.

For each approach, patients were clustered into vulnerable, canonical, and resilient groups to allow direct comparison. Given these analyses focused specifically on LATE-related regions, we expected that the vulnerable groups would be enriched in AD patients with LATE as co-pathology. Across all three approaches, the vulnerable groups were associated with LATE-related imaging features (Figure S4). As shown in Figure 5D, the vulnerable group defined by the linear T–N regression exhibited a broader extent and greater magnitude of mismatch compared with the AI-derived groups.

We next examined the stability of group assignments across approaches by assessing transitions in patient classification from the linear model to the AI models. There were 65.2% of patients that maintained consistent group membership between the linear and AI ($T \rightarrow \hat{N}_T$) models. We further examined patients with discordant group assignments between the linear and AI models ($T \rightarrow \hat{N}_T$) and found they did not show difference in global tau burden compared to with those with stable assignments ($p = 0.68$; Figure 5F), indicating that these discordances were likely not driven by tau pathology. However, the transition patterns differed in LATE-related imaging features. Patients who transitioned from the vulnerable group in the linear model to other groups under the AI model ($T \rightarrow \hat{N}_T$) exhibited higher ERC/PHC thickness ratios and larger hippocampal volumes compared with those who remained in the vulnerable group, after adjusting for tau (although without correction for multiple comparisons). In contrast, patients reassigned from the canonical group in the linear model to the vulnerable group under the AI model did not show significant differences in ERC/PHC thickness ratio or hippocampal volume compared with those who remained in the vulnerable group. The AI-derived T-N mismatch yielded results consistent with but potentially more specific than the linear regression approach.

To study if using A as an additional input to the T-N image translation improved the sensitivity of model uncovering LATE, we also compared the transition patterns between $T \rightarrow \hat{N}_T$ and $T+A \rightarrow \hat{N}_{T+A}$. There were 78.3% of the patients that showed stable group assignment. However, we did not observe significant differences in the transition patterns with respect to ERC/PHC thickness ratio or hippocampal volume, although patients who transitioned from the canonical to the vulnerable group tended to exhibit higher global tau burden compared to resilient or canonical patients (Figure S4).

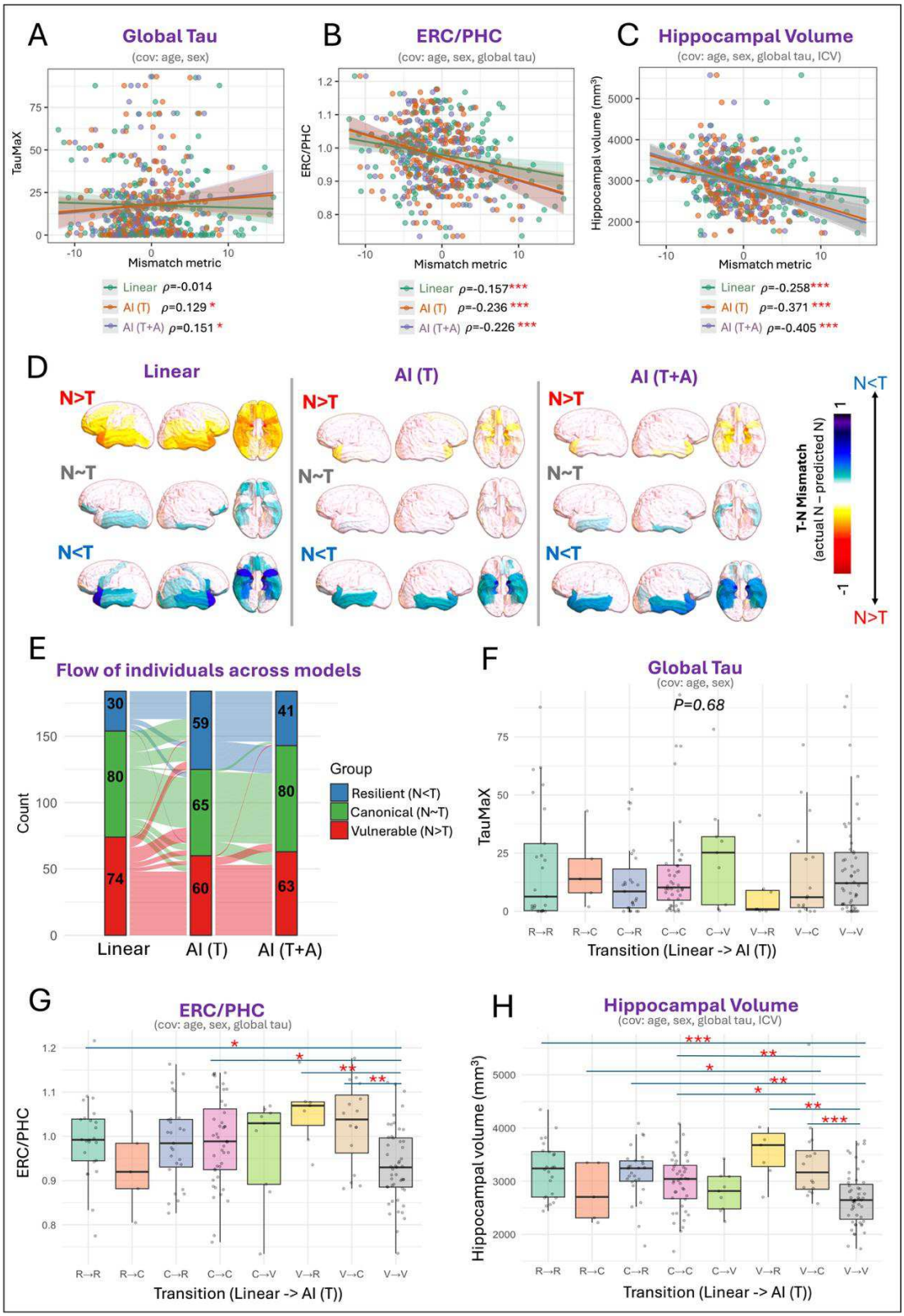


Figure 5. Comparison of linear and AI-derived T–N mismatch models for identifying LATE in AD. (A–C) Spearman associations between the calculated mismatch metric and (A) global tau, (B) ERC/PHC thickness ratio, and (C) hippocampal volume across linear, AI(T), and AI(T + A) T–N mismatch models. Covariates included in each model are indicated within the plots. (D) Spatial patterns of T–N mismatch (actual – predicted thickness) for data-driven T–N groups identified from the linear, AI(T), and AI(T + A) based T-N mismatch. (E) Flow diagram illustrating transitions

in group membership across the three models. (F–H) Transition patterns in global tau (F), ERC/PHC thickness ratio (G), and hippocampal volume (H) when moving from the linear to AI(T) model. Covariates for each comparison are shown within the corresponding panels. Comparisons were not adjusted for multiple testing. Significance levels are denoted as *p < 0.05, **p < 0.01, ***p < 0.001. Abbreviations: R= Resilient, C=Canonical, V=Resilient.

Generalization of the approach to external dataset

We applied the pretrained $T \rightarrow \hat{N}_T$ image translation models from ADNI cohort to an independent cohort of 70 patients from the Avid cohort and evaluated model performance in this external dataset as described in the method section. The MAE and Spearman correlation between predicted and actual cortical thickness predicted from the five training folds in ADNI are summarized in Figure 6A, again showing consistent performance across folds and the performance was comparable to the ADNI cohort. For each participant, the predicted thickness was averaged across the five folds and used for the following analyses. The averaged predicted thickness showed a mean MAE of 0.404 ± 0.091 mm and a Spearman correlation of 0.898 ± 0.013 with the actual cortical thickness within the gray matter cortical mask. Consistent with the ADNI findings, the regional MAE from the AI model was lower than that of the linear regression model, with the largest errors observed in the MTL. Similarly, in the Avid cohort, the AI-derived T–N mismatch metric was not significantly associated with global tau burden ($\rho = 0.13$, $p = 0.14$), but showed significant associations with lower ERC/PHC thickness ratio ($\rho = –0.34$, $p < 0.001$), reduced hippocampal volume ($\rho = –0.46$, $p < 0.001$), and greater cognitive impairment ($\rho = 0.23$, $p < 0.001$), after adjusting for age, tau and sex, which again tended to be numerically stronger than linear mismatch.

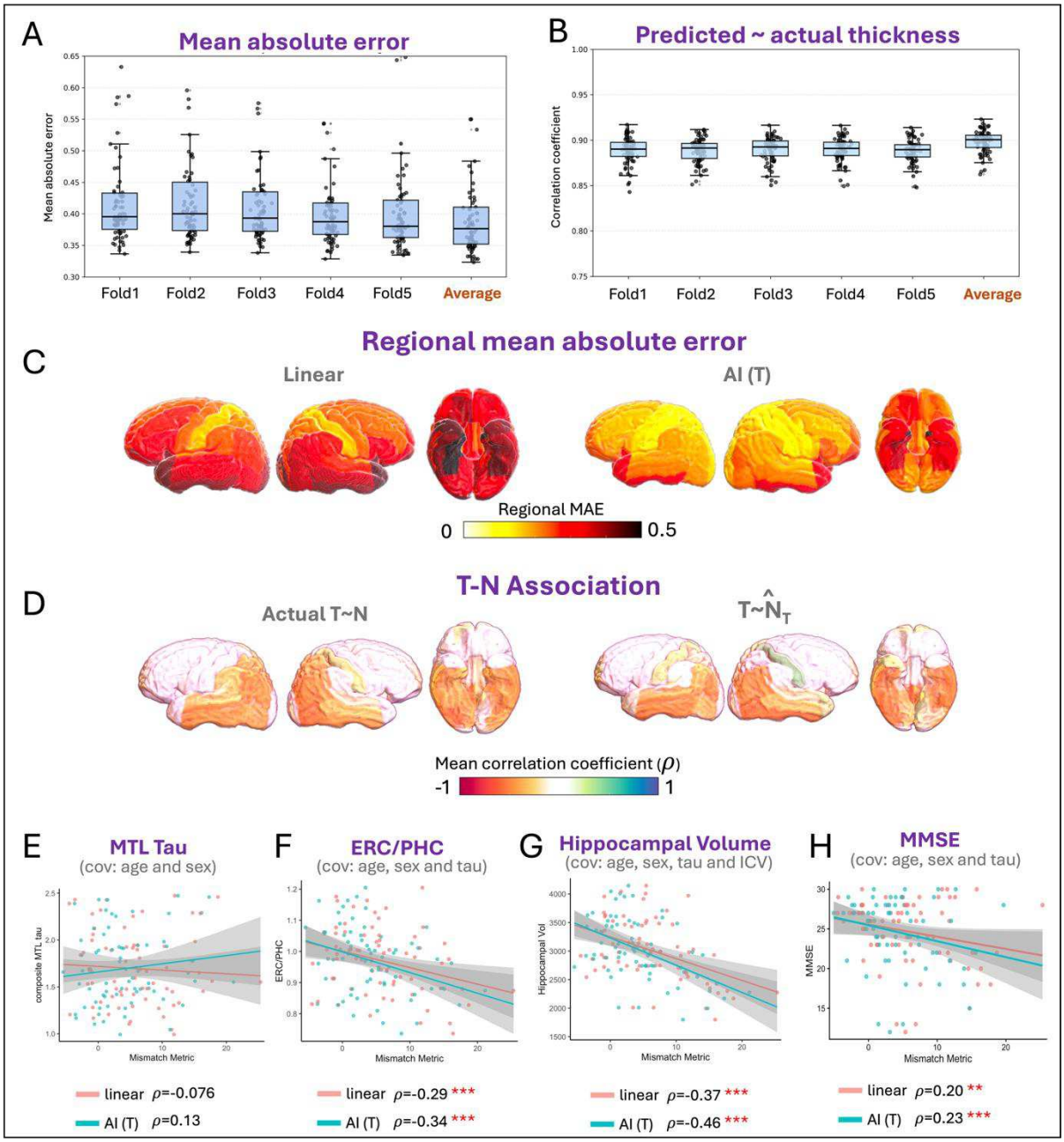


Figure 6. Generalization of AI(T)-derived T–N mismatch to an external Avid cohort (n = 70). (A–B) Voxel-wise mean absolute error (MAE) and Spearman correlation between predicted and actual cortical thickness within the gray matter mask, shown across five folds and averaged across folds. (C) Regional MAE comparison between linear T–N modeling and AI(T) modeling. (D) Regional T–N association patterns across different T–N models. (E–H) Associations between the mismatch metric and various imaging or biomarker features for both linear and AI models. Covariates for each comparison are indicated within the corresponding panels. Significant associations are denoted as *p < 0.05, **p < 0.01, ***p < 0.001.

Discussion

Deep learning has shown tremendous promise in advancing AD research[63]. One of the most active areas has been inter-modality MRI-to-PET image synthesis, mostly motivated by the high cost and limited accessibility of PET imaging. Prior studies have attempted to generate amyloid-PET, tau-PET, or ^{18}F-FDG-PET from structural MRI, leveraging anatomical context to approximate molecular pathology[31,64,65]. Although these approaches hold promise, they primarily focused on performance-driven translation rather than biological interpretation. Moreover, while these approaches achieve decent visual and quantitative fidelity, they inherently assume a one-to-one correspondence between AD pathology and downstream neurodegeneration. In reality, the atrophy among AD patients is heterogeneous. Coexisting non-AD pathologies, such as LATE, cerebrovascular disease, and α-synuclein, increase the degree of neurodegeneration beyond that due to AD pathology[19,66–68], whereas individual differences in resilience factors may attenuate it[69–72]. Therefore, while the association between tau pathology and neurodegeneration is relatively strong, variability beyond measurement noise commonly exists.

The current study extends the T-N mismatch framework[24,24,25,61,73] by modeling the T–N relationship in a nonlinear and spatially distributed manner using 3D-U-Net–based inter-modality image translation. The network learns to synthesize downstream neurodegeneration directly from tau, learning the “typical” voxel-wise T–N mapping across AD patients. Notably, this "typical" relationship does not necessarily reflect a "pure AD" relationship, given that the majority of AD cases harbor co-pathologies.

Within this framework, the reconstruction error, which is the difference between observed and synthesized neurodegeneration, serves as an AI-derived T–N mismatch, or anomaly map, reflecting deviations from the learned normative T–N relationship. Conceptually, this extends

classical anomaly detection beyond deviations from healthy controls to within-AD heterogeneity, allowing detection of non-AD comorbid effects and resilience that influences neurodegeneration in AD patients. It reframes image translation from a mere image synthesis task into a data-driven framework for identifying biologically meaningful deviations from the canonical T–N relationship in AD, which can be used for AD patient stratification.

Our findings suggest that the network generally reconstructed N from T with high accuracy, although there were still individuals with localized under- or over-estimation of N. The AI-predicted T–N relationship showed spatial patterns comparable to, and in MTL even stronger than, the actual T–N associations, suggesting that the model did indeed rely on T for estimating N to a greater extent than the actual biological relationship due to the presence of other modulators (e.g. copathologies). Moreover, beyond local effects, tau-driven remote influence on predicted thickness may be stronger than it appears in real data. To further evaluate the contribution of amyloid, we incorporated Aβ as an additional input to test whether combining both AD pathologies could better predict N. Prior studies[4,74] have debated whether N in AD is primarily driven by T alone or whether synergistic interactions between A and T jointly contribute to downstream atrophy. In our analysis, adding the global measurement of amyloid did not provide improvement in reconstruction performance or greatly alter the learned T–N relationship. These results support that tau burden alone is likely to explain most of the variance in N driven by AD, although additional contributions from spatially localized patterns of amyloid burden cannot be ruled out.

Interestingly, although the MTL exhibited one of the strongest T-N associations across the cortex, it also showed the largest reconstruction error. In fact, the MTL is not only the earliest region affected by neurofibrillary tangles (NFTs)[75] but also an early site of multiple co-pathologies, including LATE-NC[76] which can present in up to 50% of individuals with AD[58].

To further assess whether the reconstruction error reflects biologically meaningful processes, we performed clustering on regional reconstruction error maps from the ADNI cohort. This analysis identified four distinct data-driven T–N phenotypes, each characterized by unique spatial patterns of deviation between predicted and observed neurodegeneration. Notably, these subtypes were generally consistent with those identified in our prior work[24–26] using linear T–N mismatch modeling, underscoring the reproducibility of the T–N mismatch framework. Among these, the limbic vulnerable group showed greater-than-predicted neurodegeneration in temporal–limbic regions and displayed structural patterns consistent with LATE, extending anteriorly within the medial temporal lobe and across the broader limbic network[16,76,77]. This group exhibited faster cognitive decline, particularly in memory and language domains and demonstrated a lower ERC/PHC thickness ratio and reduced hippocampal volume after adjusting for global tau burden. These imaging features, previously linked to TDP-43 in postmortem studies[16,18], support the interpretation that the limbic vulnerable phenotype is likely enriched in those with LATE. The anterior vulnerable group, which exhibited greater cortical atrophy than AI-predicted, showed associations with WMH, suggesting a potential contribution of vascular pathology. In contrast, the resilient group, characterized by less-than-expected neurodegeneration, displayed reduced structural atrophy and slower decline in memory and language relative to the canonical group. Taken together, these findings indicate that the reconstruction error captured by the AI model appears to be associated with biologically and clinically meaningful variability driven by non-AD factors, rather than merely noise. In addition to vascular disease and LATE, α-synuclein pathology is another common co-pathology in AD. However, our previous findings[24,26] suggest that neurodegeneration measured using cortical thickness is less sensitive to α-synuclein pathology, whereas FDG-PET is more sensitive to its effects.

Furthermore, the AI-derived mismatch demonstrated greater potential in detecting LATE-related effects compared with the linear T–N mismatch. Although vascular disease is another common co-pathology that frequently co-occurs with LATE in AD, our previous work did not find associations between vascular imaging signatures and gray matter thinning in medial temporal lobe regions that are vulnerable to LATE. When clustering was performed using regional mismatch measurement restricted to LATE-related regions, the group assignments derived from the AI-based and linear approaches showed large overlap, which again support the validity of the general T-N mismatch concept. Nevertheless, the transitions between the two models were not explained by differences in global tau burden but rather by LATE-related features. Individuals who shifted from vulnerable under the linear-regression modeling to non-vulnerable groups under the AI-derived model showed reduced LATE-associated features compared to those who were vulnerable across methods. In the opposite direction, individuals shifting from non-vulnerable to vulnerable groups did not exhibit increased LATE-associated features. This asymmetric pattern supports that the AI-derived T-N mismatch is likely to provide a cleaner and more biologically specific separation of T–N phenotype.

Importantly, the AI-derived mismatch was generalizable to an independent Avid dataset. The model achieved comparable reconstruction performance to what was observed in the ADNI cohort, and the derived mismatch metric showed similar associations with LATE-related imaging features. These findings support the robustness and external generalizability of the AI-based T–N mismatch framework.

Our study has several limitations. First, the training sample size was modest, which may limit the generalizability of the learned T–N relationship. Because the goal of our model was to learn a “normative” mapping between T and N, a larger and more diverse training dataset would likely improve generalizability and representativeness. To mitigate overfitting, we designed a linear T–N regularization loss that discouraged the model from learning non-tau variation in

cortical thickness, such as tracer partial volume effects that may persist beyond partial volume correction or other confounding influences. We also adopted a 3D U-Net architecture, which generally requires fewer training samples than generative or diffusion-based image translation models. Importantly, the model achieved comparable performance in both internal and external testing cohorts. Nevertheless, ADNI participants are generally less heterogeneous in race, ethnicity, and disease severity than community-based populations, and future studies should validate the model in larger, independent, and more diverse cohorts. Second, amyloid was incorporated into the model using a single global Centiloid value expanded as an image rather than direct amyloid SUVR images. This choice was based on the fact that ADNI patients were scanned with different amyloid tracers, and the Centiloid scale provides a tracer-independent measure of global amyloid burden. However, using voxel-wise amyloid SUVR maps could better capture the spatial heterogeneity of amyloid deposition and may provide additional information beyond a global summary measure.

In conclusion, this study extends the concept of T–N mismatch by using inter-modality image translation to model the nonlinear and spatially distributed relationship between tau and neurodegeneration in AD. This AI-based T-N mismatch framework moves beyond image synthesis to identify meaningful deviations from the normative T–N relationship, capturing heterogeneity that reflects non-AD co-pathologies and resilience. It validates the T-N mismatch concept and offers a novel approach for AD patient stratification, understanding AD heterogeneity, and advancing precision medicine in AD.

CONFLICTS

Ilya Nasrallah has served on the scientific advisory board for Eisai and done educational speaking for Biogen. Long Xie received personal consulting fees from Galileo CDS, Inc.; was employed by Siemens Healthineers during this project; and is currently a full-time employee

of Ant Group. Sandhitsu Das has received consulting fees from Rancho Biosciences and Nia Therapeutics. David Wolk has served as a paid consultant for Eli Lilly and Beckman Coulter. He has also served on the DSMB for Functional Neuromodulation and GSK. He has received research support paid to his institution by Biogen. The rest of the authors declare no conflicts of interest.

Funding

The study was supported by the funding R01 AG072796, P30 AG072979, R01 AG056014 and RF1 AG069474.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer’s Association; Alzheimer’s Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer’s Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Competing interests

The authors report no competing interests.

References

1. Jack CR, Bennett DA, Blennow K, et al. NIA-AA Research Framework: Toward a biological definition of Alzheimer's disease. *Alzheimers Dement J Alzheimers Assoc*. 2018;14(4):535-562. doi:10.1016/j.jalz.2018.02.018

2. Jack Jr. CR, Andrews JS, Beach TG, et al. Revised criteria for diagnosis and staging of Alzheimer's disease: Alzheimer's Association Workgroup. *Alzheimers Dement*. 2024;20(8):5143-5169. doi:10.1002/alz.13859

3. Harrison TM, La Joie R, Maass A, et al. Longitudinal tau accumulation and atrophy in aging and alzheimer disease. *Ann Neurol*. 2019;85(2):229-240. doi:10.1002/ana.25406

4. Joie RL, Visani AV, Baker SL, et al. Prospective longitudinal atrophy in Alzheimer's disease correlates with the intensity and topography of baseline tau-PET. *Sci Transl Med*. 2020;12(524). doi:10.1126/scitranslmed.aau5732

5. Bejanin A, Schonhaut DR, La Joie R, et al. Tau pathology and neurodegeneration contribute to cognitive impairment in Alzheimer's disease. *Brain J Neurol*. 2017;140(12):3286-3300. doi:10.1093/brain/awx243

6. La Joie R, Visani AV, Baker SL, et al. Prospective longitudinal atrophy in Alzheimer's disease correlates with the intensity and topography of baseline tau-PET. *Sci Transl Med*. 2020;12(524):eaau5732. doi:10.1126/scitranslmed.aau5732

7. Das SR, Xie L, Wisse LEM, et al. Longitudinal and cross-sectional structural magnetic resonance imaging correlates of AV-1451 uptake. *Neurobiol Aging*. 2018;66:49-58. doi:10.1016/j.neurobiolaging.2018.01.024

8. Kapasi A, Schneider JA. Vascular contributions to cognitive impairment, clinical Alzheimer's disease, and dementia in older persons. *Biochim Biophys Acta*. 2016;1862(5):878-886. doi:10.1016/j.bbadis.2015.12.023

9. Katsumata Y, Fardo DW, Kukull WA, Nelson PT. Dichotomous scoring of TDP-43 proteinopathy from specific brain regions in 27 academic research centers: associations with Alzheimer's disease and cerebrovascular disease pathologies. *Acta Neuropathol Commun*. 2018;6(1):142. doi:10.1186/s40478-018-0641-y

10. Lamar M, Boots EA, Arfanakis K, Barnes LL, Schneider JA. Common Brain Structural Alterations Associated with Cardiovascular Disease Risk Factors and Alzheimer's Dementia: Future Directions and Implications. *Neuropsychol Rev*. 2020;30(4):546-557. doi:10.1007/s11065-020-09460-6

11. Robinson JL, Richardson H, Xie SX, et al. The development and convergence of co-pathologies in Alzheimer's disease. *Brain*. 2021;144(3):953-962. doi:10.1093/brain/awaa438

12. Kapasi A, Yu L, Boyle PA, Barnes LL, Bennett DA, Schneider JA. Limbic-predominant age-related TDP-43 encephalopathy, ADNC pathology, and cognitive decline in aging. *Neurology*. 2020;95(14):e1951-e1962. doi:10.1212/WNL.0000000000010454

13. Nelson PT, Dickson DW, Trojanowski JQ, et al. Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. *Brain*. 2019;142(6):1503-1527. doi:10.1093/brain/awz099

14. Wardlaw JM, Smith EE, Biessels GJ, et al. Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *Lancet Neurol*. 2013;12(8):822-838. doi:10.1016/S1474-4422(13)70124-8

15. Wardlaw JM, Allerhand M, Doubal FN, et al. Vascular risk factors, large-artery atheroma, and brain white matter hyperintensities. *Neurology*. 2014;82(15):1331-1338. doi:10.1212/WNL.0000000000000312

16. de Flores R, Wisse LEM, Das SR, et al. Contribution of mixed pathology to medial temporal lobe atrophy in Alzheimer's disease. *Alzheimers Dement J Alzheimers Assoc*. 2020;16(6):843-852. doi:10.1002/alz.12079

17. Agrawal S, Yu L, Leurgans SE, et al. Hippocampal neuronal loss and cognitive decline in LATE-NC and ADNC among community-dwelling older persons. *Alzheimers Dement*. 2025;21(2):e14500. doi:10.1002/alz.14500

18. Yu L, Boyle PA, Dawe RJ, Bennett DA, Arfanakis K, Schneider JA. Contribution of TDP and hippocampal sclerosis to hippocampal volume loss in older-old persons. *Neurology*. 2020;94(2):e142-e152. doi:10.1212/WNL.0000000000008679

19. Wolk DA, Nelson PT, Apostolova L, et al. Clinical criteria for limbic-predominant age-related TDP-43 encephalopathy. *Alzheimers Dement J Alzheimers Assoc*. 2025;21(1):e14202. doi:10.1002/alz.14202

20. Grothe MJ, Moscoso A, Silva-Rodríguez J, et al. Differential diagnosis of amnestic dementia patients based on an FDG-PET signature of autopsy-confirmed LATE-NC. *Alzheimers Dement J Alzheimers Assoc*. 2023;19(4):1234-1244. doi:10.1002/alz.12763

21. Chatterjee M, Özdemir S, Fritz C, et al. Plasma extracellular vesicle tau and TDP-43 as diagnostic biomarkers in FTD and ALS. *Nat Med*. 2024;30(6):1771-1783. doi:10.1038/s41591-024-02937-4

22. Siderowf A, Concha-Marambio L, Lafontant DE, et al. Assessment of heterogeneity among participants in the Parkinson's Progression Markers Initiative cohort using α-synuclein seed amplification: a cross-sectional study. *Lancet Neurol*. 2023;22(5):407-417. doi:10.1016/S1474-4422(23)00109-6

23. Kantarci K, Boeve BF, Przybelski SA, et al. FDG PET metabolic signatures distinguishing prodromal DLB and prodromal AD. *NeuroImage Clin*. 2021;31:102754. doi:10.1016/j.nicl.2021.102754

24. Lyu X, Duong MT, Xie L, et al. Tau-neurodegeneration mismatch reveals vulnerability and resilience to comorbidities in Alzheimer's continuum. *Alzheimers Dement*. 2024;20(3):1586-1600. doi:10.1002/alz.13559

25. Das SR, Lyu X, Duong MT, et al. Tau-Atrophy Variability Reveals Phenotypic Heterogeneity in Alzheimer's Disease. *Ann Neurol*. 2021;90(5):751-762. doi:10.1002/ana.26233

26. Duong MT, Das SR, Lyu X, et al. Dissociation of tau pathology and neuronal hypometabolism within the ATN framework of Alzheimer's disease. *Nat Commun*. 2022;13(1):1. doi:10.1038/s41467-022-28941-1

27. Franzmeier N, Neitzel J, Rubinski A, et al. Functional brain architecture is associated with the rate of tau accumulation in Alzheimer's disease. *Nat Commun*. 2020;11(1):347. doi:10.1038/s41467-019-14159-1

28. Ziontz J, Adams JN, Harrison TM, Baker SL, Jagust WJ. Hippocampal Connectivity with Retrosplenial Cortex is Linked to Neocortical Tau Accumulation and Memory Function. *J Neurosci*. 2021;41(42):8839-8847. doi:10.1523/JNEUROSCI.0990-21.2021

29. Iaccarino L, Tammewar G, Ayakta N, et al. Local and distant relationships between amyloid, tau and neurodegeneration in Alzheimer's Disease. *NeuroImage Clin*. 2018;17:452-464. doi:10.1016/j.nicl.2017.09.016

30. Lee WJ, Brown JA, Kim HR, et al. Regional Aβ-tau interactions promote onset and acceleration of Alzheimer's disease tau spreading. *Neuron*. 2022;110(12):1932-1943.e5. doi:10.1016/j.neuron.2022.03.034

31. Lee J, Burkett BJ, Min HK, et al. Synthesizing images of tau pathology from cross-modal neuroimaging using deep learning. *Brain*. 2023;147(3):980-995. doi:10.1093/brain/awad346

32. Dayarathna S, Islam KT, Uribe S, Yang G, Hayat M, Chen Z. Deep learning based synthesis of MRI, CT and PET: Review and analysis. *Med Image Anal*. 2024;92:103046. doi:10.1016/j.media.2023.103046

33. Yao Z, Wang Z, Zhan Y, et al. Applications of generative models in medical imaging:A review. *Eng Appl Artif Intell*. 2026;181:115349. doi:10.1016/j.engappai.2026.115349

34. Dohmen M, Klemens MA, Baltruschat IM, Truong T, Lenga M. Similarity and quality metrics for MR image-to-image translation. *Sci Rep*. 2025;15(1):3853. doi:10.1038/s41598-025-87358-0

35. Bercea CI, Wiestler B, Rueckert D, Schnabel JA. Evaluating normative representation learning in generative AI for robust anomaly detection in brain imaging. *Nat Commun*. 2025;16(1):1624. doi:10.1038/s41467-025-56321-y

36. Devous MD, Fleisher AS, Pontecorvo MJ, et al. Relationships Between Cognition and Neuropathological Tau in Alzheimer's Disease Assessed by 18F Flortaucipir PET. *J Alzheimers Dis JAD*. 2021;80(3):1091-1104. doi:10.3233/JAD-200808

37. Tustison NJ, Cook PA, Klein A, et al. Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *NeuroImage*. 2014;99:166-179. doi:10.1016/j.neuroimage.2014.05.044

38. Wang H, Suh JW, Das S, Pluta J, Altinay M, Yushkevich P. Regression-Based Label Fusion for Multi-Atlas Segmentation. *Conf Comput Vis Pattern Recognit Workshop IEEE Comput Soc Conf Comput Vis Pattern Recognit Workshop*. Published online June 20, 2011:1113-1120. doi:10.1109/CVPR.2011.5995382

39. Squire LR, Stark CEL, Clark RE. THE MEDIAL TEMPORAL LOBE*. *Annu Rev Neurosci*. 2004;27(Volume 27, 2004):279-306. doi:10.1146/annurev.neuro.27.070203.144130

40. Yushkevich PA, Pluta JB, Wang H, et al. Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Hum Brain Mapp*. 2014;36(1):258-287. doi:10.1002/hbm.22627

41. Wuestefeld A, Pichet Binette A, van Westen D, et al. Medial temporal lobe atrophy patterns in early-versus late-onset amnestic Alzheimer's disease. *Alzheimers Res Ther*. 2024;16(1):204. doi:10.1186/s13195-024-01571-z

42. Sadeghpour N, Lim SA, Wuestefeld A, et al. Developing an anatomically valid segmentation protocol for anterior regions of the medial temporal lobe for neurodegenerative diseases. *bioRxiv*. Published online February 13, 2025:2025.02.11.637506. doi:10.1101/2025.02.11.637506

43. Yushkevich PA, Ittyerah R, Li Y, et al. Morphometry of medial temporal lobe subregions using high-resolution T2-weighted MRI in ADNI3: Why, how, and what's next? *Alzheimers Dement*. 2024;20(11):8113-8128. doi:10.1002/alz.14161

44. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18(2):203-211. doi:10.1038/s41592-020-01008-z

45. Taghvaei M, Cook P, Sadaghiani S, et al. Young versus older subject diffusion magnetic resonance imaging data for virtual white matter lesion tractography. *Hum Brain Mapp*. 2023;44(10):3943-3953. doi:10.1002/hbm.26326

46. Avants BB, Epstein CL, Grossman M, Gee JC. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Med Image Anal*. 2008;12(1):26-41. doi:10.1016/j.media.2007.06.004

47. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*. 2002;17(2):825-841. doi:10.1016/s1053-8119(02)91132-8

48. Brown CA, Das SR, Cousins KAQ, et al. Tau burden is best captured by magnitude and extent: Tau-MaX as a measure of global tau. *Alzheimers Dement*. 2025;21(7):e70346. doi:10.1002/alz.70346

49. Klunk WE, Koeppe RA, Price JC, et al. The Centiloid Project: Standardizing Quantitative Amyloid Plaque Estimation by PET. *Alzheimers Dement J Alzheimers Assoc*. 2015;11(1):1-15.e4. doi:10.1016/j.jalz.2014.07.003

50. Ronneberger O, Fischer P, Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, eds. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*. Springer International Publishing; 2015:234-241. doi:10.1007/978-3-319-24574-4_28

51. Goodfellow IJ, Pouget-Abadie J, Mirza M, et al. Generative Adversarial Networks. *arXiv*. Preprint posted online June 10, 2014:arXiv:1406.2661. doi:10.48550/arXiv.1406.2661

52. Yang L, Zhang Z, Song Y, et al. Diffusion Models: A Comprehensive Survey of Methods and Applications. *arXiv*. Preprint posted online September 27, 2025:arXiv:2209.00796. doi:10.48550/arXiv.2209.00796

53. Cardoso MJ, Li W, Brown R, et al. MONAI: An open-source framework for deep learning in healthcare. *arXiv*. Preprint posted online November 4, 2022:arXiv:2211.02701. doi:10.48550/arXiv.2211.02701

54. Schluchter MD. Mean Square Error. In: *Wiley StatsRef: Statistics Reference Online*. John Wiley & Sons, Ltd; 2014. doi:10.1002/9781118445112.stat05906

55. Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13(4):600-612. doi:10.1109/TIP.2003.819861

56. On Information and Sufficiency on JSTOR. Accessed November 4, 2025. https://www.jstor.org/stable/2236703?seq=1

57. Ward JH. Hierarchical Grouping to Optimize an Objective Function. Published online 1963. doi:10.1080/01621459.1963.10500845

58. Nelson PT, Dickson DW, Trojanowski JQ, et al. Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. *Brain*. 2019;142(6):1503-1527. doi:10.1093/brain/awz099

59. Betthauser TJ, Bilgel M, Koscik RL, et al. Multi-method investigation of factors influencing amyloid onset and impairment in three cohorts. *Brain*. 2022;145(11):4065-4079. doi:10.1093/brain/awac213

60. Brown CA, Cousins KAQ, Korecka M, et al. Temporal Modeling of Amyloid and Tau Trajectories in Alzheimer's Disease Using PET and Plasma Biomarkers. *Ann Neurol*. 2026;99(6):1438-1451. doi:10.1002/ana.78194

61. Lyu X, Mundada NS, Brown CA, et al. 1 Medial temporal lobe Tau- 2 Neurodegeneration mismatch from 3 structural imaging and plasma biomarkers.

62. Nelson PT, Dickson DW, Trojanowski JQ, et al. Limbic-predominant age-related TDP-43 encephalopathy (LATE): consensus working group report. *Brain J Neurol*. 2019;142(6):1503-1527. doi:10.1093/brain/awz099

63. Mohsen S. Alzheimer's disease detection using deep learning and machine learning: a review. *Artif Intell Rev*. 2025;58(9):262. doi:10.1007/s10462-025-11258-y

64. Vega F, Addeh A, Ganesh A, Smith EE, MacDonald ME. Image Translation for Estimating Two-Dimensional Axial Amyloid-Beta PET From Structural MRI. *J Magn Reson Imaging JMRI*. 2024;59(3):1021-1031. doi:10.1002/jmri.29070

65. Karlsson L, Vogel J, Arvidsson I, et al. Machine learning prediction of tau-PET in Alzheimer's disease using plasma, MRI, and clinical data. *Alzheimers Dement*. 2025;21(2):e14600. doi:10.1002/alz.14600

66. Collij LE, Mastenbroek SE, Mattsson-Carlgren N, et al. Lewy body pathology exacerbates brain hypometabolism and cognitive decline in Alzheimer's disease. *Nat Commun*. 2024;15(1):8061. doi:10.1038/s41467-024-52299-1

67. Josephs KA, Martin PR, Weigand SD, et al. Protein contributions to brain atrophy acceleration in Alzheimer's disease and primary age-related tauopathy. *Brain*. 2020;143(11):3463-3476. doi:10.1093/brain/awaa299

68. Brown TM, James SN, Nicholas JM, et al. Association of Increase in White Matter Hyperintensity Volume With Rate of Hippocampal Atrophy in a Population-Based Study of Aging. *Neurology*. 2025;105(5):e213975. doi:10.1212/WNL.0000000000213975

69. Bartrés-Faz D, Arenaza-Urquijo E, Ewers M, et al. Theoretical frameworks and approaches used within the Reserve, Resilience and Protective Factors professional interest area of the Alzheimer's Association International Society to Advance Alzheimer's Research and Treatment. *Alzheimers Dement Amst Neth*. 2020;12(1):e12115. doi:10.1002/dad2.12115

70. Stern Y, Arenaza-Urquijo EM, Bartrés-Faz D, et al. Whitepaper: Defining and investigating cognitive reserve, brain reserve, and brain maintenance. *Alzheimers Dement*. 2020;16(9):1305-1311. doi:https://doi.org/10.1016/j.jalz.2018.07.219

71. Ossenkoppele R, Lyoo CH, Jester-Broms J, et al. Assessment of Demographic, Genetic, and Imaging Variables Associated With Brain Resilience and Cognitive Resilience to Pathological Tau in Patients With Alzheimer Disease. *JAMA Neurol*. 2020;77(5):632. doi:10.1001/jamaneurol.2019.5154

72. Lesuis SL, Hoeijmakers L, Korosi A, et al. Vulnerability and resilience to Alzheimer's disease: early life conditions modulate neuropathology and determine cognitive reserve. *Alzheimers Res Ther*. 2018;10(1):95. doi:10.1186/s13195-018-0422-7

73. Lyu X, Brown CA, Duong MT, et al. Uncovering distinct spatiotemporal trajectories of T-N mismatch subtypes with likely co-pathology in Alzheimer's disease using event-based modeling. *Alzheimers Dement*. 2026;22(5):e71474. doi:10.1002/alz.71474

74. Pascoal TA, Mathotaarachchi S, Shin M, et al. Synergistic interaction between amyloid and tau predicts the progression to dementia. *Alzheimers Dement*. 2017;13(6):644-653. doi:10.1016/j.jalz.2016.11.005

75. Braak H, Braak E. Evolution of the neuropathology of Alzheimer's disease. *Acta Neurol Scand*. 1996;94(S165):3-12. doi:10.1111/j.1600-0404.1996.tb05866.x

76. Wisse LEM, Wuestefeld A, Murray ME, Jagust W, La Joie R. Role of tau versus TDP-43 pathology on medial temporal lobe atrophy in aging and Alzheimer's disease. *Alzheimers Dement*. 2025;21(2):e14582. doi:10.1002/alz.14582

77. Lladó A, Tort-Merino A, Sánchez-Valle R, et al. The hippocampal longitudinal axis—relevance for underlying tau and TDP-43 pathology. *Neurobiol Aging*. 2018;70:1-9. doi:10.1016/j.neurobiolaging.2018.05.035

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AITNmismatchsupplementfinal.docx