Advancing Tau-PET quantification in Alzheimer's disease with machine learning: introducing THETA, a novel tau summary measure

Robel Gebre (gebre.robelkebede@mayo.edu)
Mayo Clinic https://orcid.org/0000-0002-5746-0994

Alexis Moscoso
University of Gothenburg https://orcid.org/0000-0003-0170-036X

Sheela Raghavan
Mayo Clinic

Heather Wiste
Mayo Clinic

Kohl Sparrman
Mayo Clinic

Fiona Heeman
University of Gothenburg

Alejandro Costoya-Sánchez
Universidade de Santiago de Compostela

Christopher Schwarz
Mayo Clinic https://orcid.org/0000-0002-1466-8357

Anthony Spychalla
Mayo Clinic

Val Lowe
Mayo Clinic

Jonathan Graff-Radford
Mayo Clinic

David Knopman
Mayo Clinic https://orcid.org/0000-0002-6544-066X

Ronald Petersen
Mayo Clinic Minnesota https://orcid.org/0000-0002-8178-6601

Michael Schöll
Lund University https://orcid.org/0000-0001-7800-1781

Clifford Jack
Mayo Clinic https://orcid.org/0000-0001-7916-622X

Prashanthi Vemuri

Mayo Clinic https://orcid.org/0000-0003-4286-0589

Article

Keywords:

Posted Date: October 18th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-3290598/v1

Additional Declarations: There is **NO** Competing Interest.

Advancing Tau-PET quantification in Alzheimer's disease with machine learning: introducing THETA, a novel tau summary measure

Robel K. Gebre[1], Alexis Moscoso Rial[2,3], Sheelakumari Raghavan[1], Heather J. Wiste[7], Kohl L Johnson Sparrman[1], Fiona Heeman[2,3], Alejandro Costoya-Sánchez[4,5,6], Christopher G. Schwarz[1], Anthony J. Spychalla[1], Val J. Lowe[1], Jonathan Graff-Radford[8], David S. Knopman[8], Ronald C. Petersen[7,8], Michael Schöll[2,3,6], Clifford R. Jack Jr[1], Prashanthi Vemuri[1], for the Alzheimer's Disease Neuroimaging Initiative*

[1]Department of Radiology, Mayo Clinic, Rochester, MN 55905, USA

[2]Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

[3]Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg, Gothenburg, Sweden

[4]Universidade de Santiago de Compostela, Santiago de Compostela, Spain.

[5]Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED), Instituto de Salud Carlos III, Madrid, Spain.

[6]Nuclear Medicine Department and Molecular Imaging Group, Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Travesía da Choupana s/n, Santiago de Compostela, 15706, Spain.

[7]Department of Qualitative Health Sciences, Mayo Clinic, Rochester, MN, USA

[8]Department of Neurology, Mayo Clinic, Rochester, MN 55905, USA

[9]Dementia Research Centre, Queen Square Institute of Neurology, University College London, UK

*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:

31 http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf

Working title: ML for classifying and quantifying Tau-PET scans

Word count (Without figures, tables, and references): 4786 words

Number of figures and tables: 8 figures and 2 tables.

Number of references: 26

Corresponding Authors: 1. Robel K. Gebre Ph.D.

Phone: +1 507 779 2353, **e-mail:** gebre.robelkebede@mayo.edu

2. Prashanthi Vemuri, Ph.D.

Phone: +1 507 538 0761, **Fax:** +1 507 284 9778,

e-mail: vemuri.prashanthi@mayo.edu

Mayo Clinic and Foundation

200 First Street SW, Rochester, MN 55905

Abstract (150-word limit)

Alzheimer's disease (AD) exhibits spatially heterogeneous 3R/4R tau pathology distributions across participants, making it a challenge to quantify extent of tau deposition. Utilizing Tau-PET from three independent cohorts, we trained and validated a machine learning model to identify visually positive Tau-PET scans from regional SUVR values and developed a novel summary measure, THETA, that accounts for heterogeneity in tau deposition. The model for identification of tau positivity achieved a balanced test accuracy of 95% and accuracy of ≥87% on the validation datasets. THETA captured heterogeneity of tau deposition, had better association with clinical measures, and corresponded better with visual assessments in comparison with the temporal meta-region-of-interest Tau-PET quantification methods. Our novel approach aids in identification of positive Tau-PET scans and provides a quantitative summary measure, THETA, that effectively captures the heterogeneous tau deposition seen in AD. The application of THETA for quantifying Tau-PET in AD exhibits great potential.

1. Introduction

Alzheimer's disease (AD) is characterized by the accumulation of β-amyloid (Aβ) plaques and neurofibrillary tangles (NFTs) in the brain. NFTs are composed of hyperphosphorylated tau proteins and in a majority of individuals tau progresses along predictable patterns, originating in the transentorhinal cortex and spreading to the limbic system and eventually to the neocortex. The spread of tau leads to cognitive impairment and dementia[1]. However, evidence from pathology and imaging have shed light on the heterogeneity of tau deposition in AD, suggesting that there could be distinct patterns of tau accumulation across individuals[2–4].

Current understanding of AD pathophysiology and neurodegeneration suggests that the NFT accumulation is closely correlated with clinical disease progression and precedes clinical symptoms, making tau a promising biomarker for disease diagnosis and clinical trial design[5,6]. Positron emission tomography (PET) imaging is used to visualize and assess tau deposition using radioligands that bind specifically to the paired helical filament of NFTs and can be used to detect and track tau pathology in vivo[7]. Studies using PET have shown in preclinical AD, tau deposition is spread throughout several cortical regions and there follows multiple trajectories[3]. The most common quantification methods for Tau-PET utilize meta-regions of interest (meta-ROIs), such as the temporal meta-ROI, or the more recent medial temporal lobe (MTL) and neocortical (NEO) meta-ROIs to stage disease severity[8,9]. These methods ignore the extent of tau outside these meta-ROIs and average the Tau-PET standardized uptake value ratios (SUVR) in the entire meta-ROI, which underweights any focal depositions of tau in smaller regions within the meta-ROI. In addition to the meta-ROIs, there are less commonly used quantitative methods such as the volumes-of-interest voxel-based multiblock barycentric discriminant analysis (MUBADA)[10] that have also been used to assess the clinical group separation.

The visual rating method followed in this study was based on the density and distribution of tau identified by the radiotracer [^{18}F]flortaucipir (Tauvid™) which was recently FDA-approved for AD tau pathology at B3-level (Braak stages V/VI)[11]. The visual assessment criteria consider the focal deposition of tau through the brain and could overcome the limitations of the meta-ROI methods. In this work we set out to test the hypothesis that a machine learning (ML) model can be developed to identify positive Tau-PET scans based on the clinically accepted multirater visual ratings, and improved quantification methods can be developed to incorporate the heterogeneity in spatial distribution of tau tracer signals throughout the brain. We further hypothesized that these ML-based tau quantification methods could outperform the currently used meta-ROI quantification methods and provide a more accurate and sensitive quantification of tau deposition that would map better to disease severity.

To test our hypotheses, we designed our study with three aims: 1) develop a machine learning model on a large single site dataset using regional SUVR values as inputs and visual ratings as targets and validate the model's performance on two external independent cohorts, 2) compare the performance of our ML model to temporal, MTL and NEO meta-ROI quantitative methods, and 3) develop a novel summary measure that is more sensitive to clinical disease severity by leveraging the regional heterogeneity captured by our ML model. This study aims to address the limitations in the current quantitative methods for tau deposition in AD by utilizing advanced ML approaches.

2. Results

2.1. Characteristics of study population

The study included three independent datasets: Mayo, ADNI, and OASIS-3. The Mayo dataset had 1290 participants with an average age (SD) of 67 (14) years: 55% were male, and 74% were cognitively unimpaired. The ADNI dataset had 831 participants with an average age of 72 (8) years: 48% were male, and 55% were cognitively unimpaired. The OASIS-3 dataset had 430 participants with an average age of 70 (8) years: 43% were male, and 86% were cognitively unimpaired (Table 1). The percentage of visually tau-positive cases in Mayo, ADNI, and OASIS-3 were 19%, 28%, and 14%, respectively (Table 1). The proportion of participants who were classified as tau-positive using both MTL and NEO meta-ROIs were low, highlighting the heterogeneity of the sample (14% for Mayo, 20% for ADNI, and 11% for OASIS-3) (Table 1).

111 **Table 1.** Characteristics summary of study population.

Variables	Mayo	ADNI	OASIS-3
N	1290	831	430
Age, *mean (SD) years*	67 (14)	72 (8)	70 (8)
Males, n (%)	706 (55)	399 (48)	186 (43)
Females, n (%)	584 (45)	432 (52)	244 (57)
Cognitively unimpaired (CU), n (%)	957 (74)	455 (55)	371 (86)
Mild cognitively unimpaired (MCI), n (%)	173 (13)	283 (34)	11 (3)
Alzheimer's disease (AD), n (%)	121 (9)	93 (11)	48 (11)
Dementia with Lewy Bodies (DLB), n (%)	37 (3)	-	-
APOE4+, n (%)	425 (34)	287 (40)	168 (39)
Aβ+, n (%)	512 (40)	335 (43)	133 (32)
T_{V+}, n (%)	245 (19)	230 (28)	61 (14)
T_{MTL}^{+}, n (%)	243 (19)	255 (31)	80 (19)
T_{NEO}^{+}, n (%)	202 (16)	183 (22)	57 (13)
$T_{Temporal}^{+}$, n (%)	476 (37)	418 (50)	159 (37)
T_{MTL}^{+} and T_{NEO}^{+}, n (%)	183 (14)	170 (20)	49 (11)
$T_{Temporal}^{+}$ and T_{MTL}^{+}, n (%)	235 (18)	242 (29)	74 (17)
$T_{Temporal}^{+}$ and T_{NEO}^{+}, n (%)	202 (16)	183 (22)	57 (13)

T_{V+}: Visually tau-positive

$T_{Temporal}^{+}$: Tau-positive in the temporal meta-ROI

T_{MTL}^{+}: Tau-positive in the middle temporo-lateral (MTL)

T_{NEO}^{+}: Tau-positive in the neocortex (NEO)

112

2.2. Model trained on visual ratings for predicting tau positivity

The regional SUVRs were the inputs to the ML model and the visual classifications were the predicted class (Fig. 1). The model was trained on the Mayo dataset and tested on ADNI and OASIS-3. To validate the model, we conducted multiple runs using different data splits (Fig. 2). The models' performance was consistent as indicated by a standard deviation less than 5% for all metrics (Fig. 2). We then selected the best model with the highest f1-score.

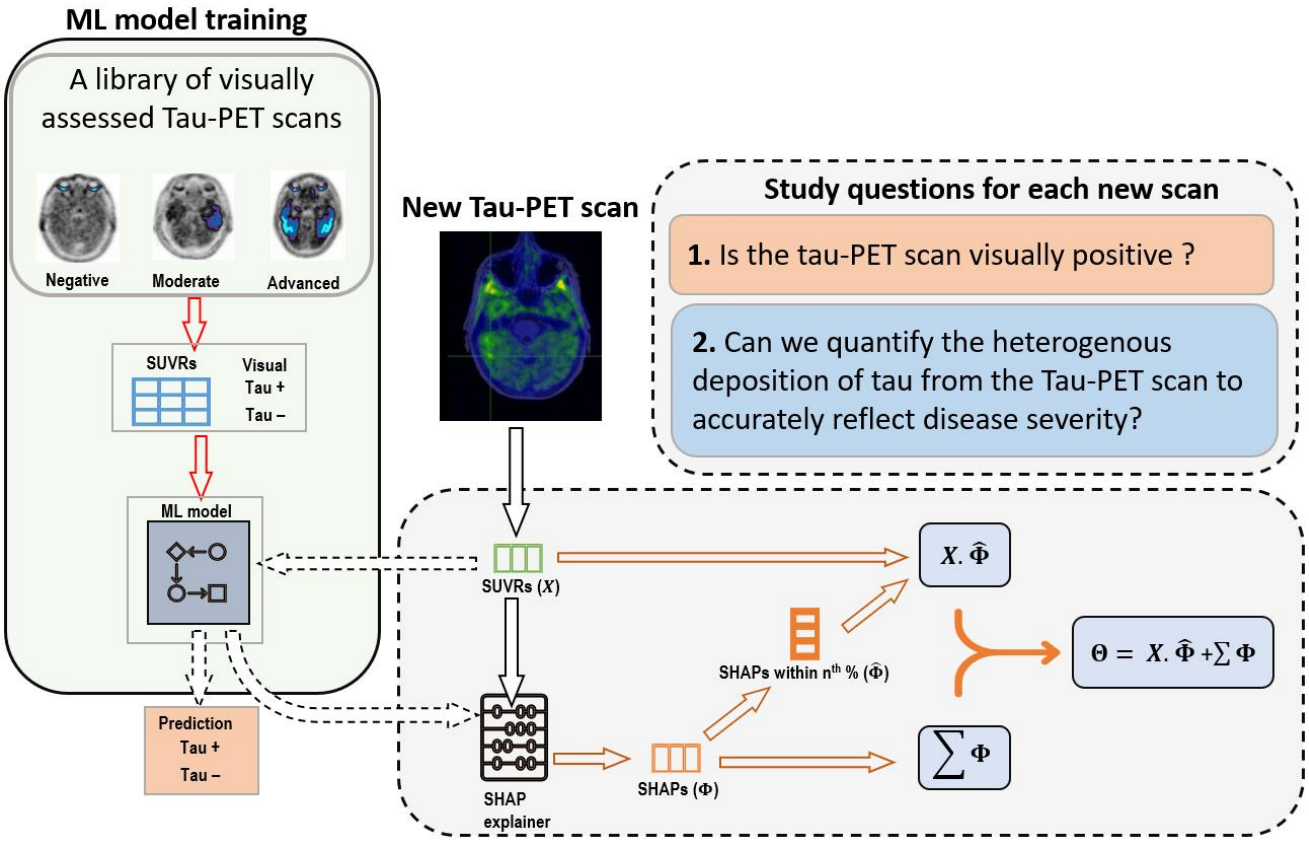

Figure 1. Study design. First, we trained a machine learning (ML) model using a library of visually assessed scans where the visual rating was used as the ground truth and the SUVRs were the inputs. Second, after training the model we applied the SHAP AI explainer to determine each region's contribution to the predicted visual rating. Lastly, we derived a summary measure we are calling tau heterogeneity evaluation in Alzheimer's disease (THETA) score using each participant's SUVR value and corresponding SHAPs.

The best model performed very well in predicting tau status on the Mayo dataset, achieving a balanced accuracy of 98.58% and 95.43% on the Mayo training and testing sets, respectively. When evaluating the model's performance on the external datasets, ADNI and OASIS-3, it achieved a balanced accuracy of 87.74% and 87.03%, respectively. The model identified tau-positive and negative participants with an AUC of 1.00 on the testing set. It also showed very good classification performance on the ADNI external dataset, with an AUC of 0.96. In contrast, the AUC was lower in the OASIS-3 dataset at 0.94 (Fig.2).

Summary of models (200 runs) with random data splits

Datasets	MCC	Bal Accuracy	F1	Precision	Recall
Mayo training† (%)	90.68 (3.71)	94.87 (2.51)	92.20 (3.14)	96.36 (2.53)	88.52 (4.87)
Mayo testing†† (%)	87.68 (3.43)	92.49 (1.99)	90.00 (2.44)	93.88 (3.60)	86.45(4.20)
ADNI testing¥ (%)	78.68 (1.69)	86.25 (1.36)	83.05 (1.60)	94.56 (2.63)	74.18 (3.28)
OASIS-3 testing‡ (%)	73.57 (2.87)	82.48 (1.95)	75.79 (2.64)	88.81 (5.19)	66.40 (4.25)

†n = 1032; ††n = 258; ¥n = 831; ‡n = 430.
MCC = Matthews correlation coefficient, Bal Accuracy = Balanced accuracy

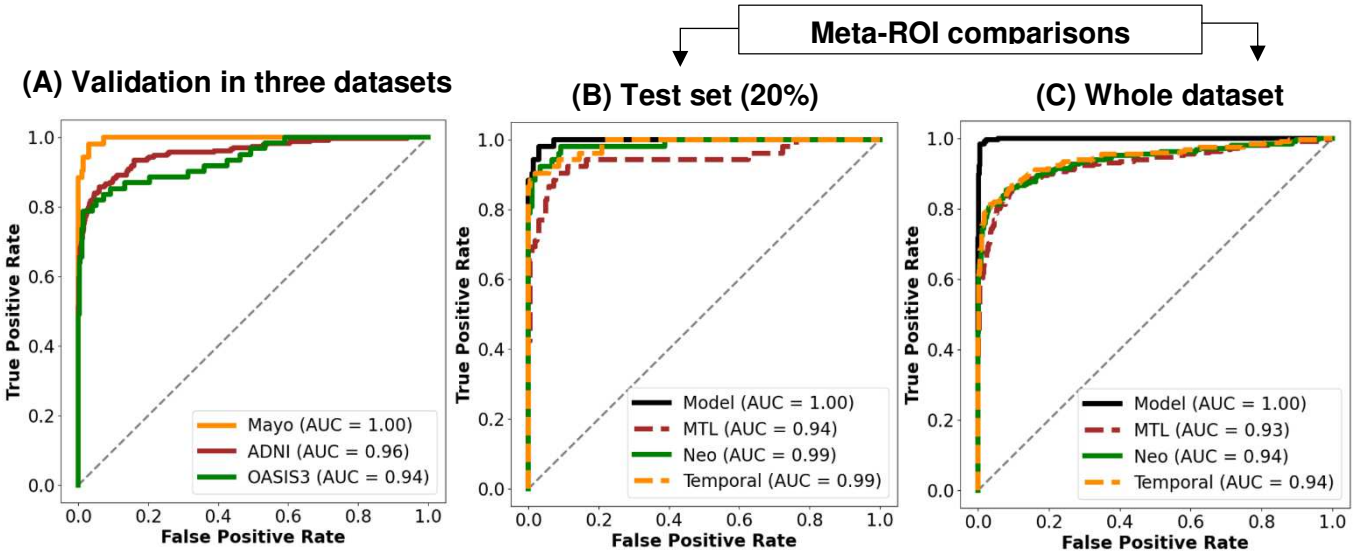

Figure 2. Model performance for binary classification of tau status based on the visual assessment from the three raters. The model was trained on the Mayo and validated on the external validation sets, ADNI and OASIS-3. The top table shows summary of the multiple runs conducted using different random splits of the training (80%) and testing (20%) sets. The metrics in the table show the mean (standard deviation). The receiver operating characteristic's area under the curve (AUC) of the model (A) compares its performance in Mayo, ADNI, and OASIS3, while (B) and (C) illustrate the comparison of the model's performance to meta-ROI classification schemes in the Mayo testing and whole dataset respectively.

126

2.3. Model performance in comparison to meta-ROI-based assessment for prediction of tau positivity

The meta-ROIs showed very similar performances in classifying tau positivity in the Mayo cohort, with an AUC of 0.99 on the test-set (20%) and 0.94 on the whole dataset (Fig. 2B and Fig. 2C). The model outperformed all three meta-ROIs when evaluating classification performance on the Mayo dataset, with a misclassification of 3.67% and 0.48% of tau-positive and negative cases, respectively. On the ADNI dataset, the model misclassified 22.17% of the tau-positive and 2.33% of the tau-negative cases and was largely outperformed by the temporal meta-ROI for tau-positive misclassification at a rate of 6.96% (Table 2). On the OASIS-3 dataset, the model performed best in classifying tau-negative cases with a misclassification rate of 1.36% and had the second-best misclassification rate of 24.59%, outperformed by the temporal meta-ROI at 18.03%. *Supplementary Tables 1 and 2* provide similar analyses for participants with CI and CU clinical diagnosis.

Table 2. Comparison of Meta-ROI-based assessments and the machine learning model predictions to the visual ratings when predicting tau positivity.

Comparisons	TPR (%)	TNR (%)	TP (*n*)	TN (*n*)	1 - TPR (%)	1 - TNR (%)
MAYO[†]						
T_V *vs* $T_{Temporal}$	0.93	0.76	227	796	7.34	23.83
T_V *vs* T_{MTL}	0.78	0.95	191	993	22.04	4.97
T_V *vs* T_{NEO}	0.76	0.98	185	1028	24.49	1.63
T_V *vs* Model	0.96	1.00	236	1040	3.67	0.48
ADNI[¥]						
T_V *vs* $T_{Temporal}$	0.93	0.66	214	397	6.96	33.94
T_V *vs* T_{MTL}	0.78	0.88	180	526	21.74	12.48
T_V *vs* T_{NEO}	0.70	0.96	161	579	30.00	3.66
T_V *vs* Model	0.78	0.98	179	587	22.17	2.33
OASIS-3[‡]						
T_V *vs* $T_{Temporal}$	0.82	0.70	50	260	18.03	29.54
T_V *vs* T_{MTL}	0.66	0.89	40	329	34.43	10.84
T_V *vs* T_{NEO}	0.66	0.95	40	352	34.43	4.61
T_V *vs* Model	0.75	0.98	46	364	24.59	1.36

[†]n = 1290; T_{V+} = 245, T_{V-} = 1045, $T_{Temporal}^{+}$ = 396, T_{MTL}^{+} = 243, T_{NEO}^{+} = 202
[¥]n = 831; T_{V+} = 230, T_{V-} = 301, $T_{Temporal}^{+}$ = 362, T_{MTL}^{+} = 255, T_{NEO}^{+} = 183
[‡]n = 430; T_{V+} = 61, T_{V-} = 369, $T_{Temporal}^{+}$ = 131, T_{MTL}^{+} = 80, T_{NEO}^{+} = 57

2.4. Spatial heterogeneity captured by the machine learning model

To assess the spatial heterogeneity captured by the model, we analyzed the SHAP (SHapley Additive exPlanations)[12] summary plots for tau in the different regions of the brain. In participants with tau positivity in the NEO region, the inferior temporal cortex region was the top predictor (Fig. 3). Conversely, in participants with tau positivity in the MTL region (the region well-known to be affected by tau deposition), the entorhinal cortex region emerged a crucial predictor (Fig. 3).

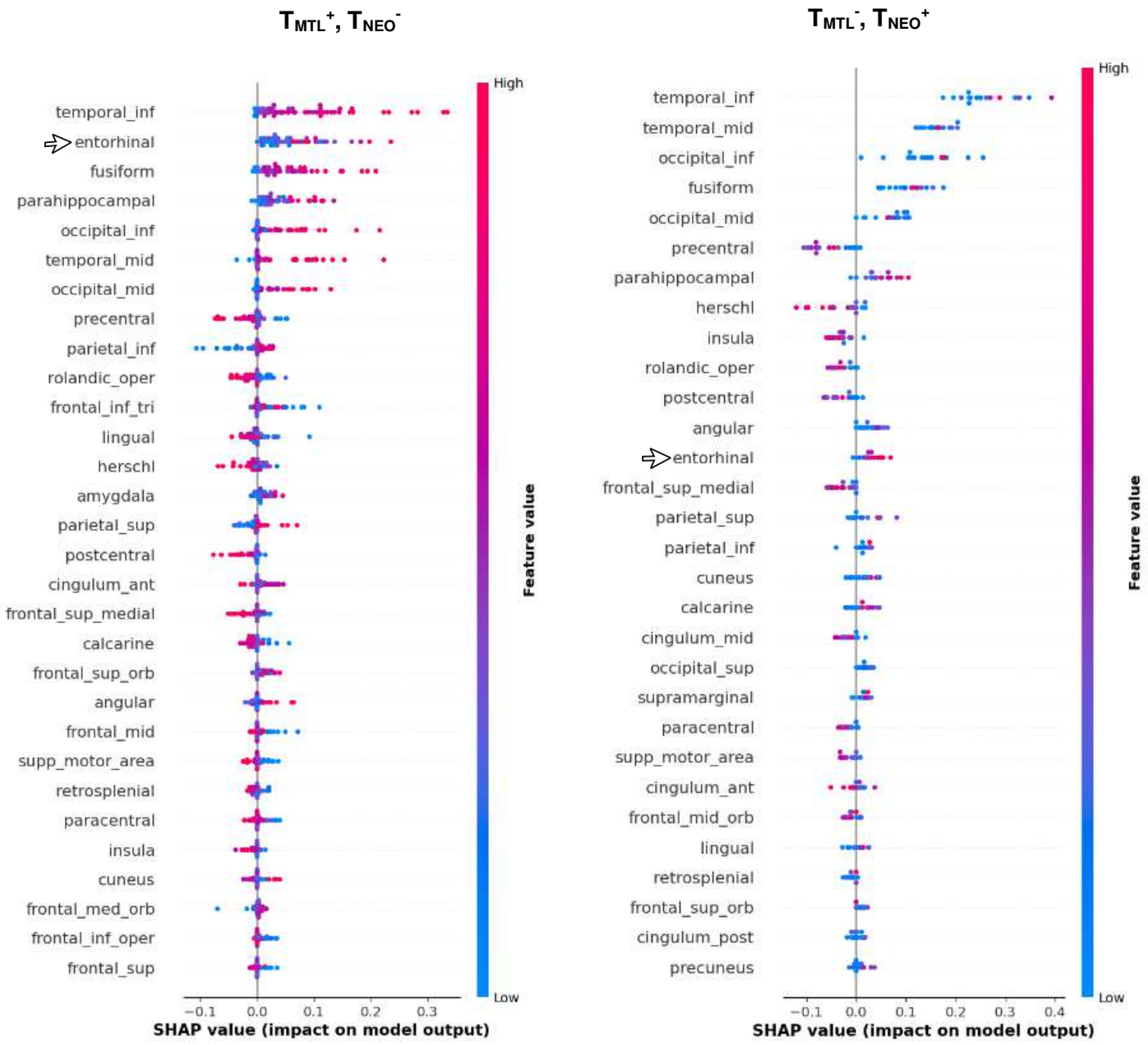

Figure 3. Feature importances for cases where tau was positive in the MTL meta-ROI only and in NEO meta-ROI only. The arrow indicates the importance of the entorhinal region changing its rank depending on the regionality for T_{MTL}^{+}, T_{NEO}^{-} (left) cases, and for T_{MTL}^{-}, T_{NEO}^{+} (right).

2.5. Novel tau summary measure – THETA score

We designed a novel tau global summary measure, THETA score (Tau Heterogeneity Evaluation in Alzheimer's Disease), that considers the spatial heterogeneity of tau deposition throughout the brain.

The THETA score considers the contribution of all the regional tau SUVRs used to the determine a tau-positive or tau-negative scan. Here we illustrate THETA in two sub-populations that highlight tau heterogeneity: discordant and concordant groups. The discordant group consist of cases where there is disagreement between the visual rating and one or more of the meta-ROI classifications while concordant group consists of cases that agree both visual and with the meta-ROIs (Fig. 4).

The THETA score, as described in Equation 2 (section 4.6), was developed to combine different regions based on their contribution to both classification and disease severity, as indicated by the SUVRs. In the tau-positive and meta-ROI negative discordant cases where the model contribution is distributed amongst different regions and not focused specifically on meta-ROI regions, the THETA formulation successfully captures the heterogenous contributions of all the regions, including those with relatively mild signals and similar contributions (Fig. 5A). On the other hand, in tau-positive concordant cases, the hotspot regions that constitute the meta-ROIs are the top predictors in our ML model. In these cases, the THETA formulation maintains the importance of the top regions, thereby preserving the spatial heterogeneity (Fig. 5B).

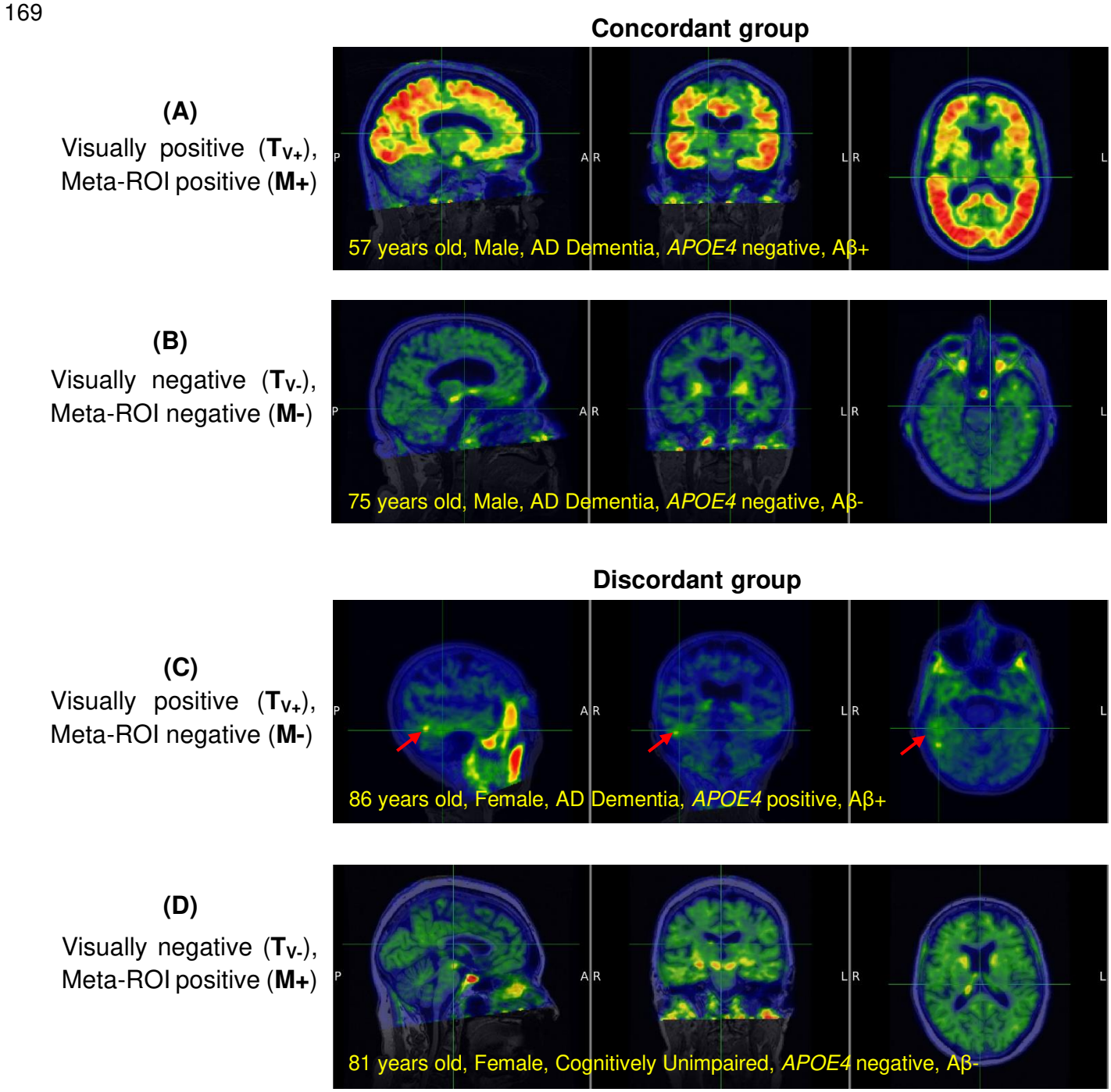

Figure 4. Examples of the concordant groups (A and B) where there is agreement between the visual rating and all three meta-ROIs while the discordant groups (C and D) have disagreement visually and with all three meta-ROIs. While the meta-ROI can miss visually positive scans where the SUVR is lower than the cutoff point (C), the visual assessment does not consider isolated increased activity in the MTL (D). The red arrows indicate where there is increased tracer uptake activity.

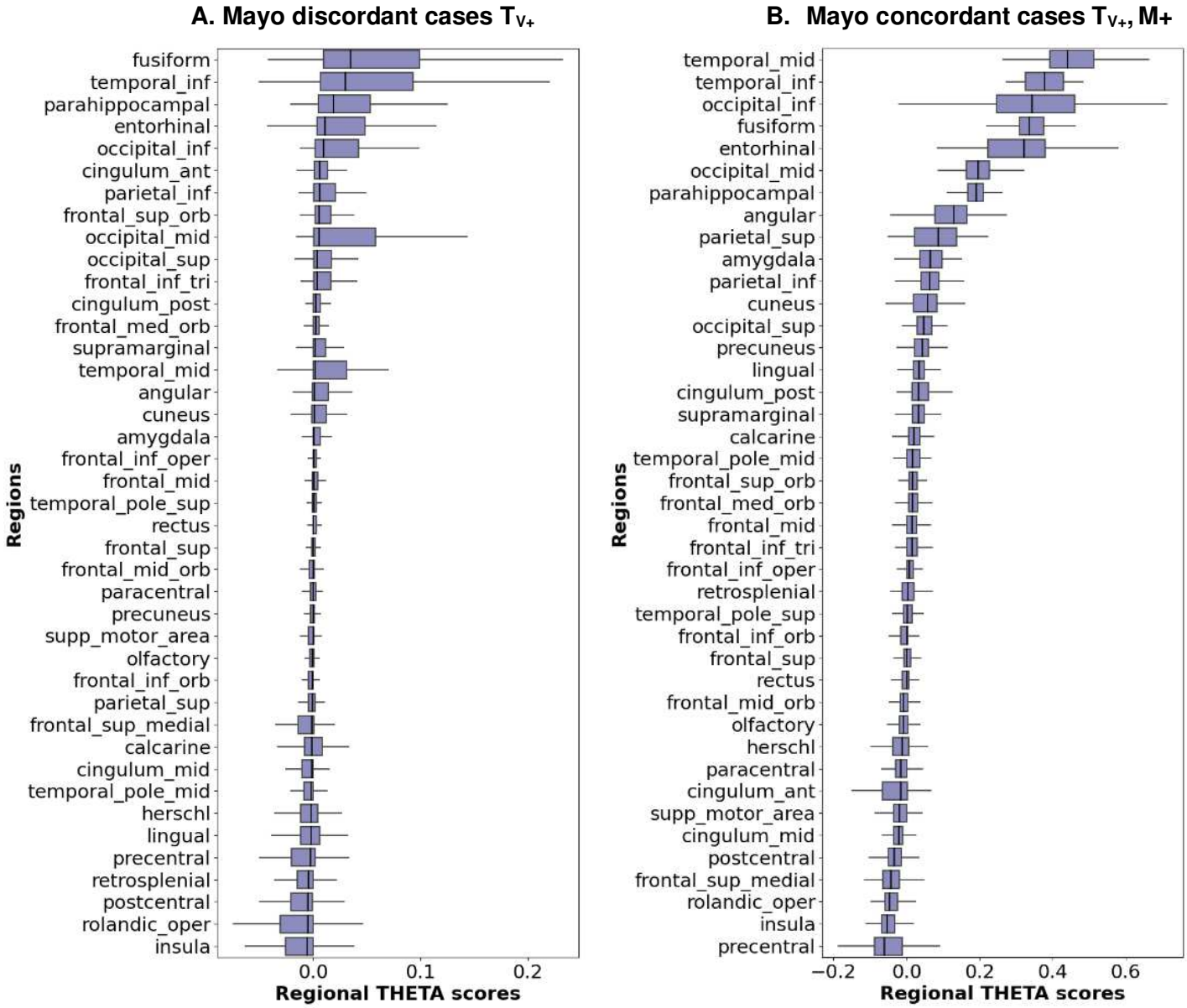

Figure 5. The average regional THETA scores ranked in ascending order by median value for discordant cases (left) and concordant cases (right). The discordant cases which were visually positive (T_{V+}) and negative with one or more meta-ROIs, and the concordant cases which were tau-positive (T_{V+} M+) both visually and all three meta-ROIs.

170

2.5.1. Performance of THETA for assessing disease severity

The performance of the tau summary score THETA for disease severity was assessed using two clinical disease severity measures, Mini-Mental State Examination (MMSE) and CDR sum of boxes (CDR-SB).

When correlation was conducted for all participants from each cohort, the performance of the THETA score and the meta-ROIs was similar (Fig. 6, OASIS-3 shown in *Supplementary Fig. 2*). When looking at the relationship of MMSE to the meta-ROIs and THETA, there was a similar trend of decreasing slope from tau-negative to tau-positive (Fig. 6). However, the THETA score provided a clearer and more distinct separation between tau-positive and negative participants (Fig. 6). This pattern was also observed in the concordant groups (Fig. 7). In contrast, for the discordant groups, THETA demonstrated a negative and significant association with MMSE and a strong positive association to CDR-SB, but the meta-ROIs were not significantly associated with MMSE (Fig. 7). Similar analysis with possible outliers excluded is shown in *Supplementary Figure 3*.

Furthermore, we compared THETA to the temporal meta-ROI for different clinical diagnostic outcomes and calculated the mean differences between tau-positive and tau-negative cases (Fig. 8). We found that for the AD Dementia participants the separation between the tau-positive and tau-negative cases created by both temporal Meta-ROI and THETA were similar in terms of statistical significance across the disease groups. However, for CU and MCI participants there was a clear overlap in tau status for the temporal Meta-ROI, whereas the THETA score showed better separation between tau-positive and tau-negative cases (Fig. 8). For instance, in the ADNI cohort, the difference between the tau-positive and tau-negative temporal Meta-ROI values for CU and MCI participants had an effect size of 3.08 (t-statistics = 16.50, $p < 0.001$) and 2.23 (t-statistics = 16.76, $p < 0.001$), respectively. In contrast, the THETA score showed a much larger

195 effect size of 10.09 (t-statistics = 54.09, $p < 0.001$) and 6.83 (51.36, $p < 0.001$), respectively (Fig.

196 8).

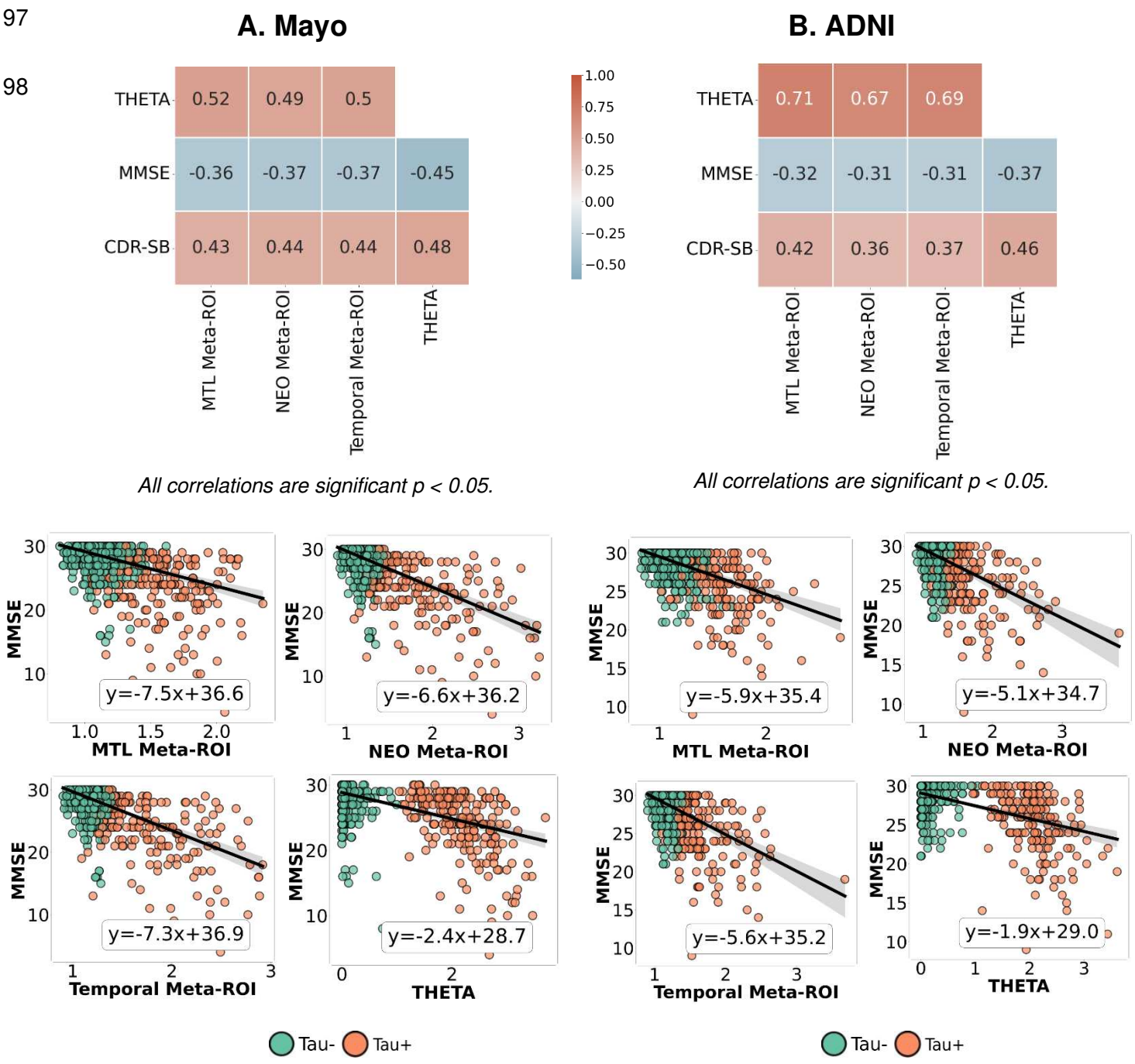

Figure 6. Comparison of the meta-ROIs and THETA score to the clinical measures MMSE and CDR-SB. The correlation coefficients are Spearman's rho and the scatter plot shows the ordinary least squares regression. Similar results for the OASIS-3 cohort are included in *Supplementary Figure 2*. Tau- and Tau+ labels indicate visual assessment status.

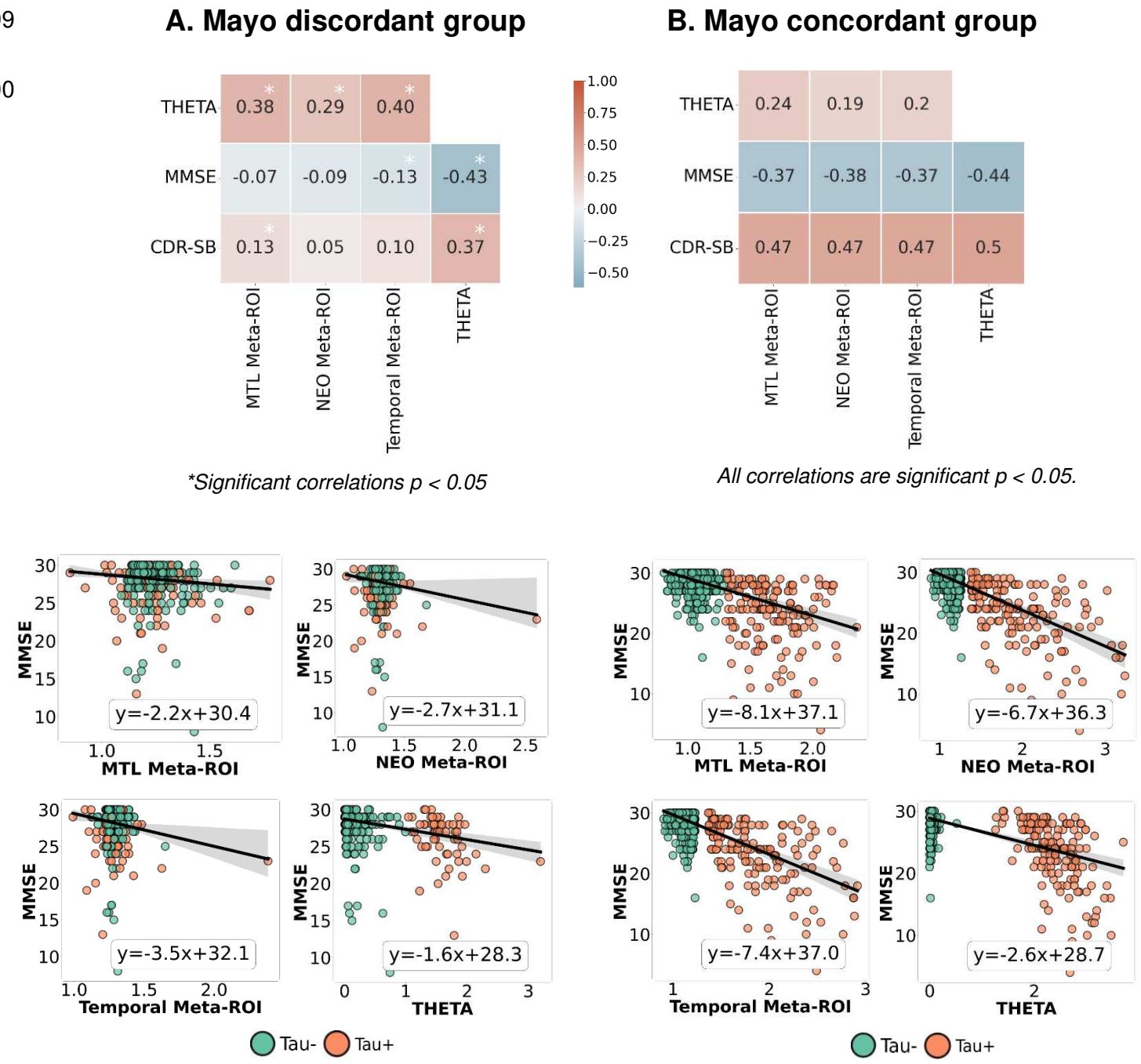

Figure 7. Comparison of the meta-ROIs and THETA to clinical scores MMSE and CDR-SB for the Mayo cohort in the discordant and concordant group. The discordant group consisted of participants with disagreement between the visual rating and one or more meta-ROIs on the tau status, and the concordant group consists of participants whose tau status had agreement between the visual and all three meta-ROI methods. A similar analysis with outliers removed is included in *Supplementary Figure 3*. Tau- and Tau+ labels indicate visual assessment status.

201

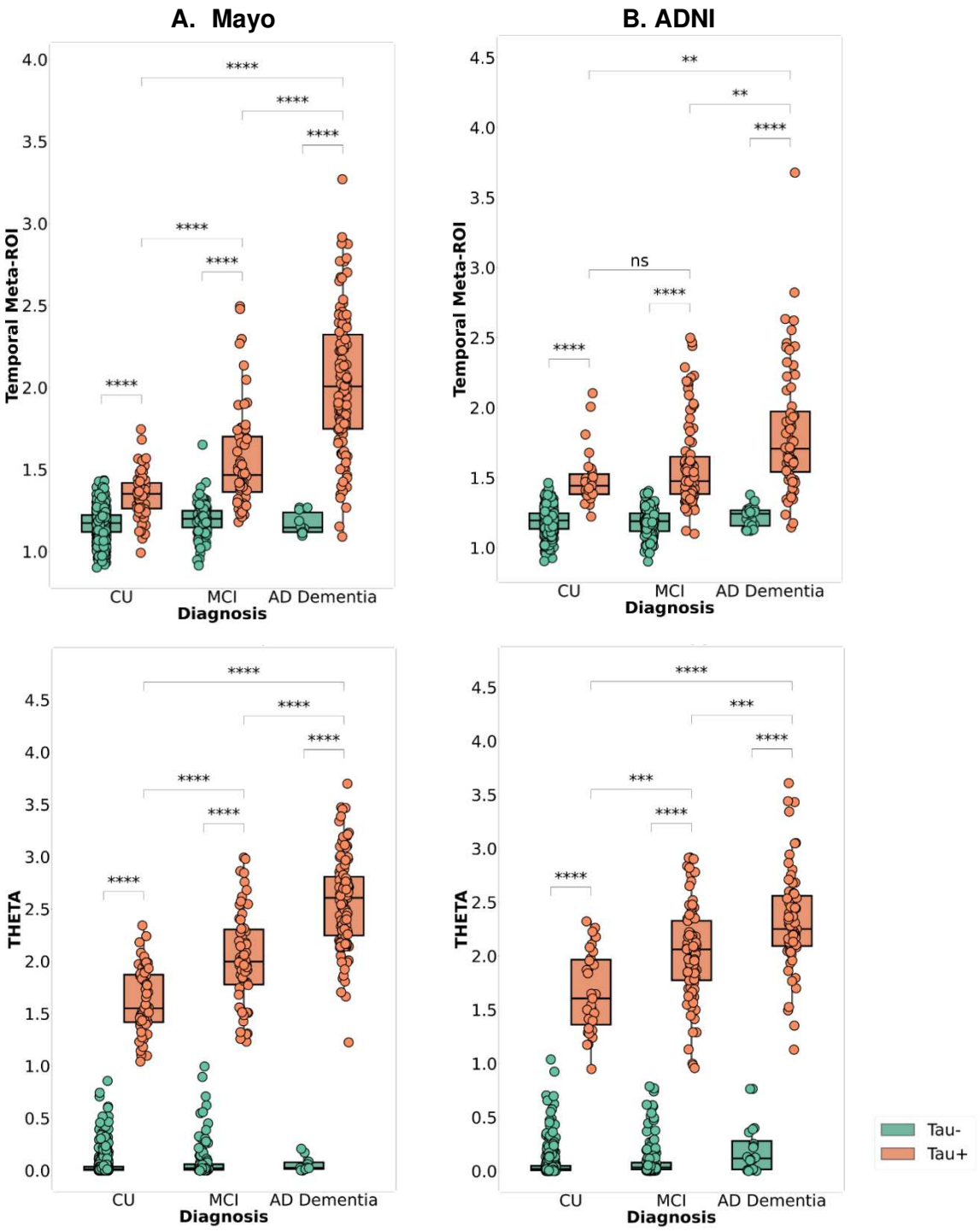

Figure 8. Comparison of the distribution of temporal meta-ROI and THETA in diagnostic groups for visually tau positive and negative participants. Mayo participants and on the right the ADNI participants are shown on the left and ADNI participants on the right. Tau- and Tau+ labels indicate visual assessment status.

ns: p <= 1.00, *: 0.001 < p <= 0.005, **: 0.0001 < p <= 0.001, ***: 0.00001 < p <= 0.0001, ****: p <= 0.00001

3. Discussion

The progression of tau pathology, as captured by Tau-PET scans, has become a key indicator of disease severity in AD. However, current methods have limitations in addressing the heterogeneity of tau deposition. They focus on a limited number of regions with typically high tau uptake while ignoring the spatial variance of tau burden within these regions. These two limitations hamper the performance of meta-ROI-based methods for accurate detection and quantification of the Tau-PET signal. Using visual assessment by three raters as the gold standard in a large single site dataset (Mayo), we developed a ML model to accurately classify the status of Tau-PET scans and validated it in two independent datasets (ADNI and OASIS-3). We then utilized the model to develop a novel tau summary measure that considers tau SUVRs across the brain and provides a metric that maps extremely well to disease progression compared to current methods.

Identification of positive Tau-PET scans

The application of deep learning and ML using Tau-PET has become common in recent years, either to improve PET image acquisition[13], to classify spatial patterns[14,15], to study the association between Aβ and Tau-PET scans[16], or to predict pathological tau accumulation from clinical measures[17,18]. ML-based indices have also been introduced such as Spatial Pattern of Abnormality for Recognition of Early Tauopathy (SPARE-Tau)[19] and Alzheimer's disease resemblance atrophy index (AD-RAI)[20]. SPARE-Tau was trained on tau SUVRs to predict clinical status (CU vs MCI/AD) while AD-RAI was trained on T1-weighted MRI volumetric measures also to predict clinical status and quantify brain atrophy. Nonetheless, our work is the first to develop and validate a ML model to identify positive Tau-PET scans using regional SUVRs from the entire brain. We validated our ML model with entirely independent datasets comprised of different population demographics and data sources. More importantly, our model was able to generalize

to both multicenter and single-center studies, with ADNI being a multicenter study while OASIS-3 is a single-center study.

Multirater visual assessment of Tau-PET is a clinically accepted standard for identifying positive Tau-PET scans as it offers the possibility of assessing tau burden in the entire brain. It can be superior to the meta-ROIs quantitative methods that rely on specific regions to quantify tau burden. While the meta-ROIs focus on the entorhinal cortex and tend to overestimate tau-positive cases, the visual assessment does not consider isolated tau deposition in the medial temporal lobe. The NEO meta-ROI's true negative rate was consistent across all three datasets while the MTL did better at identifying true negatives in Mayo and decreased in performance in ADNI and OASIS-3. On the other hand, in the Mayo cohort, all three meta-ROIs underperformed when identifying tau-positive cases compared to the visual ratings. Nonetheless, because our model was trained on the visual ratings, it showed excellent agreement with the visual ratings in the Mayo cohort.

Quantification of heterogeneity of Tau-PET signal: THETA score

Prior works have shown the spread of tau pathology to be heterogenous and to follow specific patterns across the brain. A histological study by Murray et al. has shown clinical differences between hippocampal sparing and limbic-predominant AD subtypes[21] while a recent event-based computational study by Vogel et al. has shown the presence of posterior and lateral temporal subtypes of atypical AD[2]. While heterogeneity in tau deposition is accepted in the field, there are no measures that consider the heterogeneity in the Tau-PET signal while quantifying it into a summary metric.

In this study, the novel tau summary measure, THETA, considers spatial heterogeneity across the brain, making it a better option for cognitive assessment and clinical diagnosis. Since the ML model accurately classified Tau-PET scans as tau-positive or negative by examining signals

249 throughout the entire brain, we incorporated the THETAi values to formulate our summary
250 measure. This measure was derived using SHAP values, which indicated the importance of each
251 individual region. Thus, by utilizing the heterogeneity captured by the ML model, we were able to
252 ensure that THETA captured pattern-based information. This is illustrated by the regional THETA
253 scores for the concordant or discordant subgroups (Fig. 3 and Fig. 5). Furthermore, since the
254 range of THETA scores were distinct for the tau-positive/negative cases, we were able to get a
255 clear separation between the tau-positive and negative participants for the MMSE clinical score
256 (Fig. 6 and Fig. 7) and the diagnostic groups better than the temporal Meta-ROI (Fig. 8).

257 **THETA score for assessing disease severity**

258 Tau is a proximal surrogate of clinical disease severity and Tau-PET has tremendous potential to
259 significantly impact clinical practice and clinical trials. The FDA approved [18F] flortaucipir PET
260 imaging for detecting NFT B3 corresponding to Braak stages V or IV. Hence, effectively
261 quantifying the Tau-PET signal has important implications because it provides a more accurate
262 and sensitive assessment of disease severity. Given that multirater visual assessment is the
263 clinically accepted standard in the field, developing a highly accurate model using this gold
264 standard and utilizing the model characteristics for quantification of Tau-PET signal has several
265 advantages. This is reflected in the THETA score outperforming the current methods as observed
266 in Figures 6 - 8. Additionally, the THETA scores mapped on to cognitive indices comparably or
267 better than meta-ROI-based methods.

268 THETA can be utilized with ease across multiple clinical studies. The calculation of THETA in a
269 clinical or research setting is similar to the meta-ROI calculation. Once an ML model is trained on
270 the regional SUVRs and is interpreted using the SHAP AI explainer, THETA scores can be
271 generated automatically using our formula. This process can be done for a single participant or a
272 list of participants. The training of a ML model need only be done once and the trained model can
273 be used multiple times, and the training set can constitute cohorts of different demographics as

we have demonstrated in our study. Future work will focus on validating THETA for tracking longitudinal Tau-PET changes.

Strengths and Limitations

This study has some strengths and limitations. We developed a ML model on one dataset and validated it on two independent datasets. There were some limitations in this study. First, the visual assessment of scans is subjective and can be prone to human errors. However, the visual ratings were obtained independently from three raters, and ambiguous discordant cases were reassessed by a Neuroradiologist (CRJ). Second, as expected, the model's performance was lower for the ADNI, and OASIS-3 validation sets due to differences between the cohorts. However, combining the cohorts and training a new model on the combined data solved this problem. The combined model achieved balanced accuracy between 94% and 96%, and ROCAUC greater than 0.99 for all datasets. This is shown in *Supplementary Table 3* and *Supplementary Figure 4*. Third, the THETA score exhibits high sensitivity for a given tau-status which can be strength or a limitation. While visually accurate classification can provide a better range for tau quantification, a visually inaccurate classification (< 1% cases) could force the THETA towards zero. Future studies are planned to validate its performance on longitudinal studies. Lastly, changing of the cut-points for the meta-ROIs than ones used in this study could change the results for the meta-ROI comparisons.

In conclusion, this study aimed to address the limitations of the current quantitative methods for quantifying the spread of tau deposition in Alzheimer's disease by using advanced ML approaches. We also developed a novel summary measure that captures regional heterogeneity, which can be a useful clinical tool for assessing disease progression and subtypes and identifying potential therapeutic targets. Further studies are needed to test the versatility of THETA. The ML model developed in this study performed extremely well in predicting tau status on both the MAYO dataset as well as on the external datasets. The model outperformed the three meta-ROIs in

299 classifying tau positivity on the Mayo test set and was comparable in ADNI and OASIS-3.
300 Additionally, the novel summary measure, THETA, was able to better quantify the spatial
301 heterogeneity of tau deposition and provide a more sensitive measure of clinical disease severity.
302 Overall, the study provides promising results for the use of ML models in improving the detection
303 and quantification of tau pathology in Alzheimer's disease.

4. Methods

4.1. Study participants

We included participants who had undergone a Tau-PET scan with [^{18}F]flortaucipir tracer from three studies: a combined Mayo Clinic Study of Aging (MCSA)[22] and Mayo Alzheimer's Disease Research Center (ADRC) data set (N = 1290, referred to as Mayo), Alzheimer's Disease Neuroimaging Initiative phase 2 or 3 (ADNI) (N = 831), and Open Access Series of Imaging Studies phase 3 OASIS-3 (N = 430)[23]. Individuals with frontotemporal dementia were excluded. The Mayo cohort is a population-based study of cognitive aging among residents of Olmsted County, Minnesota, while the ADRC is a longitudinal research study of individuals recruited from clinical practice, and all participants provided written informed consent. Both studies have been approved by the Mayo Clinic and Olmsted Medical Center Institutional Review Boards. The ADNI cohort initiative was as 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). The ADNI data was obtained from adni.loni.usc.edu database and for up-to-date information, see www.adni-info.org. The OASIS-3 cohort is a longitudinal study through WUSTL Knight ADRC's ongoing projects including cognitively normal adults and individuals at various stages of cognitive decline, with MR and PET scans available. The data was obtained through request at https://www.oasis-brains.org/.

4.2. Image Preprocessing and SUVR measurements

T1-weighted MRI were tissue-class segmented and divided into atlas regions using the MCALT-ADIR122 atlas[24]. Tau-PET scans were rigidly coregistered to corresponding MRI and median

values were taken for each region. Cortical and subcortical regions were referenced to the median of the cerebellar crus to form SUVR units. These regional SUVR values were used both to form the meta-ROIs and as inputs to our machine learning models (see Section 4.5).

4.3. Visual assessment of Tau-PET scans

We followed the FDA-approved official criteria for visual assessment to classify the scans in the study[11,25]. In addition, the visual assessment on Tau-PET scans in all data sets was performed independently by three trained raters. Readers examined the PET images scaled to the average counts in a 2D cerebellum ROI and assigned either a positive (increased neocortical tracer uptake isolated to the posterolateral temporal or occipital or parietal/precuneus regions with or without frontal activity) or negative (no increased neocortical activity or increased neocortical activity isolated to the mesial temporal, anterolateral temporal, and/or frontal regions) AD pattern status using a previously published visual interpretation method[25] (*Supplementary Fig. 5*).

4.4. Tau-PET status using meta-ROIs

The temporal meta-ROI was a voxel-weighted average of median uptake in the entorhinal, amygdala, parahippocampal, fusiform, inferior temporal, and middle temporal regions with the cerebral crus gray median as a reference region[8]. A cutoff point of 1.23 SUVR was used to assess tau positivity for the temporal meta-ROI. The MTL was an unweighted average of medial Tau-PET uptake in bilateral entorhinal cortex and amygdala while the NEO meta-ROI was a voxel-weighted average of bilateral middle temporal and inferior temporal gyri[9]. Meta-ROI values above 1.30 SUVR for MTL and above 1.73 for NEO were considered abnormal.

4.5. Training and interpreting the machine learning model

The inputs to the ML model were 41 cortical region SUVR values calculated as mean of the right and left hemispheres values. The final model was trained on the Mayo dataset (n = 1290) split

into 80% training (n = 1038) and 20% testing (n = 252) and was evaluated on the external datasets ADNI (n = 831) and OASIS-3 (n = 430). To validate the effect of the data splitting on the model performance, we split the data using 200 random seeds and ran the models on the different partitions (*Supplementary Fig. 1*). To account for class (tau-positive vs tau-negative) and group imbalance (discordant vs concordant) we implemented a semi-random iterative stratified data splitting technique (*Supplementary Fig. 6*).

We used a multi-layer stack ensemble machine learning technique with a repeated k-fold bagging to train our model. Repeated k-fold bagging randomly partitions the training data into k folds and then trains k models, each using a different fold as the validation set. This process is repeatedly cross validated with the folds changing each time. The final ensemble model is then created by averaging the predictions of the k models. The Autogluon package was used for this purpose[26]. We preferred this technique due to its robustness and less likelihood of overfitting[26].

In order to interpret the model we used SHAP (SHapley Additive exPlanations)[12]. SHAP is a model-agnostic approach to interpreting model predictions that assigns a value to each feature which indicates how much a feature has contributed to the final prediction[12]. To develop the new metric THETA (section 4.5), we made use of SHAP's Associative property, which states that the individual contributions sum up to the target label. In our binary problem of tau positivity, the SHAP values for each region ranged between -1 and +1 and for each tau-PET scan's regional SUVR values these SHAPs added up to either a 0 (tau-negative) or +1 (tau-positive).

4.6. Developing the novel tau summary measure

We have developed a novel summary measure which we termed as the THETA score (Tau Heterogeneity Evaluation in Alzheimer's Disease). This score is calculated as a linear combination of two components: the model outputs based on the contributions of tau SUVRs across the entire brain, and the weighted contribution of the SUVRs that fall within the 1st and the

99th percentile of SHAP values (Equation 1). The first component captures the overall feature importance by summing the SHAP values ($\sum_{i=1}^{m} \phi_i$) across m number of regions. These SHAP values represent the individual contribution of each brain region (i) to the model's prediction. The second component ($\sum_{i=1}^{m} \hat{\phi}_i x_i$) focuses on the weighted contribution of the brain regions whose SHAP values fall within the percentile range. Across this subset, the SHAP values ($\hat{\phi}_i$) and the actual values of the corresponding SUVRs (x_i) are multiplied to reflect their scaled impact. By combining these two components, the THETA score provides a comprehensive assessment of tau accumulation over the whole brain.

$$\Theta = \sum_{i=1}^{m} \phi_i + \sum_{i=1}^{m} \hat{\phi}_i x_i \quad (1)$$

Where φ_i are SHAP values, $\hat{\varphi}_i$ are the SHAP values within the percentile range, x_i are the corresponding regional SUVRs, and m is the total number of brain regions.

To assess the repeatability of the THETA scores, we calculated the intra-class correlation coefficient (ICC) of the top models. We found the smallest ICC was 0.97 and the largest ICC was 1.00 (*Supplementary Fig. 1*).

4.7. Statistical Analysis

Model performance was evaluated using Mathews correlation coefficient, balanced accuracy, precision, recall, and F1-score. Classification performance of the model and the meta-ROIs was measured on the test-set using Receiver Operating Characteristics Area Under the Curve (ROC AUC). The predicted probabilities and the raw SUVRs were used to plot the ROC AUC curve for the model and meta-ROIs, respectively. To compare the visual assessments to the meta-ROIs or to the ML model's predictions, we used the true positive rate (TPR = TP / (TP + FN)), which is also known as sensitivity, true negative rate (TNR = TN / (TN + FP)), also known as specificity, rate of tau-positive mismatch (1-TPR), and rate of tau negative mismatch (1-TNR). In addition,

we evaluated the performance of THETA on the clinical disease severity measures by calculating correlation using Spearman rho and a linear estimation of slope and intercept using ordinary least squares. Lastly, we evaluated the separation between tau-positive and tau-negative for the different clinical diagnosis groups using Cohen's d for effect size and performed mean comparison using two-tailed independent samples t-test with Bonferroni correction for multiple comparisons.

5. Acknowledgments

We would like to acknowledge Valentina Garibotto, I. Mainta, V. Camacho, M. van Essen, A. Perissinotti, and O. Rodríguez for providing visual ratings of subsets of participants in the ADNI cohorts.

This work was supported by Mayo Clinic Radiology Research fellowship, NIH grants R01 NS097495, R01 AG056366, U01 AG006786, P50 AG016574, RF1AG069052, UF1NS125417, R37 AG011378, R01 AG041851, the GHR Foundation grant, the Alexander Family Alzheimer's Disease Research Professorship of the Mayo Foundation, the Elsie and Marvin Dekelboum Family Foundation, U.S.A. and Opus building NIH grant C06 RR018898. The funding sources were not involved in the manuscript review or approval. This study used the resources of the Rochester Epidemiology Project (REP) medical records-linkage system, which is supported by the National Institute on Aging (NIA; AG 058738), by the Mayo Clinic Research Committee, and by fees paid annually by REP users. We would like to greatly thank AVID Radiopharmaceuticals, Inc., for their support in supplying AV-1451 precursor, chemistry production advice and oversight, and FDA regulatory cross-filing permission and documentation needed for this work.

In this study the part of the analysis was conducted based on the data collection and sharing 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.

Data were provided in part by OASIS [OASIS-3: Longitudinal Multimodal Neuroimaging: Principal Investigators: T. Benzinger, D. Marcus, J. Morris; NIH P30 AG066444, P50 AG00561, P30 NS09857781, P01 AG026276, P01 AG003991, R01 AG043434, UL1 TR000448, R01 EB009352. AV-45 doses were provided by Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly. And OASIS-3_AV1451: Principal Investigators: T. Benzinger, J. Morris; NIH P30 AG066444, AW00006993. AV-1451 doses were provided by Avid Radiopharmaceuticals, a wholly owned subsidiary of Eli Lilly.]

The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

6. References

1. Braak, H., Alafuzoff, I., Arzberger, T., Kretzschmar, H. & Del Tredici, K. Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta Neuropathol* **112**, 389–404 (2006).
2. Vogel, J. W. *et al.* Four distinct trajectories of tau deposition identified in Alzheimer's disease. *Nat Med* **27**, 871–881 (2021).
3. Young, C. B. *et al.* Divergent Cortical Tau Positron Emission Tomography Patterns Among Patients With Preclinical Alzheimer Disease. *JAMA Neurol* **79**, 592 (2022).
4. Mohanty, R., Ferreira, D., Nordberg, A., Westman, E., & for the Alzheimer's Disease Neuroimaging Initiative. Associations between different tau-PET patterns and longitudinal atrophy in the Alzheimer's disease continuum: biological and methodological perspectives from disease heterogeneity. *Alz Res Therapy* **15**, 37 (2023).
5. Iaccarino, L. *et al.* Local and distant relationships between amyloid, tau and neurodegeneration in Alzheimer's Disease. *NeuroImage: Clinical* **17**, 452–464 (2018).
6. La Joie, R. *et al.* Prospective longitudinal atrophy in Alzheimer's disease correlates with the intensity and topography of baseline tau-PET. *Sci. Transl. Med.* **12**, eaau5732 (2020).
7. Okamura, N. *et al.* Advances in the development of tau PET radiotracers and their clinical applications. *Ageing Research Reviews* **30**, 107–113 (2016).
8. Jack, C. R. *et al.* Defining imaging biomarker cut points for brain aging and Alzheimer's disease. *Alzheimer's & Dementia* **13**, 205–216 (2017).
9. Ossenkoppele, R. *et al.* Amyloid and tau PET-positive cognitively unimpaired individuals are at high risk for future cognitive decline. *Nat Med* **28**, 2381–2387 (2022).
10. Abdi, H. *et al.* Analysis of Regional Cerebral Blood Flow Data to Discriminate among Alzheimer's Disease, Frontotemporal Dementia, and Elderly Controls: A Multi-Block Barycentric Discriminant Analysis (MUBADA) Methodology. *JAD* **31**, S189–S201 (2012).

11. Jie, C., Treyer, V., Schibli, R. & Mu, L. Tauvid™: The First FDA-Approved PET Tracer for Imaging Tau Pathology in Alzheimer's Disease. *Pharmaceuticals* **14**, 110 (2021).
12. Lundberg, S. M. & Lee, S.-I. A Unified Approach to Interpreting Model Predictions. in.
13. Duffy, I. R., Boyle, A. J. & Vasdev, N. Improving PET Imaging Acquisition and Analysis With Machine Learning: A Narrative Review With Focus on Alzheimer's Disease and Oncology. *Mol Imaging* **18**, 153601211986907 (2019).
14. Jo, T., Nho, K., Risacher, S. L., Saykin, A. J. & for the Alzheimer's Neuroimaging Initiative. Deep learning detection of informative features in tau PET for Alzheimer's disease classification. *BMC Bioinformatics* **21**, 496 (2020).
15. Zou, J. *et al.* Deep learning improves utility of tau PET in the study of Alzheimer's disease. *Alz & Dem Diag Ass & Dis Mo* **13**, (2021).
16. Ruwanpathirana, G. P. *et al.* Mapping the association between tau-PET and Aβ-amyloid-PET using deep learning. *Sci Rep* **12**, 14797 (2022).
17. Giorgio, J. *et al.* A robust and interpretable machine learning approach using multimodal biological data to predict future pathological tau accumulation. *Nat Commun* **13**, 1887 (2022).
18. Kim, J. *et al.* Prediction of tau accumulation in prodromal Alzheimer's disease using an ensemble machine learning approach. *Sci Rep* **11**, 5706 (2021).
19. Toledo, J. B. *et al.* SPARE-Tau: A flortaucipir machine-learning derived early predictor of cognitive decline. *PLoS ONE* **17**, e0276392 (2022).
20. Zhao, L. *et al.* Risk estimation before progression to mild cognitive impairment and Alzheimer's disease: an AD resemblance atrophy index. *Aging* **11**, 6217–6236 (2019).
21. Murray, M. E. *et al.* Neuropathologically defined subtypes of Alzheimer's disease with distinct clinical characteristics: a retrospective study. *The Lancet Neurology* **10**, 785–796 (2011).
22. Roberts, R. O. *et al.* The Mayo Clinic Study of Aging: Design and Sampling, Participation, Baseline Measures and Sample Characteristics. *Neuroepidemiology* **30**, 58–69 (2008).

23. LaMontagne, P. J. *et al. OASIS-3: Longitudinal Neuroimaging, Clinical, and Cognitive Dataset for Normal Aging and Alzheimer Disease*. http://medrxiv.org/lookup/doi/10.1101/2019.12.13.19014902 (2019) doi:10.1101/2019.12.13.19014902.

24. Schwarz, C. G. *et al.* A large-scale comparison of cortical thickness and volume methods for measuring Alzheimer's disease severity. *NeuroImage: Clinical* **11**, 802–812 (2016).

25. Fleisher, A. S. *et al.* Positron Emission Tomography Imaging With [^{18} F]flortaucipir and Postmortem Assessment of Alzheimer Disease Neuropathologic Changes. *JAMA Neurol* **77**, 829 (2020).

26. Erickson, N. *et al.* AutoGluon-Tabular: Robust and Accurate AutoML for Structured Data. Preprint at http://arxiv.org/abs/2003.06505 (2020).

Supplementary Files

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

- SupplementaryTauMLGebre2023.docx