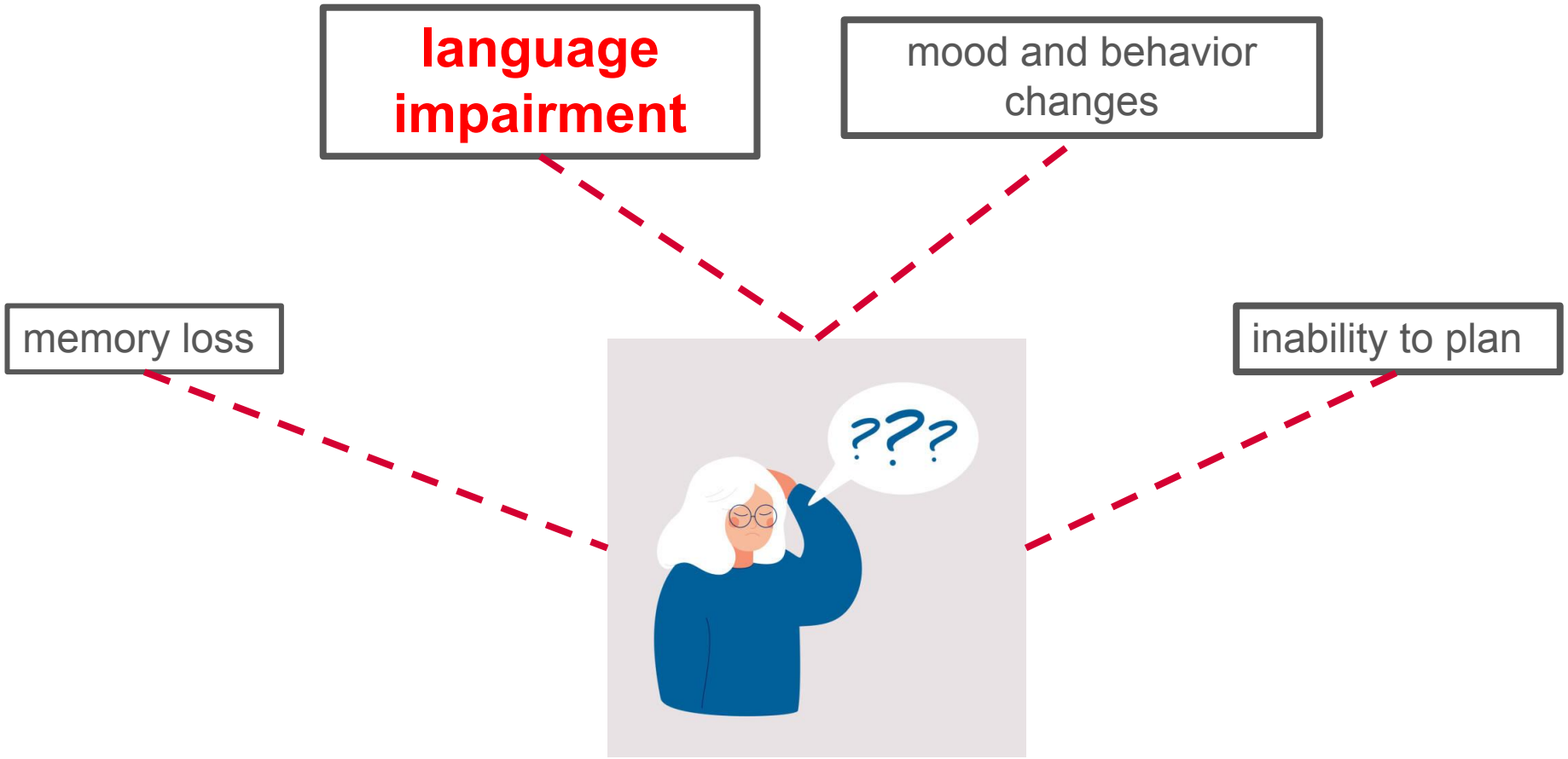
SLaCAD: A Spoken Language Corpus for Early Alzheimer's Disease Detection

Presenter:
Shahla Farzana, PhD
UIC-NLP Laboratory
Department of Computer Science, University of Illinois Chicago
1 May, 2024

UNIVERSITY OF ILLINOIS CHICAGO

Dementia and Cognitive Impairment



Early Screening of Dementia

Preclinical AD: The stage in which clinical symptoms have not yet appeared but there is evidence of AD pathogenesis via biomarkers.[1]

Goal: Explore relationships between **lexico-syntactic and acoustic features** and **neural biomarkers**

- **Can we recognize language differences after biomarkers of dementia appear, but before behavior change is observed in cognitive tests?**

[1]Marco Canevelli, Nawal Adali, Cécile Tainturier, Giuseppe Bruno, Matteo Cesari, and Bruno Vellas. 2013. Cognitive interventions targeting subjective cognitive complaints. American Journal of Alzheimer's Disease & Other Dementias®, 28(6):560–567. PMID: 23823142.

SLAcAD: A Spoken Language Corpus for Early Alzheimer's Disease Detection

SLaCAD

INV: yes but i'm gonna ask you few more questions. okay alright. can you describe when you became the leader of the dining hall please?
PAR: can i describe what?
INV: that specific day when
PAR: oh that specific day, let's see, it was my it was uh it was in my second year so that would be nineteen forty two in in this in the in the fall of forty two i would say september i can't give you the specific day but i i…

Corpus	Class	# Participants	# Transcripts	Time (Minutes)
SLaCAD	Dementia	3	3	2.99±0.98
	MCI	6	6	3.74±1.24
	Control	82	82	5.23±2.13

Preclinical AD variables

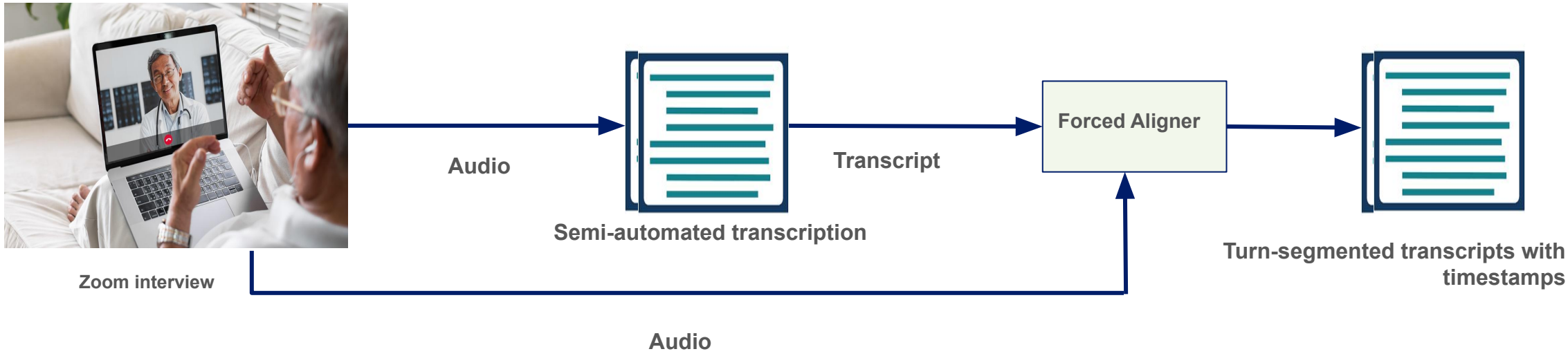
CSF based biomarker	CSF based biomarker	Blood based biomarker
$tTau/A\beta_{42}$ Positivity: **$tTau/A\beta_{42} \geq 0.609$**[1]	**$A\beta_{42}/A\beta_{40}$** Positivity: **$A\beta_{42}/A\beta_{40} \leq 0.056$**[1]	**Plasma Tau ($pTau_{181}$)** Positivity : **$pTau_{181}$** concentration **≥4.09 pg/ml**

Chappelle, S.D., Gigliotti, C., Leger, G.C., Peavy, G.M., Jacobs, D.M., Banks, S.J., Little, E.A., Galasko, D.R. and Salmon, D.P. (2022), Comparison of the Telephone-Montreal Cognitive Assessment (T-MoCA) and Telephone Interview for Cognitive Status (TICS) as Remote Screening Tests for Early Alzheimer's Disease.. Alzheimer's Dement., 18: e065917.

Preclinical AD variables

	$tTau/A\beta_{42}$		$A\beta_{42}/A\beta_{40}$		$pTau_{181}$	
	Negative n=51	**Positive n=12**	**Negative n=44**	**Positive n=19**	**Negative n=50**	**Positive *n*=27**
Age	74.22± 4.63	77.66± 7.13	74.16± 4.74	76.53± 6.00	74.46± 4.63	78.44± 5.89
Education (in year)	17.56± 2.05	17.42± 2.84	17.52 ± 2.14	17.58± 2.28	17.10± 2.37	17.93± 1.69
Gender (F/M)	30/20	7/20	19/25	10/9	23/28	6/6

Data Collection and Transcription



Content-Free Features

- POS tags
- CFG features
- Syntactic complexity features
- NER tags
- Vocabulary richness features
- SUBTL scores
- Consecutive utterance similarity scores
- Acoustic features

Modeling and Feature Selection

Models:

- RandomForest
- XGBoost

Feature selection:

- ANOVA F-values
- mutual information (MI) values
- RF-based frequency

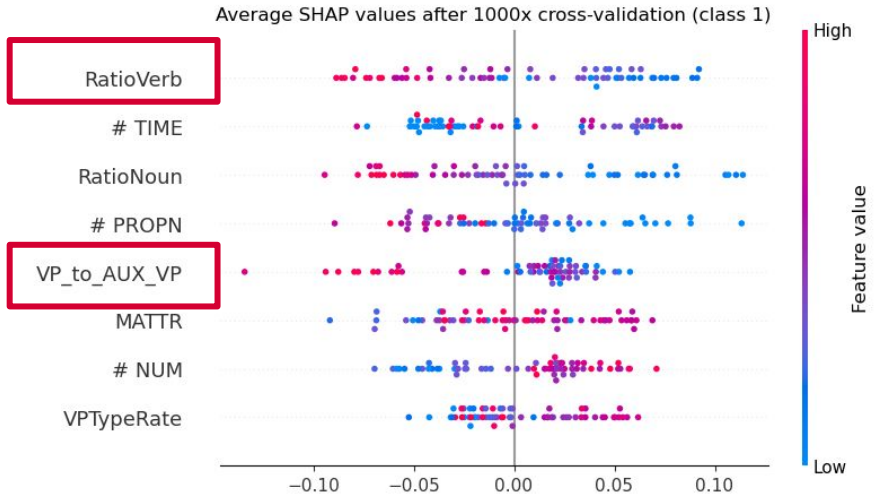
Result

- Results averaged across *1000 stratified 5-fold cross-validation runs* of Random Forest (RF)
- Results are maximized using the top eight features determined via RF-based feature selection

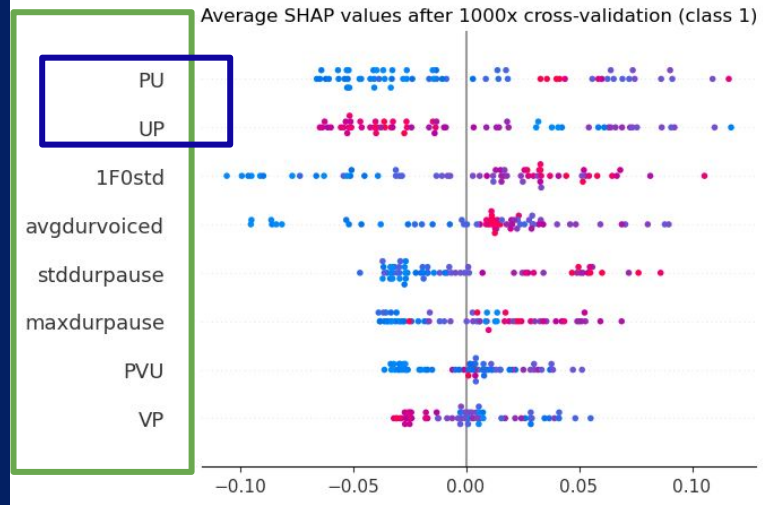
Target Variable	#Feature	Acc.	ROC-AUC
CSF $tTau/A\beta_{42}$	Verbal fluency (baseline)	0.80	0.70
	2	0.67	0.59
	4	0.71	0.61
	8	**0.84**	**0.72**
	16	0.83	0.71
CSF $A\beta_{42}/A\beta_{40}$	Verbal fluency (baseline)	0.72	0.68
	2	0.71	0.68
	4	0.72	0.68
	8	**0.75**	**0.70**
	16	0.71	0.63
Plasma $pTau_{181}$	Verbal fluency (baseline)	0.54	0.43
	2	0.66	0.64
	4	0.66	0.65
	8	0.73	**0.71**
	16	**0.74**	0.70

Feature Analysis

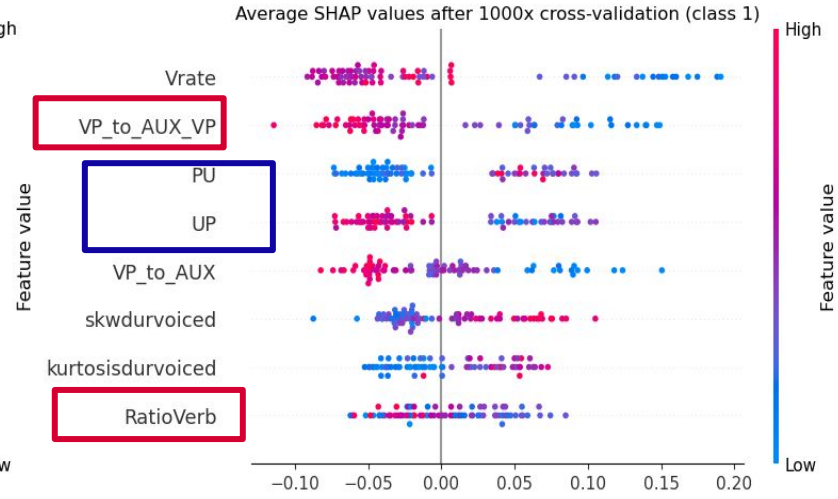
- For tTau/Aβ_{42}, the most predictive features are **acoustic**
- **Lexica**l and **context-free grammar** features are commonly selected across the Aβ_{42}/Aβ_{40} and pTau
- pTau and tTau/Aβ_{42} are positively associated with the **acoustic** feature pause ratio



CSF Aβ_{42}/Aβ_{40} target variable with RF features.



CSF tTau/Aβ_{42} target variable with RF features.



Plasma pTau$_{181}$ target variable with RF features.

Demographic Variables Analysis

tTau/Aβ$_{42}$ (r^2):-0.7751

Aβ$_{42}$/Aβ$_{40}$ (r^2):-0.5057

pTau$_{181}$ (r^2): -0.5863

tTau/Aβ$_{42}$ (r^2):-0.4719

Aβ$_{42}$/Aβ$_{40}$ (r^2):-0.5700

pTau$_{181}$ (r^2): -0.5421

Top feature values of the respective target variables are not significantly different across different genders

Takeaway

Shahla Farzana, Edoardo Stoppa, Alex Leow, Tamar Gollan, Raeanne Moore, David Salmon, Douglas Galasko, Erin Sundermann, Natalie Parde. **SLaCAD: A Spoken Language Corpus for Early Alzheimer's Disease Detection.** To appear in the Proceedings the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation, May 20-25, Torino, Italy, 2024.

- **SLaCAD**, a novel spoken language corpus
  - Wide range of **clinical biomarkers**
  - Cognitive status from **preclinical AD** to **AD pathology**
- Findings are encouraging; positive results motivate future exploration
- Speech and language biomarkers emerged
  - Sensitive to both AD pathology and **preclinical AD!**

Thank You

Acknowledgements

- Creation of SLaCAD was funded in part by a seed grant from the University of California San Diego's Alzheimer's Disease Research Center.
- Authors Shahla Farzana and Natalie Parde were partially funded by the NSF under Grant No. 2125411. Any opinions, findings, and conclusions or recommendations are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Authors

Shahla Farzana: sfarza3@uic.edu

Natalie Parde: parde@uic.edu

edoardo Stoppa: estopp2@uic.edu

Alex Leow: weihliao@uic.edu

Tamar Gollan: tgollan@health.ucsd.edu

Raeanne Moore: r6moore@health.ucsd.edu

David Salmon: dsalmon@health.ucsd.edu

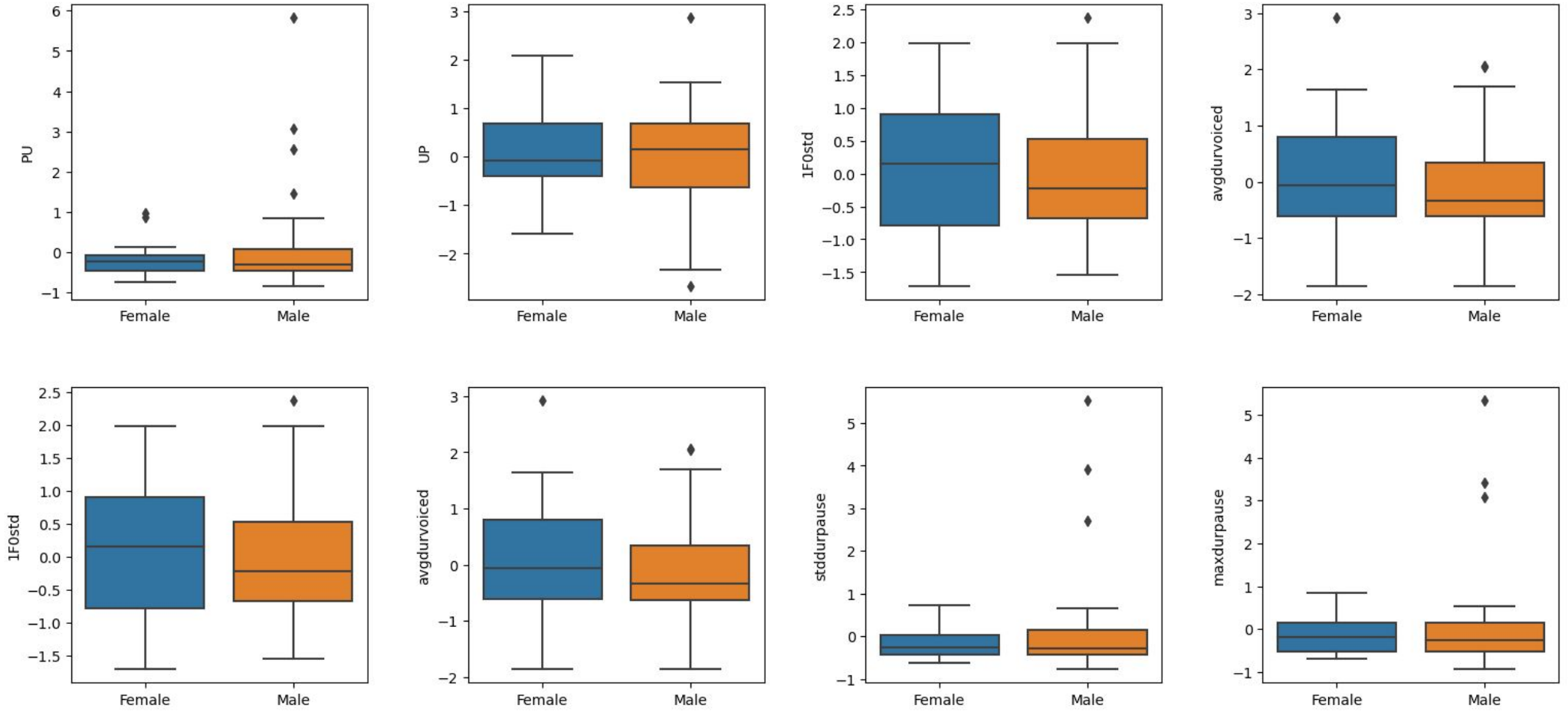
Douglas Galasko: dgalasko@health.ucsd.edu

Erin Sundermann: esundermann@health.ucsd.edu

Web

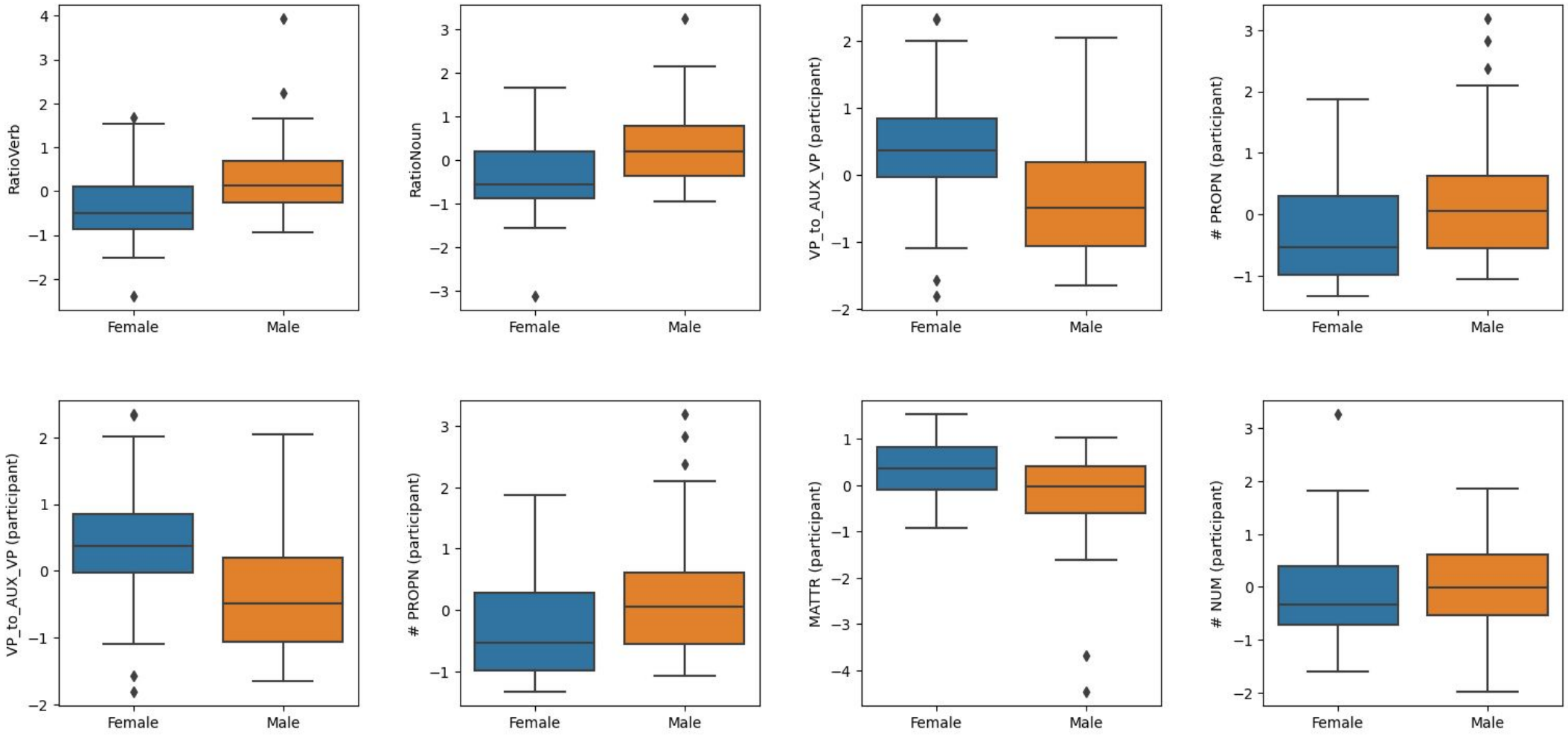
https://nlp.lab.uic.edu

Backup Slides



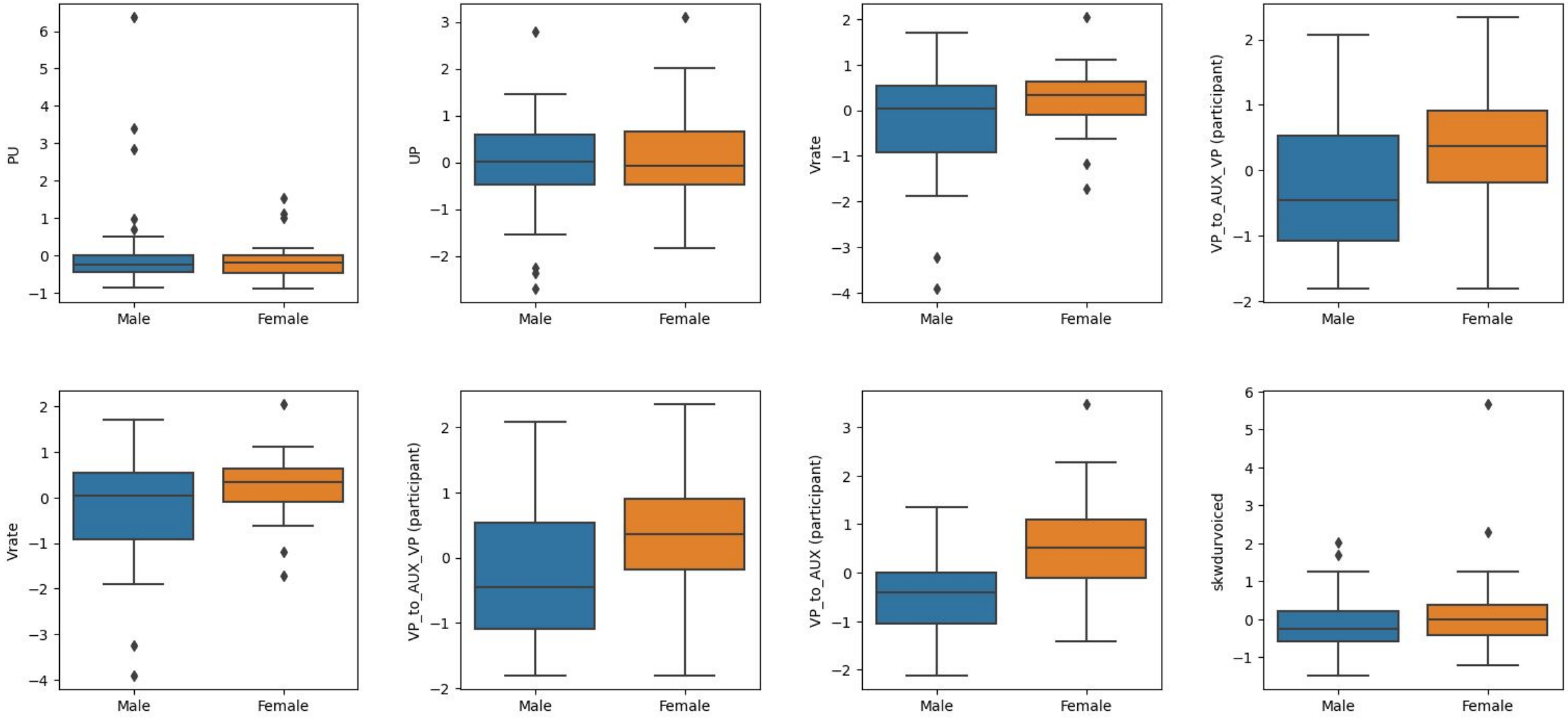
Association of top 8 selected features of **tTau/Aβ_{42}** variable with *sex*.

.



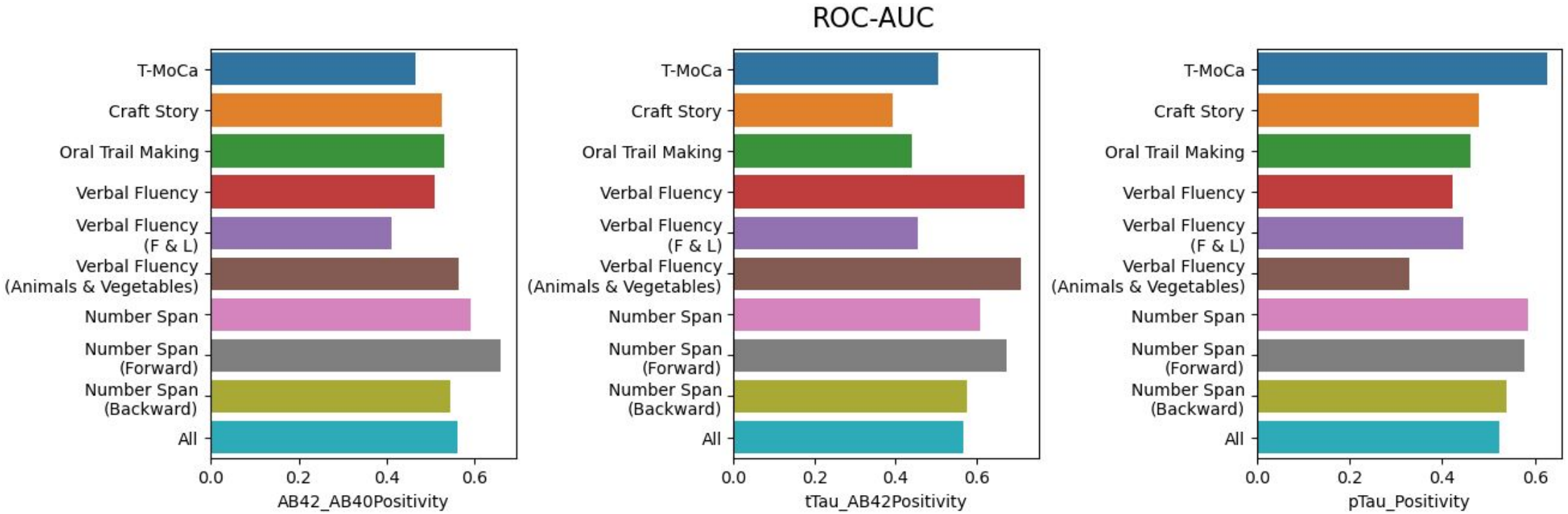
Association of top 8 selected features of **Aβ42/Aβ40** variable with *sex*.

.



Association of top 8 selected features of **pTau181** variable with *sex.*

.



ROC-AUC metric of the standard cognitive tests in predicting the target variables related to preclinical AD.