Subject Harmonization of Digital Biomarkers: Improved Detection of Mild Cognitive Impairment from Language Markers

Bao Hoang,Yijiang Pang, Hiroko Dodge, ***Jiayu Zhou***
Illidan Lab @ Michigan State University

illidan
Intelligent Data Analytics Lab

Applications	Biomarkers	Disease Progression	Drug Discovery
Methodology	Transfer Learning Multi-task, Few-Shot,Adaptation	Multi-Modality Fusion	Robust Learning against Missing and Noisy Data
Infrastructure	Privacy Protection, Federated Learning (Data Heterogeneity, Resource Heterogeneity, Availability), Fairness, Distributed Optimization		

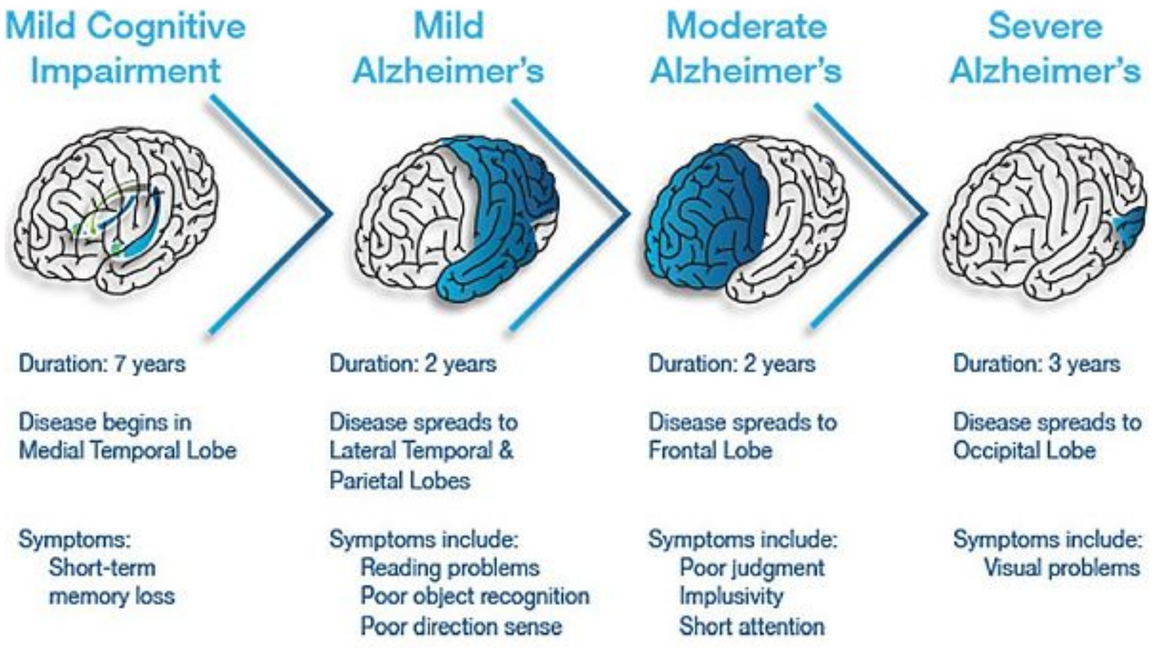
Alzheimer’s disease is officially listed as the seventh leading cause of death in the United States in 2020 and 2021 with 32.4% and 31.0% death rates, respectively.

S. L. Murphy, K. D. Kochanek, J. Xu and E. Arias, Mortality in the united states, 2020 (2021)
https://www.cdc.gov/nchs/data/databriefs/db427.pdf

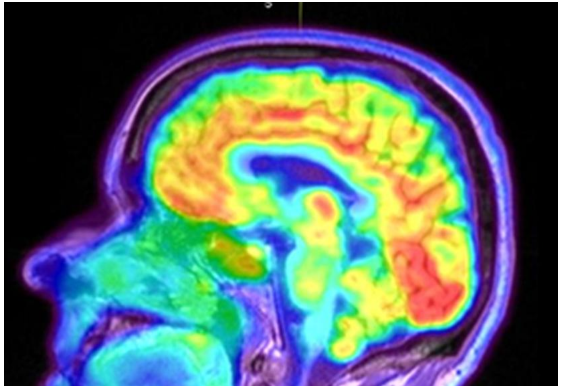
Neurodegenerative Disease - Alzheimer's

Early Stage

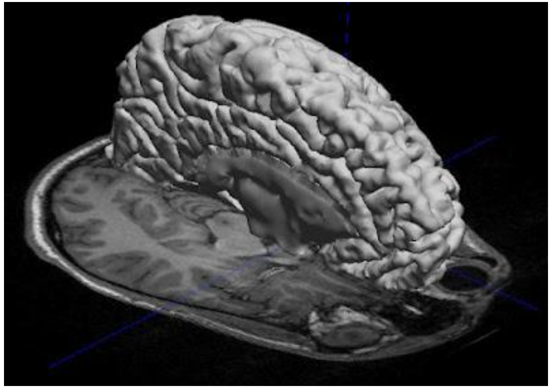
MCI



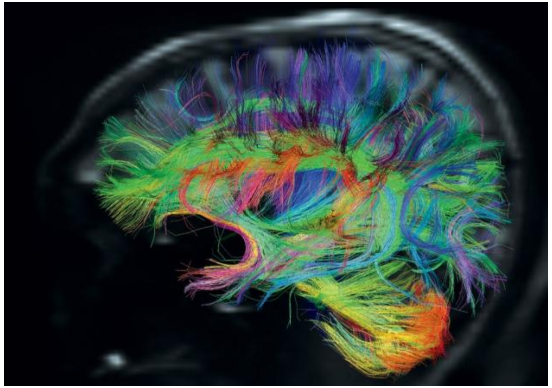
Early Diagnosis and Biomarkers



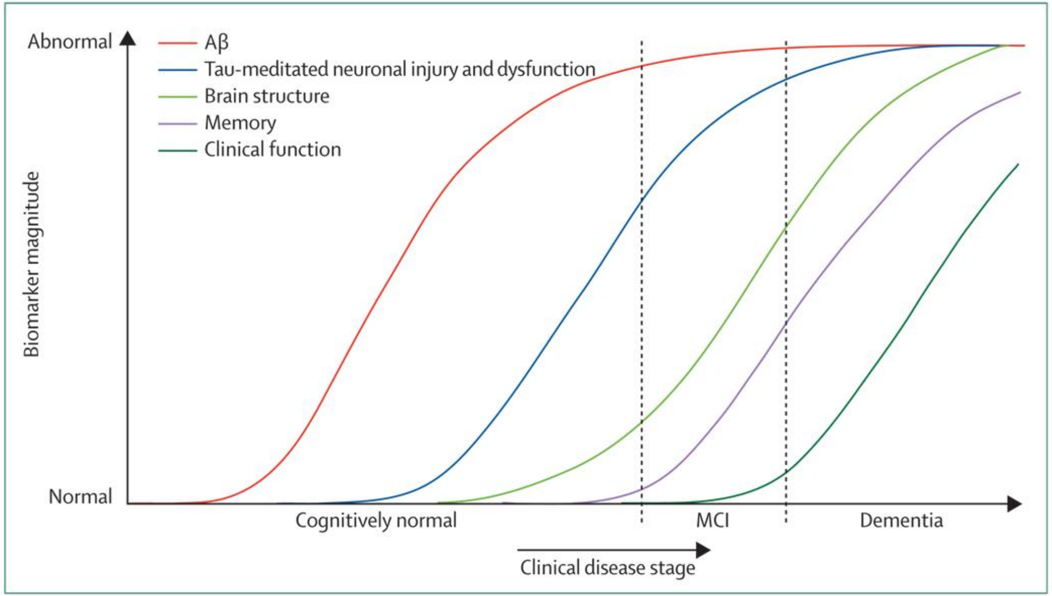
Huang, Shuai, Jing Li, Liang Sun, Jieping Ye, Adam Fleisher, Teresa Wu, Kewei Chen, Eric Reiman, and Alzheimer's Disease NeuroImaging Initiative. "*Learning brain connectivity of Alzheimer's disease by sparse inverse covariance estimation*." NeuroImage 50, no. 3 (2010): 935-949.



Zhou, Jiayu, Jun Liu, Vaibhav A. Narayan, Jieping Ye, and Alzheimer's Disease Neuroimaging Initiative. "*Modeling disease progression via multi-task learning*." NeuroImage 78 (2013): 233-248.



Wang, Qi, Liang Zhan, Paul M. Thompson, Hiroko H. Dodge, and Jiayu Zhou. "*Discriminative fusion of multiple brain networks for early mild cognitive impairment detection*." In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pp. 568-572. IEEE, 2016.



Biomarkers are effective, but it is already too late when brain markers are obtained from a patient.

Jack Jr, Clifford R., David S. Knopman, William J. Jagust, Leslie M. Shaw, Paul S. Aisen, Michael W. Weiner, Ronald C. Petersen, and John Q. Trojanowski. "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade." *The Lancet Neurology* 9, no. 1 (2010): 119-128.

Language Markers

- An early detection approach of MCI that is **affordable and accessible**.
- Extract language markers from conversations to build predictive models
- Semantic, Syntactic, and Lexical features are used for language markers

"... i see a german shepherd deluge with two children..."

Subject's Interview Transcription

Extract Linguistic Feature

Semantic

Syntactic

Lexical

Concatenate Semantic, Syntactic, and Lexical

Language Marker of Conversation

> Interspeech. 2021 Aug-Sep:2021:3830-3834. doi: 10.21437/interspeech.2021-2002.

Automatic Detection of Alzheimer's Disease Using Spontaneous Speech Only

Jun Chen [1], Jieping Ye [1], Fengyi Tang [2], Jiayu Zhou [2]

Affiliations + expand

PMID: 35493062 PMCID: PMC9056005 DOI: 10.21437/inte

Article | Open access | Published: 31 March 2020

Scalable diagnostic screening of mild cognitive impairment using AI dialogue agent

Fengyi Tang, Ikechukwu Uchendu, Fei Wang, Hiroko H. Dodge & Jiayu Zhou

Scientific Reports **10**, Article number: 5732 (2020) | Cite this article

Front. Digit. Health, 11 February 2022
Sec. Connected Health
Volume 3 - 2021 |
https://doi.org/10.3389/fdgth.2021.702772

This article is part of the R
Bridging the Gap: Advancing C
Digital World
View all 11 Articles >

The Joint Effects of Acoustic and Linguistic Markers for Early Identification of Mild Cognitive Impairment

Fengyi Tang[1] Jun Chen[2] Hiroko H. Dodge[3] Jiayu Zhou[1]

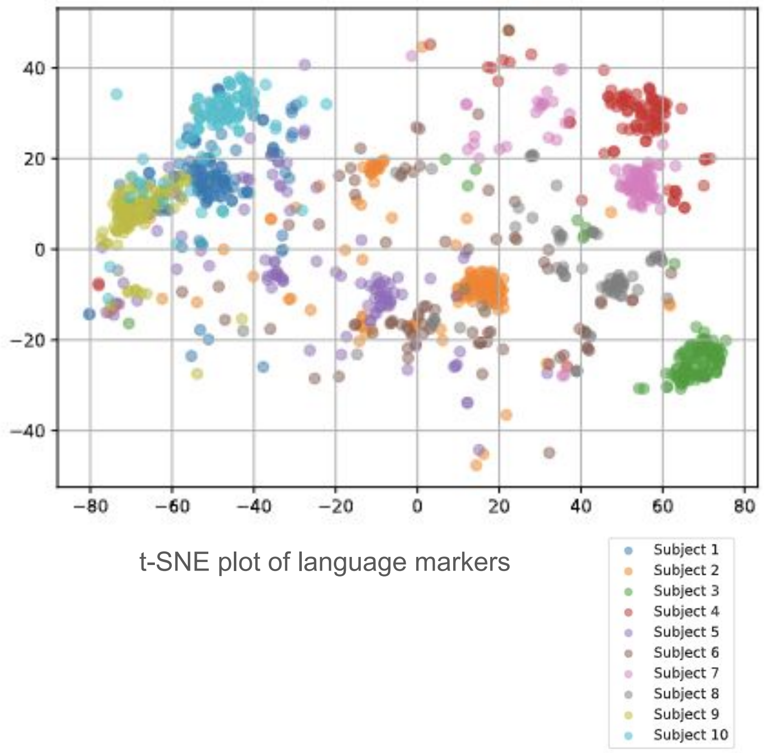
Pac Symp Biocomput. 2023; 28: 7–18.

Detection of Mild Cognitive Impairment from Language Markers with Crossmodal Augmentation

Guangliang Liu,[1] Zhiyu Xue,[1] Liang Zhan,[2] Hiroko H. Dodge,[3] and Jiayu Zhou[1,*]

Problems of Language Markers (any digital ones)

- The data points are **not** independent and identically distributed (i.i.d.)
- One subject will have multiple conversations, and these conversations' language markers form a **cluster**
  - The way people speak can be drastically different
- Such differences are much **more outstanding** than subtle differences characterizing cognitive capability (MCI or NL)



t-SNE plot of language markers

Challenges from Non-IID/Subject Bias

- A classifier may naturally pick up "**easier**" features during learning
  - Easier features are ones that separate subjects rather than MCI/NL
  - Performance degradation because they are not useful features
- A **larger dataset** with more subjects may be helpful …
  - But not available :(
- Harmonization cannot be used due to **unseen subjects** in the testing

Study Design

Study Type :	Interventional (Clinical Trial)
Actual Enrollment :	186 participants
Allocation:	Randomized
Intervention Model:	Parallel Assignment
Masking:	Single (Outcomes Assessor)
Masking Description:	Study assessors will be blinded to the subject study arm assignment.
Primary Purpose:	Prevention
Official Title:	Internet-based Conversational Engagement Clinical Trial
Actual Study Start Date :	June 1, 2018
Actual Primary Completion Date :	August 31, 2021
Actual Study Completion Date :	August 31, 2021

Web-enabled social interaction to delay cognitive decline among seniors with MCI:

Project Number	Former Number	Contact PI/Project Leader	Awardee Or
1R01AG056102-01A1	1R01AG056113-01A1	DODGE, HIROKO HAYAMA	OREGON HE SCIENCE UI

Can we harmonize language markers to remove subject bias?

so as to further improve cognitive predictive power

Subject Harmonization

- Key Idea
  - Harmonized features should **not** be able to **differentiate** subjects under classifiers
  - Harmonized features should be **close** to the original feature as possible
- Approach
  - We train a feature harmonization network $f_{FH}(.)$: $x \longrightarrow \bar{x}$ with parameter θ_{FH} where x is original feature and $\bar{x}$ is harmonized feature
  - The objective function is:

Subject classifier

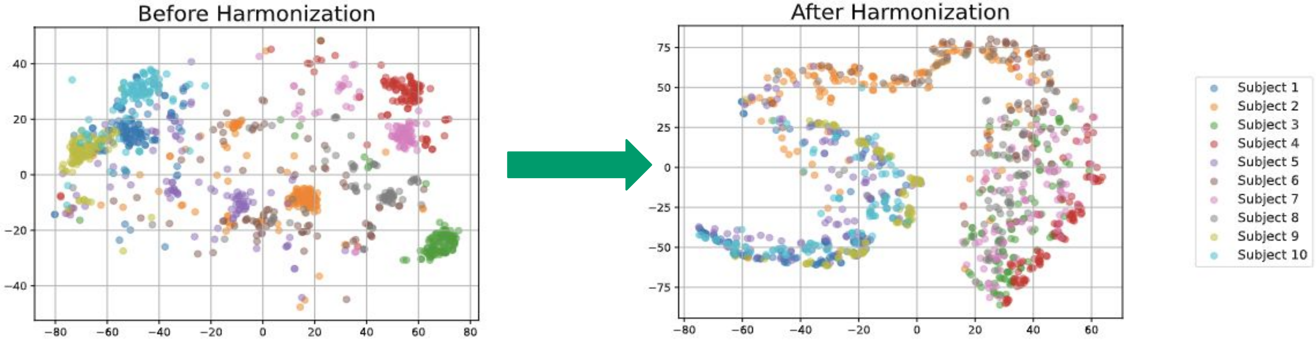
Subject label

$$\min_{\theta_{FH},\theta_s} \frac{1}{M} \sum_{i=1}^{M} -\ell_{ent}(f_s \circ f_{\mathrm{FH}}(\mathbf{x}_i), \mathbf{y}_i^s) + \ell_{mse}(f_{\mathrm{FH}}(\mathbf{x}_i), \mathbf{x}_i),$$

Encourage the harmonized features **cannot** be **differentiate** by subject identities

Encourage the **similarity** between harmonized features and original features

Key Result for Subject Harmonization



t-SNE plot of language markers

Subject clusters are successfully **destroyed** by harmonization

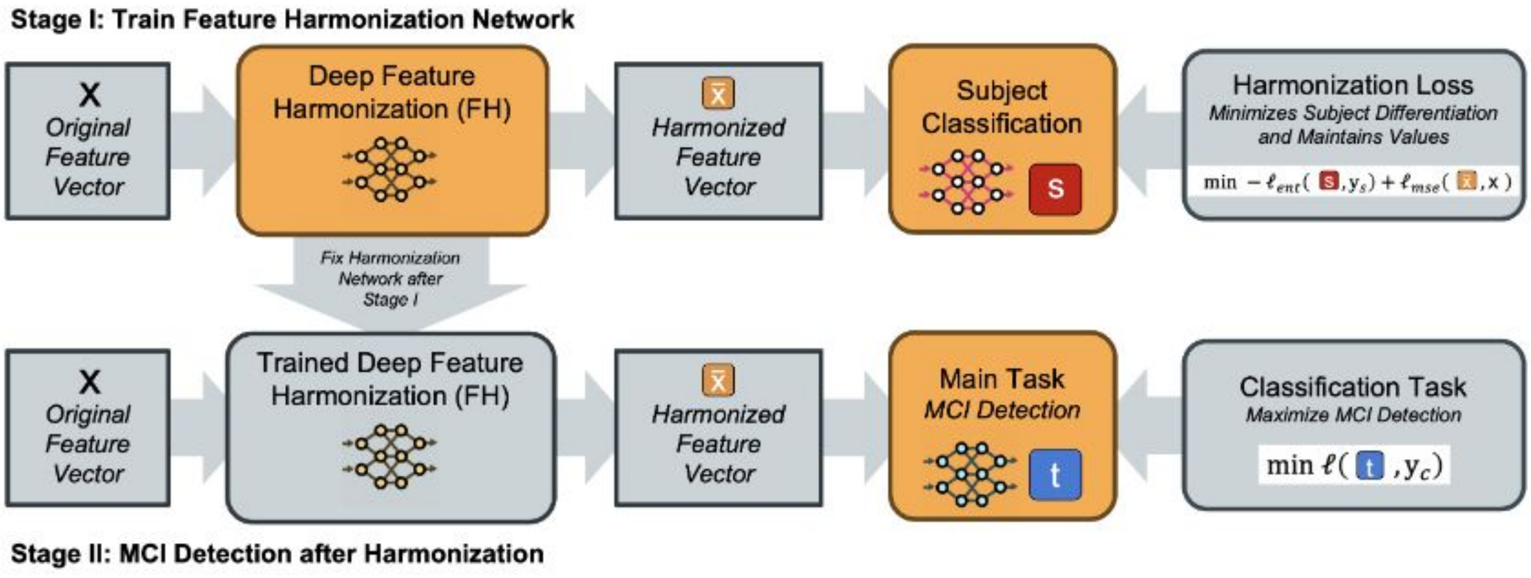
Classifier	Before harmonization	After harmonization
Logistic Regression	0.921±0.007	0.221±0.012
Multi-layer Perceptron	0.905±0.007	0.219±0.038

Subject Classification Task Accuracy

Classifier cannot predict **subject identity** of conversations after harmonization

Subject Harmonization Process

- Stage 1: Harmonize features
- Stage 2: Use harmonized features to predict cognitive status (MCI or NL)



Key Quantitative Results

	Task Classifier	AUC	F1	Sensitivity	Specificity
Conversation classification					
Before harmonization	LR	0.583±0.098	0.557±0.092	0.570±0.123	0.557±0.101
	MLP	0.594±0.092	0.556±0.088	0.545±0.116	0.611±0.091
After harmonization	LR	0.640±0.097	0.581±0.089	0.575±0.129	0.625±0.132
	MLP	0.646±0.092	0.558±0.101	0.541±0.136	0.640±0.126
Subject classification					
Before harmonization	LR	0.591±0.124	0.579±0.126	0.593±0.166	0.568±0.169
	MLP	0.626±0.122	0.593±0.124	0.576±0.153	0.649±0.159
After harmonization	LR	0.649±0.121	0.592±0.115	0.575±0.157	0.652±0.162
	MLP	0.657±0.113	0.571±0.118	0.546±0.152	0.655±0.152

Take away: Subject Harmonization improves predictive performance on both cognitive classification tasks (conversation-wise and subject-wise)

Discussions and Future Work

- Validate Subject Harmonization on **different modality** such as speech or brain imaging other than language markers.
- Can we deploy subject harmonization in privacy-aware **collaborative learning** (e.g., federated learning)?

Thanks!

http://illidanlab.github.io

Acknowledgement: This material is based in part upon work supported by the National Science Foundation under Grant IIS-2212174, IIS-1749940, Office of Naval Research N00014-20-1-2382, and National Institute on Aging (NIA) RF1AG072449, R01AG051628, and R01AG056102.