# Multi-Source Sparse Learning to Identify MCI and Predict Decline

**Principal Investigators**

YE, JIEPING

**Institution**

UNIVERSITY OF MICHIGAN

**Contact information of lead PI
Country**

USA

**Title of project or programme**

Multi-Source Sparse Learning to Identify MCI and Predict Decline

**Source of funding information**

NIH (NIA)

**Total sum awarded (Euro)**

€ 2,582,902.75

**Start date of award**

01/06/2016

**Total duration of award in years**

1

**The project/programme is most relevant to:**

Alzheimer's disease & other dementias

**Keywords**

Acquired Cognitive Impairment... Aging... Alzheimer's Disease... Alzheimer's Disease including Alzheimer's Disease Related Dementias (AD/ADRD)... Behavioral and Social Science... Bioengineering... Biomedical Informatics Research... Biomedical Information Resources... Biomedical Information Resources and Informatics Research... Brain Disorders... Dementia... Diagnostic Radiology... Mental Health... Networking and Information Technology R&D... Neurodegenerative... Neurosciences... Prevention

## Research Abstract

? DESCRIPTION (provided by applicant): Patients with Mild Cognitive Impairment (MCI) are at high risk of progression to dementia. MCI offers an opportunity to target the disease process early. Clinicians and researchers are intensifying their efforts to detect MCI pre-symptomatically in order to develop preventive treatments. These efforts generate a large amount of data – brain images of multiple modalities, and proteomics, genetic, and neurocognitive data that provide unprecedented opportunities to investigate MCI-related questions with greater precision and predictive power. Understanding its importance, NIH in 2003 funded the Alzheimer's Disease Neuroimaging Initiative (ADNI) to facilitate scientific evaluation of various biomarkers for the onset and progression of MCI and AD. To realize such an ambitious vision, there is an urgent need for multi-source fusion and disease biomarker discovery frameworks. While promising, large volumes of incomplete data from multiple heterogeneous data sources pose huge challenges to scientists and engineers. For instance, the ADNI-1 data (like many other large datasets) exhibit a block-wise missing pattern: most subjects have MRI, genetic information; about half of the subjects have CSF measures; a different half of the subjects have FDG-PET; and some subjects have proteomics data. Although many bioinformatics tools are available, no existing tools offer an effective way to fuse multi-source incomplete data for disease biomarker discovery. Here we aim to develop a novel computational framework to integrate and analyze multiple, heterogeneous, large volume, incomplete biomedical data for early detection of MCI. Our 4 primary aims are: (1) Develop novel structured sparse learning formulations for multi-source fusion. The computational methods will identify biomarkers to correlate multi-source data with MCI. Novel sparse screening methods will be developed to scale the proposed formulations to very high-dimensional data. (2) Develop computational methods to integrate network data. We will develop novel methods for incorporating existing biological knowledge such as pathways represented as networks into the prediction model. The network structure will be used as prior knowledge to constrain model parameters, to further improve predictive power. (3) Develop computational methods to integrate multiple incomplete data sources. The proposed computational framework will integrate multiple heterogeneous data with a block-wise missing pattern. The proposed framework formulates the multiple incomplete data source fusion problem as a multi-task learning problem by first decomposing the prediction problem into a set of tasks, then building the models for all tasks simultaneously. (4) Develop and disseminate software tools for multi-source fusion and biomarker identification. The software tools will be used for early detection of MCI and will be validated by several clinical research projects. Our open source software will be made freely available to the research communities, including our large community of existing users. One of our current packages, SLEP, has ~4,500 active users from ~25 countries. Our software tools will be easily adaptable for analyzing multi-source data from other neurological and psychiatric disorders.

## Lay Summary

PUBLIC HEALTH RELEVANCE: This project focuses on an integrated analysis of multiple,

heterogeneous, large volume, incomplete data in the Alzheimer's Disease Neuroimaging Initiative (ADNI) database, and the tools will be applied to large biomedical databases in general. Our research will accelerate the discovery of more sensitive and integrated biomarkers for detecting MCI and predicting brain and cognitive decline.

**Further information available at:**

**Types:**
Investments > €500k

**Member States:**
United States of America

**Diseases:**
Alzheimer's disease & other dementias

**Years:**
2016

**Database Categories:**
N/A

**Database Tags:**
N/A