

Charles Phelps Taft Research Center
at the University of Cincinnati
Summer Research Fellowship Application

Each section (I-V) should be placed at the start of a new page. All required materials must be included in a single document, uploaded to the electronic submissions system, no later than 5PM on the published day of the deadline. This program requires a letter of support from the applicant's department head. Applicants should submit their application with enough time for the letter to be uploaded to the submissions system prior to the close of the deadline. Taft does not accept an obligation to review applications that have not received the required documents by the close of the deadline.

I General Information

- (a) Name: XXX
- (b) M#: XXXXXX
- (c) Department: XXXXX
- (d) Position: XXXXXXX
- (e) Project Title: Aberrant structure MRI in Parkinson's disease based on tensor regression analysis
- (f) Probable Results of a Grant (such as external funding, publications, or presentations): Scientific output from the project is intended for publications in high impact peer-reviewed journals such as Annals of Applied Statistics, Neuroimage and for presentations both within and outside university network, such as Women in Statistics and Data Science Conference, Joint Statistical Meetings, ENAR Conference, etc.
- (g) Other Funding Applied For or Received for This Project (list source and amounts requested and awarded): I have not applied for other funding for this project.

II Budget

- (a) Requested Research Supplement: Not applicable. The PI is not applying for additional research supplement.

III Project Proposal

III.I Introduction

Parkinson's disease (PD) is a major neurodegenerative disease influenced by both genetic and environmental factors. As the second most common neurodegenerative disorder, PD is characterized by the degeneration of dopamine-producing cells in the brain resulting in motor and nonmotor features. Nonmotor features can appear in the earliest phase of the disease even before clinical motor impairment. Depression is a prominent nonmotor feature that is highly prevalent early in the disease process and has a significant impact on quality of life and disability. Available diagnostic tools are better at detecting motor symptoms than nonmotor symptoms including depression. Hence, the neural and patho-physiologic mechanisms for progression of depression in PD (DPD) remain unclear and are key research priorities. How to understand the inner working mechanisms in order to discover biomarkers of DPD is one of the most intriguing scientific questions.

Structure magnetic resonance imaging (MRI) has received more research focus with better stability and repeatability for predicting PD and DPD compared to Resting-state functional MRI, where there are concerns about accuracy due to noise. Machine learning and artificial intelligence are recognized as booming and promising methods used to detect abnormal brain structures in MRI. Most recently, a computer-based technique utilizing convolutional neural networks to create prognostic and diagnostic biomarkers has generated a lot of attention. However, these methods typically require significantly large memory and extensive computation time. In addition, the intuitions behind these machine learning methods are not apparent as the model parameters could not be explicitly interpreted.

On the contrary, tensor (i.e. multi-dimensional array) regression model is a regression framework that treats clinical outcome as response, and images as covariates in the form of multi-dimensional array. These tensor regression methods not only could resolve the computational and modeling challenges of large-scale imaging data, but also could achieve perfect accuracy even in smaller sample sizes.

In this project, the PI will aim to build and validate a tensor regression based framework that incorporated both three-dimensional MRI data and clinical information using the method of maximum likelihood estimation to predict the individual diagnosis of PD and the development of DPD in PD patients. In addition, the proposed framework will identify regions of interest in PD and DPD that are relevant to the onset of PD or DPD such that physicians could get an early-diagnosis in time for available treatment.

III.II Participants and MRI preprocessing

This study was approved by the Medical Research Ethical Committee of Nanjing Brain Hospital (Nanjing, China) in accordance with the Declaration of Helsinki, and written informed consent was obtained from all subjects. Sixty-nine PD patients including twenty-one DPD patients and fifty healthy controls (HCs) were recruited. All the demographic

characteristics and clinical symptom ratings were collected before MRI scanning and all patients were in the ON state during the MRI scan.

Three-dimensional T1-weighted images from both PD patients and HCs were then normalized using Statistical Parametric Mapping on the Matlab platform. The detailed step included spatial normalization to the Montreal Neurological Institute space using the transformation parameters estimated via a unified segmentation algorithm (Fig. 1). Individual images for all subjects were therefore mapped into a common reference space. As a result, the images of original size of (512, 512, 128) were converted into images of size (79, 95, 79) such that the complexity of the following analysis were dramatically reduced without lost of relevant information.

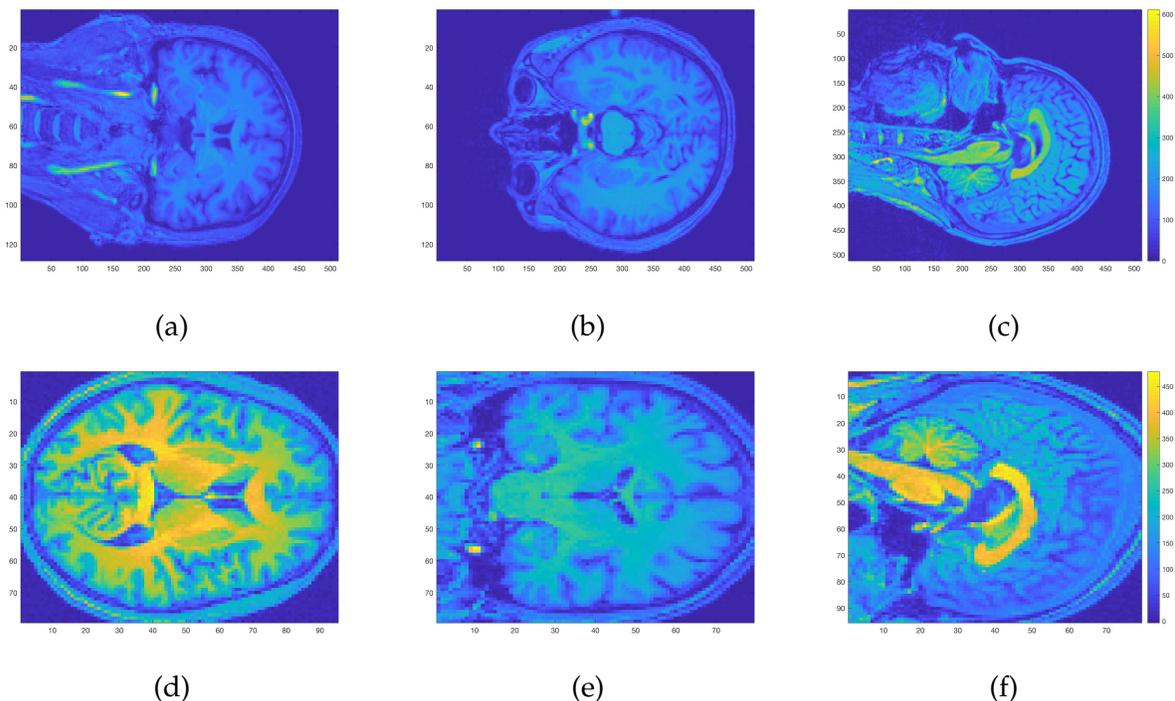


Figure 1: Original MRI images (a-c) and normalized MRI images (d-f).

III.III Proposed research: tensor regression leveraging MRI

Modern applications in medical imaging generate covariates of more complex form such as multidimensional arrays (tensors). Traditional statistical and computational methods including deep learning are proving either computationally infeasible or expensive for dealing with these ultra high-dimensional data with complex structure. For example, the normalized three-dimensional MRI images in our cases have $79 \times 95 \times 79 = 592,895$ voxels, which yield 592,895 parameters to estimate, each representing value of one voxel. To address this issue, this project will utilize the following family of tensor regression models that incorporate both clinical data and the special structure of tensor covariates encoded in these MRI images. The curse of dimensionality is diminished by imposing a

low rank approximation to the extremely high-dimensional full coefficient array, which allows development of a fast estimation algorithm and regularization.

Let Y again be a binary vector indicating whether the subject has PD or DPD, Z be a vector-valued covariate containing clinical information and $X \in \mathbb{R}^{79 \times 95 \times 79}$ be a three-dimensional array variate. The tensor classification model can be expressed as

$$\begin{aligned} Y &\sim \text{Bernoulli}(\mu), \\ g(\mu) &= \alpha + \gamma^\top Z + \langle B, X \rangle, \end{aligned} \tag{1}$$

where $g(\mu) = \log \frac{\mu}{1-\mu}$, α is the intercept, γ is the coefficient vector for covariate Z , and $\langle B, X \rangle$ represents the inner product of tensor B and X . $B \in \mathbb{R}^{79 \times 95 \times 79}$ is a weight tensor in the following form $B = \beta_1 \circ \beta_2 \circ \beta_3$, where $\beta_1 \in \mathbb{R}^{79}$, $\beta_2 \in \mathbb{R}^{95}$ and $\beta_3 \in \mathbb{R}^{79}$ are three vector components for each of the three dimensions respectively, and $\beta_1 \circ \beta_2 \circ \beta_3$ denotes the outer products for three vectors. Equivalently, $\beta_1 \circ \beta_2$, $\beta_1 \circ \beta_3$ and $\beta_2 \circ \beta_3$ demonstrate the coefficient matrices for each of three directions respectively and each entry of B stands for the log odds ratio of each voxel in the MRI image. Thus, by re-parameterizing, the multilinear model (1) is only $= 79 + 95 + 79 = 253$ -dimensional. The parameters α, γ, B can be obtained by maximizing the log likelihood with a block relaxation algorithm [de Leeuw, 1994]. These parameters will then be plugged in (1) with both clinical and image covariates for one particular subject to obtain the probability of having PD. In practice, the cutoff thresholding value 0.5 will be applied on the resulting probability for diagnosis and prediction.

References

Jan de Leeuw. Block-relaxation algorithms in statistics. In Hans-Hermann Bock, Wolfgang Lenski, and Michael M. Richter, editors, *Information Systems and Data Analysis*, pages 308–324, Berlin, Heidelberg, 1994. Springer Berlin Heidelberg.

IV Taft Grant History

Domestic Conference Travel Grant \$721 for presenting the paper entitled “Consistent Bayesian Joint Variable and DAG Selection in High Dimensions” in ENAR Conference, March 2019.

V Curriculum Vitae

