Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2013, Article ID 428385, 8 pages
http://dx.doi.org/10.1155/2013/428385

Research Article

Group Factor Analysis for Alzheimer's Disease

Wei-Chen Cheng, Philip E. Cheng, and Michelle Liou

Institute of Statistical Science, Academia Sinica, Taipei 11529, Taiwan

Correspondence should be addressed to Wei-Chen Cheng; wccheng@stat.sinica.edu.tw

Received 21 December 2012; Accepted 18 January 2013

Academic Editor: Kumar Durai

For any neuroimaging study in an institute, brain images are normally acquired from healthy controls and patients using a single track of protocol. Traditionally, the factor analysis procedure analyzes image data for healthy controls and patients either together or separately. The former unifies the factor pattern across subjects and the latter deals with measurement errors individually. This paper proposes a group factor analysis model for neuroimaging applications by assigning separate factor patterns to control and patient groups. The clinical diagnosis information is used for categorizing subjects into groups in the analysis procedure. The proposed method allows different groups of subjects to share a common covariance matrix of measurement errors. The empirical results show that the proposed method provides more reasonable factor scores and patterns and is more suitable for medical research based on image data as compared with the conventional factor analysis model.

1. Introduction

Modern medical imaging techniques are capable of measuring human brain in vivo [1]. For instance, magnetic resonance (MR) imaging measures nuclei of atoms, and positron emission tomography detects the positron-emitting radionuclides to construct three-dimensional images. The imaging procedures are designed and settled before medical or cognitive experiments. Once the protocol is established, the laboratory and the hospital begin to recruit a variety of subjects of interest into experimental sessions. Errors resulting from individual scans are actually generated from common sources, such as the scanner, protocol, and software. Initial classification of subjects into groups can be realized by using clinical diagnosis, which may be uncertain to some extent, provided by physicians along with subjects' anamnesis.

Conventional factor analysis [2] models reduce high-dimensional data into a few latent variables and assume that data $\mathbf{x}$ were generated by a set of unobserved independent unit-variance Gaussian source $\mathbf{f}$ plus uncorrelated zero-mean Gaussian random noise $\mathbf{u}$, $\mathbf{x} = L\mathbf{f} + \mathbf{u}$, where L is the factor loading matrix. The sample covariance of $\mathbf{x}$ can be expressed as $LL^T + \Psi$, where Ψ is a diagonal covariance matrix of random noises. The goal of factor analysis is to find L and Ψ that maximally fit the sample covariance [3–5]. The EM algorithm was proposed to estimate the matrices [6]. Factor analysis is commonly applied to the dataset as a whole or to different groups of data separately, which may result in factor patterns hard to interpret and limit the potential use of the method in a wider range of medical applications. In this study, we propose a mixture factor analysis model (MFAM) to assign a common covariance matrix of noises or measurement errors to different groups of subjects but to allow individual groups having their own latent structures. In the empirical application, we analyzed an Alzheimer's disease (AD) dataset by first extracting the volumetric information from MR anatomical images for both healthy controls and the patients suffering either AD or mild cognitive impairment, followed by applying the proposed MFAM to the volumetric data.

2. Material and Method

2.1. The Model. Let M be the number of subject groups. To find multiple sets of factor loadings, $\{L_j; j = 1, \ldots, M\}$, with

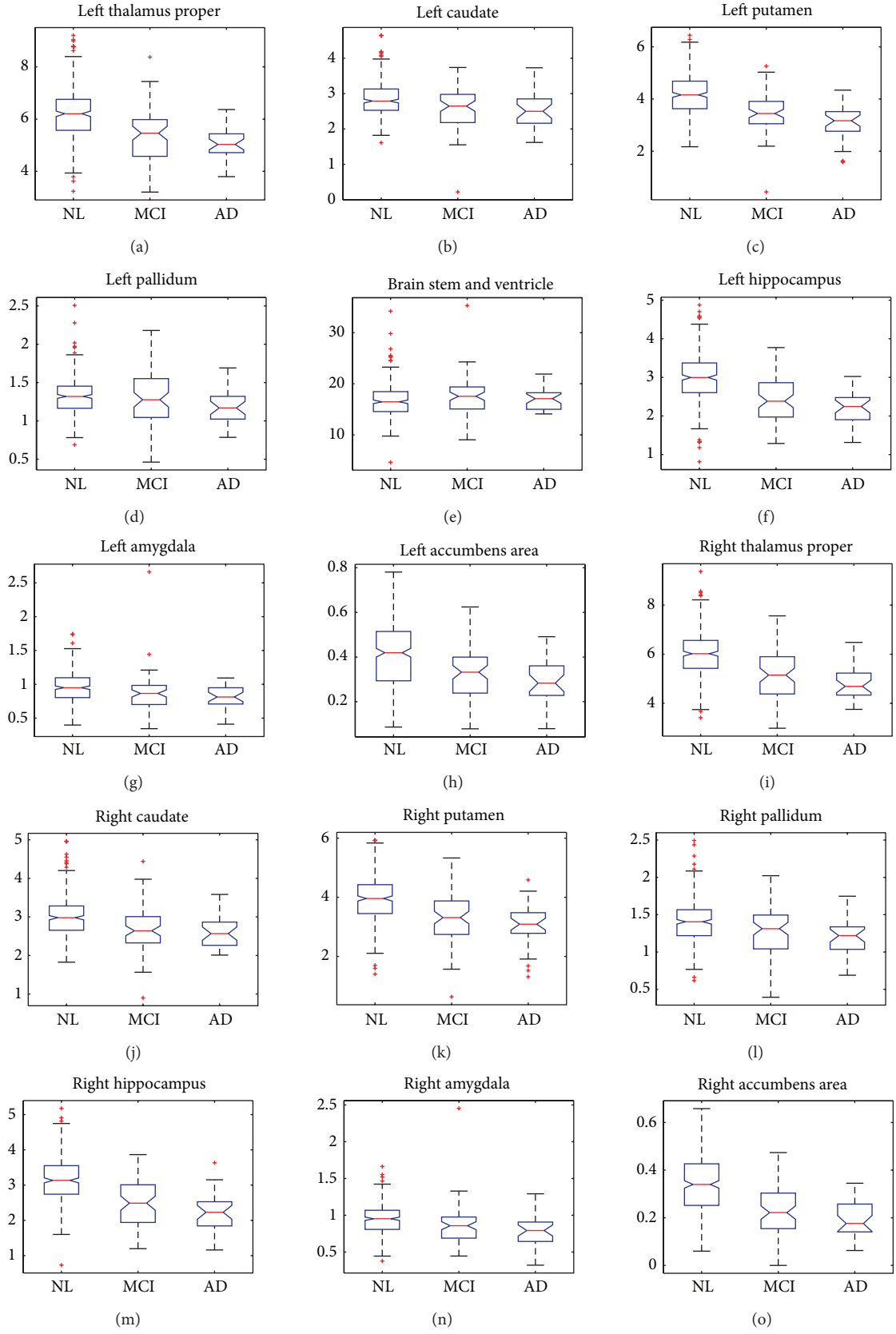


Figure 1: The plots of means and standard deviations for the three groups in different subcortical structures.

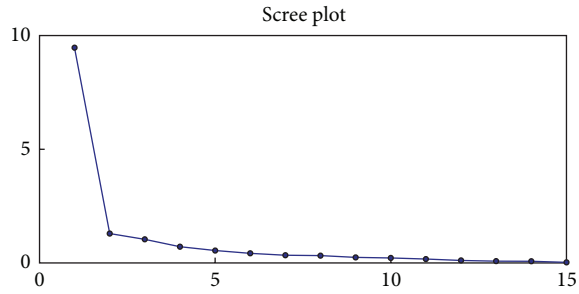


FIGURE 2: The scree plot for the ordered eigenvalues.

the $\mathbf{f}$ scores distributed as Gaussian within each group, the data vector can be decomposed into a linear combination of factor loadings for each group [7, 8], that is, $L_j \in R^{D\times K}$,

$$\boldsymbol{x} = \sum_j \pi_j \left(\boldsymbol{\mu}_j + L_j \times \mathbf{f} \mid w_j\right) + \mathbf{u}, \tag{1}$$

where $\mathbf{x}$ is D-dimensional and each factor scores $\mathbf{f} \mid w_j$ has K variables, that is, $\mathbf{f} \in R^K$. The parameter $\boldsymbol{\pi}$ is associated with the proportion of subjects in the jth group, $\pi_j = p(w_j)$. The indicator variable w is one, $w_j = 1$, when the data belongs to jth group, otherwise w is set to zero, $w_j = 0$. The formula (1) using $\boldsymbol{\pi}$ introduces the main difference from previous mixture models of factor analysis. The data vector $\mathbf{x}$ need not be centered and the mean of the jth group data is $\boldsymbol{\mu}_j$. The covariance matrix of residuals $\mathbf{u}$ is a diagonal matrix $\Psi = \text{diag}[\Psi_1, \Psi_2, \ldots, \Psi_D]$. The data distribution can be expressed as

$$P(\mathbf{x}) = \sum_{j=1}^{M} \int P\left(\mathbf{x} \mid \mathbf{f}, w_j\right) P\left(\mathbf{f} \mid w_j\right) p\left(w_j\right) d\mathbf{f}. \tag{2}$$

In this work, capitalized P denotes the probability function of a vector or a matrix and lowercase p denotes the probability function of a scalar. The factor scores are assumed to be distributed as Gaussian

$$P\left(\mathbf{f} \mid w_j\right) = N(0, I), \quad \forall j. \tag{3}$$

The notation I is the identity matrix of order D. The distribution of data $\mathbf{x}$ in each group is given by

$$P\left(\mathbf{x} \mid \mathbf{f}, w_j\right) = N\left(\boldsymbol{\mu}_j + L_j\mathbf{f}_j, \Psi\right). \tag{4}$$

Based on the MFAM (2), the likelihood function Q is as follows:

$$Q = E\left[\prod_{i=1}^{N}\prod_{j=1}^{M}\left\{(2\pi)^{-D/2}|\Psi|^{-1/2}\ \exp\left[-\left(\mathbf{x}_i - \boldsymbol{\mu}_j - L_j\mathbf{f}_i\right)^T\Psi^{-1} \times\left(\mathbf{x}_i - \boldsymbol{\mu}_j - L_j\mathbf{f}_i\right)\right]\right\}^{w_j}\right], \tag{5}$$

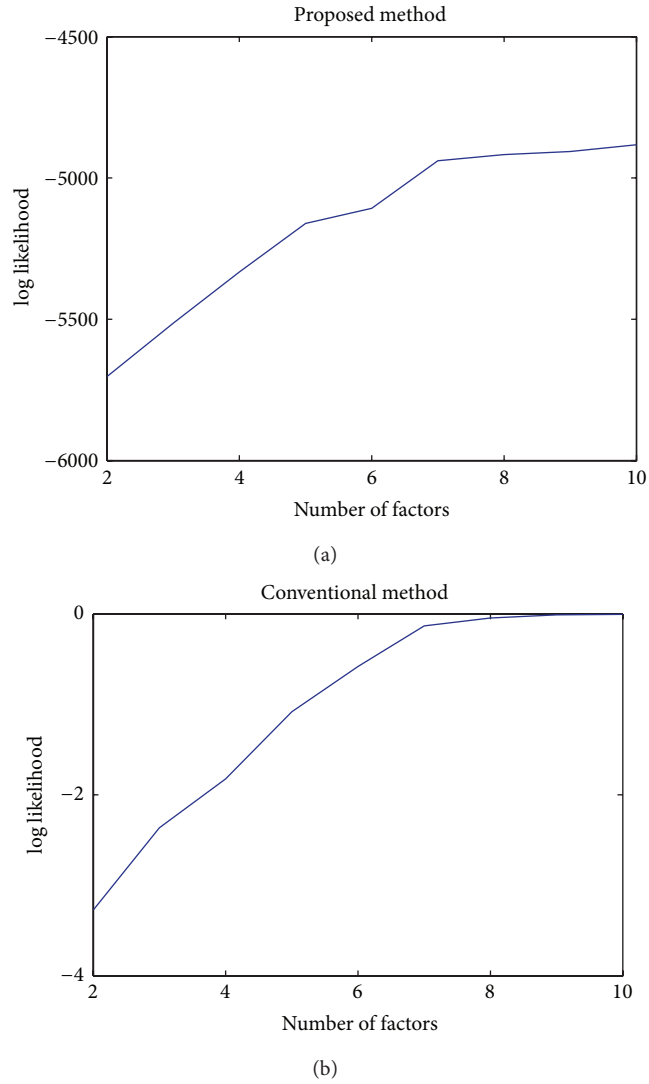


FIGURE 3: The two curves record the approximate value of log likelihood for two methods.

where E denotes the expectation. The N is the number of data vectors (subjects) with subscript i for the ith subject. We need to compute the expectation of the variables,

$$E\left(w_j\mathbf{f}_i \mid \mathbf{x}_i\right) = E\left(w_j \mid \mathbf{x}_i\right) E\left(\mathbf{f}_i \mid w_j, \mathbf{x}_i\right). \tag{6}$$

To estimate Q in (5), the posterior probability of the jth group is calculated as

$$P\left(w_j \mid \mathbf{x}\right) = \frac{P\left(\mathbf{x} \mid w_j\right) P\left(w_j\right)}{P(\mathbf{x})} = \frac{\pi_j N\left(\mathbf{x} - \boldsymbol{\mu}_j, L_jL_j^T + \Psi\right)}{\sum_u \pi_u N\left(\mathbf{x} - \boldsymbol{\mu}_u, L_uL_u^T + \Psi\right)}, \tag{7}$$

where the probability of $\mathbf{x}$ given w_j is

$$P\left(\mathbf{x} \mid w_j\right) = N\left(\mathbf{x} - \boldsymbol{\mu}_j, L_jL_j^T + \Psi\right). \tag{8}$$

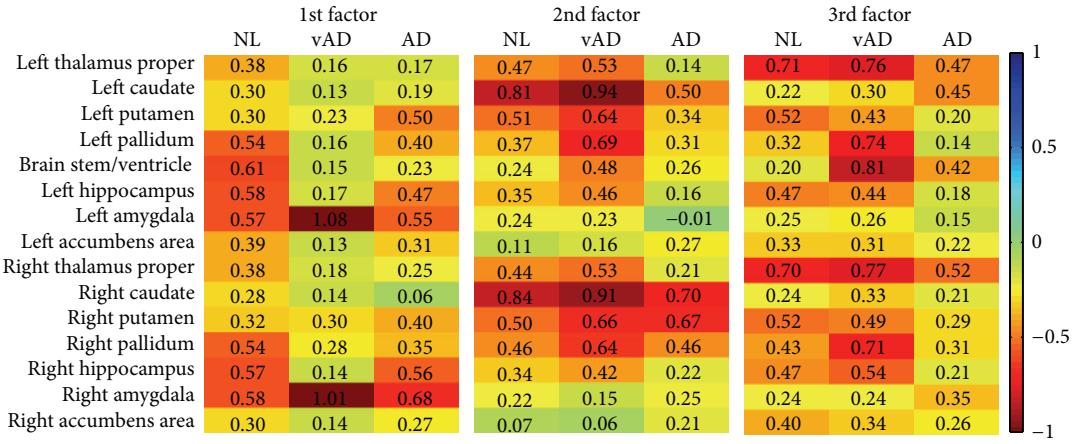


Figure 4: The factor loading matrices for the three groups.

The parameters $\boldsymbol{\pi}$ in (7) is the prior probability derived from the clinical diagnosis. Therefore, the expectation of w_j given $\mathbf{x}_i$ in (6) is proportional to the numerator in (7),

$$h_{ij} = E\left[w_j \mid \mathbf{x}_i\right] \propto \pi_j N\left(\mathbf{x}_i - \boldsymbol{\mu}_j, L_j L_j^T + \Psi\right). \tag{9}$$

To calculate (6), we consider that the posterior probability of $\mathbf{f}$ given $\mathbf{x}$ is

$$P(\mathbf{f} \mid \mathbf{x}) = \frac{P(\mathbf{x} \mid \mathbf{f}) P(\mathbf{f})}{P(\mathbf{x})} \propto \exp\left(-\left[\mathbf{f}^T \left(L^T \Psi^{-1} L + I\right) \mathbf{f} - 2\mathbf{f}^T L^T \Psi^{-1} \mathbf{x}\right]\right). \tag{10}$$

After some arithmetic calculation, $P(\mathbf{f} \mid \mathbf{x})$ can be expressed as

$$P(\mathbf{f} \mid \mathbf{x}) \sim N\left(R^{-1} L^T \Psi^{-1} \mathbf{x}, R\right), \tag{11}$$

where $R = (L^T \Psi^{-1} L + I)$. Hence, the expectation of $\mathbf{f}$ given $\mathbf{x}$ is

$$E[\mathbf{f} \mid \mathbf{x}] = R^{-1} L^T \Psi^{-1} \mathbf{x}. \tag{12}$$

From above, $E(\mathbf{f}_i \mid w_j, \mathbf{x}_i)$ in (6) is calculated as

$$E\left(\mathbf{f}_i \mid w_j, \mathbf{x}_i\right) = R_j^{-1} L_j^T \Psi^{-1} \left(\mathbf{x}_i - \boldsymbol{\mu}_j\right), \tag{13}$$

where $R_j = (L_j^T \Psi^{-1} L_j + I)$, according to (12).

There is no constraint on those factor loadings L_j. The estimation of L_j is simply the maximum of Q. A convenient way to express Q in (5) is achieved by setting $\widetilde{\mathbf{f}}_i = [\mathbf{f}_i^T \quad 1]^T$ and $\widetilde{L}_j = [L_j \quad \boldsymbol{\mu}_j]$. The expected log likelihood function can be expressed as

$$\begin{aligned} &E[\log Q] \\ &= E\left[\log \prod_{i=1}^{N} \prod_{j=1}^{M} \left\{(2\pi)^{-D/2} |\Psi|^{-1/2} \exp\left[-\frac{1}{2}\left(\mathbf{x}_i - \widetilde{L}_j \widetilde{\mathbf{f}}_i\right)^T \Psi^{-1} \times \left(\mathbf{x}_i - \widetilde{L}_j \widetilde{\mathbf{f}}_i\right)\right]\right\}^{w_j}\right] \\ &= -\frac{D \times N}{2} \times \log(2\pi) - \frac{N}{2} \log|\Psi| \\ &\quad - \sum_{i,j} \frac{1}{2} h_{ij} \mathbf{x}_i^T \Psi^{-1} \mathbf{x}_i - h_{ij} \mathbf{x}_i^T \Psi^{-1} \widetilde{L}_j E\left[\widetilde{\mathbf{f}}_i \mid \mathbf{x}_i, w_j\right] \\ &\quad + \frac{1}{2} h_{ij} \times \operatorname{trace}\left[\widetilde{L}_j^T \Psi^{-1} \widetilde{L}_j E\left[\widetilde{\mathbf{f}}_i \widetilde{\mathbf{f}}_i^T \mid \mathbf{x}_i, w_j\right]\right]. \end{aligned} \tag{14}$$

To maximize Q with respect to $\widetilde{L}_j$, we equate the derivative of (14) to zero,

$$\begin{aligned} \frac{\partial \log E[Q]}{\partial \widetilde{L}_j} &= -\sum_i h_{ij} \Psi^{-1} \mathbf{x}_i E\left[\widetilde{\mathbf{f}}_i \mid \mathbf{x}_i, w_j\right]^T \\ &\quad + h_{ij} \Psi^{-1} \widetilde{L}_j E\left[\widetilde{\mathbf{f}}_i \widetilde{\mathbf{f}}_i^T \mid \mathbf{x}_i, w_j\right]^T = 0 \\ \Longrightarrow \widetilde{L}_j &= \left(\sum_i h_{ij} \mathbf{x}_i E\left[\widetilde{\mathbf{f}}_i \mid \mathbf{x}_i, w_j\right]^T\right) \\ &\quad \times \left(\sum_s h_{sj} E\left[\widetilde{\mathbf{f}}_i \widetilde{\mathbf{f}}_i^T \mid \mathbf{x}_s, w_j\right]^T\right)^{-1}, \end{aligned} \tag{15}$$

where

$$E\left[\widetilde{\mathbf{f}}_i \mid w_j, \mathbf{x}_i\right] = \left[E\left[\mathbf{f}_i \mid w_j, \mathbf{x}_i\right]^T \quad 1\right]^T. \tag{16}$$

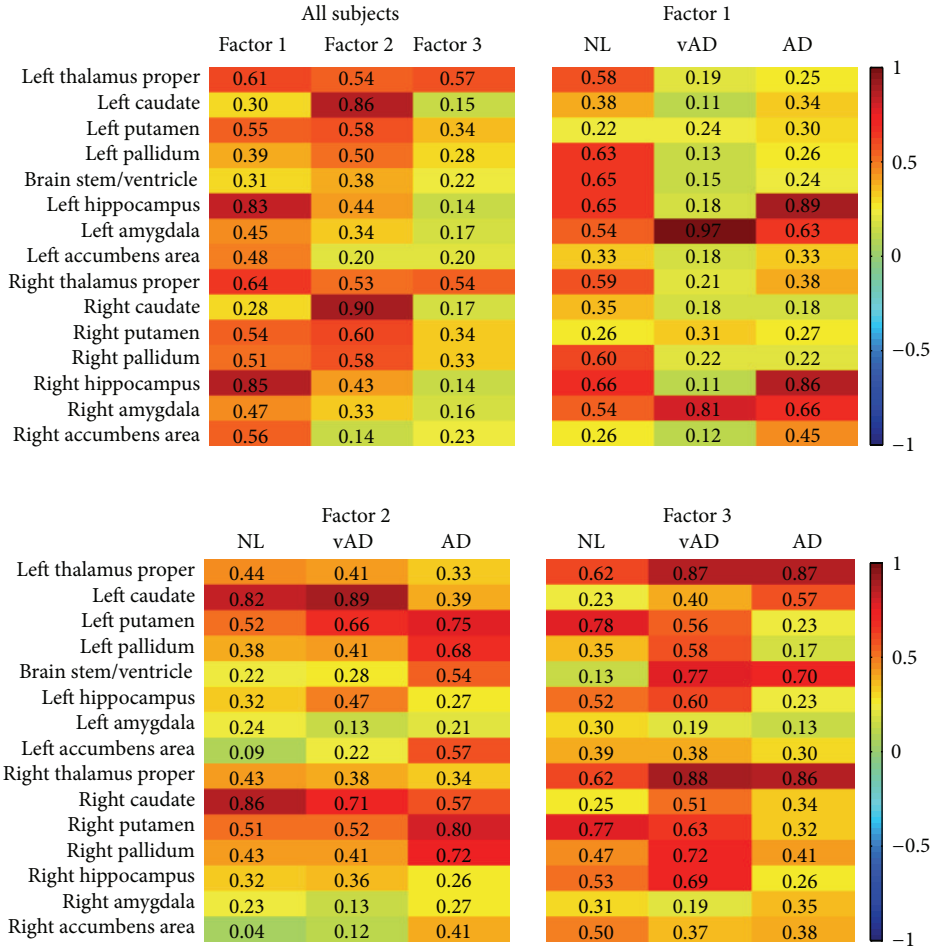


FIGURE 5: The result of maximum likelihood method and rotated by varimax with Kaiser normalization. The results of healthy controls are not unique and unstable.

All the variables are estimated by the EM algorithm. In the E-step, the algorithm computes the expectation of the factor scores in (6) and the second moment of the scores,

$$E\left[w_j \mathbf{f}\mathbf{f}^T \mid \mathbf{x}\right] = E\left[w_j \mid \mathbf{x}\right] E\left[\mathbf{f}\mathbf{f}^T \mid w_j, \mathbf{x}\right], \tag{17}$$

by

$$\begin{aligned} E\left[w_j \boldsymbol{f}_i \boldsymbol{f}_i^T \mid \boldsymbol{x}_i\right] &= h_{ij}\mathrm{Cov}\left(\boldsymbol{f}_i \mid w_j, \boldsymbol{x}_i\right) \\ &\quad + h_{ij} E\left[\boldsymbol{f}_i \mid w_j, \boldsymbol{x}_i\right] E\left[\boldsymbol{f}_i \mid w_j, \boldsymbol{x}_i\right]^T. \end{aligned} \tag{18}$$

The covariance matrix of residual, Ψ, can be estimated by its inverse matrix,

$$\begin{aligned} \frac{\partial Q}{\partial \Psi^{-1}} &= \frac{-N}{2}\Psi - \frac{1}{2}\sum_{i,j} h_{ij}\mathbf{x}_i\mathbf{x}_i^T + h_{ij}\mathbf{x}_i E\left[\widetilde{\mathbf{f}}_i \mid \mathbf{x}_i, w_j\right]^T \widetilde{L}_j^T \\ &\quad - \frac{1}{2}h_{ij}\widetilde{L}_j E\left[\widetilde{\mathbf{f}}_i\widetilde{\mathbf{f}}_i^T \mid \mathbf{x}_i, w_j\right]^T \widetilde{L}_j^T = 0. \end{aligned} \tag{19}$$

Substituting (15) for $\widetilde{L}_j$ and making constraints on the diagonal of Ψ, we obtain

$$\Psi = \frac{1}{N}\,\mathrm{diag}\left(\sum_{i,j} h_{ij}\left(\mathbf{x}_i - \widetilde{L}_j E\left[\widetilde{\mathbf{f}}_i \mid \mathbf{x}_i, w_j\right]\right)\mathbf{x}_i^T\right). \tag{20}$$

The prior probability $p(w_j)$ should be proportional to the clinical diagnosis such that the estimation of the factor loadings and the factor scores can capture the latent factors of different disease groups. The proposed model also carries the same indeterminacy problem associated with factor patterns; that is there exist numerous orthogonal transformations to rotate the matrix of factor loadings without changing the maximum of Q [9]. Considering H be any $K \times K$ orthogonal matrix, $HH^T = H^T H = I$. Equation (1) can be written

$$\begin{aligned} \mathbf{x} &= \sum_j \pi_j\left(\boldsymbol{\mu}_j + L_j \times \left(HH^T\right) \times \mathbf{f} \mid w_j\right) + \mathbf{u} \\ &= \sum_j \pi_j\left(\boldsymbol{\mu}_j + L_j^* \times \mathbf{f}^* \mid w_j\right) + \mathbf{u}, \end{aligned} \tag{21}$$

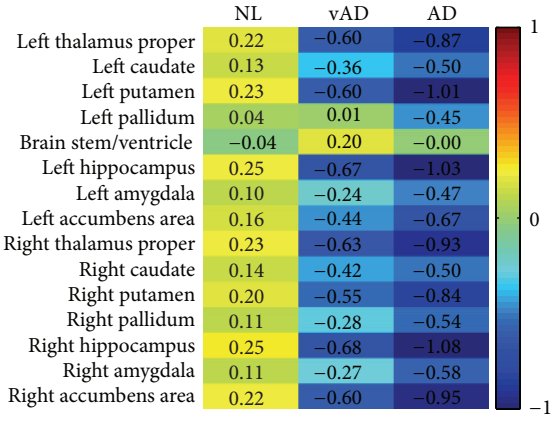


Figure 6: The cluster means of the three groups.

where $L_j^* = L_j H$ and $\mathbf{f}^* \mid w_j = H^T \times \mathbf{f} \mid w_j$. The assumption, $\mathbf{f}^* \mid w_j \sim N(0, I)$, is kept. The covariance of $\mathbf{x}$ is $L_j^*(L_j^*)^T + \Psi = L_j H H^T L_j^T + \Psi = L_j L_j^T + \Psi$, which remains the same. Therefore, there are infinite equivalent solutions to satisfy the maximum of (5). Imposing reasonable constraints to identify a set of model parameters can make the factor loadings scientifically interpretable. A widely used approach for a simple factor structure is realized by setting some factor loadings to hypothetical values such as zeros.

The permutation and changing the sign of columns in the factor loading matrix with factor scores does not affect the model at all and the algorithm will yield the same solution. In order to realize consistent, interpretable, and comparable results, we suggest to recursively test all combinations to find the one of them that has the highest similarity among M factor loading matrices so that we can find a coherent interpretation for different groups of subjects. Each pair of factor loading and factor scores can be multiplied by either +1 or −1. The M sets of loadings has $(2^K)^M$ combinations. The possible permutation of the M set of loadings is the factorial of K. The complexity of the recurrence is therefore $2^{KM} \times (K!)^M$. The problem can be formulated as a bipartite matching and the Hungarian algorithm can find the match in a lower complexity.

2.2. Data Description. The T1-weighted MR images of 416 subjects were downloaded from the Open Access Series of Imaging Studies [10], which is publically available for analysis. All the T1-weighted images were acquired on a 1.5-T Siemens Vision scanner. Among all 416 subjects, there are 316 normal subjects (average age: 45.09 ± 23.90), 70 subjects who have been clinically diagnosed with very mild AD (average age: 76.21 ± 7.19), and 30 are with moderate AD (average age: 78.03 ± 6.91). The proportions of each type of subject are $\boldsymbol{\pi} = [75.96\%, 16.83\%, 7.21\%]^T$. Multiple intrasession acquisitions provide extremely high signal-to-noise ratio, making the data amenable to our analysis. The available images were provided skull stripped, gain field corrected, and registered to the atlas space of Talairach and Tournoux [11] with a 12-parameter rigid affine transform. The resolution of the images is 176 × 208 × 176. The number of voxels, which is more than six million, is much larger than the number of subjects. We extracted the clinically and psychologically interested regions instead of processing whole voxels in the image. The subcortical structures are extracted by the segmentation method [12] which uses manually labeled image data as priori information for a Bayesian framework that utilizes the principles of the active shape and appearance models. The size of a subcortical region was calculated by multiplying the voxel size and the number of voxels in the region. Fifteen subcortical regions were successfully extracted.

Table 1: The ANOVA results for the three groups in different subcortical structures.

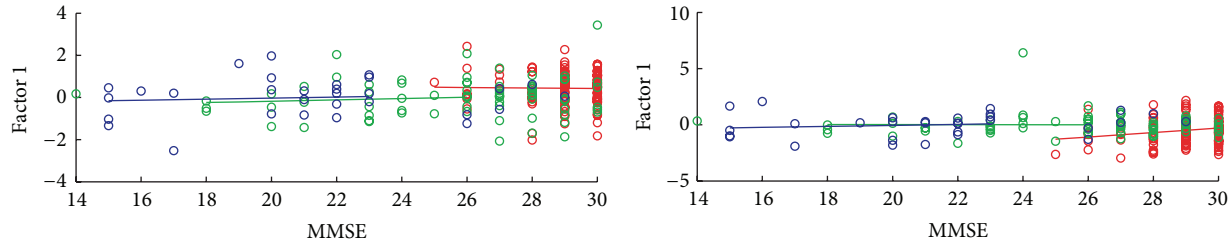
Structure	p-value
Left thalamus proper	2.0183×10^{-15}
Left caudate	1.3736×10^{-5}
Left putamen	6.9145×10^{-18}
Left pallidum	0.0351
Brain stem and ventricle	0.1707
Left hippocampus	1.5656×10^{-20}
Left amygdala	9.1999×10^{-4}
Left accumbens area	1.0958×10^{-8}
Right thalamus proper	2.6864×10^{-17}
Right caudate	1.6556×10^{-6}
Right putamen	1.6447×10^{-13}
Right pallidum	8.8771×10^{-5}
Right hippocampus	3.5498×10^{-22}
Right amygdala	6.5971×10^{-5}
Right accumbens area	8.7323×10^{-17}

According to a demographic study by the National Institute on Aging and Alzheimer's Association based on the data collected in the Chicago Health and Aging Project, the prevalence of dementia among individuals aged 71 and older was 13.9%, and AD (Alzheimers disease) was 9.7% [13]. The study was based on a sample of 856 individuals. The $\boldsymbol{\pi}$ was estimated to be $[76.4\%, 13.9\%, 9.7\%]^T$ which is close to the statistics in our empirical data. The data vector of each subject had fifteen dimensions, each corresponding to the volume size of a subcortical structure divided by the estimated total intracranial volume. The average size of all of the intracranial volume is 1480.5 cm^3. The intracranial volume is estimated by the linear registration from a manually measured intracranial volume of a standard brain to the individual brain [14]. The analysis of variance (ANOVA) of the data for each structure were calculated and shown in Table 1 and Figure 1. The smaller p value indicates high probability of inequality of the structure size among the three groups.

We subtracted the mean from the data and used the remainder for analysis. Using the covariance matrix of the data to estimate the factor scores would cause that a few structures dominate the factor loadings; therefore, we divided each dimension by its standard deviation to compel each of them to have unit variance. After the algorithm converged, we used varimax rotation [15], which transforms the loadings

Table 2: The normality test of factor score by Kolmogorov-Smirnov test.

p value	Proposed method			Traditional method		
	Factor 1	Factor 2	Factor 3	Factor 1	Factor 2	Factor 3
NL	0.20	0.12	0.53	0.51	0.19	0.82
vAD	0.00	0.55	0.29	0.54	0.44	0.30
AD	0.02	0.14	0.14	0.81	0.89	0.83



(a) Proposed method (b) Traditional method

Figure 7: Correlations of factor scores with the MMSE scores. The red color denotes healthy subjects; the green color denotes very mild AD patients; the blue color denotes the moderate AD patients. The regression lines of the three groups by the (a) proposed method and (b) conventional method are also plotted in the figures.

into the space that maximizes the variance, to rotate the factor loadings. Given data, the expectation of its type was set to

$$[h_{i1}, h_{i2}, h_{i3}] = \begin{cases} [1,0,0] & \text{if subject } i \text{ belongs to NL} \\ [0,1,0] & \text{if subject } i \text{ belongs to vAD} \\ [0,0,1] & \text{if subject } i \text{ belongs to AD.} \end{cases} \quad (22)$$

3. Results

Figure 3 shows the trend of the likelihood climbs as adding the number of factors in the analysis. In the scree plot in Figure 2, three eigenvalues of the covariance matrix of the whole dataset are greater than one and the cumulative percentage of variance from the largest three eigenvalues reaches 78%. Thus we set $K = 3$ in this analysis.

The factor loadings for the three groups are shown in Figure 4, in which the vertical axis marks the fifteen regions. The vAD denotes the group of very mild AD. The log likelihood in (5) after the algorithm converges is −5475.814. The loading of structures has symmetric property and usually the right and left structures have similar loadings. Using the factor loadings to estimate $\boldsymbol{\pi}$ and the expected group information given $\mathbf{x}$ by (9), we obtain the adjusted and turned proportions as $[82.89\%, 7.97\%, 9.14\%]^T$. This may suggest the underlying variation among different groups of subject and need further investigation. Note that the reestimated proportion h_{ij} is not binary anymore.

We show the results of conventional factor analysis in Figure 5 as a comparison. The program run on the mild AD patients in the dataset cannot achieve reproducible results; therefore, the quantity of mild AD's results in Figure 5 varies from time to time. The analysis for the AD group cannot converge, however the factor loadings are reproducible. The distance of whole factor loading matrices among the three groups for conventional factor analysis is 5.28 while the proposed method is 4.54. The correlation of the three-factor loading matrix estimated by conventional factor analysis methods is $[C_{12}, C_{13}, C_{23}] = [0.3879, 0.1698, 0.3633]$. The correlation by proposed method is $[0.5388, 0.5564, 0.4986]$. Table 2 lists the p values for all factors by the Kolmogorov-Smirnov test [16] on the factor scores against a Gaussian distribution. The test examines the difference between input distributions and a Gaussian distribution. The smaller the p values, the more strongly the test rejects the Gaussian assumption. The algorithm tries to search for loads with normally distributed factor score, hence large p indicates the factor fit well to Gaussian distribution.

The means (centers) of the clusters are shown in Figure 6. The means are near the origin and include negative value because the data are standardized by the subtraction of the overall mean of the data in the preprocess stage. The yellow color in the first column indicates that healthy controls have larger sizes in subcortical structures, and the second and the third columns indicate that the patients have smaller sizes in different subcortical regions in general. The AD patient has very small thalamus, putamen, and hippocampus. The hippocampus is related to memory and learning. The putamen is a structure involved in the regulation of voluntary movement. The abnormal pallidum in Figure 4 can cause movement disorders. Figure 7 shows the associations of first factor scores with the score of minimental state examination (MMSE) by both methods.

4. Conclusions

The proposed method finds closer and more correlated factor loadings than the conventional method because it considers the same error matrix for different groups of data. The result of conventional factor analysis having higher normality

for AD patients than normal subjects is less convincing. Conventional factor analysis that decomposes the observed data together intermixes the latent factors. Taking the data apart will misseparate the noise. This work proposed using a mixture model of factor analysis method for neurodegenerative disease research by showing highly correlated factor loading across different groups of subjects and together with proper normality of the factor scores.

Acknowledgment

This work was supported by National Science Council under project NSC101-2811-M-001-082.

References

[1] W. C. Cheng, P. E. Cheng, and M. Liou, "Modeling local distortion in shape for brain MR images," in *Proceedings of the 17th Annual Meeting of the Organization for Human Brain Mapping*, 2011.

[2] C. Spearman, "General intelligence, objectively determined and measured," *American Journal of Psychology*, vol. 15, no. 1, pp. 201–293, 1904.

[3] D. N. Lawley, "The estimation of factor loadings by the method of maximum likelihood," *Proceedings of the Royal Society of Edinburgh*, vol. 60, pp. 64–82, 2014.

[4] K. Jöreskog, "Some contributions to maximum likelihood factor analysis," *Psychometrica*, vol. 32, pp. 443–482, 1967.

[5] K. Jöreskog and D. N. Lawley, "New methods in maximum likelihood factor analysis," *Mathematical and Statistical Psychology*, vol. 21, pp. 85–96, 1968.

[6] D. B. Rubin and D. T. Thayer, "EM algorithms for ML factor analysis," *Psychometrika*, vol. 47, no. 1, pp. 69–76, 1982.

[7] Z. Ghahramani and G. E. Hinton, "The EM algorithm for mixtures of factor analyzers," Tech. Rep. CRG-TR-96-1, Department of Computer Science, University of Toronto, 1996.

[8] C. Y. Liou and W. C. Cheng, "Manifold construction by local neighborhood preservation," *Lecture Notes in Computer Science*, vol. 4985, no. 2, pp. 683–692, 2008.

[9] R. I. Jennrich and P. F. Sampson, "Rotation for simple loadings," *Psychometrika*, vol. 31, no. 3, pp. 313–323, 1966.

[10] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, "Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults," *Journal of Cognitive Neuroscience*, vol. 19, no. 9, pp. 1498–1507, 2007.

[11] J. Talairach and P. Tournoux, *Co-Planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System—An Approach to Cerebral Imaging*, Thieme Medical Publishers, New York, NY, USA, 1988.

[12] B. Patenaude, S. M. Smith, D. N. Kennedy, and M. Jenkinson, "A Bayesian model of shape and appearance for subcortical brain segmentation," *NeuroImage*, vol. 56, no. 3, pp. 907–922, 2011.

[13] B. L. Plassman, K. M. Langa, G. G. Fisher et al., "Prevalence of dementia in the United States: the aging, demographics, and memory study," *Neuroepidemiology*, vol. 29, no. 1-2, pp. 125–132, 2007.

[14] R. L. Buckner, D. Head, J. Parker et al., "A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume," *NeuroImage*, vol. 23, no. 2, pp. 724–738, 2004.

[15] H. F. Kaiser, "The varimax criterion for analytic rotation in factor analysis," *Psychometrika*, vol. 23, no. 3, pp. 187–200, 1958.

[16] F. J. Massey, "The kolmogorov-smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.