We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

CLARIVATE ANALYTICS
BOOK CITATION INDEX
INDEXED

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Chapter

Supervised Sparse Components Analysis with Application to Brain Imaging Data†

Atsushi Kawaguchi

Abstract

We propose a dimension-reduction method using supervised (multi-block) sparse (principal) component analysis. The method is first implemented through basis expansion of spatial brain images, and the scores are then reduced through regularized matrix decomposition to produce simultaneous data-driven selections of related brain regions, supervised by univariate composite scores representing linear combinations of covariates. Two advantages of the proposed method are that it identifies the associations between brain regions at the voxel level and that supervision is helpful for interpretation. The proposed method was applied to a study on Alzheimer's disease (AD) that involved using multimodal whole-brain magnetic resonance imaging (MRI) and positron emission tomography (PET). For illustrative purposes, we demonstrate cases of both single- and multimodal brain imaging and longitudinal measurements.

Keywords: data-driven approach, dimension reduction, principal component analysis, multimodal, multi-measurement

1. Introduction

Recently, multiple neuroimaging data sets per subject have become obtainable due to the remarkable development of imaging techniques such as magnetic resonance imaging (MRI) and positron emission tomography (PET), as well as computer resources and technologies. Vandenberghe and Marsden [1] provide a review on the use of PET and MRI integration technology, such as integrated scanning devices, rather than data analysis. Other modalities such as diffusion MRIs (dMRIs) and functional MRIs (fMRIs) are also useful in collecting brain-related information. These multimodal imaging data sets have the potential to provide rich information about human health and behavior, such as brain function and structure, from different perspectives. From multiple measurements of a single-modal (or multimodal) technique, longitudinal changes in the status and combination of neuro

---

biomarkers can be observed to support the prediction and early diagnosis of disease and the classification of disease subtypes.

Multimodal brain imaging analysis is important in brain-related disease studies. Arbabshirani et al. [2] provide many reviews on the subject. Imaging data analysis makes a substantial contribution to the study of mental disorders. Most single-modal or multimodal imaging studies concern dementia leading to Alzheimer's disease (AD) [3] (around 300 of the AD imaging studies searched in Ref. [2]). Modalities considered in there are structural MRIs (sMRIs), fMRIs, dMRIs, fluorodeoxyglucose PETs, and Amyloid/Tau PETs. In a recent study, Ref. [4] examined sMRI and cerebrospinal fluid (CSF) markers. Magnetoencephalography (MEG) is also useful as AD biomarker, and its localization using sMRI has high accuracy [5]. Schizophrenia is the second most studied disorder after dementia. Shah et al. [6] provide an example of multimodal meta-analysis. For Huntington's disease, white matter is evaluated using dMRI [7]. For mood disorders (depressive disorder and bipolar disorder), Refs. [8, 9] provide a review of the machine learning method. Moeller and Paulus [10] studied the longitudinal prediction of relapse for substance-related disorders using MRI, fMRI, EEG, and PET. Moser et al. [11] studied schizophrenia and bipolar disorder using multimodal imaging data analysis. dMRI is also effective for analyzing these conditions [12]. For developmental disabilities, Ref. [13] investigated volume reductions in attention-deficit hyperactivity disorder (ADHD) with 1713 participants. Aoki et al. [14] reviewed dMRI studies and conducted meta-analysis for ADHD. Li et al. [15] provide a review of imaging studies in autism spectrum disorder. For anxiety disorder, Ref. [16] applied support vector machine (SVM) to multimodal data. They used clinical questionnaires and measured cortisol release, and gray and white matter volumes in subjects with generalized anxiety disorder and major depression and in healthy subjects. Steiger et al. [17] investigated cortical volume, diffusion tensor imaging, and network-based statistics using multimodal analysis for social anxiety disorder. For borderline personality disorder, Ref. [18] conducted an imaging-based meta-analysis of 10 studies. In cancer research, especially that on glioblastoma multiforme, multimodal imaging analysis is useful for identifying some types of tumors and evaluating patient prognosis (for more details, see [19]). Genome-related data can be regarded as a modality and called imaging genetics when analyzed in combination with imaging data [20].

One important technique for single- and multimodal imaging analysis is prediction, which is useful for the support of disease diagnosis and the selection of treatments [21]. SVM is the most used method not only in neuroimaging but also in the life sciences in high-dimensional data analysis. The random forest method is also useful due to their capability for complex interactions based on the tree model [22, 23]. For multimodal analysis, multiple kernel learning [24] and (multimodal) deep learning [25, 26] have been developed. Janssen et al. [27] reviewed machine learning methods for psychiatric prognosis. Related statistical methodology appeared as multi-omics in bioinformatics, and Ref. [28] reviewed these methods while introducing an R package, mixOmics.

Analysis for such discovery and evaluation is based on the detection of the buried signal in the noise (irrelevant information). Statistical analysis is useful for this purpose; however, it suffers from the ultrahigh dimensional and complex structure of this data, and appropriate dimension reduction is therefore required. Even if a machine learning method is used, appropriate input (feature) should be specified to obtain interpretable results because the method is feasible for high-dimensional procedures but not ultrahigh dimensional ones. A region-of-interest-based analysis was the leading approach. In contrast, whole-brain analysis is more informative, and if it is combined with a data-driven approach, it can potentially

obtain undiscovered knowledge. In [29], by using ReliefF [30], features such as the fractional amplitude of low-frequency fluctuations from resting-state fMRIs, segmented gray matter from sMRIs, and fractional anisotropy from dMRIs were extracted. Component analysis based on low-rank approximation is a successful data-driven approach in the fields of not only neuroimaging but also other biological and medical big data analyses, including principal component analysis, partial lease squares, canonical correlation analysis (CCA), independent component analysis (ICA), and nonnegative matrix factorization. These methods are organized into a matrix decomposition framework consisting of score and loading (weight) matrices. The score matrix, with same row length as the number of subjects, is regarded as dimension-reduced data and is suitable for application to statistical models. The weight matrix, with the same column length as the number of features in the imaging data, is regarded as the basis images. All these methods, except for ICA, have a derivation sparse approach with a regularized matrix decomposition to pose small weights to zeros, which helps estimation by avoiding irrelevant information. In addition, the resulting weights can be interpreted to mean that the corresponding features with nonzero weights contribute to the basis image, specifically to produce data-driven selections of brain regions related to that component.

These methods also consider another direction in which the application of multimodal imaging data can be extended. Supplementary information from another data set can also be useful for the interpretation of the output. For this purpose, appropriate data fusion or integration techniques are required and are useful for multisite studies. In neuroimaging data analysis, multimodal CCA (mCCA) [31] and mCCA + joint ICA [32] have been developed on the schizophrenia study. Multivariate data fusion approaches were categorized by [33] into asymmetric or symmetric data and blind or semi-blind data in symmetric approach. The asymmetric approach is a regression-type approach and includes specific modalities such as dMRI and electroencephalography. The symmetric approach is a correlation-type approach and allows relationships in both directions. Kawaguchi [19] constructed a risk score for glioblastomas based on MRI data and proposed a two-step dimension-reduction method using a radial basis function-supervised multi-block sparse principal component analysis (SMS-PCA) method. Kawaguchi and Yamashita [34] proposed a more general case including a PLS or CCA framework and applied it to MRI, PET, and SNP data sets. Yoshida et al. [4] analyze imaging and non-imaging data with network structure by using the PLS.

In this chapter, we applied SMS-PCA to MRI and PET data sets and a longitudinal MRI data set. One of the key features in the analysis is a multi-block technique which can achieve structural dimension reduction with interpretable parameters (weights for each data set and the possibility of combining them). Although it is not the focus of this chapter, the dimension reduction prior to SMS-PCA is conducted using 3D basis functions. Specifically, our dimension reduction takes place in two steps, and, as described in [35] which applied these techniques to longitudinal study, this two-step approach yields a composite basis function expression with a flexible shape. The organization of this chapter is as follows. Section 2 describes the methodology of the SMS-PCA, which is applied to real data in Section 3. The characteristics of the method, found through its application, are discussed in Section 4.

2. Methods

We describe the proposed method in this section. The contents are similar to Ref. [19].

2.1 Priory dimension reduction

$\mathbf{S} = \{s_\alpha\}_{\alpha=1,\,\ldots,\,n}$ is the $n \times N$ matrix whose column corresponds to the vectorized original image data. As the dimensions for each mth image are the same, we use the same basis function to reduce the dimension from N to q. $\mathbf{X} = \mathbf{SB}$ is the $n \times q$ matrix, where $\mathbf{B}$ is the $N \times q$ matrix whose jth column corresponds to the vectorized basis function with the jth knot being the center. Note that knots are pre-specified to span the space equally, as shown in **Figure 1**. In this example, four-pixel equal spanning knots are applied.

2.2 Objective function

Dimension reduction using the basis function is then followed by the SMS-PCA method, considering (sample) correlations based on data values. We consider score t for $n \times q$ matrices $\mathbf{X}_m$, where $m = 1, 2, \ldots, M$ with the following multi-block structure:

$$t = \sum_{m=1}^{M} b_m \mathbf{X}_m \boldsymbol{w}_m = \sum_{m=1}^{M} b_m \boldsymbol{t}_m, \tag{1}$$

where $\boldsymbol{w}_m$ is the weight vector for the mth sub-block $\mathbf{X}_m$ and b_m is the weight for the superblock. Here, it should be noted that the scores in Eq. (1) are referred to as the super scores, whereas $\boldsymbol{t}_m = \mathbf{X}_m \boldsymbol{w}_m$ is referred to as the block score. **Figure 2** schematically describes the score structure for the case of $M = 2$.

Thus, the super score has a hierarchical structure for each individual and can be used in an application such as the construction of a diagnosis score.

When matrix $\mathbf{X}_m$ is normalized by its columns, the weights $\boldsymbol{w} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_M)^\top$ and $\boldsymbol{b} = (b_1, b_2, \ldots, b_M)^\top$ are estimated by maximizing the function

$$\mathrm{L}(\boldsymbol{b}, \boldsymbol{w}) = (1 - \mu)\,\boldsymbol{t}^\top \boldsymbol{t} + \mu\,\boldsymbol{t}^\top \mathbf{Z} - \sum_{m=1}^{M} \mathrm{P}_{\lambda_m}(\boldsymbol{w}_m) \tag{2}$$
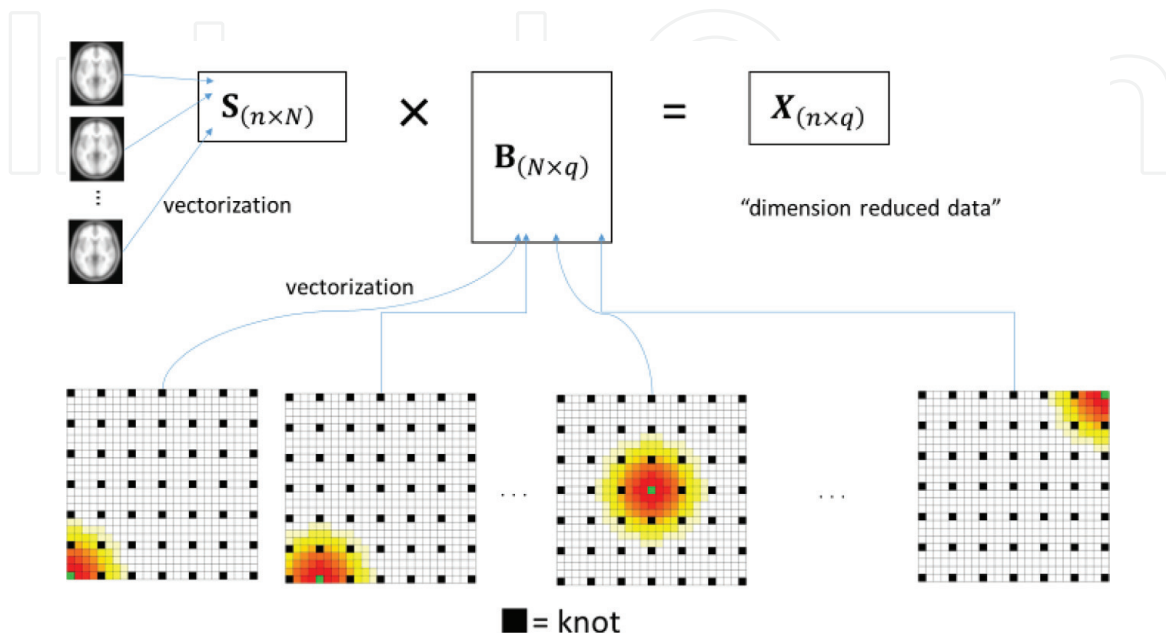

Figure 1.
Dimension reduction via basis function.
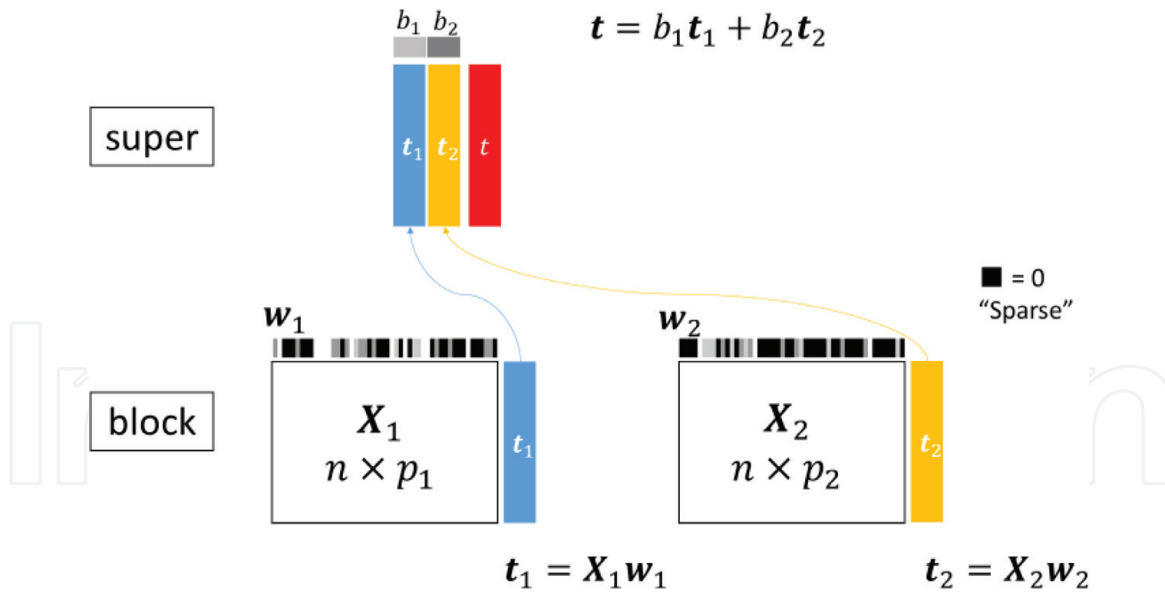
Figure 2.
Score structure.

---

1. Initialize t with $\|t\|_2 = 1$.
2. Repeat until convergence:
   2.1. Set $\widetilde{w}_m = h_{\lambda_m}\left(b_m X_m^\top \{(1-\mu)t + \mu Z\}\right)$, where $h_\lambda(y) = \text{sign}(y)\,(|y| > \lambda)_+$, and normalize as $\hat{w}_m = \widetilde{w}_m / \|\widetilde{w}_m\|_2$ $(m = 1, 2, ..., M)$.
   2.2. Set $t_m = X_m \hat{w}_m$ and $\widetilde{b}_m = t_m^\top\{(1-\mu)t + \mu Z\}$; then set $\widetilde{b} = \left(\widetilde{b}_1, \widetilde{b}_2, ..., \widetilde{b}_M\right)^\top$ and normalize as $\hat{b} = \widetilde{b} / \|\widetilde{b}\|_2$.
   2.3. Set $t = \sum_{m=1}^M \hat{b}_m X_m \hat{w}_m$.
3. (Deflation step) Set $p_m = X_m^\top t_m / t_m^\top t_m$ and $\hat{X}_m = t_m p_m^\top$, and $X_m \leftarrow X_m - \hat{X}_m$.

Table 1.
Algorithm for SMS-PCA method.

subject to $\|w_m\|_2^2 = 1$ and $\|b\|_2^2 = 1$ with $\|\cdot\|_2$ as the L2 norm, where $0 \leq \mu \leq 1$ is the proportion of the supervision, $P_\lambda(x)$ is the penalty function, $[P_\lambda(x) = 2\lambda|x|$ is used in this study], and $\lambda > 0$ is the regularized parameter that is used to control the sparsity. The larger value of the regularization parameter λ_m has many nonzero elements in w_m.

2.3 Optimization

The algorithm given in **Table 1** is used to estimate the weights in Eq. (1) by maximizing L in Eq. (2). The rationality behind this approach is provided in [19].

Note that the deflation step yields multiple components and has several alternatives; that is, through K time iteration for step. 1 to 3 of the algorithm, we can obtain K component super scores $t^{(1)}, ..., t^{(K)}$ with $t^{(k)} = \left(t_1^{(k)}, ..., t_M^{(k)}\right)$.

2.4 Parameter selection

The optimal value for $\lambda = (\lambda_1, ..., \lambda_M)^\top$ is selected by minimizing the Bayesian information criterion (BIC):

$$\text{BIC}(\lambda) = \log\left(\frac{\sum_{m=1}^{M}\left\|\hat{\boldsymbol{X}}_m^{(r)} - \boldsymbol{X}_m\right\|^2}{nMq}\right) + \frac{\log(nMq)}{nMq}(\# \text{ nonzero elements in } \boldsymbol{w}_m),$$

where $\hat{\boldsymbol{X}}_m^{(r)} = \boldsymbol{T}_m^{(r)}\boldsymbol{P}_m^{(r)}$ with $\boldsymbol{T}_m^{(r)} = \left[\boldsymbol{t}_m^{(1)}, ..., \boldsymbol{t}_m^{(r)}\right]$ and $\boldsymbol{P}_m^{(r)} = \left[\boldsymbol{p}_m^{(1)}, ..., \boldsymbol{p}_m^{(r)}\right]$ is obtained from r deflation steps (the projection of $\boldsymbol{X}_m$ onto the r-dimensional subspace). There are several search strategies for optimization, and these are introduced in the software options below.

2.5 Software

The statistical software R package msma is provided to implement the method described in Ref. [34] where the SMS-PCA method is a part of the package and the PLS type can also be implemented. The package is available from the Comprehensive R Archive Network (CRAN) at http://CRAN.R-project.org/package=msma. Four-parameter search methods are available. Here, the parameters are λ_m and the number of components. The "simultaneous" method identifies the number of components by searching the regularized parameters in each component. The "regpara1st" method identifies the regularized parameters by fixing the number of components and then searching for the number of components with the selected regularized parameters. The "ncomp1st" method identifies the number of components with a regularized parameter of 0 and then searches for the regularized parameters with the selected number of components. The "regparaonly" method searches for the regularized parameters with a fixed number of components.

In this chapter, the "ncomp1st" method was applied with nonzero sparsity when the number of components was selected because, in our experience, the BIC value suffered from the high dimensionality of the data. The basic R code for this method is as follows:

tuneparams = optparasearch(X=X, Z=Z, search.method="ncomp1st", maxpct4ncomp=0.5, muX=0.5)

where the argument maxpct4ncomp = 0.5 means that 0.5 $\lambda_{\max}$ is used as the regularized parameter when the number of components is searched and where $\lambda_{\max}$ is the maximum of the regularized parameters among the possible candidates. In order to obtain the final fit result with optimized parameters, the following code should be implemented:

fit1 = msma(X=X, Z=Z, comp=tuneparams$optncomp, lambdaX=tuneparams $optlambdaX, lambdaY=tuneparams$optlambdaY, muX = 0.5)

For more details, please see the package manual.

3. Application

In this section, we apply the SMS-PCA described in the previous section to real data. The data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and

early Alzheimer's disease (AD). We use two types of data set: baseline measurement with multimodal MRI and PET images and repeated measuring MRI images.

3.1 Multimodality

3.1.1 Data

Baseline imaging data were collected from 106 subjects with mean ages of 75.2 years for the 54 normal cognitive subjects and 72.9 years for the 27 patients with dementia. This data set was somewhat larger than that of [34] because in this study single-nucleotide polymorphism (SNP) was not considered and subjects with missing SNP data were included. **Table 2** summarizes the characteristics of these patients.

We consider imaging data from two modalities, MRI X_1 and PET X_2, namely, $M = 2$. The preprocessing method is the same as that used in [34]. For the basis function, we used four-voxel (therefore, $h = \sqrt{3 \times 4^2} = 6.93$) equal spacing knots because of the results of our simulation study. The clinical outcome to supervise is given by $Z = 3.17 \times CDR + 0.11 \times ADAS13 - 0.57 \times MMSE$ where CDR is the clinical dementia rating score, ADAS13 is the Alzheimer's disease assessment scale-cognitive subscale, and MMSE is the mini-mental state examination score. These coefficients were the same as in [34]. The SMSMA method was applied to the data (X_1, X_2, Z) with parameters $\mu = 0, 0.25, 0.5,$ and 0.75.

3.1.2 Results

The original data with dimensions of 2,122,945 (= 121 × 145 × 121) was reduced to 7,162 using the basis functions for each imaging data set. The number of components were selected as 8 for all $\mu = 0, 0.25, 0.5, 0.75, 1$. **Figure 3** shows the correlation matrix from the dataset with the binary outcome, AD or Normal, and the resulting super scores for each μ.

The correlations between the super scores were small except for $\mu = 1$, and for $\mu = 0$, the second component had a high correlation with the outcome. In contrast, for $\mu > 0$, the first component had the highest correlation with the outcome.

Table 3 shows the results for the multiple logistic regression model with AD or normal as the outcomes and the super scores as predictors for each μ. The numbers

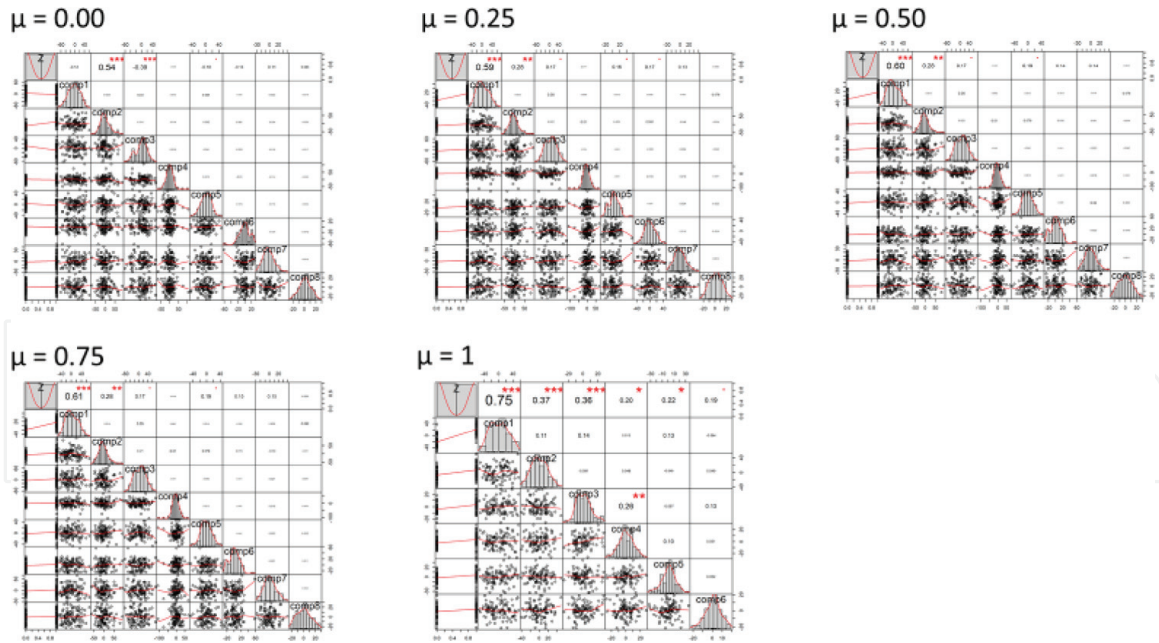|	Dementia	Normal	p
n	52	54	
Age (mean [sd])	75.41 (7.18)	74.93 (4.89)	0.684
PTGENDER = Male (%)	31 (59.6)	36 (66.7)	0.582
APOE4 (%)			<0.001
0	17 (32.7)	39 (72.2)	
1	29 (55.8)	13 (24.1)	
2	6 (11.5)	2 (3.7)	
PTEDUCAT (mean [sd])	14.19 (3.04)	15.89 (2.99)	0.005
CDRSB (mean [sd])	4.54 (1.73)	0.03 (0.12)	<0.001
ADAS11 (mean [sd])	18.70 (5.63)	6.56 (3.28)	<0.001
ADAS13 (mean [sd])	28.94 (6.30)	10.08 (4.30)	<0.001
MMSE (mean [sd])	23.38 (2.07)	28.87 (1.24)	<0.001

Table 2.
Characteristic for data set 1.

Figure 3.
Correlations between super scores.

of 5% statistically significant components were 3, 4, 3, 3, and 0 for μ = 0, 0.25, 0.5, 0.75, and 1, respectively. The minimum numbers of nonzero subweights were 552, 581, 574, 523, and 1075, respectively.

Figure 4 shows the reconstructed subweights Bw_1 and Bw_2 for the MRI and PET data, respectively, overlying a structural brain image shown for the most correlated components with the binary outcome from each of $\mu = 0$, 0.5, 0.75, and 1. The images for $\mu = 0.25$ were similar to those of $\mu = 0.5$, 0.75 and are not shown here.

Figure 5 shows the reconstructed subweights Bw_1 and Bw_2 overlying a structural brain image and bar plots for the super-weights (right bottom) in the case of $\mu = 0.5$ for all components.

In each component, the negative and positive sides are represented. These can be interpreted by looking at the sign of the super-weight. Most cases remain on one side of 0 (positive or negative), except for components 5 to 8. The super-weights are similar between MRI and PET.

A 10-fold cross validated ROC analysis (**Figure 6A**) was conducted to evaluate the diagnostic probabilities estimated from the multivariable logistic regression mode whose coefficients and p-values are shown in **Table 3**. For comparison, the single modalities, MRI (**Figure 6B**) and PET (**Figure 6C**), were also analyzed.

In the case of the multimodal MRI and PET (**Figure 6A**), $\mu = 1$ had the highest AUC value (0.984) following by $\mu = 0.75$ (AUC = 0.880). In the case of the single-modal MRI (**Figure 6B**), all values were below the AUC values of the multimodal case. In the case of the single-modal PET (**Figure 6C**), $\mu = 1$ and 0.75 outperformed the multimodal case, and the other values ($\mu = 0$, 0.25, and 0.5) did not.

3.2 Multi-measurements

3.2.1 Data

The second data set was a collection of repeated measured imaging data from 68 patients with mild cognitive impairment (MCI). There were two groups, the conversion to dementia MCI (cMCI) group and the stable MCI (not converted to dementia, sMCI) group. MRI data measured at four time points were used. For the cMCI group, the four time points were just before diagnosis of conversion. For the sMCI group, the

	μ = 0.00		μ = 0.25		μ = 0.50		μ = 0.75		μ = 1	
	Estimate	**Pr(>|z|)**	**Estimate**	**Pr(>|z|)**	**Estimate**	**Pr(>|z|)**	**Estimate**	**Pr(>|z|)**	**Estimate**	**Pr(>|z|)**
comp1	−0.0210	0.0615	0.0827	<0.0001	0.0832	<0.0001	0.0857	<0.0001	4.287	0.9982
comp2	0.0882	<0.0001	0.0460	0.0030	0.0458	0.0031	0.0451	0.0039	2.555	0.9989
comp3	−0.0621	0.0001	0.0180	0.0923	0.0180	0.0920	0.0181	0.0952	4.633	0.9995
comp4	−0.0072	0.6126	0.0037	0.7953	0.0044	0.7583	0.0064	0.6574	1.827	0.9987
comp5	−0.0424	0.0228	0.0452	0.0432	0.0430	0.0203	0.0431	0.0206	3.905	0.9988
comp6	−0.0364	0.0900	0.0396	0.0320	0.0403	0.0715	0.0425	0.0658	4.994	0.9984
comp7	0.0446	0.0891	0.0505	0.0636	0.0510	0.0619	0.0535	0.0549		
comp8	0.0336	0.2517	−0.0228	0.3804	−0.0226	0.3816	−0.0218	0.3904		

Table 3.
Results for multivariable logistic regression analysis.
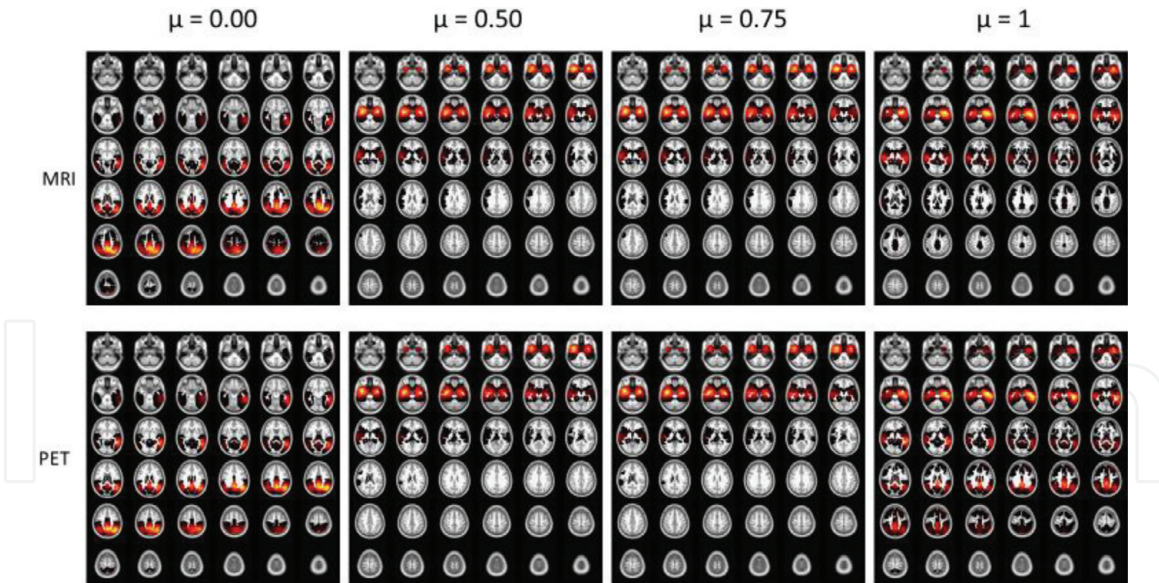
Figure 4.
Subweights.
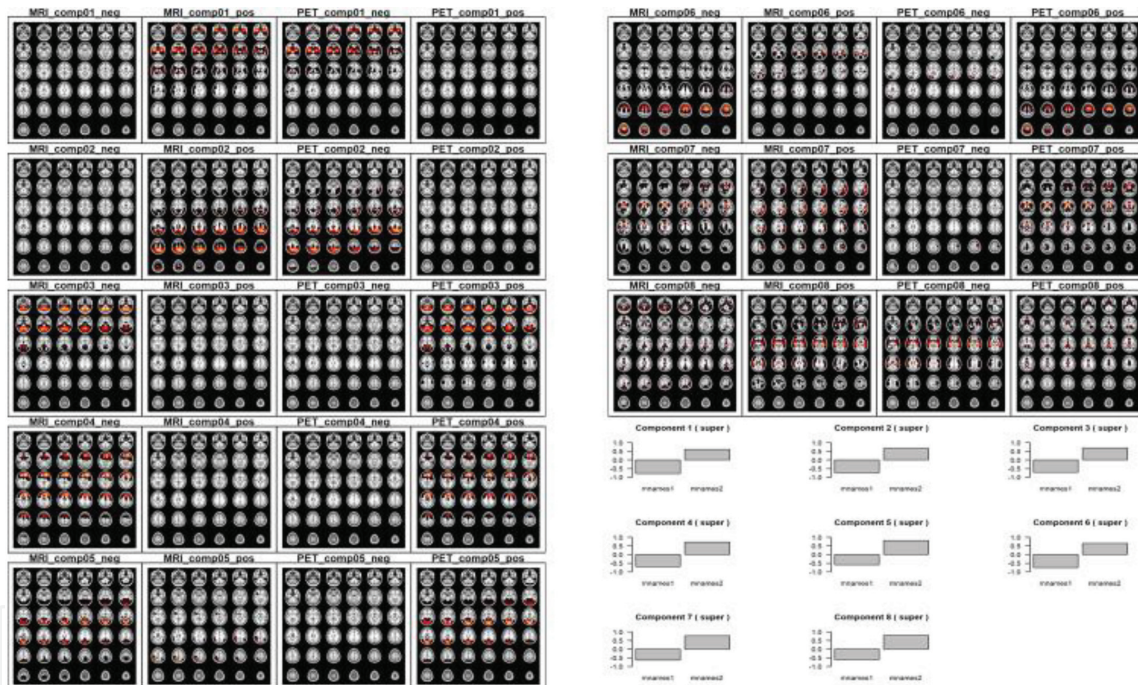
Figure 5.
Sub- and super-weights for all components of $\mu = 0.5$.

four time points were from the baseline for the entire period of the study. Groups were matched for age, gender, and intracranial volume. **Table 4** summarizes the characteristics of these patients at baseline (at the first image observation).

For imaging data processing, we used the VBM8 toolbox. For the basis function, we used four-voxel equal spacing knots, as in the first study in the previous section. The clinical outcome is given by
$Z = 0.44 \times CDR + 0.12 \times ADAS13 - 0.11 \times MMSE$. The coefficients were different from those in the first study because the target population was different.

3.2.2 Results

The original data with dimensions of 2,122,945 (=121 × 145 × 121) was reduced to 7162 using basis functions for each imaging data set. The number of components
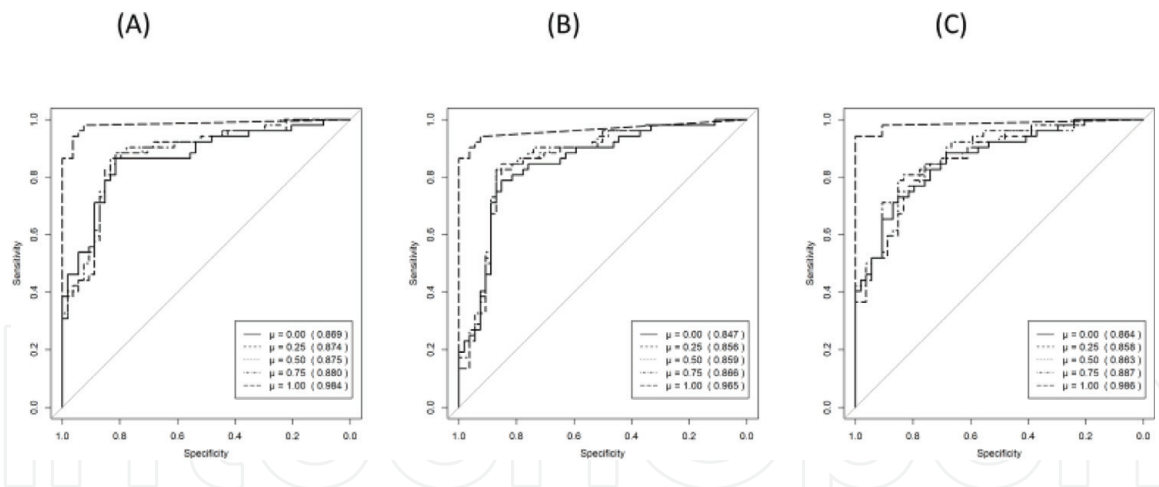
Figure 6.
Results for cross-validated ROC analysis for (A) MRI and PET, (B) MRI, and (C) PET.

	cMCI	sMCI	p
n	34	34	
Age (mean [sd])	76.06 (5.94)	75.91 (5.90)	0.922
PTGENDER = 2 (%)	10 (29.4)	10 (29.4)	1.000
APOE4 (%)			0.040
0	12 (35.3)	22 (64.7)	
1	18 (52.9)	11 (32.4)	
2	4 (11.8)	1 (2.9)	
PTEDUCAT (mean [sd])	16.15 (3.06)	15.50 (2.86)	0.371
CDRSB (mean [sd])	1.76 (1.07)	1.32 (0.73)	0.051
ADAS11 (mean [sd])	12.09 (3.49)	9.40 (4.08)	0.005
ADAS13 (mean [sd])	19.65 (4.31)	15.93 (6.10)	0.005
MMSE (mean [sd])	26.71 (1.71)	27.88 (1.70)	0.006

Table 4.
Characteristic for data set 2.

selected was 6 for $\mu = 0, 0.25, 0.5, 0.75$ and 4 for $\mu = 1$. **Table 5** shows the results for the multiple logistic regression model with cMCI or sMCI as the outcomes and the super scores as the predictors for each μ. The numbers of 5% statistically significant components were 2, 3, 3, 3, and 2 for $\mu = 0, 0.25, 0.5, 0.75, 1$, respectively. The minimum numbers of nonzero subweights were 724, 736, 749, 753, and 852, respectively.

A tenfold cross validated ROC analysis (**Figure 7**) was conducted to evaluate the diagnostic probabilities estimated from the multivariable logistic regression mode whose coefficients and p-values are shown in **Table 5**.

For comparison, the single-modal analysis for each time point was conducted. The fourth time point (MRI4), which is closest to the MCI conversion diagnosis time, had the highest AUC values, and these were higher than the multimodal values (**Figure 8**).

Figure 9 shows the first component subweights, $\boldsymbol{Bw}_m$ ($m = 1, 2, 3, 4$), for the four time points for $\mu = 0$ and 0.5. In the case of $\mu = 0.5$, the hippocampus area was related to the components, and in the case of $\mu = 0$, the parietal lobe was.

	μ = 0.00		μ = 0.25		μ = 0.50		μ = 0.75		μ = 1	
	Estimate	**Pr(>|z|)**	**Estimate**	**Pr(>|z|)**	**Estimate**	**Pr(>|z|)**	**Estimate**	**Pr(>|z|)**	**Estimate**	**Pr(>|z|)**
comp1	−0.0132	0.0139	0.0142	0.0215	0.0142	0.0212	0.0143	0.0204	0.0331	0.0018
comp2	0.0083	0.1714	0.0198	0.0199	0.0095	0.1298	0.0097	0.1240	0.0359	0.0049
comp3	−0.0125	0.1481	−0.0093	0.1410	0.0195	0.0215	0.0196	0.0214	0.0172	0.2833
comp4	−0.0333	0.0042	−0.0073	0.4505	−0.0073	0.4490	−0.0070	0.4709	−0.0157	0.5288
comp5	0.0001	0.9963	0.0458	0.0022	0.0460	0.0021	0.0469	0.0019		
comp6	−0.0203	0.1617	0.0077	0.6205	0.0074	0.6313	0.0078	0.6139		

Table 5.
Results for multivariable logistic regression analysis.
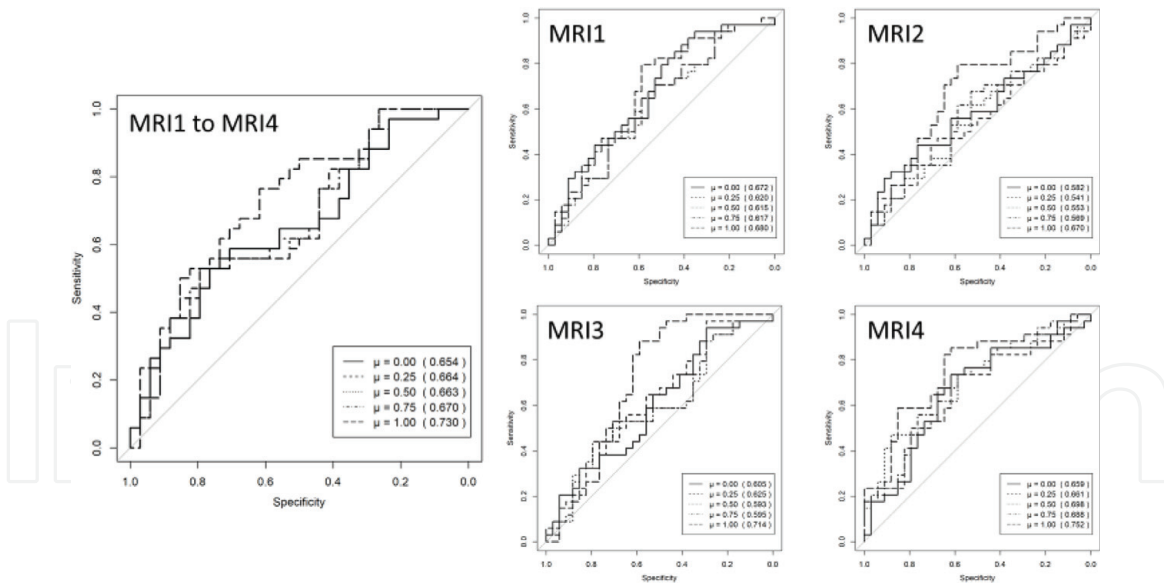
Figure 7.
Results for cross validated ROC analysis.
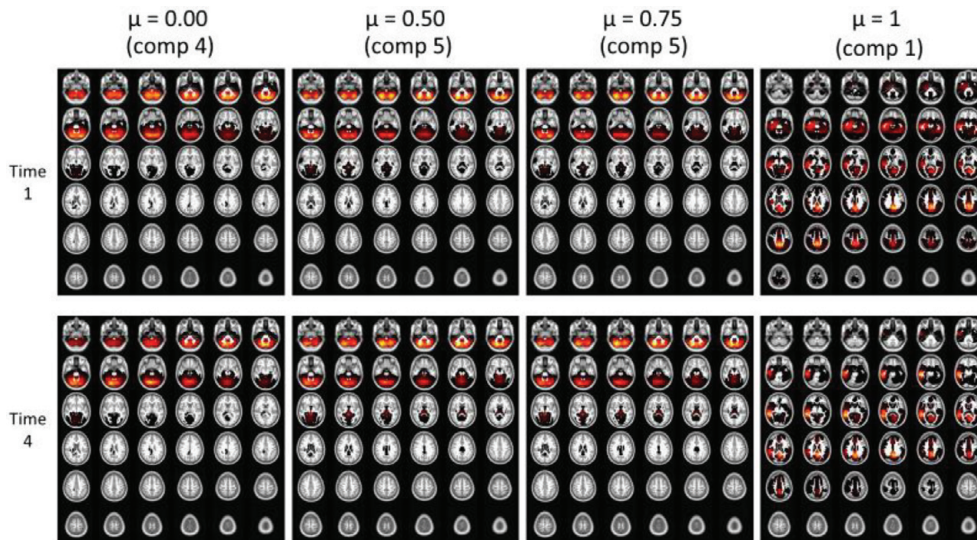
Figure 8.
Subweights for times 1 and 4.
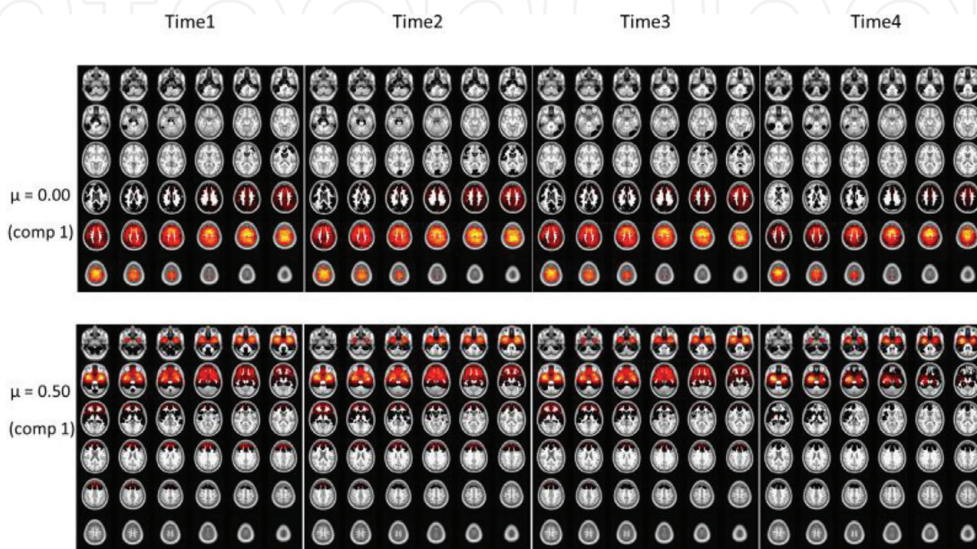
Figure 9.
Subweights for all time points for μ = 0 and 0.5.
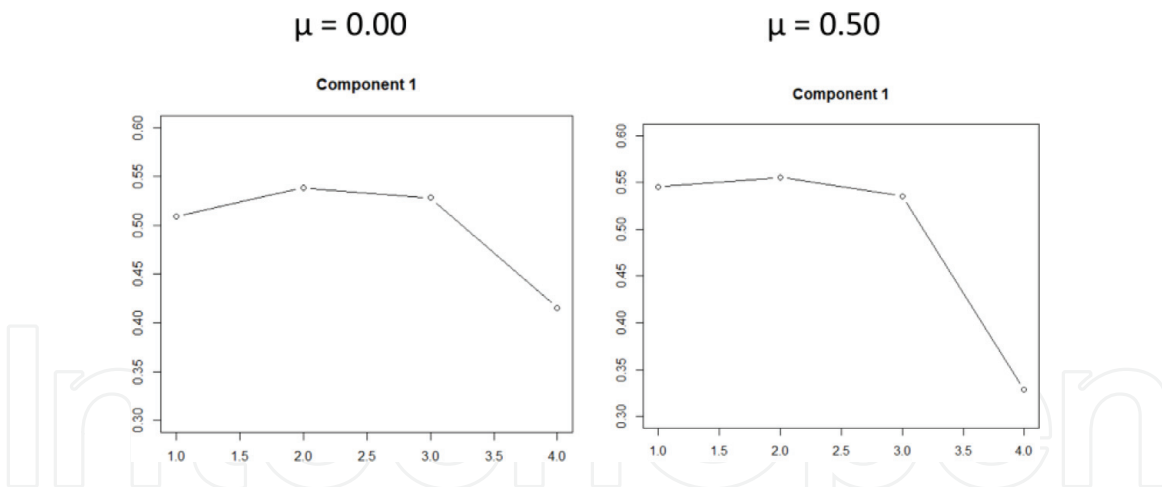
Figure 10.
Super-weights.

Figure 10 shows the corresponding super-weights. This result should be carefully interpreted. For time 4, the sparsest block weights were obtained, and thus the weight values were larger than those of times 1 to 3, which were balanced by the small super-weight. As a result, the super score for this component has the mean value of the block scores.

4. Discussion

In this chapter, the SMS-PCA method was introduced and applied to multiple measured neuroimaging data sets. The first data set consisted of two different types of images, MRI and PET. The second data set consisted of repeated MRI measurements (the same type of image). These imaging data have many voxels per person which were reduced using the basis function prior to conducting the SMS-PCA. The multi-block feature of the SMS-PCA also caused further reduction in each block, and their summary was obtained in the super level where the weights were the relationship and the scores were used in the prediction model.

One of the key features in the SMS-PCA is that it is supervised and its proportion to (self) variance is parametrized by μ. In each study, the impact of μ was studied. The case of $\mu = 1$ resulted in only supervision, that is, only the correlation between the score and the outcome, without the variance of the score. As in an original PCA, maximizing the variance of the score corresponds to $\mu = 0$, and the correlated variables (voxels) have relatively high weights for each component. Thus, the messy maps for the block weights overlaying the brain in the case of $\mu = 1$ were reasonable. In both applications, because $\mu = 0.25, 0.5$, and 0.75 had similar results, a possible large value in $\mu < 1$, or the median value $\mu = 0.5$ with a trade-off, can be selected as optimal.

Repeated measured imaging data analysis was studied in [35] which reduced the imaging dimensions using basis functions but did this independent for each image. In contrast, in this study, the correlation between measurements at different time points is considered. That is, simultaneous temporal and spatial correlation was considered. This approach was limited by the need that the number of images for each individual be the same, and this will be improved in future work. In addition, the method introduced in this chapter can incorporate modalities such as network models which would need to summarize the information into the component. This research is in progress.

5. Conclusion

Although there is room for improvement in this method, this study showed reasonable results when the method was applied to the dementia study. In conclusion, this data-driven approach would be helpful for exploratory neuroimaging data analysis.

Acknowledgements

Conflict of interest

None declared.

Author details

Atsushi Kawaguchi†
Faculty of Medicine, Saga University, Saga, Japan

*Address all correspondence to: akawa@cc.saga-u.ac.jp

†For the Alzheimer's Disease Neuroimaging Initiative

IntechOpen

References

[1] Vandenberghe S, Marsden PK. PET-MRI: A review of challenges and solutions in the development of integrated multimodality imaging. Physics in Medicine and Biology. 2015; **60**:R115-R154

[2] Arbabshirani MR, Plis S, Sui J, Calhoun VD. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. NeuroImage. 2017;**145**:137-165

[3] Rondina JM, Ferreira LK, De Souza Duran FL, et al. Selecting the most relevant brain regions to discriminate Alzheimer's disease patients from healthy controls using multiple kernel learning: A comparison across functional and structural imaging modalities and atlases. NeuroImage: Clinical. 2018;**17**:628-641

[4] Yoshida H, Kawaguchi A, Yamashita F, Tsuruya K. The utility of a network-based clustering method for dimension reduction of imaging and non-imaging biomarkers predictive of Alzheimer's disease. Scientific Reports. 2018;**8**:2807

[5] Josef Golubic S, Aine CJ, Stephen JM, Adair JC, Knoefel JE, Supek S. MEG biomarker of Alzheimer's disease: Absence of a prefrontal generator during auditory sensory gating. Human Brain Mapping. 2017;**38**:5180-5194

[6] Shah C, Zhang W, Xiao Y, et al. Common pattern of gray-matter abnormalities in drug-naive and medicated first-episode schizophrenia: A multimodal meta-analysis. Psychological Medicine. 2017;**47**: 401-413

[7] Liu W, Yang J, Burgunder J, Cheng B, Shang H. Diffusion imaging studies of Huntington's disease: A meta-analysis. Parkinsonism & Related Disorders. 2016;**32**:94-101

[8] Kim YK, Na KS. Application of machine learning classification for structural brain MRI in mood disorders: Critical review from a clinical perspective. Progress in Neuro-Psychopharmacology & Biological Psychiatry. 2018;**80**:71-80

[9] Librenza-Garcia D, Kotzian BJ, Yang J, et al. The impact of machine learning techniques in the study of bipolar disorder: A systematic review. Neuroscience & Biobehavioral Reviews. 2017;**80**:538-554

[10] Moeller SJ, Paulus MP. Toward biomarkers of the addicted human brain: Using neuroimaging to predict relapse and sustained abstinence in substance use disorder. Progress in Neuro-Psychopharmacology & Biological Psychiatry. 2018;**80**:143-154

[11] Moser DA, Doucet GE, Lee WH, et al. Multivariate associations among behavioral, clinical, and multimodal imaging phenotypes in patients with psychosis. JAMA Psychiatry. 2018;**75**: 386-395

[12] O'Donoghue S, Holleran L, Cannon DM, McDonald C. Anatomical dysconnectivity in bipolar disorder compared with schizophrenia: A selective review of structural network analyses using diffusion MRI. Journal of Affective Disorders. 2017;**209**:217-228

[13] Hoogman M, Bralten J, Hibar DP, et al. Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: A cross-sectional mega-analysis. Lancet Psychiatry. 2017; **4**:310-319

[14] Aoki Y, Cortese S, Castellanos FX. Research review: Diffusion tensor imaging studies of attention-deficit/ hyperactivity disorder: Meta-analyses

and reflections on head motion. Journal of Child Psychology and Psychiatry. 2018;**59**:193-202

[15] Li D, Karnath HO, Xu X. Candidate biomarkers in children with autism spectrum disorder: A review of MRI studies. Neuroscience Bulletin. 2017;**33**: 219-237

[16] Hilbert K, Lueken U, Muehlhan M, Beesdo-Baum K. Separating generalized anxiety disorder from major depression using clinical, hormonal, and structural MRI data: A multimodal machine learning study. Brain and Behavior: A Cognitive Neuroscience Perspective. 2017;**7**:e00633

[17] Steiger VR, Brühl AB, Weidt S, et al. Pattern of structural brain changes in social anxiety disorder after cognitive behavioral group therapy: A longitudinal multimodal MRI study. Molecular Psychiatry. 2017;**22**:1164-1171

[18] Schulze L, Schmahl C, Niedtfeld I. Neural correlates of disturbed emotion processing in borderline personality disorder: A multimodal meta-analysis. Biological Psychiatry. 2016;**79**:97-106

[19] Kawaguchi A. Supervised dimension reduction methods for brain tumor image data analysis. In: Matsui S, Crowley J, editors. Frontiers of Biostatistical Methods and Applications in Clinical Oncology. Singapore: Springer; 2017. pp. 401-412

[20] Shen L, Thompson PM, Potkin SG, et al. Genetic analysis of quantitative phenotypes in AD and MCI: Imaging, cognition and biomarkers. Brain Imaging and Behavior. 2014;**8**:183-207

[21] Stephan KE, Schlagenhauf F, Huys QJM, et al. Computational neuroimaging strategies for single patient predictions. NeuroImage. 2017;**145**:180-199

[22] Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review. Frontiers in Aging Neuroscience. 2017;**9**:329

[23] Dimitriadis SI, Liparas D. How random is the random forest? Random forest algorithm on the service of structural imaging biomarkers for Alzheimer's disease: From Alzheimer's disease neuroimaging initiative (ADNI) database. Neural Regeneration Research. 2018;**13**:962-970

[24] Ahmed OB, Benois-Pineau J, Allard M, et al. Recognition of Alzheimer's disease and mild cognitive impairment with multimodal image-derived biomarkers and multiple kernel learning. Neurocomputing. 2017;**220**: 98-110

[25] Vieira S, Pinaya WH, Mechelli A. Using deep learning to investigate the neuroimaging correlates of psychiatric and neurological disorders: Methods and applications. Neuroscience & Biobehavioral Reviews. 2017;**74**:58-75

[26] Shen D, Wu G, Suk HI. Deep learning in medical image analysis. Annual Review of Biomedical Engineering. 2017;**19**:221-248

[27] Janssen RJ, Mourão-miranda J, Schnack HG. Making individual prognoses in psychiatry using neuroimaging and machine learning. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. 2018

[28] Rohart F, Gautier B, Singh A, Lê Cao KA. mixOmics: An R package for 'omics feature selection and multiple data integration. PLoS Computational Biology. 2017;**13**:e1005752

[29] Meng X, Jiang R, Lin D, et al. Predicting individualized clinical measures by a generalized prediction framework and multimodal fusion of MRI data. NeuroImage. 2017;**145**: 218-229

[30] Stokes ME, Visweswaran S. Application of a spatially-weighted relief algorithm for ranking genetic predictors of disease. BioData Mining. 2012;**5**:20

[31] Correa NM, Adali T, Li YO, Calhoun VD. Canonical correlation analysis for data fusion and group inferences: Examining applications of medical imaging data. IEEE Signal Processing Magazine. 2010;**27**:39-50

[32] Sui J, He H, Pearlson GD, et al. Three-way (N-way) fusion of brain imaging data based on mCCA+jICA and its application to discriminating schizophrenia. NeuroImage. 2013;**66**: 119-132

[33] Calhoun VD, Sui J. Multimodal fusion of brain imaging data: A key to finding the missing link(s) in complex mental illness. Biological Psychiatry: Cognitive Neuroscience and Neuroimaging. 2016;**1**:230-244

[34] Kawaguchi A, Yamashita F. Supervised multiblock sparse multivariable analysis with application to multimodal brain imaging genetics. Biostatistics. 2017;**18**:651-665

[35] Kawaguchi A. Diagnostic probability modeling for longitudinal structural brain MRI data analysis. In: Truong KY, editor. Statistical Techniques for Neuroscientists. Boca Raton, Florida: CRC Press; 2016. pp. 361-374