

Supervised multiblock sparse multivariable analysis with application to multimodal brain imaging genetics

ATSUSHI KAWAGUCHI*

*Center for Comprehensive Community Medicine, Faculty of Medicine, Saga University,
5-1-1 Nabeshima, Saga 849-8501, Japan
akawa@cc.saga-u.ac.jp*

FUMIO YAMASHITA

*Division of Ultrahigh Field MRI, Institute for Biomedical Sciences, Iwate Medical University,
Yahaba, Iwate 028-3694, Japan
for the Alzheimer's Disease Neuroimaging Initiative*

SUMMARY

This article proposes a procedure for describing the relationship between high-dimensional data sets, such as multimodal brain images and genetic data. We propose a supervised technique to incorporate the clinical outcome to determine a score, which is a linear combination of variables with hierarchical structures to multimodalities. This approach is expected to obtain interpretable and predictive scores. The proposed method was applied to a study of Alzheimer's disease (AD). We propose a diagnostic method for AD that involves using whole-brain magnetic resonance imaging (MRI) and positron emission tomography (PET), and we select effective brain regions for the diagnostic probability and investigate the genome-wide association with the regions using single nucleotide polymorphisms (SNPs). The two-step dimension reduction method, which we previously introduced, was considered applicable to such a study and allows us to partially incorporate the proposed method. We show that the proposed method offers classification functions with feasibility and reasonable prediction accuracy based on the receiver operating characteristic (ROC) analysis and reasonable regions of the brain and genomes. Our simulation study based on the synthetic structured data set showed that the proposed method outperformed the original method and provided the characteristic for the supervised feature.

Keywords: Neuroscience; Dimension reduction; Basis expansion; Matrix decomposition; Scoring.

1. INTRODUCTION

The analysis of neuroimaging data is useful for the study of brain diseases such as Alzheimer's disease, schizophrenia, and Parkinson's disease. Early studies mainly focused on using a single modality as a biomarker, such as magnetic resonance imaging (MRI). Structural MRI (sMRI) is useful to characterize the brain structure; however, there exist other modalities to characterize the brain condition, for example,

*To whom correspondence should be addressed.

fluorodeoxyglucose positron emission tomography (FDG-PET) as a glucose metabolism, functional MRI (fMRI) as a BOLD response, and diffusion tensor imaging (DTI) as a fiber tract. These techniques provide complementary information about the mechanism of brain disease. Thus, recent research has focused on multimodalities analysis, in which some of these modalities are used simultaneously (Grossman, 2015).

The different role of neuroimaging data in the biomarker study of brain disease is considered as “intermediate phenotypes”. Clinical outcomes (endpoints) for diagnosis are mainly based on diagnoses by medical doctors, cognitive tests, and questionnaires. However, these kinds of clinical outcomes often contain heterogeneity, thereby reducing the power to detect a potential biomarker. This would often occur in a genetic study. As a consequence, the research area “imaging genetics”, in which neuroimaging data are considered as “intermediate phenotypes”, has been growing rapidly with the aim of overcoming these disadvantages. Moreover, the advantage of using imaging data is that it provides rich and objective information rather than clinical outcomes.

Actually, multimodal imaging and genomic data contain potentially useful information about a disease. Especially, AD, the most common form of dementia, is a highly prevalent neurodegenerative disease, which causes the memory and other cognitive functions to decline progressively over time. The detection and diagnosis of the onset and progression of AD in its earliest stages based on a neuroimaging biomarker is invaluable and is the target of intensive research worldwide. A large number of multimodality imaging studies to investigate AD (Teipel *and others*, 2015; Ritter *and others*, 2015) as well as genetic studies (review in Liu and Calhoun, 2014; Shen *and others*, 2014) have been conducted. In addition, Sui *and others* (2012) studied schizophrenia, and Long *and others* (2012) focused on Parkinson’s disease. Our work aims to develop a comprehensive method to analyze data obtained from studies such as these and we refer to these data as muBIG (Multimodal Brain Imaging Genetics) data.

Brain imaging data consist of image intensities of millions of voxels in a 3D array. In the case of sMRI, the value of a voxel represents brain morphometry, and in the case of FDG-PET it represents the brain glucose metabolism. Thus, sMRI is used to evaluate patients’ atrophy, whereas FDG-PET evaluates their hypometabolism. One voxel value, which is the minimum unit in imaging analysis, corresponds to one variable in a statistical term, which means the data is high dimensional. Although region of interest (ROI)-based methods, which are defined anatomically and have often been used to evaluate brain abnormality in earlier studies, can significantly reduce the feature dimensionality and are robust to noise and registration error, ROI-based regional features are generally very coarse and thus not sensitive to detecting small changes related to brain diseases. This led to the consideration of whole brain voxel-based approaches as an alternative with a different approach of dimension reduction. In addition, single nucleotide polymorphism (SNP) is often used for genomic data and in genome-wide association studies. Gene level analysis based on database knowledge can reduce the dimension but lose information as ROI analysis pertaining to brain imaging. Thus, whole-brain and genome-wide imaging genetics on the voxel level and on the SNP level for the purpose of exploratory analysis yields ultra-high-dimensional data sets with relational information.

A rapidly growing area of neuroimaging research as well as genome research by way of high-dimensional data analysis is the use of machine learning and regularized regression procedures (Lin *and others*, 2014). Multimodal studies for neuroimaging data analysis have seen the development of multi-task learning (Zhang *and others*, 2012) and subsequent studies. These methods consider both imaging and genomic data as predictors and would not be appropriate for an exploratory association study between imaging and genomic data. Instead, matrix factorization methods, such as independent component analysis (ICA), canonical component analysis (CCA), partial least squares (PLS), and reduced rank regression (RRR), have been developed because of their applicability and usefulness in association studies between multiple variables. These matrix factorization methods were reviewed by Liu and Calhoun (2014) who pointed out that parallel ICA is a representative method for imaging genetics, but has not been studied to incorporate sparse modeling, which is useful for specifying the associated region and genes for imaging and genomic data, respectively. On the other hand, CCA, PLS, and RRR are applicable to sparse modeling

and are associated with each other; thus, the difference between the results obtained using these techniques would not be significant.

One of the perspectives of matrix factorization is to consider the score, which is a linear combination of the variables and their weights, as determined by certain criteria (correlation is most used). Occasionally, this score is not meaningful because it only uses information based on the correlation; hence, the clinical interpretation of the result would be problematic. Gross and Tibshirani (2015) proposed the supervised CCA method and it was extended by Luo and others (2016) to the case with more than two modalities. These methods, which were not applied to imaging data, considered the pairwise correlation among modalities to have no structure to enable modalities to be distinguished within image and genomic data. The multiblock structure, which has two major (super (or global) and sub (or block)) levels and the sub (block) level corresponds to each modality, would be useful because the super level, which is a linear combination of sub levels, can be used in the prediction score. The other reason for their usefulness is their multivariate fashion against pairwise correlation, and the contribution of sublevels (association between modalities within image phenotype or genotype) on the super level can be evaluated through the loading (weight). As far as we know, the application of the multiblock framework to multimodal imaging analysis has not received much attention. In this regard, we propose a method involving supervised multiblock sparse matrix analysis (SMSMA) in which to incorporate the clinical outcome to determine its direction and to be applicable in the muBIG study.

In our previous study, two-step dimension reduction was studied for imaging analysis from an application perspective (Araki and others, 2013; Yoshida and others, 2013). Two-step dimension reduction was originally proposed by Reiss and Ogden (2010) to functionalize spatial data. The method was first functionalized through the basis expansion of the original data, and the scores were then represented as the linear combination of the reduced data by performing principal component analysis. Because their approach was to use a regression model, the method was regarded as constituting functional data analysis with the regression coefficients being the functional form. In our interpretation for our extension, this method is able to reduce the dimension by taking into account two kinds of correlation. The first is structural correlation, which originates from the data structure, whereas the other is the sample correlation, which is computed from the data values by using a kind of correlation coefficient. Taking brain-imaging analysis as our example, in the first step, taking into account correlation in the form of structural correlation, the dimension is reduced by basis expansion. This would be reasonable because the spatial neighborhood voxels rather than isolated voxels should be carefully considered, as disease-induced brain structural or functional change often happens in local regions. Then, the SMSMA method is applied to reduce the dimension as a second step by taking into account the sample correlation.

This emphasizes that the development of the method presented in this article can be studied with respect to the following aspects: the complementary usage of multimodality, including not only imaging but also genomic data, which are high dimensional and exhibit correlation within and between modality, in addition to the use of selection regions that affect disease and their related SNPs and their usefulness for prediction, that is, as biomarkers.

This article is organized as follows. Section 2 describes the proposed method. In Section 3, the proposed method is applied to real data from the Alzheimer’s Disease Neuroimaging Initiative (ADNI). Section 4 presents a simulation study. Section 5 summarizes and concludes the article.

2. METHODS

In this section, we present the proposed SMSMA method. Consider n subjects and M_X and M_Y blocks (modalities) for the predictor and response, respectively. We assume that modalities are given and discuss the modality selection in Section 5. \mathbf{X}_m is the $n \times p_m$ ($m = 1, 2, \dots, M_X$) and \mathbf{Y}_m is the $n \times q_m$ ($m = 1, 2, \dots, M_Y$) for the m -th block (modality) predictor and response matrix, respectively. Each block

corresponds to the modality, for example, \mathbf{Y}_1 is MRI and \mathbf{Y}_2 is PET. Subjects also have a univariate outcome, and the n -dimensional outcome vector is denoted by \mathbf{Z} . We consider scores \mathbf{t} and \mathbf{u} for the predictor and response, respectively, with the following multiblock structure.

$$\mathbf{t} = \sum_{m=1}^{M_X} b_{X,m} \mathbf{X}_m \mathbf{w}_{X,m}, \quad \mathbf{u} = \sum_{m=1}^{M_Y} b_{Y,m} \mathbf{Y}_m \mathbf{w}_{Y,m}, \quad (2.1)$$

where $\mathbf{w}_{X,m}$ and $\mathbf{w}_{Y,m}$ are weight vectors for the m -th sub-blocks \mathbf{X}_m and \mathbf{Y}_m , respectively, and $b_{X,m}$ and $b_{Y,m}$ are the weights for the super-block. Here, it should be noted that the scores in (2.1) are referred to as the super score, whereas $\mathbf{t}_m = \mathbf{X}_m \mathbf{w}_{X,m}$ and $\mathbf{u}_m = \mathbf{Y}_m \mathbf{w}_{Y,m}$ are referred to as the block score. Thus, the super score, which is used in an application such as a prediction, has a hierarchical structure.

This scoring has a low dimensional representation of not only the data in each block but also the relationship between blocks. Thus, this would be useful for modeling statistically such as construction of the prediction model as in our application by using the multimodality of images and genetics whose importance is discussed in the previous section. We could also evaluate the effects of modalities through $b_{X,m}$ and $b_{Y,m}$. For an application perspective in our case, it is necessary to detect the pair of images and genomic component with the block feature, in which the image is used as marker. Our aim is to determine the associated genomic information with the marker. Thus, the association between two data sets (e.g. images and SNPs) is evaluated by maximizing the correlation between the super scores \mathbf{t} and \mathbf{u} supervised by the outcome. Actually, when matrices \mathbf{X}_m and \mathbf{Y}_m are normalized by their column, the weights $\mathbf{w}_X = (\mathbf{w}_{X,1}, \mathbf{w}_{X,2}, \dots, \mathbf{w}_{X,M_X})^\top$, $\mathbf{w}_Y = (\mathbf{w}_{Y,1}, \mathbf{w}_{Y,2}, \dots, \mathbf{w}_{Y,M_Y})^\top$, $\mathbf{b}_X = (b_{X,1}, b_{X,2}, \dots, b_{X,M_X})^\top$, and $\mathbf{b}_Y = (b_{Y,1}, b_{Y,2}, \dots, b_{Y,M_Y})^\top$ are estimated by maximizing the following function.

$$L(\mathbf{w}_X, \mathbf{w}_Y, \mathbf{b}_X, \mathbf{b}_Y) = \mu_{XY} \mathbf{t}^\top \mathbf{u} + \mu_{XZ} \mathbf{t}^\top \mathbf{Z} + \mu_{YZ} \mathbf{u}^\top \mathbf{Z} - \sum_{m=1}^{M_X} P_{\lambda_{X,m}}(\mathbf{w}_{X,m}) - \sum_{m=1}^{M_Y} P_{\lambda_{Y,m}}(\mathbf{w}_{Y,m}) \quad (2.2)$$

subject to $\|\mathbf{w}_{X,m}\|_2^2 = 1$, $\|\mathbf{w}_{Y,m}\|_2^2 = 1$, for any m , $\|\mathbf{b}_X\|_2^2 = 1$, $\|\mathbf{b}_Y\|_2^2 = 1$, where $0 \leq \mu_{XY}, \mu_{XZ}, \mu_{YZ} \leq 1$ and $\mu_{XY} + \mu_{XZ} + \mu_{YZ} = 1$ is the proportion of the supervision, $P_\lambda(x)$ is the penalty function, ($P_\lambda(x) = 2\lambda|x|$ is used in the application of this article because of the simplicity and the possibility for the sparsity of the weights), and $\lambda > 0$ is the regularized parameter that is used to control the sparsity. Note that it would be possible for the penalty function to have other structures, but these structures are beyond the scope of this article. Note that when $\mu_{XZ} = 0$ and $\mu_{YZ} = 0$ the regularized multiblock PLS or CCA is induced, and when $\mathbf{X}_m = \mathbf{Y}_m$ the supervised regularized PCA framework (regularized PLS) is induced. These could fit flexibly to several applications according to clinical interests. The proposed method is closely related to that of [Li and others \(2012\)](#), which is the unsupervised version and was applied to multi-genomic data analysis without imaging data.

We estimate the weights in (2.1) by stating the following proposition.

PROPOSITION 1 The solution of optimization problem (2.2) satisfies

$$\begin{aligned} \tilde{\mathbf{b}}_{X,m} &= \mathbf{t}_m^\top \{\mu_{XY} \mathbf{u} + \mu_{XZ} \mathbf{Z}\}, \quad \tilde{\mathbf{b}}_{Y,m} = \{\mu_{XY} \mathbf{t}^\top + \mu_{YZ} \mathbf{Z}^\top\} \mathbf{u}_m \\ \tilde{\mathbf{w}}_{X,m} &= h_{\lambda_{X,m}}(b_{X,m} \mathbf{X}_m^\top \{\mu_{XY} \mathbf{u} + \mu_{XZ} \mathbf{Z}\}), \quad \tilde{\mathbf{w}}_{Y,m} = h_{\lambda_{Y,m}}(b_{Y,m} \{\mu_{XY} \mathbf{t}^\top + \mu_{YZ} \mathbf{Z}^\top\} \mathbf{Y}_m), \end{aligned}$$

where $h_\lambda(y) = \text{sign}(y)(|y| - \lambda)_+$ is the sparse function. The proof is shown in the [supplementary materials](#) available at *Biostatistics* online.

This leads to the following algorithm to estimate the weights in (2.1) by maximizing L in (2.2).

1. Initialize \mathbf{t} and \mathbf{u} and normalize the super scores as follows.

$$\mathbf{t} \leftarrow \mathbf{t} / \|\mathbf{t}\|_2, \quad \mathbf{u} \leftarrow \mathbf{u} / \|\mathbf{u}\|_2$$

where \leftarrow means “replaced with”.

2. Repeat until convergence.

(a) For fixed \mathbf{u} , $\tilde{\mathbf{w}}_{X,m} = h_{\lambda_{X,m}}(b_{X,m}\mathbf{X}_m^\top\{\mu_{XY}\mathbf{u} + \mu_{XZ}\mathbf{Z}\})$ and normalize $\hat{\mathbf{w}}_{X,m} = \tilde{\mathbf{w}}_{X,m} / \|\tilde{\mathbf{w}}_{X,m}\|_2$ ($m = 1, 2, \dots, M_X$).

(b) Putting $\mathbf{t}_m = \mathbf{X}_m\hat{\mathbf{w}}_{X,m}$, $\tilde{b}_{X,m} = \mathbf{t}_m^\top\{\mu_{XY}\mathbf{u} + \mu_{XZ}\mathbf{Z}\}$ then putting $\tilde{\mathbf{b}}_X = (\tilde{b}_{X,1}, \tilde{b}_{X,2}, \dots, \tilde{b}_{X,M_X})^\top$ and normalize as $\hat{\mathbf{b}}_X = \tilde{\mathbf{b}}_X / \|\tilde{\mathbf{b}}_X\|_2$

(c) For $\mathbf{t} = \sum_{m=1}^{M_X} b_{X,m}\mathbf{X}_m\hat{\mathbf{w}}_{X,m}$, $\tilde{\mathbf{w}}_{Y,m} = h_{\lambda_{Y,m}}(b_{Y,m}\{\mu_{XY}\mathbf{t}^\top + \mu_{YZ}\mathbf{Z}^\top\}\mathbf{Y}_m)$ then normalize $\hat{\mathbf{w}}_{Y,m} = \tilde{\mathbf{w}}_{Y,m} / \|\tilde{\mathbf{w}}_{Y,m}\|_2$ ($m = 1, 2, \dots, M_Y$).

(d) Putting $\mathbf{u}_m = \mathbf{Y}_m\hat{\mathbf{w}}_{Y,m}$, $\tilde{b}_{Y,m} = \{\mu_{XY}\mathbf{t}^\top + \mu_{YZ}\mathbf{Z}^\top\}\mathbf{u}_m$ then putting $\tilde{\mathbf{b}}_Y = (\tilde{b}_{Y,1}, \tilde{b}_{Y,2}, \dots, \tilde{b}_{Y,M_Y})^\top$ and normalize as $\hat{\mathbf{b}}_Y = \tilde{\mathbf{b}}_Y / \|\tilde{\mathbf{b}}_Y\|_2$.

(e) set $\mathbf{u} = \sum_{m=1}^{M_Y} \hat{b}_{Y,m}\mathbf{u}_m$ where $\hat{b}_{Y,m}$ is the m -th element of $\hat{\mathbf{b}}_Y$.

3. Set $\mathbf{p}_m = \mathbf{X}_m^\top\mathbf{t}_m / \mathbf{t}_m^\top\mathbf{t}_m$ and $\mathbf{q}_m = \mathbf{Y}_m^\top\mathbf{u}_m / \mathbf{u}_m^\top\mathbf{u}_m$, $\mathbf{X}_m \leftarrow \mathbf{X}_m - \mathbf{t}_m\mathbf{p}_m^\top$ and $\mathbf{Y}_m \leftarrow \mathbf{Y}_m - \mathbf{u}_m\mathbf{q}_m^\top$.

Note that the deflation step yields multiple components and has several alternatives. The convergence of this algorithm was supported by the proposition provided in the [supplementary materials](#) available at *Biostatistics* online. The larger value of the regularization parameter $\lambda_{X,m}$ or $\lambda_{Y,m}$ has many non-zero elements in $\mathbf{w}_{X,m}$ or $\mathbf{w}_{Y,m}$. The method for the regularization parameters selection was also provided in the [supplementary materials](#) available at *Biostatistics* online.

3. APPLICATION

In this section, we provide the method we applied to perform the muBIG data analysis. The proposed method is dimension reduced and describes the association between multimodal images and genetics. Our previous work showed that a single dimension reduction was insufficient and this led us to propose the use of two reduction steps. In this section, we use this method with the SMSMA method as a second step ([Araki and others, 2013](#); [Yoshida and others, 2013](#)). We first review the first step briefly and start by providing some notations. Consider the data of n patients with the notation:

$$\{(Z_\alpha, \mathbf{s}_{1\alpha}, \mathbf{s}_{2\alpha}, \mathbf{s}_{3\alpha}); \alpha = 1, \dots, n\},$$

where Z is the scalar clinical outcome. The brain imaging data (sMRI, FDG-PET) is represented by $\mathbf{s}_m = (s_m(\mathbf{v}_1), s_m(\mathbf{v}_2), \dots, s_m(\mathbf{v}_{N_m}))^\top$ ($m = 1, 2$), which is the vectorized image data, where $\mathbf{v}_j \in \mathbb{Z}^3$ ($j = 1, 2, \dots, N_m$) is the voxel location, and N_m is the number of voxels assuming the same number among modalities, that is, $N_1 = N_2$. For $m = 3$, \mathbf{s}_m is the SNP data vector and N_m is the number of SNPs.

3.1. Dimension reduction

In the first phase, the number of dimensions for imaging data is reduced by applying the basis expansion. As the dimensions of the sMRI and FDG-PET data are the same, we consider the same basis function to reduce the dimension to q . $\mathbf{Y}_m = \mathbf{S}_m\mathbf{B}$ is the $n \times q$ matrix with the $n \times N$ matrix $\mathbf{S}_m = (\mathbf{s}_{m,1}, \dots, \mathbf{s}_{m,n})^\top$,

the $N \times q$ matrix $\mathbf{B} = \{\phi_k(\mathbf{v}_j)\}_{j=1,\dots,N,k=1,\dots,q}$ with the element being the radial B-spline function (Saranli and Baykal, 1998)

$$\phi_k(\mathbf{v}) = \frac{1}{4h^2} \begin{cases} h^3 + 3h^2(h - d_k(\mathbf{v})) + 3h(h - d_k(\mathbf{v}))^2 - 3(h - d_k(\mathbf{v}))^3 & (d_k(\mathbf{v}) \leq h) \\ (2h - d_k(\mathbf{v}))^3 & (h < d_k(\mathbf{v}) \leq 2h) \\ 0 & (d_k(\mathbf{v}) > 2h) \end{cases},$$

where $d_k(\mathbf{v}) = \|\mathbf{v} - \boldsymbol{\kappa}_k\|$, $\boldsymbol{\kappa}_k \in \mathbb{Z}^3$ ($k = 1, \dots, q$) is the pre-specified knot, $h > 0$ is the distance between knots, and $\|\cdot\|$ is the Euclidean distance. The basis function reduces the number of parameters from N to q , that is, by converting each voxel to a component of the function. The number of basis functions q is determined by the number of voxels and the distance between pre-specified knots. In this work, we used four-voxel (therefore, $h = \sqrt{3 \times 4^2} = 6.93$) equal spacing knots because our simulation study (Yoshida and others, 2013) showed that the possible smaller distance between knots improved the accuracy. It should be noted that the use of a basis expansion of this nature as a preprocessing procedure for reducing the number of dimensions was useful for neuroimaging analysis as in, for example, Reiss and Ogden (2010), Araki and others (2013), and Yoshida and others (2013). Dimension reduction based on the basis function is followed by the use of the SMSMA method. Thus, the result is obtained as the linear combination of the basis functions $\psi_m(\mathbf{v}) = \sum_{k=1}^q \phi_k(\mathbf{v}) \mathbf{w}_{Y,m}$, which is referred to as a composite function. $\psi_1(\cdot)$ and $\psi_2(\cdot)$ are paired because of the hierarchical structure. This can be regarded as a paired composite function and can be interpreted as representing related regions in the brain between modalities. It would be possible for the composite basis $\psi_m(\mathbf{v})$ to represent a flexible shape, whereas the original basis function $\phi_k(\mathbf{v})$ has a spherical shape.

This method can be applied to SNP data by defining the distance as the physical distance between SNPs. However, the incorporation of the specific biological information required the reduction of the dimension of the SNP data by performing grouping based on the information of linkage disequilibrium (LD), that is, the haplotype block (Gabriel and others, 2002). Thus, in this application, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p)$, where \mathbf{x}_j is the n -dimensional mean of SNPs within the block j .

3.2. Real data analysis

We assessed the validity of the proposed method by applying it to real data. The data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (<http://adni.loni.usc.edu/>). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in our analysis or in writing this report. The ADNI was launched in 2003 as a five-year public-private partnership of several research institutions and private pharmaceutical companies. The ADNI database includes results from various types of imaging examinations and from clinical and neuropsychological assessments; therefore, it is expected that effective results can be obtained by combining data from this source of information to measure the progression of mild cognitive impairment (MCI) and early AD (for the details of study procedures, see <http://adni.loni.usc.edu/methods/documents/>).

For confirmatory purposes, we predicted AD from both 48 normal and 52 AD subjects ($n = 100$ in total) in the ADNI-1 cohort, which has already been studied well. Thus, we could evaluate the proposed method by referencing the existing results. We preprocessed imaging data by using the common software, Statistical Parametric Mapping (SPM). Three-dimensional T1-weighted images of a subject were first segmented into gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) space, and this was followed by anatomical normalization to a template image using DARTEL (Ashburner, 2007). In the case of PET, images were co-registered to the bias-corrected T1 image and the same warping parameter as for the T1 images was applied, following an intensity normalization to the GM for each subject using the same

method that was used for the global normalization in SPM. For SNP, we used PLINK for screening and haplotype block estimation. SNPs with a minor allele frequency = 0.05, the Hardy–Weinberg equilibrium test p.level = 0.001, and call rate per subject = 0.95 were excluded. Median imputation for missing data was applied.

The clinical outcome is given by $Z = 3.17 \times CDRSB + 0.11 \times ADAS13 - 0.57 \times MMSE$, 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 only computed based on the logistic regression model from data with 1684 subjects as a more general population both with and without images, and genetics data not only in the ADNI-1 but also in the ADNI-2 and ADNI-GO cohorts. Note that although the diagnosis information is available, we used this continuous clinical outcome taking into account the circumstances without the diagnosis.

The SMSMA method was applied to data (X, Y_1, Y_2, Z) with parameters $\mu_{XZ} = 0$ and $\mu_{YZ} = 0, 0.2, 0.5, 0.8$ and the number of components was 10. The case of $\mu_{YZ} = 0$ corresponds to the original method (multiblock PLS). As a result of the 1st step of the dimension reduction, the dimensions were 299 336 and 7162 for X and Y’s, respectively. The original dimensions were 549 709 and 2 122 945 ($=121 \times 145 \times 121$) for X and Y’s, respectively.

We evaluated the methods by applying the multiple logistic regression model with the diagnosis (AD or Normal) as a binary response and selected the super scores of Y as predictors by using the Akaike Information Criterion (AIC), that is, $\text{logit Pr}(AD|T) = T\beta$ where T is the matrix of the super score for Y with n rows for patients and the components as the columns. Then, Figure 1 shows the 10-fold cross-validated Receiver Operating Characteristic (ROC) curve and its area under the curve (AUC) based on $T\hat{\beta}$ as risk score, where $\hat{\beta}$ is the maximum likelihood estimator in the training super score set and AD or normal are the binary outcomes in the test dataset. It shows the proposed supervised feature works well and the score of Y in the case of $\mu_{YZ}=0.5$ is most predictive with $AUC=0.952$ (for other cases, as displayed in Figure 1). As in our situation, the dementia status is predicted using image data only and the related SNPs are obtained as supportive information. Thus, using the same $\hat{\beta}$ as in Y, the ROC analysis

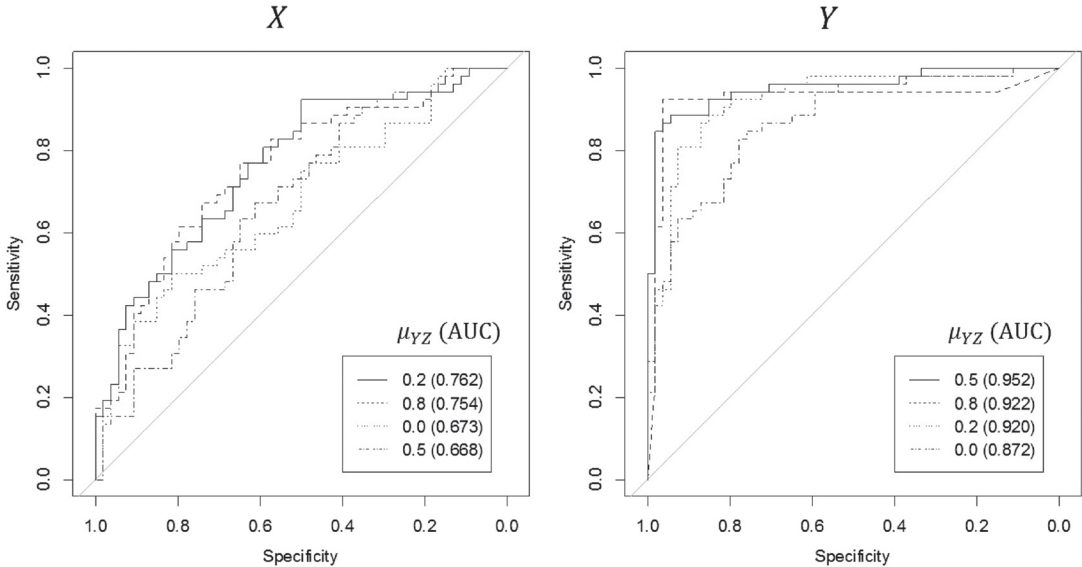


Fig. 1. ROC curves for the super score of X and Y.

Table 1. Non-zero numbers for $\mathbf{w}_{X,1}$, $\mathbf{w}_{Y,1}$, and $\mathbf{w}_{Y,2}$ and the estimates for parameters $b_{Y,1}$ and $b_{Y,2}$. The adjusted variances (the details were provided in the supplementary materials) were multiplied by 100.

Comp	Non-zero numbers			Adjusted variance			Estimates		Logistic regression		
	$\mathbf{w}_{X,1}$	$\mathbf{w}_{Y,1}$	$\mathbf{w}_{Y,2}$	X	Y_1	Y_2	$b_{Y,1}$	$b_{Y,2}$	Est	S.E.	p -value
1	27	1057	1022	1.5	23.6	14.4	-0.654	0.757	0.057	0.015	0.0002
2	11	1028	1140	0.9	8.6	11.5	-0.63	0.777			
3	16	1243	1214	1.3	10.3	10.7	-0.617	0.787	0.097	0.025	0.0001
4	1	1028	995	1.6	3.3	4.2	-0.712	0.702	0.09	0.034	0.0083
5	9	835	768	1.3	2.8	4.1	-0.567	0.823			
6	18	956	836	1.0	4.4	4.0	-0.715	0.699	0.079	0.022	0.0004
7	17	860	830	1.0	3.7	2.1	-0.855	0.519			
8	9	995	965	1.3	1.8	3.1	-0.631	0.775			
9	17	916	765	0.9	1.5	2.0	-0.615	0.788	0.078	0.038	0.0394
10	24	952	677	1.0	1.3	2.5	-0.518	0.855			

Comp, component index; Est, Estimates; S.E., standard error; Z stat, Z statistic.

based on $\mathbf{U}\hat{\boldsymbol{\beta}}$, where \mathbf{U} is the matrix of the super score for X was conducted and the result is shown in the left side of Figure 1, which shows the case of $\mu_{YZ} = 0.2$ is most predictive with $\text{AUC} = 0.762$. The average of the AUCs between X and Y was highest for $\mu_{YZ} = 0.2$. Thus, the details of application result for $\mu_{YZ} = 0.2$ are presented as follows.

Table 1 lists the non-zero numbers for $\mathbf{w}_{X,1}$, $\mathbf{w}_{Y,1}$, and $\mathbf{w}_{Y,2}$ and the estimates for parameters $b_{Y,1}$ and $b_{Y,2}$ in the result of the SMSMA method with $\mu_{YZ} = 0.2$. In the multivariate logistic model, five components were selected and estimated coefficients $\hat{\boldsymbol{\beta}}$, standard errors, Z statistics, and p -values are also presented in Table 1. All selected components by using AIC from the logistic regression model were significant at the 5% level.

$\mathbf{B}\mathbf{w}_{Y,m}\hat{\boldsymbol{\beta}}$ ($m = 1, 2$) is shown in Figure 2 for $\mu_{YZ}=0.2$. The results were divided into two parts according to the sign of $\mathbf{B}\mathbf{w}_{Y,m}\hat{\boldsymbol{\beta}}$. The negative part means atrophy in MRI and hypometabolism in PET. The selected regions show a flexible shape due to the composite basis. For the MRI result, the medial temporal lobe including amygdala and hippocampus, precuneus, and lateral parieto-temporal regions were detected as negative regions, which were well consistent with previous studies (Matsuda, 2013; Whitwell and others, 2012). The positive regions showed complementary results, showing a peak at the boundary of the caudate nucleus, which was considered to reflect the widening of ventricular space. For the PET result, lateral parieto-temporal regions were more prominent than in the MRI result, whereas the weights in the medial temporal regions were lessened. These results were also in good agreement with the previous reports (Jagust and others, 2010; Ishii and others, 1998).

The vector \mathbf{w}_X is expanded to the SNP level from the block level, which is easy to implement by taking the mean value and its absolute values and displaying them as a Manhattan plot in Figure 3. These values represent the association of SNP with the structural and functional regions shown in Figure 2. The vertical dotted line represents the top 10 AD candidate genes listed on the AlzGene database (www.alzgene.org), and the number is the rank.

This application found an association among the top 10 genes, except for ABCA7, CR1, and PICALM. The gene with the highest coefficient was near the BIN1 on chromosome 2, which is a well-validated AD risk gene ranked second in the AlzGene database. There were several genes detected near APOE, the most established AD risk gene (Bertram and others, 2007). Although most likely, our result didn't completely

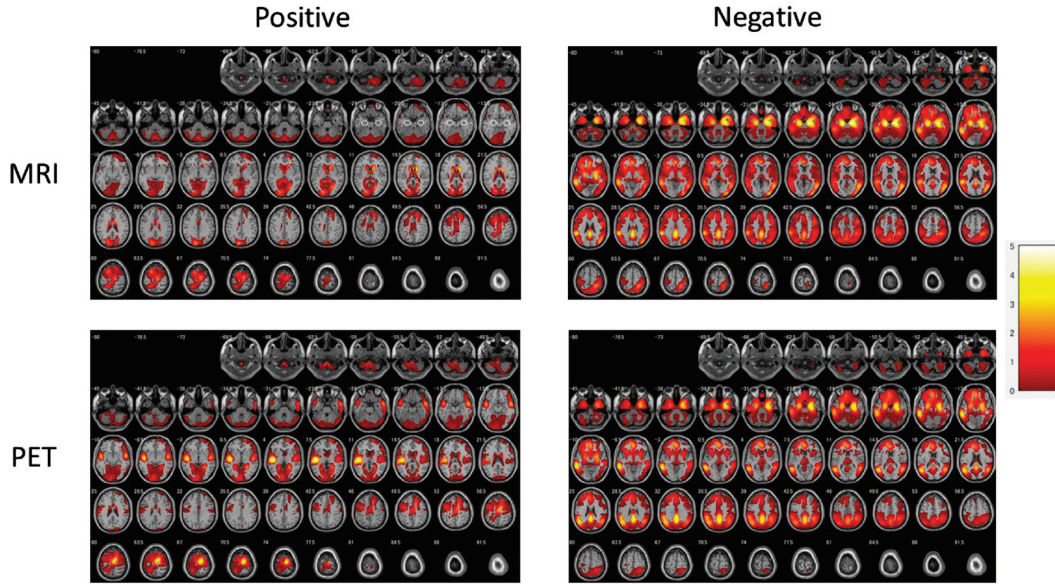


Fig. 2. Result for identified regions of MRI and PET. The circle in the MRI positive represents the hippocampus region.

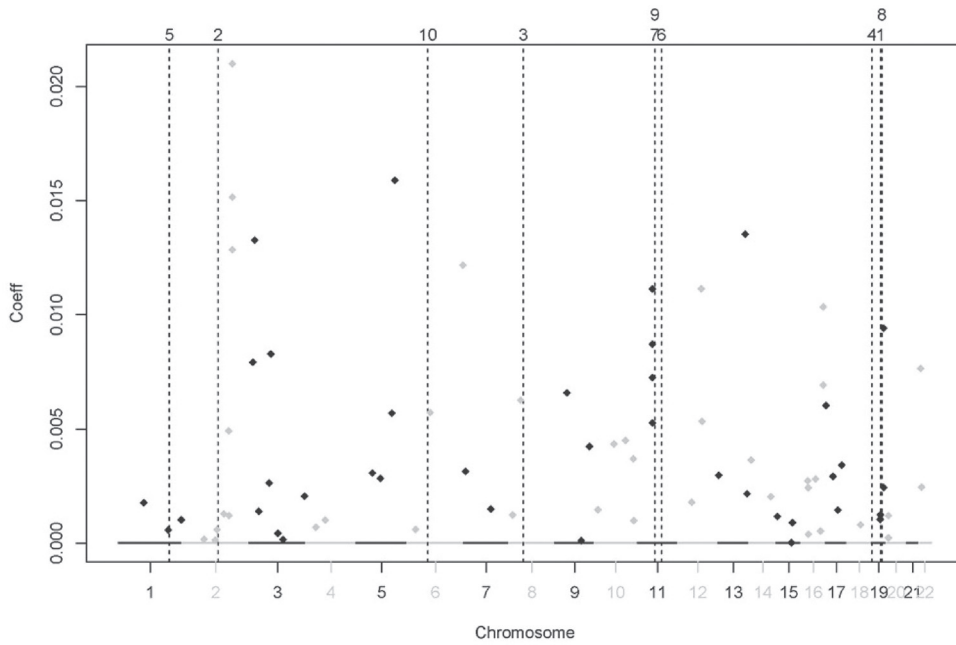


Fig. 3. Manhattan plot. Vertical lines represent the top 10 genes for AD; 1=APOE, 2=BIN1, 3=CLU, 4=ABCA7, 5=CR1, 6=PICALM, 7=MS4A6A, 8=CD33, 9=MS4A4E, 10=CD2AP.

reflect the review, for example, APOE has the highest association. We inferred that the reason originated from the difficulty of GWAS as high dimensional analysis and there was a possible improvement in the first dimension reduction step prior to application of our method. Further discussion is provided in Section 5. Other identified regions were close to the top genes defined in the AlzGene database, CCR2 on chromosome 3, LOC651924 on chromosome 6, and IL33 on chromosome 9. Other SNPs may be close to those identified in previous studies, which are included in the AlzGene database. Further details are not shown here.

4. SIMULATION STUDY

In this section, the characteristics of the proposed method are shown based on the synthetic data. Specially, the aim is to show how the supervised feature works by mainly comparing the original method and impact on the value of μ_{YZ} , simply, mainly for fixed $\mu_{XZ}=0$ and also studied the influence of μ_{XZ} . Note that, as described in Section 2, the case of $\mu_{YZ}=0$ corresponds to the original method.

4.1. Setting

We consider one and two modalities for the predictor and response, respectively, that is, $M_X=1$ and $M_Y=2$. The number of subjects is $n = 50$, and the number of components is 2. Further, \mathbf{t} is the $n \times 2$ score vector for X obtained by generating the multivariate normal random number from $N(0, \mathbf{I}_2)$. The predictor is given by

$$\mathbf{X}_m = \mathbf{t} \mathbf{w}_{X,m}^-$$

with $\mathbf{w}_{X,m} = [\mathbf{I}_2 \otimes \mathbf{1}_{10}, \mathbf{0}_{90,2}]^\top$ being the 100×2 matrix, $\mathbf{1}_a$ is the a -dimensional vector with the element of 1 and $\mathbf{0}_{a,b}$ is the $a \times b$ matrix with the element of 0, and \mathbf{x}^- is the generalized inverse for \mathbf{x} . The relationship between scores are given by $\mathbf{u} = \mathbf{t} \mathbf{w}_u^\top + \boldsymbol{\varepsilon}$, $\mathbf{w}_u = (1, w_{u,0})^\top$, $w_{u,0} = 0.2$ or 0.8 , $\boldsymbol{\varepsilon} \sim N(0, \mathbf{I}_2)$. In the second component, relationships between scores are considered to be either weak or strong. The weight for images $\mathbf{s}_m = (s_m(\mathbf{v}_1), s_m(\mathbf{v}_2), \dots, s_m(\mathbf{v}_N))^\top$ was 100×100 pixels (i.e. $N = 10\,000$) defined as in the left side of Figure 4 or 5 and $\mathbf{w}_{Y,m} = \mathbf{s}_m \mathbf{w}_Y$ with $w_Y=1$ or 5 . The response is given by

$$\mathbf{Y}_m = \mathbf{u} \mathbf{w}_{Y,m}^- + \mathbf{E},$$

where $\mathbf{E} = e_0 \mathbf{E}_0$ by considering two noise levels $e_0=0.001$ and 0.005 and $\mathbf{E}_0 = \{e_{0,i,j}\}$ with $e_{0,i,j} \sim N(0, 1)$ ($i = 1, \dots, n, j = 1, \dots, N$). Thus, the case $w_Y=1$ represents the strong effect of the score on the observed data and the case $w_Y=5$ ($w_Y^{-1}=0.2$) is the weak effect. The outcome is given by $\mathbf{Z} = \mathbf{u} \mathbf{w}_z^\top$ with $w_z = (0, 1)^\top$. Thus, only the score of Y for the second component is associated with the outcome.

The SMSMA method is applied to data $(\mathbf{X}, \mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Z})$ with parameters $\mu_{XZ}=0$ and $\mu_{YZ}=0, 0.2, 0.5, 0.8$ by using two components, and the radial B-spline function with 2-pixel equal spacing knots. The iteration for generating data and applying the method was conducted 100 times for each parameter pair.

4.2. Result

All the resulting weights from the SMSMA method were binarized with the weight being equal (unselected) or not 0 (selected) and the c-index was computed with true images and variables. The c-index is defined by the c-index = sensitivity – (1 – specificity). The average value of the c-index of 50 simulations and blocks by parameters $w_{u,0}$, w_Y , and μ_{YZ} are listed in Table 2 and Table S1 of the [supplementary materials](#) available at *Biostatistics* online. for $e_0=0.001$ and $e_0=0.005$, respectively. Note that to facilitate interpretation, w_Y is displayed as the inverse w_Y^{-1} . The result is divided by X and Y , and their mean is provided in

Table 2. Simulation results for $e_0=0.001$. Mean values of the c-index. The bold fonts represent the maximum value among μ_{YZ} .

Parameters			X			Y		
$w_{u,0}$	w_Y^{-1}	μ_{YZ}	comp1	comp2	mean	comp1	comp2	mean
0.2	0.2	0	0.979	0.803	0.891	0.722	0.114	0.418
		0.2	0.976	0.849	0.912	0.712	0.135	0.424
		0.5	0.959	0.907	0.933	0.663	0.401	0.532
		0.8	0.909	0.896	0.902	0.555	0.915	0.735
	1.0	0	0.971	0.806	0.888	0.880	0.112	0.496
		0.2	0.973	0.849	0.911	0.860	0.164	0.512
		0.5	0.976	0.880	0.928	0.649	0.380	0.515
		0.8	0.92	0.884	0.902	0.720	0.892	0.806
	0.8	0	0.958	0.956	0.957	0.812	0.850	0.831
		0.2	0.959	0.956	0.957	0.807	0.847	0.827
		0.5	0.978	0.961	0.969	0.759	0.736	0.747
		0.8	0.961	0.946	0.953	0.794	0.932	0.863
0.8	1.0	0	0.956	0.959	0.957	0.905	0.869	0.887
		0.2	0.953	0.956	0.954	0.897	0.874	0.885
		0.5	0.972	0.951	0.962	0.754	0.742	0.748
		0.8	0.970	0.954	0.962	0.904	0.898	0.901
	0.2	0	0.958	0.956	0.957	0.812	0.850	0.831

the last three columns. The highest mean value among μ_{YZ} for each $w_{u,0}$ and w_Y is displayed in bold style. Figure 4 shows 2-D displays of the resulting probability for Y for $e_0=0.001$ and $e_0=0.005$, respectively. The probability is computed from the pixel-wise mean value of binary images, which represents whether a pixel is selected or not.

For Table 2, in the case of $w_{u,0} = 0.2$, which means a lower association between the scores u and t of X and Y , respectively, the over performance of the supervised method with $\mu_{YZ} > 0$ against the original method with $\mu_{YZ}=0$ is shown. The difference appeared in the second component of Y, which only has an association with the outcome. For the higher values of μ_{YZ} , the mean c-indexes in the second component of Y were higher than those of $\mu_{YZ}=0$, whereas we observed that for the reverse results for the first component of Y, that is, the mean c-indexes, the higher values of μ_{YZ} were lower than those of $\mu_{YZ}=0$. This might be suitable if the purpose is to extract only the associated components with the outcome. More importantly, the loss of accuracy was also observed in X from both tables. The mean values of the c-index among components of X were the highest except for $\mu_{YZ}=0.8$, which yielded $\mu_{XY}=0.2$ from equation (2.2). This suggests that placing too much emphasis on the supervision of the outcome should be avoided. In the case of $w_{u,0}=0.8$, which indicates a higher association between the scores of X and Y, the difference among μ_{YZ} may not be essential. In that case, we also observed that in the case of $w_{u,0}=0.2$ it was necessary to use the higher value of μ_{YZ} and in the case of $w_{u,0}=0.8$ it was not, although the difference may not be essential. For the noisier case ($e_0=0.005$) in Table S1 of the [supplementary materials](#) available at *Biostatistics* online, the mean values in the strong effect of the score on Y, that is, in the case of $w_Y=1$ the values were generally higher than those of the case of $w_Y=5$, where the effect is weak.

In the cases of $w_{u,0}=0.2$, the detection of true images for the second component in Figure 4 was difficult for the unsupervised version and the lower value of $\mu_{YZ}=0.2$. On the other hand, for the proposed supervised version $\mu_{YZ}=0.5$ or 0.8 . These are mainly because of the data structure of the simulation, where only the second component was associated with the outcome. However, the over-supervised ($\mu_{YZ}=0.8$) resulted in poor reconstruction of the true image for the first component especially in the noisier case (Figure S1 of the [supplementary materials](#) available at *Biostatistics* online, $e_0=0.005$).

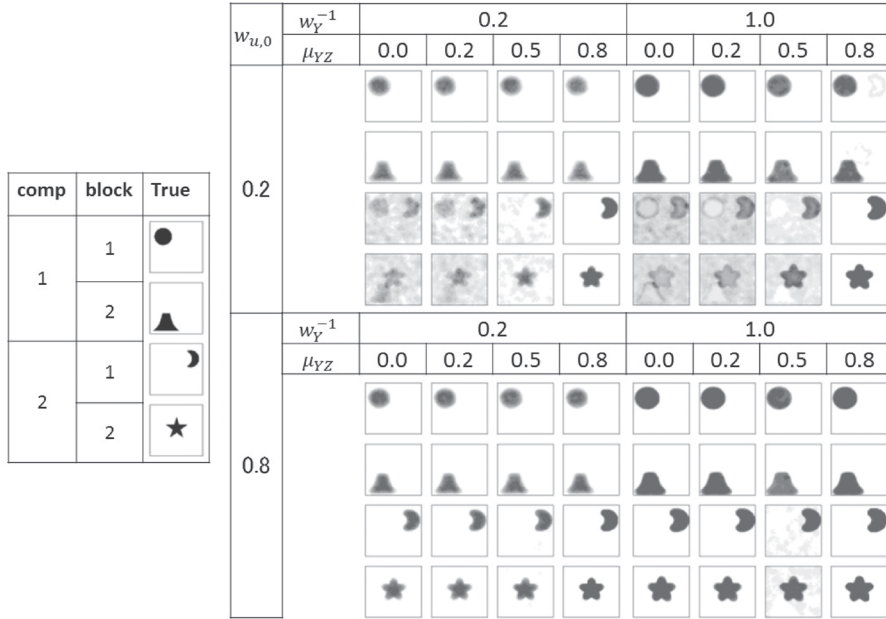


Fig. 4. Simulation results for $e_0=0.001$. Probability map for selected pixels.

We performed an additional simulation to evaluate the local optimization and the estimation with misspecified parameters. Results are shown in the [supplementary materials](#) available at *Biostatistics* online. [Table S2](#) in the [supplementary materials](#) available at *Biostatistics* online showed the maximum number of results having same c-index from the applied method by 100 times with replacing initial values to one data for each parameters. In this case, the large values mean that the algorithm met the global maximization. The balanced case ($\mu_{YZ}=0.5$) presented relatively less global maximization, whereas the case of $\mu_{YZ}=0.8$ showed acceptable convergence although these were relative evaluations because we didn't use the true values. We also provided the proposition about the convergence in the [supplementary materials](#). [Table S3](#) of [supplementary material](#) available at *Biostatistics* online shows the mean c-indices for the method with $\mu_{XZ} > 0$ applied to the same data in [Table 2](#). Because $\mu_{XZ} = 0$ was correct from the simulation setting, general cases were worse but some cases in [Table S3](#) of [supplementary material](#) available at *Biostatistics* online had a similar result as in [Table 2](#). The parameter selection would be important and will be discussed below.

5. DISCUSSION

The work we present in this article involves the development of a method to analyze high-dimensional data sets. From a methodological perspective, we developed a supervised multiblock matrix analysis (SMSMA) method capable of determining the association between block structured multivariable data sets incorporating clinical outcomes. From an application perspective, we developed a method to predict dementia from brain images and to identify associations with SNPs by referring to muBIG data with the following three key features: (i) A paired composite basis function with a data-driven shape, (ii) Data-driven brain regions and selection of SNPs, (iii) Available for brain-wide and genome-wide analysis. The proposed method is capable of constructing a prediction model by using multimodalities imaging as well as genetic information and is expected to generate new knowledge from an exploratory analysis.

The most characteristic aspect of the proposed method is the supervised feature, in which the selection of the tuning parameters μ_{YZ} may be problematic. The best performance differed depending on the simulation scenario, and we suggest addressing this issue by using moderate values, that is, $\mu_{YZ}=0.2$ or 0.5 or by testing for candidates by using an appropriate measure such as predictive accuracy as in our application. In this work, for simplification purposes, μ_{XZ} was fixed as $\mu_{XZ}=0$, even though we could obtain reasonable results in the application of our method. The use of $\mu_{XZ} > 0$ in the method would be helpful if it were necessary to obtain a more predictive score for X. However, over-supervised larger values of μ_{YZ} and μ_{XZ} yield a loss of accuracy in the association between X and Y as shown in our simulation study. Thus, the criteria for selecting μ_{YZ} and μ_{XZ} should evaluate not only the predictivity and the reproducibility, but also their balance. This would have to be further investigated in future.

There are some possible extensions. It would be desirable if the connection of our dimension reduction to the prediction model was to appear in the application. In our application, the scores obtained from our method were given following the construction of the prediction formulation based on the traditional logistic regression model. One extension is to use the regularization version of the logistic model, and the tuning parameters in (2.2) are selected taking into account the generalization in the prediction model, which may yield more sparse weights in Figure 2. The uncertainty of the parameter estimate in the super and block weights could be considered by using the statistical model and be incorporated into the inference in the prediction model. In addition, although cross-validation was used for the evaluation, it may be insufficient for the generalization and can be extended: for example, the usage of completely independent test data which was impossible in our study because of the limited sample size and the normalization of data. As an additional simulation study, local optimization has been evaluated. Although it showed that the convergence was not completed, which would be general in these kinds of high dimensional data analysis methods, this could be solved by using the multiple initialization technique as in [Kawaguchi and Truong \(2011\)](#). The proposed method includes a regularization estimation procedure that enabled us to select the effective brain and genetic regions. In this work, we used the L1 type penalization because of the simplicity, but, it would be straightforward to apply other penalty functions such as SCAD and MC+. These functions will be included in the package and compared in future work. In the application, although we already studied the image data well previously, there may be aspects of the genome side that could be improved based on our result. One of them is to reduce the dimension prior to the application of our method, which is based on the LD block. It would not be sufficient reduction which was a half of the original (549709 to 299336). Not only the number but also the usefulness such as reproductivity would be need to be considered. To achieve this, we need more studies and additional genetic information by collaborating with experts in the future.

We showed one application, which offers the possibility for various application studies; in other words, it would be possible to extend the proposed method in response to application problems. In our application, only the block in Y was considered for imaging data. Although we only used SNP in X, genomic data could also be considered on other levels such as for gene expression, transcriptome, and proteomics. The proposed method could analyze such data by incorporating the blocks into X. For Y, other neuroimaging data, such as fMRI, DTI, and the other PET, could be considered. These selections would depend on the clinical interests and measurability. One of the problems associated with a multimodality study is the acquisition cost and the burden to patients. This concern could be addressed by applying the penalty function at the super level, which would be straightforward. Thus, the modalities considered effective for a particular disease could be selected to lighten the burden imposed on subjects. Another application could be to consider the relationship between images. The neuropathological diagnosis of AD relies on the deposition of amyloid beta ($A\beta$) in the brain tissue. In living patients, this can be measured by the administration of the radiotracer ^{11}C Pittsburgh Compound B (PiB) and subsequent positron emission tomography to assess the spatial pattern of $A\beta$ deposition ([Adlard and others, 2014](#)). Structural brain changes are known to occur after $A\beta$ deposition and are more relevant to the clinical diagnosis of AD.

Thus, a regression model with PiB data as the predictor and MRI data as the response would have to be considered. This work has already been done and reasonable results were obtained.

In conclusion, our method is foreseen to facilitate the construction of a multimodality study with high-dimensional data mainly for the study of dementia as well as other diseases or research (except for medicine such as psychology) by expecting the result to be predictive or interpretable. Further, this kind of study would also be useful from a biomedical informatics perspective.

6. SOFTWARE

The R package `msma` is provided to implement the method described in Section 2 and is available from the Comprehensive R Archive Network (CRAN) at <http://CRAN.R-project.org/package=msma>.

SUPPLEMENTARY MATERIAL

[Supplementary material](http://biostatistics.oxfordjournals.org) is available at <http://biostatistics.oxfordjournals.org>.

ACKNOWLEDGMENTS

The authors are deeply grateful to the referees for constructive comments. The acknowledgment for ADNI data is in the supplementary materials available at *Biostatistics* online.

Conflict of Interest: None declared.

FUNDING

This study was supported in part by Intramural Research Grant (27-8) for Neurological and Psychiatric Disorders of NCNP and a Grant-in-Aid from the Ministry of Education, Culture, Sport, Science and Technology of Japan (24700286). For this research work we used the supercomputer of ACCMS, Kyoto University.

REFERENCES

- ADLARD, P. A., TRAN, B. A., FINKELSTEIN, D. I., DESMOND, P. M., JOHNSTON, L. A., BUSH, A. I. AND EGAN, G. F. (2014). A review of β -amyloid neuroimaging in alzheimer's disease. *Frontiers in Neuroscience* **8**, 327.
- ARAKI, Y., KAWAGUCHI, A. AND YAMASHITA, F. (2013). Regularized logistic discrimination with basis expansions for the early detection of alzheimer's disease based on three-dimensional mri data. *Advances in Data Analysis and Classification* **7**, 109–119.
- ASHBURNER, J. (2007). A fast diffeomorphic image registration algorithm. *Neuroimage* **38**, 95–113.
- BERTRAM, L., MCQUEEN, M. B., MULLIN, K., BLACKER, D. AND TANZI, R. E. (2007). Systematic meta-analyses of alzheimer disease genetic association studies: the alzgene database. *Nature Genetics* **39**, 17–23.
- GABRIEL, S. B., SCHAFFNER, S.F., NGUYEN, H., MOORE, J. M., ROY, J., BLUMENSTIEL, B., HIGGINS, J., DEFELICE, M., LOCHNER, A., FAGGART, M. AND OTHERS. (2002). The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229.
- GROSS, S. M. AND TIBSHIRANI, R. (2015). Collaborative regression. *Biostatistics* **16**, 326.
- GROSSMAN, M. (2015). Integrated multimodal imaging in neurodegenerative disease. *The Lancet Neurology* **14**(10), 973–975.
- ISHII, K., SASAKI, M., YAMAJI, S., SAKAMOTO, S., KITAGAKI, H. AND MORI, E. (1998). Relatively preserved hippocampal glucose metabolism in mild alzheimer's disease. *Dementia and Geriatric Cognitive Disorders* **9**, 317–322.

- JAGUST, W. J., BANDY, D., CHEN, K., FOSTER, N. L., LANDAU, S. M., MATHIS, C. A., PRICE, J. C., REIMAN, E. M., SKOVRONSKY, D., KOEPPE, R. A. AND OTHERS. (2010). The alzheimer's disease neuroimaging initiative positron emission tomography core. *Alzheimer's & Dementia* **6**, 221–229.
- KAWAGUCHI, A. AND TRUONG, K. Y. (2011). Logspline independent component analysis. *Bulletin of Informatics and Cybernetics* **43**, 83–94.
- LI, W., ZHANG, S., LIU, C.-C. AND ZHOU, X. J. (2012). Identifying multi-layer gene regulatory modules from multi-dimensional genomic data. *Bioinformatics* **28**, 2458–2466.
- LIN, D., CAO, H., CALHOUN, V. D. AND WANG, Y.-P. (2014). Sparse models for correlative and integrative analysis of imaging and genetic data. *Journal of neuroscience methods* **237**, 69–78.
- LIU, J. AND CALHOUN, V. D. (2014). A review of multivariate analyses in imaging genetics. *Frontiers in neuroinformatics* **8**, 29.
- LONG, D., WANG, J., XUAN, M., GU, Q., XU, X., KONG, D. AND ZHANG, M. (2012). Automatic classification of early parkinson's disease with multi-modal mr imaging, *PloS one*, **7**(11), e47714.
- LUO, C., LIU, J., DEY, D. K. AND CHEN, K. (2016). Canonical variate regression. *Biostatistics*, **17**(3), 468–483.
- MATSUDA, H. (2013). Voxel-based morphometry of brain mri in normal aging and alzheimer's disease. *Aging and Disease* **4**, 29.
- REISS, P. T. AND OGDEN, R. T. (2010). Functional generalized linear models with images as predictors. *Biometrics* **66**, 61–69.
- RITTER, K., SCHUMACHER, J., WEYGANDT, M., BUCHERT, R., ALLEFELD, C., HAYNES, J.-D., INITIATIVE ALZHEIMER'S DISEASE NEUROIMAGING AND OTHERS. (2015). Multimodal prediction of conversion to alzheimer based on incomplete biomarkers. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* **1**, 206–215.
- SARANLI, A. AND BAYKAL, B. (1998). Complexity reduction in radial basis function (rbf) networks by using radial b-spline functions. *Neurocomputing* **18**, 183–194.
- SHEN, L., THOMPSON, P. M., POTKIN, S. G., BERTRAM, L., FARRER, L. A., FOROUD, T. M., GREEN, R. C., HU, X., HUENTELMAN, M. J., KIM, S. AND OTHERS. (2014). Genetic analysis of quantitative phenotypes in ad and mci: imaging, cognition and biomarkers. *Brain Imaging and Behavior* **8**, 183–207.
- SUI, J., ADALI, T., YU, Q., CHEN, J. AND CALHOUN, V. D. (2012). A review of multivariate methods for multimodal fusion of brain imaging data. *Journal of Neuroscience Methods* **204**, 68–81.
- TEIPEL, S., DRZEZGA, A., GROTHE, M. J., BARTHEL, H., CHÉTELAT, G., SCHUFF, N., SKUDLARSKI, P., CAVEDO, E., FRISONI, G. B., HOFFMANN, W. AND OTHERS. (2015). Multimodal imaging in alzheimer's disease: validity and usefulness for early detection. *The Lancet Neurology* **14**, 1037–1053.
- WHITWELL, J. L., DICKSON, D. W., MURRAY, M. E., WEIGAND, S. D., TOSAKULWONG, N., SENJEM, M. L., KNOPMAN, D. S., BOEVE, B. F., PARISI, J. E., PETERSEN, R. C. AND OTHERS. (2012). Neuroimaging correlates of pathologically defined subtypes of alzheimer's disease: a case-control study. *The Lancet Neurology* **11**, 868–877.
- YOSHIDA, H., KAWAGUCHI, A. AND TSURUYA, K. (2013). Radial basis function-sparse partial least squares for application to brain imaging data. *Computational and Mathematical Methods in Medicine* **2013**, Article ID 591032, 7 pages.
- ZHANG, D., SHEN, D., INITIATIVE ALZHEIMER'S DISEASE NEUROIMAGING AND OTHERS. (2012). Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease. *Neuroimage* **59**, 895–907.

[Received April 18, 2016; revised January 17, 2017; accepted for publication February 6, 2017]