

Integrative analysis of multi-omics and imaging data with incorporation of biological information via structural Bayesian factor analysis

Jingxuan Bao[†], Changgee Chang[†], Qiyiwen Zhang, Andrew J. Saykin, Li Shen, Qi Long and for the Alzheimer's Disease Neuroimaging Initiative

Corresponding authors: Jingxuan Bao, B301 Richards Building, 3700 Hamilton Walk, Philadelphia, PA 19104. Tel: 267-815-4242;

E-mail: jingxuan.bao@penntestmed.upenn.edu; Changgee Chang, 410 W. 10th Street HITS 3000 Indianapolis, IN 46202. E-mail: changgee@iu.edu

[†]Jingxuan Bao and Changgee Chang has equal contribution

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.ucla.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 analysis or writing of this report. A complete listing of ADNI investigators can be found at: (http://adni.loni.usc.edu/wp-content/uploads/how_to_apply/ADNI_Acknowledgement_List.pdf).

Abstract

Motivation: With the rapid development of modern technologies, massive data are available for the systematic study of Alzheimer's disease (AD). Though many existing AD studies mainly focus on single-modality omics data, multi-omics datasets can provide a more comprehensive understanding of AD. To bridge this gap, we proposed a novel structural Bayesian factor analysis framework (SBFA) to extract the information shared by multi-omics data through the aggregation of genotyping data, gene expression data, neuroimaging phenotypes and prior biological network knowledge. Our approach can extract common information shared by different modalities and encourage biologically related features to be selected, guiding future AD research in a biologically meaningful way.

Method: Our SBFA model decomposes the mean parameters of the data into a sparse factor loading matrix and a factor matrix, where the factor matrix represents the common information extracted from multi-omics and imaging data. Our framework is designed to incorporate prior biological network information. Our simulation study demonstrated that our proposed SBFA framework could achieve the best performance compared with the other state-of-the-art factor-analysis-based integrative analysis methods.

Results: We apply our proposed SBFA model together with several state-of-the-art factor analysis models to extract the latent common information from genotyping, gene expression and brain imaging data simultaneously from the ADNI biobank database. The latent information is then used to predict the functional activities questionnaire score, an important measurement for diagnosis of AD quantifying subjects' abilities in daily life. Our SBFA model shows the best prediction performance compared with the other factor analysis models.

Availability: Code are publicly available at <https://github.com/JingxuanBao/SBFA>.

Contact: qlong@upenn.edu

Keywords: structural Bayesian factor analysis, multi-omics, Alzheimer's disease, biological network

Introduction

With the rapid advances in modern biotechnologies such as high-throughput genotyping and sequencing, and multimodal neuroimaging, researchers can investigate disease mechanisms from multiple perspectives simultaneously. Multi-omics data integration enables us to discover new insights into biological mechanism of complex diseases such as cancer and Alzheimer's disease (AD) [1].

AD diagnosis prediction or AD-related cognitive score prediction can help uncover novel therapeutic targets for AD. Many studies seek to predict the AD diagnosis or cognitive scores using only imaging data [2–8] or only genetics data [9]. However, due to the pathological heterogeneity of AD [10, 11], overly relying on one type of data may undermine the power of analyses. To

boost the power of the existing AD research, studying AD from both imaging and genetics/genomics perspectives may improve prediction accuracy and reduce false negative signals [12–17]. In our study, we proposed a novel structural Bayesian factor analysis (SBFA) framework to extract the common information shared by both imaging and genetics/genomics information, which can be used to guide future AD research in a biologically meaningful way.

Factor analysis is a popular statistical method for integrative analysis of multi-modal data. The factor analysis aims to describe variability among high-dimensional variables in terms of a substantially lower number of unobserved variables called latent factors. Many state-of-the-art factor-analysis-based methods have been proposed and applied to multi-omics data [18],

Jingxuan Bao is a Ph.D. student co-advised by Dr. Li Shen and Dr. Qi Long at the University of Pennsylvania Perelman School of Medicine.

Changgee Chang is an assistant professor at Indiana University School of Medicine.

Qiyiwen Zhang is a postdoctoral fellow in Dr. Qi Long's lab at the University of Pennsylvania Perelman School of Medicine.

Andrew J. Saykin is the director of the Center for Neuroimaging and Indiana Alzheimer's Disease Research Center. He is a professor at Indiana University School of Medicine.

Li Shen is a Professor of Informatics at the University of Pennsylvania Perelman School of Medicine.

Qi Long is a Professor of Biostatistics at the University of Pennsylvania.

Received: October 31, 2022. Revised: January 14, 2023. Accepted: February 10, 2023

© The Author(s) 2023. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com

for example to integratively define the subtypes of diseases such as breast cancer [19–23] and lung cancer [19, 20]. The iCluster [19] is a widely used factor analysis framework based on the joint latent variable model. It can flexibly model different types of genomics data in a single framework. It successfully models the tumor subtypes as latent variables integrating the information from multiple datasets. Later, Mo et al. [24] extended the iCluster to the iCluster+ framework by allowing the simultaneous integration of continuous and discrete datasets. Another highly cited integrative factor analysis method based on the non-negative matrix factorization is joint and individual variation explained (JIVE) factor analysis [25], which extends iCluster by adding a data-specific term [21]. A recently published integrative factor analysis method, structural learning and integrative decomposition of multi-view data (SLIDE) [26], represents each data view via shared and individual components. SLIDE has been successfully applied to multiple applications including exploratory dimension reduction, association analysis and consensus clustering.

There has been limited work on integrative analysis of multi-omics and imaging data in a principled modeling framework in AD studies. Some existing AD studies applied factor analysis models trying to define AD subtypes solely in terms of neuropathology [27, 28], molecular signatures [29, 30] or solely through imaging phenotype [31, 32]. Additionally, none of these studies takes prior biological network information into consideration. Network information encourages the selection of related features (e.g. genes belonging to the same pathway) and produces a more biologically interpretable result.

In AD studies, the interactions or associations within or across molecular pathways can be treated as network information. Many databases can provide pathway-level interactions and associations such as Integrative Multi-species Prediction [33] and Kyoto Encyclopedia of Genes and Genomes [34–36]. Some databases can provide tissue-specific interactions and associations such as the HumanBase [37–40]. Several existing studies achieved the incorporation of the biological network information using network-based penalties [41, 42], or proper priors in the Bayesian statistics framework [43–46]. One factor-analysis-based Bayesian integrative study of multi-omics data proposed a generalized Bayesian factor analysis (GBFA) framework by extending the iCluster+ framework [47]. They incorporate the prior biological network information by employing the Markov random field prior. However, one limitation of the GBFA framework is the phase transition problem [48, 49], which may require a massive hyperparameter tuning process.

In this paper, we propose an SBFA model that assigns the Laplace prior to the loading matrix for feature selection. It also assigns a log-normal prior to the shrinkage parameter in the Laplace prior and assigns a graph Laplacian prior [50, 51] to the precision matrix in the log-normal prior. Our model formulation enables smoothing shrinkage parameters for variables that are connected in the biological graph. Our major contributions are summarized as follows:

- We propose a novel structural Bayesian factor analysis framework as a dimension reduction technique that yields latent factors by integrating the genotyping data, gene expression data, region-level brain imaging amyloid deposition data and biological network information. Our SBFA framework is biologically interpretable. As such, the common information shared among multi-omics data learned from our model can be used for various subsequent analyses.

- Our model can handle both continuous (e.g. gene expression and imaging phenotypic traits) and discrete [e.g. genotyping values for single nucleotide polymorphisms (SNPs)] data simultaneously. We successfully incorporate prior biological information without suffering the phase transition problem of the existing GBFA framework [47].
- Our work bridges the gap that there are few multi-omics studies applied to AD. We extracted more biologically meaningful information from multiple modalities, and our framework achieves the best prediction accuracy for the subsequent machine learning prediction problems compared with the other state-of-the-art factor-analysis-based methods.

The rest of the paper is organized as follows. We introduce our proposed method including model formulation and computation in Section 2. We present and discuss the analyses of real data from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) in Section 3. We present our simulation experiment results in Section 4 and conclude the paper with some discussion remarks in Section 5.

Methodology

We propose a hierarchical Structural Bayesian Factor Analysis model named SBFA (Figure 1). Our model serves as a dimension reduction technique to extract common information from multi-modal data.

Notation

All vectors and matrices in this paper are denoted by bold lowercase letters or bold Greek letters. We use unbold lowercase letters or Greek letters to denote the corresponding elements of the vector/matrix. The columns and rows of the matrix are indicated by the subscript. For example, a_i where $i = 1, 2, \dots, n$ denotes the i th element of vector \mathbf{a} of length n ; a_{ji} where $j = 1, 2, \dots, p$ and $i = 1, 2, \dots, n$ denotes the element at the j th row and i th column of the matrix \mathbf{A} with dimension $p \times n$; $\mathbf{a}_i = (a_{1i}, a_{2i}, \dots, a_{pi})^\top$ represents the i th column of matrix \mathbf{A} and the $\bar{\mathbf{a}}_j = (a_{j1}, a_{j2}, \dots, a_{jn})^\top$ represents the j th row of the matrix \mathbf{A} .

Structural Bayesian factor analysis model

Suppose we have multi-modal data with H modalities generated from H different technologies. We denote each data modality by X_h for $h = 1, 2, \dots, H$, where X_h can be any omics data such as genotyping data, gene expression data, DNA methylation data and metabolomics data. Additionally, it can be imaging phenotypic traits such as brain gray matter density, brain volume, protein aggregates and metabolism level in a specific area of the brain. We denote the vertical concatenation of the multi-omics data to be

$$\mathbf{X} = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_H \end{bmatrix} = (x_{ji}) \in \mathbb{R}^{p \times n},$$

where $p = \sum_{h=1}^H p_h$ and p_h is number of features in dataset h , $j = 1, 2, \dots, p$ representing the j th feature, and $i = 1, 2, \dots, n$ representing the i th subject.

We assume that each feature of the dataset x_{ji} has the same underlying distribution family π_j governed by the mean parameter μ_{ji} for each sample i . Under the assumption that every subject is independent of each other and every omic dataset is independent

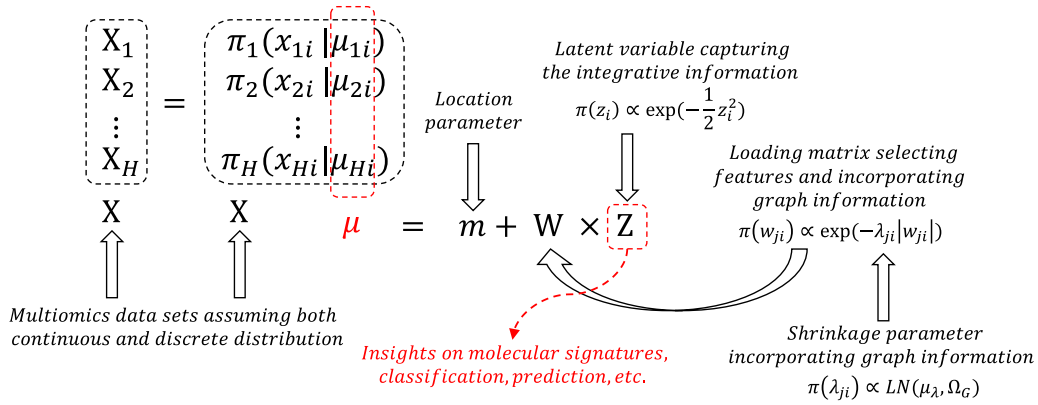


Figure 1. Structural Bayesian factor analysis framework. H represents the number of multi-modal data sets with each dataset denoted as X_h , $h = 1, 2, \dots, H$. The concatenation of all H data sets is denoted by X . We assume both continuous and discrete underlying distribution for each dataset dominated by mean parameter μ_{hi} , $h = 1, 2, \dots, H$; $i = 1, 2, \dots, n$ where n is number of samples. z_i represents the i th column of factor matrix Z where $i = 1, 2, \dots, n$. w_{ji} and λ_{ji} represent the j th row and i th column of loading matrix W and shrinkage parameter for Laplacian prior where $j = 1, 2, \dots, p$; $i = 1, 2, \dots, n$ and p is the total number of features. LN represents the log-normal distribution. μ_λ and Ω_G denote the mean and precision matrix for log-normal prior, respectively.

of each other conditioning on μ , we can write the following likelihood function:

$$\pi(\mathbf{X}|\mu) = \prod_j \prod_i \pi_j(x_{ji}|\mu_{ji}).$$

Our SBFA framework can integrate multi-modal datasets with both continuous and discrete data types simultaneously, including Gaussian data, binomial data and negative binomial data.

- **Gaussian distribution:**
- Continuous x_{ji} is assumed to be normally distributed with mean μ_{ji} and precision ρ_j ,

$$\pi_j(x_{ji}|\mu_{ji}) = \frac{\rho_j^{1/2}}{\sqrt{2\pi}} e^{-\frac{1}{2}\rho_j(x_{ji}-\mu_{ji})^2}, \quad x_{ji} \in \mathbb{R},$$

where a gamma prior $\mathcal{G}(\frac{\xi_j}{2}, \frac{\xi_j}{2})$ will be assigned to ρ_j .

- **Binomial distribution:**
- If the random variable x_{ji} follows a binomial distribution with parameters n_j trials and success probability

$$p_{ji} = \frac{e^{\mu_{ji}}}{1 + e^{\mu_{ji}}},$$

we will use the logit model with probability mass function (pmf) shown as follows:

$$\pi_j(x_{ji}|\mu_{ji}) = \binom{n_j}{x_{ji}} \frac{e^{\mu_{ji}x_{ji}}}{(1 + e^{\mu_{ji}})^{n_j}}, \quad x_{ji} = 0, 1, \dots, n_j. \quad (1)$$

The parameter n_j is determined by data. Bernoulli distribution is a special case for the binomial distribution with n_j to be 1.

- **Negative binomial distribution:**
- To account for over-dispersion in counting data, we accommodate the negative binomial distribution

$$\pi_j(x_{ji}|\mu_{ji}) = \binom{r_j + x_{ji} - 1}{x_{ji}} \frac{e^{\mu_{ji}x_{ji}}}{(1 + e^{\mu_{ji}})^{r_j + x_{ji}}}, \quad x_{ji} \geq 0, \quad (2)$$

where parameter r_j is provided by data. The geometric distribution is a special case for the negative binomial distribution with r_j to be 1.

The mean parameter μ_{ji} are modeled by the generalized factor analysis framework as

$$\mu_{ji} = m_j + \tilde{\mathbf{w}}_j^T \mathbf{z}_i,$$

which, in matrix form, can be written as follows:

$$\mu = \mathbf{m}\mathbf{1}^T + \mathbf{WZ},$$

where $\mathbf{Z} \in \mathbb{R}^{L \times n}$ is the $L \times n$ latent factor shared by all modalities; $\mathbf{W} \in \mathbb{R}^{p \times L}$ is the $p \times L$ factor loading matrix; $\tilde{\mathbf{w}}_j \in \mathbb{R}^{L \times 1}$ is the j th row of \mathbf{W} and $\mathbf{z}_i \in \mathbb{R}^{L \times 1}$ is the i th column of \mathbf{Z} ; $\mathbf{m} \in \mathbb{R}^{p \times 1}$, and $\mathbf{1} \in \mathbb{R}^{n \times 1}$. Our goal is to estimate a biologically meaningful low-rank representation ($L \ll p$) of μ to reduce the dimension of original multi-modal data while preserving the common information shared by multiple modalities. Such dimension reduction can reduce computational expenses and prevent the potential overfitting issue.

\mathbf{Z} represents a low-dimensional latent driver for all H omics and imaging modalities and can be used for subsequent analysis in the applications on AD. For example, it can be used to cluster the subjects to define different AD subtypes. On the other hand, it can also serve as a predictor to predict the diagnosis-related cognitive traits such as the functional assessment questionnaire (FAQ) in the ADNI database, which will be described in Section 3.

Priors

Priors for W and Z

To obtain sparse estimates for \mathbf{W} , we employ the Laplace prior for \mathbf{W}

$$\log \pi(\mathbf{W}|\lambda) = C + \sum_{j=1}^p \sum_{l=1}^L \log \lambda_{jl} - \sum_{j=1}^p \sum_{l=1}^L \lambda_{jl} |w_{jl}|,$$

where λ_{jl} is the shrinkage parameter. The prior distributions for the entries in \mathbf{Z} are the standard Gaussian distribution

$$\log \pi(\mathbf{Z}) = C - \frac{1}{2} \sum_{l=1}^L \sum_{i=1}^n z_{li}^2.$$

Prior for λ : incorporating graphical knowledge

One key contribution of our work is to incorporate biological knowledge $\mathcal{G} = \langle P, E \rangle$ via a structured prior for the shrinkage parameter $\lambda = (\lambda_{jl})_{1 \leq j \leq p, 1 \leq l \leq L}$. In our setting, the graph information for each omics dataset \mathbf{X}_h is represented by $\langle P_h, E_h \rangle$, where $P_h = \{1, \dots, p_h\}$ denotes the nodes and E_h denotes the edges. The presence of an edge indicates that the corresponding pair of nodes are correlated. We denote the \mathbf{G}_h as the adjacency matrix representation of $\langle P_h, E_h \rangle$. We combine the $\langle P_h, E_h \rangle$ to get the graph information for the multiomics data $\langle P, E \rangle$ by diagonally concatenating the adjacency matrix for each omics data \mathbf{X}_h . Now, the graph information of multiomics data can be represented by $\mathcal{G} = \langle P, E \rangle$ where $P = \{1, 2, \dots, p\}$ with $p = \sum_h p_h$ and $E = \{(\mathcal{I}(h, j), \mathcal{I}(h, k)) : (j, k) \in E_h, 1 \leq h \leq H\}$ with $\mathcal{I}(h, j) \in P$ referring to the index of the j th feature of modality h in the concatenated graph \mathcal{G} . Such a setup can help us to achieve the goal that given independent latent factors, the only way a pair of variables can be correlated conditionally on the rest is that they must load at least one common factor. Mathematically, if \mathbf{x}_j and \mathbf{x}_k are connected in \mathbf{G} and $w_{jl} \neq 0$ for some l , then w_{kl} is encouraged to be nonzero. To this end, we employed a log-normal prior to λ_{jl} with graph Laplacian on the precision matrix

$$\log \pi(\boldsymbol{\alpha} | \boldsymbol{\Omega}) = C_{v_2} + \frac{L}{2} \log |\boldsymbol{\Omega}| - \frac{1}{2v_2} \sum_{l=1}^L (\boldsymbol{\alpha}_l - v_1 \mathbf{1})^\top \boldsymbol{\Omega} (\boldsymbol{\alpha}_l - v_1 \mathbf{1}), \quad (3)$$

where scalar v_1 , the mean of $\boldsymbol{\alpha}_l$, is a tuning parameter that controls the overall shrinkage level of \mathbf{W} . Scalar v_2 , variance of $\boldsymbol{\alpha}_l$, controls the extent of individual shrinkage parameters $\alpha_{jl} = \log \lambda_{jl}$ adapted to \mathbf{W} and \mathcal{G} . If $v_2 = 0$, we have $\lambda_{jl} = e^{v_1}$ and the prior reduces to the Bayesian lasso. As v_2 increases, λ_{jl} are more adapted to \mathbf{W} and \mathcal{G} . In our study, we fix the relative variance, known as the adaptivity, to one to allow sufficient variability for λ_{jl} for all shrinkage levels [47], i.e.

$$\text{adaptivity} = \frac{\sigma_\lambda^2}{\mu_\lambda^2} = e^{v_2} - 1 = 1,$$

or equivalently,

$$v_2 = \log 2.$$

The intuition behind the choice of log-normal prior is as follows: (i) we employ the log-normal distribution for λ to incorporate the graph information via correlation and (ii) the shrinkage parameter λ needs to be positive.

Prior for $\boldsymbol{\Omega}$: incorporating biological knowledge

We use the graph Laplacian prior [45] for $\boldsymbol{\Omega}$ to incorporate the graph information \mathcal{G} , which is a symmetric and diagonally dominant matrix

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 + \sum_{j \neq 1} \omega_{1j} & -\omega_{12} & \cdots & -\omega_{1p} \\ -\omega_{21} & 1 + \sum_{j \neq 2} \omega_{2j} & \cdots & -\omega_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ -\omega_{p1} & -\omega_{p2} & \cdots & 1 + \sum_{j \neq p} \omega_{pj} \end{bmatrix},$$

where the elements in matrix $\boldsymbol{\Omega}$, $\boldsymbol{\omega} = \{\omega_{jl} : j < k\}$, are assigned by the following prior:

$$\pi(\boldsymbol{\omega}) \propto |\boldsymbol{\Omega}|^{-\frac{1}{2}} \prod_{\substack{(j,k) \in E \\ 1 \leq j, k \leq p}} \omega_{jk}^{\alpha_\omega - 1} e^{-b_\omega \omega_{jk}} \mathbb{I}(\omega_{jk} > 0) \prod_{\substack{(j,k) \notin E \\ 1 \leq j, k \leq p}} \delta_0(\omega_{jk}),$$

where $\delta_0(\cdot)$ is the Dirac delta function concentrated at 0 and $\mathbb{I}(\cdot)$ is the indicator function. E represents all connected edges in prior biological graph information. In real data application, the graph information can be summarized from gene-gene interaction networks for genetics data and from brain functional pathway information for neuroimaging modality.

There are several mathematical intuitions behind the choice of $\boldsymbol{\Omega}$. First of all, the diagonal dominance of the precision matrix $\boldsymbol{\Omega}$ ensures its positive definiteness. Second, to ensure the positiveness of $\boldsymbol{\omega}$ on the connected pairs, we incorporate a gamma-like distribution on all the connected pairs and 0 on those pairs without an edge. The overall tendency of positiveness of the variance-covariance matrix ensures that increasing ω_{jk} for some $1 \leq j, k \leq p$ encourages an increase of the corresponding correlation.

Model for heterogeneous data types

Our model can handle heterogeneous data types simultaneously. In our framework, we use the exponential family for modeling heterogeneous data types in \mathbf{X} . For example, gene expression and neuroimaging data can be modeled as continuous data with normal distribution. SNP data can be modeled as two Bernoulli random variables where each SNP feature can be represented by the presence of homozygous major allele and the presence of homozygous minor allele, respectively.

With three different underlying data distributions (Gaussian, binomial, negative binomial), we can integrate the discrete exponential family random variables by unifying the likelihood function via the following identity [52]:

$$\frac{e^{\mu_{ji} x_{ji}}}{(1 + e^{\mu_{ji}})^{b_{ji}}} = 2^{-b_{ji}} e^{\kappa_{ji} \mu_{ji}} \int_0^\infty e^{-\rho_{ji} \mu_{ji}^2 / 2} \pi_{ji}(\rho_{ji}) d\rho_{ji},$$

where $\kappa_{ji} = x_{ji} - \frac{b_{ji}}{2}$ and $\pi(\rho_{ji}) = \mathcal{PG}(b_{ji}, 0)$ (Pólya-Gamma distribution). By applying the identity to the binomial (Eq. (1)) and negative binomial (Eq. (2)) likelihood functions, we obtain the unified likelihood function:

$$\pi_j(\tilde{\mathbf{x}}_j | \tilde{\boldsymbol{\mu}}_j) \propto e^{-\frac{1}{2} \sum_i \rho_{ji} (\mu_{ji} - \psi_{ji})^2 + \sum_i \kappa_{ji} \mu_{ji}} \pi_j^*(\tilde{\boldsymbol{\rho}}_j),$$

where the unknown parameters are shown in Table 1. Now, the full posterior density can be written as

$$\pi(\boldsymbol{\lambda}, \boldsymbol{\rho}, \mathbf{W}, \mathbf{Z}, \mathbf{X}) = \pi(\mathbf{W} | \boldsymbol{\lambda}) \pi(\log \boldsymbol{\lambda} | \boldsymbol{\Omega}) \pi(\boldsymbol{\omega}) \pi(\mathbf{Z}) \cdot \prod_j e^{-\frac{1}{2} \sum_i \rho_{ji} (\mu_{ji} - \psi_{ji})^2 + \sum_i \kappa_{ji} \mu_{ji}} \pi_j^*(\tilde{\boldsymbol{\rho}}_j),$$

where $\boldsymbol{\omega} = \{\omega_{jk}\}_{1 \leq j < k \leq p}$.

Table 1. Parameters for unified likelihood.

Data Type	ψ_{ji}	κ_{ji}	b_{ji}	$\pi_j^*(\rho_j)$
Gaussian	X_{ji}	0	NA	$\rho_j = \rho_j \sim \mathcal{G}(\frac{\xi_j+n}{2}, \frac{\xi_j}{2})$
Binomial	0	$X_{ji} - \frac{n_j}{2}$	n_j	$\rho_j \sim \mathcal{PG}(n_j, 0)$
Negative Binomial	0	$\frac{X_{ji}-r_j}{2}$	$X_{ji} + r_j$	$\rho_j \sim \mathcal{PG}(X_{ji} + r_j, 0)$

Feature representation

In our model, new feature representation is learned from genotyping, gene expression and imaging datasets with biological graph information. The high dimensional information is summarized into the low dimensional latent variable \mathbf{Z} for subsequent analyses.

For example, our SBFA framework estimated low dimensional latent features, which can serve as a predictor for machine learning algorithms. For unsupervised learning such as clustering, the low dimensional features derived from our algorithm are considered to be less noisy. We can apply the existing clustering algorithm to cluster the subjects based on the new feature representation. Supervised learning such as prediction often suffers from overfitting with high dimensional input data. With the low dimension feature representation derived from our algorithm, we can reduce the risk of overfitting. Compared with the other dimension reduction techniques, our SBFA is more biologically interpretable.

Parameter estimation and computation

We estimate the factor loading matrix \mathbf{W} and the location parameter \mathbf{m} by finding the mode of their marginal posterior density, i.e. the maximum a posteriori (MAP) estimator. Specifically, denote the parameters to be estimated by $\Theta = (\mathbf{W}, \mathbf{m})$ and the nuisance parameters by $\Xi = (\alpha, \rho, \mathbf{Z}, \omega)$. Then, the MAP estimator for \mathbf{W} and \mathbf{m} is defined as

$$\hat{\Theta} = \operatorname{argmax}_{\Theta} \int \pi(\Theta, \Xi) d\Xi,$$

where $\pi(\Theta, \Xi)$ is the (joint) posterior density. To find the MAP, we could utilize the expectation maximization (EM) algorithm [53], which iteratively maximizes conditional expectations as follows:

$$\begin{aligned} Q^{(t)}(\Theta) &= \mathbb{E}(\log \pi(\Theta^{(t)}, \Xi) | \Theta^{(t)}), \\ \Theta^{(t+1)} &= \operatorname{argmax}_{\Theta} Q^{(t)}(\Theta), \end{aligned}$$

where \mathbb{E} denotes the expectation with respect to the conditional distribution $\pi(\Xi | \Theta^{(t)})$.

Unfortunately, the conditional expectation of $\log \pi(\Theta, \Xi)$ in our model does not have a closed-form solution. Therefore, we choose to use the variational EM algorithm [54, 55], where the conditional distribution $\pi(\Xi | \Theta)$ is approximated by a more tractable distribution family $\hat{\pi}(\Xi)$ parameterized by Υ . The variational parameter Υ can be determined by minimizing the Kullback–Leibler (KL) divergence

$$\Upsilon^{(t)} = \operatorname{argmin}_{\Upsilon} \left(-\hat{\mathbb{E}} \log \pi(\Xi | \Theta^{(t)}) + \hat{\mathbb{E}} \log \hat{\pi}(\Xi) \right),$$

where $\hat{\mathbb{E}}$ denotes the expectation under $\hat{\pi}(\Xi)$. Thus, the variational EM is equivalent to maximizing

$$Q(\Theta, \Upsilon) = \hat{\mathbb{E}} \log \pi(\Theta, \Xi) - \hat{\mathbb{E}} \log \hat{\pi}(\Xi),$$

which is called the evidence lower bound (ELBO) [55] as it always satisfies $Q(\Theta, \Upsilon) \leq \log \pi(\Theta)$, with respect to Θ and Υ .

In our model, we use a product measure on $\alpha, \rho, \mathbf{Z}, \omega$ as the variational (approximating) measure, i.e. $\hat{\pi}(\alpha, \rho, \mathbf{Z}, \omega) = \hat{\pi}(\alpha) \hat{\pi}(\rho) \hat{\pi}(\mathbf{Z}) \hat{\pi}(\omega)$ where

$$\begin{aligned} \hat{\pi}(\alpha) &\propto \prod_{l=1}^L e^{-\frac{1}{2}(\alpha_l - \mu_{\alpha,l})^T \Sigma_{\alpha,l}^{-1}(\alpha_l - \mu_{\alpha,l})} \\ \hat{\pi}(\omega) &\propto \prod_{j=1}^p \prod_{k=j+1}^p w_{jk}^{a_{jk}-1} e^{-b_{jk} w_{jk}} \\ \hat{\pi}(\rho) &\propto \prod_{j=1}^p e^{-\frac{1}{2} \sum_{i=1}^n \rho_j \varphi_{ji}^2} \pi_j^*(\tilde{\rho}_j) \\ \hat{\pi}(\mathbf{Z}) &\propto \prod_{i=1}^n e^{-\frac{1}{2}(\mathbf{z}_i - \mu_{z,i})^T \Sigma_{z,i}^{-1}(\mathbf{z}_i - \mu_{z,i})}, \end{aligned}$$

where a_{jk} 's, b_{jk} 's, $\mu_{\alpha,l}$'s, $\Sigma_{\alpha,l}$'s, φ_{ji} 's, $\mu_{z,i}$'s and $\Sigma_{z,i}$'s are variational parameters. Note that $\hat{\pi}(\rho)$ is called tilted Pólya-Gamma distribution.

E-step for ω : We update the variational parameters a_{jk} and b_{jk} by matching the form of the equations between the expectation for log variational measure and the expectation for log posterior measure

$$\begin{aligned} a_{jk} &= a_{\omega} \\ b_{jk} &= b_{\omega} + \frac{1}{2\nu_2} \sum_{l=1}^L \left(\Sigma_{\alpha,l(k,k)} - 2\Sigma_{\alpha,l(j,k)} + \Sigma_{\alpha,l(j,j)} \right) \\ &\quad + \frac{1}{2\nu_2} \sum_{l=1}^L \left(\mu_{\alpha,lk}^2 - 2\mu_{\alpha,lj}\mu_{\alpha,lk} + \mu_{\alpha,lj}^2 \right), \end{aligned}$$

where $j = 1, 2, \dots, p$ and $k = 1, 2, \dots, p$; $\mu_{\alpha,lk}$ is the mean of the variational measure for the k th element of α_l ; $\Sigma_{\alpha,l(j,k)}$ is the i th row and j th column of the covariance matrix of the variational measure for α_l .

We then update the expectation of the ω using the mean formula for gamma distribution by

$$\mu_{\omega,jk} = \frac{a_{jk}}{b_{jk}},$$

for $j = 1, 2, \dots, p$ and $k = 1, 2, \dots, p$. Now, we have a graph Laplacian matrix

$$\mathbf{\Omega} = \begin{bmatrix} 1 + \sum_{j \neq 1} \mu_{\omega,1j} & -\mu_{\omega,12} & \cdots & -\mu_{\omega,1p} \\ -\mu_{\omega,21} & 1 + \sum_{j \neq 2} \mu_{\omega,2j} & \cdots & -\mu_{\omega,2p} \\ \vdots & \vdots & \ddots & \vdots \\ -\mu_{\omega,p1} & -\mu_{\omega,p2} & \cdots & 1 + \sum_{j \neq p} \mu_{\omega,pj} \end{bmatrix}.$$

E-Step for α : Since there exists an $e^{\alpha_{jl}}$ term in the log posterior distribution but not in the log variational measure, we cannot update the variational parameters by matching the form of equations as we did when updating the ω (E-step for ω). Instead, we need to maximize the ELBO function and the results give us that

$$\mu_{\alpha,kl} = -W \left(\frac{d_{kl}}{r_{kl}} e^{-\frac{c_{kl}}{r_{kl}}} \right) - \frac{c_{kl}}{r_{kl}},$$

where $k = 1, 2, \dots, p$, $l = 1, 2, \dots, L$, W represents the Lambert W function [56] and

$$\begin{aligned} d_{kl} &= |w_{kl}| e^{\frac{1}{2} \text{diag}(\mathbf{\Sigma}_{\alpha,l})_k}; \\ r_{kl} &= \frac{1}{v_2} \left(1 + \sum_{\substack{j=1 \\ j \neq k}}^p \mu_{\omega,kj} \right); \\ c_{kl} &= -\frac{1}{v_2} \sum_{\substack{j=1 \\ j \neq k}}^p \mu_{\omega,kj} \mu_{\alpha,jl} - \frac{v_1 + v_2}{v_2}, \end{aligned}$$

and $\text{diag}(\cdot)$ is an operator taking the diagonal elements of a square matrix as a column vector; w_{kl} represents the k th row and l th column element of the matrix \mathbf{W} ; μ_{α} represents the mean matrix for α ; $\mathbf{\Sigma}_{\alpha,l}$ represents the covariance matrix for l th column of α ; $\text{diag}(\mathbf{\Sigma}_{\alpha,l})_k$ represents the k th element of the vector $\text{diag}(\mathbf{\Sigma}_{\alpha,l})$.

The update rule for the covariance matrix of α , i.e., $\mathbf{\Sigma}_{\alpha,l}$, is separated into two parts. In the first part, we update the off-diagonal element of the $\mathbf{\Sigma}_{\alpha,l}$ by

$$\mathbf{\Sigma}_{\alpha,l(-j,j)} = C_j \mathbf{\Sigma}_{\alpha,l(-j,-j)} \mathbf{\Omega}_{(-j,j)} \quad (4)$$

for each $j = 1, 2, \dots, p$, where

$$C_j = \frac{v_2 - \sqrt{v_2^2 + 4 \text{tr}(\mathbf{\Sigma}_{\alpha,l(-j,-j)} \mathbf{\Omega}_{(j,-j)} \mathbf{\Omega}_{(-j,j)}) \mathbf{\Sigma}_{\alpha,l(j,j)}}}{2 \text{tr}(\mathbf{\Sigma}_{\alpha,l(-j,-j)} \mathbf{\Omega}_{(j,-j)} \mathbf{\Omega}_{(-j,j)})}.$$

The $\text{tr}(\cdot)$ operator defines the trace of the matrix, and the minus sign in the subscript represents the matrix with the corresponding row or column removed, or vector with the corresponding element removed.

In the second part, we update the diagonal element of the $\mathbf{\Sigma}_{\alpha,l}$ as follows:

$$\mathbf{\Sigma}_{\alpha,l(j,j)} = t_j + \frac{1}{c_j} W_{r_j e^{-\zeta_j t_j}} (c_j a_j e^{-\zeta_j t_j}) \quad (5)$$

for each $j = 1, 2, \dots, p$, where the r -Lambert function W_r is defined as the inverse of the function $x e^x + rx$ [57] and

$$\begin{aligned} c_j &= \frac{1}{2} \\ t_j &= \mathbf{\Sigma}_{\alpha,l(j,-j)} \mathbf{\Sigma}_{\alpha,l(-j,-j)}^{-1} \mathbf{\Sigma}_{\alpha,l(-j,j)} \\ r_j &= \frac{\mathbf{\Omega}_{(j,j)}}{v_2 |w_{jl}| e^{\mu_{\alpha,jl}}} \\ a_j &= \frac{1}{|w_{jl}| e^{\mu_{\alpha,jl}}}, \end{aligned}$$

where \cdot^{-1} is the inverse operator for a square matrix; w_{jl} is the j th row and l th column of the loading matrix \mathbf{W} . We iteratively update $\mathbf{\Sigma}_{\alpha,l}$ for each $j = 1, 2, \dots, p$ and $l = 1, 2, \dots, L$.

E-step for \mathbf{z} : We update the precision matrix of the i th column of factor matrix \mathbf{Z} , i.e. $\mathbf{\Sigma}_{z,i}^{-1}$, by matching the form of equations between log posterior distribution and the log variational measure. For each $i = 1, 2, \dots, n$, we have

$$\mathbf{\Sigma}_{z,i}^{-1} = \mathbf{I} + \sum_{j=1}^p \mu_{\rho_j} \tilde{\mathbf{w}}_j \tilde{\mathbf{w}}_j^T,$$

where \mathbf{I} is identity matrix; μ_{ρ_j} is the mean for parameter ρ_j and $\tilde{\mathbf{w}}_j$ is the j th row of loading matrix \mathbf{W} . For each $i = 1, 2, \dots, n$, we update the mean parameter for \mathbf{z}_i by

$$\mu_{z,i} = \mathbf{\Sigma}_{z,i} \sum_{j=1}^p (\mu_{\rho_j} \psi_{ji} + \kappa_{ji} - \mu_{\rho_j} m_j) \tilde{\mathbf{w}}_j.$$

E-step for ρ : By matching the form of expectation of log posterior and expectation of log variational measure, the variational parameter φ_{ji} for each $j = 1, 2, \dots, p$ and $i = 1, 2, \dots, n$ for ρ can be updated by

$$\varphi_{ji} = \left(\tilde{\mathbf{w}}_j^T \mathbf{\Sigma}_{z,i} \tilde{\mathbf{w}}_j + (m_j + \tilde{\mathbf{w}}_j^T \mu_{z,i} - \psi_{ji})^2 \right)^{\frac{1}{2}},$$

where $\tilde{\mathbf{w}}_j$ is the j th row of loading matrix \mathbf{W} and $\mathbf{\Sigma}_{z,i}$ is the variance-covariance matrix for variational measure of \mathbf{z} .

If \mathbf{X} is a discrete random variable, we have

$$\mu_{\rho_j} = \frac{b_{ji}(e^{\varphi_{ji}} - 1)}{2\varphi_{ji}(e^{\varphi_{ji}} + 1)},$$

whereas if \mathbf{X} follows Gaussian distribution, we have

$$\mu_{\rho_j} = \frac{\zeta_j + n}{\zeta_j + \sum_{i=1}^n \varphi_{ji}^2}.$$

M-step for $\tilde{\mathbf{w}}$: We update the loading matrix row by row via maximizing the ELBO function with respect to $\tilde{\mathbf{w}}_j$ for each $j = 1, 2, \dots, p$. The optimization problem is solved by the dynamic weighted lasso algorithm [58].

M-step for \mathbf{m} : The location parameter \mathbf{m} is updated by maximizing the ELBO function with respect to m_j for each $j = 1, 2, \dots, p$

$$m_j = \frac{1}{\sum_{i=1}^n \mu_{\rho_j}} \left(\sum_{i=1}^n \kappa_{ji} - \sum_{i=1}^n \mu_{\rho_j} \tilde{\mathbf{w}}_j^T \mu_{z,i} + \sum_{i=1}^n \mu_{\rho_j} \psi_{ji} \right).$$

Table 2. Participant characteristics. Total number of subjects, age and sex are shown in this table. The mean \pm sd for the age of all subjects within each diagnosis group is reported. The number of male/female subjects within each diagnosis group is also shown.

Diagnosis	GN	EMCI	LMCI	AD	Overall
Number	199	198	165	42	604
Age (mean \pm sd)	74.54 \pm 5.50	71.27 \pm 7.22	73.14 \pm 7.16	75.45 \pm 9.43	73.15 \pm 7.01
Sex (M/F)	98/101	109/89	101/64	26/16	334/270

Initialization

For the initialization, we define the matrix \mathbf{Y} as an approximation of the mean matrix $\boldsymbol{\mu}$ as follows [47]:

$$y_{ji} = \begin{cases} x_{ji}, & \text{for } x_{ji} \sim \text{Gaussian}; \\ \text{logit}\left(\frac{x_{ji}+1}{n_j+2}\right), & \text{for } x_{ji} \sim \text{Binomial}; \\ \text{logit}\left(\frac{x_{ji}+1}{r_j+x_{ji}+2}\right), & \text{for } x_{ji} \sim \text{Negative Binomial}. \end{cases}$$

The location vector \mathbf{m} can be initialized by a fixed pre-specified constant vector. If not, for each $j = 1, 2, \dots, p$, m_j can be initialized by the trimmed mean, median or mean of $\hat{\mathbf{y}}_j$.

In our setting, we handle the problems by choosing \mathbf{W} and \mathbf{Z} matrices close to orthogonal matrices via singular value decomposition. Specifically, we have

$$\begin{aligned} \mathbf{Y} - \mathbf{m}\mathbf{1}^\top &= \mathbf{U}\mathbf{D}\mathbf{V}^\top \\ \mathbf{W} &= \mathbf{U}\mathbf{D} \\ \mathbf{Z} &= \mathbf{V}^\top, \end{aligned}$$

where \mathbf{U} and \mathbf{V} are orthonormal matrices and \mathbf{D} is diagonal matrix.

Hyperparameter tuning

We used the Bayesian information criterion (BIC) to tune the hyperparameters. We select the hyperparameters with the lowest BIC criteria

$$\text{BIC} = -2 \sum_h l_h(\mathbf{X}_h, \hat{\boldsymbol{\mu}}_h) + \log n \times df,$$

where $\hat{\boldsymbol{\mu}}_h$ is the estimated mean matrix of \mathbf{X}_h ; n is the number of subjects and df denotes the degrees of freedom, which is the number of non-zero elements in estimated factor loading matrix. We try different sets of hyperparameters (L, v_1, a_w) in both simulation and real data analysis.

Analysis of ADNI Data

To demonstrate the usefulness of our proposed method in real data analysis, we aimed to predict the FAQ cognitive score using the latent factors learned through integrative analysis of the genotyping, gene expression, brain regional level amyloid deposition, gene-gene interaction network and brain functional network data from the ADNI biobank database. We also performed the same analysis using existing factor analysis models (JIVE [25], SLIDE [26], iCluster+ [24], GBFA [47]) as the benchmark.

Data description and data preprocessing

The genotyping data, gene expression data, demographic data and imaging data used in the preparation of this article were obtained from the ADNI database (adni.loni.usc.edu) [59–61]. 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, 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 AD. Up-to-date information about the ADNI is available at www.adni-info.org.

After data preprocessing steps and the common subjects matching in each omics data, we have 604 subjects left with main characteristics summarized in Table 2.

Demographic data preprocessing and cognitive score

Our demographic data and the cognitive score are extracted from Quantitative Templates for the Progression of Alzheimer’s Disease project in the Alzheimer’s Disease Modeling Challenge (www.pi4cs.org/qt-pad-challenge). We use baseline age, gender, education and the presence of APOE4 as our covariates. A test of the subjects’ daily living activities, the FAQ [62], is extracted as our dependent variable of real data analysis. All subjects with missing values are excluded from our study.

Genotyping data preprocessing

For the genotyping data, we performed quality control (QC) using the following criteria: genotyping call rate $> 95\%$, minor allele frequency $> 5\%$ and Hardy–Weinberg Equilibrium $> 1e - 6$. After the QC, rs429358 (APOE) SNP was added to the genotyping data. We prioritized the SNP markers based on the existing findings from the major AD GWAS articles [63–67]. Our raw data were obtained by additive recoding the number of minor alleles (–recode A in PLINK1.9 [68]). Given that we want to model our genotyping data as Bernoulli distribution, we further recoded the genotyping data by the presence of homozygous major allele and the presence of homozygous minor allele, which doubled the number of input features for the genotyping dataset.

Gene expression data preprocessing

Gene expression data were produced using Affymetrix Human Genome U219 Array (Affymetrix, Santa Clara, CA) for expression profiling where the blood sample extraction, hybridization and QC process could be found in [59]. In the raw data, one gene can correspond to multiple probes. In our real data analysis, we averaged all the probes within each gene across the subjects to obtain the corresponding gene expression. We prioritize the AD-related genes according to Disease Gene Network (DisGeNet) [69]. We extract all the genes that have gene–disease association scores [69] greater than 0.15.

Imaging data preprocessing

For the neuroimaging data, participants with complete [^{18}F] florbetapir (AV45) PET data (measuring amyloid burden) were included in our analysis. The AV45 PET scans were registered to the Montreal Neurological Institute space, and the standard uptake value ratio was computed by intensity normalization using the cerebellar curv reference region. Region-of-interest-level (ROI-level) AV45 measures were extracted based on the

Algorithm 1 EM Algorithm: SBFA

- 1: Initialization: $\mathbf{m}, \mathbf{W}, \mathbf{Z}, \boldsymbol{\Omega}, \boldsymbol{\mu}_\alpha, \boldsymbol{\Sigma}_\alpha, \boldsymbol{\mu}_\rho, \boldsymbol{\varphi}$, and $\boldsymbol{\mu}_z$;
- 2: **while** Not Converge **do**
- 3: For $j = 1, 2, \dots, p$ and $k = 1, 2, \dots, p$, set

$$a_{jk} = a_\omega; \quad b_{jk} = b_\omega + \frac{1}{2\nu_2} \sum_{l=1}^L (\boldsymbol{\Sigma}_{\alpha,l(k,k)} - 2\boldsymbol{\Sigma}_{\alpha,l(j,k)} + \boldsymbol{\Sigma}_{\alpha,l(j,j)}) + \frac{1}{2\nu_2} \sum_{l=1}^L (\mu_{\alpha,lk}^2 - 2\mu_{\alpha,lj}\mu_{\alpha,lk} + \mu_{\alpha,lj}^2); \quad \mu_{\omega,jk} = \frac{a_{jk}}{b_{jk}}$$

- 4: For $l = 1, 2, \dots, L$ and $k = 1, 2, \dots, p$, set

$$d_{kl} = |w_{kl}| e^{\frac{1}{2} \text{diag}(\boldsymbol{\Sigma}_{\alpha,l})_k}; \quad r_{kl} = \frac{1}{\nu_2} \left(1 + \sum_{\substack{j=1 \\ j \neq k}}^p \mu_{\omega,kj} \right); \quad c_{kl} = -\frac{1}{\nu_2} \sum_{\substack{j=1 \\ j \neq k}}^p \mu_{\omega,kj} \mu_{\alpha,jl} - \frac{\nu_1 + \nu_2}{\nu_2}; \quad \mu_{\alpha,kl} = -W \left(\frac{d_{kl}}{r_{kl}} e^{-\frac{c_{kl}}{r_{kl}}} \right) - \frac{c_{kl}}{r_{kl}}$$

- 5: For $l = 1, 2, \dots, L$ and $j = 1, 2, \dots, p$, set

$$C_j = \frac{\nu_2 - \sqrt{\nu_2^2 + 4\text{tr}(\boldsymbol{\Sigma}_{\alpha,l(-j,-j)} \boldsymbol{\Omega}_{(j,-j)} \boldsymbol{\Omega}_{(-j,j)} \boldsymbol{\Sigma}_{\alpha,l(j,j)})}}{2\text{tr}(\boldsymbol{\Sigma}_{\alpha,l(-j,-j)} \boldsymbol{\Omega}_{(j,-j)} \boldsymbol{\Omega}_{(-j,j)})}$$

$$\boldsymbol{\Sigma}_{\alpha,l(-j,j)} = C_j \boldsymbol{\Sigma}_{\alpha,l(-j,-j)} \boldsymbol{\Omega}_{(-j,j)}$$

$$c_j = \frac{1}{2}; \quad t_j = \boldsymbol{\Sigma}_{\alpha,l(j,-j)} \boldsymbol{\Sigma}_{\alpha,l(-j,-j)}^{-1} \boldsymbol{\Sigma}_{\alpha,l(-j,j)}; \quad r_j = \frac{\boldsymbol{\Omega}_{(j,j)}}{\nu_2 |w_{ji}| e^{\mu_{\alpha,jl}}}; \quad a_j = \frac{1}{|w_{ji}| e^{\mu_{\alpha,jl}}}$$

$$\boldsymbol{\Sigma}_{\alpha,l(j,j)} = t_j + \frac{1}{c_j} W_{r_j e^{-c_j t_j}} (c_j a_j e^{-c_j t_j})$$

- 6: For $i = 1, 2, \dots, n$, set

$$\boldsymbol{\Sigma}_{z,i}^{-1} = \mathbf{I} + \sum_{j=1}^p \mu_{\rho_j} \tilde{\mathbf{w}}_j \tilde{\mathbf{w}}_j^\top; \quad \boldsymbol{\mu}_{z,i} = \boldsymbol{\Sigma}_{z,i} \sum_{j=1}^p (\mu_{\rho_j} \psi_{ji} + \kappa_{ji} - \mu_{\rho_j} m_j) \tilde{\mathbf{w}}_j$$

- 7: For $j = 1, 2, \dots, p$ and $i = 1, 2, \dots, n$, set

$$\varphi_{ji} = \left(\tilde{\mathbf{w}}_j^\top \boldsymbol{\Sigma}_{z,i} \tilde{\mathbf{w}}_j + (m_j + \tilde{\mathbf{w}}_j^\top \boldsymbol{\mu}_{z,i} - \psi_{ji})^2 \right)^{\frac{1}{2}};$$

$$\mu_{\rho_{ji}} = \frac{b_{ji}(e^{\varphi_{ji}} - 1)}{2\varphi_{ji}(e^{\varphi_{ji}} + 1)} \text{ for discrete } x_{ji}; \quad \mu_{\rho_{ji}} = \frac{\xi_j + n}{\xi_j + \sum_{i=1}^n \varphi_{ji}^2} \text{ for Gaussian } x_{ji};$$

- 8: For $j = 1, 2, \dots, p$, set

$$\hat{\mathbf{w}}_j = \text{argmax}_{\tilde{\mathbf{w}}_j} \left(\frac{1}{2} \tilde{\mathbf{w}}_j^\top \left(\sum_{i=1}^n \mu_{\rho_{ji}} (\boldsymbol{\Sigma}_{z,i} + \boldsymbol{\mu}_{z,i} \boldsymbol{\mu}_{z,i}^\top) \right) \tilde{\mathbf{w}}_j - \sum_{i=1}^n (\kappa_{ji} - \mu_{\rho_{ji}} (m_j - \psi_{ji})) \boldsymbol{\mu}_{z,i}^\top \tilde{\mathbf{w}}_j + (e^{\mu_{\alpha,j} + \frac{1}{2} \text{diag} \boldsymbol{\Sigma}_{\alpha,j}})^\top |\tilde{\mathbf{w}}_j| \right)$$

- 9: For $j = 1, 2, \dots, p$, set

$$m_j = \frac{1}{\sum_{i=1}^n \mu_{\rho_{ji}}} \left(\sum_{i=1}^n \kappa_{ji} - \sum_{i=1}^n \mu_{\rho_{ji}} \tilde{\mathbf{w}}_j^\top \boldsymbol{\mu}_{z,i} + \sum_{i=1}^n \mu_{\rho_{ji}} \psi_{ji} \right)$$

- 10: **end while**

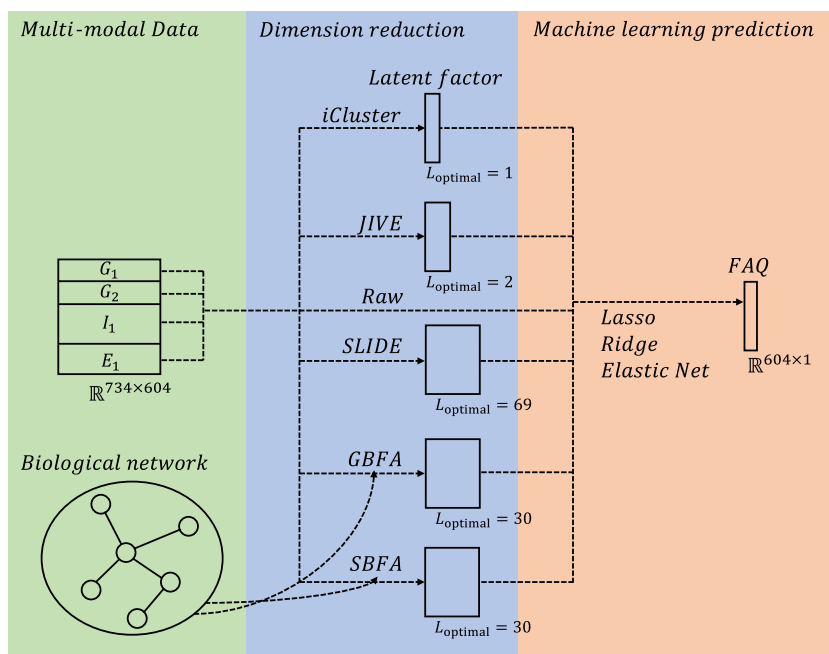


Figure 2. Workflow for real data analysis. G_1 and G_2 are genotyping data encoded by the presence of homozygous major allele and the presence of homozygous alternative allele, respectively. I_1 and E_1 are normalized QTs representing the gene expression and neuroimaging data. Starting from the multi-modal data, different factor analysis methods are used to extract the latent factors. The number of latent dimensions is tuned either through the implemented function from the package or using BIC. After obtaining the latent factor, we use the latent factor as the input predictor to predict the FAQ score. Linear regression with lasso, ridge and elastic net regularizations are used. The raw data (transpose of multi-omics data), covariates and raw data plus covariates are also used for comparison purpose.

Glasser atlas [70], where 360 ROI-level quantitative traits (QTs) were obtained by averaging all the voxel-level measures within each ROI.

Biological network

We download the tissue-specific gene–gene interaction network from GIANT [40] in HumanBase database (hb.flatironinstitute.org). The imaging network is obtained from Glasser et al. [70], which utilizes a supervised machine learning classifier to automatically delineate and identify each cortical area from its neighbors across a large majority of individual subjects based on multi-modal information.

Real data analysis

In our real data analysis, we evaluate our SBFA framework by predicting the FAQ score using our extracted latent factors, which can be considered as a low-dimensional representation of the multi-omics data sets. Our real data analysis pipeline is shown in Figure 2.

Our framework starts from the dimension reduction of multi-omics data including indicator variables for homozygous dominant genotyping data G_1 , indicator variables for homozygous recessive genotyping data G_2 , AV45 imaging measurements I_1 and gene expression E_1 . Specifically, G_1 and G_2 are binary matrices with dimension $\mathbb{R}^{111 \times 604}$, the entries of which are assumed to follow the Bernoulli distribution. I_1 and E_1 are normalized QTs with dimension $\mathbb{R}^{360 \times 604}$ and $\mathbb{R}^{152 \times 604}$, which are assumed to be normally distributed.

We estimate the latent factors from the multi-omics and imaging dataset using our proposed SBFA method, in comparison with several existing methods including iCluster+ [19, 71], JIVE [25], SLIDE [26] and GBFA [47]. For the hyperparameters, we fix ($v_2 = \log(2)$, $b_\omega = 1$) and tune (a_ω, L, v_1). The best tuning results are obtained by minimizing the BIC criteria, which is presented

in Figure 2. The best-tuned L is indicated by L_{optimal} , which is the number of columns of the corresponding latent factors. For our proposed SBFA model, the best-tuned hyperparameter set is ($L_{\text{optimal}} = 30$, $v_1 = 5.5$, $a_\omega = 0.5$). We then apply linear regression with lasso, ridge and elastic net regularizations models to predict the FAQ score using the latent factors learned from the aforementioned methods as predictors. For comparison purposes, we also include three other approaches that use (i) the demographic covariates, (ii) the multi-omics and imaging data and (iii) the multi-omics and imaging data together with the demographic covariates as the input of machine learning models, respectively; see Table 3.

For our training process, we first evenly split the subjects into five folds where 4-folds are treated as the training-validation set and 1-fold is treated as the testing set. Next, we apply 4-fold cross-validation in our training-validation set to select the best tuning hyperparameters according to the mean squared error (MSE). The best tuning models are obtained from the validation set and applied to the testing set to get the testing performance. We repeat the splitting and tuning process 100 times by randomly splitting the data 100 times with 100 different random seeds. We summarized our real data analysis results in Table 3. The real data analysis performance is reported using the mean of the training, validation and testing MSE across 100 random splits. The standard errors for the training and testing MSE are rounded to zero after preserving three digits.

As shown in Table 3, compared with the results using original multi-omics and imaging data and clinical variables, our SBFA method does a better job in mitigating the overfitting problem. In particular, the three prediction models trained using estimated factors from our SBFA method outperform the models trained using the original multi-omics, imaging and demographic variables, in spite of the fact that our factor analysis used only multi-omics and imaging data, not demographic variables.

Table 3. Real data performance. The averaged MSE between the predicted FAQ score and the true FAQ score is reported across 100 random splittings. All standard errors are rounded to 0 after preserving 3 digits. For comparison purposes, we also include three other approaches that use (A) the demographic covariates, (B) the multi-omics and imaging data and (C) the multi-omics and imaging data together with the demographic covariates as the input of machine learning models.

Feature\Method	Lasso			Ridge			ElasticNet		
	Train	Validation	Test	Train	Validation	Test	Train	Validation	Test
Raw data and clinical variables									
Demographic covariates	0.626	0.955	0.953	0.592	0.934	0.949	0.596	0.938	0.945
Multi-omics and imaging	0.626	0.958	0.959	0.334	0.883	0.921	0.516	0.864	0.940
Multi-omics, imaging and covariates	0.622	0.957	0.956	0.333	0.886	0.933	0.517	0.864	0.940
Latent factors									
JIVE	0.638	0.964	0.963	0.638	0.964	0.963	0.638	0.964	0.963
SLIDE	0.638	0.964	0.963	0.478	0.920	0.908	0.638	0.963	0.963
iCluster+	0.638	0.964	0.963	0.637	0.974	0.961	0.638	0.964	0.963
GBFA	0.528	0.894	0.918	0.452	0.803	0.911	0.483	0.842	0.918
SBFA (proposed)	0.538	0.896	0.893	0.454	0.850	0.881	0.466	0.853	0.895

Compared with other existing factor analysis methods, our SBFA method yields the best performance (smallest MSE) across all comparisons (i.e. linear regression model with lasso, ridge and elastic net regularizations).

Simulation Study

We compared the performance of our proposed SBFA framework with other factor analysis frameworks including JIVE [25], SLIDE [26] and GBFA model with spike and slab prior and Markov random field prior [47] in the simulation study.

Simulation design

We evaluated our proposed SBFA framework in two different simulation scenarios—low dimensional scenario and high dimensional scenario with the number of features $p = 150$ and $p = 300$, respectively. We used the same number of samples in both simulation scenarios with $n = 160$. To simulate multi-omics data, we used the following model:

$$\boldsymbol{\mu}_h = \mathbf{m}_h \mathbf{1}^\top + \mathbf{W}_h \mathbf{Z}, \quad h = 1, 2, \dots, H, \quad (6)$$

where $\boldsymbol{\mu}_h \in \mathbb{R}^{p_h \times n}$, $\mathbf{m}_h \in \mathbb{R}^{p_h \times 1}$, $\mathbf{W}_h \in \mathbb{R}^{p_h \times L}$ and $\mathbf{Z} \in \mathbb{R}^{L \times n}$. We choose the ground truth latent variable $L = 3$ and the number of datasets $H = 6$.

We design our ground-truth loading matrix \mathbf{W} as shown in Figure 3, where each row represents a pathway ($h = 1, 2, \dots, 6$). The shaded areas are generated by continuous uniform distribution with $[1.5, 2.5]$ and every non-zero element of the loading matrix \mathbf{W} has a probability of 0.5 to flip its sign.

The ground truth factor matrix $\mathbf{Z} = [\mathbf{z}_i]_{i=1, \dots, n}$ is simply generated by

$$\mathbf{z}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{I}),$$

where $\mathbf{0}$ represents the zero vector, and \mathbf{I} represents the identity matrix.

The ground truth mean parameter for data \mathbf{X} , i.e. $\boldsymbol{\mu} = [\boldsymbol{\mu}_1^\top \boldsymbol{\mu}_2^\top \boldsymbol{\mu}_3^\top \boldsymbol{\mu}_4^\top \boldsymbol{\mu}_5^\top \boldsymbol{\mu}_6^\top]^\top$, is obtained by

$$\boldsymbol{\mu}_h = \mathbf{W}_h \mathbf{Z}, \quad h = 1, 2, \dots, 6,$$

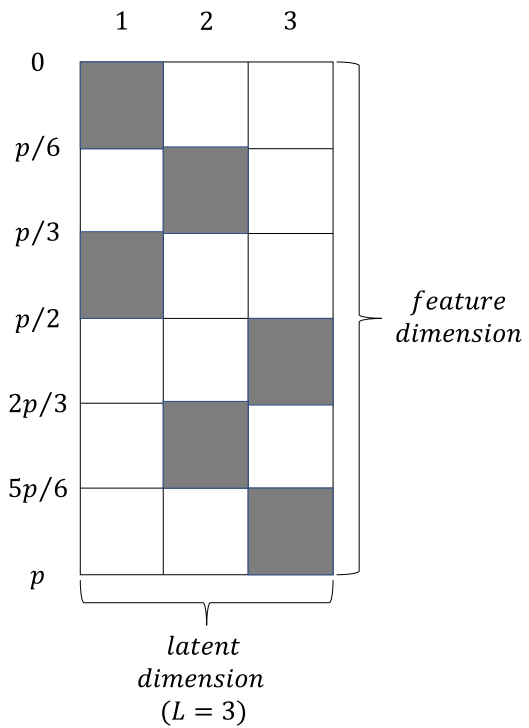


Figure 3. Groundtruth setting for factor loading matrix \mathbf{W} . Numbers labeled on each column represent the latent variable. Numbers labeled on rows represent the number of rows from the first row to the corresponding row. All shaded area represents elements generated by $\mathcal{U}(1.5, 2.5)$ with probability 0.5 to be multiplied by -1 . All the other area contains only zero elements.

where we assume location parameter $\mathbf{m} = \mathbf{0}$. For the first setting in each simulation scenario, we assume only Gaussian distribution with mean $\boldsymbol{\mu}_h$ and standard deviation 2 for all $h \in \{1, 2, \dots, 6\}$. For the second setting in each simulation scenario, we assume the first two omic datasets with $\frac{p}{3}$ features following Bernoulli distribution with each element having probability $\mathbf{p}^{(1)} = \frac{1}{1+e^{-\boldsymbol{\mu}^{(1)}}}$ to be 1. It mimics the data structure for the allele homozygosity encoded genotyping dataset. For the second two omics datasets with $\frac{p}{3}$ features, we assume that they are generated by a binomial distribution with trail parameter independently sampled from the set $\{n \in \mathbb{R} : 1 \leq n \leq 15\}$. For each feature and each subject, we assume the success probability parameter $\mathbf{p}^{(2)} = \frac{1}{1+e^{-\boldsymbol{\mu}^{(2)}}}$. For the

Table 4. Simulation performance results for Gaussian only setting. ‘Dimension’ denotes our simulation scenario where ‘Low’ indicates the simulation scenario with $p = 150$ and ‘High’ indicates the simulation scenario with $p = 300$. ‘Edge’ represents different graph information where ‘Within’ denotes the ideal case that we use the graph information which only connects the features within each pathway, and ‘Cross’ denotes the case that we use the graph information that connects the features not only within each pathway but also some randomly selected nodes between pathways within each modality. ‘RRE (mean)’ is the average of the reconstruction error between estimated $\hat{\mu}$ and the ground truth μ ratio to the true μ across 100 simulations. ‘RRE (se)’ is the standard error of the reconstruction error between estimated $\hat{\mu}$ and the ground truth μ across 100 simulations.

Method	Dimension	Edge	RRE (mean)	RRE (se)
JIVE	low		0.236552	0.006583
SLIDE	low		0.167271	0.000722
iCluster	low		0.221568	0.000935
GBFA ($\eta = 0$)	low	Within	0.198160	0.000848
GBFA ($\eta = 1$)	low	Within	0.190166	0.001088
SBFA (No \mathcal{G})	low	Within	0.182882	0.000743
SBFA (With \mathcal{G})	low	Within	0.180844	0.000699
GBFA ($\eta = 0$)	low	Cross	0.198160	0.000848
GBFA ($\eta = 1$)	low	Cross	0.185953	0.001015
SBFA (No \mathcal{G})	low	Cross	0.182882	0.000743
SBFA (With \mathcal{G})	low	Cross	0.181908	0.000704
JIVE	high		0.180514	0.001741
SLIDE	high		0.154596	0.000727
iCluster	high		0.188864	0.000806
GBFA ($\eta = 0$)	high	Within	0.166760	0.000663
GBFA ($\eta = 1$)	high	Within	0.166246	0.000772
SBFA (No \mathcal{G})	high	Within	0.153240	0.000686
SBFA (With \mathcal{G})	high	Within	0.151277	0.000665
GBFA ($\eta = 0$)	high	Cross	0.166760	0.000663
GBFA ($\eta = 1$)	high	Cross	0.158246	0.000882
SBFA (No \mathcal{G})	high	Cross	0.153240	0.000686
SBFA (With \mathcal{G})	high	Cross	0.152455	0.000672

Table 5. Simulation performance results for mixture distribution setting. ‘Dimension’ denotes our simulation scenario where ‘Low’ indicates the simulation scenario with $p = 150$ and ‘High’ indicates the simulation scenario with $p = 300$. ‘Edge’ represents different graph information where ‘Within’ denotes the ideal case that we use the graph information which only connects the features within each pathway, and ‘Cross’ denotes the case that we use the graph information that connects the features not only within each pathway but also some randomly selected nodes between pathways within each modality. ‘RRE (mean)’ is the average of the reconstruction error between estimated $\hat{\mu}$ and the ground truth μ ratio to the true μ across 100 simulations. ‘RRE (se)’ is the standard error of the reconstruction error between estimated $\hat{\mu}$ and the ground truth μ across 100 simulations.

Method	Dataset	Edge	RRE (mean)	RRE (se)
JIVE	low		1.539543	0.014274
SLIDE	low		1.519234	0.010254
iCluster	low		1.407917	0.011129
GBFA ($\eta = 0$)	low	Within	0.220563	0.000961
GBFA ($\eta = 1$)	low	Within	0.220918	0.000993
SBFA (No \mathcal{G})	low	Within	0.197850	0.000877
SBFA (With \mathcal{G})	low	Within	0.196208	0.000864
GBFA ($\eta = 0$)	low	Cross	0.220563	0.000961
GBFA ($\eta = 1$)	low	Cross	0.217769	0.001006
SBFA (No \mathcal{G})	low	Cross	0.197850	0.000877
SBFA (With \mathcal{G})	low	Cross	0.199828	0.000874
JIVE	high		1.551844	0.007424
SLIDE	high		1.508521	0.006791
iCluster	high		1.397968	0.007480
GBFA ($\eta = 0$)	high	Within	0.190666	0.000886
GBFA ($\eta = 1$)	high	Within	0.188583	0.000733
SBFA (No \mathcal{G})	high	Within	0.175308	0.000773
SBFA (With \mathcal{G})	high	Within	0.173823	0.000733
GBFA ($\eta = 0$)	high	Cross	0.190666	0.000886
GBFA ($\eta = 1$)	high	Cross	0.189722	0.000890
SBFA (No \mathcal{G})	high	Cross	0.175308	0.000773
SBFA (With \mathcal{G})	high	Cross	0.177781	0.000764

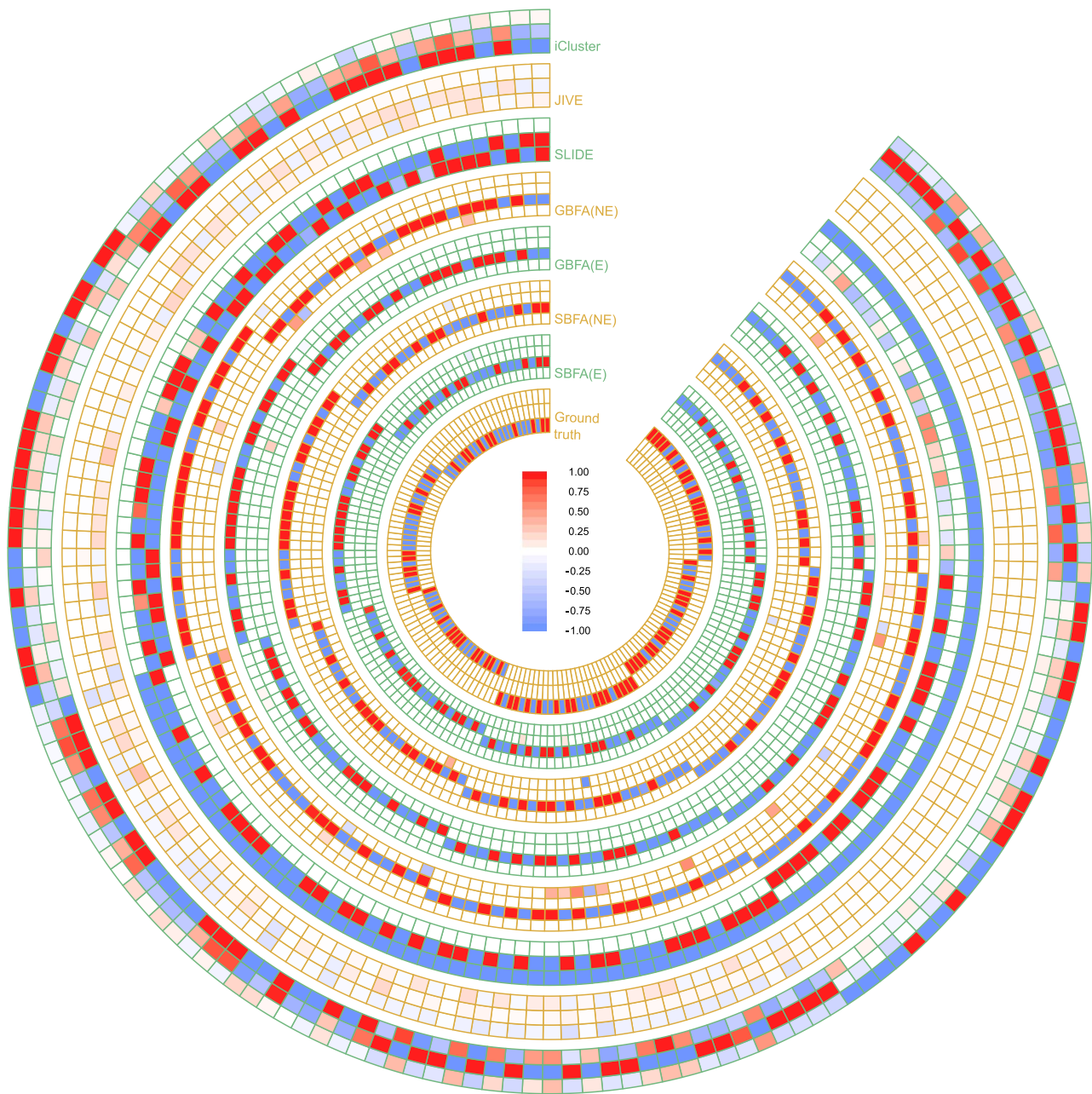


Figure e1. Estimated factor loadings for the simulated dataset. The heatmap shows the factor loadings from iCluster, JIVE, SLIDE and GBFA without graph information (GBFA(NE)), GBFA with graph information (GBFA(E)), SBFA without graph information (SBFA(NE)), SBFA with graph information (SBFA(E)) and the ground truth. The results are derived from one simulated dataset in the low-dimensional scenario ($p = 150$). The optimal latent dimension (L_{optimal}) for each model is indicated by the number of circles in the corresponding method track. Numbers beyond the range of $(-1, 1)$ have the same color as the boundaries $(-1$ and $1)$.

third two omics datasets with $\frac{p}{2}$ features, we assume that they are generated to mimic the gene expression and imaging data, which follows the Gaussian distribution with mean $[\mu_5^T \ \mu_6^T]^T$ and standard deviation 2.

For the graph information, we evaluate our algorithm with 2 different settings:

- *Within-pathway edges only:* For the first graph information setting, we randomly generate $\frac{p}{2}$ edges within each pathway and do not add across pathway edges (i.e. in total, we have $3p$ within pathway edges for p features). This is a simulation of an ideal case.
- *Within-pathway edges and cross pathway edges:* For the second graph information setting, besides the $3p$ within pathway

edges generated in the first setting, we also randomly generate $\frac{p}{2}$ cross-pathway edges within each modality (i.e. in total, we have $3p$ within pathway edges and $1.5p$ cross-pathway edges for p features).

The graph with within-pathway edges only represents the ideal case where the graphical structure coincides with the sparsity structure in W , while the cross-pathway edges mimic the real data.

Results

In our simulation study, we seek to reconstruct the mean matrix from the extracted low-rank representation. The performance

is evaluated by the relative reconstruction error (RRE) between estimated $\hat{\boldsymbol{\mu}}$ and true $\boldsymbol{\mu}$ using Frobenius norm

$$\text{RRE} = \frac{\|\boldsymbol{\mu} - \hat{\boldsymbol{\mu}}\|_F}{\|\boldsymbol{\mu}\|_F}, \quad (7)$$

where $\boldsymbol{\mu}$ denotes the ground truth mean parameter; $\hat{\boldsymbol{\mu}}$ represents the estimated mean parameter.

To reduce the randomness of the performance, we generate 100 different datasets and apply our proposed algorithm to obtain the new feature representation for each dataset. We fixed ($v_2 = \log 2, b_\omega = 1$) and tuned the (a_ω, L, v_1) parameters using BIC. We compare our SBFA framework with four different baseline factor analysis algorithms including JIVE, SLIDE, iCluster+ and GBFA.

We first evaluate our results by calculating the mean and the standard error of the RRE across 100 simulation datasets. It is worth noting that the best-tuned L_{optimal} based on BIC is always equal to or larger than the ground truth. For those L_{optimal} greater than the ground truth, compared with the other baseline methods, the estimated loadings by SBFA for the extra latent dimension(s) are exactly 0 except for a small number of nonzero entries (supplementary information Figure 1). The best-tuned results for each method and each simulation setting are summarized in Tables 4 and 5. Our proposed SBFA model always outperforms the other baseline methods in both the Gaussian distribution setting (Table 4) and the mixture distribution setting (Table 5).

In the Gaussian distribution setting (Table 4), without incorporating the graph information, our proposed model (SBFA No \mathcal{G}) shows significant improvement compared with JIVE, SLIDE, iCluster+ and GBFA without graph information (GBFA $\eta = 0$). By taking the graph information into consideration, our proposed model (SBFA With \mathcal{G}) can further achieve a smaller RRE and significantly outperforms the GBFA model (GBFA $\eta = 1$) both with and without cross pathway noise.

In the mixture distribution setting (Table 5), JIVE, SLIDE and iCluster yield a large RRE compared with GBFA and SBFA models. This is expected because those methods handle the discrete data by treating them as continuous random variables. Although the GBFA model and our proposed SBFA model can both handle discrete and continuous data simultaneously, our SBFA model outperforms the GBFA model in both low-dimensional simulation and high-dimensional simulation scenarios with or without incorporating the graph information.

In summary, our proposed SBFA model can handle discrete and continuous data simultaneously. Our simulation study demonstrates that our proposed SBFA model can outperform the most state-of-the-art factor analysis models. Moreover, our SBFA model can incorporate the biological graph information as prior knowledge so that our model can produce biologically interpretable results.

Conclusion

In this article, we proposed a structural Bayesian factor analysis model which decomposes the underlying mean parameters of the data into a sparse factor loading matrix and factor matrix. Our model employed the Laplace prior to achieve the sparse estimation of the loading matrix and used a graph Laplacian prior to incorporate the biological network information. Our method can handle both discrete and continuous data simultaneously. We derived an efficient variational EM algorithm to obtain the MAP estimators of the parameters of interest.

Our SBFA framework is a useful dimension reduction technique that yields latent factors that are biologically interpretable. As such, the common information shared among multi-modal data learned from our model can be used for various downstream machine learning tasks including, but not limited to, prediction. For example, another potential application is to use learned factors for clustering which can be used to define disease subtypes based on genotyping data, imaging modalities and biological pathways.

Our simulation study demonstrates that our proposed SBFA framework achieves the best performance compared with the other state-of-the-art factor-analysis-based integrative methods. Our work bridges the gap that there are few multi-omics studies applied to AD. We used the multi-omics data from the ADNI dataset to predict the FAQ score, which is an important evaluation for the diagnosis of AD. The real data analysis results show the advantages of our framework by comparing the prediction performance between our SBFA framework and other state-of-the-art factor-analysis-based integrative methods. Our method exhibited the best prediction performance.

Key Points

- We propose a novel structural Bayesian factor analysis framework as a dimension reduction technique that yields latent factors by integrating the genotyping data, gene expression data, region-level brain imaging amyloid deposition data and biological network information. Our SBFA framework is biologically interpretable. The common information shared among multi-omics data learned from our model can be used for various subsequent analyses such as supervised and unsupervised machine learning.
- Our structural Bayesian factor analysis model can handle both continuous (e.g. gene expression and imaging phenotypic traits) and discrete [e.g. genotyping values for single nucleotide polymorphisms (SNPs)] datatypes simultaneously.
- Our work bridges the gap that there are few multi-omics studies applied to Alzheimer's disease. We applied our framework to extract biologically meaningful information shared by multiple modalities in the ADNI database. We performed the subsequent machine learning analyses to predict the Alzheimer's disease-related cognitive score using the extracted latent factors. Our structural Bayesian factor analysis model achieves the best prediction accuracy for the subsequent machine learning prediction problems compared with the other state-of-the-art factor-analysis-based methods.

Data availability

The source code is available through GitHub (<https://github.com/JingxuanBao/SBFA>). The genotyping data, gene expression data, demographic data, and imaging data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database. The data are available after application through the ADNI website (<http://adni.loni.usc.edu>).

Supplementary data

Supplementary data are available online at <http://bib.oxfordjournals.org/>.

Acknowledgements

This study was partly supported by NIH grants RF1AG063481, R01AG071174, R01 AG071470, U01 AG066833 and U01 AG068057. Data used in this study were obtained from the ADNI database (adni.loni.usc.edu), which was funded by NIH U01AG024904.

Author contributions statement

Jingxuan Bao and Dr. Changgee Chang take full responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Jingxuan Bao, Dr. Changgee Chang and Dr. Qi Long.

Acquisition, analysis or interpretation of data: Jingxuan Bao, Dr. Changgee Chang, Dr. Andrew J. Saykin, Dr. Li Shen and Dr. Qi Long.

Drafting of the manuscript: Jingxuan Bao, Dr. Changgee Chang and Dr. Qi Long.

Critical revision of the manuscript for important intellectual content: All authors.

Study supervision: Dr. Changgee Chang and Dr. Qi Long.

References

- Subramanian I, Verma S, Kumar S, et al. Multi-omics data integration, interpretation, and its application. *Bioinform Biol Insights* 2020; **14**:1177932219899051.
- Trambaiolli LR, Lorena AC, Fraga FJ, et al. Improving alzheimer's disease diagnosis with machine learning techniques. *Clin EEG Neurosci* 2011; **42**(3): 160–5.
- Kim M, Bao J, Liu K, et al. Structural connectivity enriched functional brain network using simplex regression with graphnet. In: *Machine Learning in Medical Imaging*. Cham: Springer International Publishing, 2020, 292–302.
- Kim M, Bao J, Liu K, et al. A structural enriched functional network: an application to predict brain cognitive performance. *Med Image Anal* 2021; **71**:102026.
- Kochunov P, Zavaliangos-Petropulu A, Jahanshad N, et al. A white matter connection of schizophrenia and Alzheimer's disease. *Schizophr Bull* 2020; **47**(1): 197–206.
- Sendi MSE, Zendeirouh E, Miller RL, et al. Alzheimer's disease projection from normal to mild dementia reflected in functional network connectivity: a longitudinal study. *Front Neural Circuits* 2021; **14**: 593263.
- Sendi MSE, Zendeirouh E, Fu Z, et al. Disrupted dynamic functional network connectivity among cognitive control networks in the progression of alzheimer's disease. *Brain Connect* 2021.
- Wan J, Zhang Z, Rao BD, et al. Identifying the neuroanatomical basis of cognitive impairment in alzheimer's disease by correlation- and nonlinearity-aware sparse bayesian learning. *IEEE Trans Med Imaging* 2014; **33**(7): 1475–87.
- Madar IH, Sultan G, Tayubi IA, et al. Identification of marker genes in alzheimer's disease using a machine-learning model. *Bioinformatics* 2021; **17**(2): 348–55.
- Ferreira D, Nordberg A, Westman E. Biological subtypes of alzheimer disease. *Neurology* 2020; **94**(10): 436–48.
- Jellinger KA. Pathobiological subtypes of alzheimer disease. *Dement Geriatr Cogn Disord* 2021; **49**:321–33.
- Lin D, Cao H, Calhoun VD, et al. Sparse models for correlative and integrative analysis of imaging and genetic data. *J Neurosci Methods* 2014; **237**:69–78.
- Zhu X, Yao J, Xiao G, et al. Imaging-genetic data mapping for clinical outcome prediction via supervised conditional gaussian graphical model. In: *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. New York City, USA: Institute of electrical and electronics engineers (IEEE), pp. 455–9, 2016.
- Batmanghelich NK, Dalca A, Quon G, et al. Probabilistic modeling of imaging, genetics and diagnosis. *IEEE Trans Med Imaging* 2016; **35**(7): 1765–79.
- Shen L, Thompson PM. Brain imaging genomics: integrated analysis and machine learning. *Proc IEEE* 2020; **108**(1): 125–62.
- Shen L, Kim S, Qi Y, et al. Identifying neuroimaging and proteomic biomarkers for mci and ad via the elastic net. *Multimodal Brain Image Analysis* 2011; **7012**:27–34.
- Wang H, Nie F, Huang H, et al. Identifying disease sensitive and quantitative trait-relevant biomarkers from multidimensional heterogeneous imaging genetics data via sparse multimodal multitask learning. *Bioinformatics* 2012; **28**(12): i127–36.
- Wang D, Gu J. Integrative clustering methods of multi-omics data for molecule-based cancer classifications. *Quant Biol* 2016; **4**(1): 58–67.
- Shen R, Olshen AB, Ladanyi M. Integrative clustering of multiple genomic data types using a joint latent variable model with application to breast and lung cancer subtype analysis. *Bioinformatics* 2009; **25**(22): 2906–12.
- Shen R, Wang S, Mo Q. Sparse integrative clustering of multiple omics data sets. *Ann Appl Stat* 2013; **7**(1): 269–94.
- Chauvel C, Novoloaca A, Veyre P, et al. Evaluation of integrative clustering methods for the analysis of multi-omics data. *Brief Bioinform* 2019; **21**(2): 541–52.
- Chalise P, Fridley BL. Integrative clustering of multi-level 'omic data based on non-negative matrix factorization algorithm. *PLoS One* 2017; **12**(5): 1–18.
- Ahmed KT, Sun J, Cheng S, et al. Multi-omics data integration by generative adversarial network. *Bioinformatics* 2021; **38**(1): 179–86.
- Mo Q, Wang S, Seshan VE, et al. Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proc Natl Acad Sci U S A* 2013; **110**(11): 4245–50.
- Lock EF, Hoadley KA, Marron JS, et al. Joint and individual variation explained (JIVE) for integrated analysis of multiple data types. *Ann Appl Stat* 2013; **7**(1): 523–42.
- Gaynanova I, Li G. Structural learning and integrative decomposition of multi-view data. *Biometrics* 2019; **75**(4): 1121–32.
- Murray ME, Graff-Radford NR, Ross OA, et al. Neuropathologically defined subtypes of alzheimer's disease with distinct clinical characteristics: a retrospective study. *Lancet Neurol* 2011; **10**(9): 785–96.
- Whitwell JL, Dickson DW, Murray ME, et al. Neuroimaging correlates of pathologically defined subtypes of alzheimer's disease: a case-control study. *Lancet Neurol* 2012; **11**(10): 868–77.
- Neff RA, Wang M, Vatansever S, et al. Molecular subtyping of alzheimer's disease using rna sequencing data reveals novel mechanisms and targets. *Sci Adv* 2021; **7**(2): eabb5398.
- Di Fede G, Catania M, Maderna E, et al. Molecular subtypes of alzheimer's disease. *Sci Rep* 2018; **8**(1): 3269.
- Levin F, Ferreira D, Lange C, et al. Data-driven fdg-pet subtypes of alzheimer's disease-related neurodegeneration. *Alzheimer's Res Ther* 2021; **13**(1): 49.

32. Qiang W, Yau WM, Jun-Xia L, et al. Structural variation in amyloid- β fibrils from alzheimer's disease clinical subtypes. *Nature* 2017; **541**(7636): 217–21.
33. Wong AK, Krishnan A, Yao V, et al. IMP 2.0: a multi-species functional genomics portal for integration, visualization and prediction of protein functions and networks. *Nucleic Acids Res* 2015; **43**(W1): W128–33.
34. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of genes and genomes. *Nucleic Acids Res* 2000; **28**(1): 27–30.
35. Kanehisa M. Toward understanding the origin and evolution of cellular organisms. *Protein Sci* 2019; **28**(11): 1947–51.
36. Kanehisa M, Furumichi M, Sato Y, et al. KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res* 2020; **49**(D1): D545–51.
37. Zhou J, Troyanskaya OG. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat Methods* 2015; **12**(10): 931–4.
38. Zhou J, Theesfeld CL, Yao K, et al. Deep learning sequence-based ab initio prediction of variant effects on expression and disease risk. *Nat Genet* 2018; **50**(8): 1171–9.
39. Krishnan A, Zhang R, Yao V, et al. Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nat Neurosci* 2016; **19**(11): 1454–62.
40. Greene CS, Krishnan A, Wong AK, et al. Understanding multicellular function and disease with human tissue-specific networks. *Nat Genet* 2015; **47**(6): 569–76.
41. Li C, Li H. Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics* 2008; **24**(9): 1175–82.
42. Pan W, Xie B, Shen X. Incorporating predictor network in penalized regression with application to microarray data. *Biometrics* 2010; **66**(2): 474–84.
43. Li F, Zhang NR. Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics. *J Am Stat Assoc* 2010; **105**(491): 1202–14.
44. Stingo FC, Chen YA, Tadesse MG, et al. Incorporating biological information into linear models: a bayesian approach to the selection of pathways and genes. *Ann Appl Stat* 2011; **5**(3): 1978–2002.
45. Chang C, Kundu S, Long Q. Scalable bayesian variable selection for structured high-dimensional data. *Biometrics* 2018; **74**(4): 1372–82.
46. Rockova V, Lesaffre E. Incorporating grouping information in Bayesian variable selection with applications in genomics. *Bayesian Anal* 2014; **9**(1): 221–58.
47. Min EJ, Chang C, Long QI. Generalized bayesian factor analysis for integrative clustering with applications to multi-omics data. In: *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*. New York City, USA: Institute of electrical and electronics engineers (IEEE), pp. 109–19, 2018.
48. Coniglio A, Nappi CR, Peruggi F, et al. Percolation and phase transitions in the ising model. *Commun Math Phys* 1976; **51**(3): 315–23.
49. Dubes RC, Jain AK, Nadabar SG, Chen CC. Mrf model-based algorithms for image segmentation. In: *[1990] Proceedings. 10th International Conference on Pattern Recognition*. New York City, USA: Institute of electrical and electronics engineers (IEEE), pp. 808–14, vol. **1**, 1990.
50. Chung FR. *Spectral graph theory*. Rhode Island, US: American Mathematical Soc, 1997.
51. Ng AY, Jordan MI, Weiss Y. On spectral clustering: Analysis and an algorithm. In: *Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic, NIPS'01*. Cambridge, Massachusetts, USA: Massachusetts Institute of Technology (MIT), pp. 849–56. MIT Press, 2001.
52. Polson NG, Scott JG, Windle J. Bayesian inference for logistic models using pólya-gamma latent variables. *J Am Stat Assoc* 2013; **108**(504): 1339–49.
53. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the em algorithm. *J R Stat Soc B Methodol* 1977; **39**(1): 1–38.
54. Girolami M. A Variational method for learning sparse and overcomplete representations. *Neural Comput* 2001; **13**(11, 11): 2517–32.
55. Blei DM, Kucukelbir A, McAuliffe JD. Variational inference: a review for statisticians. *J Am Stat Assoc* 2017; **112**(518): 859–77.
56. Corless RM, Gonnet GH, Hare DEG, et al. On the lambert w function. *Adv Comput Math* 1996; **5**(1): 329–59.
57. Mezo I, Baricz A. On the generalization of the lambert w function. *Trans Am Math Soc* 2017; **7**(369): 7917–34.
58. Chang C, Tsay RS. Estimation of covariance matrix via the sparse cholesky factor with lasso. *J Stat Plan Inference* 2010; **140**(12): 3858–73.
59. Saykin AJ, Shen L, Yao X, et al. Genetic studies of quantitative MCI and AD phenotypes in ADNI: progress, opportunities, and plans. *Alzheimers Dement* 2015; **11**(7): 792–814.
60. Shen L, Thompson PM, Potkin SG, et al. Genetic analysis of quantitative phenotypes in AD and MCI: imaging, cognition and biomarkers. *Brain Imaging Behav* 2014; **8**(2): 183–207.
61. Weiner MW, Veitch DP, Aisen PS, et al. Recent publications from the Alzheimer's Disease Neuroimaging Initiative: reviewing progress toward improved AD clinical trials. *Alzheimers Dement* 2017; **13**(4): e1–85.
62. Pfeffer RI, Kurosaki TT, Harrah CH, Jr, et al. Measurement of functional activities in older adults in the community. *J Gerontol* 1982; **37**(3): 323–9.
63. Seshadri S, Fitzpatrick AL, Ikram MA, et al. Genome-wide analysis of genetic loci associated with Alzheimer disease. *JAMA* 2010; **303**(18): 1832–40.
64. Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; **45**(12): 1452–8.
65. Jansen IE, Savage JE, Watanabe K, et al. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet* 2019; **51**(3): 404–13.
66. Kunkle B, Benjamin G-B, Rebecca S, et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates A β , tau, immune and lipid processing. *Nat Genet* 2019; **51**:414–30.
67. Bellenguez C, Kucukali F, Jansen I, et al. New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat Genet* 2020; **54**: 412–36.
68. Chang CC, Chow CC, Tellier LCAM, et al. Second-generation plink: rising to the challenge of larger and richer datasets. *GigaScience* 2015; **4**:7.
69. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res* 2019; **48**(D1): D845–55.
70. Glasser MF, Coalson TS, Robinson EC, et al. A multi-modal parcellation of human cerebral cortex. *Nature* 2016; **536**(7615): 171–8.
71. Mo Q, Shen R. iClusterPlus: integrative clustering of multi-type genomic data. *R package version* 2021; **1.28.0**. <http://52.71.54.154/packages/release/bioc/html/iClusterPlus.html>.