Published in final edited form as:

J Neurosci Methods. 2018 November 01; 309: 161–174. doi:10.1016/j.jneumeth.2018.08.027.

A Kernel Machine Method for Detecting Higher Order Interactions in Multimodal Datasets: Application to Schizophrenia

Md. Ashad Alam^{a,*}, Hui-Yi Lin^b, Hong-Wen Deng^c, Vince D. Calhoun^d, and Yu-Ping Wang^a

^aDepartment of Biomedical Engineering, Tulane University, New Orleans, LA 70118, USA.

^bBiostatistics Program, School of Public Health, Louisiana State University Health Sciences Center, New Orleans, LA 70112, USA.

^cCenter for Bioinformatics and Genomics, Department of Global Biostatistics and Data Science, Tulane University, New Orleans, LA 70112, USA.

^dDepartment of Electrical and Computer Engineering, The University of New Mexico, Albuquerque, NM 87131, USA.

Abstract

Background: Technological advances are enabling us to collect multimodal datasets at an increasing depth and resolution while with decreasing labors. Understanding complex interactions among multimodal datasets, however, is challenging.

New Method: In this study, we tested the interaction effect of multimodal datasets using a novel method called the kernel machine for detecting higher order interactions among biologically relevant multimodal data. Using a semiparametric method on a reproducing kernel Hilbert space, we formulated the proposed method as a standard mixed-effects linear model and derived a score-based variance component statistic to test higher order interactions between multimodal datasets.

Results: The method was evaluated using extensive numerical simulation and real data from the Mind Clinical Imaging Consortium with both schizophrenia patients and healthy controls. Our method identified 13-triplets that included 6 gene-derived SNPs, 10 ROIs, and 6 gene-specific DNA methylations that are correlated with the changes in hippocampal volume, suggesting that these triplets may be important for explaining schizophrenia-related neurodegeneration.

Comparison with Existing Method(s): The performance of the proposed method is compared with the following methods: test based on only first and first few principal components followed by multiple regression, and full principal component analysis regression, and the sequence kernel association test.

Conclusions: With strong evidence (p -value = 0.000001), the triplet (**MAGI2**, **CRBLCrus1.L**, **FBXO28**) is a significant biomarker for schizophrenia patients. This novel method can be applicable to the study of other disease processes, where multimodal data analysis is a common task.

*Corresponding author malam@tulane.edu.

Keywords

Multimodal datasets; Higher order interaction; Kernel machine methods; Imaging genetics and epigenetics; Schizophrenia

1. Introduction

The advancement in biomedical technology over the last decade has produced huge volume of multimodal data, providing a comprehensive way for disease diagnosis. A central goal of multimodal data integration is to understand the interaction effects of different features. The integration of multimodal data information (e.g., imaging, genetics and epigenetics), however, continues to be a challenging problem.

One goal of imaging (epi)-genetics is the modeling and understanding of how (epi)-genetic variations influence the structure and function of the brain. This can be achieved by using multimodal data including functional magnetic resonance imaging (fMRI), structural MRI (sMRI), positron emission tomography (PET) scans, diffusion tensor imaging (DTI), along with single nucleotide polymorphisms (SNPs), deoxyribonucleic acid (DNA) methylations, and gene expression (GE) factors. To date, both genetic and brain imaging techniques have played a substantial role in detecting disease phenotypes. For example, by correlating imaging with genetic data, it has been shown that some genes affect specific brain functions, connectivity, and serve as risk predictors for certain diseases (Jahanshad et al., 2012; Lin et al., 2014; Bis et al., 2012; Jahanshad et al., 2013). In another example, Bis et al. (2012) have identified genetic variants affecting the volume of the hippocampus, which can be used as predictors of cognitive decline and dementia (Jahanshad et al., 2013).

As shown in (Wen et al., 2017), accurate identification of Tourette's syndrome in children has notably improved using multimodal features as compared to relying solely on one view of data. Accumulating evidence also shows that the inherent genetic variations for complex traits can sometimes be explained by joint analysis of multiple genetic and environmental factors. In addition, numerous studies have suggested that these different factors do not act in isolation, but rather interact at multiple levels and depend on one another in an intertwined manner (Calhoun and Sui, 2016; Pearlson et al., 2015). Figure 1 illustrates how the interaction effects of different data sets can be used to model and predict human illnesses. Extracting the interaction effects from within and among data sets, however, remains a challenge for multimodal data analysis (Li et al., 2015; Chekouo et al., 2016; Zheng et al., 2015; Zhao et al., 2016; Liu et al., 2016).

The use of multimodal imaging and genomic data is particularly popular for the study of schizophrenia (SZ). SZ is a complex brain disorder that affects how a person thinks, feels and acts, which is thought to be caused through an interplay of genetic effects, brain region, and DNA methylation abnormalities (Richfield et al., 2017). Studies using neurological tests and brain imaging technologies (fMRI and PET) have been used to examine functional differences in brain activity that seem to arise within the frontal lobes, hippocampus and temporal lobes (Van and Kapur, 2009; Kircher and Renate, 2005). Many researchers have shown that genetic alterations at the mRNA and SNP level, however, also play a significant

role in SZ (Chang et al., 2013; Lencz et al., 2007). Thus, only focusing on brain imaging data is not sufficient for the identification of the related risk factors for SZ (Potkin et al., 2015). To address this issue, Chekouo et al. (2016) have developed the ROI-SNP network for the selection of discriminatory markers using brain imaging and genetic information. A number of studies suggest that epigenetics also has a role in SZ disease susceptibility. Genome-wide DNA methylation analysis of human brain tissue from SZ patients shows a heritable epigenetic modification that can regulate gene expression. The cell specific differences in chromatin structure that influence cell development, including DNA methylation, have emerged as a potential explanation for the non-Mendelian inheritance of SZ (Wockner et al., 2014). There is also evidence on epigenetic alterations in the blood and central nervous system of patients with SZ, and it has been shown that methylation status in brain tissue from SZ patients varies significantly from controls (Aberg et al., 2014; Montano et al., 2016). In the last decade, a number of statistical methods have been used to detect gene-gene interactions (GGIs). Logistic regression, multifactor dimensionality reduction, linkage disequilibrium and entropy based statistics are examples of such methods (Hieke et al., 2014; Wan et al., 2010). While most of these methods are based on the unit association of the SNPs, testing the associations between the phenotype and SNPs has limitations and is not sufficient for the interpretation of GGIs (Yuan et al., 2012).

A number of multimodal fusion methods such as co-training, multi-view learning, subspace learning, multi-view embedding, and kernel multiple learning, have been developed to analyze multimodal data of biological relevance (Xu et al., 2013). Recently, positive definite kernel based methods have become an effective tool in imaging genetics. For example, they have been used for identifying genes associated with diseases (Li and Cui, 2012; Ge et al., 2015; Alam et al., 2016a,b). Kernel methods offer useful ways to learn how a large collection of genetic variants are associated with complex phenotypes, to help explore the relationship between genetic markers and a disease state (Camps-Valls et al., 2007; S. Yu and Moreau, 2011; Alam, 2014; Alam and Fukumizu, 2015; Schölkopf et al., 1998; Kung, 2014). Linear, kernel, and robust canonical correlation based U statistics have been utilized to identify gene-gene co-associations (Peng et al., 2010; Alam et al., 2016b). In Li and Cui (2012) a model-based kernel machine method was proposed for detecting GGIs. In addition, in Ge et al. (2015) authors have also proposed a kernel machine method for detecting effects of interactions between multi-variable sets. This is an extended model of (Li and Cui, 2012) to jointly model the genetics and non-genetic features, and their interactions. While these methods could ultimately shed light on novel features of the etiology of complex diseases, they cannot be reliably used for multimodal datasets. By checking the properties of the test statistic such as the consistency in testing, these methods are not rendered more effective. Also, these methods cannot be generalized for multi-modal datasets, because each modality has a specific statistical property. In contrast, multiple-modality based association test can gain additional power by considering joint effects of multiple variants, promising for improved testing power.

The goal of this paper is therefore to develop a novel kernel based method for the study of high-order interactions. First, we propose a novel semiparametric method, namely, a kernel machine method for detecting higher order interactions (KMDHOI), which includes the pairwise and higher order Hadamard product of the features from different views. Second,

we formulate the problem as a mixed-effect linear model to derive a score-based variance component test for the higher order interactions. The proposed method offers a flexible framework to account for the main (single), pairwise, and other higher order effects. Third, we validate the proposed method on both simulated and real MCIC data (Chen et al., 2012; Gollub et al., 2013).

The remainder of the paper is organized as follows. In Section 2, we propose a linear mixed-effects model to derive score-based variance component test. In Section 3, we propose statistical testing for higher order interaction effects. The relevant methods are discussed in Section 4. In Section 5, we describe the experiments conducted on both synthesized and real imaging genetics data. We conclude the paper with a discussion on major findings and future research in Section 6. Details for the analysis of the proposed method, Satterthwaite approximation to the score test, and the applications to MCIC data can be found in the supplementary material.

2. Method

Kernel methods map data from a high dimensional space to a feature space using a nonlinear feature map. The main advantage of these methods is to combine statistics and geometry in an effective way (Hofmann et al., 2008; Richfield et al., 2017; Alam and Fukumizu, 2014). In kernel methods, the nonlinear feature map is defined by a **positive definite kernel**. It is known (Aronszajn, 1950) that a positive definite kernel k is associated with a Hilbert space $\mathcal{H}$, called **reproducing kernel Hilbert space** (RKHS), consisting of functions on $\mathcal{X}$ so that the function value is reproduced by the kernel. Namely, for any function $f \in \mathcal{H}$ and a point $X \in \mathcal{X}$, the function value $f(X)$ is $f(X) = \langle f(\cdot), k(\cdot, X) \rangle_{\mathcal{H}}$, where $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is the inner product of $\mathcal{H}$ is called the reproducing property. Replacing f with $k(\cdot, \tilde{X})$ yields $k(X, \tilde{X}) = \langle k(\cdot, X), k(\cdot, \tilde{X}) \rangle_{\mathcal{H}}$ for any $X, \tilde{X} \in \mathcal{X}$. A symmetric kernel $k(\cdot, \cdot)$ defined on a space $\mathcal{X}$ is called **positive definite**, if for an arbitrary number of points $X_1, X_2, \dots, X_n \in \mathcal{X}$ the Gram matrix $(k(X_i, X_j))_{ij}$ is positive semi-definite. To transform data for extracting nonlinear features, the mapping $\Phi: \mathcal{X} \rightarrow \mathcal{H}$ is defined as $\Phi(X) = k(\cdot, X)$, which is a function of the first argument. This map is called the **feature map**, and the vector $\Phi(X)$ in $\mathcal{H}$ is called the **feature vector**. The inner product of two feature vectors is then $\langle \Phi(X), \Phi(\tilde{X}) \rangle_{\mathcal{H}} = k(X, \tilde{X})$.

This is known as the **kernel trick**. By this trick the kernel can evaluate the inner product of any two feature vectors efficiently without knowing an explicit form of $\Phi(\cdot)$.

2.1. Model setting

Assuming that we have n independent identical distributed (IID) subjects under investigation; y_i ($i = 1, 2, \dots, n$) with $(q - 1)$ covariates $X_i = [X_{i1}, X_{i2}, \dots, X_{i(q-1)}]^T$ and m -modal datasets, $\mathbf{M}_i^{(1)}, \mathbf{M}_i^{(2)}, \dots, \mathbf{M}_i^{(m)}$, we associate the output y_i with covariates and m -view datasets in the following semiparametric model:

$$y_i = X_i^T \boldsymbol{\beta} + f(\mathbf{M}_i^{(1)}, \mathbf{M}_i^{(2)}, \dots, \mathbf{M}_i^{(m)}) + \epsilon_i, \quad (1)$$

where X_i is a vector of q covariates including intercept for the i -th subject, β is a vector of q fixed effects, f is an unknown function on the product domain,

$\mathcal{M} = \mathcal{M}^{(1)} \otimes \mathcal{M}^{(2)} \otimes \dots \otimes \mathcal{M}^{(m)}$ with and $\mathbf{M}_i^{(\ell)} \in \mathcal{M}_\ell$, $\ell = 1, 2, \dots, m$ the error ϵ_i 's are IID as normal with mean zero and variance σ^2 , $\epsilon_i \sim \text{NIID}(0, \sigma^2)$. According to the ANOVA decomposition, the function, f can be extended as:

$$\begin{aligned} f(\mathbf{M}_i^{(1)}, \mathbf{M}_i^{(2)}, \dots, \mathbf{M}_i^{(m)}) = & \sum_{\ell=1}^m h_{\ell}(\mathbf{M}_i^{(\ell)}) + \sum_{\ell_1 > \ell_2} h_{\ell_1 \times \ell_2}(\mathbf{M}_i^{(\ell_1)}, \mathbf{M}_i^{(\ell_2)}) + \quad (2) \\ & \sum_{\ell_1 > \ell_2 > \ell_3} h_{\ell_1 \times \ell_2 \times \ell_3}(\mathbf{M}_i^{(\ell_1)}, \mathbf{M}_i^{(\ell_2)}, \mathbf{M}_i^{(\ell_3)}) + \dots \\ & + h_{\ell_1 \times \ell_2 \times \ell_3 \times \dots \times \ell_m}(\mathbf{M}_i^{(\ell_1)}, \mathbf{M}_i^{(\ell_2)}, \dots, \mathbf{M}_i^{(\ell_m)}), \end{aligned}$$

where $h_{\ell}(\mathbf{M}_i^{(\ell)})$'s ($\ell = 1, 2, \dots, m$) are the main effects for the respective dataset,

$h_{\ell_1 \times \ell_2}(\mathbf{M}_i^{(\ell_1)}, \mathbf{M}_i^{(\ell_2)})$ are pairwise interactions effects, $h_{\ell_1 \times \ell_2 \times \ell_3}(\mathbf{M}_i^{(\ell_1)}, \mathbf{M}_i^{(\ell_2)}, \mathbf{M}_i^{(\ell_3)})$ are the interactions effects of the three dataset and so on. The functional space, RKHS, is decomposed as:

$$\begin{aligned} \mathcal{H} = & \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \dots \oplus \mathcal{H}_m \oplus \mathcal{H}_{1 \times 2} \oplus \mathcal{H}_{1 \times 3} \oplus \dots \oplus \mathcal{H}_{1 \times m} \oplus \mathcal{H}_{2 \times 3} \quad (3) \\ & \oplus \dots \oplus \mathcal{H}_{1 \times 2 \times 3} \oplus \dots \\ & \oplus \mathcal{H}_{1 \times 2 \times \dots \times m}, \end{aligned}$$

equipped with an inner product $\langle \cdot, \cdot \rangle$ and a norm $\left\| \cdot \right\|_{\mathcal{H}}$. If $m = 1$, Eq. (1) becomes simple

semiparametric regression model as shown in (Liu et al., 2007). Li and Cui (2012) and Ge et al. (2015) have proposed similar models (special case of Eq. (1), $m = 2$) for detecting interaction effects among multidimensional variable sets.

Specifically, in our case we have three data sets (see Section 5.1.2 for the explanation of the data set). To do this, we assume that we have n IID subjects under investigation: y_i ($i = 1, 2, \dots, n$) is a quantitative trait for the i -th subject (say, hippocampal volume derived from structural MRI scan). It is associated with the clinical covariates (e.g., age, weight, and height) and three data sets (e.g., gene-derived SNPs, ROIs, and gene-specific DNA methylation). Let $\mathbf{M}_i^{(1)}, \mathbf{M}_i^{(2)}, \mathbf{M}_i^{(3)}$ be the three-modal data sets and $\mathbf{X}_i$ denote the $(q - 1)$ covariates, where $X_{ij}; j = 1, 2, \dots, (q - 1)$ is a measure of the i -th subject. For each i -th subject, let $M_i^{(1)} = [M_{i1}^{(1)}, M_{i2}^{(1)}, \dots, M_{is}^{(1)}]$, $M_i^{(2)} = [M_{i1}^{(2)}, M_{i2}^{(2)}, \dots, M_{ir}^{(2)}]$ and $M_i^{(3)} = [M_{i1}^{(3)}, M_{i2}^{(3)}, \dots, M_{id}^{(3)}]$ be

the gene-derived SNPs with s SNP markers, the ROI with r voxels in the fMRI scan, and the gene-specific DNA methylation with d methylation profiles, respectively. Under this setting, Eq. (1), Eq. (2) and Eq. (3) become:

$$y_i = X_i^T \beta + f(\mathbf{M}_i^{(1)}, \mathbf{M}_i^{(2)}, \mathbf{M}_i^{(3)}) + \epsilon_i, \quad (4)$$

$$\begin{aligned} f(\mathbf{M}_i^{(1)}, \mathbf{M}_i^{(2)}, \mathbf{M}_i^{(3)}) = & h_1(\mathbf{M}_i^{(1)}) + h_2(\mathbf{M}_i^{(2)}) + h_3(\mathbf{M}_i^{(3)}) + h_{1 \times 2}(\mathbf{M}_i^{(1)}, \mathbf{M}_i^{(2)}) + \\ & h_{1 \times 3}(\mathbf{M}_i^{(1)}, \mathbf{M}_i^{(3)}) + h_{2 \times 3}(\mathbf{M}_i^{(2)}, \mathbf{M}_i^{(3)}) \\ & + h_{1 \times 2 \times 3}(\mathbf{M}_i^{(1)}, \mathbf{M}_i^{(2)}, \mathbf{M}_i^{(3)}), \end{aligned} \quad (5)$$

and

$$\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2 \oplus \mathcal{H}_3 \oplus \mathcal{H}_{1 \times 2} \oplus \mathcal{H}_{1 \times 3} \oplus \mathcal{H}_{2 \times 3} \oplus \mathcal{H}_{1 \times 2 \times 3}, \quad (6)$$

respectively. Here $\mathcal{H}_1, \mathcal{H}_2$ and $\mathcal{H}_3, \mathcal{H}_{1 \times 2}, \mathcal{H}_{1 \times 3}, \mathcal{H}_{2 \times 3}$, and $\mathcal{H}_{1 \times 2 \times 3}$ are RKHSs functions on $\mathcal{M}_1, \mathcal{M}_2$ and $\mathcal{M}_3, \mathcal{M}_1 \times \mathcal{M}_2, \mathcal{M}_1 \times \mathcal{M}_3$ and $\mathcal{M}_2 \times \mathcal{M}_3$ and $\mathcal{M}_1 \times \mathcal{M}_2 \times \mathcal{M}_3$, respectively. The notation $\oplus$ is a direct sum of RKHS.

2.2. Model estimation

We can estimate the function $f \in \mathcal{H}$ by minimizing the penalized squared error loss function of Eq. (4) as follows:

$$\mathcal{L}(\mathbf{y}, \beta, f) = \frac{1}{2} \sum_{i=1}^n [y_i - X_i^T \beta - f(\mathbf{M}_i^{(1)}, \mathbf{M}_i^{(2)}, \mathbf{M}_i^{(3)})]^2 + \frac{\lambda}{2} \mathcal{J}(f) \quad (7)$$

where $\mathcal{J}(\cdot) = \|\cdot\|_{\mathcal{H}}^2$ is a roughness penalty with tuning parameter λ . It is known that the complete function space of Eq. (6), $\mathcal{H}$, has the orthogonal decomposition. Hence the function $\mathcal{J}(\cdot)$ can be decomposed accordingly. Eq. (7) then becomes:

$$\begin{aligned}
\mathcal{L}(\mathbf{y}, \boldsymbol{\beta}, f) &= \frac{1}{2} \sum_{i=1}^n \left[y_i - X_i^T \boldsymbol{\beta} - h_1(\mathbf{M}_i^{(1)}) - h_2(\mathbf{M}_i^{(2)}) - h_3(\mathbf{M}_i^{(3)}) - h_{1 \times 2}(\mathbf{M}_i^{(1)}, \mathbf{M}_i^{(2)}) \right. \\
&\quad \left. - h_{1 \times 3}(\mathbf{M}_i^{(1)}, \mathbf{M}_i^{(3)}) - h_{2 \times 3}(\mathbf{M}_i^{(2)}, \mathbf{M}_i^{(3)}) - h_{1 \times 2 \times 3}(\mathbf{M}_i^{(1)}, \mathbf{M}_i^{(2)}, \mathbf{M}_i^{(3)}) \right]^2 \\
&\quad + \frac{\lambda^{(1)}}{2} \|h_1\|^2 + \frac{\lambda^{(2)}}{2} \|h_2\|^2 + \frac{\lambda^{(3)}}{2} \|h_3\|^2 + \frac{\lambda^{(1 \times 2)}}{2} \|h_{1 \times 2}\|^2 + \frac{\lambda^{(1 \times 3)}}{2} \|h_{1 \times 3}\|^2 \\
&\quad + \frac{\lambda^{(2 \times 3)}}{2} \|h_{2 \times 3}\|^2 + \frac{\lambda^{(1 \times 2 \times 3)}}{2} \|h_{1 \times 2 \times 3}\|^2 \\
&= [\mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{h}_1 - \mathbf{h}_2 - \mathbf{h}_3 - \mathbf{h}_{1 \times 2} - \mathbf{h}_{1 \times 3} \\
&\quad - \mathbf{h}_{2 \times 3} - \mathbf{h}_{1 \times 2 \times 3}]^2 + \frac{\lambda^{(1)}}{2} \|h_1\|^2 + \frac{\lambda^{(2)}}{2} \|h_2\|^2 + \frac{\lambda^{(3)}}{2} \|h_3\|^2 \\
&\quad + \frac{\lambda^{(1 \times 2)}}{2} \|h_{1 \times 2}\|^2 + \frac{\lambda^{(1 \times 3)}}{2} \|h_{1 \times 3}\|^2 + \frac{\lambda^{(2 \times 3)}}{2} \|h_{2 \times 3}\|^2 \\
&\quad + \frac{\lambda^{(1 \times 2 \times 3)}}{2} \|h_{1 \times 2 \times 3}\|^2,
\end{aligned}
\tag{8}$$

where

$$\begin{aligned}
\mathbf{y} &= [y_1, y_2, \dots, y_n]^T, \quad \mathbf{X} = [X_1, X_2, \dots, X_n]^T, \\
\mathbf{h}_1 &= [h_1(\mathbf{M}_1^{(1)}), h_1(\mathbf{M}_2^{(1)}), \dots, h_1(\mathbf{M}_n^{(1)})]^T, \\
\mathbf{h}_2 &= [h_2(\mathbf{M}_1^{(2)}), h_2(\mathbf{M}_2^{(2)}), \dots, h_2(\mathbf{M}_n^{(2)})]^T, \\
\mathbf{h}_3 &= [h_3(\mathbf{M}_1^{(3)}), h_3(\mathbf{M}_2^{(3)}), \dots, h_3(\mathbf{M}_n^{(3)})]^T, \\
\mathbf{h}_{1 \times 2} &= [h_{1 \times 2}(\mathbf{M}_1^{(1)}, \mathbf{M}_1^{(2)}), h_{1 \times 2}(\mathbf{M}_2^{(1)}, \mathbf{M}_2^{(2)}), \dots, h_{1 \times 2}(\mathbf{M}_n^{(1)}, \mathbf{M}_n^{(2)})]^T, \\
\mathbf{h}_{1 \times 3} &= [h_{1 \times 3}(\mathbf{M}_1^{(1)}, \mathbf{M}_1^{(3)}), h_{1 \times 3}(\mathbf{M}_2^{(1)}, \mathbf{M}_2^{(3)}), \dots, h_{1 \times 3}(\mathbf{M}_n^{(1)}, \mathbf{M}_n^{(3)})]^T, \\
\mathbf{h}_{2 \times 3} &= [h_{2 \times 3}(\mathbf{M}_1^{(2)}, \mathbf{M}_1^{(3)}), h_{2 \times 3}(\mathbf{M}_2^{(2)}, \mathbf{M}_2^{(3)}), \dots, h_{2 \times 3}(\mathbf{M}_n^{(2)}, \mathbf{M}_n^{(3)})]^T, \\
\mathbf{h}_{1 \times 2 \times 3} &= [h_{1 \times 2 \times 3}(\mathbf{M}_1^{(1)}, \mathbf{M}_1^{(2)}, \mathbf{M}_1^{(3)}), h_{1 \times 2 \times 3}(\mathbf{M}_2^{(1)}, \mathbf{M}_2^{(2)}, \mathbf{M}_2^{(3)}), \dots, h_{1 \times 2 \times 3}(\mathbf{M}_n^{(1)}, \mathbf{M}_n^{(2)}, \mathbf{M}_n^{(3)})]^T,
\end{aligned}$$

also $\lambda^{(1)}, \lambda^{(2)}, \lambda^{(3)}, \lambda^{(1 \times 2)}, \lambda^{(1 \times 3)}, \lambda^{(2 \times 3)}$ and $\lambda^{(1 \times 2 \times 3)}$ are the positive tuning parameters that make trade-off between the model fitting and the complexity.

By the representer theorem (Kimeldorf and Wahhba, 1971; Schölkopf and Smola, 2002) and the fact that the reproducing kernel of a product of an RKHS is the product of the reproducing kernels (Aronszajn, 1950), the expanded functions of f in Eq. (8) for arbitrary $\widetilde{\mathbf{M}}^{(1)} \in \mathcal{M}^{(1)}, \widetilde{\mathbf{M}}^{(2)} \in \mathcal{M}^{(2)}$ and $\widetilde{\mathbf{M}}^{(3)} \in \mathcal{M}^{(3)}$ can be written as:

$$\begin{aligned}
h_1 &= \sum_{i=1}^n \alpha_i^{(1)} k^{(1)}(\bar{\mathbf{M}}^{(1)}, \mathbf{M}_i^{(1)}), \\
h_2 &= \sum_{i=1}^n \alpha_i^{(2)} k^{(2)}(\bar{\mathbf{M}}^{(2)}, \mathbf{M}_i^{(2)}), \\
h_3 &= \sum_{i=1}^n \alpha_i^{(3)} k^{(3)}(\bar{\mathbf{M}}^{(3)}, \mathbf{M}_i^{(3)}), \\
h_{1 \times 2} &= \sum_{i=1}^n \alpha_i^{(1 \times 2)} k^{(1)}(\bar{\mathbf{M}}^{(1)}, \mathbf{M}_i^{(1)}) k^{(2)}(\bar{\mathbf{M}}^{(2)}, \mathbf{M}_i^{(2)}), \\
h_{1 \times 3} &= \sum_{i=1}^n \alpha_i^{(1 \times 3)} k^{(1)}(\bar{\mathbf{M}}^{(1)}, \mathbf{M}_i^{(1)}) k^{(3)}(\bar{\mathbf{M}}^{(3)}, \mathbf{M}_i^{(3)}), \\
h_{2 \times 3} &= \sum_{i=1}^n \alpha_i^{(2 \times 3)} k^{(2)}(\bar{\mathbf{M}}^{(2)}, \mathbf{M}_i^{(2)}) k^{(3)}(\bar{\mathbf{M}}^{(3)}, \mathbf{M}_i^{(3)}), \\
h_{1 \times 2 \times 3} &= \sum_{i=1}^n \alpha_i^{(1 \times 2 \times 3)} k^{(1)}(\bar{\mathbf{M}}^{(1)}, \mathbf{M}_i^{(1)}) k^{(2)}(\bar{\mathbf{M}}^{(2)}, \mathbf{M}_i^{(2)}) k^{(3)}(\bar{\mathbf{M}}^{(3)}, \mathbf{M}_i^{(3)}).
\end{aligned}$$

For each data view, we can define the kernel matrices: $\mathbf{K}^{(1)} = (k^{(1)}(\mathbf{M}_i^1, \mathbf{M}_j^1))_{ij}$, $\mathbf{K}^{(2)} = (k^{(2)}(\mathbf{M}_i^2, \mathbf{M}_j^2))_{ij}$, $\mathbf{K}^{(3)} = (k^{(3)}(\mathbf{M}_i^3, \mathbf{M}_j^3))_{ij}$, $\mathbf{K}^{(1 \times 2)} = \mathbf{K}^{(1)} \odot \mathbf{K}^{(2)}$, $\mathbf{K}^{(1 \times 3)} = \mathbf{K}^{(1)} \odot \mathbf{K}^{(3)}$, $\mathbf{K}^{(2 \times 3)} = \mathbf{K}^{(2)} \odot \mathbf{K}^{(3)}$ and $\mathbf{K}^{(1 \times 2 \times 3)} = \mathbf{K}^{(1)} \odot \mathbf{K}^{(2)} \odot \mathbf{K}^{(3)}$, where $\odot$ denotes the element-wise product of two matrices. Now we have

$$\begin{aligned}
\mathbf{h}_1 &= \mathbf{K}^{(1)} \boldsymbol{\alpha}^{(1)}, \mathbf{h}_2 = \mathbf{K}^{(2)} \boldsymbol{\alpha}^{(2)}, \mathbf{h}_3 = \mathbf{K}^{(3)} \boldsymbol{\alpha}^{(3)}, \mathbf{h}_{1 \times 2} = \mathbf{K}^{(1 \times 2)} \boldsymbol{\alpha}^{(1 \times 2)}, \\
\mathbf{h}_{1 \times 3} &= \mathbf{K}^{(1 \times 3)} \boldsymbol{\alpha}^{(1 \times 3)}, \mathbf{h}_{2 \times 3} = \mathbf{K}^{(2 \times 3)} \boldsymbol{\alpha}^{(2 \times 3)}, \mathbf{h}_{1 \times 2 \times 3} = \mathbf{K}^{(1 \times 2 \times 3)} \boldsymbol{\alpha}^{(1 \times 2 \times 3)},
\end{aligned} \quad (9)$$

where $\boldsymbol{\alpha}^{(1)} = [\alpha_1^{(1)}, \alpha_2^{(1)}, \dots, \alpha_n^{(1)}]^T$, $\boldsymbol{\alpha}^{(2)} = [\alpha_1^{(2)}, \alpha_2^{(2)}, \dots, \alpha_n^{(2)}]^T$, $\boldsymbol{\alpha}^{(3)} = [\alpha_1^{(3)}, \alpha_2^{(3)}, \dots, \alpha_n^{(3)}]^T$, $\boldsymbol{\alpha}^{(1 \times 2)} = [\alpha_1^{(1 \times 2)}, \alpha_2^{(1 \times 2)}, \dots, \alpha_n^{(1 \times 2)}]^T$, $\boldsymbol{\alpha}^{(1 \times 3)} = [\alpha_1^{(1 \times 3)}, \alpha_2^{(1 \times 3)}, \dots, \alpha_n^{(1 \times 3)}]^T$, $\boldsymbol{\alpha}^{(2 \times 3)} = [\alpha_1^{(2 \times 3)}, \alpha_2^{(2 \times 3)}, \dots, \alpha_n^{(2 \times 3)}]^T$ and $\boldsymbol{\alpha}^{(1 \times 2 \times 3)} = [\alpha_1^{(1 \times 2 \times 3)}, \alpha_2^{(1 \times 2 \times 3)}, \dots, \alpha_n^{(1 \times 2 \times 3)}]^T$.

Substituting $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \mathbf{h}_{1 \times 2}, \mathbf{h}_{1 \times 3}, \mathbf{h}_{2 \times 3}$ and $\mathbf{h}_{1 \times 2 \times 3}$ into Eq. (8), and applying the reproducing kernel properties, we get

$$\begin{aligned}
\mathcal{L}(\mathbf{y}, \boldsymbol{\beta}, \boldsymbol{\alpha}) &= \frac{1}{2} \boldsymbol{\epsilon}^T \boldsymbol{\epsilon} + \frac{\lambda^{(1)}}{2} [\boldsymbol{\alpha}^{(1)}]^T \mathbf{K}^{(1)} \boldsymbol{\alpha}^{(1)} + \frac{\lambda^{(2)}}{2} [\boldsymbol{\alpha}^{(2)}]^T \mathbf{K}^{(2)} \boldsymbol{\alpha}^{(2)} + \frac{\lambda^{(3)}}{2} [\boldsymbol{\alpha}^{(3)}]^T \mathbf{K}^{(3)} \boldsymbol{\alpha}^{(3)} \\
&+ \frac{\lambda^{(1 \times 2)}}{2} [\boldsymbol{\alpha}^{(1 \times 2)}]^T \mathbf{K}^{(1 \times 2)} \boldsymbol{\alpha}^{(1 \times 2)} + \frac{\lambda^{(1 \times 3)}}{2} [\boldsymbol{\alpha}^{(1 \times 3)}]^T \mathbf{K}^{(1 \times 3)} \boldsymbol{\alpha}^{(1 \times 3)} \\
&+ \frac{\lambda^{(2 \times 3)}}{2} [\boldsymbol{\alpha}^{(2 \times 3)}]^T \mathbf{K}^{(2 \times 3)} \boldsymbol{\alpha}^{(2 \times 3)} \\
&+ \frac{\lambda^{(1 \times 2 \times 3)}}{2} [\boldsymbol{\alpha}^{(1 \times 2 \times 3)}]^T \mathbf{K}^{(1 \times 2 \times 3)} \boldsymbol{\alpha}^{(1 \times 2 \times 3)}
\end{aligned} \quad (10)$$

where $\epsilon = \mathbf{y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{K}^{(1)}\boldsymbol{\alpha}^{(1)} - \mathbf{K}^{(2)}\boldsymbol{\alpha}^{(2)} - \mathbf{K}^{(3)}\boldsymbol{\alpha}^{(3)} - \mathbf{K}^{(1 \times 2)}\boldsymbol{\alpha}^{(1 \times 2)} - \mathbf{K}^{(1 \times 3)}\boldsymbol{\alpha}^{(1 \times 3)} - \mathbf{K}^{(2 \times 3)}\boldsymbol{\alpha}^{(2 \times 3)} - \mathbf{K}^{(1 \times 2 \times 3)}\boldsymbol{\alpha}^{(1 \times 2 \times 3)}$ and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}^{(1)}, \boldsymbol{\alpha}^{(2)}, \boldsymbol{\alpha}^{(3)}, \boldsymbol{\alpha}^{(1 \times 2)}, \boldsymbol{\alpha}^{(1 \times 3)}, \boldsymbol{\alpha}^{(2 \times 3)}, \boldsymbol{\alpha}^{(1 \times 2 \times 3)})$.

The gradients of $\mathcal{L}$ with respect to the parametric coefficients $\boldsymbol{\beta}$ and nonparametric coefficients $\boldsymbol{\alpha}'$ s are

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}} &= \mathbf{X}^T \boldsymbol{\epsilon}, \quad \frac{\partial \mathcal{L}}{\partial \boldsymbol{\alpha}^{(1)}} = [\mathbf{K}^{(1)}]^T \boldsymbol{\epsilon} + \lambda^{(1)} \mathbf{K}^{(1)} \boldsymbol{\alpha}^{(1)}, \\ \frac{\partial \mathcal{L}}{\partial \boldsymbol{\alpha}^{(2)}} &= [\mathbf{K}^{(2)}]^T \boldsymbol{\epsilon} + \lambda^{(2)} \mathbf{K}^{(2)} \boldsymbol{\alpha}^{(2)}, \quad \frac{\partial \mathcal{L}}{\partial \boldsymbol{\alpha}^{(3)}} = [\mathbf{K}^{(3)}]^T \boldsymbol{\epsilon} + \lambda^{(3)} \mathbf{K}^{(3)} \boldsymbol{\alpha}^{(3)}, \\ \frac{\partial \mathcal{L}}{\partial \boldsymbol{\alpha}^{(1 \times 2)}} &= [\mathbf{K}^{(1 \times 2)}]^T \boldsymbol{\epsilon} + \lambda^{(1 \times 2)} \mathbf{K}^{(1 \times 2)} \boldsymbol{\alpha}^{(1 \times 2)}, \quad \frac{\partial \mathcal{L}}{\partial \boldsymbol{\alpha}^{(1 \times 3)}} = [\mathbf{K}^{(1 \times 3)}]^T \boldsymbol{\epsilon} \\ &+ \lambda^{(1 \times 3)} \mathbf{K}^{(1 \times 3)} \boldsymbol{\alpha}^{(1 \times 3)}, \\ \frac{\partial \mathcal{L}}{\partial \boldsymbol{\alpha}^{(2 \times 3)}} &= [\mathbf{K}^{(2 \times 3)}]^T \boldsymbol{\epsilon} + \lambda^{(2 \times 3)} \mathbf{K}^{(2 \times 3)} \boldsymbol{\alpha}^{(2 \times 3)}, \quad \frac{\partial \mathcal{L}}{\partial \boldsymbol{\alpha}^{(1 \times 2 \times 3)}} = [\mathbf{K}^{(1 \times 2 \times 3)}]^T \boldsymbol{\epsilon} \\ &+ \lambda^{(1 \times 2 \times 3)} \mathbf{K}^{(1 \times 2 \times 3)} \boldsymbol{\alpha}^{(1 \times 2 \times 3)} \end{aligned}$$

(11)

By setting the gradients to be zeros, this first-order condition is given by the following linear systems:

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

$$\begin{array}{cccccccc}
 \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{K}^{(1)} & \mathbf{X}^T \mathbf{K}^{(2)} & \mathbf{X}^T \mathbf{K}^{(3)} & \mathbf{X}^T \mathbf{K}^{(1 \times 2)} & \mathbf{X}^T \mathbf{K}^{(1 \times 3)} & \mathbf{X}^T \mathbf{K}^{(2 \times 3)} & \mathbf{X}^T \mathbf{K}^{(1 \times 2 \times 3)} \\
 \left[\mathbf{K}^{(1)} \right]^T \mathbf{X} & A & \left[\mathbf{K}^{(1)} \right]^T \mathbf{K}^{(2)} & \left[\mathbf{K}^{(1)} \right]^T \mathbf{K}^{(3)} & \left[\mathbf{K}^{(1)} \right]^T \mathbf{K}^{(1 \times 2)} & \left[\mathbf{K}^{(1)} \right]^T \mathbf{K}^{(1 \times 3)} & \left[\mathbf{K}^{(1)} \right]^T \mathbf{K}^{(2 \times 3)} & \left[\mathbf{K}^{(1)} \right]^T \mathbf{K}^{(1 \times 2 \times 3)} \\
 \left[\mathbf{K}^{(2)} \right]^T \mathbf{X} & \left[\mathbf{K}^{(2)} \right]^T \mathbf{K}^{(1)} & B & \left[\mathbf{K}^{(2)} \right]^T \mathbf{K}^{(3)} & \left[\mathbf{K}^{(2)} \right]^T \mathbf{K}^{(1 \times 2)} & \left[\mathbf{K}^{(2)} \right]^T \mathbf{K}^{(1 \times 3)} & \left[\mathbf{K}^{(2)} \right]^T \mathbf{K}^{(2 \times 3)} & \left[\mathbf{K}^{(2)} \right]^T \mathbf{K}^{(1 \times 2 \times 3)} \\
 \left[\mathbf{K}^{(3)} \right]^T \mathbf{X} & \left[\mathbf{K}^{(3)} \right]^T \mathbf{K}^{(1)} & \left[\mathbf{K}^{(3)} \right]^T \mathbf{K}^{(2)} & C & \left[\mathbf{K}^{(3)} \right]^T \mathbf{K}^{(1 \times 2)} & \left[\mathbf{K}^{(3)} \right]^T \mathbf{K}^{(1 \times 3)} & \left[\mathbf{K}^{(3)} \right]^T \mathbf{K}^{(2 \times 3)} & \left[\mathbf{K}^{(3)} \right]^T \mathbf{K}^{(1 \times 2 \times 3)} \\
 \left[\mathbf{K}^{(1 \times 2)} \right]^T \mathbf{X} & \left[\mathbf{K}^{(1 \times 2)} \right]^T \mathbf{K}^{(1)} & \left[\mathbf{K}^{(1 \times 2)} \right]^T \mathbf{K}^{(2)} & \left[\mathbf{K}^{(1 \times 2)} \right]^T \mathbf{K}^{(3)} & D & \left[\mathbf{K}^{(1 \times 2)} \right]^T \mathbf{K}^{(1 \times 3)} & \left[\mathbf{K}^{(1 \times 2)} \right]^T \mathbf{K}^{(2 \times 3)} & \left[\mathbf{K}^{(1 \times 2)} \right]^T \mathbf{K}^{(1 \times 2 \times 3)} \\
 \left[\mathbf{K}^{(1 \times 3)} \right]^T \mathbf{X} & \left[\mathbf{K}^{(1 \times 3)} \right]^T \mathbf{K}^{(1)} & \left[\mathbf{K}^{(1 \times 3)} \right]^T \mathbf{K}^{(2)} & \left[\mathbf{K}^{(1 \times 3)} \right]^T \mathbf{K}^{(3)} & \left[\mathbf{K}^{(1 \times 3)} \right]^T \mathbf{K}^{(1 \times 2)} & E & \left[\mathbf{K}^{(1 \times 3)} \right]^T \mathbf{K}^{(2 \times 3)} & \left[\mathbf{K}^{(1 \times 3)} \right]^T \mathbf{K}^{(1 \times 2 \times 3)} \\
 \left[\mathbf{K}^{(2 \times 3)} \right]^T \mathbf{X} & \left[\mathbf{K}^{(2 \times 3)} \right]^T \mathbf{K}^{(1)} & \left[\mathbf{K}^{(2 \times 3)} \right]^T \mathbf{K}^{(2)} & \left[\mathbf{K}^{(2 \times 3)} \right]^T \mathbf{K}^{(3)} & \left[\mathbf{K}^{(2 \times 3)} \right]^T \mathbf{K}^{(1 \times 2)} & \left[\mathbf{K}^{(2 \times 3)} \right]^T \mathbf{K}^{(1 \times 3)} & F & \left[\mathbf{K}^{(2 \times 3)} \right]^T \mathbf{K}^{(1 \times 2 \times 3)} \\
 \left[\mathbf{K}^{(1 \times 2 \times 3)} \right]^T \mathbf{X} & \left[\mathbf{K}^{(1 \times 2 \times 3)} \right]^T \mathbf{K}^{(1)} & \left[\mathbf{K}^{(1 \times 2 \times 3)} \right]^T \mathbf{K}^{(2)} & \left[\mathbf{K}^{(1 \times 2 \times 3)} \right]^T \mathbf{K}^{(3)} & \left[\mathbf{K}^{(1 \times 2 \times 3)} \right]^T \mathbf{K}^{(1 \times 2)} & \left[\mathbf{K}^{(1 \times 2 \times 3)} \right]^T \mathbf{K}^{(1 \times 3)} & \left[\mathbf{K}^{(1 \times 2 \times 3)} \right]^T \mathbf{K}^{(2 \times 3)} & G
 \end{array}$$

$$\begin{bmatrix}
 \beta \\
 \alpha^{(1)} \\
 \alpha^{(2)} \\
 \alpha^{(3)} \\
 \alpha^{(1 \times 2)} \\
 \alpha^{(1 \times 3)} \\
 \alpha^{(2 \times 3)} \\
 \alpha^{(1 \times 2 \times 3)}
 \end{bmatrix}$$

$$= \begin{bmatrix}
 \mathbf{X}^T \mathbf{y} \\
 \mathbf{K}^{(1)} \mathbf{y} \\
 \mathbf{K}^{(2)} \mathbf{y} \\
 \mathbf{K}^{(3)} \mathbf{y} \\
 \mathbf{K}^{(1 \times 2)} \mathbf{y} \\
 \mathbf{K}^{(1 \times 3)} \mathbf{y} \\
 \mathbf{K}^{(2 \times 3)} \mathbf{y} \\
 \mathbf{K}^{(1 \times 2 \times 3)} \mathbf{y}
 \end{bmatrix}$$

(12)

where $A = [\mathbf{K}^{(1)}]^T \mathbf{K}^{(1)} + \lambda^{(1)} \mathbf{K}^{(1)}$, $B = [\mathbf{K}^{(2)}]^T \mathbf{K}^{(2)} + \lambda^{(2)} \mathbf{K}^{(2)}$, $C = [\mathbf{K}^{(3)}]^T \mathbf{K}^{(3)} + \lambda^{(3)} \mathbf{K}^{(3)}$,
 $D = [\mathbf{K}^{(1 \times 2)}]^T \mathbf{K}^{(1 \times 2)} + \lambda^{(1 \times 2)} \mathbf{K}^{(1 \times 2)}$, $E = [\mathbf{K}^{(1 \times 3)}]^T \mathbf{K}^{(1 \times 3)} + \lambda^{(1 \times 3)} \mathbf{K}^{(1 \times 3)}$,
 $F = [\mathbf{K}^{(2 \times 3)}]^T \mathbf{K}^{(2 \times 3)} + \lambda^{(2 \times 3)} \mathbf{K}^{(2 \times 3)}$ and
 $G = [\mathbf{K}^{(1 \times 2 \times 3)}]^T \mathbf{K}^{(1 \times 2 \times 3)} + \lambda^{(1 \times 2 \times 3)} \mathbf{K}^{(1 \times 2 \times 3)}$. Following many derivations in the literature (e.g., (Liu et al., 2007; Li and Cui, 2012; Ge et al., 2015)), we can show that the first-order linear system is equivalent to the normal equation of the linear mixed effects model:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{h}_{1 \times 2} + \mathbf{h}_{1 \times 3} + \mathbf{h}_{2 \times 3} + \mathbf{h}_{1 \times 2 \times 3} + \boldsymbol{\epsilon}, \quad (13)$$

where $\boldsymbol{\beta}$ is the coefficient vector of fixed effects, $\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3, \mathbf{h}_{1 \times 2}, \mathbf{h}_{1 \times 3}, \mathbf{h}_{2 \times 3}$ and $\mathbf{h}_{1 \times 2 \times 3}$ are independent random effects with distribution as $\mathbf{h}_1 \sim N(0, \tau^{(1)} \mathbf{K}^{(1)})$, $\tau^{(1)} = \frac{\sigma^2}{\lambda^{(1)}}$,

$$\mathbf{h}_2 \sim N(0, \tau^{(2)} \mathbf{K}^{(2)}), \tau^{(2)} = \frac{\sigma^2}{\lambda^{(2)}}, \mathbf{h}_3 \sim N(0, \tau^{(3)} \mathbf{K}^{(3)}), \tau^{(3)} = \frac{\sigma^2}{\lambda^{(3)}},$$

$$\mathbf{h}_{1 \times 2} \sim N(0, \tau^{(1 \times 2)} \mathbf{K}^{(1 \times 2)}), \tau^{(1 \times 2)} = \frac{\sigma^2}{\lambda^{(1 \times 2)}}, \mathbf{h}_{1 \times 3} \sim N(0, \tau^{(1 \times 3)} \mathbf{K}^{(1 \times 3)}), \tau^{(1 \times 3)} = \frac{\sigma^2}{\lambda^{(1 \times 3)}},$$

$$\mathbf{h}_{2 \times 3} \sim N(0, \tau^{(2 \times 3)} \mathbf{K}^{(2 \times 3)}), \tau^{(2 \times 3)} = \frac{\sigma^2}{\lambda^{(2 \times 3)}},$$

$$\mathbf{h}_{1 \times 2 \times 3} \sim N(0, \tau^{(1 \times 2 \times 3)} \mathbf{K}^{(1 \times 2 \times 3)}), \tau^{(1 \times 2 \times 3)} = \frac{\sigma^2}{\lambda^{(1 \times 2 \times 3)}}. \boldsymbol{\epsilon} \text{ is also an independent random}$$

variable with the distribution $\boldsymbol{\epsilon} \sim N(0, \sigma^2 \mathbf{I})$, where $\mathbf{I}$ is an identity matrix. This relationship ensure that all of the effects extracted by minimizing the loss function in Eq. (7), are the same as the best linear unbiased predictors (BLUPs) of the linear mixed effects model in Eq. (13). To test the higher order interaction effect, testing the null hypothesis is equivalent to testing the variance component. It is possible to estimate the variance components using the restricted maximum likelihood (ReML) approach (see the supplementary material for details). The solution of the linear system in Eq. (12) gives the coefficients of the fixed effect, $\boldsymbol{\beta}$, and the coefficients for the random effect, $\mathbf{a}$. By inserting $\mathbf{a}$ into Eq. (9), we can estimate the random effects $\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2, \hat{\mathbf{h}}_3, \hat{\mathbf{h}}_{1 \times 2}, \hat{\mathbf{h}}_{1 \times 3}, \hat{\mathbf{h}}_{2 \times 3}$ and $\hat{\mathbf{h}}_{1 \times 2 \times 3}$, respectively.

3. Statistical testing

Based on the above regression model in the kernel space, we can derive the statistics to test the interactions among each gene-derived SNP, ROI, and gene-specific DNA methylation. In the following subsections, we study the test statistic for both the overall effect and higher order interaction effects.

3.1. Testing overall effect

We know that the test of overall effect $H_0 : h_1(\cdot) = h_2(\cdot) = h_3(\cdot) = h_{1 \times 2}(\cdot) = h_{1 \times 3}(\cdot) = h_{2 \times 3}(\cdot) = h_{1 \times 2 \times 3}(\cdot) = 0$ is equivalent to testing the variance components in Eq. (13), $H_0 : \tau^{(1)} = \tau^{(2)} = \tau^{(3)} = \tau^{(1 \times 2)} = \tau^{(1 \times 3)} = \tau^{(2 \times 3)} = \tau^{(1 \times 2 \times 3)} = 0$.

Unfortunately, under the null hypothesis, the asymptotic distribution of a likelihood ratio test (LRT) statistic does not follow a chi-square distribution or a mixture chi-square distribution. Because the parameters in the variance components analysis are laid on the boundary of the parameter space when the null hypothesis is true and kernel matrices are not block-diagonal, S. Li and Cui (2012) have proposed a score test statistic based on the restricted likelihood. In this paper, we have constructed a score test statistic for the multimodal datasets in Eq. (13). Assuming that the linear mixed model in Eq. (13) has multivariate normal distribution with mean $\mathbf{X}\boldsymbol{\beta}$ and variance-covariance matrix $\Theta(\boldsymbol{\theta}) = \sigma^2 \mathbf{I} + \tau^{(1)} \mathbf{K}^{(1)} + \tau^{(2)} \mathbf{K}^{(2)} + \tau^{(3)} \mathbf{K}^{(3)} + \tau^{(1 \times 2)} \mathbf{K}^{(1 \times 2)} + \tau^{(1 \times 3)} \mathbf{K}^{(1 \times 3)} + \tau^{(2 \times 3)} \mathbf{K}^{(2 \times 3)} + \tau^{(1 \times 2 \times 3)} \mathbf{K}^{(1 \times 2 \times 3)}$, where $\boldsymbol{\theta} = (\sigma^2, \tau^{(1)}, \tau^{(2)}, \tau^{(3)}, \tau^{(1 \times 2)}, \tau^{(2 \times 3)}, \tau^{(1 \times 2 \times 3)})$ are the variance components, the restricted log-likelihood function of Eq. (13) can be written as

$$\ell_R(\boldsymbol{\theta}) = -\frac{1}{2} \ln(|\Theta(\boldsymbol{\theta})|) - \frac{1}{2} \ln(|\mathbf{X}^T \boldsymbol{\Theta}^{-1}(\boldsymbol{\theta}) \mathbf{X}|) - \frac{1}{2} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T \boldsymbol{\Theta}^{-1}(\boldsymbol{\theta}) (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (14)$$

The estimates of the variance components are obtained by the partial derivative of Eq. (14) with respect to each of the variance components (see supplementary materials for more detail). By considering that the true value of σ^2 under the null hypothesis is σ_0^2 , under the ReML, the score of test statistic is defined as

$$S(\sigma_0^2) = \frac{1}{2\sigma_0^2} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})^T \mathbf{K} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (15)$$

where $\mathbf{K} = \mathbf{K}^{(1)} + \mathbf{K}^{(2)} + \mathbf{K}^{(3)} + \mathbf{K}^{(1 \times 2)} + \mathbf{K}^{(1 \times 3)} + \mathbf{K}^{(2 \times 3)} + \mathbf{K}^{(1 \times 2 \times 3)}$, $\hat{\boldsymbol{\beta}}$ is the maximum likelihood estimator (MLE) of the regression coefficient under the null hypothesis $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}_0$, σ_0^2 is the variance of $\boldsymbol{\epsilon}_0$, and $S(\sigma_0^2)$ is the quadratic function for the variable $\mathbf{y}$, which follows a mixture of the chi-square distribution under the null hypothesis. By the Satterthwaite method (Satterthwaite, 1946), we can approximate the distribution of $S(\sigma_0^2)$ to a scaled chi-square distribution, i.e., $S(\sigma_0^2) \sim \gamma \chi_\nu^2$, where the scale parameter γ and the degrees of freedom ν can be measured by the method of moments (MOM). The mean and variance of the test statistic $S(\sigma_0^2)$ are

$$\mathbb{E}[S(\sigma_0^2)] = \mathbb{E}[\gamma \chi_\nu^2] = \gamma \nu, \quad \text{and} \quad \text{Var}[S(\sigma_0^2)] = \text{Var}[\gamma \chi_\nu^2] = 2\gamma^2 \nu.$$

respectively. By solving the above two equations, we have $\hat{\gamma} = \frac{\text{Var}[S(\sigma_0^2)]}{2E[S(\sigma_0^2)]}$ and $\hat{\nu} = \frac{2E[S(\sigma_0^2)]^2}{\text{Var}[S(\sigma_0^2)]}$.

In practice, σ_0^2 is unknown but we can replace it by its ReML under the null model, which is denoted by $\hat{\sigma}_0^2$. Lastly, the p -value of an experimental score statistic $S(\hat{\sigma}_0^2)$ is obtained using the scaled chi-square distribution $\hat{\gamma}\chi_{\hat{\nu}}^2$.

3.2. Testing higher order interaction effect

To test the higher order interaction effect, we show that testing the null hypothesis $H_0 : h_{1 \times 2 \times 3}(\cdot) = 0$ is equivalent to testing the variance component: $H_0 : \tau^{(1 \times 2 \times 3)} = 0$. Let $\Sigma = \sigma^2 \mathbf{I} + \tau^{(1)} \mathbf{K}^{(1)} + \tau^{(2)} \mathbf{K}^{(2)} + \tau^{(3)} \mathbf{K}^{(3)} + \tau^{(1 \times 2)} \mathbf{K}^{((1 \times 2))} + \tau^{(1 \times 3)} \mathbf{K}^{(1 \times 3)} + \tau^{(2 \times 3)} \mathbf{K}^{(2 \times 3)}$, and $\tau^{(1)}, \tau^{(2)}, \tau^{(3)}, \tau^{(1 \times 2)}, \tau^{(1 \times 3)}, \tau^{(2 \times 3)}$, and σ^2 are model parameters under the null model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{h}_1 + \mathbf{h}_2 + \mathbf{h}_3 + \mathbf{h}_{1 \times 2} + \mathbf{h}_{1 \times 3} + \mathbf{h}_{2 \times 3} + \boldsymbol{\epsilon}$. We formulate the test statistic as:

$$S_I(\tau_I) = \frac{1}{2\sigma_0^2} \mathbf{y}^T \mathbf{W}_{01} \mathbf{K}^{(1 \times 2 \times 3)} \mathbf{W}_{01} \mathbf{y}, \quad (16)$$

where $\tau \mathbf{I} = (\sigma^2, \tau^{(1)}, \tau^{(2)}, \tau^{(3)}, \tau^{(1 \times 2)}, \tau^{(1 \times 3)}, \tau^{(2 \times 3)})$, and $\mathbf{W}_{01} = \Sigma^{-1} - \Sigma^{-1} \mathbf{X} (\mathbf{X}^T \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Sigma^{-1}$ is the projection matrix under the null hypothesis. Similar to the overall effect test, we can use the Satterthwaite method to approximate the distribution for the higher order interaction test statistic $S_I(\tau_I)$ by a scaled chi-square distribution with scaled γ_I and degree of freedom ν_I , i.e., $S_I(\tau_I) \sim \gamma_I \chi_{\nu_I}^2$. The scaled parameter and degree of freedom are estimated by the MOM,

$\hat{\gamma}_I = \frac{\text{Var}[S_I(\tau_I)]}{2E[S_I(\tau_I)]}$ and $\hat{\nu}_I = \frac{2E[S_I(\tau_I)]^2}{\text{Var}[S_I(\tau_I)]}$, respectively. In practice, the unknown model parameters $\tau^{(1)}, \tau^{(2)}, \tau^{(3)}, \tau^{(1 \times 2)}, \tau^{(1 \times 3)}, \tau^{(2 \times 3)}$, and σ^2 are estimated by their respective ReML estimates $\hat{\tau}^{(1)}, \hat{\tau}^{(2)}, \hat{\tau}^{(3)}, \hat{\tau}^{(1 \times 2)}, \hat{\tau}^{(1 \times 3)}, \hat{\tau}^{(2 \times 3)}$, and $\hat{\sigma}^2$ under the null hypothesis. Lastly, the p -value for the observed higher order interaction effect (score statistic $S_I(\tau_I)$) is obtained using the scaled chi-square distribution $\hat{\gamma}_I \chi_{\hat{\nu}_I}^2$.

3.3. Kernel choice

In kernel methods, choosing a suitable kernel is indispensable. Most kernel methods suffer from poor selection of a suitable kernel. It is often the case that the kernel has parameters which may strongly influence the results. Assuming $k: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is a positive definite kernel, then for any $X, \tilde{X} \in \mathcal{X}$, a linear positive definite kernel on $\mathbb{R}$ is defined as

$$k(X, \tilde{X}) = \langle X, \tilde{X} \rangle = X^T \tilde{X}.$$

The linear kernel is used by the underlying Euclidean space to define the similarity measure. Whenever the dimensionality of $\mathbf{X}$ is high, it may allow for more complexity in the function class than what we could measure and assess otherwise. The polynomial kernel is defined as

$$k(X, \tilde{X}) = (X^T \tilde{X} + c)^d, c \geq 0, d \in \mathbb{N}$$

where c and d are two free parameters. c is a trade off between higher-order and lower-order in the polynomial. The polynomial kernel is called homogeneous when $c = 0$. d is the degree of the polynomial and a larger degree tends to lead to overfitting. If d increases, the polynomial kernel tends to zero when $(X^T \tilde{X} + c) < 0$ or tends to infinity when $(X^T \tilde{X} + c) > 0$. Using the polynomial kernel makes it possible to use higher order correlations between data. This kernel incorporates every polynomial interaction up to the degree d (provided that $c > 0$). For instance, if we want to take only the mean and variance into account, we only need to consider $d = 2$ and $c = 1$. For more emphasis on the mean, we need to increase the constant offset c . Polynomial kernels only map data into a finite dimensional space. Due to the finite bounded degree, the given kernel will not provide us with guarantees for a good dependency measure. In addition, both linear and polynomial kernels are unbounded.

Many radial basis function kernels, such as the Gaussian kernel, map X into a infinite dimensional space (Schölkopf and Smola, 2002). The Gaussian kernel is defined as:

$$k(X, \tilde{X}) = e^{-\frac{1}{2\sigma^2} \|X - \tilde{X}\|^2}, (\sigma > 0).$$

While the Gaussian kernel has a free parameter (bandwidth), it still follows a number of theoretical properties such as boundedness, consistence, universality, robustness, etc. It is the most applicable one among the kernel based methods (Sriperumbudur et al., 2009). For the Gaussian kernel, we can use the median of the pairwise distance as a bandwidth (Gretton et al., 2008; Song et al., 2012). This choice is heuristic. A large sigma means that the RBF kernel is very wide. Let us assume that it is so wide that the RBF kernel is still sufficiently positive for every data point. But for the small σ , the RBF kernel is very narrow. This is not desirable, which may give the optimizer a hard job. By considering this issue, the heuristics is a good choice. We must realize that a certain value of σ determines a boundary for the RBF kernel, in which the kernel will be larger than it (one σ quintile for the normal distribution). By choosing the σ according to quintiles on the pairwise distance we ensure that a certain presence of the data points lies within that boundary.

For genome-wide association study (GWAS), a kernel captures the pairwise similarity across a number of SNPs in each gene. Kernel projects the genotype data from original high dimensional space to a feature space. One popular kernel used for genomics similarity is the identity-by-state (IBS) kernel (nonparametric function of the genotypes) (L. C. Kwee, 2008):

$$k(\mathbf{M}_i, \mathbf{M}_j) = 1 - \frac{1}{2s} \sum_{b=1}^s |M_{ib} - M_{jb}|,$$

where s is the number of SNP markers of the corresponding gene. The IBS kernel does not need any assumption on the type of genetic interactions. Thus, in principle, it can capture any genetic effects on the phenotype. In this paper, we used the Gaussian kernel for imaging and epigenetics data integration, while the IBS kernel for genetic data analysis.

4. Relevant methods

Li and Cui (2012) have proposed a linear PCA (LPCA) based method for detecting the interaction effect between two genes, which is possible to extend to three datasets. Let $\mathbf{M}^{(1)} = [M_1^1, M_2^1, \dots, M_s^1]$, $\mathbf{M}^{(2)} = [M_1^2, M_2^2, \dots, M_r^2]$, and $\mathbf{M}^{(3)} = [M_1^3, M_2^3, \dots, M_d^3]$ be the data matrix for the genetics, imaging and epigenetics, respectively. Using the PCA we can compute the first ℓ principle components: $U_1^1, U_2^1, \dots, U_{s_\ell}^1, U_1^2, U_2^2, \dots, U_{r_\ell}^2$, and $U_1^3, U_2^3, \dots, U_{d_\ell}^3$ with $s_\ell \leq s$, $r_\ell \leq r$, and $d_\ell \leq d$, for the corresponding data matrix, respectively. An interaction method using a partial least squares was developed by Wang et al. (2009) for the binary disease traits. They compared their method with a regression-based principle component analysis. Specifically, assuming an additive model for each marker, the singular value decomposition (SVD) can be applied to both gene matrices (Wang et al., 2009). We then compared the numerical, simulation and real data analysis with the following methods: test based on only first and first few principal components followed by multiple regression, i.e., partial principal component analysis regression (pPCAR), and full principal component analysis regression (fPCAR), respectively.

4.1. Principal component multiple regression

By considering only the first principal component, the 3rd order interaction model (i.e., pPCA) can be stated as:

$$\begin{aligned} \mathbf{y} = \mathbf{X}\beta + \sum_{a=1}^s \alpha_a \mathbf{M}_a^{(1)} + \sum_{b=1}^r \alpha_b \mathbf{M}_b^{(2)} + \sum_{c=1}^d \alpha_c \mathbf{M}_c^{(3)} + \eta_1 U_1^{(1)} U_1^{(2)} + \eta_2 U_1^{(1)} U_1^{(3)} + \eta_3 U_1^{(2)} U_1^{(3)} \\ + \eta_4 U_1^{(1)} U_1^{(2)} U_1^{(3)}. \end{aligned}$$

(17)

This model is called partial PCA regression (pPCAR). Using the multiple regression in Eq. (17), the interaction of $\mathbf{M}^{(1)} \times \mathbf{M}^{(2)} \times \mathbf{M}^{(3)}$ is assessed by testing $H_0: \eta_4 = 0$. To consider all

possible interactions of the selected principal components, we can also replace the main effects by the first ℓ principal components. The number of principal components ℓ is selected based on the proportion of variation explained by the principal components, which can explain the major variations (say, 85%). The model in Eq. (17) then becomes

$$\begin{aligned} \mathbf{y} = \mathbf{X}\beta + \sum_{a=1}^{s_\ell} \alpha_a U_a^{(1)} + \sum_{b=1}^{r_\ell} \alpha_b U_b^{(2)} + \sum_{c=1}^{d_\ell} \alpha_c U_c^{(3)} + \sum_{a=1}^{s_\ell} \sum_{b=1}^{r_\ell} \eta_{12} U_a^{(1)} U_b^{(2)} + \sum_{a=1}^{s_\ell} \sum_{c=1}^{d_\ell} \eta_{13} U_a^{(1)} U_c^{(3)} \\ + \sum_{b=1}^{r_\ell} \sum_{c=1}^{d_\ell} \eta_{23} U_b^{(2)} U_c^{(3)} \\ + \sum_{a=1}^{s_\ell} \sum_{b=1}^{r_\ell} \sum_{c=1}^{d_\ell} \eta_{123} U_a^{(1)} U_b^{(2)} U_c^{(3)}, \end{aligned} \quad (18)$$

Using the multiple regression in Eq. (18), the interaction of $\mathbf{M}^{(1)} \times \mathbf{M}^{(2)} \times \mathbf{M}^{(3)}$ is assessed by testing $H_0 : \eta_{123} = 0$.

4.2. Principal component sequence kernel association test

Over the past several years, the sequence kernel association test (SKAT) approach has been widely used in GWAS due to its flexibility and computational efficiency. The SKAT is based on a SNP-set (e.g., a gene or a region) level test for the association between a set of variants and dichotomous or quantitative phenotypes. This method aggregates individual test statistics of SNPs and efficiently computes SNP-set level p-values, while adjusting for covariates such as principal components to account for population stratification (Wu et al., 2011; Ionita-Laza et al., 2013). We applied SKAT to gene-derived SNPs, ROIs, and gene-specific DNA methylations data. To do this, we use SKAT in Eq. (18) and the interaction of $\mathbf{M}^{(1)} \times \mathbf{M}^{(2)} \times \mathbf{M}^{(3)}$ is assessed by testing $H_0 : \eta_{123} = 0$.

5. Experiments

We conducted experiments on both the simulation and real imaging (epi)-genetics data from the SZ study. We considered the IBS kernel for the genetic data and the Gaussian kernel for all other data. For the Gaussian kernel, we used the median of the pairwise distance as the bandwidth (Gretton et al., 2008). The solution of the model is based on the ReML algorithm (Fisher's scoring algorithm). The ReML algorithm converged in less than 50 iterations (the difference between successive log ReML values was smaller than 10^{-04}), and in most cases it converged quickly with 10 iterations, taking only a few seconds with an R-program. The iteration of the ReML may be trapped in a local minima. To avoid this problem, we use a set of initial points (0, 0.00001, 0.0001, 0.001, 0.01, 0.1, 1) for the optimization algorithm and choose the best one (maximized ReML).

5.1. Simulation studies

The goal of these simulation studies is to evaluate the performance of the proposed method and the accuracy of the score tests. To synthesize quantitative phenotypes, we applied the following model:

$$y_i = \mathbf{X}_i^T \boldsymbol{\beta} + \alpha_1 [h_S(S_i) + h_T(T_i) + h_C(C_i)] + \alpha_2 [h_{S \times T}(S_i, T_i) + h_{S \times C}(S_i, C_i) + h_{T \times C}(T_i, C_i)] + \alpha_3 [h_{S \times T \times C}(S_i, T_i, C_i)] + \sigma \epsilon_i \quad (19)$$

where $\mathbf{X}_i$ is a vector of covariates including an intercept (e.g., age, height, weight) of i th subject ($i = 1, 2, \dots, n$) and $\boldsymbol{\beta}$ s are the coefficient, ϵ_i is a random error that follows the Gaussian distribution with mean zero and unit variance, i.e., $\epsilon_i \sim \mathcal{N}(0, 1)$, and σ is the standard deviation of the error and is fixed to 10^{-02} . Here S_i , T_i and C_i represent three different data sets. For each function, we designed the following form

$$h_S(S_i) = \sum_{a=1}^{10} S[i, a] \cos(S[i, a]), \quad h_T(T_i) = \sum_{b=1}^2 2T[i, b] \sin(T[i, b]), \quad h_C(C_i) = \sum_{c=1}^{10} \frac{i}{\sqrt{2 \exp(C[i, c])}},$$

$$h_{S \times T}(S_i, T_i) = h_S(S_i) 2h_T(T_i), \quad h_{S \times C}(S_i, C_i) = h_S(S_i) 3h_C(C_i), \quad h_{T \times C}(T_i, C_i) = 2h_T(T_i) 3h_C(C_i),$$

$$h_{S \times T \times C}(S_i, T_i, C_i) = h_S(S_i) 2h_T(T_i) 3h_C(C_i).$$

In simulation-I and simulation-II, we generated data under different values of $(\alpha_1, \alpha_2, \alpha_3)$ to evaluate the performance of the test. In other words, for $\alpha_1 = \alpha_2 = \alpha_3 = 0$ both main effects and all interaction effects vanish and we examined the false positive rate of the score test for the overall effect. For $\alpha_1 \neq 0, \alpha_2 = 0 (\alpha_2 \neq 0)$ and $\alpha_3 = 0$, there are main effects (2nd order interaction effects) but no higher order interaction effects, hence we can evaluate the power of the score test. We also set $(\alpha_1, \alpha_2, \alpha_3)$ to a variety of values to test the power of both score tests. In each setting, 5000 simulations were performed for the consistence of the results.

5.1.1. Simulation-I (numerical data)—In this simulation, we generated two covariates (height and weight) and three views (genetics, topological, and categorical data). We generated the height and weight by the regular sequencing of the interval (50,80) and (60,225) with an increment of 2.05 and 4.7 for the $n = 500$ subjects, respectively. Then, we added the noise $3\mathcal{N}(0,1)$ to each of the variables. The element of coefficient vector $\boldsymbol{\beta}$ was fixed to 0.5. For the genetic data, we simulated a gene with 10 SNPs for 500 subjects using the latent model as in (Parkhomenko et al., 2009; Alam et al., 2016b). We generated data using three circles of different radii with small noise for topological features (Alam and Fukumizu, 2014):

$$T_i = r_i \begin{pmatrix} \cos(\mathbf{R}_i) \\ \sin(\mathbf{R}_i) \end{pmatrix} + \epsilon_i,$$

where $r_i = 1, 0.5$ and 0.25 , for $i = 1, 2, \dots, n_1$, $i = n_1 + 1, \dots, n_2$, and $i = n_2 + 1, \dots, n_3$ ($n = n_1 + n_2 + n_3 = 500$), respectively, $\mathbf{R}_i \sim U[-1, 1]$ and $\epsilon_i \sim \mathcal{N}(0, I_2)$ independently. For the categorical data, we considered 10 categories with probability $1/10$ and converted these features into the dummy features with levels of zero and one.

In addition, in plotting the receiver operating characteristic (ROC), the data were generated by fixing $\alpha_1 = 1$, $\alpha_2 = 1$ and α_3 was allocated with probability 0.5 for each run, a random number was either uniformly distributed on $[0, 1]$ or at 0. We also fixed $\alpha_1 = 1$ and for each run α_3 ($\alpha_2 = \alpha_3$) was allocated with probability 0.5; a random number is uniformly distributed on either $[0, 1]$ or at 0. We considered three sample sizes $n \in \{100, 500, 1000\}$ and compared the ROC curves of the proposed method with the three state-of-the-art methods in identifying the interaction effects.

Results (simulation-I): Table 1 presents the simulation-I results for testing the overall effects and higher order interactions (HOI). The nominal p -value threshold was fixed to 0.05. By observing this table, we can see that when $\alpha_1 = \alpha_2 = \alpha_3 = 0$, the size of the overall score test is close to the nominal p -value threshold. When $\alpha_1 = 0$, $\alpha_2 = 0$ (or $(\alpha_2 = 0)$) and $\alpha_3 = 0$, the false positive rate of the test for higher order interaction effects is also controlled. For the power analysis ($\alpha_3 = 0$) we found that the power of the interaction test by the proposed method quickly exceeds 0.85 and 0.90. While the SKAT method has higher power when compared to other methods (pPCAR and fPCA), it has lower power than the proposed method in this simulation. We observed that dimension reduction methods (pPCAR and fPCA) can significantly inflate the false positive rates and dramatically lose power when compared with the proposed method and SKAT.

Figure 2 shows the ROC of the proposed method and three alternative methods to detect interactions using the simulation-III with three sample sizes, $n \in \{100, 500, 1000\}$ in the following cases: (a) third parameter is random; (b) second and third parameter values are random. The sensitivity is plotted against (1- specificity) for the p -value in the range of 0 – 1 with a step size 0.0001. The power gain of the proposed method relative to the alternative ones is evident in all situations. When the sample size was increased, and the coefficient for the second order interaction was one, a higher power was observed. We also observed extremely high power for the higher order interactions.

5.1.2. Simulation-II (Mind Clinical Imaging Consortium's schizophrenia data)

—To validate Eq. (19) under different values of $(\alpha_1, \alpha_2, \alpha_3)$, we consider hybrid data based on real experiments. This simulation was based on the SZ data collected by the MCIC (Chen et al., 2012; Liu et al., 2014; Chekouo et al., 2016). These are 208 subjects including 92 schizophrenic patients (age: 34 ± 11 , 22 females) and 116 (age: 32 ± 11 , 44 females) healthy controls. All participants' symptoms were evaluated by the scale for the assessment of positive symptoms and negative symptoms (Andreasen, 1984). By filtering missing data, the

number of subjects was reduced to 182 subjects (79 SZ patients and 103 healthy controls). We considered the age, height, and weight as the covariates and used gene-derived SNPs, ROIs with voxels, and gene-specific DNA methylation information.

Genetics: For each subject (SZ patients and healthy controls) a blood sample was taken and DNA was extracted. Genotyping was performed for all subjects at the Mind Research Network using the Illumina Infinium HumanOmni1-Quad assay covering 1140419 SNP loci. To form the final genotype calls and to perform a series of standard quality control procedures, the bead studio and PLINK software packages were applied, respectively. The final dataset spans 722177 loci with 22442 genes of 182 subjects. Genotypes “AA” (non-minor allele), “Aa” (one minor allele) and “aa” (two minor alleles) were coded as 0, 1 and 2 for each SNP, respectively (Alam et al., 2016b). The top 75 genes for the SZ are listed in the SZ genes database (<https://bioinfo.uth.edu/SZGR/>).

Imaging: Participants’ fMRI data were collected during a block design motor response for auditory stimulation. State-of-the-art approaches using participant feedback and expert observation were used. The aim was to continuously monitor the patients while acquiring images with the parameters (TR=2000 ms, TE= 30ms, field of view=22cm, slice thickness=4mm, 1 mm skip, 27 slices, acquisition matrix 64×64, flip angle=90°) on a Siemens 3T Trio Scanner and 1.5 T Sonata. The data come from four different sites (& scanners) with echo-planar imaging (EPI). Data were pre-processed with SPM software and were realigned spatially, normalized and resliced to 3×3×3 mm. They were smoothed with a 10×10×10 mm³ Gaussian kernel and then analyzed by multiple regression that considered the stimulus and their temporal derivatives plus an intercept term as a regressors. Finally the stimulus-on versus stimulus-off contrast images were extracted. Next, 41236 voxels were extracted from 116 ROIs based on the AAL brain atlas for analysis (Alam et al., 2016a). For imaging features (ROIs), we considered 116 ROIs. The name for the ROIs is given by the automated anatomical labeling (AAL) template (Yan and Zang, 2010).

Epigenetics: DNA methylation is one of the main epigenetic mechanisms to regulate gene expression, and may be involved in the development of SZ. For this paper, we investigated 27481 DNA methylation markers in blood from SZ patients and healthy controls. DNA from blood samples were measured by the Illumina Infinium Methylation27 Assay. The methylation value is calculated by taking the ratio of the methylated probe and the total probe intensity.

We treated each gene-derived SNPs, region of interest (ROI) and gene-specific DNA methylations as a single testing unit. Each gene-derived SNP, ROI and gene-specific DNA methylation consists of genetic features (SNPs), imaging features (voxels) and epigenetics features (methylations). Gene-derived SNPs, ROI, or gene-specific DNA methylation based association analysis would provide more biologically interpretable results than the single SNP, voxel or methylation based analysis. In addition, this analysis is statistically appealing. In this paper, the top 72 genes (from <https://bioinfo.uth.edu/SZGR/>, where genes have more than one SNPs), 116 ROIs, and 129 gene-specific DNA methylations (> 5 methylations) are used, respectively.

Results (simulation-II): Table 2 presents the simulation-II results for testing the overall effects and higher order interactions with the same nominal p -value threshold. By this table, we can also see that when $\alpha_1 = \alpha_2 = \alpha_3 = 0$, the size of the overall score test is close to the nominal p -value threshold. When $\alpha_1 = 0$, $\alpha_2 = 0$ (or $(\alpha_2 = 0)$) and $\alpha_3 = 0$, the false positive rate of the test for higher order interaction effects is also controlled. For the power analysis ($\alpha_3 = 0$) we found that the power of the interaction test for the proposed method also quickly exceeds 0.90. Similar to the results of simulation-I, while the SKAT method has higher power than other methods (e.g., pPCAR and fPCA), it has lower power than the proposed method. But the dimension reduction methods can significantly inflate the false positive rates and dramatically lose power when compared with the proposed method and SKAT. For the proposed method, we get statistically significant results in 9 out of 10 scenarios, except for one situation $(\alpha_1, \alpha_2, \alpha_3) = (0, 0.5, 0)$. This setting is dependent on the sample size. For large sample size, the empirical results tend to approximate the theoretical ones.

5.2. Application to imaging genetics and epigenetics with schizophrenia

Here we demonstrate the power of our proposed kernel method on real imaging (epi)-genomics study of SZ from MCIC. The goal here is to characterize the underlying interactions between genetic makers (gene-derived SNPs), human brain regions (ROIs) and epigenetic factors (gene-specific methylations) along with the covariates (age, height, weight) for their association with hippocampal volume derived from structural MRI scans.

By considering 72 genes-derived SNPs, 116 ROIs and 129 gene-specific DNA methylations, we have $1077408(72 \times 116 \times 129)$ triplets. By the overall tests, we obtained 15436 significant triplets ($p = 0.05$). Figure 3 visualizes the plot of $-\log_{10}(p)$ for 15436 triplets. The vertical solid, dotted and double dotted lines correspond to the p -values of 0.05, 0.01, 0.001, respectively. Based on these lines, we observed that 272, 72, and 13 triplets turn out to have significantly higher order interactions at 0.05, 0.01 and 0.001 levels, respectively.

Table 3 presents the ReML estimates of σ^2 , $\tau^{(1)}$, $\tau^{(2)}$, $\tau^{(3)}$, $\tau^{(1 \times 2)}$, $\tau^{(1 \times 3)}$, $\tau^{(2 \times 3)}$, $\tau^{(1 \times 2 \times 3)}$ and the p -values for both the proposed and SKAT methods for each of the 13 triplets, which are identified to have significant interactions at a level of 0.001. At this p -value, we have 6 gene-derived SNPs (**IL1B**, **MAGI2**, **NRG1**, **PDLIM5**, **SLC18A1**, **TDRD3**), 10 ROIs (**CRBL8.L**, **CRBLCrus1.L**, **ORBSUP.R**, **LING.L**, **CAU.R**, **IPL.L**, **IPL.R**, **PoCG.L**, **ITG.R**, **VER54**), and 6 gene-specific DNA methylations (**CRABP1**, **FBXO28**, **DUSP1**, **FHIT**, **PLAGL1**, **TFPI2**) that have significant effects on the hippocampal volume of SZ brains.

Figure 4 shows the interaction networks within and between genetics, imaging and epigenetics factors. Each node represents the gene-derived SNP, ROI and gene-specific DNA methylation, respectively and the interactions between nodes are denoted with lines. The thickness of the line indicates the strength of the interaction. Those selected SNPs and methylations show the interactions between several other genes. The selected ROIs also show their interactions (shown in Figure 4) as well as with other ROIs (not shown in the figure). Consistent with many studies in the literature, the identified gene-derived SNPs are susceptible to SZ disease. For example, **IL1B** gene has been implicated in the pathophysiology of SZ (Siawa et al., 2016); **MAGI2** gene is associated with increased risk for cognitive impairment in SZ (Koide et al., 2013); **NRG1** is one of the leading SZ

susceptibility gene (Harrison and Law, 2006); **PDLIM5** variants have been highly linked to SZ (Moselhy et al., 2015); **SLC18A1** has been shown to be abundantly expressed in the human adrenal medulla (Bly, 2005); and **TDRD3** associated with the mediator protein tudor domain-containing protein 3 participates in two gene expression processes of transcription and translation (Shibuya et al., 2013).

Recent research has also shown that the 10 ROIs selected by the proposed method have a critical role in brain diseases (Suk et al., 2016; Chen et al., 2013; Wu et al., 2013). In addition, we conducted a detailed study on the connectivity network based on these 10 ROIs. To do this, we evaluated the differences of networks between the SZ and healthy controls. Table 4 presents the transitivity, degree and global efficiency (the global efficiency is defined as the shortest path length between two nodes) of each ROI for the SZ and healthy control, respectively. From this table, we observed that the transitivity (measuring the probability that the adjacent vertices of a vertex are connected) of the SZ group is larger than that of the healthy control. This suggests that the connectivity network of SZ tends to have more transitive triples. The degree (the number of edges incident to the vertex) of the SZ group is also larger than in the healthy control group for all of the ROIs. Finally, the global efficiency of the network for the SZ group is also different from the healthy control, indicating their different functional activities in these regions.

To confirm the differences in the selected ROIs between the SZ candidate and healthy control group, we evaluated correlation matrices and networks analysis separately. Figure 5 is the visualization of correlation matrices and axial view of all networks for the SZ candidate and healthy control group, respectively. From Figure 5, it can be observed that the ROIs in the SZ candidate groups are more correlated and connected than in the healthy controls. Figure 7 shows axial view of the selected 10 ROIs (see supplementary material).

Table 5 lists the selected significant gene-derived SNPs, ROIs and gene-specific methylations using the proposed method (KMDHOI) and SKAT at $p = 0.01$. We found that 31 genes-derived SNP, 35 ROIs and 20 genes-specific methylations from 72 triplets were identified to be significantly associated with the hippocampal volume. We also observed that 6 gene-derived SNPs, 10 ROIs and 6 gene-specific methylations were significant at $p = 0.001$. The underlined elements in Table 5 indicated that they have significant interactions. Table 5 & 6 (in the supplementary materials) list 72 triplets, which were significant at $p = 0.001$.

Although the interactions do not appear to be significant, some of them show promising and consistent results. According to the p -values, we can determine the gene-derived SNPs, ROIs, and gene-specific methylations that are associated with a highly significant hippocampal volume. We observed that gene-derived SNP (**MAGI2**, **NRG1**, **SLC18A1**, **TDRD3**), ROIs (**CRBL8.L**, **CRBLCrus1.L**, **ORBSUP.R**, **LING.L**, **IPL.L**, **IPL.R**) and gene-specific DNA methylations (**CRABP1**, **FBXO28**, **FHIT**, **PLAGL1**) at $p = 0.0001$, gene-derived SNPs (**MAGI2**, **NRG1**, **TDRD3**), ROIs (**CRBL8.L**, **CRBLCrus1.L**, **IPL.L**, **IPL.R**) and gene-specific DNA methylations (**FBXO28**, **PLAGL1**) at $p = 0.00001$, and gene-derived SNPs (**MAGI2**), ROIs (**CRBLCrus1.L**), and gene-specific DNA methylation

(**FBXO28**) at $p = 0.000001$, are identified to have strong associations with hippocampal volumes of SZ brains.

To confirm this discovery, we used DAVID, and gene ontology (GO) enrichment analysis to find the most relevant GO terms associated with the selected 31 genes. The selected genes are associated with a set of annotation terms. We compared 5 annotation categories, including literature, disease, gene ontology, pathways and protein interactions using DAVID (Huang et al., 2009). Table 7 (in the supplementary materials) presents five annotation categories of the 31 selected genes. From this table, we observed that the selected genes have had remarkable literature review in the past studies. According to the disease annotation, the selected genes are highly associated with complex diseases including SZ, cognitive function, bipolar disorder, and others. By GO annotation, the selected genes have significant relationship to single-organism processes, response to stimuli, developmental processes and etc. From the table, we also observed that the selected genes have a significant pathway to facilitate biological interpretation in a network context. Moreover, protein interaction annotations show that the selected genes have been discussed in many biomedical papers (Sanders et al., 2008; Gerhard et al., 2004; Strausberg et al., 2002).

Genes do not function alone; rather, they interact with each other. When genes share a similar set of GO annotation terms, they are most likely to be involved with similar biological mechanisms. To verify this, we extracted the (gene-derived SNPs)-(gene-specific DNA methylations) networks using STRING (Szklarczyk et al., 2007). STRING imports protein association knowledge from databases of both physical interactions and curated biological pathways. In STRING, the simple interaction unit is the functional relationship between two proteins/genes that can contribute to a common biological purpose. Figure 8 (in the supplementary material) shows the network (e.g., gene-derived SNPs and gene-specific methylations) based on the protein interactions between the combined 31 gene-derived SNPs and 20 gene-specific DNA methylations (see supplementary file). In this figure, the color saturation of the edges represents the confidence score of a functional association. Further network analysis shows that the number of nodes, number of edges, average node degree, clustering coefficient, PPI enrichment p -values are 51, 93, 300, 11.8, 0.603 for $p = 0 \times 10^{-16}$, respectively (Szklarczyk et al., 2007). This network of genes has significantly more interactions than expected, which indicates that they may function in a concerted effort.

In addition, the three datasets were used to classify the SZ patients from the healthy controls via the proposed KMDHOI, multiple canonical correlation analysis (Multiple CCA) and multiple kernel canonical correlation analysis (Multiple KCCA) followed by the two classifiers: the k-nearest neighbors (KNN) and linear support vector machine (SVM). For the proposed approach, we considered triplets that have significant (at $p < 0.01$) effects on the hippocampal volume of SZ brains: 6 gene-derived SNPs (**IL1B**, **MAGI2**, **NRG1**, **PDLIM5**, **SLC18A1**, **TDRD3**), 10 ROIs (**CRBL8.L**, **CRBLCrus1.L**, **ORBSUP.R**, **LING.L**, **CAU.R**, **IPL.L**, **IPL.R**, **PoCG.L**, **ITG.R**, **VER54**), and 6 gene-specific DNA methylations (**CRABP1**, **FBXO28**, **DUSP1**, **FHIT**, **PLAGL1**, **TFPI2**). Both multiple CCA and multiple KCCA serve as a feature extraction tool, based on which the classifier is used to separate patients from healthy controls. Table 6 presents the classification error using the cross-validation (2-fold and 5-fold). From these results, it is evident that the KMDHOI based

classification is significantly more accurate than using multiple CCA and Multiple KCCA, demonstrating that KMDHOI is a better tool for feature selection.

Lastly, we conducted standard logistic regression analysis with the covariates of age, gender, and BMI on the outcome of SZ disease (SZ vs healthy control). We found that BMI is a significant covariate for the SZ vs healthy control at $p = 0.0353$. Thus, BMI is one of the risk factors of SZ disease. For a BMI ≥ 25 , we considered the subject to be at a high risk. Based on this risk, we divided the estimated higher order interaction effect $\hat{\mathbf{h}}_{\mathbf{M}^{(1)} \times \mathbf{M}^{(2)} \times \mathbf{M}^{(3)}}$

values into four regimes: SZ with high BMI risk, SZ with low BMI risk, healthy control with high BMI risk, and healthy control with low BMI risk. Figure 6 shows the box-plots of the estimated interaction effects within each of the four regimes for the most significant triplet (**MAGI2**, **CRBLCrus1.L**, and **FBXO28**). The small variation indicates a higher risk of the interaction effect on hippocampal volume. This figure shows that higher interactions are associated with higher SZ and BMI risk and vice versa.

6. Discussion and future research

The identification of multi-omics interactions is becoming a common challenge to multidimensional imaging and genetics data analysis. Fundamental works in kernel machine method have been boldly pursued by Liu et al. (2007), where a single modal dataset was used to test for a genetic pathway effect. Li and Cui (2012) have proposed a kernel machine based method for gene-gene interaction. They treated each gene as a testing unit for gene-gene interactions. A kernel machine method was also proposed by Ge et al. (2015) for detecting multiple factor interactions, where a smoothing spline-ANOVA decomposition method was adopted. However, these approaches only use single or pairwise datasets. The extension of such methods to three or multiple data sets is not straightforward, and poses a significant challenge.

In this paper, we have proposed a semiparametric kernel based method for higher order interactions (i.e., KMDHOI) among multimodal datasets. This is a generalized approach of Liu et al. (2007); Li and Cui (2012); Ge et al. (2015) and can apply to multimodal datasets. Compared with the PCA based multiple regression and SKAT methods, the proposed KMDHOI shows a more flexible and biologically plausible approach to model higher order epistasis factors among genetic, imaging, and epigenetic data. While kernel based methods provide more powerful and reproducible outputs, the interpretation of the results remains challenging. Incorporating biological knowledge information (e.g., GO) can provide additional evidences on the results. The performance of the KMDHOI method was evaluated on both simulated and real MCIC data. The extensive simulation studies show the power gain of the proposed method relative to the alternative methods such as dimension reduction based multiple regression and SKAT methods.

The utility of the proposed KMDHOI method is further demonstrated with the application to imaging (epi)-genetics study of SZ. According to the p -values, the proposed method is able to rank the triplets (gene-derived SNPs)-ROI-(gene-specific DNA methylations) and a subset of triplets are identified to be highly related to SZ disease. At $p = 0.01$ the proposed KMDHOI method extracts the unique 31 gene-derived SNPs, 35 ROIs and 20 gene-specific

DNA methylations from 72 triplets, which are considered to have significant impact on hippocampal volume of SZ patients. By conducting gene ontology, pathway analysis, and network analysis including visualization, we find evidences that the selected gene-derived SNPs, region of interest (ROI) and gene-specific DNA methylations have a significant influence on the manifestation of SZ disease and can serve as a distinct feature for the classification of SZ patients from the healthy controls.

It must be emphasized that choosing a suitable kernel is indispensable. Although the linear kernel does not have any free parameters, the linear kernel has some limitations. Using the polynomial kernel makes it possible to detect higher order correlations. In addition, both linear and polynomial kernels are unbounded. The Gaussian kernel has a free parameter (bandwidth) and a number of properties (e.g., boundedness). In this study, we used the median of the pairwise distance as the bandwidth for the Gaussian kernel, which appears to be practical. While the proposed KMDHOI method was applied here for the study of three data sets (e.g., SNPs, imaging voxels, and methylations), they can also be applied to many multi-omics data from other diseases, which are ubiquitous in biomedical research.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors wish to thank the department of energy (DE-FG02-99ER62764), the National Institutes of Health (1RC1MH089257, P20GM103472, R01GM109068, R01MH104680, R01MH107354) and the National Science foundation EPSCoR program (1539067) for the support.

References

- Aberg KA, McClay JL, Nerella S, et al., S. C., 2014 Methylome-wide association study of schizophrenia identifying blood biomarker signatures of environmental insults. *JAMA Psychiatry* 71(3), 255–264. [PubMed: 24402055]
- Alam MA, 2014 Kernel Choice for Unsupervised Kernel Methods. PhD. Dissertation, The Graduate University for Advanced Studies, Japan.
- Alam MA, Calhoun V, Wang YP, 2016a Influence function of multiple kernel canonical analysis to identify outliers in imaging genetics data. *Proceedings of 7th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM BCB)*, Seattle, WA, USA, 210–2198.
- Alam MA, Fukumizu K, 2014 Hyperparameter selection in kernel principal component analysis. *Journal of Computer Science* 10(7), 1139–1150.
- Alam MA, Fukumizu K, 2015 Higher-order regularized kernel canonical correlation analysis. *International Journal of Pattern Recognition and Artificial Intelligence* 29(4), 1551005(1–24).
- Alam MA, Komori O, Calhoun V, Wang YP, 2016b Robust kernel canonical correlation analysis to detect gene-gene interaction for imaging genetics data. *Proceedings of 7th ACM Conference on Bioinformatics, Computational Biology, and Health Informatics (ACM BCB)*, Seattle, WA, USA, 279–288.
- Andreasen NC, 1984 Scale for the assessment of positive symptoms (SAPS). Springer, Iowa City, University of Iowa.
- Aronszajn N, 1950 Theory of reproducing kernels. *Transactions of the American Mathematical Society* 68, 337–404.
- Bis JC, DeCarli C, et al., A. S., 2012 Common variants at 12q14 and 12q24 are associated with hippocampal volume. *Nature Genetics* 44(5), 545–551. [PubMed: 22504421]

- Bly M, 2005 Mutation in the vesicular monoamine gene, *slc18a1*, associated with schizophrenia. *Schizophrenia Research* 78, 337–338. [PubMed: 15961286]
- Calhoun VD, Sui J, 2016 Multimodal fusion of brain imaging data: A key to finding the missing link(s) in complex mental illness. *Biol Psychiatry Cogn Neurosci Neuroimaging* 1, 230–244. [PubMed: 27347565]
- Camps-Valls G, Rojo-Alvarez JL, Martinez-Romon M, 2007 *Kernel Methods in Bioengineering, Signal and Image*. Idea Group publishing, London.
- Chang B, Kruger U, Kustra R, Zhang J, 2013 Canonical correlation analysis based on hilbert-schmidt independence criterion and centered kernel target alignment. *Proceedings of the 30th International Conference on Machine Learning*, Atlanta, Georgia, USA.
- Chekouo T, Stingo FC, Guindani M, Do KA, 2016 A bayesian predictive model for imaging genetics with application to schizophrenia. *The Annals of Applied Statistics* 10(3), 1547–1571.
- Chen J, Calhoun VD, Pearlson GD, Ehrlich S, Turner JA, Ho BC, Wassink TH, Michale A, Liu J, 2012 Multifaceted genomic risk for brain function in schizophrenia. *NeuroImage* 61, 866–875. [PubMed: 22440650]
- Chen Z, Liu M, Gross DW, Beaulieu C, 2013 Graph theoretical analysis of developmental patterns of the white matter network. *Frontiers in Human Neuroscience* 7, 199–211. [PubMed: 23734115]
- Ge T, Nichols TE, Ghosh D, Mormino EC, Smoller JW, a. M. R. S., the Alzheimer's Disease Neuroimaging Initiative, 2015 A kernel machine method for detecting effects of interaction between multidimensional variable sets: An imaging genetics application. *NeuroImage* 109, 505–514. [PubMed: 25600633]
- Gerhard DS, Wagner L, Feingold EA, et al., 2004 The status, quality, and expansion of the nih full-length cDNA project: the mammalian gene collection (MGC). *The American Journal of Psychiatry* 14(10B), 2121–7.
- Gollub RL, Shoemaker JM, King MD, White T, Ehrlich S, Sponheim SR, Clark VP, Turner JA, Mueller BA, Magnotta V, O'Leary D, Ho BC, Brauns S, Manoach DS, Seidman L, Bustillo JR, Lauriello J, Bockholt J, Lim KO, Rosen BR, Schulz SC, Calhoun VD, Andreasen NC, 2013 The mcic collection: a shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia. *Front Genet* 11, 367–38.
- Gretton A, Fukumizu K, Teo CH, Song L, Schölkopf B, Smola A, 2008 A kernel statistical test of independence. In *Advances in Neural Information Processing Systems* 20, 585–592.
- Harrison PJ, Law AJ, 2016 Neuregulin 1 and schizophrenia: Genetics, gene expression, and neurobiology. *BIOL PSYCHIATRY* 60, 132–140.
- Hieke S, Binder H, Nieters A, Schumacher M, 2014 Convergence analysis of kernel canonical correlation analysis: theory and practice. *Computational Statistics* 29(1–2), 51–63.
- Hofmann T, Schölkopf B, Smola JA, 2008 Kernel methods in machine learning. *The Annals of Statistics* 36, 1171–1220.
- Huang D, Sherman BR, Lempicki RA, 2009 Systematic and integrative analysis of large gene lists using david bioinformatics resources. *Nature Protocols* 4(1), 44–57. [PubMed: 19131956]
- Ionita-Laza I, Lee S, Makarov V, Buxbaum JD, Lin X, 2013 Sequence kernel association tests for the combined effect of rare and common variants. *American Journal of Human Genetics* 92, 841–853. [PubMed: 23684009]
- Jahanshad N, Hibar DP, Ryles A, Toga AW, McMahon KL, de Zubicaray GI, Hansell NK, Montgomery GW, N. G. M., Wright MJ, Thompson PM, 2012 Discovery of genes that affect human brain connectivity: A genome-wide analysis of the connectome. In *Proceeding IEEE Int Symp Biomed Imaging*, 542–545.
- Jahanshad N, Rajagopalan P, Hua X, Hibar DP, Nir TM, Toga AW, Jack CRJ, Saykin AJ, Green RC, Weiner MW and, S. E. M., Montgomery GW, Hansell NK, McMahon KL, Zubicaray GI, Martin NG, Wright MJ, Thompson PM, Initiative, A. D. N., 2013 Genome-wide scan of healthy human connectome discovers *SPON1* gene variant influencing dementia severity. In *Proceedings of the National Academy of Sciences* 110(12), 4768–73.
- Kimeldorf G, Wahba G, 1971 Some results on tchebycheffian spline functions. *Journal of mathematical analysis and applications* 33(1), 82–95.

- Kircher T, Renate T, 2005 Functional brain imaging of symptoms and cognition in schizophrenia. *Progress in Brain Research* 150, 299308.
- Koide T, Banno M, Aleksic B, et al., 2013 Common variants in magi2 gene are associated with increased risk for cognitive impairment in schizophrenic patients. *PLoS ONE* 7(9), e36836.
- Kung SY, 2014 *Kernel Methods and Machine Learning*. Cambridge University Press, New York.
- Kwee LC, Liu D, X. L. D. G. M. P. E., 2008 A powerful and flexible multilocus association test for quantitative traits. *Annals of Human Genetics* 82(2), 386–397.
- Lencz T, Morgan TV, Athanasiou M, Dain B, Reed CR, Kane JM, Kucher-Lapati R, Malhotra AK, 2007 Converging evidence for a pseudoautosomal cytokine receptor gene locus in schizophrenia. *Molecular Psychiatry* 12, 572–580. [PubMed: 17522711]
- Li J, Huang D, Guo M, Liu X, Wang C, Teng Z, Zhang R, Jiang Y, Lv H, Wang L, 2015 A gene-based information gain method for detecting genegene interactions in casecontrol studies. *European Journal of Human Genetics* 23, 1566–1572. [PubMed: 25758991]
- Li S, Cui Y, 2012 Gene-centric gene-gene interaction: a model-based kernel machine method. *The Annals of Applied Statistics* 6(3), 1134–1161.
- Lin D, Callhoun VD, Wang YP, 2014 Correspondence between fmri and snp data by group sparse canonical correlation analysis. *Medical Image Analysis* 18, 891–902. [PubMed: 24247004]
- Liu D, Lin X, Ghosh D, 2007 Semiparametric regression of multidimensional genetics pathway data: least squares kernel machines and linear mixed model,. *Biometrics* 63(4), 1079–1088.
- Liu J, Chen J, Ehrlich S, Walton E, White T, N. P. B., Bustillo J, Turner JA, Calhoun VD, 2014 Methylation patterns in whole blood correlate with symptoms in schizophrenia patients. *Schizophrenia Bulletin* 40(4), 769–776. [PubMed: 23734059]
- Liu M, Min R, Gao Y, D. Z., Shen D, 2016 Multitemplate-based multiview learning for alzheimers disease diagnosis machine learning and medical imaging. *Machine Learning and Medical Imaging*, 259–297.
- Montano C, Tauband MA, Jaffe A, Briem E, et al., 2016 Association of dna methylation differences with schizophrenia in an epigenome-wide association study. *JAMA Psychiatry* 73(5), 506–514. [PubMed: 27074206]
- Moselhy H, Eapenb V, Akawi NA, Younis A, et al., 2015 Secondary association of pdlim5 with paranoid schizophrenia in emirati patients. *Meta Gene* 5, 135–139. [PubMed: 26925374]
- Parkhomenko E, Tritchler D, Beyene J, 2009 Sparse canonical correlation analysis with application to genomic data integration. *Statistical Applications in Genetics and Molecular Biolog* 8(1), 1–34.
- Pearlson GD, Liu J, Calhoun VD, 2015 An introductory review of parallel independent component analysis (p-ica) and a guide to applying p-ica to genetic data and imaging phenotypes to identify disease-associated biological pathways and systems in common complex disorders. *Front Genet* 6, 276. [PubMed: 26442095]
- Peng QN, Zhao J, Xue F, 2010 A gene-based method for detecting genegene co-association in a casecontrol association study. *European Journal of Human Genetics* 18, 582–587. [PubMed: 20029457]
- Potkin SG, Van, E. TGM, Ling S, Macciardi F, Xie X, 2015 *Unanticipated Genes and Mechanisms in Serious Mental Illness: GWAS based Imaging Genetics Strategies*. Vol. 209 Oxford University Press, London.
- Richfield O, Alam MA, Calhoun V, Wang YP, 2017 Learning schizophrenia imaging genetics data via multiple kernel canonical correlation analysis. *Proceedings - 2016 IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2016, Shenzhen, China* 5, 507–5011.
- Yu S, Tranchevent L-C, B. D. M., Moreau Y, 2011 *Kernel-based Data Fusion for Machine Learning*. Springer, Verlag Berlin Heidelberg.
- Sanders AR, Duan J, Levinson DF, et al., 2008 No significant association of 14 candidate genes with schizophrenia in a large european ancestry sample: implications for psychiatric genetics. *The American Journal of Psychiatry* 165(10), 1359–1368.
- Satterthwaite FE, 1946 An approximate distribution of estimates of variance components. *Biometrics Bulletin* 2(6), 110–114. [PubMed: 20287815]
- Schölkopf B, Smola AJ, 2002 *Learning with Kernels*. MIT Press, Cambridge MA.

- Schölkopf B, Smola AJ, Müller K-R, 1998 Nonlinear component analysis as a kernel eigenvalue problem. *Neural Computation*. 10, 1299–1319.
- Shibuya M, Watanabe Y, Nunokawa A, Egawa J, Kaneko N, Igeta H, Someya T, 2013 Interleukin 1 beta gene and risk of schizophrenia: detailed case-control and family-based studies and an updated meta-analysis. *Human Psychopharmacology* 29, 31–37.
- Siawa GE-L, Liuc I-F, Linc PY, Beend MD, Hsiehc T, 2016 Dna and rna topoisomerase activities of top3 are promoted by mediator protein tudor domain-containing protein 3. *Proc Natl Acad Sci USA* 113, 5544–5551.
- Song L, Smola A, Gretton A, Bedo J, Borgwardt K, 2012 Feature selection via dependence maximization. *Journal of Machine Learning Research* 13, 1393–1434.
- Sriperumbudur BK, Fukumizu K, Gretton A, Lanckriet GRG, Schölkopf B, 2009 Kernel choice and classifiability for rkhs embeddings of probability distributions. *Advances in Neural Information Processing Systems* 21, 1750–1758.
- Strausberg RL, Feingold EA, Grouse LH, et al., 2002 Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proceedings of the National Academy of Sciences, USA* 99(26), 16899–903.
- Suk H, Wee C, Lee S, Shen D, 2016 State-spacemodel with deep learning for functional dynamics estimation in resting-state fmri. *NeuroImage* 129, 292–307. [PubMed: 26774612]
- Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, Simonovic M, h. A. Rot, Santos A, Tsafou KP, and, M. K., Bork P, Jensen LJ, von Mering C, 2007 STRING v10: Proteinprotein interaction networks, integrated over the tree of life. *Nucleic Acids Research* 43, 531–543.
- Van SO, Kapur S, 2009 Schizophrenia. *Lancet* 374 (9690), 635645.
- Wan X, Yang C, Yang Q, Xue H, Fan X, Tang NL, Yu W, 2010 Boost: A fast approach to detecting gene-gene interactions in genome-wide case-control studies. *The American Journal of Human Genetics* 87, 325–340. [PubMed: 20817139]
- Wang T, Ho G, K. Y., Strickler H, Elston RC, Schaid DJ, 2009 A partial least-square approach for modeling genegene and gene-environment interactions when multiple markers are genotyped. *Genet. Epidemiol* 33, 615.
- Wen H, Liu Y, Rekik I, Wang S, Chen Z, Zhang J, Zhang Y, Peng Y, He H, 2017 Multi-modal multiple kernel learning for accurate identification of tourette syndrome children. *Pattern Recognition* 63, 601–611.
- Wockner LF, Noble EP, Lawford BR, Young RM, Morris CP, Whitehall VLJ, Voisey J, 2014 Genome-wide dna methylation analysis of human brain tissue from schizophrenia patients. *Transl Psychiatry* 4 (e339), 1–8.
- Wu K, Taki Y, Sato K, Oi H, Kawashima R, Fukuda H, 2013 A longitudinal study of structural brain network changes with normal aging. *Frontiers in Human Neuroscience* 7, 225–236. [PubMed: 23761742]
- Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X, 2011 Rare variant association testing for sequencing data using the sequence kernel association test (SKAT). *American Journal of Human Genetics* 89, 82–93. [PubMed: 21737059]
- Xu C, Tao D, Xu C, 2013 A survey of multi-view machine learning. *Neural Computation and Applications* 23(7–8), 2031–2038.
- Yan C, Zang Y, 2010 DPARSF: a MATLAB toolbox for pipeline data analysis of resting-state fMRI. *Frontiers in Systems Neuroscience* 4 (13), 1–7. [PubMed: 20204156]
- Yuan Z, Gao Q, He Y, Zhang X, Li F, Zhao J, Xue F, 2012 Detection for gene-gene co-association via kernel canonical correlation analysis. *BMC Genetic* 13:83.
- Zhao F, Qiao L, Shi F, P. T. Y., Shen D, 2016 Feature fusion via hierarchical supervised local cca for diagnosis of autism spectrum disorder. *Brain Imaging and Behavior*, 1–11. [PubMed: 25749917]
- Zheng S, Cai X, Ding CH, Nie F, Hung H, 2015 A closed form solution to multi-view lowrank regression. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI-15)*, 1973–1979.

Highlights:

- The proposed method, a kernel machine method for detecting higher order interactions (KMDHOI), is a multimodal semiparametric method on a reproducing kernel Hilbert space.
- KMDHOI is apply to multimodal datasets for detecting higher order interactions (e.g. genetics, brain imaging, and epigenetic data).
- The identified triplets (gene-derived SNPs, ROIs, and gene-specific DNA methylations) have explained Schizophrenia-related neurodegeneration.

Figure 1:

An illustration of different imaging and (epi)-genetics data along with their interaction effect on human behavior. Note, SNP: single nucleotide polymorphism, DNA: deoxyribonucleic acid methylations, PET: positron emission tomography (PET), fMRI: functional magnetic resonance imaging (fMRI), sMRI: structural MRI, GE: gene expression.

Figure 2:

The Receiver operating characteristics (ROC) of the kernel methods and relevant ones for higher order interaction detection with three sample sizes, $n \in \{100, 500, 1000\}$ for (a) third parameter value is random, (b) second and third parameter values are random. The sensitivity is plotted against (1- specificity) with the p -value threshold in the range of (0 – 1) with a step size 0.0001.

Figure 3:
The plot of $-\log_{10}(p)$ with 15436 triplets.

Figure 4:

The graph shows the network of interactions within and between different views ($p < 0.001$).

Each node represents gene-derived SNPs (G1: IL1B, G2: MAGI2, G3: NRG1, G4: PDLIM5, G5: SLC18A1, G6: TDRD), ROIs (R1: CRBL8.L, R2: CRBLCrus1.L, R3: ORBSUP.R, R4: LING.L, R5: CAU.R, R6: IPL.L, R7: IPL.R, R8: PoCG.L, R9: ITG.R, R10: VER54) and gene-specific DNA methylation (E1: CRABP1, E2: FBXO28, E3: DUSP1, E4:FHIT, E5: PLAGL1, E6: TFPI2).

Figure 5:
The visualization of selected 10 ROIs for schizophrenia and healthy control groups: (a) correlation matrices, (b) axial view of all networks.

Figure 6:

The boxplot of significant interaction effects in different regimes (SZ and high BMI risk, SZ and low BMI risk, healthy control and high BMI risk, healthy control and low BMI risk) for the most significant triplet (**MAGI2**, **CRBL Crus1.L**, **FBXO28**).

Table 1:

The power of the overall and higher order interaction (HOI) score tests, and the alternative methods for interaction detection using dimension reduction regression (e.g., pPCAR, fPCAR) and sequence kernel association test (SKAT). The nominal p -value threshold was fixed to 0.05.

Parameters $(\alpha_1, \alpha_2, \alpha_3)$	Simulation - I				
	KMDHOI		State-of-the-art methods		
	Overall	HOI	pPCAR	fPCAR	SKAT
			HOI	HOI	HOI
(0, 0, 0)	0.048	–	0.044	0.036	0.052
(0.5, 0, 0)	1.00	0.049	0.137	0.041	0.127
(1, 0, 0)	1.00	0.048	0.060	0.015	0.143
(0, 0.5, 0)	1.00	0.052	0.137	0.053	0.160
(0, 0.5, 0.5)	1.00	0.764	0.161	0.093	0.410
(0, 0.5, 1)	1.00	0.853	0.157	0.082	0.478
(0, 0, 1)	1.00	0.823	0.173	0.080	0.787
(0, 0, 1)	1.00	0.872	0.213	0.110	0.748
(0.5, 0.5, 0.5)	1.00	0.765	0.174	0.107	0.403
(1, 1, 1)	1.00	0.858	0.207	0.111	0.526

Table 2:

The power of overall and higher order interaction (HOI) score tests, and the alternative methods for interaction detection using dimension reduction regression (e.g., pPCAR, fPCAR) and sequence kernel association test (SKAT). The nominal p -value threshold was fixed to 0.05.

Parameters $(\alpha_1, \alpha_2, \alpha_3)$	Simulation - II				
	KMDHOI		State-of-the-art methods		
	Overall	HOI	pPCAR HOI	fPCAR HOI	SKAT HOI
(0, 0, 0)	0.053	–	0.046	0.056	0.045
(0.5, 0, 0)	0.999	0.051	0.172	0.077	0.188
(1, 0, 0)	0.999	0.049	0.199	0.077	0.176
(0, 0.5, 0)	1.000	0.052	0.189	0.095	0.261
(0, 0.5, 0.5)	1.000	0.899	0.199	0.107	0.288
(0, 0.5, 1)	1.000	0.904	0.173	0.092	0.305
(0, 0, 1)	1.000	0.935	0.231	0.132	0.362
(0, 0, 1)	1.000	0.924	0.283	0.115	0.351
(0.5, 0.5, 0.5)	1.000	0.857	0.151	0.117	0.381
(1, 1, 1)	1.000	0.906	0.190	0.110	0.405

Table 3:

The selected significant genes-derived SNP, ROIs and gene-specific DNA methylation using the proposed method (KMDHOI) and SKAT. The p -value was set to be 0.001.

Genetics	Imaging	Epigenetics	σ^2	$\tau^{(1)}$	$\tau^{(2)}$	KMDHOI					OVA	HOI	SKAT
						$\tau^{(3)}$	$\tau^{1 \times 2}$	$\tau^{1 \times 3}$	$\tau^{2 \times 3}$	$\tau^{1 \times 2 \times 3}$			HOI
IL1B	CAU.R	FBXO28	0.6755	0.0038	0.0229	0.1225	1.1013	0.0000	0.1307	1.3606	0.0383	2E-04	0.4943
IL1B	PoCG.R	FBXO28	0.5837	0.0189	0.1827	0.1247	1.8403	0.0000	0.3469	1.0603	0.0202	7E-04	0.2871
MAGI2	CRBL Crus1.L	FBXO28	0.1833	0.3246	0.0000	0.2693	1.1963	2.2426	1.3683	0.0100	0.0288	0E-06	0.01891
MAGI2	LING.L	CRABP1	0.1813	0.4366	0.0000	0.0885	1.5370	2.6299	1.0271	0.0100	0.0470	0E-05	0.4030
MAGI2	IPL.R	FBXO28	0.1833	0.3778	0.0000	0.3044	0.9808	2.3888	1.2203	0.0100	0.0457	0E-05	0.5592
NRG1	IPL.L	PLAGL1	0.3682	0.0024	0.2162	0.1270	1.1227	2.7930	0.0000	0.2056	0.0284	0E-05	0.6664
PDLIM5	IPL.L	DUSP1	0.3648	0.0000	0.0804	0.2182	1.5177	1.9409	1.0276	0.0100	0.0183	5E-04	0.0173
PDLIM5	PoCG.R	DUSP1	0.3598	0.0000	0.2256	0.0139	0.9189	1.7853	1.2796	0.0100	0.0498	Ee-04	0.2761
SLC18A1	ORBsup.R	FHIT	0.4096	0.0065	0.5356	0.2276	1.3456	0.0000	1.0003	0.1323	0.0495	1E-04	0.0522
SLC18A1	ORBsup.R	PLAGL1	0.2869	0.0000	0.3909	0.1933	1.0186	1.3148	1.3676	0.0100	0.0373	3E-04	0.0149
SLC18A1	Vermis45	TFPI2	0.5571	0.0447	0.0815	0.0020	0.0000	1.0458	0.5579	0.0100	0.0354	7E-04	0.4234
TDRD3	CRBL8.L	FBXO28	0.5856	0.2284	0.0022	0.0702	1.1722	0.7667	0.0000	0.0100	0.0052	0E-05	0.0807
TDRD3	ITG.R	CRABP1	0.5291	0.2318	0.0000	0.0223	0.7052	0.6038	0.6240	0.0100	0.0033	4E-04	0.0359

Table 4:

The network measurements (e.g., transitivity, degree, and global efficiency) of the selected 10 ROIs for schizophrenia and healthy control groups, respectively.

ROIs	Transitivity		Degree		Global efficiency	
	Schizophrenia	Healthy	Schizophrenia	Healthy	Schizophrenia	Healthy
R1 = CRBL8.L	0.571	1.000	7	2	0.800	0.317
R2 = CRBL Crus1.L	0.600	0.333	6	3	0.750	0.400
R3 = ORBSUP.R	0.667	0.000	3	1	0.583	0.100
R4 = LING.L	0.667	0.333	7	3	0.800	0.400
R5 = CAU.R	0.700	0.000	5	1	0.683	0.100
R6 = IPL.L	1.000	1.000	2	2	0.500	0.317
R7 = IPL.R	0.667	1.000	4	2	0.650	0.317
R8 = PoCG.L	0.667	1.000	7	2	0.800	0.316
R9 = ITG.R	0.900	0.000	5	1	0.683	0.100
R10 = VER45	0.800	0.000	6	1	0.750	0.100

Table 5:

The selected significant gene-derived SNPs, ROIs and gene-specific DNA methylations using the proposed KMDHOI method and SKAT at $p = 0.01$. The bold indicates the significance level at $p = 0.001$. Note: the name of ROI is given by the AAL template.

Genetics	IL1B	MAGI2	NRG1	PDLIM5	SLC18A1	TDRD3	BDNF	CHGA	CHGB	CLINT1
	COMTD1	DAOA	DISC1	DRD2	DTNBP1	ERBB4	GABBR1	GABRB2	GRIN2B	GRM3
	HTR2A SNAP29	IL10RA	MAGI1	MICB	NOS1AP	NOTCH4	NR4A2	NUMBL	PLXNA2	PPP3CC
Imaging	CRBL8.L	CRBLCrus1.L	ORBSUP.R	LING.L	CAU.R	IPL.L	IPL.R	PoCG.L	ITG.R	VER45
	AMYG.L	CRBL10.R	CRBL10.L	CRBL3.R	CRBL3.R	CRBL45.L	CRBL6.L	CRBL8.R	CRBLCrus2.R	CRBLCrus2.L
	DCG.R	DCG.L	PCG.R	ORBsup.L	ORBmid.R	LING.R	ROL.R	SMA.R	TPOsup.R	TPOsup.L
	STG.L	ITG.L	Vermis10	Vermis3	MTG.R					
Epigenetic	CRABP1	FBXO28	DUSP1	FHIT	PLAGL1	TFPI2	CCND2	CDKN1A	EDNRB	ESR1
	EYA4	FEN1	GPSN2	HOXA9	HOXB4	PTGS2	RB1	SRF	WDR37	ZNF512

Table 6:

The classification error of discriminating schizophrenia patients from healthy controls with cross-validations.

Methods	Classifier	2-folds	5-fold
Multiple CCA	KNN	22.571 ± 0.278	35.169 ± 0.194
	SVM	19.489 ± 0.459	31.258 ± 0.401
Multiple KCCA	KNN	15.269 ± 0.562	30.625 ± 0.475
	SVM	14.785 ± 0.627	28.156 ± 0.617
KMDHOI	KNN	11.865 ± 0.275	21.627 ± 0.284
	SVM	10.329 ± 0.372	20.022 ± 0.278