JOURNAL OF COMPUTATIONAL BIOLOGY
Volume 29, Number 8, 2022
© Mary Ann Liebert, Inc.
Pp. 892–907
DOI: 10.1089/cmb.2021.0598

Open camera or QR reader and scan code to access this article and other resources online.

SDGCCA: Supervised Deep Generalized Canonical Correlation Analysis for Multi-Omics Integration

SEHWAN MOON,[1,*] JEONGYOUNG HWANG,[2,*] and HYUNJU LEE[1,2]

ABSTRACT

Integration of multi-omics data provides opportunities for revealing biological mechanisms related to certain phenotypes. We propose a novel method of multi-omics integration called supervised deep generalized canonical correlation analysis (SDGCCA) for modeling correlation structures between nonlinear multi-omics manifolds that aims at improving the classification of phenotypes and revealing the biomarkers related to phenotypes. SDGCCA addresses the limitations of other canonical correlation analysis (CCA)-based models (such as deep CCA, deep generalized CCA) by considering complex/nonlinear cross-data correlations between multiple ($\geq$2) modalities. Although there are a few methods to learn nonlinear CCA projections for classifying phenotypes, they only consider two views. Methods extended to multiple views either do not perform classification or do not provide feature ranking. In contrast, SDGCCA is a nonlinear multi-view CCA projection method that performs classification and ranks features. When we applied SDGCCA in predicting patients with Alzheimer's disease (AD) and discrimination of early- and late-stage cancers, it outperformed other CCA-based and other supervised methods. In addition, we demonstrate that SDGCCA can be applied for feature selection to identify important multi-omics biomarkers. On applying AD data, SDGCCA identified clusters of genes in multi-omics data, well known to be associated with AD.

Keywords: Alzheimer's disease, canonical correlation analysis, deep neural networks, multi-omics, supervised learning.

[1]School of Electrical Engineering and Computer Science, Gwangju Institute of Science and Technology, Gwangju, South Korea.
[2]Artificial Intelligence Graduate School, Gwangju Institute of Science and Technology, Gwangju, South Korea.
*These authors contributed equally to this work.

1. INTRODUCTION

The advent of sequencing technology has facilitated the collection of genome-wide data for different molecular processes (such as gene expression [GE], DNA methylation [ME], and microRNA [miRNA] expression [MI]), resulting in multi-omics data analysis from the same set of individuals or biospecimens. Exploring molecular mechanisms using multi-omics data is expected to improve our current knowledge of diseases, leading to further improvements in disease diagnosis, prognosis, and personalized treatment. Although single-omics analysis can only capture a part of the biological complexity of the disease, integration of multi-omics data is required to provide a comprehensive overview of the underlying biological mechanisms.

Various methods such as unsupervised data integration models based on matrix factorization and correlation-based analysis, supervised data integration models based on network-based methods and multiple kernel learning, and Bayesian methods have been proposed for multi-omics data integration (Huang et al, 2017). For example, multi-omics factor analysis (Argelaguet et al, 2018) is a Bayesian method for multi-omics integration that extracts the shared axes of variation between the different omics. Sparse generalized canonical correlation analysis (sGCCA) (Tenenhaus et al, 2014) is a generalization of regularized canonical correlation analysis (CCA) with an L1-penalty model that selects co-expressed variables from omics datasets.

Recently, researchers have been interested in multi-omics biomarkers that can explain or characterize a known phenotype. Data Integration Analysis for Biomarker discovery using Latent cOmponents (DIABLO) (Singh et al, 2019) extends sGCCA to a supervised framework for identifying shared molecular patterns that can explain phenotypes across multi-omics. However, most of these methods are linear representations that cannot capture complex biological processes.

The CCA (Hotelling, 1992) is a well-known multivariate model for capturing the associations between any two sets of data. The CCA and its variations have been applied in several studies (Tenenhaus et al, 2014; Mandal and Maji, 2017; Jendoubi and Strimmer, 2019; Singh et al, 2019) because of its advantages in biological interpretation. However, a drawback of CCA is that it can only consider the linear relationship of two modalities to correlate them maximally. Generalized canonical correlation analysis (GCCA) (Kettenring, 1971) extends the CCA to consider more than two modalities.

To complement the GCCA, deep generalized canonical correlation analysis (DGCCA) (Benton et al, 2017) considers the nonlinear relationship learning of more than two modalities. Further, supervised deep CCA (Liu et al, 2017) and task-optimal CCA (Couture et al, 2019) have been proposed for supervised learning while considering nonlinear maximal correlation, but they can be only applied to two modalities.

In this study, we proposed a supervised deep generalized canonical correlation analysis (SDGCCA), a nonlinear supervised learning model integrating with multiple modalities for discriminating phenotypic groups. The SDGCCA identifies the common and correlated information between multiple omics data, crucial for discriminating phenotypic groups. The SDGCCA is also based on a deep neural network (DNN), allowing the powerful capturing of the nonlinear part of the biological complexity. After training SDGCCA, we utilized Shapley additive explanation (SHAP) (Lundberg and Lee, 2017) to identify correlated biomarkers contributing to classification.

2. RELATED WORK

In this section, we briefly review relevant previous studies. Table 1 presents all the notions considered throughout the study.

2.1. Canonical correlation analysis

The CCA is a representative method for dimension reduction that can consider the correlation between two modalities. It is trained to maximize the correlation between two mapped matrices using the projection matrices of each of the two modalities. The objective function of CCA is as follows:

$$(V_1^*,\ V_2^*) = \underset{V_1,\ V_2}{\operatorname{argmax}}\ corr(V_1^\top X_1,\ V_2^\top X_2) = \underset{V_1,\ V_2}{\operatorname{argmax}} \frac{V_1^\top \Sigma_{12} V_2}{\sqrt{V_1^\top \Sigma_{11} V_1 V_2^T \Sigma_{22} V_2}}, \tag{1}$$

Table 1. Notations Used in Equations 1–12

Notation	*Dimension*	*Description*
n	—	Number of samples
m	—	Number of modalities
k	—	Dimensions of the shared representation
c	—	Number of label categories
d_i	—	Dimensions of the ith modality.
$\overline{d_i}$	—	Output dimensions of a DNN of the i-th modality.
$f_i(\cdot)$	—	Deep neural network of the i-th modality.
θ_i	—	Parameters of $f_i(\cdot)$.
X_i	$d_i \times n$	i-th modality
V_i	$d_i \times k$	Projection matrix for X_i
U_i	$\overline{d_i} \times k$	Projection matrix for $f_i(X_i)$
Y	$c \times n$	Label
U_y	$c \times k$	Projection matrix for Y
$U_y^{\dagger}$	$c \times k$	Pseudo inverse of U_y
G	$k \times n$	Shared representation

DNN, deep neural network.

where X_i denotes the i-th modality; Σ_{11} and Σ_{22} denote covariance matrices of X_1 and X_2, respectively; Σ_{12} denotes a cross-covariance matrix; V_i denotes a projection matrix for the i-th modality; and V_1^* and V_2^* can be adopted to select the relevant features in both modalities. Since the objective function above is invariant for scaling of V_1 and V_2, the final objective function is expressed as follows by adding the constraints of the unit variance.

$$\begin{aligned}(V_1^*,\ V_2^*) &= \underset{V_1,\ V_2}{\operatorname{argmax}}\ V_1^{\top}\Sigma_{12}V_2,\\ \text{s.t. } V_1^{\top}\Sigma_{11}V_1 &= V_2^{\top}\Sigma_{22}V_2 = I\end{aligned} \tag{2}$$

However, the CCA is limited to mapping linear relationships and can only leverage two modalities.

2.2. Deep canonical correlation analysis

A deep canonical correlation analysis (DCCA) (Andrew et al, 2013) is used to overcome the limitations of CCA that extracts only the linear relationship. In DCCA, a DNN is applied to each modality to consider a nonlinear relationship. The DCCA is learned by maximizing the correlation between the projected representations of modalities. The objective function of DCCA is as follows:

$$(\theta_1{}^*,\ \theta_2{}^*,\ U_1{}^*,\ U_2{}^*) = \underset{U_1,\ U_2}{\operatorname{argmax}}\ corr(U_1^{\top}f_1(X_1),\ U_2^{\top}f_2(X_2)), \tag{3}$$

where $f_i(.)$ is a DNN function for the i-th modality, U_i indicates a projection matrix for $f_i(X_i)$, and θ_i is a parameter of $f_i(.)$. $\theta_i(.)$ is trained via back-propagation that maximizes the objective function of DCCA. However, DCCA maximizes the correlation between the representations projected by each DNN output unlike CCA. Therefore, it cannot directly extract correlated features in both modalities. The DCCA also cannot be applied to more than two modalities.

2.3. Generalized canonical correlation analysis

The GCCA is used to extend the CCA to more than two modalities. The GCCA learns projection metrics that map each modality to a shared representation. The objective function of GCCA is as follows:

$$\begin{aligned}\underset{V_1,\ldots,V_m,\,G}{\text{minimize}} &\sum_{i=1}^{m} \|\ G - V_i^{\top}X_i\ \|_F^2\ ,\\ \text{s.t. } GG^{\top} &= I,\end{aligned} \tag{4}$$

where G denotes the shared representation, and V_i indicates a projection matrix for X_i. To solve the objective function of GCCA, an eigen decomposition of an $n \times n$ matrix is required, which increases

quadratically with sample size and leads to memory constraints. Also, unlike DCCA, nonlinear associations between modalities cannot be considered.

2.4. Deep generalized canonical correlation analysis

The DGCCA is a model that addresses the limitations of CCA by including the advantages of GCCA and DCCA. The DGCCA learns projection metrics that map each output of DNN to a shared representation. The objective function of DGCCA is as follows:

$$\begin{aligned} \underset{U_1, \ldots, U_m, G}{\text{minimize}} & \sum_{i=1}^{m} \| G - U_i^\top f_i(X_i) \|_F^2 , \\ \text{s.t.} & GG^\top = I. \end{aligned} \tag{5}$$

U_i and G are trained to reduce the reconstruction error of GCCA, and to update θ_i, gradients are back-propagated through the neural network. The gradient propagating to $f_i(X_i)$ is defined as $2U_iG - 2U_iU_i^\top f_i(X_i)$, and θ_i can be updated with back-propagation to minimize the objective function of DGCCA. As θ_i is updated, the value of $f_i(X_i)$ is changed.

Therefore, to solve the objective function of DGCCA, updating U_i and G and updating θ_i are alternately performed. The DGCCA has the advantage of obtaining the nonlinear relationship of each modality. It can also consider the correlation between more than two modalities.

2.5. Data Integration Analysis for Biomarker discovery using Latent cOmponents

DIABLO extends sGCCA, which is a GCCA with L1-penalty. It is different from sGCCA as (1) the correlation between linear combinations of multi-omics data is changed to covariance; and (2) unlike sGCCA, which is an unsupervised method, it is a supervised framework that is capable of classification by maximizing the covariance between multiple omics datasets, including phenotype information. The objective function of DIABLO is as follows:

$$\begin{aligned} \underset{V_1, \ldots, V_m, U_y}{\text{maximize}} & \sum_{i,j=1; i \neq j}^{m} D_{i,j}\, cov(V_i^\top X_i,\ V_j^\top X_j) + \sum_{l=1}^{m} D_{l,y}\, cov(V_l^\top X_l,\ U_y^\top Y), \\ \text{s.t.} & \ \| V_i \|_2 = 1 \text{ and } \| V_i \|_1 \leq \lambda_i,\ \| U_y \|_2 = 1 \text{ and } \| U_y \|_1 \leq \lambda_y, \end{aligned} \tag{6}$$

where $D = \{D_{i,j}\} \in \mathbb{R}^{(m+1) \times (m+1)}$ is a design matrix that determines whether datasets should be connected. However, DIABLO has a limitation that only assumes a linear relationship between the selected features to explain the phenotype.

3. METHODS

3.1. The SDGCCA method

The SDGCCA proposed in this study integrates ideas from DGCCA and DIABLO. The SDGCCA incorporates the phenotypes of samples for supervised learning and selects the significant features based on CCA. It uses DNN to consider nonlinear interactions between multi-omics data as well as phenotypes (Fig. 1). The SDGCCA makes it possible to predict the phenotypes of samples by adding two elements to DGCCA.

First, the correlation between each modality and that with the labels was considered. Thus, the shared representation G can be trained to obtain the label information. The correlation loss function is defined as follows:

$$\begin{aligned} L_{corr} = & \| G - U_y^\top Y \|_F^2 + \sum_{i=1}^{m} \| G - U_i^\top f_i(X_i) \|_F^2 , \\ \text{s.t. } & GG^\top = I, \end{aligned} \tag{7}$$

where U_y denotes the projection matrix for label Y. Second, cross-entropy (De Boer et al, 2005), which is widely used in supervised models, is used to enable the propagation of label information directly to the DNN of each modality. The projection matrix U_y obtained from Eq. (7) can map the label to the shared representation. In addition, a projection matrix U_i maps each modality to a shared representation. Using the pseudo-inverse matrix of a projection matrix U_y, label Y can be approximated as follows:

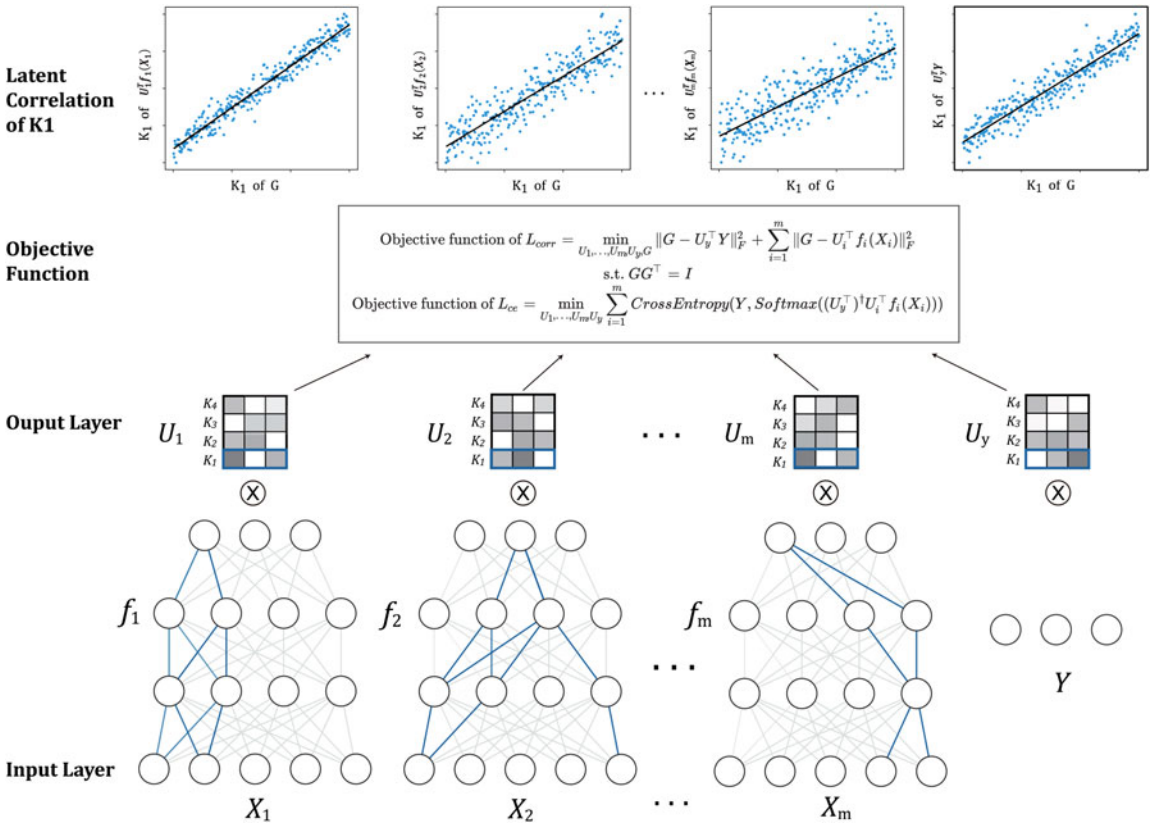

FIG. 1. A schematic of SDGCCA. $X_1, \ldots, X_m$ are m modality, and Y is the label information. DNNs $f_1, \ldots, f_m$ operate on $X_1, \ldots, X_m$. The outputs of each modality and Y are multiplied by each projection matrix ($U_1, \ldots, U_m, U_y$). Two objective functions search for the optimal network $f_1, \ldots, f_m$ and projection matrices, which provide both the highest correlation and the lowest prediction error. DNNs, deep neural networks; SDGCCA, supervised deep generalized canonical correlation analysis.

$$\begin{aligned} G &\approx U_i^\top f_i(X_i) \approx U_y^\top Y, \\ Y &\approx (U_y^\top)^\dagger U_i^\top f_i(X_i), \end{aligned} \tag{8}$$

where $U_y^\dagger$ denotes the pseudo-inverse of U_y, Let $\hat{Y}_i = (U_y^\top)^\dagger U_i^\top f_i(X_i)$. By applying the softmax function to $\hat{Y}_i$, the model was trained using cross-entropy. Classification loss is defined as follows:

$$L_{ce} = \sum_{i=1}^{m} CrossEntropy\ (Y,\ Softmax(\hat{Y}_i)). \tag{9}$$

The final label prediction of SDGCCA uses soft voting of the label representation ($\hat{Y}_i$) of each modality. The label prediction of SDGCCA is defined as follows:

$$\hat{Y} = Softmax((\sum_{i=1}^{m} \hat{Y}_i)/m), \tag{10}$$

where m denotes the number of modalities, The optimization of the proposed model consists of three main steps. First, $U_i, \ldots, U_m, U_y$, and G were trained using the correlation loss function (L_{corr}). Here,G is obtained by solving the eigenvalue problem. Let $C_{ii} = f_i(X_i)f_i(X_i)^\top$, s.t. $i = 1, \ldots, m$, $C_{(m+1)(m+1)} = YY^\top$, $P_i = f_i(X_i)^\top C_{ii}^{-1} f_i(X_i) \in \mathbb{R}^{n\times n}$, $P_{m+1} = Y^\top C_{(m+1)(m+1)}^{-1} Y \in \mathbb{R}^{n\times n}$, and $M = \sum_{i=1}^{m+1} P_i$.

Subsequently, the rows of $G \in \mathbb{R}^{k\times n}$ are orthonormal as the top k eigenvectors of M. After such G is obtained, the following can be easily obtained: $U_i = C_{ii}^{-1} f_i(X_i) G^\top$ and $U_y = C_{(m+1)(m+1)}^{-1} YG^\top$. Second, θ_i of $f_i(.)$ was trained using L_{corr}. It can be updated by selecting only the part related to θ_i in L_{corr} and finding gradients to back-propagate to $f_i(X_i)$ as follows:

$$
\begin{aligned}
&\| G - U_y^\top Y \|_F^2 + \sum_{i=1}^{m} \| G - U_i^\top f_i(X_i) \|_F^2 , \\
&= \| G - GY^\top C_{(m+1)(m+1)}^{-1} Y \|_F^2 + \sum_{i=1}^{m} \| G - G f_i(X_i)^\top C_{ii}^{-1} f_i(X_i) \|_F^2 , \\
&= \sum_{i=1}^{m+1} \| G(I_n - P_i)) \|_F^2 , \\
&= \sum_{i=1}^{m+1} \mathrm{Tr}[G(I_n - P_i)G^\top], \\
&= \sum_{i=1}^{m+1} \mathrm{Tr}(I_k) - \mathrm{Tr}(GMG^\top), \\
&= (m+1)k - \mathrm{Tr}(GMG^\top).
\end{aligned}
\tag{11}
$$

As mentioned earlier, L_{corr} can be minimized by maximizing $Tr(GMG^\top)$, and the derivative of $Tr(GMG^\top)$ with respect to $f_i(X_i)$ is demonstrated in the DGCCA (Benton et al, 2017) as $2U_iG - 2U_iU_i^\top f_i(X_i)$. Finally, after substituting the U_i and U_y obtained earlier into Eq. 7, θ_i is trained using L_{ce}. A detailed algorithm for training SDGCCA using Eq. (7–11) is summarized in Algorithm 1.

Algorithm 1: Training the proposed model

Input: Training dataset $X = [X_1, X_2, \ldots, X_m]$, regularization rate α, learning rate β, and max iterations T
Output: Projection matrices $U_1, \ldots, U_m, U_y$, parameters θ_i of f_i

Initialize θ_i by He initialization
$t = 1$
while: Validation loss does not converge or t $\leqq$ T
  Step 1. Calculate $U_1, \ldots, U_m, U_y, G$
    $L_{corr} = \| G - U_y^\top Y \|_F^2 + \sum_{i=1}^{m} \| G - U_i^\top f_i(X_i) \|_F^2$
    $U_1, \ldots, U_m, U_y, G = \underset{U_1, \ldots, U_m, U_y, G}{\mathrm{argmin}} L_{corr}$
      G is obtained by solving an eigenvalue problem.
      $U_i = C_{ii}^{-1} f_i(X_i) G^\top$, s.t. $C_{ii} = f_i(X_i) f_i(X_i)^\top$
      $U_y = C_{(m+1)(m+1)}^{-1} YG^\top$, s.t. $C_{(m+1)(m+1)} = YY^\top$
  Step 2. Training θ_i by using L_{corr}
    $\nabla_{f_i(X_i)} L_{corr} = U_i U_i^\top f_i(X_i) - U_i G$
    $\theta_i \leftarrow (1 - \alpha)\theta_i - \beta \nabla_{\theta_i} \nabla_{f_i(X_i)} L_{corr}$
  Step 3. Training θ_i by using L_{ce}
    $\hat{Y}_i = (U_y^\top)^\dagger U_i^\top f_i(X_i)$
    $L_{ce} = \sum_{i=1}^{m} CrossEntropy(Y, Softmax(\hat{Y}_i))$
    $\theta_i \leftarrow (1 - \alpha)\theta_i - \beta \nabla_{\theta_i} L_{ce}$
  t $\leftarrow t + 1$
end while

3.2. Identification of multi-omics biomarkers

The SDGCCA is trained by maximizing the correlation between the shared representation and projected representation for each modality, and a projection matrix can be used to select the most correlated dimension of the DNN output of each modality. Because SDGCCA uses eigen decomposition to obtain a projection matrix as a CCA-based model, it can be observed that the correlation value of the first component ($U_i[:, 1]$) is the largest among the values mapped to the shared representation through each modality.

Therefore, the most correlative dimension of each DNN output corresponds to the highest coefficient of the projection matrix (argmax$|U_i[:, 1]|$), which is

$$
f_i(\cdot)[\mathrm{argmax}\ |U_i[:, 1]|, :] \tag{12}
$$

However, unlike CCA and GCCA, the model is difficult to interpret because of the DNN. Thus, we used SHAP to select features related to the most correlative dimension of each DNN output. In Lundberg and Lee (2017), SHAP calculated the feature importance using the SHAP value that satisfied the desirable properties (local accuracy [ACC], missingness, and consistency) for each prediction.

Specifically, we used deep SHAP, which is tailored to the DNN and effectively combines the SHAP values calculated for smaller components of a DNN into SHAP values for the entire DNN. In deep SHAP, feature importance is calculated as the extent to which the output of the arbitrary input deviates from that of the network of the reference input, computing SHAP values recursively for smaller components of the model.

4. RESULTS

4.1. Datasets

We applied the proposed method to an Alzheimer's disease (AD) classification task using multi-omics data. Three types of omics data (mRNA, ME, and miRNA) and clinical data were obtained from the Religious Order Study and Rush Memory and Aging Project (ROSMAP) cohort in the AMP-AD Knowledge Portal (https://adknowledgeportal.synapse.org/). We downloaded mRNA data that were normalized with quantile normalization to fragments per kilobase of transcript per million mapped reads (FPKM) and removed potential batch effects using Combat (Johnson et al, 2007).

The β-values of the downloaded ME data were measured using the Infinium HumanMethylation450 BeadChip, and the missing β-values were imputed using a k-nearest neighbor algorithm. We downloaded miRNA data normalized using a variance stabilization normalization method (Huber et al, 2002) and removed potential batch effects using Combat. Patients with AD ($n = 207$) and normal controls ($n = 169$) with GE, ME, and MI profiles were included.

The normalized FPKM values of the GE profiles were log2-transformed. For ME data, CpG sites located in promoter regions (TSS200 or TSS1500) were mapped to the corresponding gene, and the β-values of the promoter CpG sites for the same genes were averaged to create a score per gene. Finally, 18,164 GE, 19,353 ME, and 309 MI features were obtained.

To further measure the performance of the proposed method, we used kidney renal clear cell carcinoma (KIRC) data collected from The Cancer Genome Atlas (TCGA) for the early- and late-stage classification. The TCGA level-3 data on GE (Illumina mRNAseq), ME (Illumina HumanMethylation450 BeadArray), and MI (Illumina HiSeq miRNAseq) were obtained. The ME data used in this study were pre-processed according to Ma et al (2020). Finally, KIRC data comprised 313 samples (184 and 129 early- and late-stage samples, respectively) on 16,406 GE, 16,459 ME, and 342 MI features.

4.2. Existing methods for performance comparison

We compared the classification performance of SDGCCA using the following 10 existing methods. We selected widely utilized machine- or deep learning models and CCA-based multi-omics integration methods to determine the mechanism by which SDGCCA can contribute to the CCA framework. We selected (1) support vector machine (SVM), (2) extreme gradient boosting (XGB) (Chen et al., 2016), (3) logistic regression (LR), and (4) random forest (RF) as a method of machine learning, and (5) DNN as methods of deep learning.

For CCA-based methods, (6) GCCA, (7) DGCCA, and (8) DIABLO (Singh et al, 2019) were selected. The SVM was used as an additional classification model, because GCCA and DGCCA are unsupervised learning models. In addition, the performance was compared with (9) Multi-Omics Graph cOnvolutional NETworks (MOGONET) (Wang et al, 2021) and (10) multimodal self-paced learning with a soft weighting scheme (SMSPL) (Yang et al, 2020), which are recently released multi-omics integration algorithms although they are not CCA-based models.

The performances of all combinations of GE, ME, and MI in ROSMAP, such as GE+ME, GE+MI, ME+MI, and GE+ME+MI, were compared. In addition, the performance of the GE+ME+MI of KIRC was compared. We used ACC, F1 score (F1), area under the receiver operating characteristic curve (AUC), and Matthews correlation coefficient (MCC) (Elith et al, 2006) as metrics for evaluating the classification performance. The mean and standard deviation for fivefold cross-validation (CV) were calculated for all metrics.

Each CV used 60% of the samples as the training set, 20% as the validation set, and 20% as the test set. The hyperparameters of all models were selected based on the MCC of the validation set, because MCC is a more reliable statistical measure for evaluating binary classifications than ACC and F1 (Chicco et al, 2021).

For DNN, DGCCA, and SDGCCA methods, a four-layer fully connected neural network with {256, 64, 16} neurons in the hidden layers was used for GE and ME and a three-layer fully connected neural network with {64, 16} neurons in the hidden layers was used for MI. For hyperparameters in SDGCCA, learning

Table 2. Performance Comparison of Alzheimer's Disease Classification Using Gene Expression+DNA Methylation in ROSMAP Multi-Omics Data

Method	*ACC*	*F1*	*AUC*	*MCC*
SVM	0.676 ± 0.044	0.711 ± 0.036	0.751 ± 0.055	0.346 ± 0.095
XGB	0.643 ± 0.063	0.686 ± 0.059	0.697 ± 0.053	0.275 ± 0.131
LR	0.674 ± 0.067	0.674 ± 0.072	0.750 ± 0.071	0.363 ± 0.133
RF	0.602 ± 0.058	0.687 ± 0.047	0.678 ± 0.059	0.179 ± 0.134
DNN	0.697 ± 0.037	0.695 ± 0.035	0.785 ± 0.038	0.412 ± 0.079
GCCA+SVM	0.665 ± 0.054	0.699 ± 0.046	0.710 ± 0.067	0.323 ± 0.111
DGCCA+SVM	0.609 ± 0.035	0.700 ± 0.021	0.673 ± 0.072	0.194 ± 0.080
DIABLO	0.633 ± 0.059	0.637 ± 0.062	0.702 ± 0.050	0.277 ± 0.120
MOGONET	0.670 ± 0.022	0.698 ± 0.050	0.698 ± 0.042	0.332 ± 0.034
SMSPL	0.683 ± 0.071	0.723 ± 0.056	0.751 ± 0.084	0.356 ± 0.155
SDGCCA-G_{corr}	0.721 ± 0.050	0.724 ± 0.055	**0.788 ± 0.043**	0.453 ± 0.095
SDGCCA-G_{clf}	0.691 ± 0.034	0.693 ± 0.034	0.765 ± 0.046	0.396 ± 0.07
SDGCCA	**0.729 ± 0.035**	**0.728 ± 0.037**	0.782 ± 0.019	**0.474 ± 0.069**

The best performances are marked in bold.

ACC, accuracy; AUC, area under the receiver operating characteristic curve; DGCCA, deep generalized canonical correlation analysis; DIABLO, Data Integration Analysis for Biomarker discovery using Latent cOmponents; F1, F1 score; GCCA, generalized canonical correlation analysis; LR, logistic regression; MCC, Matthews correlation coefficient; MOGONET, Multi-Omics Graph cOnvolutional NETworks; RF, random forest; ROSMAP, Religious Order Study and Rush Memory and Aging Project; SDGCCA, supervised deep generalized canonical correlation analysis; SMSPL, multimodal self-paced learning with a soft weighting scheme; SVM, support vector machine; XGB, extreme gradient boosting.

rate from the set {1e−4, 1e−5}, L2 regularization term on weights from the set {0, 1e−2, 1e−4}, and the dimension of the shared representation from the set {1, 2, . . ., 10} were selected using the validation set. Details regarding the hyperparameters of the other models and fivefold CV are described in the Supplementary Data and Supplementary Figure S1.

The SDGCCA was trained using correlation and classification losses. To see how each loss affects the classification and feature selection, we performed ablation studies by measuring the performance of two additional models. First, SDGCCA-G_{corr} is a model that excludes Step 2 of Algorithm 1 in the training process. Second, SDGCCA-G_{clf} excludes Step 3 of Algorithm 1 in the training process.

4.3. Evaluation of classification performances

The results of the classification of patients with AD and normal controls are summarized in Tables 2–5. The SDGCCA showed the best performance in 10 of 16 cases, except for the performance of AUC in

Table 3. Performance Comparison of Alzheimer's Disease Classification Using Gene Expression+microRNA Expression in ROSMAP Multi-Omics Data

Method	*ACC*	*F1*	*AUC*	*MCC*
SVM	0.679 ± 0.042	0.714 ± 0.036	0.755 ± 0.054	0.351 ± 0.089
XGB	0.647 ± 0.062	0.689 ± 0.057	0.704 ± 0.057	0.283 ± 0.130
LR	0.680 ± 0.069	0.681 ± 0.070	0.758 ± 0.070	0.375 ± 0.140
RF	0.602 ± 0.054	0.683 ± 0.046	0.678 ± 0.056	0.181 ± 0.126
DNN	0.689 ± 0.048	0.695 ± 0.049	0.765 ± 0.065	0.387 ± 0.095
GCCA+SVM	0.648 ± 0.044	0.700 ± 0.044	0.693 ± 0.060	0.288 ± 0.092
DGCCA+SVM	0.633 ± 0.071	0.714 ± 0.047	0.617 ± 0.095	0.244 ± 0.154
DIABLO	0.662 ± 0.060	0.672 ± 0.063	0.736 ± 0.066	0.330 ± 0.116
MOGONET	0.696 ± 0.055	**0.722 ± 0.055**	0.759 ± 0.040	0.387 ± 0.112
SMSPL	0.691 ± 0.079	0.719 ± 0.056	0.760 ± 0.062	0.378 ± 0.169
SDGCCA-G_{corr}	0.697 ± 0.047	0.702 ± 0.053	0.757 ± 0.051	0.404 ± 0.091
SDGCCA-G_{clf}	0.667 ± 0.039	0.692 ± 0.030	0.739 ± 0.043	0.331 ± 0.082
SDGCCA	**0.699 ± 0.017**	0.697 ± 0.015	**0.796 ± 0.033**	**0.416 ± 0.035**

The best performances are marked in bold.

Table 4. Performance Comparison of Alzheimer's Disease Classification Using DNA Methylation+microRNA Expression in ROSMAP Multi-Omics Data

Method	*ACC*	*F1*	*AUC*	*MCC*
SVM	0.678 ± 0.040	0.713 ± 0.036	0.753 ± 0.052	0.349 ± 0.085
XGB	0.653 ± 0.059	0.697 ± 0.055	0.708 ± 0.054	0.296 ± 0.123
LR	0.683 ± 0.064	0.684 ± 0.065	0.758 ± 0.067	**0.380 ± 0.130**
RF	0.597 ± 0.051	0.682 ± 0.043	0.670 ± 0.058	0.169 ± 0.120
DNN	0.644 ± 0.033	0.637 ± 0.045	0.741 ± 0.031	0.305 ± 0.061
GCCA+SVM	0.631 ± 0.044	0.699 ± 0.037	0.672 ± 0.065	0.249 ± 0.096
DGCCA+SVM	0.561 ± 0.031	0.681 ± 0.027	0.548 ± 0.071	0.073 ± 0.079
DIABLO	0.686 ± 0.048	0.701 ± 0.051	0.755 ± 0.072	0.374 ± 0.095
MOGONET	0.668 ± 0.030	0.708 ± 0.040	0.708 ± 0.028	0.329 ± 0.048
SMSPL	**0.686 ± 0.032**	**0.724 ± 0.025**	0.747 ± 0.054	0.365 ± 0.068
SDGCCA-G_{corr}	0.678 ± 0.050	0.679 ± 0.066	**0.764 ± 0.052**	0.369 ± 0.093
SDGCCA-G_{clf}	0.662 ± 0.012	0.681 ± 0.021	0.733 ± 0.029	0.325 ± 0.027
SDGCCA	0.684 ± 0.046	0.693 ± 0.051	**0.764 ± 0.039**	0.372 ± 0.089

The best performances are marked in bold.

GE+ME, F1 in GE+MI, ACC, F1, MCC in ME+MI, and F1 in GE+ME+MI. In addition, for SDGCCA, integrating all three omics datasets (GE+ME+MI) outperformed that of the two omics datasets. Notably, the integration of ME and MI showed different results from combining the other two omics datasets.

For ME+MI, LR performed better than the other machine-learning models (SVM, XGB, and RF) and had the highest MCC values. In addition, SMSPL was the best-performing model for ACC and FI measurements. If we consider that LR extracts the linear relationship between multi-omics data and SMSPL as an LR-based model, the importance of nonlinearity in ME+MI is less than that in other combinations of omics data.

In all the experiments, SVM that used the original input data performed better than GCCA+SVM and DGCCA+SVM. In addition, SDGCCA performed better than GCCA+SVM and DGCCA+SVM, except for F1 in GE+MI. This result indicates a risk of losing information related to classification when dimension reduction is performed by considering only the correlation. In most cases, the performance of SDGCCA-G_{corr} was better than that of SDGCCA-G_{clf}, and the performance improved when both the correlation and classification losses were combined.

The results of the classification of early- and late-stage KIRC are shown in Table 6. The SDGCCA showed the best performance in two out of four cases, except F1 and AUC. The F1 and AUC were the highest in LR-based SMSPL, and LR also had a higher performance than other machine-learning models

Table 5. Performance Comparison of Alzheimer's Disease Classification Using Gene Expression+DNA Methylation+microRNA Expression in ROSMAP Multi-Omics Data

Method	*ACC*	*F1*	*AUC*	*MCC*
SVM	0.679 ± 0.040	0.714 ± 0.035	0.756 ± 0.050	0.352 ± 0.084
XGB	0.655 ± 0.060	0.698 ± 0.055	0.711 ± 0.055	0.299 ± 0.124
LR	0.683 ± 0.061	0.683 ± 0.063	0.759 ± 0.064	0.380 ± 0.124
RF	0.603 ± 0.050	0.684 ± 0.041	0.672 ± 0.055	0.181 ± 0.116
DNN	0.707 ± 0.039	0.701 ± 0.037	0.779 ± 0.043	0.437 ± 0.079
GCCA+SVM	0.628 ± 0.042	0.702 ± 0.033	0.669 ± 0.065	0.240 ± 0.094
DGCCA+SVM	0.569 ± 0.018	0.680 ± 0.037	0.615 ± 0.055	0.104 ± 0.034
DIABLO	0.673 ± 0.060	0.679 ± 0.064	0.739 ± 0.044	0.354 ± 0.117
MOGONET	0.684 ± 0.040	0.736 ± 0.012	0.692 ± 0.059	0.359 ± 0.086
SMSPL	0.699 ± 0.047	0.726 ± 0.027	0.777 ± 0.068	0.397 ± 0.110
SDGCCA-G_{corr}	**0.731 ± 0.035**	**0.742 ± 0.031**	0.797 ± 0.034	0.469 ± 0.075
SDGCCA-G_{clf}	0.678 ± 0.047	0.682 ± 0.050	0.753 ± 0.061	0.367 ± 0.089
SDGCCA	**0.731 ± 0.050**	0.729 ± 0.056	**0.805 ± 0.043**	**0.479 ± 0.094**

The best performances are marked in bold.

Table 6. Performance Comparison of Early- and Late-Stage Classification Using Gene Expression+DNA Methylation+microRNA Expression in Kidney Renal Clear Cell Carcinoma Multi-Omics Data

Method	*ACC*	*F1*	*AUC*	*MCC*
SVM	0.713 ± 0.040	0.708 ± 0.039	0.790 ± 0.035	0.401 ± 0.082
XGB	0.693 ± 0.055	0.688 ± 0.057	0.778 ± 0.066	0.362 ± 0.125
LR	0.738 ± 0.053	0.738 ± 0.052	0.784 ± 0.039	0.480 ± 0.106
RF	0.687 ± 0.024	0.661 ± 0.032	0.770 ± 0.031	0.340 ± 0.054
DNN	0.687 ± 0.023	0.715 ± 0.025	0.763 ± 0.054	0.418 ± 0.041
GCCA+SVM	0.652 ± 0.057	0.615 ± 0.073	0.678 ± 0.086	0.247 ± 0.159
DGCCA+SVM	0.665 ± 0.067	0.642 ± 0.081	0.684 ± 0.106	0.287 ± 0.167
DIABLO	0.719 ± 0.052	0.760 ± 0.044	0.791 ± 0.030	0.425 ± 0.117
MOGONET	0.661 ± 0.095	0.728 ± 0.087	0.745 ± 0.061	0.327 ± 0.123
SMSPL	0.710 ± 0.069	**0.763 ± 0.052**	**0.808 ± 0.067**	0.394 ± 0.151
SDGCCA-G_{corr}	0.741 ± 0.063	0.742 ± 0.062	0.800 ± 0.058	0.479 ± 0.129
SDGCCA-G_{clf}	0.735 ± 0.060	0.734 ± 0.057	0.794 ± 0.061	0.472 ± 0.122
SDGCCA	**0.745 ± 0.035**	0.745 ± 0.034	0.793 ± 0.084	**0.484 ± 0.069**

The best performances are marked in bold.

(SVM, XGB, and RF) and DNN. Consistent with the results of ROSMAP, all performances of DGCCA+SVM were better than those of GCCA+SVM, and all performances of SDGCCA-G_{corr} were better than those of SDGCCA-G_{clf}.

To statistically estimate the performance of our model against other models, we performed a paired *t*-test using fivefold CV classification results in MCC values for GE+ME+MI of ROSMAP and KIRC (Table 7). We observed that SDGCCA statistically outperformed its competing methods in 15 out of 20 cases (p-value <0.05).

We projected omics data and concatenated multi-omics data on a low-dimensional space through dimension reduction using *t*-distributed Stochastic Neighbor embedding. Figure 2 shows the projections of the multi-omics data for each method. As expected, the supervised learning-based methods clearly separated classes. Among the supervised learning-based models, nonlinear SDGCCA classifies classes more clearly than linear DIABLO.

To further demonstrate the effects of the hyperparameter k (dimension of shared representation) on the SDGCCA, we trained SDGCCA under a wide range of k using the ROSMAP data. Supplementary Figure S2 shows the embedding performance, correlation sum, and classification performance of SDGCCA when k varied from 1 to 10. We observed that the hyperparameter k did not influence the embedding and classification performance of SDGCCA as the performance fluctuated with the change in k. However, we observed that the correlation sum peaked at $k=7$ and decreased thereafter. This experiment is described in detail in the Supplementary Data.

Table 7. Statistical Significances of Performance Improvements of Supervised Deep Generalized Canonical Correlation Analysis Against Other methods

Methods	*ROSMAP*	*KIRC*
SVM	6.57E-02	**7.22E-04**
XGB	**2.82E-02**	**2.28E-02**
LR	**1.68E-03**	4.40E-01
RF	5.65E-02	**1.09E-02**
DNN	**8.39E-03**	**3.28E-02**
GCCA+SVM	9.26E-02	**4.17E-03**
DGCCA+SVM	**1.72E-02**	**1.35E-02**
DIABLO	**1.67E-03**	**2.36E-02**
MOGONET	**1.59E-02**	**3.51E-02**
SMSPL	**1.59E-02**	5.24E-02

Values with a p-value <0.05 are marked in bold.
KIRC, kidney renal clear cell carcinoma.

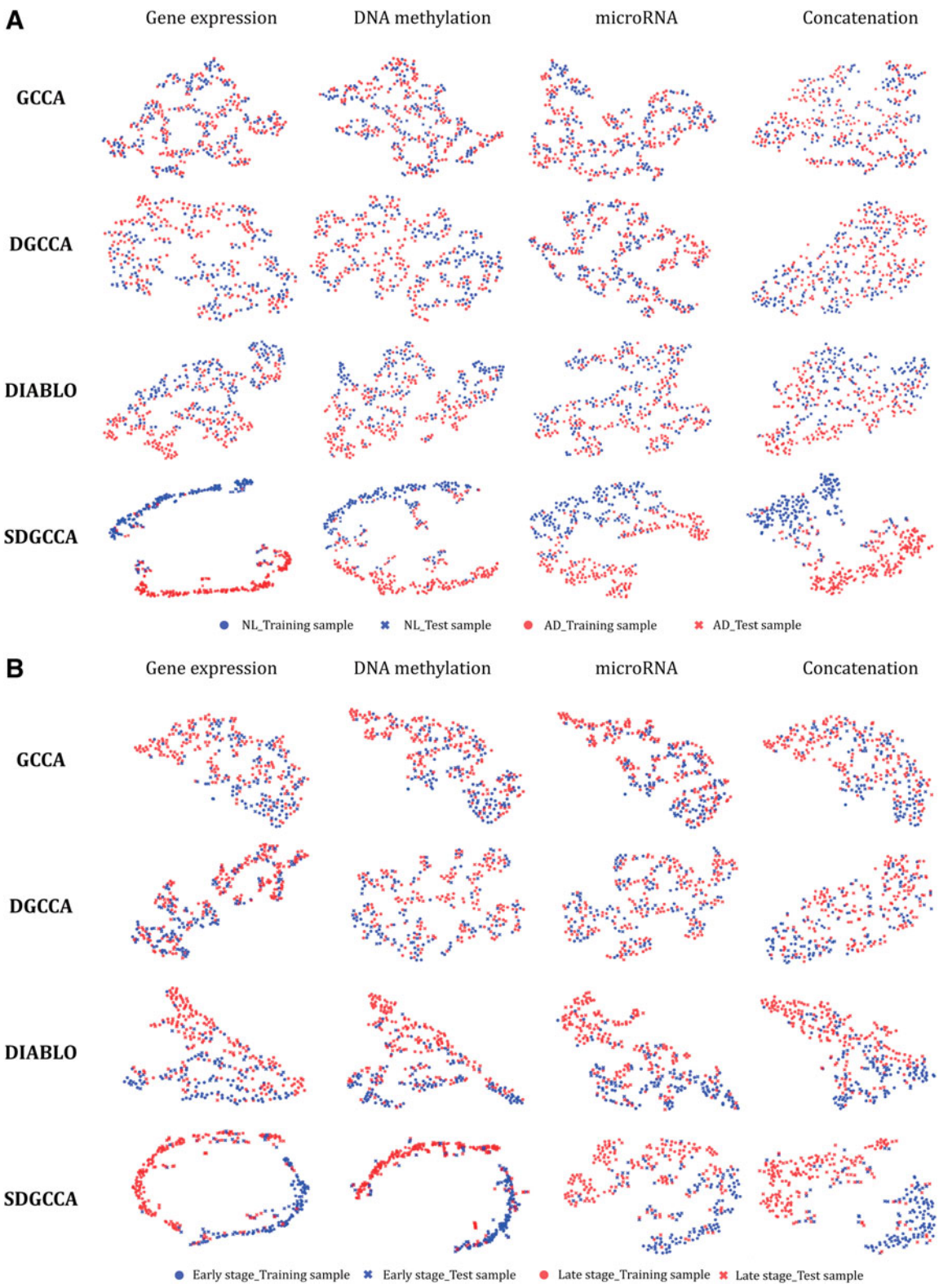

FIG. 2. *t*-SNE plots for CCA-based methods, including GCCA, DGCCA, DIABLO, and SDGCCA. **(A)** ROSMAP data and **(B)** TCGA KIRC data. Each method was used to compute projections for the GE, ME, and miRNA data. The circle and cross symbols represent training and test samples, respectively. Samples are colored according to labels. CCA, canonical correlation analysis; DGCCA, deep generalized canonical correlation analysis; DIABLO, Data Integration Analysis for Biomarker discovery using Latent cOmponents; GCCA, generalized canonical correlation analysis; GE, gene expression; KIRC, kidney renal clear cell carcinoma; ME, DNA methylation; miRNA, microRNA; ROSMAP, Religious Order Study and Rush Memory and Aging Project; *t*-SNE, *t*-distributed Stochastic Neighbor embedding; TCGA, The Cancer Genome Atlas.

4.4. Classification performance of the identified biomarkers

We compared the feature selection performance of the CCA-based method to demonstrate that the set of relevant features of DNN output with a high correlation between each modality is effective in classification. For each CV of the ROSMAP dataset, SHAP selected 300 out of 18,164 features for GE, 300 out of 19,353 features for ME, and 30 out of 309 features for MI using only the training data. Here, correlated features were selected using only the training data.

The performance was evaluated using LR, which was the best performance among the machine-learning models in the GE+ME+MI experiments. To confirm the importance of features selected by SDGCCA in the classification, LRs with these features were compared with that of randomly selected features and features selected by CCA-based models, GCCA and DGCCA. Random feature samplings were repeated 100 times, and the performances were averaged.

Table 8 presents the classification performance of the critical features selected by all competing methods in the ROSMAP data. All features and feature sets obtained from DGCCA and SDGCCA performed better than the randomly selected features, whereas the others did not. A feature set from SDGCCA showed better performance than using all features except AUC.

Thus, it can be observed that SDGCCA can identify essential features in AD classification using multi-omics data. Performance comparison experiments with additional settings, including different numbers of features and different parameter tuning approaches on the ROSMAP and KIRC datasets, were described in Supplementary Tables S1–S5 in Supplementary Data. In most cases, SDGCCA showed performance similar to that of all features.

The SDGCCA-G_{corr} and SDGCCA-G_{clf} performed lower than when randomly selected features were used. Regarding SDGCCA-G_{corr}, the gradient associated with the correlation between the shared representation and the projected representation for each modality is not propagated to the weight and bias of the DNN of each modality, indicating that the ability to select the correlative features between multi-omics is weak.

When we calculated the average of correlations between the first components of the shared representation of each modality in the training set, SDGCCA-G_{corr} had a correlation coefficient of 0.462, which is much lower than the correlation coefficient of 0.954 from SDGCCA-G_{clf} and a correlation coefficient of 0.956 from SDGCCA. Regarding SDGCCA-G_{clf}, the correlation value is slightly lower than that of SDGCCA but shows lower performance. Accordingly, L_{corr} cannot propagate sufficient information about the label to the DNN for each modality, and it is crucial to use L_{clf} together.

4.5. Pathway analysis using the SHAP values

To further illustrate the applicability of the proposed method, we performed pathway analysis. For pathway analysis, all ROSMAP samples were used for training the SDGCCA, where hyperparameters with the highest MCC values for fivefold on average in the fivefold CV were used. We clustered features with similar patterns using all samples and features with variable SHAP values (Fig. 3A). Pathway enrichment analysis was performed based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa and Goto, 2000) using the GE and ME features of each cluster.

Figure 3A illustrates enriched KEGG pathways with adjusted p-values <0.05. Cluster H was significantly enriched in the KEGG pathway related to olfactory transduction (adjusted p-values = 2E-37). Previous

Table 8. ROSMAP Classification Performance Comparison of Important Features Selected by Canonical Correlation Analysis-Based Methods

Feature set	*ACC*	*F1*	*AUC*	*MCC*
All features	0.683 ± 0.061	0.683 ± 0.063	**0.759 ± 0.064**	0.380 ± 0.124
Random features	0.661 ± 0.043	0.674 ± 0.047	0.727 ± 0.045	0.328 ± 0.086
GCCA	0.630 ± 0.043	0.638 ± 0.036	0.693 ± 0.045	0.269 ± 0.089
DGCCA	0.669 ± 0.040	0.678 ± 0.048	0.739 ± 0.037	0.345 ± 0.076
SDGCCA-G_{corr}	0.646 ± 0.045	0.661 ± 0.048	0.716 ± 0.053	0.294 ± 0.092
SDGCCA-G_{clf}	0.650 ± 0.021	0.650 ± 0.020	0.739 ± 0.040	0.315 ± 0.048
SDGCCA	**0.689 ± 0.045**	**0.698 ± 0.042**	0.755 ± 0.043	**0.386 ± 0.095**

The best performances are marked in bold.

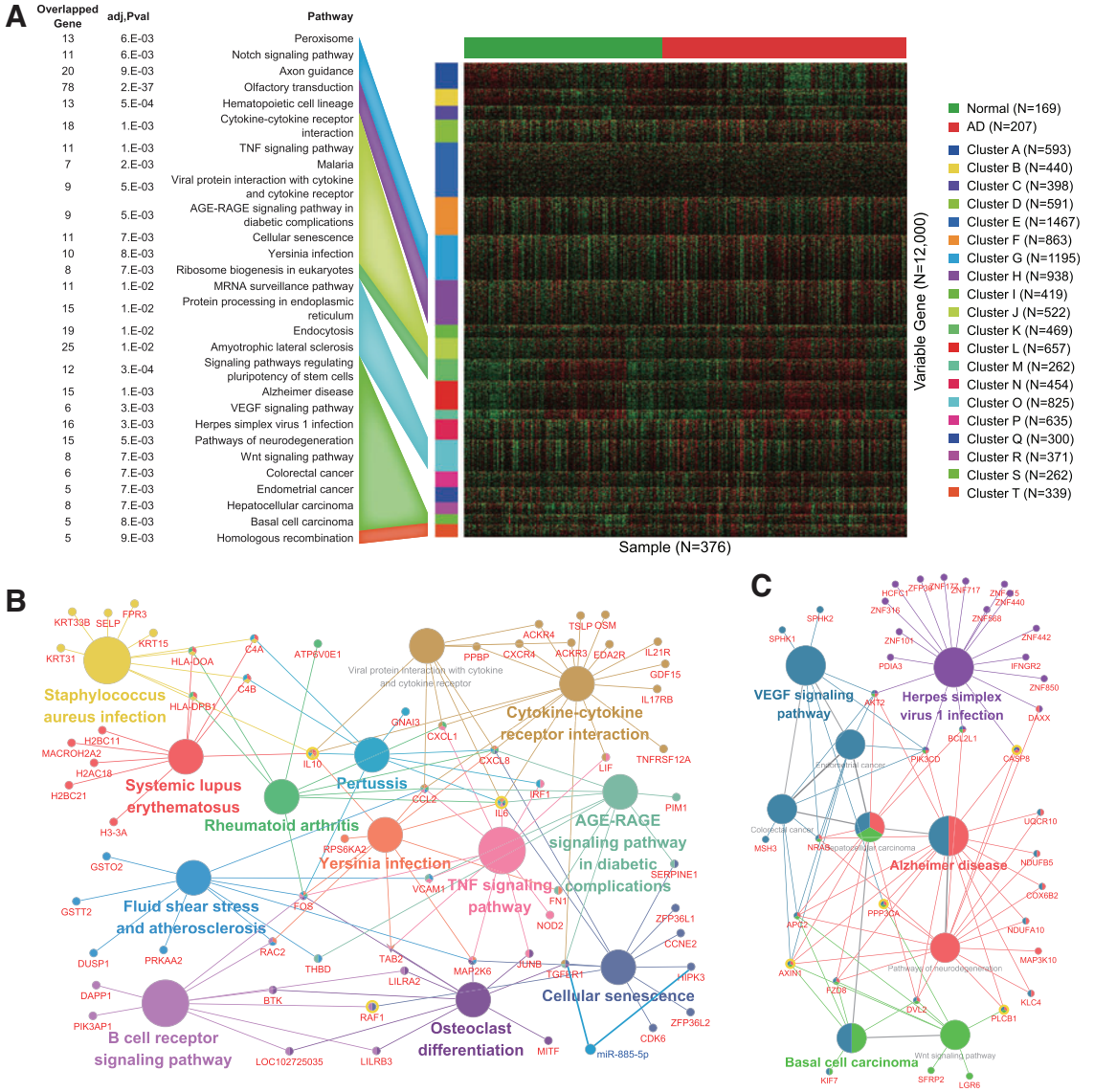

FIG. 3. The results of pathway analysis. **(A)** Clustered heatmap of SHAP value with the red color denoting increase and green color denoting decrease. The KEGG pathways enriched in each of the 20 clusters are represented with a heatmap. **(B)** The pathway network of cluster J. **(C)** The pathway network of cluster S. Yellow circles denote well-known AD-related genes. AD, Alzheimer's disease; KEGG, Kyoto Encyclopedia of Genes and Genomes; SHAP, Shapley additive explanation.

studies (Zou et al, 2016) have revealed that AD is closely related to olfactory dysfunction. We analyzed clusters J and S in detail using ClueGO (Bindea et al, 2009) Cytoscape plugins to show the relationship (Fig. 3B, C). In cluster J, we identified that two genes, *HIPK3* and *TGFBR1*, related to cellular senescence, and miR-885-5p targeting them were clustered. Cellular senescence is known to be associated with AD (Boccardi et al, 2015; Masaldan et al, 2019; Reddy et al, 2017).

In addition, interleukin (IL)6, IL10, and RAF1 in the cluster J network are also well-known AD closely related genes (Arosio et al, 2004; Mei et al, 2006). Cluster S was significantly enriched in the KEGG pathway related to AD (adjusted p-values = 1E-03) and neurodegeneration (adjusted p-values = 5E-03). *CASP8* and *PLCB1* are known AD biomarkers (Rehker et al, 2017; Shimohama and Matsushima, 1995). *AXIN1* and *PPP3CA* are AD-related genes (Lloret et al, 2011; Whelan et al, 2019).

5. CONCLUSION

In this study, we proposed a CCA-based SDGCCA, an integration method of multi-omics data, to classify and identify significant multi-omics biomarkers. The SDGCCA was trained to consider the nonlinear/complex interaction between multi-omics using the loss of DGCCA, which maximizes the correlation between the shared representation and the projected representation for each modality. In addition, the label can be predicted using a projection matrix. The model can be trained to propagate the label information to each DNN using cross-entropy.

The SDGCCA performed better in the AD prediction and early- and late-stage KIRC classification task using GE, ME, and MI than the other machine-learning models, DNN, DIABLO, MOGONET, and SMSPL. We demonstrated that SDGCCA could select a vital feature set related to a phenotype by comparing it with other feature selection models. Using SHAP values, we clustered features in multi-omics data and showed that it applies to AD-related biomarker discovery using pathway analysis.

We expect that SDGCCA can be extended to predict quantitative phenotypes by using the mean squared error loss function for regression problems instead of using the cross-entropy for the classification loss. In conclusion, SDGCCA is a multi-omics integration algorithm with high classification performance and the ability to select a set of mutually contributing features from different multi-omics datasets.

AUTHORS' CONTRIBUTIONS

H.L. initiated and supervised the study. S.M. and J.H. developed the method, investigated the experiments, and implemented software. All authors participated in writing the article. Both authors have read and approved the article.

SOFTWARE

The source codes of SDGCCA are available at https://github.com/DMCB-GIST/SDGCCA.

AUTHOR DISCLOSURE STATEMENT

The authors declare they have no conflicting financial interests.

FUNDING INFORMATION

This work was supported by the Bio and Medical Technology Development Program of NRF, funded by the Korean government (MSIT) (NRF-2018M3C7A1054935), and the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korean government (MSIT) (No.2019-0-01842, Artificial Intelligence Graduate School Program (GIST)).

ACKNOWLEDGMENTS

The authors thank Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, Illinois, for making their data available. Data collection was supported through funding by NIA (Grants P30AG10161, RO1AG15819, RO1AG17917, RO1AG30146, RO1AG36042, and RO1AG36836); the Illinois Department of Public Health; and the Translational Genomics Research Institute.

SUPPLEMENTARY MATERIAL

Supplementary Data
Supplementary Figure S1

Supplementary Figure S2
Supplementary Table S1
Supplementary Table S2
Supplementary Table S3
Supplementary Table S4
Supplementary Table S5

REFERENCES

Andrew G, Arora R, Bilmes J, et al. Deep canonical correlation analysis. In: International Conference on Machine Learning PMLR; 2013; pp. 1247–1255.

Argelaguet R, Velten B, Arnol D, et al. Multi-omics factor analysis–a framework for unsupervised integration of multi-omics data sets. Mol Syst Biol 2018;14(6):e8124; doi: 10.15252/msb.20178124.

Arosio B, Trabattoni D, Galimberti L, et al. Interleukin-10 and interleukin-6 gene polymorphisms as risk factors for Alzheimer's disease. Neurobiol Aging 2004;25(8):1009–1015; doi: 10.1016/j.neurobiolaging.2003.10.009.

Benton A, Khayrallah H, Gujral B, et al. Deep generalized canonical correlation analysis. arXiv preprint arXiv:170202519 2017.

Bindea G, Mlecnik B, Hackl H, et al. ClueGO: A cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 2009;25(8):1091–1093; doi: 10.1093/bioinformatics/btp101.

Boccardi V, Pelini L, Ercolani S, et al. From cellular senescence to Alzheimer's disease: The role of telomere shortening. Age Res Rev 2015;22:1–8; doi: 10.1016/j.arr.2015.04.003.

Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining; 2016; pp. 785–794.

Chicco D, TÃűtsch N, Jurman G. The Matthews correlation coefficient (Mcc) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. BioData Min 2021;14(1):1–22; doi: 10.1186/s13040-021-00244-z.

Couture HD, Kwitt R, Marron J, et al. Deep multi-view learning via task-optimal Cca. arXiv preprint arXiv:190707739 2019.

De Boer P-T, Kroese DP, Mannor S, et al. A tutorial on the cross-entropy method. Ann Operat Res 2005;134(1):19–67; doi: 10.1007/s10479-005-5724-z.

Elith J, Graham CH, Anderson RP, et al. Novel methods improve prediction of species' distributions from occurrence data. Ecography 2006;29(2):129–151; doi: 10.1111/j.2006.0906-7590.04596.x.

Hotelling H. Relations between two sets of variates. In: Breakthroughs in Statistics. (Kotz S, Johnson NL. eds.) Springer: New York, NY; 1992; pp. 162–190.

Huang S, Chaudhary K, Garmire LX. More is better: Recent progress in multi-omics data integration methods. Front Genet 2017;8:84; doi: 10.3389/fgene.2017.00084.

Huber W, Von Heydebreck A, Sültmann H, et al. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. Bioinformatics 2002;18(Suppl_1):S96–S104; doi: 10.1093/bioinformatics/18.suppl_1.s96.

Jendoubi T, Strimmer K. A whitening approach to probabilistic canonical correlation analysis for omics data integration. BMC Bioinform 2019;20(1):1–13; doi: 10.1186/s12859-018-2572-9.

Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics 2007;8(1):118–127; doi: 10.1093/biostatistics/kxj037.

Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res 2000;28(1):27–30; doi: 10.1093/nar/28.1.27.

Kettenring JR. Canonical analysis of several sets of variables. Biometrika 1971;58(3):433–451; doi: 10.1093/biomet/58.3.433.

Liu Y, Li Y, Yuan Y-H, et al. Supervised deep canonical correlation analysis for multiview feature learning. In: International Conference on Neural Information Processing Springer; 2017; pp. 575–582.

Lloret A, Badia M-C, Giraldo E, et al. Amyloid-β toxicity and tau hyperphosphorylation are linked via Rcan1 in Alzheimer's disease. J Alzheimers Dis 2011;27(4):701–709; doi: 10.3233/JAD-2011-110890.

Lundberg SM, Lee S-I. A Unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems; 2017; pp. 4768–4777.

Ma B, Meng F, Yan G, et al. Diagnostic classification of cancers using extreme gradient boosting algorithm and multi-omics data. Comput Biol Med 2020;121:103761; doi: 10.1016/j.compbiomed.2020.103761.

Mandal A, Maji P. FaRoC: Fast and robust supervised canonical correlation analysis for multimodal omics data. IEEE Trans Cybern 2017;48(4):1229–1241; doi: 10.1109/TCYB.2017.2685625.

Masaldan S, Belaidi AA, Ayton S, et al. Cellular senescence and iron dyshomeostasis in Alzheimer's disease. Pharmaceuticals 2019;12(2):93; doi: 10.3390/ph12020093.

Mei M, Su B, Harrison K, et al. Distribution, levels and phosphorylation of Raf-1 in Alzheimer's disease. J Neurochem 2006;99(5):1377–1388; doi: 10.1111/j.1471-4159.2006.04174.x.

Reddy PH, Williams J, Smith F, et al. MicroRNAs, aging, cellular senescence, and Alzheimer's disease. Progr Mol Biol Transl Sci 2017;146:127–171; doi: 10.1016/bs.pmbts.2016.12.009.

Rehker J, Rodhe J, Nesbitt RR, et al. Caspase-8, association with Alzheimer's disease and functional analysis of rare variants. PLoS One 2017;12(10):e0185777; doi: 10.1371/journal.pone.0185777.

Shimohama S, Matsushima H. Signal transduction mechanisms in Alzheimer disease. Alzheimer Dis Assoc Disord 1995;9(suppl 2):15–22; doi: 10.1097/00002093-199501002-00004.

Singh A, Shannon CP, Gautier B, et al. DIABLO: An integrative approach for identifying key molecular drivers from multi-omics assays. Bioinformatics 2019;35(17):3055–3062; doi: 10.1093/bioinformatics/bty1054.

Tenenhaus A, Philippe C, Guillemot V, et al. Variable selection for generalized canonical correlation analysis. Biostatistics 2014;15(3):569–583; doi: 10.1093/biostatistics/kxu001.

Wang T, Shao W, Huang Z, et al. MOGONET integrates multi-omics data using graph convolutional networks allowing patient classification and biomarker identification. Nat Commun 2021;12(1):1–13; doi: 10.1038/s41467-021-23774-w.

Whelan CD, Mattsson N, Nagle MW, et al. Multiplex proteomics identifies novel Csf and plasma biomarkers of early Alzheimer's disease. Acta Neuropathol Commun 2019;7(1):1–14; doi: 10.1186/s40478-019-0795-2.

Yang Z, Wu N, Liang Y, et al. SMSPL: Robust multimodal approach to integrative analysis of multiomics data. IEEE Trans Cybern 2020;1–14; doi: 10.1109/TCYB.2020.3006240.

Zou Y-m, Da Lu L-pL, Zhang H-h, et al. Olfactory dysfunction in Alzheimer's disease. Neuropsych Dis Treat 2016;12:869; doi: 10.2147/NDT.S104886.

Address correspondence to:
Dr. Hyunju Lee
School of Electrical Engineering and Computer Science
Gwangju Institute of Science and Technology
Gwangju 61005
South Korea

E-mail: hyunjulee@gist.ac.kr