

A New Method to Address Singularity Problem in Multimodal Data Analysis

Ankita Mandal^{} and Pradipta Maji^{}

Biomedical Imaging and Bioinformatics Lab, Machine Intelligence Unit,
Indian Statistical Institute, Kolkata, India
{amandal, pmaji}@isical.ac.in

Abstract. In general, the ‘small sample (n)-large feature (p)’ problem of bioinformatics, image analysis, high throughput molecular screening, astronomy, and other high dimensional applications makes the features highly collinear. In this context, the paper presents a new feature extraction algorithm to address this ‘large p small n ’ issue associated with multimodal data sets. The proposed algorithm judiciously integrates the concept of both regularization and shrinkage with canonical correlation analysis to extract important features. To deal with the singularity problem, the proposed method increases the diagonal elements of covariance matrices by using regularization parameters, while the off-diagonal elements are decreased by shrinkage coefficients. The concept of hypercuboid equivalence partition matrix of rough hypercuboid approach is used to compute both significance and relevance measures of a feature. The importance of the proposed algorithm over other existing methods is established extensively on real life multimodal omics data set.

1 Introduction

Unimodal based pattern recognition and analysis systems usually provide low level of performance due to the noisy nature and drastic variation of the acquired signals, which lead to inaccurate and insufficient pattern representation of the perception of interest. On the other hand, multimodal data contains more information, which is expected to provide potentially more discriminatory and complete description of the intrinsic characteristics of the pattern, which leads to improve system performance than single modality only [5].

Canonical correlation analysis (CCA) [4] finds the best linear transformation to achieve the maximum correlation between two multidimensional data sets. The modern technology has enabled more directions on data streams, which ensues in very high dimensional feature spaces (p), while the number of training samples (n) is usually limited. When the number of samples (n) is very less than the number of features (p), the features in both data sets tend to be highly collinear, which leads to ill-conditioned of the covariance matrices of the

This work is partially supported by the Department of Electronics and Information Technology, Government of India PhD-MLA/4(90)/2015-16).

data sets. In effect, their inverses are no longer reliable, resulting in an invalid computation of CCA. There are two ways to overcome this problem. The first possible approach is regularized CCA (RCCA) [11], where the diagonal elements of covariance matrices are increased using a grid search optimization. However, the off-diagonal elements of these matrices remain constant. The another method of regularization algorithm is based on the optimal estimate of the correlation matrices and is known as fast RCCA (FRCCA) [1]. In FRCCA, shrinkage coefficients [10] are estimated to invert the covariance matrices. These shrinkage coefficients reduce the values of off-diagonal elements of covariance matrices, while the values of diagonal elements remain same. However, CCA, RCCA and FRCCA all are unsupervised in nature and fail to take complete advantage of available class label information [1, 2]. To incorporate the class information, some supervised versions of RCCA have been proposed, termed as supervised RCCA (SRCCA) [2]. It includes available class label information to select maximally correlated features using grid search optimization.

One of the main problems in omics data analysis is uncertainty. Rough set theory [9] is an effective paradigm to deal with uncertainty, vagueness, and incompleteness. It provides a mathematical framework to capture uncertainties associated with the data [9]. In this context, a feature extraction algorithm, termed as CuRSaR [7], has been introduced. It judiciously integrates the merits of SRCCA and rough sets, to extract maximally correlated features from two multidimensional data sets. In [8], another method, named as FaRoC, has been proposed to generate canonical variables sequentially using rough hypercuboid based maximum relevance-maximum significance criterion. However, all these existing methods fail to produce the optimal set of features.

In general, RCCA increases the diagonal elements, whereas FRCCA decreases the off-diagonal elements to deal with the singularity issue of covariance matrices. So, it is expected to give better results if both can be done concurrently. In this regard, the paper presents a new feature extraction algorithm, which integrates the advantages of both RCCA and FRCCA to handle the ill-conditioned of the covariance matrices. The effectiveness of the proposed method, along with a comparison with other methods, is demonstrated on several real life data sets.

2 Basics of Canonical Correlation Analysis

Canonical correlation analysis (CCA) [4] obtains a linear relationship between two multidimensional variables. The objective of CCA is to extract latent features from two data sets $\mathcal{X} \in \mathbb{R}^{p \times n}$ and $\mathcal{Y} \in \mathbb{R}^{q \times n}$. Here p and q are the number of features of \mathcal{X} and \mathcal{Y} , respectively, whereas n is the number of samples. CCA obtains two directional basis vectors $w_x \in \mathbb{R}^p$ and $w_y \in \mathbb{R}^q$ such that, the correlation between $\mathcal{X}^T w_x$ and $\mathcal{Y}^T w_y$ is maximum. The correlation coefficient ρ is given as

$$\rho = \max_{w_x, w_y} \frac{w_x^T C_{xy} w_y}{\sqrt{w_x^T C_{xx} w_x w_y^T C_{yy} w_y}} \quad (1)$$

where $C_{xx} \in \mathbb{R}^{p \times p}$ and $C_{yy} \in \mathbb{R}^{q \times q}$ are covariance matrices of X and \mathcal{Y} , respectively, while $C_{xy} \in \mathbb{R}^{p \times q}$ is the cross-covariance matrix of X and \mathcal{Y} . The basis vectors w_x and w_y are the eigenvectors of matrices \mathcal{H} and $\tilde{\mathcal{H}}$, respectively, with eigenvalue ρ , where

$$\mathcal{H} = C_{xx}^{-1} C_{xy} C_{yy}^{-1} C_{yx}; \quad \text{and} \quad \tilde{\mathcal{H}} = C_{yy}^{-1} C_{yx} C_{xx}^{-1} C_{xy}. \quad (2)$$

If the number of features p and q of X and \mathcal{Y} , respectively, is larger than n , the covariance matrices C_{xx} and C_{yy} are ill-conditioned, which make the computation of CCA invalid. That means, the inverses of C_{xx} and C_{yy} do not make any sense [3]. To overcome this problem, RCCA [11] increases the diagonal elements of C_{xx} and C_{yy} by adding small positive quantities, \mathbf{r}_x and \mathbf{r}_y , which are known as regularization parameters. The optimal parameter set of \mathbf{r}_x and \mathbf{r}_y is selected for which the Pearson's correlation is maximum. On the other hand, FRCCA [1] decreases the off-diagonal elements of C_{xx} and C_{yy} by subtracting the shrinkage parameters s_x and s_y . To find the minimum mean squared error estimator of cross-covariance matrix C_{xy} , the shrinkage parameter s_{xy} is used.

3 Proposed Method

This section presents a new feature extraction algorithm, integrating judiciously the advantages of both RCCA and FRCCA to take care of the singularity problem of covariance matrices. The proposed method also incorporates the available class label information to make it supervised. It extracts new features from two multidimensional data sets by maximizing their relevance with respect to class label and significance with respect to already-extracted features. Prior to describing the proposed method for multimodal data analysis, some important analytical formulations are reported next.

To deal with this singularity problem, the proposed method integrates the advantages of both RCCA and FRCCA. Here, regularization parameters \mathbf{r}_x and \mathbf{r}_y are varied within a range $[\mathbf{r}_{min}, \mathbf{r}_{max}]$, with common differences, d_x and d_y for \mathbf{r}_x and \mathbf{r}_y , respectively. To address this singularity issue, the covariance and cross-covariance matrices can be formulated as

$$[\tilde{C}_{xx}]_{ij} = \begin{cases} (1 - s_x)[C_{xx}]_{ij}; & \text{where } i \neq j \\ [C_{xx}]_{ij} + (\mathbf{r}_x + l d_x); & \text{where } i = j \end{cases}$$

and $[\tilde{C}_{xy}]_{ij} = (1 - s_{xy})[C_{xy}]_{ij}; \quad \forall i, j \quad (3)$

where $\forall l \in \{1, 2, \dots, \mathbf{t}_x\}$. Similarly, $[\tilde{C}_{yy}]_{ij}$ can be computed $\forall l \in \{1, 2, \dots, \mathbf{t}_y\}$. The parameters \mathbf{t}_x and \mathbf{t}_y denote the number of possible values of \mathbf{r}_x and \mathbf{r}_y , respectively. The best estimator of the shrinkage parameters s_x , s_y and s_{xy} , which minimize the risk function of the mean squared error, can be calculated as [1]

$$s_x = \frac{\sum_{i \neq j} \hat{\mathcal{V}}([C_{xx}]_{ij})}{\sum_{i \neq j} [C_{xx}^2]_{ij}}; \quad s_y = \frac{\sum_{i \neq j} \hat{\mathcal{V}}([C_{yy}]_{ij})}{\sum_{i \neq j} [C_{yy}^2]_{ij}}; \quad \text{and} \quad s_{xy} = \frac{\sum_i \sum_j \hat{\mathcal{V}}([C_{xy}]_{ij})}{\sum_i \sum_j [C_{xy}^2]_{ij}}; \quad (4)$$

where $\hat{\mathcal{V}}([C_{xx}]_{ij})$, $\hat{\mathcal{V}}([C_{yy}]_{ij})$ and $\hat{\mathcal{V}}([C_{xy}]_{ij})$ are the unbiased empirical variance of $[C_{xx}]_{ij}$, $[C_{yy}]_{ij}$ and $[C_{xy}]_{ij}$, respectively. Let us assume that Λ_x and Λ_y be the diagonal matrices, where diagonal elements are the eigenvalues of \tilde{C}_{xx} and \tilde{C}_{yy} , respectively and the corresponding orthonormal eigenvectors are in the columns of Ψ_x and Ψ_y , respectively. If τ_x and τ_y are varied within a range with arithmetic progression, then following two relations can be established, based on the theoretical analysis, reported in [8],

$$\begin{aligned} \mathcal{H}_{\kappa\ell} &= \Psi_x(\Lambda_x + (\kappa - 1)d_x I)^{-1} \Psi_x^T \tilde{C}_{xy} \Psi_y(\Lambda_y + (\ell - 1)d_y I)^{-1} \Psi_y^T \tilde{C}_{yx}; \\ \tilde{\mathcal{H}}_{\kappa\ell} &= \Psi_y(\Lambda_y + (\ell - 1)d_y I)^{-1} \Psi_y^T \tilde{C}_{yx} \Psi_x(\Lambda_x + (\kappa - 1)d_x I)^{-1} \Psi_x^T \tilde{C}_{xy}. \end{aligned} \quad (5)$$

As non-zero eigenvalues of $\mathcal{H}_{\kappa\ell}$ are same as non-zero eigenvalues of $\tilde{\mathcal{H}}_{\kappa\ell}$, one of the matrices is enough to compute the eigenvector of $\mathcal{H}_{\kappa\ell}$ or $\tilde{\mathcal{H}}_{\kappa\ell}$ [7], which are the basis vectors $w_{x\kappa\ell}$ and $w_{y\ell\kappa}$, respectively.

To compute both the relevance and significance of an extracted feature, the hypercuboid equivalence partition matrix of rough hypercuboid approach [6] is used. The regularization parameters are optimized through computing the relevance and significance measures [7]. Hence, the problem of extracting a relevant and significant feature set \mathbb{S} from all possible combinations of regularization parameters τ_x and τ_y is equivalent to maximize the average relevance of all extracted features as well as to maximize the average significance among them. To solve this problem, the following greedy algorithm is used.

1. Compute two covariance matrices C_{xx} and C_{yy} , of X and \mathcal{Y} , respectively.
2. Compute the cross-covariance matrix C_{xy} , of X and \mathcal{Y} .
3. Determine the values of s_x , s_y and s_{xy} using (4).
4. Compute \tilde{C}_{xx} , \tilde{C}_{yy} and \tilde{C}_{xy} using (3).
5. Calculate eigenvalues $\Lambda_x \in \mathbb{R}^p$ and $\Lambda_y \in \mathbb{R}^q$ of \tilde{C}_{xx} and \tilde{C}_{yy} , respectively, along with corresponding eigenvectors Ψ_x and Ψ_y .
6. Repeat the following six steps for all (κ, ℓ) -th regularization parameters of τ_x and τ_y , where $\forall \kappa \in \{1, 2, \dots, \mathbf{t}_x\}$ and $\forall \ell \in \{1, 2, \dots, \mathbf{t}_y\}$.
 - (i) Initialize $\mathbb{C}_{\kappa\ell} \leftarrow \emptyset$.
 - (ii) Compute $\mathcal{H}_{\kappa\ell}$ or $\tilde{\mathcal{H}}_{\kappa\ell}$ using (5).
 - (iii) Calculate all \mathcal{D} basis vectors $w_{x\kappa\ell}$ and $w_{y\ell\kappa}$, which are the eigenvectors of $\mathcal{H}_{\kappa\ell}$ and $\tilde{\mathcal{H}}_{\kappa\ell}$, respectively, where $\mathcal{D} = \min(p, q)$

$$\mathcal{H}_{\kappa\ell} w_{x\kappa\ell} = \rho w_{x\kappa\ell}; \quad \text{and} \quad \tilde{\mathcal{H}}_{\kappa\ell} w_{y\ell\kappa} = \rho w_{y\ell\kappa}. \quad (6)$$

- (iv) Calculate the \mathcal{D} canonical variables $\mathcal{U}_{\kappa\ell}$ and $\mathcal{V}_{\kappa\ell}$,

$$\mathcal{U}_{\kappa\ell} = w_{x\kappa\ell}^T X; \quad \text{and} \quad \mathcal{V}_{\kappa\ell} = w_{y\ell\kappa}^T \mathcal{Y}. \quad (7)$$

- (v) The extracted feature $\mathcal{A}_{k\ell}$ can be calculated using $\mathcal{A}_{k\ell} = \mathcal{U}_{k\ell} + \mathcal{V}_{k\ell}$.
 - vi) $\mathbb{C}_{k\ell} = \mathbb{C}_{k\ell} \cup \mathcal{A}_{k\ell}$.
7. Initialize $\mathbb{S} \leftarrow \emptyset$.
 8. Repeat the following three steps until $t \leq \mathcal{D}$, where $\mathcal{D} = \min(p, q)$.
 - (i) If $t = 1$, calculate the relevance $\gamma_{\mathcal{A}_{k\ell}}(\mathbb{D})$, otherwise calculate the significance $\sigma_{\{\mathcal{A}_{k\ell}, \mathcal{A}_i\}}(\mathbb{D}, \mathcal{A}_{k\ell})$ where $\mathcal{A}_i \in \mathbb{S}$, of all t -th extracted feature $\mathcal{A}_{k\ell} \in \mathbb{C}_{k\ell}$, $\forall k \in \{1, 2, \dots, t_x\}$ and $\forall \ell \in \{1, 2, \dots, t_y\}$. \mathbb{D} denotes the decision attribute set. Discard that $\mathcal{A}_{k\ell}$, if it has zero significance with at least one of the selected features of \mathbb{S} .
 - (ii) If $t = 1$, select a feature $\mathcal{A}_{k\ell}$ as t -th feature for which $\gamma_{\mathcal{A}_{k\ell}}(\mathbb{D})$ is maximum. Otherwise, the feature $\mathcal{A}_{k\ell}$ has to be selected as optimal for which $\gamma_{\mathcal{A}_{k\ell}}(\mathbb{D}) + \frac{1}{t-1} \sum_{\mathcal{A}_i \in \mathbb{S}} \sigma_{\{\mathcal{A}_{k\ell}, \mathcal{A}_i\}}(\mathbb{D}, \mathcal{A}_{k\ell})$ is maximum.
 - (iii) $\mathbb{S} = \mathbb{S} \cup \mathcal{A}_{k\ell}$ and $t = t + 1$.
 9. Stop.

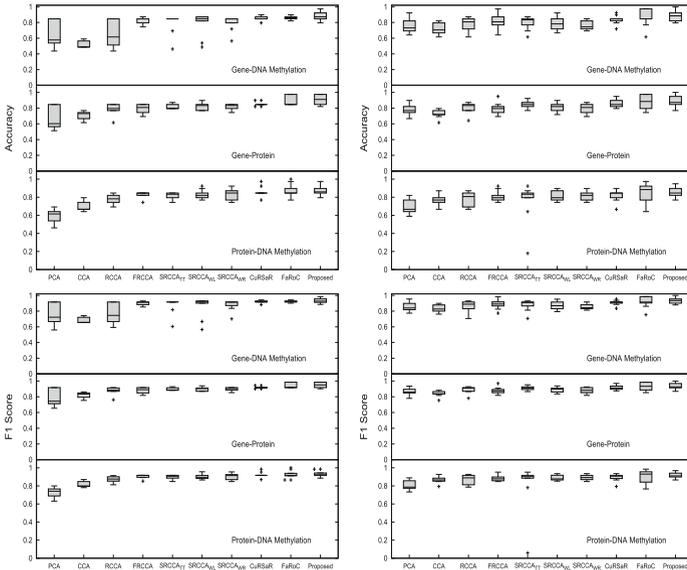
4 Experimental Results and Discussion

In the current research work, the multimodal data set, named Ovarian Serous Cystadenocarcinoma (OV), is used with three different modalities, namely, gene expression, protein expression, and DNA methylation. The data set is downloaded from TCGA. It has 379 samples with two categories: 51 samples of grade 2 and 328 samples of grade 3. This data set contains expressions of 17,814 genes and 222 proteins, and β values of 27,578 methylated DNAs. Total 2000 top-ranked features, based on their variances, are taken from both gene and methylation data in the current study. To evaluate the performance of different algorithms, both support vector machine (SVM) and nearest neighbor algorithm (NNA) are used. To compute the classification accuracy and F1 score of different approaches, 10-fold cross-validation is performed. A set of correlated features is first generated for each training set. Then, both SVM and NNA are trained with this feature set. The correlated features which are selected for the training set are used to generate test set. Finally, the class label of the test sample is predicted using the SVM and NNA. Twenty five top-ranked correlated features are selected for the analysis.

In the proposed method, both relevance and significance of an extracted feature are computed using rough hypercuboid (RH) approach [6]. However, mutual information (MI) as well as rough sets (RS) with fuzzy discretization can also be used to compute these two measures. In order to establish the importance of rough hypercuboid (RH) approach over rough sets (RS) and mutual information (MI), extensive experimental results are reported in Table 1 for three pairs of modalities. All the results reported in Table 1 establish the fact that the rough hypercuboid equivalence partition matrix based approach performs better than other approaches, irrespective of the pair of modalities and classifier used. In 12 cases, out of total 24 cases, rough hypercuboid equivalence partition matrix based approach performs significantly better and in remaining 12 cases

Table 1. Classification accuracy and F1 score of the proposed algorithm

Measures	Classifier used	Different indices	Gene-Protein			Protein-DNA Methy.			Gene-DNA Methy.		
			Mean	StDv	p-value	Mean	StDv	p-value	Mean	StDv	p-value
Accuracy	SVM	MI	0.859	0.028	1.76E-02	0.864	0.030	1.85E-01	0.851	0.047	8.88E-02
		RS	0.854	0.036	3.31E-02	0.831	0.028	1.12E-02	0.844	0.025	7.65E-02
		RH	0.908	0.071		0.879	0.058		0.879	0.060	
	NNA	MI	0.838	0.068	4.50E-02	0.828	0.078	8.10E-02	0.756	0.084	5.55E-04
		RS	0.849	0.072	5.77E-02	0.838	0.077	1.61E-01	0.846	0.077	3.84E-02
		RH	0.882	0.069		0.856	0.054		0.879	0.066	
F1 Score	SVM	MI	0.923	0.014	1.79E-02	0.925	0.016	1.79E-01	0.917	0.027	8.94E-02
		RS	0.919	0.020	2.75E-02	0.907	0.017	1.11E-02	0.914	0.014	1.01E-01
		RH	0.949	0.038		0.934	0.031		0.932	0.034	
	NNA	MI	0.905	0.040	3.63E-02	0.897	0.047	5.48E-02	0.851	0.051	5.08E-04
		RS	0.913	0.041	5.24E-02	0.906	0.045	1.39E-01	0.910	0.046	3.59E-02
		RH	0.933	0.038		0.918	0.031		0.931	0.037	

**Fig. 1.** Box and whisker plots for accuracy and F1 score (left: SVM; right: NNA)

it is better but not significant. This analysis establishes the importance of rough hypercuboid approach over other methods in proposed method.

The performance of the proposed feature extraction algorithm is compared in this section with that of some existing CCA based algorithms, namely, principal component analysis (PCA), CCA, RCCA, FRCCA, several variants of SRCCA using t -test (SRCCA_{TT}) [2], Wilcoxon rank sum test (SRCCA_{WR}) [2], and Wilks's lambda test (SRCCA_{WL}) [2], CuRSaR [7] and FaRoC [8]. The regularization parameters τ_x and τ_y are varied within $[0.0, 1.0]$ with 0.1 as common difference.

The box and whisker plots, tables of means, standard deviations, and p-value computed through paired- t (one-tailed) test are used to study the performance of different algorithms and the proposed algorithm. Fig. 1 shows the box and whisker plots for classification accuracy and F1 score. On the other hand, the means, standard deviations and p-values of accuracy and F1 score for all the methods are reported in Tables 2 and 3. The best mean values are marked in bold in these tables. The experimental results are presented on three pairs of modalities, namely, gene-protein, gene-DNA methylation, and protein-DNA methylation. All the results, presented in Fig. 1 and Tables 2 and 3, establish the fact that the proposed method attains the best mean classification accuracy and F1 score in all the cases, irrespective of the pairs of modalities, and classifiers used. The results, reported in Fig. 1 and Tables 2 and 3, demonstrate that the proposed algorithm performs significantly better than other algorithms in 83 cases out of total 108 cases, considering 0.05 as the level of significance. In remaining 25 cases, it is better but not significant.

Table 2. Classification accuracy of the proposed and other methods

Classifier used	Different algorithms	Gene-Protein			Protein-DNA Methy.			Gene-DNA Methy.		
		Mean	StDv	p-value	Mean	StDv	p-value	Mean	StDv	p-value
SVM	PCA	0.677	0.149	7.85E-04	0.595	0.080	7.32E-09	0.654	0.171	9.04E-04
	CCA	0.715	0.050	2.06E-05	0.695	0.055	2.50E-05	0.515	0.046	6.72E-09
	RCCA	0.787	0.068	6.00E-03	0.774	0.052	3.37E-04	0.667	0.164	6.39E-04
	FRCCA	0.792	0.059	2.32E-03	0.828	0.032	4.23E-03	0.823	0.044	2.31E-02
	SRCCA _{TT}	0.821	0.034	7.93E-04	0.813	0.042	3.73E-04	0.792	0.126	3.13E-02
	SRCCA _{WL}	0.823	0.046	4.36E-03	0.828	0.051	1.48E-02	0.787	0.146	4.56E-02
	SRCCA _{WR}	0.813	0.042	9.70E-04	0.828	0.062	1.37E-02	0.797	0.092	4.50E-02
	CuRSaR	0.851	0.026	2.10E-02	0.859	0.054	2.09E-01	0.851	0.026	1.35E-01
	FaRoC	0.897	0.066	2.31E-01	0.867	0.075	3.37E-01	0.859	0.028	1.76E-01
Proposed	0.908	0.071		0.879	0.058		0.879	0.060		
NNA	PCA	0.677	0.149	9.99E-04	0.692	0.075	2.18E-05	0.759	0.090	5.62E-04
	CCA	0.726	0.050	5.48E-05	0.772	0.053	1.84E-04	0.726	0.072	5.21E-06
	RCCA	0.808	0.068	2.59E-02	0.782	0.083	2.72E-02	0.782	0.083	9.77E-03
	FRCCA	0.797	0.070	8.66E-05	0.805	0.061	1.08E-03	0.808	0.092	3.42E-04
	SRCCA _{TT}	0.841	0.050	3.16E-03	0.759	0.216	1.16E-01	0.797	0.082	2.18E-02
	SRCCA _{WL}	0.810	0.065	7.98E-04	0.813	0.051	9.60E-03	0.790	0.091	2.31E-04
	SRCCA _{WR}	0.797	0.060	4.36E-03	0.818	0.052	7.48E-03	0.762	0.057	2.06E-04
	CuRSaR	0.854	0.050	1.46E-01	0.821	0.065	6.88E-02	0.838	0.055	3.06E-02
	FaRoC	0.877	0.093	4.41E-01	0.849	0.113	4.12E-01	0.864	0.113	3.68E-01
Proposed	0.882	0.069		0.856	0.054		0.879	0.066		

Table 3. F1 score of the proposed and other methods

Classifier used	Different algorithms	Gene-Protein			Protein-DNA Methy.			Gene-DNA Methy.		
		Mean	StDv	p-value	Mean	StDv	p-value	Mean	StDv	p-value
SVM	PCA	0.793	0.110	1.11E-03	0.731	0.057	1.64E-08	0.765	0.139	1.67E-03
	CCA	0.823	0.035	9.40E-06	0.814	0.034	1.02E-05	0.679	0.039	5.41E-09
	RCCA	0.878	0.045	6.01E-03	0.871	0.034	3.07E-04	0.776	0.130	1.24E-03
	FRCCA	0.882	0.038	2.19E-03	0.904	0.021	2.96E-03	0.901	0.026	2.58E-02
	SRCCA _{TT}	0.898	0.019	5.24E-04	0.895	0.027	2.08E-04	0.876	0.100	5.47E-02
	SRCCA _{WL}	0.899	0.025	2.62E-03	0.902	0.029	9.43E-03	0.860	0.131	6.21E-02
	SRCCA _{WR}	0.895	0.025	8.16E-04	0.901	0.037	8.63E-03	0.881	0.069	5.90E-02
	CuRSaR	0.918	0.014	2.01E-02	0.923	0.030	2.10E-01	0.918	0.016	1.71E-01
	FaRoC	0.944	0.035	2.39E-01	0.925	0.043	2.96E-01	0.922	0.015	2.01E-01
Proposed	0.949	0.038		0.934	0.031		0.932	0.034		
NNA	PCA	0.793	0.110	9.35E-04	0.806	0.052	2.28E-05	0.853	0.057	5.00E-04
	CCA	0.838	0.035	5.23E-05	0.867	0.035	1.82E-04	0.837	0.046	7.90E-06
	RCCA	0.891	0.044	3.31E-02	0.870	0.056	2.71E-02	0.867	0.069	1.68E-02
	FRCCA	0.879	0.040	3.95E-05	0.885	0.035	1.35E-03	0.884	0.057	4.16E-04
	SRCCA _{TT}	0.908	0.027	1.43E-03	0.814	0.269	1.38E-01	0.878	0.070	3.66E-02
	SRCCA _{WL}	0.887	0.037	4.73E-04	0.889	0.028	5.67E-03	0.874	0.056	1.57E-04
	SRCCA _{WR}	0.881	0.036	3.54E-03	0.893	0.030	7.09E-03	0.856	0.037	2.08E-04
	CuRSaR	0.913	0.029	9.62E-02	0.892	0.040	5.58E-02	0.908	0.031	3.98E-02
	FaRoC	0.930	0.052	4.40E-01	0.904	0.077	2.79E-01	0.918	0.072	3.22E-01
Proposed	0.933	0.038		0.918	0.031		0.931	0.037		

5 Conclusion

In present days, the ‘large p small n ’ problem becomes a common issue in genetics research, medical studies, risk management, and other fields. If n is very small compared to p , the features become highly collinear, which leads to ill-conditioned of the covariance matrix. The current research work deals with this ‘small n large p ’ problem to overcome the singularity issue of the covariance matrices. The effectiveness of the proposed algorithm, along with a comparison with other algorithms, has been demonstrated considering three different modalities, namely, gene expression, protein expression, and DNA methylation. The concept of hypercuboid equivalence partition matrix is found to be successful in extracting relevant and significant features from high dimensional multimodal real-life data sets.

References

1. Cruz-Cano, R., Lee, M.T.: Fast regularized canonical correlation analysis. *Comput. Stat. Data Anal.* **70**, 88–100 (2014)
2. Golugula, A., Lee, G., Master, S.R., Feldman, M.D., Tomaszewski, J.E., Speicher, D.W., Madabhushi, A.: Supervised regularized canonical correlation analysis: integrating histologic and proteomic measurements for predicting biochemical recurrence following prostate surgery. *BMC Bioinform.* **12**, 483 (2011)
3. Gou, Z., Fyfe, C.: A canonical correlation neural network for multicollinearity and functional data. *Neural Netw.* **17**(2), 285–293 (2004)

4. Hotelling, H.: Relations between two sets of variates. *Biometrika* **28**(3/4), 321–377 (1936)
5. Lanckriet, G.R.G., Bie, T.D., Cristianini, N., Jordan, M.I., Noble, W.S.: A statistical framework for genomic data fusion. *Bioinformatics* **20**(16), 2626–2635 (2004)
6. Maji, P.: A rough hypercuboid approach for feature selection in approximation spaces. *IEEE Trans. Knowl. Data Eng.* **26**(1), 16–29 (2014)
7. Maji, P., Mandal, A.: Multimodal omics data integration using max relevance-max significance criterion. *IEEE Trans. Biomed. Eng.* (2016). doi:[10.1109/TBME.2016.2624823](https://doi.org/10.1109/TBME.2016.2624823)
8. Mandal, A., Maji, P.: FaRoC: fast and robust supervised canonical correlation analysis for multimodal omics data. *IEEE Trans. Cybern.* (2017). doi:[10.1109/TCYB.2017.2685625](https://doi.org/10.1109/TCYB.2017.2685625)
9. Pawlak, Z.: *Rough Sets: Theoretical Aspects of Reasoning about Data*. Kluwer Academic Publishers, Dordrecht, Boston and London (1991)
10. Schafer, J., Strimmer, K.: A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Stat. Appl. Genet. Mol. Biol.* **4**(1), 1–32 (2005)
11. Vinod, H.D.: Canonical ridge and econometrics of joint production. *J. Econometrics* **4**(2), 147–166 (1976)