MULTIMODAL IMAGING FEATURE EXTRACTION WITH REFERENCE CANONICAL CORRELATION ANALYSIS UNDERLYING INTELLIGENCE

Ram Sapkota [1,3], Bishal Thapaliya[1,3], Pranav Suresh[1,3], Bhaskar Ray[1,3], Vince Calhoun[1,2,3], Jingyu Liu[1,3]
[1]Georgia State University, [2] Georgia Institute of Technology, [3]TReNDs Center

TReNDS Center for Translational Research in Neuroimaging & Data Science

Georgia State University

INTRODUCTON

During adolescence, rapid brain development is accompanied by cognitive maturation and behavioral changes. In this study, we investigated brain gray matter and white matter images from a large developmental cohort: the Adolescent Brain Cognitive Development (ABCD) Study, in hope to unveil brain characteristics underlying intelligence. Specifically, we developed reference Canonical Correlation Analysis (RCCA) which finds the correlated latent features between gray matter and white matter, while incorporating a reference carrying intelligence information. We systemically compared feature extraction by Principle Component Analysis (PCA), Sparse CCA (SCCA), and our developed RCCA.

PARTICIPANTS AND DATA

We have utilized baseline data from two image modalities: Gray matter density (GM) and White matter fractional anisotropy (FA) images.
Participants: 7,874 (Male: 4,100 and Female: 3,774)
Cognitive Assessments: Fluid intelligence, Crystallized intelligence, and Total composite scores.
Gray Matter Image Preprocessing (T1-weighted MRI): Preprocessing using SPM12 toolbox, smoothed by a 6 mm Gaussian kernel, registered in the MNI space, masked by voxels with mean GM density >0.2.
White Matter Image Preprocessing (Diffusion MRI): Preprocessed using FSL (v6.0.1) tools, and diffusion tensor calculation for FA maps, registered to MNI space using nonlinear ANTs function, masked by voxels with mean FA >0.2.

METHODS

PCA Feature Extraction: PCA was applied to the GM and FA data separately, to derive 100 PC features for each, capturing over 80% variance for GM and over 65% for FA, respectively.
CCA Feature Extraction: CCA is a straightforward way of getting the correlated features between two datasets. Consider two datasets $X \in \mathbb{R}^{n \times p}$ and $Y \in \mathbb{R}^{n \times q}$, n is the number of samples, p and q are numbers of variables for each dataset. Let Σ_{XY} be the covariances of X, Y. CCA finds the linear transformation of X and Y such that the latent features or canonical covariates Xu and Yv are maximally correlated. The standard objective function for optimization is shown in Equation (1).

$$max\ u^T \Sigma_{XY} v = min\ ||Xu - Yv||^2$$
$$s.t.\ ||Xu||_2 = ||Yv||_2 = 1 \tag{1}$$

When the dimensionality of the dataset is high, SCCA is implemented that incorporates the sparsity. The objective function of a typical SCCA is as shown in equation (2). We used the Iterative Penalized Least Square (IPLS) SCCA [1].

$$\min\ : ||Xu - Yv||^2 + \lambda_1||u||_1 + \lambda_2||v||_1$$
$$\text{s.t. } ||u||_1 \le 1, ||v||_1 \le 1, ||Xu||_2 = ||Yv||_2 = 1, \tag{2}$$

Reference CCA Feature Extraction: RCCA model incorporates a reference variable into the typical CCA model. The objective function for the RCCA is as shown in equation (3).

$$min : ||Xu - Yv||^2 + \lambda_3||Xu - ref||^2 + \lambda_4||Yv - ref||^2 +$$
$$\lambda_1||u||_1 + \lambda_2||v||_1 \tag{3}$$
$$s.t.\ ||Xu||_2 = 1, ||Yv||_2 = 1,$$

where λ_1 ,λ_2 λ_3 ,λ_4 are the regularization terms, *ref* is the normalized reference vector. We used each intelligence score as the reference, and λ_3 ,λ_4 were assigned as 0.5
Prediction of intelligence: LASSO regression was used for the prediction of intelligence scores for each types of features derived from PCA, SCCA and RCCA.
Implementation: Samples were divided into training (80%) and testing(20%), and 5-fold cross-validation within the training samples.

RESULT

SCCA: 30 pairs of latent features were significantly correlated. The correlations of the first five pairs are shown in Table 1.

Table 1: Correlation of SCCA features

Score	1st comp	2nd comp	3rd comp	4th comp	5th comp
Training	0.86	0.77	0.77	0.69	0.47
Testing	0.80	0.75	0.74	0.68	0.46

RCCA: The numbers of significant pairs of features extracted by RCCA models are 31, 33, and 31 for Fluid intelligence, Crystallized intelligence, and Total composite scores, respectively.

Table 2: Correlation of RCCA model (Fluid)

	Score of	1st comp	2nd comp	3rd comp	4th comp	5th comp
Training	GM and FA	0.58	0.61	0.68	0.76	0.8
	GM and Ref	0.09	0.11	0.18	0.12	0.11
	FA and Ref	-0.23	-0.19	-0.11	-0.1	-0.06
Testing	GM and FA	0.58	0.58	0.66	0.75	0.77
	GM and Ref	0.06	0.11	0.14	0.07	0.15
	FA and Ref	-0.22	-0.14	-0.11	-0.1	-0.03

Table 3: Correlation of RCCA model (Crystallized)

	Score of	1st comp	2nd comp	3rd comp	4th comp	5th comp
Training	GM and FA	0.49	0.54	0.68	0.8	0.81
	GM and Ref	0.16	0.17	0.21	0.14	0.12
	FA and Ref	-0.26	-0.23	-0.12	-0.09	-0.06
Testing	GM and FA	0.47	0.49	0.65	0.78	0.79
	GM and Ref	0.15	0.15	0.19	0.12	0.14
	FA and Ref	-0.25	-0.22	-0.1	-0.08	-0.07

RESULT

Table 4: Correlation of RCCA model (Total Composite)

	Score of	1st comp	2nd comp	3rd comp	4th comp	5th comp
Training	GM and FA	0.48	0.53	0.67	0.8	0.82
	GM and Ref	0.16	0.18	0.22	0.14	0.12
	FA and Ref	-0.28	-0.23	-0.12	-0.09	-0.06
Testing	GM and FA	0.47	0.49	0.64	0.79	0.79
	GM and Ref	0.15	0.18	0.2	0.1	0.16
	FA and Ref	-0.28	-0.2	-0.14	-0.09	-0.04

Intelligence prediction: Table 5 compares the prediction results of three intelligence scores.

Table 5: Intelligence prediction using PCA, SCCA and RCCA features

	Fluid			**Crystallized**			**Total Composite**		
	PCA	SCCA	RCCA	PCA	SCCA	RCCA	PCA	SCCA	RCCA
r^2	0.13	0.07	0.13	0.2	0.16	0.2	0.21	0.14	0.21
MSE	0.86	0.92	0.86	0.77	0.81	0.77	0.77	0.84	0.77
Feature number	128	75	11	130	81	10	169	77	11

For illustration purposes, we plotted brain regions linked to total composite score using RCCA features, including six GM features and five FA features. GM regions positively related to intelligence include cerebellar tonsil, insula, culmen, and cingulate in red, and GM regions negatively related are lingual gyrus, middle occipital gyrus, thalamus, anterior cingulate in green. FA in corpus callosum is positively related to intelligence.

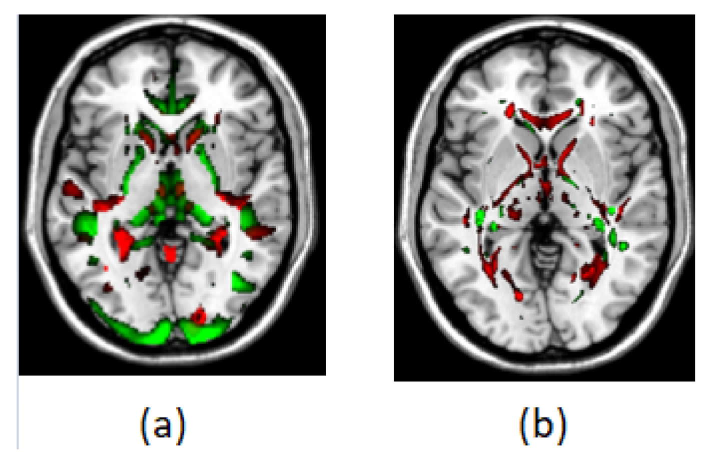

Fig 1: single slice representing the top brain regions of (a) GM (b) FA

DISCUSSION AND CONCLUSION

RCCA features achieve higher r-squares than SCCA features and approximately same r-squares as PCA features by using a much smaller number of features. Therfore, RCCA efficiently extracts informative features across multiple modalities, aiding in identifying cognitively sensitive brain regions.

REFERENCE

[1] J. Chapman and H.-T. Wang, “Cca-zoo: A collection of regularized, deep learning based, kernel, and probabilistic cca methods in a scikit-learn style framework”, vol.6, 2021.