Combination of FMRI-SMRI-EEG Data Improves Discrimination of Schizophrenia Patients by Ensemble Feature Selection

Jing Sui*, *Senior Member IEEE,* Eduardo Castro, Hao He, David Bridwell, Yuhui Du, Godfrey D Pearlson, Tianzi Jiang and Vince D Calhoun, *Fellow IEEE*

***Abstract*— Multimodal brain imaging data fusion is a scientifically interesting and clinically important topic; however, there is relatively little work on N-way data fusion. In this paper, we applied multi-set canonical correlation analysis (MCCA) to combine data of resting state fMRI, EEG and sMRI, in order to elucidate the abnormalities that underlie schizophrenia patients and also covary across multiple modalities. We also tested whether the identified group-discriminative components can be used for feature selection in group classification. MCCA is demonstrated to be an effective feature selection technique, especially in multimodal fusion. We also proposed an ensemble feature selection scheme by combining two sample t-test, MCCA and support vector machine with recursive feature elimination (SVM-RFE), resulting in optimal group-discriminating features for each modality. Finally, we compared the classifying power between two groups based on the above selected features via 7 modality-combinations. Results show that the fMRI-sMRI-EEG combination derives the best classification accuracy in training (91%) and predication rate (100%) in testing data, validating the effectiveness and advantages of multimodal fusion in discriminating schizophrenia.**

I. Introduction

Multimodal brain imaging techniques are playing increasingly important roles in elucidating structural and functional properties in normal and diseased brains, as well as providing the conceptual glue to bind together data from multiple types or levels of analysis. However, most current approaches have focused on pair-wise fusion and there is still relatively little work on N-way data fusion and examination of the full relationships among multiple data types. Given the availability of more powerful MR scanners, there are typically more than 2 imaging modalities available for one participant. Hence, we believe the joint multivariate analysis of multiple data types will improve our ability to understand brain diseases and show promise for biomarker identification.

Till now, Most multivariate *N*-way fusion models are based on canonical correlation analysis(CCA) or independent component analysis (ICA), *e.g.* multi-set CCA (MCCA) [1] and "mCCA+joint ICA" [2, 3], which have been successfully applied to combine 3 modalities, *e.g.*, fMRI, sMRI and DTI. Both methods are able to identify both modal-common and modal-unique group discriminative patterns for schizophrenia patients (SZ) versus healthy controls (HC), which makes discoveries of changes in one modality causing related alterations in distant, but connected regions in other modalities possible.

To the best of our knowledge, there has been no report combining resting state fMRI, resting state EEG and sMRI together to study schizophrenia. Additionally, whether the identified group-differentiating components from the proposed *N*-way fusion model can be used for classification? Which combination of the above 3 modalities achieves the greatest discriminating power? These questions are still to be determined.

In this project, first we applied the 3-way MCCA model to examine the joint and modality-unique group differences between SZ and HC by using resting state fMRI, resting state EEG and sMRI data. Then we proposed an ensemble feature selection strategy to determine the optimal features for each modality, by taking advantages of 3 feature selection methods: MCCA, two sample t-test and support vector machine with recursive feature elimination (SVM-RFE). Finally, we compared the group prediction power of 7 modality-combinations based on above selected features.

The remaining of this paper is organized as follows: section 2 describes the MCCA algorithm, the proposed feature-selection and classification framework and a brief introduction of SVM-RFE. In section 3, the used real human data and its preprocessing are described. Section 4 and 5, the real data application results are introduced and discussed, as well as the future work.

II. Approaches

A. Multimodal Fusion by MCCA

The basic strategy of mCCA+jICA is shown in Figure 1. We assume the multimodal fusion input matrix $\mathbf{X}_k$ (subjects-by-voxels) is a linear mixture of *M* sources given by $\mathbf{S}_k$, mixed with a nonsingular mixing matrix $\mathbf{A}_k$:

$$\mathbf{X}_k = \mathbf{A}_k \cdot \mathbf{S}_k \qquad k = 1, 2, 3 \qquad (1)$$

MCCA aims to project $\mathbf{X}_k$ into a space so that correlations among mixing profiles $\mathbf{A}_k$ of *n* modalities are jointly maximized (in the sense of sum of squared correlations in this study)[4]. As shown in Figure 1, the resulting canonical variants (CVs) $\mathbf{A}_k$ are correlated from high to low pair-wisely only on corresponding columns. Namely, MCCA can associate multiple modalities with flexible linkages (correlation) in their mixing matrices:

$$E\{\mathbf{D}_k^T \mathbf{D}_k\} = \mathbf{I} \quad E\{\mathbf{D}_i^T \mathbf{D}_j\} \approx diag(r_{ij}^1, r_{ij}^2 ... r_{ij}^M) \qquad (2)$$

And the corresponding components $\mathbf{S}_k$ can be derived as:

*Resrach supported by NIH and Chinese Academy of Sciences (CAS)

Jing Sui and Tianzi Jiang are with the Brainnetome center & National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences(corresponding author e-mail: kittysj@gmail.com).

Jing Sui, Eduardo Castro, Hao He, Yuhui Du and Vince D Calhoun are with the Mind Research Network/LBERI and Dept. of ECE, University of New Mexico, Albuquerque, NM, 87106. (vcalhoun@unm.edu).

Godfrey D Pearlson is with Olin Neuropsychiatry Research Center, Hartford, CT and Dept. of Psychiatry, Yale University, New Haven, CT, 06519.

$$\mathbf{S}_k = pinv(\mathbf{A}_k) \cdot \mathbf{X}_k \qquad (pinv\text{: pseudo inverse}) \quad (3)$$

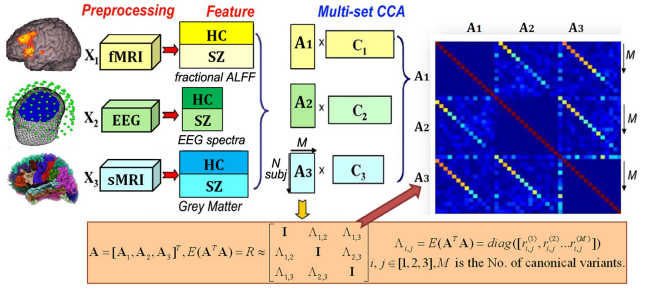

Figure 1. Three-way fusion strategy of MCCA

Two-sample t-tests were then performed column-wisely on $\mathbf{A}_k$ to test which components are significantly different between HC and SZ. If components of the same index show group differences in more than one modality, they are called modality-common group-discriminative components, otherwise, they are modality-unique discriminative ones.

B. Classification based on MCCA components

To test the potential use of the identified group discriminative components, we used them to generate features and train a classifier, to see whether they are able to predict diagnosis or serve as potential biomarkers, which may prove the great significance for multimodal analysis.

For each modality, we transferred the group-discriminating components into *Z* values and set a threshold at |Z|>3 to generating masks. The masks of the same modality were then combined and applied to the raw input matrix of each modality, which served as the input to the further classification based on single-modal and multimodal features. Each individual was assigned one of two class memberships. We then trained linear support vector machine with recursive feature elimination [5] by 10-fold cross-validation (trained on 90% of the randomly chosen data samples and tested on the other 10%) 100 times for each of the 7 modal combinations and recorded sensitivity and specificity, in order to find out which combination is optimal for SZ discrimination when using the features generated by MCCA.

C. Ensemble feature selection

After verifying that MCCA is an effective feature selection method, we proposed an ensemble feature selection strategy, by combining three techniques: two sample t-test (with p<0.01), MCCA(with |Z score|>3) and SVM-RFE[5], as shown in Figure 2.

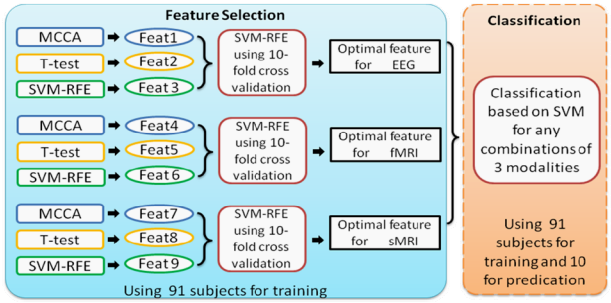

Figure 2 flowchart to select optimal features&modal-combinations

Here we trained SVM using recursive feature elimination (SVM-RFE), which is able to determine a value of each feature, i.e. frequency values, by training a SVM using training samples with class labels to identify a determinative subset. The frequency values ranged from 0 to 1; the higher the value, the more relevant a specific brain feature is for the classification. One or more features having the smallest values are removed and an updated kernel matrix is generated using the remaining features. The process is repeated until a predetermined number of features remain which are capable of accurately separating the data into two classes [6]. Note that SVM-RFE can be used for both feature selection and classification. Therefore the subset of the EEG spectra/ ALFF/ GM that achieves a minimum validation error is chosen as the set of most informative features for each modality.

There are two stages as shown in Fig 2. In the first stage (feature selection), we used only 91 subjects, and the other 5 HC and 5 SZ were set aside and used for prediction in the second stage (classification). Both stages were repeated 10 times for a random set of 10 testing subjects. For each modality, after we got features of interest by 3 methods, all of them were combined and trained by SVM-RFE again via 10-fold cross-validation. Those features with frequency values >0.5 were selected as optimal features. The corresponding specificity, sensitivity, and predication rates (for the 10 testing subjects) for each single modality were recorded, see part *B* in Results.

D. Determination of optimal modality-combination

After obtaining the optimal features for each single modality, we tested the group classification (for 91 training subjects) and predication power (for 10 testing subjects) of 7 modal-combinations (3 single, 3 pair-wise, 1 three-way) by a linear SVM classifier again with 10-fold cross-validation. The specificity, sensitivity, classification accuracy (for 91), and predication rates (for 10) are displayed in Results section.

III. DATA

A. Human brain data

Participants were recruited at the Olin Neuropsychiatric Research Center, CT and were scanned by a 3T Siemens scanner for resting state fMRI and sMRI. EEG was recorded with a 64-channel (sampling rate = 1000 Hz) during a 5 minute resting-state with their eyes open. Electrodes were placed according to the standard 10-20 placement. An additional two channels recorded ocular artifacts. All subjects gave written, informed, Hartford hospital IRB approved consent. Table 1 lists their demographics. Before doing two sample t-test on mixing coefficients of MCCA, we regressed age and gender factors out to remove the potential influence of these variables on SZ-HC difference.

Table 1. Demographics of 101 participants

	Num	*Age*	*Gender*	*Ethnic*	*Hands*
SZ	48	30.5+-11.5	36M 12 F	40 white	4 left hand
HC	53	36.7+-12.4	23M 30F	50 white	4 left hand
p		0.012	0.001	0.73	0.8

B. Data preprocessing

The high-dimensional neuroimaging data is typically very noisy, thus redundancy reduction and denoising is important to facilitate multimodal fusion [7] and to provide a simpler representative space for each modality, *e.g.,* the fractional amplitude of low-frequency fluctuation (ALFF) of resting state fMRI, the gray matter (GM) segmentation image from sMRI, the frequency spectra of resting-state EEG and voxel-wise DTI fractional anisotropy *etc*. we used the first 3 measures in this project.

FMRI preprocessing: SPM8 software package was employed (http://www.fil.ion.ucl.ac.uk/spm /software/spm8) Ten first time points were removed and images were realigned. The fMRI data were then spatially normalized and slightly down sampled to 3×3×3 mm^3. To remove the motion effect, we further regressed out the 6 motion parameters for each slice and continued to spatial smoothing (8×8×8 mm^3). Finally, we extracted the voxel wise ALFF map, which is calculated by computing the fast Fourier transform (FFT) of each voxel time series, taking the square root of the power spectrum to obtain amplitude, and averaging amplitude in [0.01, 0.1] Hz. Prior to computing ALFF, the original 4-D fMRI data sets were divided by their global mean (over time and space) to normalize differences in scan intensity units[8].

SMRI preprocessing: sMRI data were also preprocessed using the SPM8 software package which was used to segment the brain into white matter (WM), GM, and cerebral spinal fluid with un-modulated normalized parameters via the unified segmentation method [9]. After segmentation, the GM images were down-sampled and smoothed similar to the fMRI data. Subject outlier detection was further performed using a spatial correlation with the template image, to ensure that all subjects were properly segmented, (for details, see [10]).

EEG preprocessing: was conducted using both EEGLAB toolbox (http://sccn.ucsd.edu/eeglab) and custom functions. Data was band pass filtered (0.01- 0.5 Hz), down-sampled to 250Hz and average referenced. Individual channels were excluded if the voltage deviated by ±50 mv (4 % of channels were excluded). Then EEG data was segmented into 5 individual 60-second epochs and epochs were excluded if more than 30 channels exceeded ± 500 mv (47 epochs were excluded out of 550). Eye blink and muscle artifacts were attenuated by conducting a temporal ICA decomposition via EEGLAB. The number of components equaled the number of good channels within the segment. 15 artifactual components were removed by visual inspection. The ICA reconstructed data was segmented into 2 second epochs without overlap. Each epoch was decomposed into frequencies with the discrete Fourier transform (Δf=0.5Hz) implemented with the fast Fourier transform function. The complex valued Fourier data was absolute valued (i.e. converted to amplitude), logged, averaged across epochs, and the response at excluded electrodes was interpolated (inverse distance weighting). The resulting [freq × channel] matrix was converted to a vector for each subject for MCCA analysis.

IV. Results

A. Group difference identified by MCCA

The components with significantly different mixing coefficients between HC and SZ are shown in Figure 3, where the joint group-discriminative ones are framed in green color (asterisk indicates FDR correction for multiple comparison)

fMRI				10 *
p-value				0.0076
sMRI	1 *	2	4	
p-value	0.0084	0.043	0.024	
EEG	1		4	10
p-value	0.024		0.032	0.033

Figure 3. Group- discriminative components between SZ and HC

Fig 4 presents the overlapped spatial maps (extracted by |Z|>3) of the components (fMRI #10, sMRI #1&4, EEG #1&4))that were used for classification. It shows that prefrontal, occipital and motor regions are emphasized in resting EEG spectra (Fig 4b) and the low frequency bands are highly activated. For sMRI (Fig 4c), superior temporal gyrus, parahippocampal gyri and inferior parietal cortex are significant. In ALFF map (Fig 4d), the default mode network, middle frontal gyrus and middle occipital gyrus are displayed; All these findings are consistent with previous reports [11], while our method is able to further link these co-varied alterations among modalities.

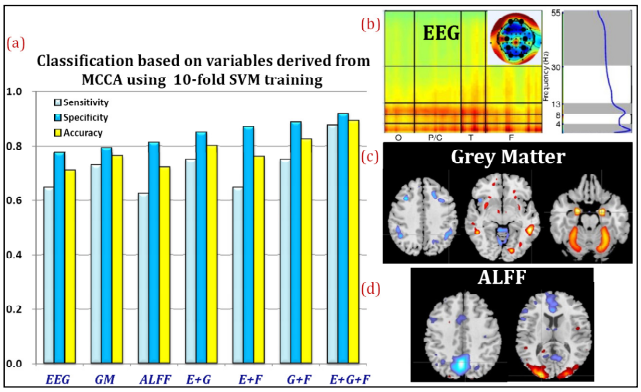

Figure 4. Spatial maps of the selected features by MCCA components and the classification results based on these features.

Fig 4(a) indicates the classification results based on the features extracted by MCCA. It is notable that combination of EEG-GM-ALFF achieves the best classification accuracy at 90%, both sensitivity and specificity increase about 15% compared to single modalities, suggesting that multimodal fusion does improve the diagnosis prediction, in accordance with [12, 13], and the MCCA method is promising for identification of potential biomarkers.

B. Optimal feature and modality-combination

Figure 5 indicates the effectiveness of the proposed ensemble feature selection scheme for every modality. It's clear that SVM-RFE is an excellent feature selection method and it greatly improves the classification and prediction after being used in the combined features (T-test + MCCA + SVM-RFE) to further remove the less discriminative ones, see the last row in each sub-table. In this unbiased test, sensitivity, specificity and classification accuracy all

increased considerably compared using only one technique. Moreover, the prediction rates for 10 blind test subjects were all more than 80%. In addition, sMRI and fMRI appears to be more group-discriminative than EEG spectrum, and fMRIshows 90% accuracyin testing.

EEG	training (N=91)			testing(N=10)
Feat selected	sensitivity	specificity	accuracy	Prediction
ttest2	0.56	0.75	0.66	0.5
svm-rfe	0.63	0.75	0.70	0.6
mcca	0.57	0.71	0.69	0.8
combined	**0.72**	**0.75**	**0.74**	**0.8**
FMRI	**training (N=91)**			**testing (N=10)**
Feat selected	sensitivity	specificity	accuracy	Prediction
ttest2	0.74	0.69	0.72	0.8
svm-rfe	0.84	0.96	0.90	0.6
mcca	0.63	0.79	0.71	0.7
combined	**0.79**	**0.88**	**0.84**	**0.9**
SMRI	**training (N=91)**			**testing(N=10)**
Feat selected	sensitivity	specificity	accuracy	Prediction
ttest2	0.88	0.98	0.93	0.7
svm-rfe	0.72	0.85	0.79	0.7
mcca	0.67	0.81	0.75	0.6
combined	**0.77**	**0.94**	**0.86**	**0.8**

Figure 5. Classification results using various selected features.

Figure 6 indicates the most discriminative feature maps for each modality, as well as the specificity, sensitivity, classification accuracy (for 91), and predication rates (for 10) for all 7 modality combinations. The 3-way combination again achieves the best performance with 91% accuracy and 100% prediction rate, confirming the strengths of multimodal fusion. Furthermore, the implicated brain regions for each modality are in well accordance with previous reports, all demonstrating brain deficits in schizophrenia underlying frontal lobe. Generally speaking, the proposed method takes advantage of different imaging modalities to help improve the diagnosis of mental illness.

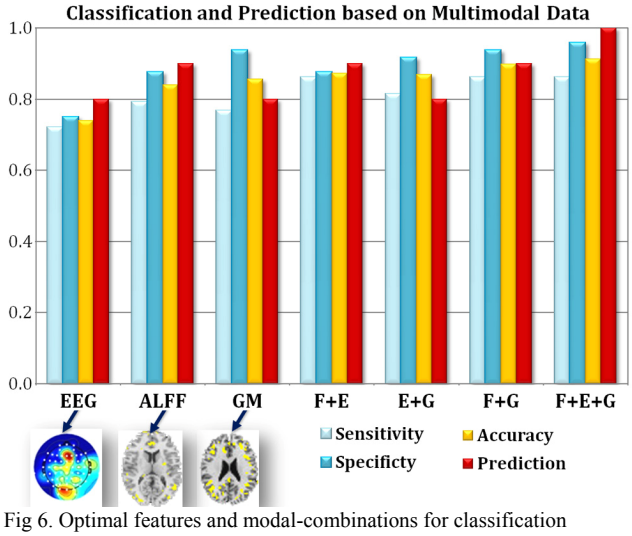

Fig 6. Optimal features and modal-combinations for classification

V. Conclusion

This project is to our knowledge the first attempt to combine resting-state fMRI, resting state EEG and sMRI data to discriminate schizophrenia. Our results suggest that multimodal fusion of the selected group-discriminative MCCA components enhances the diagnosis prediction substantially. Moreover, we proposed an ensemble feature selection strategy, in which SVM-RFE acts as a core and capitalizes the strengths of multiple methods, leading to the discovery of the most group differentiating features for each single modality, and the optimal modal- combination for classification.

Though we found fMRI-sMRI-EEG combination is the most group distinguishing in this study, we note that fusing as many modalities as possible does not guarantee the best classification rates, as reported in [3]. In order to detect potential biomarkers for several brain disorders, the proposed analysis could be used to compare all possible combinations when multiple modalities are available. We plan to pursue this possibility in future work by using larger data sets and various modalities, which aims to have bigger effect size and achieve higher accuracy.

Acknowledgment

This work was supported by the Strategic Priority Research Program of the Chinese Academy of Sciences (Grant No. XDB02030300), the "100 Talents Plan" of Chinese Academy of Sciences (to Sui J) and the National Institutes of Health grants R01EB 006841 and R01EB005846 (to Calhoun VD).

References

[1] N. M. Correa, Y. O. Li, T. Adali et al., "Fusion of fMRI, sMRI, and EEG data using canonical correlation analysis " in ICASSP 2009, Taipei, Taiwan, 2009, pp. 385-388.

[2] J. Sui, H. He, G. D. Pearlson et al., "Three-way (N-way) fusion of brain imaging data based on mCCA+jICA and its application to discriminating schizophrenia," Neuroimage, vol. 2, no. 66, pp. 119-132, Oct 26, 2012.

[3] J. Sui, H. He, Q. Yu et al., "Combination of Resting State fMRI, DTI, and sMRI Data to Discriminate Schizophrenia by N-way MCCA + jICA," Front Hum Neurosci, vol. 7, pp. 235, 2013.

[4] Y. O. Li, T. Adali, W. Wang et al., "Joint Blind Source Separation by Multi-set Canonical Correlation Analysis," IEEE Trans Signal Process, vol. 57, no. 10, pp. 3918-3929, Oct 1, 2009.

[5] I. Guyon, Weston, J., Barnhill, S., Vapnik, V., "Gene selection for cancer classification using support vector machines," Mach. Learn., vol. 46, pp. 389, 2002.

[6] E. Castro, M. Martinez-Ramon, G. Pearlson et al., "Characterization of groups using composite kernels and multi-source fMRI analysis data: Application to schizophrenia," Neuroimage, vol. 58, no. 2, pp. 526-36, Sep 15, 2011.

[7] V. D. Calhoun, and T. Adali, "Feature-based fusion of medical imaging data," IEEE Trans Inf Technol Biomed, vol. 13, no. 5, pp. 711-20, Sep, 2009.

[8] E. B. Erhardt, E. A. Allen, E. Damaraju et al., "On network derivation, classification, and visualization: a response to Habeck and Moeller," Brain Connect, vol. 1, no. 2, pp. 1-19, Jun 8, 2011.

[9] J. Ashburner, and K. J. Friston, "Unified segmentation," Neuroimage, vol. 26, no. 3, pp. 839-51, Jul 1, 2005.

[10] J. M. Segall, J. A. Turner, T. G. van Erp et al., "Voxel-based morphometric multisite collaborative study on schizophrenia," Schizophr Bull, vol. 35, no. 1, pp. 82-95, Jan, 2009.

[11] P. N. Jayakumar, G. Venkatasubramanian, B. N. Gangadhar et al., "Optimized voxel-based morphometry of gray matter volume in first-episode, antipsychotic-naive schizophrenia," Progress in neuro-psychopharmacology & biological psychiatry, vol. 29, no. 4, pp. 587-91, May, 2005.

[12] H. Yang, J. Liu, J. Sui et al., "A Hybrid Machine Learning Method for Fusing fMRI and Genetic Data: Combining both Improves Classification of Schizophrenia," Front Hum Neurosci, vol. 4, pp. 192, 2010.

[13] J. Sui, T. Adali, G. D. Pearlson et al., "An ICA-based method for the identification of optimal FMRI features and components using combined group-discriminative techniques," Neuroimage, vol. 46, no. 1, pp. 73-86, May 15, 2009.