Comparison of techniques for classification of patients with schizophrenia and healthy controls

JL Winterburn,[a,b,c,d] **AN Voineskos,**[a,d,e,f] **G Devenyi,**[b] **BH Mulsant,**[d,f] **JP Lerch,**[g] **MM Chakravarty**[b,h]

a- Kimel Family Translational Imaging-Genetics Research Laboratory, Research Imaging Centre, Centre for Addiction and Mental Health, Toronto, Canada; b- Cerebral Imaging Centre, Douglas Mental Health University Institute, McGill University, Montreal, Canada; c- Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, Canada; d- Campbell Family Mental Health Research Institute, Centre for Addiction and Mental Health, Toronto, Canada; e- Department of Psychiatry, University of Toronto, Toronto, Canada; f- Division of Geriatric Psychiatry, Centre for Addiction and Mental Health, Toronto, Canada; g- Program in Neuroscience and Mental Health, The Hospital for Sick Children, Toronto, Canada; h- Department of Psychiatry, McGill University, Montreal, Canada

Objective

Machine learning (ML) is a computer-assisted pattern recognition tool that can automatically identify individuals suffering from a neuropsychiatric disorder such as schizophrenia[1] using only neuroanatomical features (ex. tissue density) derived from magnetic resonance (MR) images.

We applied three common ML methods (support vector machine, logistic regression, and linear discriminant analysis) to a single dataset to determine which method was best at distinguishing schizophrenia (SZ) patients from healthy controls (HC). We then repeated this study design on two replication datasets to confirm our findings.

Datasets

1. CAMH (Centre for Addiction and Mental Health, Toronto)[2]

T1-weighted images collected on 1.5T GE Echospeed MR scanner; 0.78mm x 0.78mm x 1.5mm voxel dimensions

Demographic	Schizophrenia Patients (n=88)		Healthy Controls (n=103)	
	Mean	SD	Mean	SD
Age	36.6	12.5	35.2	12.6
Education (years) [a]	13.3	2.32	15.5	1.89
Parental Education (years)	15.6	3.91	16.8	4.75
WTAR (IQ) [b]	108.8	15.8	117	8.27
MMSE [c]	28.9	1.76	29.5	0.8
CIRS-G [d]	1.7	0.76	0.79	0.63
Age of onset	23.8	6.73	NA	NA
Illness Duration (years)	13.3	12.5	NA	NA
Chlorpr. Equiv (mg)	266.6	292.4	NA	NA
PANSS				
Positive	13.5	5.36	NA	NA
Negative	14.2	6.18	NA	NA
General	24.8	7.38	NA	NA
	N		N	
Diagnosis	51 SCZ	21 SA	NA	
Antipsychotic Treatment	5 1°	54 2°	NA	
Gender	57 M	30 F	57 M	43 F
Handedness	66 R	5 L	80 R	5 L

Significance in independent samples t-test: [a]p<<0.0001 [b]p<0.001 [c]p<0.01 [d]p<<0.0001

WTAR = Wechsler Test of Adult Reading
MMSE= Mini Mental State Examination
CIRS-G = Cumulative Illness Rating Scale for Geriatrics
PANSS = Positive and Negative Syndrome Scale

2. Mexico (Instituto Nacional de Neurologia y Neurocirugia, Mexico City)[3]

All patients were recruited during their first nonaffective psychosis episode, and were antipsychotic naive; T1-weighted images collected on 3T GE Healthcare MR scanner; 0.5mm x 0.5mm x 1.2mm voxel dimensions

Demographic	FEP Group (n=60)		Healthy Controls (n=59)	
	Mean	SD	Mean	SD
Age	24.7	7.68	35.2	12.6
Education (years) [a]	11.5	3.13	15	2.9
Illness Duration (wks)	33	52.7	NA	NA
PANSS				
Positive	24.1	4.97	NA	NA
Negative	24.3	5.66	NA	NA
General	48.8	8.38	NA	NA
	N		N	
Diagnosis	14 BPD/ 21 SFD/ 25 SZ		NA	
Gender	37 M	23 F	38 M	21 F

Significance in independent samples t-test: [a]p<0.05

PANSS = Positive and Negative Syndrome Scale
BPD = Brief Psychotic Disorder
SFD = Schizophreniform Disorder
SZ = Schizophrenia

3. NUSDAST (Northwestern University Schizophrenia Data and Software Tool) [4]

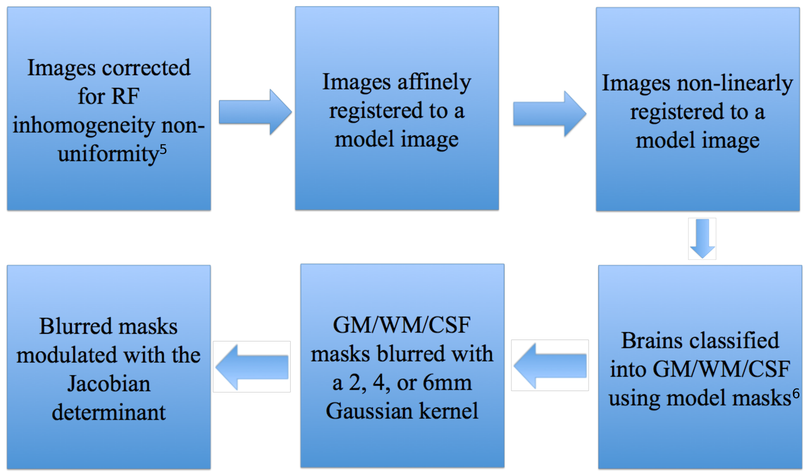
T1-weighted images collected on 1.5T Seimens MR scanner; 1mm x 1mm x 1mm dimensions

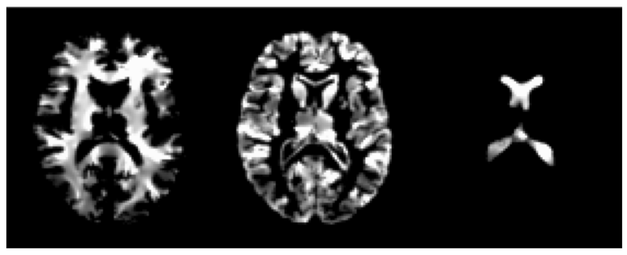
Demographic	Schizophrenia Patients (n=147)		Healthy Controls (n=137)	
	Mean	SD	Mean	SD
Age	34.8	12.9	33.1	14.1
Education (years) [a]	12.2	2.44	14.5	2.64
Parental Education (years)	12.7	3.27	13.3	2.96
PANSS				
Positive	23.1	17.3	NA	NA
Negative	29.6	17.8	NA	NA
	N		N	
Gender	37 M	23 F	95 M	52 F
Handedness	129R	15L	123 R	13L

Significance in independent samples t-test: [a]p<0.05

PANSS = Positive and Negative Syndrome Scale

Image Processing

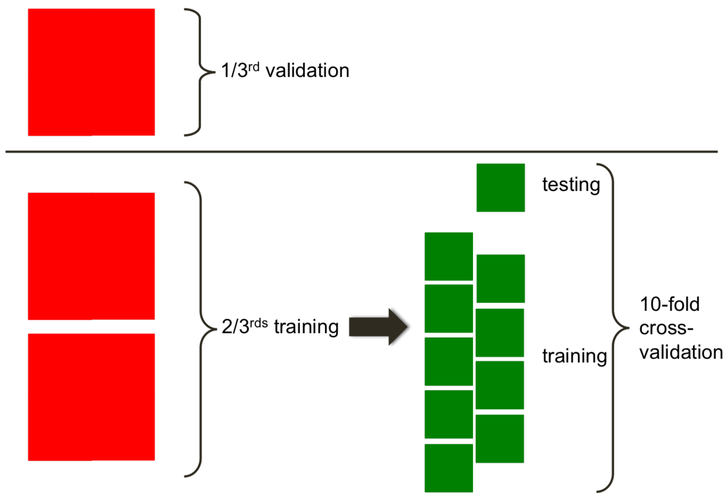




Images are masked into tissue type: (from left): grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF).[7]

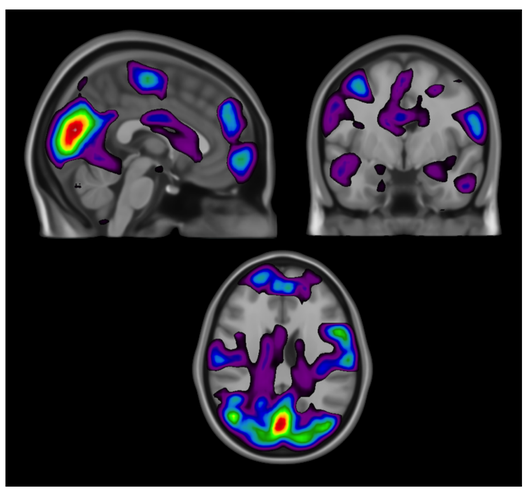
Results

Validation



Principal Component Analysis (PCA)

Data was first reduced with a principal component analysis. CAMH: 191 PCs; Mexico: 119 PCs; NUSDAST: 284 PCs



The first PC from the CAMH dataset (of 191 PCs total) of the GM density dataset explains 32% of the variance in the data and distinguishes between SZ and HC (R^2= 0.061, p=0.020). This PC describes GM reductions in the superior frontal, medial orbito-frontal, occipital (lingual/peri-calcarine area), cingulate, medial and superior temporal lobes, and the precentral gyrus in patients relative to controls.

Measuring Performance

Sensitivity = # of true positives (eg. SZ classified as SZ)
Specificity = # of true negatives (eg. HC classified as HC)
Accuracy = Sensitivity + Specificity

Support Vector Machine (SVM)

Linear Kernel

	Sensitivity	Specificity	Accuracy
CAMH	65%	34%	51%
Mexico	59%	52%	52%
NUSDAST	47%	55%	54%

Radial Basis Function Kernel

	Sensitivity	Specificity	Accuracy
CAMH	62%	48%	53%
Mexico	55%	70%	60%
NUSDAST	60%	45%	55%

Logistic Regression (LR)

	Sensitivity	Specificity	Accuracy
CAMH	75%	88%	81%
Mexico	60%	81%	72%
NUSDAST	62%	79%	77%

Linear Discriminant Analysis (LDA)

	Sensitivity	Specificity	Accuracy
CAMH	51%	53%	52%
Mexico	45%	60%	52%
NUSDAST	50%	55%	53%

Conclusions

-> Logistic Regression is the most effective classification technique among the three tested for classifying schizophrenia patients from healthy controls based on tissue density
-> Although higher accuracies are reported in places in the literature, our methods are consistent across algorithms, and we avoid over-tuning
-> Understanding the most effective classifier may offer insight into the biological underpinnings of schizophrenia

References

1. Zarogianni, E., Moorhead, T.W.J., & Lawrie, S.M. (2013). Towards the identification of imaging biomarkers in schizophrenia using multivariate pattern classification at a single-subject level. *NeuroImage. Clinical, 3,* 279-89. doi:10.1016/j.nicl.2013.09.003; **2.** Wheeler, A. L., Chakravarty, M. M., Lerch, J. P., Pipitone, J., Daskalakis, Z. J., Rajji, T. K., … Voineskos, A. N. (2013). Disrupted Prefrontal Interhemispheric Structural Coupling in Schizophrenia Related to Working Memory Performance. Schizophrenia Bulletin, 1–11. doi:10.1093/schbul/sbt100; 3. De la Fuente-Sandoval, C., León-Ortiz, P., Azcárraga, M., Stephano, S., Favila, R., Díaz-Galvis, L., … Graff-Guerrero, A. (2013). Glutamate levels in the associative striatum before and after 4 weeks of antipsychotic treatment in first-episode psychosis: a longitudinal proton magnetic resonance spectroscopy study. JAMA Psychiatry, 70(10), 1057–66. doi:10.1001/jamapsychiatry.2013.289; **4.** Wang, L., Kogan, A., Cobia, D., Alpert, K., Kolasny, A., Miller, M. I., & Marcus, D. (2013). Northwestern University Schizophrenia Data and Software Tool (NUSDAST). Frontiers in Neuroinformatics, 7(November), 25. doi:10.3389/fninf.2013.00025; **5.** Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametricmethod for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17 (1), 87–97. **6.** Ashburner, J., & Friston, K. J. (2000). Voxel-based morphometry--the methods. NeuroImage, 11(6 Pt 1), 805–21. doi:10.1006/nimg.2000.0582; **7.** Fan, Y., Shen, D., & Gur, R. (2007). COMPARE: classification of morphological patterns using adaptive regional elements. Medical Imaging, IEEE …, 26(1), 93–105.

Acknowledgements: MMC: The W. Garfield Weston Foundation, Canadian Institutes of Health Research (CIHR); Douglas Mental Health University Institute; ANV: Canadian Institutes of Health Research (CIHR); National Alliance for Research on Schizophrenia and Depression (NARSAD), University of Toronto, Centre for Addiction and Mental Health (CAMH), Ontario Mental Health Foundation (OMHF), and CAMH Foundation (Koerner New Scientist Program and Paul Garfinkel New Investigator Catalyst Fund)