Comparison of techniques for classification of schizophrenia patients and healthy controls based on cortical thickness

JL Winterburn,[a,b,c,d] **AN Voineskos,**[a,d,e,f] **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. cortical thickness).[2,3]

We applied six ML methods to a single subject set to determine which method was best at distinguishing schizophrenia (SZ) patients from healthy controls (HC). These ML methods include:

1. Principal Component Analysis
2. Naive Bayes
3. Logistic Regression
4. Support Vector Machine
5. Linear Discriminant Analysis
6. Neural Networks

Subject Population

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: WTAR = Wechsler Test of Adult Reading
[a]p<<0.0001 [b]p<0.001 MMSE= Mini Mental State Examination
[c]p<0.01 [d]p<<0.0001 CIRS-G = Cumulative Illness Rating Scale for Geriatrics
PANSS = Positive and Negative Syndrome Scale

-> T1-weighed *in vivo* images collected on 1.5T GE Echospeed MR scanner
-> Final image resolution of 0.78 mm x 0.78 mm x 1.5 mm

Cortical Thickness with CIVET

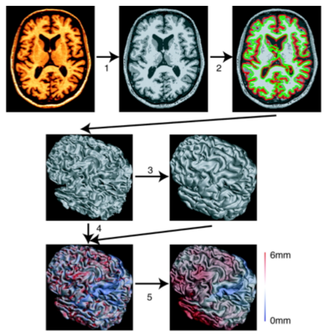

Figure 1: CIVET pipeline: images are 1) Classified into white matter (WM)/ grey matter (GM)/cerebral spinal fluid (CSF); 2) Fit with a WM surface; 3) Fit with a GM surface; 4) Cortical thickness (distance between WM and GM surfaces) is measured at 81924 vertices; 5) Thicknesses are blurred with a 20mm surface-based kernel.[4]

Classification Techniques

Measuring Performance

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

Validation

66% train/33% test split; 10-fold cross-validation in the train set

1. Principal Component Analysis (PCA)

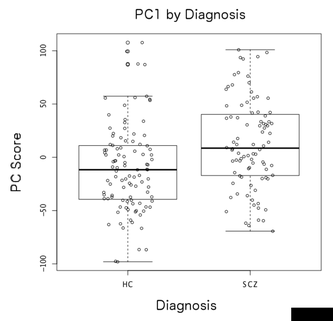

Figure 2: The first PC (of 191 PCs total) explains 32% of the variance in the data and distinguishes between SZ and HC.

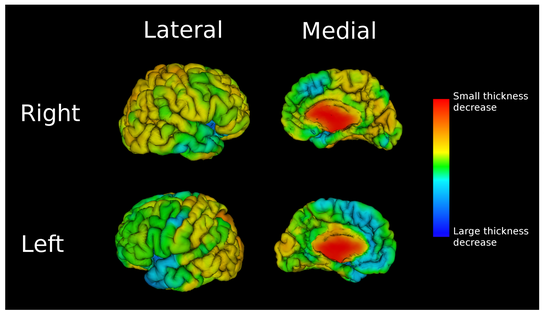

Figure 3: Decreases in cortical thickness represented by the first PC: from red to blue, smaller to larger thickness decreases. Affected structures include medial and superior temporal lobe; superior frontal, medial orbito-frontal lobes; insular cortex; left anterior cingulate cortex.

2. Naive Bayes (NB)

NB is a probabilistic classifier based on Bayes' theorem, which uses conditional probabilities to relate a class estimate to input evidence.
Sens. = 55%; Spec. = 57%; Acc. = 56%

3. Logistic Regression (LR)

LR is a discriminatory binary multivariate linear classifier.
Sens. = 76%; Spec. = 82%; Acc. = 80%

4. Support Vector Machine (SVM)

An SVM is a non-probabilistic linear classifier that uses the most difficult subjects to define the classification boundary.

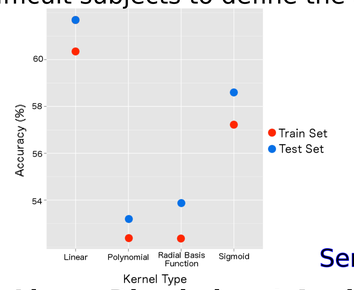

Figure 4: Kernels are used to represent non-linear relationships in data. The linear kernel outperformed all others.

Sens. = 62%; Spec. = 58%; Acc. = 60%

5. Linear Discriminant Analysis (LDA)

An LDA uses a linear combination of the input features to determine the output class. Sens. = 64%; Spec. = 59%; Acc. = 60%

6. Neural Networks (NNs)

NNs use a sytem of interconnected nodes and adaptive weights to perform classifications. Data is represented in an abstract "hidden layer" before being pushed through to the output layer.

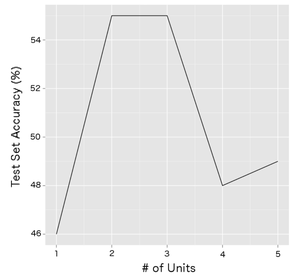

Figure 5: Effect of the number of units in the hidden layer on classification accuracy. The optimum is two or three hidden units.

Sens. = 53%; Spec. = 58%; Acc. = 55%

Conclusions

-> Logistic Regression is the most effective classification technique among the six tested for classifying schizophrenia patients from healthy controls based on cortical thickness
-> Further tuning of the algorithms is required to get accuracies that are competitive with the literature; also consider medication effects
-> Understanding the most effective classifier may offer insight into the biological underpinnings of schizophrenia

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)

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.** Narr, K.L., Bilder, R.M., Toga, A.W., Woods, R.P., Rex, D.E., Szeszko, P.R., ... Thompson, P.M. (2005). Mapping cortical thickness and gray matter concentration in first episode schizophrenia. *Cerebral Cortex (New York, N.Y.:1991), 15*(6), 708-19. doi:10.1093/cercor/bhh172; **3.** 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. **4.** Lerch, J.P, Pruessner, J.C., Zijdenbos, A., Hampel, H., Teipel, S.J., & Evans, A.C. (2005). Focal decline of cortical thicnkess in Alzheimer's disease identified by computational neuroanatomy. *Cerebral Cortex (New York, N.Y.: 1991), 15*(7), 995-1001, doi:10.1093/cercor/bhh200.