AS PREDICTED DRAFT

Exploring the Predictive Value of Neurocognitive Measures on Neuroanatomical Subtypes of Depression and Psychosis

Authors:

1. Gülce Lale (Graduate School of Systemic Neurosciences, LMU)
   - Guelce.Lale@med.uni-muenchen.de
2. Clara Sophie Vetter (Department for Precision Psychiatry, Hospital of the University of Munich)
   - Clara.Vetter@med.uni-muenchen.de

Created on: 04/24/2023

1) Have any data been collected for this study already?

- Yes. For this study, we will use the data from The Personalised Prognostic Indicators for early Psychosis management (PRONIA) study (https://www.pronia.eu/) which is a Collaboration Project funded by the European Union under the Seventh Framework Programme under Grant Agreement No. 602152. PRONIA study recruited participants in six European centers and one in Australia. For this study, we will include the data of participants who are at the state of Recent-Onset Depression, Recent-Onset Psychosis, and Clinical High-Risk for Psychosis.

2) What's the main question being asked, or hypothesis being tested in this study?

- In this study we intend to investigate the utility of the non-linear semi-supervised machine learning technique termed as HYDRA (heterogeneity through discriminative analysis) [1] in identifying subtypes of the disorder spectrum in the PRONIA dataset. Furthermore, we seek to employ neurocognitive data obtained from the PRONIA study to predict the subtype classification.

  Our main hypothesis is that the PRONIA dataset contains distinct neuroanatomical subtypes, similar to the subtypes of schizophrenia identified in previous studies that employed different datasets [2]. We also propose that by utilizing neurocognitive data, we can predict these subtypes.

  We will ask the following questions:
  I. Are there replicable neuroanatomical subtypes across patients with Recent-Onset Depression, Recent-Onset Psychosis, and Clinical High-Risk for Psychosis? Can these subtypes be identified in PRONIA dataset using a previously established machine learning model HYDRA that is trained on PHENOM subsample?

  II. How effective and accurate is using neurocognitive data to predict subtype membership revealed by HYDRA model across patients with Recent-Onset Depression, Recent-Onset Psychosis, and Clinical High-Risk for Psychosis?

3) Describe the key dependent variable(s) specifying how they will be measured.

- The dependent variable in this study will be a categorical variable with multiple levels, representing the different possible subtypes present in the data. To solve this multi-class classification problem, we will cluster disease effects by modelling differences from healthy controls using the HYDRA model. Additionally, we aim to predict subtype membership from the neurocognitive data, using the predicted subtype as the dependent variable for that analysis.

4) How many and which conditions will participants be assigned to?

- NA

5) Specify exactly which analyses you will conduct to examine the main question/hypothesis.

- The PHENOM data is processed using the MUSE pipeline for multi-site dataset ROI segmentation [3]. Therefore, a multi-site data analysis will be applied to the ROI data, as the PRONIA Consortium includes data from different locations around the world with different scanning designs. To remove the site effects, we chose to apply fastICA and neuroCombat (https://github.com/ncullen93/neuroCombat) [4] methods using R and Python. The effectiveness of the two methods will be assessed by ANOVA to determine if there are statistically significant differences in each of the mean ROI values between the PRONIA sites. Moreover, a visual comparison will be made between trying to predict the Age from ROI data for each site. Additionally, the impact of the choice of harmonisation method on subtyping results will be evaluated. After the site effect correction, a second harmonisation will be applied to adjust the PRONIA data to the training data that is used by the HYDRA model. Next, the ROI data will be adjusted for age and sex. Afterwards, the HYDRA model will be applied to the site calibrated PRONIA data to discover subtypes within the subjects. To evaluate the predictive value of these brain subtypes for neurocognitive measures, various machine learning algorithms such as Support Vector Machines (SVM) can be used. Furthermore, the accuracy of neurocognitive measures in predicting anatomical brain subtypes can be measured in terms of area under the receiver operating characteristic curve (ROC), balanced accuracy (BAC), sensitivity, and specificity.

6) Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

- Subjects will be excluded in case of missing neurocognitive data.
- Subjects who do not have the diagnosis type information in the dataset will be excluded.
- Only subjects between the age of 16 and 50 will be included in the analyses.

7) How many observations will be collected or what will determine sample size? No need to justify decision but be precise about exactly how the number will be determined.

- PRONIA dataset as mentioned in question 1 will be used to complete this study. The HYDRA model was trained using a repeated hold out cross-validation strategy such as for training, it applies 1000 repetitions for 80% of the data in each repetition.

8) Anything else you would like to pre-register? (e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

References:

1. Varol, E., Sotiras, A., Davatzikos, C., & Alzheimer's Disease Neuroimaging Initiative (2017). HYDRA: Revealing heterogeneity of imaging and genetic patterns through a multiple max-margin discriminative analysis framework. *NeuroImage*, *145*(Pt B), 346–364. https://doi.org/10.1016/j.neuroimage.2016.02.041
2. Chand, G. B., Dwyer, D. B., Erus, G., Sotiras, A., Varol, E., Srinivasan, D., Doshi, J., Pomponio, R., Pigoni, A., Dazzan, P., Kahn, R. S., Schnack, H. G., Zanetti, M. V., Meisenzahl, E., Busatto, G. F., Crespo-Facorro, B., Pantelis, C., Wood, S. J., Zhuo, C., Shinohara, R. T., … Davatzikos, C. (2020). Two distinct neuroanatomical subtypes of schizophrenia revealed using machine learning. *Brain : a journal of neurology*, *143*(3), 1027–1038. https://doi.org/10.1093/brain/awaa025
3. Doshi, J., Erus, G., Ou, Y., Resnick, S. M., Gur, R. C., Gur, R. E., Satterthwaite, T. D., Furth, S., Davatzikos, C., & Alzheimer's Neuroimaging Initiative (2016). MUSE: MUlti-atlas region Segmentation utilizing Ensembles of registration algorithms and parameters, and locally optimal atlas selection. *NeuroImage*, *127*, 186–195. https://doi.org/10.1016/j.neuroimage.2015.11.073
4. Fortin, J. P., Cullen, N., Sheline, Y. I., Taylor, W. D., Aselcioglu, I., Cook, P. A., Adams, P., Cooper, C., Fava, M., McGrath, P. J., McInnis, M., Phillips, M. L., Trivedi, M. H., Weissman, M. M., & Shinohara, R. T. (2018). Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage*, *167*, 104–120. https://doi.org/10.1016/J.NEUROIMAGE.2017.11.024