Schizophrenia Research 192 (2018) 167–171

Contents lists available at ScienceDirect

Schizophrenia Research

journal homepage: www.elsevier.com/locate/schres

Multisite generalizability of schizophrenia diagnosis classification based on functional brain connectivity

Pierre Orban [a,b,c,*,1], Christian Dansereau [a,d,1], Laurence Desbois [b], Violaine Mongeau-Pérusse [b], Charles-Édouard Giguère [b], Hien Nguyen [e], Adrianna Mendrek [b,f], Emmanuel Stip [b,c,g], Pierre Bellec [a,d]

[a] Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, Québec, Canada
[b] Centre de Recherche de l'Institut Universitaire en Santé Mentale de Montréal, Montréal, Québec, Canada
[c] Département de Psychiatrie, Université de Montréal, Montréal, Québec, Canada
[d] Département d'Informatique et de Recherche Opérationnelle, Université de Montréal, Montréal, Québec, Canada
[e] Department of Mathematics and Statistics, La Trobe University, Bundoora, Australia
[f] Department of Psychology, Bishop's University, Sherbrooke, Québec, Canada
[g] Centre Hospitalier Universitaire de Montréal, Montréal, Québec, Canada

ARTICLE INFO

Article history:
Received 10 January 2017
Received in revised form 23 May 2017
Accepted 24 May 2017
Available online 24 August 2017

Keywords:
Schizophrenia
fMRI
Machine learning
Classification
Multisite
Cognition

ABSTRACT

Our objective was to assess the generalizability, across sites and cognitive contexts, of schizophrenia classification based on functional brain connectivity. We tested different training-test scenarios combining fMRI data from 191 schizophrenia patients and 191 matched healthy controls obtained at 6 scanning sites and under different task conditions. Diagnosis classification accuracy generalized well to a novel site and cognitive context provided data from multiple sites were used for classifier training. By contrast, lower classification accuracy was achieved when data from a single distinct site was used for training. These findings indicate that it is beneficial to use multisite data to train fMRI-based classifiers intended for large-scale use in the clinical realm.

1. Introduction

Psychiatrists and other mental health professionals could benefit in the not-so-far future from neuroimaging-based classification tools to assist diagnosis and prognosis in mental illness (Huys et al., 2016). Recent developments in the neuroimaging field have led to a shift from group comparisons based on averaging across subjects to machine learning techniques making prediction at the individual level (Dubois and Adolphs, 2016). In this approach, the emphasis is put on the ability of an algorithm to classify individuals into clinical categories with good generalizability to unseen subjects. Over the last decade, hundreds of studies have successfully classified various psychiatric and neurological disorders based on in vivo brain imaging (reviewed in Arbabshirani et al., 2017; Wolfers et al., 2015). For instance, Arbabshirani et al. (2017) identified 30 published studies that distinguished schizophrenia patients from healthy controls with an average accuracy of 83% using functional magnetic resonance imaging (fMRI), either under task or rest states.

To date, however, the vast majority of classification works in mental illness were performed in a research context, using data from single sites of acquisition. Such findings may not generalize to large-scale clinical settings, with patients being scanned at widely-spread sites and possibly under various mental states. In most cases, the performance of classifiers was only assessed for unseen, test subjects with the exact same characteristics as the sample used for training. Yet, using gender as a proof-of-concept target variable, there was initial evidence that classifiers only poorly generalize to data drawn from other site samples (Huf et al., 2014). The inclusion of data from multiple sites during training improved the classifier performance for data of unseen sites.

In schizophrenia, a study pooling fMRI data from two distinct scanning sites reported similar prediction accuracy levels irrespective of whether test data were drawn from the dataset used for training or not, thus suggesting good generalizability (Skåtun et al., 2016). However, this result appears at odds with a recent fMRI study in autism that showed poorer accuracy for inter-site than intra-site training/test configurations, depending on the ratio of training set used (Abraham et al., 2017). In the case of inter-site testing, data pooled from 4 sites were used for training the classifier, which was tested on data from a fifth site. Yet, none of these two studies specifically evaluated whether using multisite training data could compensate to some extent for the

* Corresponding author at: Centre de Recherche de l'Institut Universitaire en Santé Mentale de Montréal, 7331 Hochelaga, Montréal H1N 3V2, QC, Canada.
E-mail address: pierre.orban@umontreal.ca (P. Orban).
[1] Equally contributed to this work.

http://dx.doi.org/10.1016/j.schres.2017.05.027
0920-9964/

deleterious effect of inter-site testing, by assuming the actual presence of such an effect. In the present work, we sought to address this question based on fMRI brain connectivity in schizophrenia. Since it is impossible to completely control the variations in mental states in realistic clinical situations, we further promoted the complexity of the classification problem by including data obtained in distinct cognitive task conditions across sites. Mass univariate findings have indicated that cognitive state does not further impact on the nature of functional brain connectivity alterations in schizophrenia (Kaufmann et al., 2017; Orban et al., 2017). However, the potential influence of cognitive context on classification performance in a multivariate analysis should not be rejected.

2. Methods

2.1. Datasets

Brain imaging data from 6 independent studies were obtained through either the SchizConnect and OpenfMRI data sharing platforms (http://schizconnect.org; https://openfmri.org) or local scanning (Çetin et al., 2014; Gollub et al., 2013; Kogan et al., 2016; Orban et al., 2017; Poldrack et al., 2016; Wang et al., 2016). The 6 datasets differed in terms of both scanning site and cognitive context during fMRI data acquisition (resting-state, emotional memory, Sternberg item recognition paradigm, N-back, task-switching and oddball tasks). Classification analyses included fMRI data from 382 subjects, 191 patients diagnosed with schizophrenia and 191 healthy controls. Subjects provided informed consent to participate in their respective studies and ethics approval was obtained at the site of secondary analysis (Centre de Recherche de l'Institut Universitaire de Gériatrie de Montréal, Montréal, Canada).

2.2. Subjects matching

Sample size differed between sites (N = 84, 82, 70, 62, 50 and 34). Site samples were obtained after subjects were selected in order to ensure even proportions of schizophrenia patients and controls within each site (N = 42, 41, 35, 31, 25 and 17 subjects per group) and to reduce between-group differences with regards to gender ratio (75% vs 73% males in controls vs. schizophrenia patients), age distribution (32.3 ± 9.8 vs. 33.4 ± 9.5 years old) and motion levels (average frame displacement = 0.15 ± 0.05 vs 0.17 ± 0.06, see Data preprocessing). Matching of schizophrenia and control subjects was achieved based on propensity scores, using the Optmatch R library version 0.9-7 (https://cran.r-project.org/web/packages/optmatch/index.html). The propensity score associated with each participant was defined by the conditional probability of being in the clinical or control group given the confounding covariates (gender, age and motion). Propensity scores were then used to balance those covariates in the two groups. Although we took great care in matching participants with respect to these factors of no interest, it is very likely that other confounds such as medication in schizophrenia patients impacted the reported findings.

2.3. Data preprocessing

Brain imaging data preprocessing and extraction of functional brain connectomes were performed with the NeuroImaging Analysis Kit version 0.12.17 (NIAK, http://niak.simexp-lab.org). Briefly, preprocessing included slice timing correction, estimation of rigid-body motion within the functional runs, nonlinear coregistration of the structural scan in stereotaxic space, individual coregistration between structural and functional scans, resampling of the functional scans at 3 mm isotropic resolution in stereotaxic space, scrubbing of volumes with excessive motion (frame displacement greater > 0.5 mm), regression of confounds (slow time drifts, average of conservative white matter and cerebrospinal fluid masks and motion parameters), and smoothing of functional volumes with a 6 mm isotropic Gaussian blurring kernel. A detailed description of the preprocessing pipeline can be found at http://niak.simexp-lab.org/pipe_preprocessing.html.

Individual functional connectomes included 2016 functional connections between 64 brain parcels. The functional brain parcellation was previously obtained by conducting a bootstrap analysis of stable clusters (BASC, Bellec et al., 2010) on an independent fMRI dataset of 200 healthy young subjects (https://doi.org/10.6084/m9.figshare.1285615.v1). In each schizophrenia or control participant, the time series of a brain parcel consisted in the average of the voxel signals in the parcel. Connectivity measures between pairs of parcels were defined by Pearson product-moment correlation coefficients. Individual connectomes were parcel by parcel (64 × 64) symmetrical matrices that summarized connectivity levels in the whole brain. Lower triangular matrices were then vectorized for all subjects in order to form a subject by connections (382 × 2016) matrix.

2.4. Data analysis

Classification analyses were performed with a linear support vector machine (SVM) algorithm, as implemented in the SciKit-Learn python library version 0.18.1 (Abraham et al., 2014). The SVM classifier, a supervised classification algorithm, represented subjects as points in space, mapped so that the subjects of the separate clinical labels were divided by a clear gap (called a margin) that was as wide as possible. The hyperparameter C of the SVM was optimized using nested cross-validation. Each model used the residuals from a regression of confounding variables (gender, age and motion parameters) across connections estimated from the subjects selected for training the model. The evaluation metrics were computed using four main values, namely the number of true and false positive (TP, FP) as well as true and false negatives (TN, FN). Sensitivity was defined as TP / (TP + FN), specificity as TN / (TN + FP) and accuracy as (TP + TN) / (TP + FP + TP + FN). The main analyses evaluated the impact on classification accuracy of the number of site(s) (1, 2, 3, 4 or 5) included in the training set. We evaluated this impact in situations where the test set included only subjects from the same site(s) used during training (intra-site test with 10-fold cross validation) or, alternatively, situations where the test set included only subjects from sites not used during training (inter-site test with "leave-site-out" cross validation). Cross validation ensured that the subjects used for training were never used in the test phase.

The statistical significance of changes in accuracy levels as a function of the number of sites used for training and whether data used for testing were drawn from the same dataset(s) used for training (intra-site vs inter-site) was assessed with binary logistic regressions using the GLM function in R version 3.2.5. These analyses relied on the prediction of categorical outcomes (hit/miss data) based on predictor variables (number of sites used for training, intra-site vs inter-site). Significance threshold in the different contrasts was set at $p < 0.05$.

Complementary analyses were conducted. First, we explored differences in whole brain connectivity between schizophrenia patients and controls using mass univariate statistics for the various training site combinations. Similarly for multivariate classification analyses, we extracted feature weights separately for all site combinations. We then examined the level of correspondence across site combinations for both univariate and multivariate analyses. Second, we aimed at demonstrating the presence of multivariate site effects on functional brain connectivity. To this end, we determined accuracy levels for the classification of scanning sites by performing separate SVM analyses for all pairs of sites, using 10-fold cross validation as in the main analyses.

3. Results

3.1. Correspondence across site combinations

We first report patterns of functional brain dysconnectivity in schizophrenia patients based on mass univariate statistics. For the